Research Paper

System level characterization of small molecule drugs and their affected long noncoding RNAs

Haixiu Yang[1,*], Yanan Jiang[2,3,*], Yunpeng Zhang[1], Yanjun Xu[1], Chunlong Zhang[1], Junwei Han[1], Fei Su[1], Xiaoqi Liu[4], Kai Mi[1], Bing Liu[2,3], Desi Shang[1]

[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China
[2]Translational Medicine Research and Cooperation Center of Northern China, Heilongjiang Academy of Medical Sciences, Harbin 150081, Heilongjiang, China
[3]Department of Pharmacology (State-Province Key Laboratories of Biomedicine-Pharmaceutics of China, Key Laboratory of Cardiovascular Research, Ministry of Education), College of Pharmacy, Harbin Medical University, Harbin 150081, Heilongjiang, China
[4]Department of Orthopedic Surgery, The Second Affiliated Hospital, Harbin Medical University, Harbin 150001, China
*Equal contribution

Correspondence to: Desi Shang, Bing Liu; **email:** sds_46@163.com, liubingdoctor123@126.com
Keywords: small molecule drugs, long noncoding RNAs, network, pharmacological analysis, tissue-specificity
Received: July 8, 2019 **Accepted:** November 26, 2019 **Published:** December 18, 2019

ABSTRACT

Long noncoding RNAs (lncRNAs) have multiple regulatory roles and are involved in many human diseases. A potential therapeutic strategy based on targeting lncRNAs was recently developed. To gain insight into the global relationship between small molecule drugs and their affected lncRNAs, we constructed a small molecule lncRNA network consisting of 1206 nodes (1033 drugs and 173 lncRNAs) and 4770 drug-lncRNA associations using LNCmap, which reannotated the microarray data from the Connectivity Map (CMap) database. Based on network biology, we found that the connected drug pairs tended to share the same targets, indications, and side effects. In addition, the connected drug pairs tended to have a similar structure. By inferring the functions of lncRNAs through their co-expressing mRNAs, we found that lncRNA functions related to the modular interface were associated with the mode of action or side effects of the corresponding connected drugs, suggesting that lncRNAs may directly/indirectly participate in specific biological processes after drug administration. Finally, we investigated the tissue-specificity of drug-affected lncRNAs and found that some kinds of drugs tended to have a broader influence (e.g. antineoplastic and immunomodulating drugs), whereas some tissue-specific lncRNAs (nervous system) tended to be affected by multiple types of drugs.

INTRODUCTION

Increasing research has shown that long noncoding RNAs (lncRNAs) are pivotal regulators in many biological processes, as well as in the generation and progression of various diseases [1–3]. Many lncRNAs are aberrantly expressed in different pathological states [4], and the restoration of lncRNA expression has dramatic therapeutic potential on diseases, making it a novel therapeutic strategy [5].

Currently, pharmacotherapy is the most effective strategy in the treatment of some diseases. Small molecules are widely used due to their simple structure. In recent years, small molecules are found

to either have therapeutic effects or induce side-effects/toxicity through the regulation of lncRNAs [6, 7]. Meanwhile, the differently expressed lncRNAs are also indicators of predicting drug sensitivity and resistance, especially in treatment of cancer [8, 9]. Therefore, a lncRNA-based therapeutic strategy hopefully could make personalized medicine become more realistic. However, structure-based approach such as molecular docking has been difficult to achieve as predicting the exact structure of lncRNAs remains a challenge. Another approach based on the transcriptional response might be appropriate to investigate the global relationships between small molecules (drugs) and lncRNAs.

Enormous array-based expression profile resources have promoted the development of research on small molecules. For example, Connectivity Map (CMap), which is a genome-wide transcriptional expression dataset of selected human cells treated with bioactive small molecules, pioneered the systemic transcriptional response-based approach [10]. Compared to the array-based expression profiles, there are not enough publicly available RNA-seq datasets of drug treatments yet, and this has been a limitation to the lncRNA-targeting therapy [11]. This limitation could be alleviated by a novel method called probe re-annotation in which microarray probes are reannotated for investigating lncRNA expression [12]. This approach has been successfully used for the functional annotation of lncRNAs in various studies [11]. We have reannotated the microarray data from the CMap database in our previous work named LNCmap, which successfully characterized the connections among diseases, lncRNAs and small molecules [13].

To investigate the global relationships between small molecules (drugs) and lncRNAs, we obtained the lncRNA expression profiles affected by small molecules from the LNCmap. We constructed a global small molecule lncRNA network (SMLN) in which nodes represented drugs or lncRNAs. Starting from the bipartite SMLN, we generated two biologically relevant network projections: the small molecule-small molecule network (SSN), in which nodes represented drugs, and two drugs were connected if they shared at least one lncRNA; and the lncRNA-lncRNA network (LLN), in which nodes were lncRNAs and two lncRNAs were connected if they significantly shared small molecules. Then we (i) investigated the pharmacological similarities of linked small molecule pairs in the SSN, (ii) explored the function of modules of the LLN in the responses to drug treatment, and (iii) analyzed the tissue specificity of lncRNAs after drug treatment.

RESULTS

Construction of a small molecule lncRNA network (SMLN)

To construct the global SMLN, we retrieved lncRNA expression profiles affected by small molecules from LNCmap [13]. LNCmap reannotated the genome-wide transcriptional expression data from the CMap database, which contains 1,309 bioactive small molecules corresponding to 6100 instances (experiments). Among them, 5,916 microarray profiles were reannotated in LNCmap, including 237 lncRNA signatures and 1,262 small molecule drugs. We then used fold-change analysis to identify differentially expressed lncRNAs (DEL) for every instance with |log2fold change|>1 from the corresponding treatment and control microarray profiles. The DELs were merged if the corresponding experiments belonged to the same small molecules, and these lncRNAs were considered the affected lncRNAs for this small molecule (Figure 1). Then, we obtained a bipartite SMLN consisting of 1,005 small molecules and 173 lncRNAs (Figure 2 and Supplementary Dataset 1). We generated two biologically relevant network projections: the SSN and the LLN. If most small molecules specifically affected a single lncRNA, the LLN would consist of isolated nodes with few or no edges between them. Instead, the LLN displayed close connections between different lncRNAs. One reason for this phenomenon might be that there were over 1,000 small molecules compared with only 173 lncRNAs in the SMLN. The number of shared small molecules between different lncRNA pairs spanned a wide range. To improve the specificity of lncRNA pairs, we adopted a hypergeometric test to generate a more specific LLN (see Materials and Methods).

The basic properties of the SMLN

The SMLN was composed of 1,206 nodes (1,033 small molecules and 173 lncRNAs) and 4,770 edges (2,628 downregulated and 2,142 upregulated) (Figure 2). All nodes were in one giant component, suggesting that the small molecules and lncRNAs were closely connected in the SMLN. The degree of distribution of the small molecule and lncRNA nodes followed power law distributions with a slope of -0.947 and -0.850, respectively, and R^2=0.874 and 0.532, respectively (Supplementary Figure 1A , 1B and Supplementary Dataset 2). Thus, the SMLN was scale-free [14].

The degree of small molecule nodes spanned a wide range from 1 to 87. The highest degree node was trichostatin A (TSA), an organic compound that serves as an antifungal antibiotic and selectively inhibits class I and II mammalian histone deacetylases (HDACs) [15].

TSA can broadly alter gene expression by interfering with the removal of acetyl groups from histones [16, 17]. It is also a member of a larger class of histone deacetylase inhibitors that have a broad spectrum of epigenetic activities [16, 17]. The second highest degree small molecule node (degree=46) was emetine, an anti-malaria drug that was recently found to have broad anticancer activity in many types of malignancies including breast, colon, prostate, skin, and lymphoid tumors by inhibiting NF-κB signaling or regulating the RNA splicing of members of the Bcl-2 family [18, 19]. Although there are no specific reports about emetine and lncRNAs, it was linked to many lncRNAs, partly because of its broad anticancer effects. Interestingly, we found that other highly connected nodes, namely anisomycin and idoxuridine (degree: 39 and 38, respectively) could inhibit protein/DNA synthesis. Anisomycin is a potent apoptosis inducer that functions

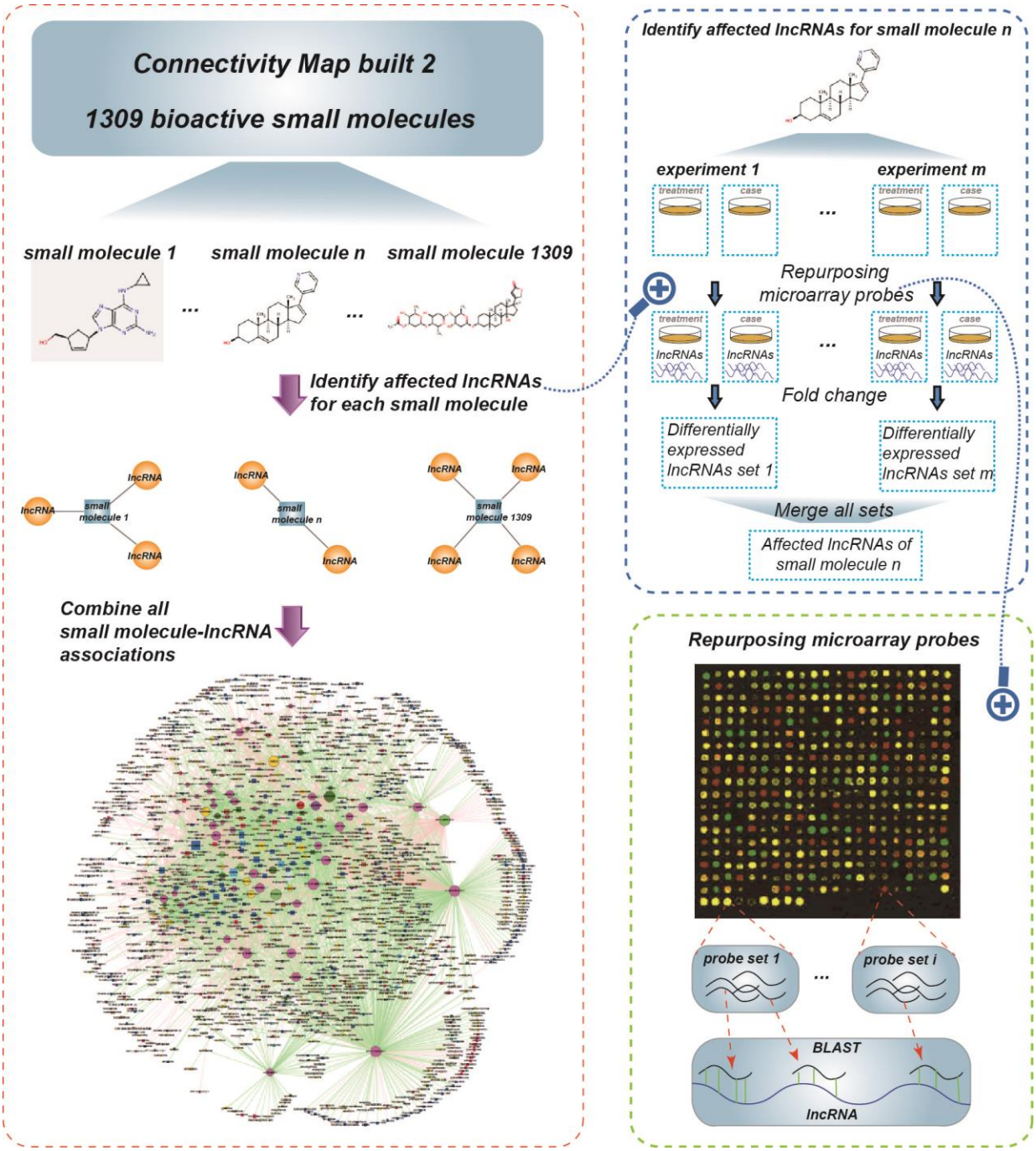


Figure 1. Schematic data flowchart of SMLN.

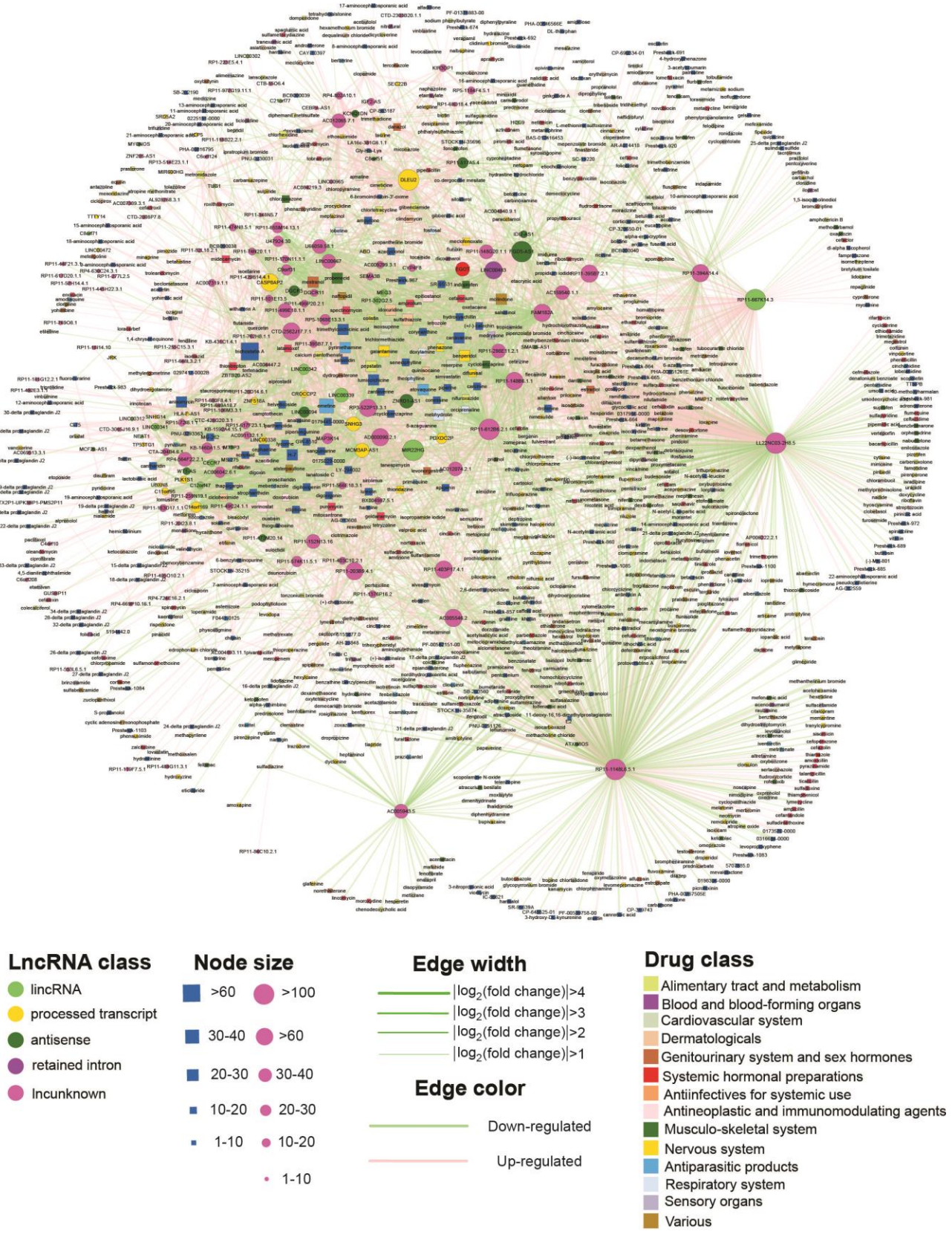


Figure 2. The SMLN network. The rectangles and circles in the network correspond to small molecules and lncRNAs, respectively. A small molecule and a lncRNA are connected by an edge if the lncRNA differentially expressed when treated with this small molecule. Colors represent different lncRNA and small molecule classes.

by activating JNK/SAPK and inhibiting protein/DNA synthesis during translation [20, 21]. Idoxuridine, which is used as an antiviral agent, is an analog of deoxyuridine, an inhibitor of viral DNA synthesis [22]. The high connectivity may have been due to their activity related to apoptosis and the inhibition of protein/DNA synthesis.

Similar to the small molecule nodes, the lncRNA nodes also displayed evident differences in connection (range, 1–366). The lncRNA node with the highest degree was RP11-1148L6.5.1. There are no functional studies about this lncRNA. To date, few lncRNAs have been functionally annotated. Of seven lncRNAs with a degree >100, only *DLEU2* (Deleted in Lymphocytic leukemia 2) is well studied. It encodes a pair of critical pro-apoptotic microRNAs, miR-15a/16-1, which are critical for the increased survival exhibited by chronic lymphocytic leukemia cells [23]. Chen et al. indicated that the HDAC inhibitor TSA, the most-connected small molecule in the SMLN, could upregulate the expression of miR-15a/16-1, residing in the host tumor suppressor *DLEU2* gene [24]. Furthermore, in our SMLN, TSA could also upregulate *DLEU2* (log2 fold change = 1.4), suggesting that our SMLN could identify a promising cancer therapy via targeting lncRNAs [23]. In addition, we showed that the fold change value varied within a wide range (Figure 3A). All fold-change values of LINC00667 were positive, indicating that the expression of this lncRNA is always upregulated in response to drug treatment (Figure 3B). The function of this lncRNA has not been well-studied. Thus, pathway enrichment analysis was used to examine the function of LINC00667. The results showed that it was related to purine metabolism, which shows clear relevance to various cancers such as bladder cancer, kidney cancer and prostate cancer (Figure 3C) [25].

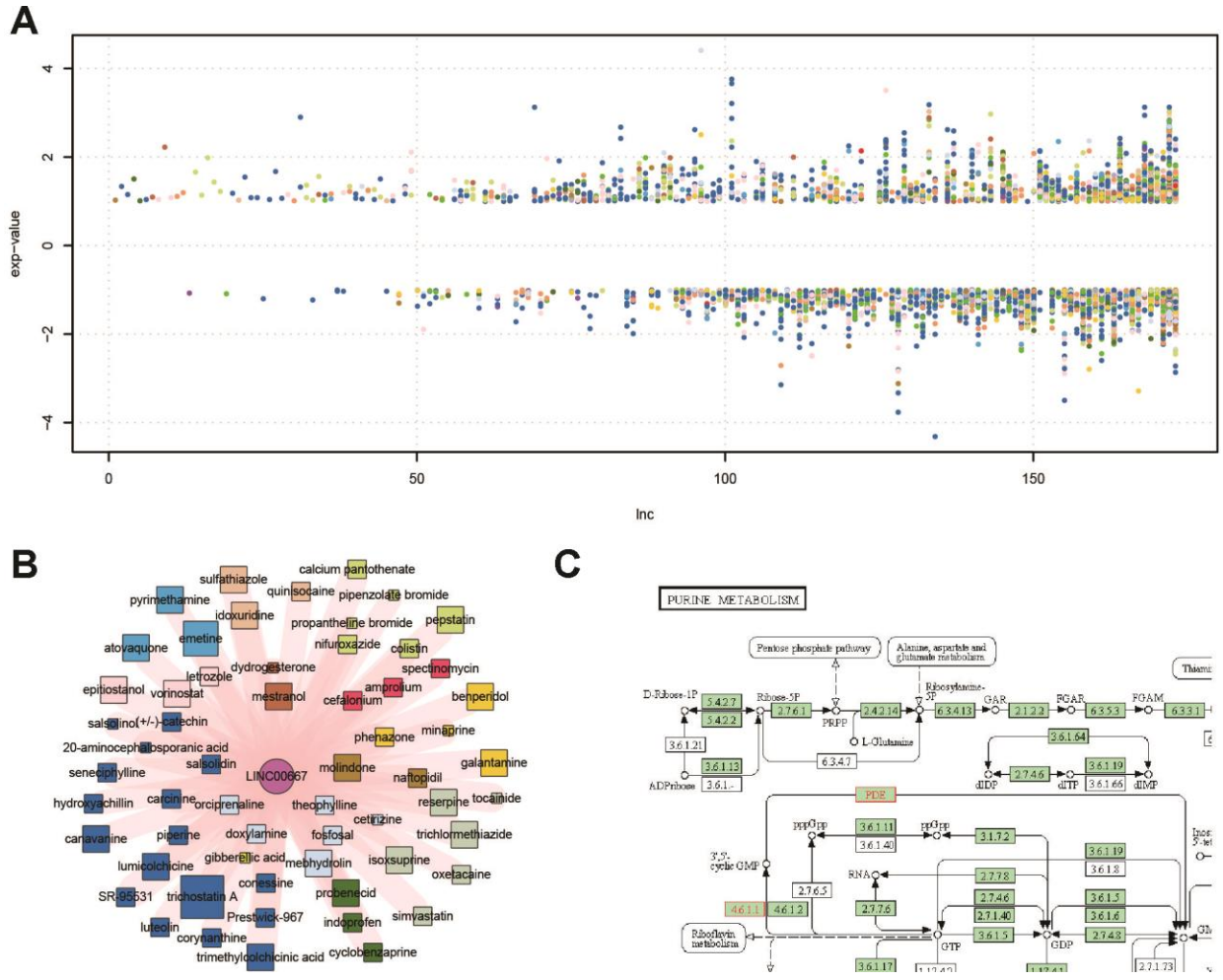


Figure 3. LncRNA expression values and functional characteristics. (**A**) Fold change value of lncRNAs affected by drugs; colors represent different ATC codes of drugs affected by the specific lncRNA. (**B**) Sub-network of LINC00667 and the related drugs: LINC00667 was always up-regulated after drug treatment. (**C**) Functional characteristics of LINC00667 by pathway enrichment with its co-expressed protein-coding genes.

Pharmacological similarity of small molecule pairs

In our SMLN, the connections between small molecules and lncRNAs revealed the non-coding transcriptional responses after drug treatment. Studies have indicated that drugs leading to similar transcriptional responses tend to have similar pharmacological properties [10, 26]. Thus, we first constructed a SSN in which nodes represent drugs, and two small molecules are connected if they share at least one lncRNA (Supplementary Figure 2 and Supplementary Dataset 3). Then, we investigated whether the connected small molecules (drugs) tend to have similar pharmacological properties.

We examined a total of four properties of connected drug pairs in SSN: indications, targets, side effects, and 2-D structural similarity (Figure 4). Firstly, we investigated whether connected drug pairs tend to share the same indications (treat the same disease). For this purpose, we first downloaded the dataset of drug–indication association from the report published by Yildirim et al. (see Materials and Methods) [27]. Of these connected drug pairs, 417 shared the same indications. We then generated randomized drug pairs 1,000 times. We found that in the 1,000 random times, the number of randomized drug pairs sharing the same indications were <417, suggesting that connected pairs tended to share the same indications ($P = 0$) (Figure 4A, left). Some drugs such as acetohexamide and gliclazide were connected to the same lncRNAs, and they were all used for the treatment of diabetes (Figure 4A, right). Based on this result, we questioned whether connected drug pairs tend to share the same targets. We found that some connected drugs such as minaprine and thioridazine both target serotonin receptor 2A (HTR2A), a protein associated with the susceptibility to schizophrenia and obsessive-compulsive disorder (Figure 4B, right) [28]. To further test this, we downloaded the drug-target association from the DrugBank database [29]. In the SMLN, 1,066 unique connected drug pairs targeted the same proteins. Similar to the aforementioned indications, we generated randomized drug pairs 1,000 times and there were no instances in which the number of randomized drug pairs sharing the same targets were more than 1,066, suggesting that connected drug pairs tended to share the same targets (Figure 4B, left). Then, we downloaded the public and accurate side-effect records from the side effect resource (SIDER), including 997 drugs corresponding to 4,492 side effects [30]. In the SMLN, there were 303 drugs recorded in the SIDER database. In the SIDER database, however, some side effects, such as dizziness and nausea, were caused by most drugs. To improve the specificity of similarity, we calculated the number of side effects shared by drug pairs rather than the number of drug pairs sharing the same side effects [26]. We found that the ratio of side effects shared by connected drug pairs was significantly higher than the number of side effects shared by other drug pairs in the SIDER database ($P = 2.2e^{-16}$, Wilcoxon rank-sum test; Figure 4C, left), suggesting that two drugs connected by the same lncRNAs tended to cause the same side effects. Unlike the indications and targets, some connected drug pairs (atovaquone and galantamine), despite belonging to different categories, could cause the same side effects (Figure 4C, right), indicating that the non-coding transcriptional response might also capture such "heterogeneous" similarity.

Previously, a study indicated that pharmacologically similar drugs tend to have a similar structure [27]. Thus, we tested whether lncRNA-connected drugs tend to have a similar structure. For this purpose, we downloaded the SMILES files of small molecules in SSN and calculated the Tanimoto Coefficients (TC) of connected drug pairs and other small molecule pairs. We found that the TC scores of connected drug pairs were significantly higher than those of other small molecule pairs ($P = 6.2e^{-5}$), Wilcoxon rank-sum test; Figure 4D, left), suggesting that the connected drugs tend to have a similar structure. Some connected small molecules pairs showing high structural similarity are shown in the right part of Figure 4D.

Functional interface of drug-induced lncRNA modules

Similar to the SSN, we generated another biologically relevant network projection, namely the LLN, in which two lncRNAs are connected if they share significant numbers of small molecules (Supplementary Figure 3 and Supplementary Dataset 4). The connected lncRNA pairs affected by the same small molecules in the LLN might tend to have similar functions. Thus, the lncRNAs in one community of the LLN were considered to function "synergistically" because they were affected by the same or similar small molecules. We further proposed that lncRNAs, as members of more than one community, were more important and may be involved in key pathways related to therapeutic effects or the indication of corresponding drugs, because lncRNAs in multiple communities could be considered to be at the "interface" of different biological processes. Here, we investigated the functions of such interface lncRNAs and the relations to their connected drugs.

First, we inferred the functions of lncRNAs through their co-expressing mRNAs across all the re-annotated microarrays in the CMap according to the "Guilt By Association" principle [12]. We then defined a lncRNA-lncRNA module as a clique, which is a maximal complete subgraph using Cfinder [31]. Each module

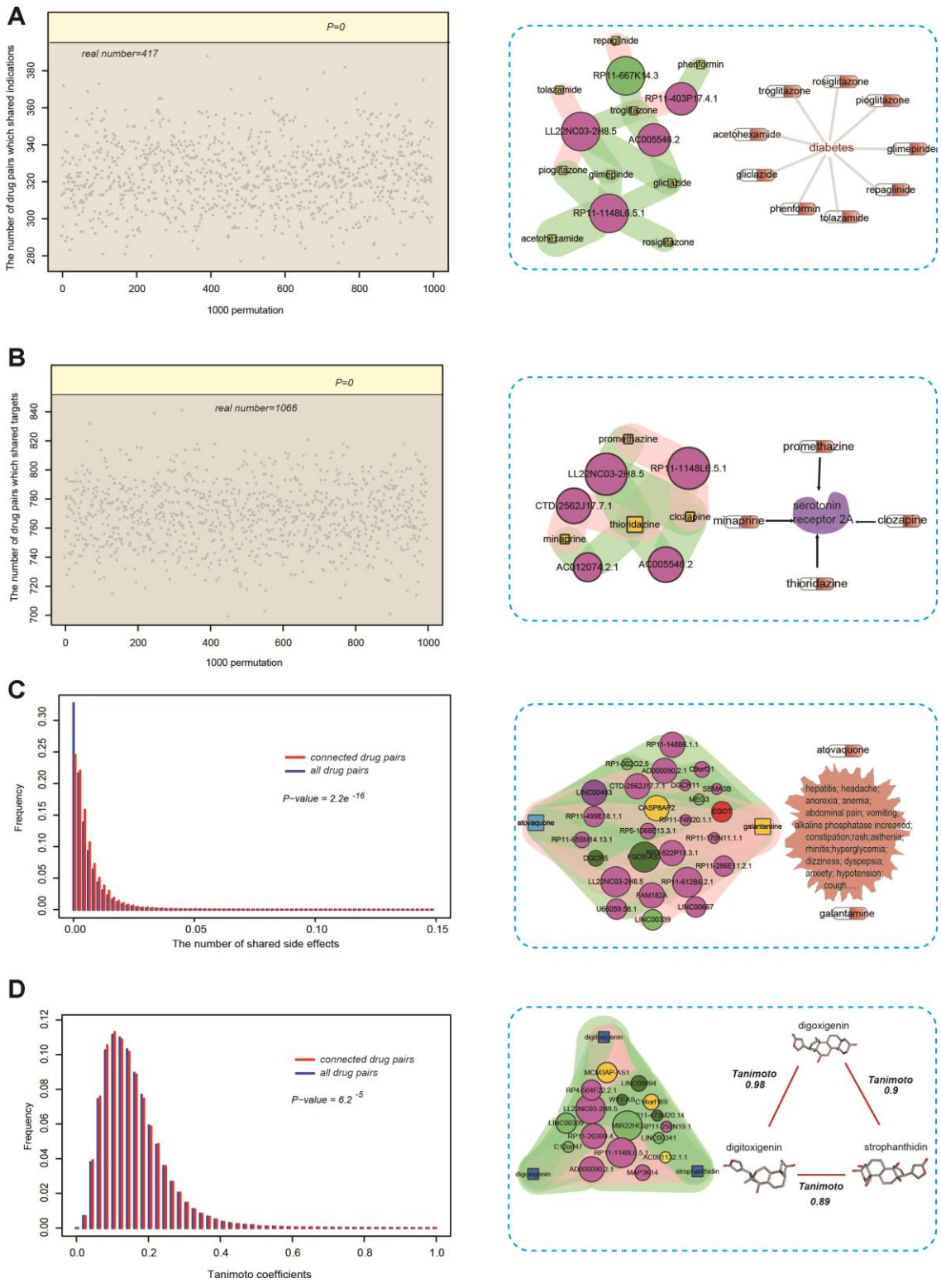


Figure 4. Pharmacological properties of connected drug pairs in the SSN. (**A**, left) 417 drug pairs with the same lncRNAs shared the same indications, compared with 1000 permutations. (**A**, right) Acetohexamide and gliclazide were connected to the same lncRNAs and they were all used for the treatment of diabetes. (**B**, left) 1066 drug pairs with the same lncRNAs shared the same drug targets, compared with 1000 permutations. (**B**, right) Minaprine and thioridazine shared the same lncRNA and both target the serotonin receptor 2A (HTR2A). (**C**, left) The proportion of shared side effects by drug pairs with the same lncRNAs (red), compared with the proportion of shared side effects among the total drug pairs in the SIDER database (blue). (**C**, right) Atovaquone and galantamine shared the same lncRNAs, although they belong to different categories, and could cause many of the same side effects. (**D**) Drug pairs with the same lncRNAs had higher TC scores.

had a unique composition of lncRNAs, allowing the same lncRNAs or the same pairs to occur in more than one module. Here, we adopted K=8, 9, and 10, because a low k-value generated numbers of extensive and less tightly connected lncRNA modules, leading to a high degree of overlap [14]. We displayed the modules of the LLN and the interface between them when the k-value was 8, 9, and 10 (the left part of Figure 5). The lncRNAs are colored according to the majority of their connected drug ATC classes. The right part of Figure 5 shows the significantly enriched pathways of co-expressing mRNAs of the "interface" lncRNAs. When the k-value was 8, there were two communities and 5 lncRNAs located at the interface of two modules (Figure 5A). The interface lncRNAs were mainly affected by drugs of "Alimentary tract and metabolism", and these lncRNAs were involved in key pathways such as the calcium signaling pathway, the PI3K-Akt signaling pathway, and the GABAergic synapse. We found that these pathways (such as "Gastric acid secretion" and "Pancreatic secretion") were not only related to "Alimentary tract and metabolism" drugs, but also related to the other two kinds of drugs ("Respiratory system" and "Cardiovascular system"). For example, cardiovascular disease pathways such as "Dilated cardiomyopathy" and "Hypertrophic cardiomyopathy" were identified. Furthermore, some key pathways such as the "PI3K-Akt signaling pathway" have comprehensive biological roles and are involved in the indications of these kinds of drugs [32, 33].

When the k-value was 9, two communities had three interface lncRNAs. We also found that these interface lncRNAs were involved in key pathways related to the drugs of "Alimentary tract and metabolism", "Systemic hormonal preparations", and "Nervous system"(Figure 5B). For example, the pathway "Cytokine-cytokine receptor interaction" plays a crucial role as intercellular regulator and mobilizer of cells engaged in innate as well as adaptive inflammatory host defenses, cell growth, differentiation, cell death, angiogenesis, and development and repair processes [34], and it is also related to many indications of the above three kinds of drugs. Another pathway, "Drug metabolism-other enzymes", was also detected. We found that this pathway was not only related to drug metabolism, but also contained enzymes encoded by genes related to metabolic diseases (i.e. xanthinuria and hyperbilirubinemia). When the k-value was 10, there were five interface lncRNAs that were all affected by the drug TSA. We found that the functions of these lncRNAs focused on the nervous system, and TSA had some side effects affecting the nervous system (Figure 5C). These results indicated that these interface lncRNAs may be directly/indirectly affected by drugs and broadly participated in the process of drug therapeutics, side effects, and metabolism.

Tissue-specificity of drug-affected lncRNAs

Previous studies indicated that lncRNA expression may be limited to selected tissues or subpopulations of cells [35, 36]. The exclusively tissue-specific expression pattern of lncRNAs provides a unique opportunity for specific regulation by lncRNA-targeting therapeutics [35, 36]. In our study, lncRNAs in the SMLN were induced by different drugs, which were classified according to ATC (Anatomical Therapeutic Chemical Classification System) categories. This raises the issue of whether drugs tend to affect the lncRNAs expressed in their corresponding target tissues.

To obtain a global overview of the anatomical distribution of the lncRNAs affected by different drug classes, we first reannotated the microarray dataset of GSE1133 obtained for 176 lncRNAs across 79 healthy tissues, and identified tissue-specific lncRNAs for each tissue (Supplementary Table 1) [37]. Then, we classified these 79 tissues into 11 anatomical classes according to ATC categories. We calculated the Jaccard coefficient of numbers of overlapped lncRNAs between 13 drug classes and tissues of 11 anatomical classes (Figure 6A). We found that drugs from all drug classes do not affect lncRNAs expressed in the corresponding anatomical classifications. Instead, some drugs tended to affect broadly tissue-specific lncRNAs. For example, anti-infectives (J) and antineoplastic and immunomodulating (L) drugs were more spread than other kinds of drugs. This might be due to the systemic therapeutic use or side effects of these two kinds of drugs. Certain types of drugs and tissues had obviously higher Jaccard coefficients [for example, (L) drugs and nervous system; (C) drugs (cardiovascular system) and nervous system]. On the other hand, we found that some tissue-specific lncRNAs (for example, nervous system specific lncRNAs) were prone to be affected by many kinds of drugs. The highest Jaccard coefficient was found between alimentary tract and metabolism drugs and the nervous system, suggesting that this kind of drug might predominantly affect nervous system-specific lncRNAs. To ensure the robustness of our results, we also calculated the Jaccard coefficients of overlapped lncRNAs between 13 drug classes and 16 tissues by processing the RNA-seq data of 16 normal human individual tissues from the ArrayExpress database (ERP000546) [38] (see Materials and Methods). Although the tissues were different from those of GSE1133, some similar results were observed (Figure 6B). For example, drugs of (J) and (L) codes were more spread; furthermore, higher Jaccard coefficients were also observed in (L) drugs and the nervous system (brain).

To better investigate the mechanism of tissue-specificity of lncRNAs induced by drugs, we extracted the sub-

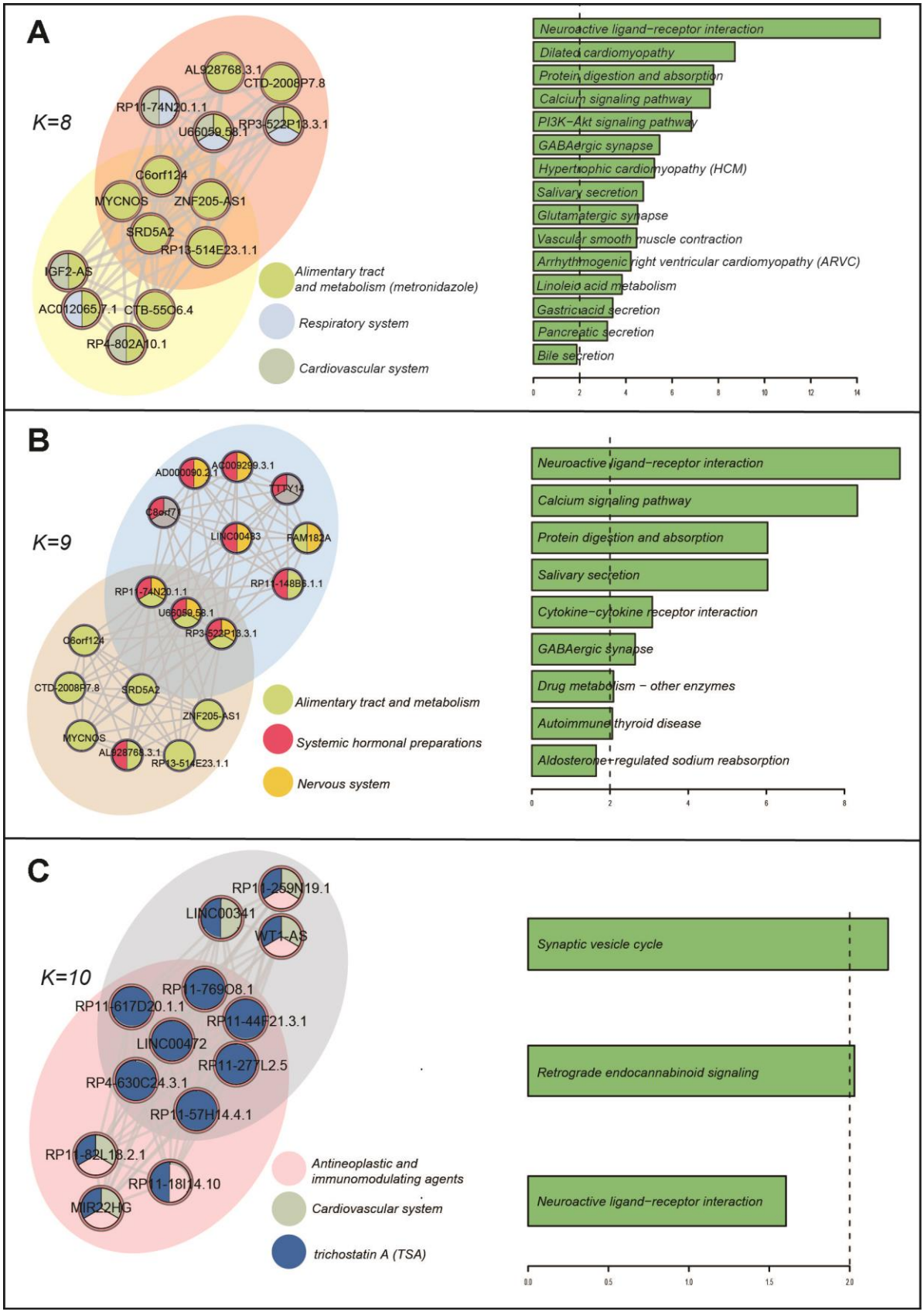


Figure 5. Drug-induced lncRNA modules and enriched KEGG pathways. (**A**) k=8. (**B**) k=9. (**C**) k=10. The K represents the number of the nodes in the modules.

network from the SMLN of drugs belonging to the (L) code and their affected lncRNAs (Figure 6C). Many lncRNAs were expressed in the nervous system of GSE1133 or in the brain of ERP000546. Some of the LncRNAs were related to cancers and expressed in the brain. For example, DLEU2 (black dotted circle), a potential therapeutic target of chronic lymphocytic leukemia [23], was expressed in the brain and associated with axon degeneration in the brain [39].

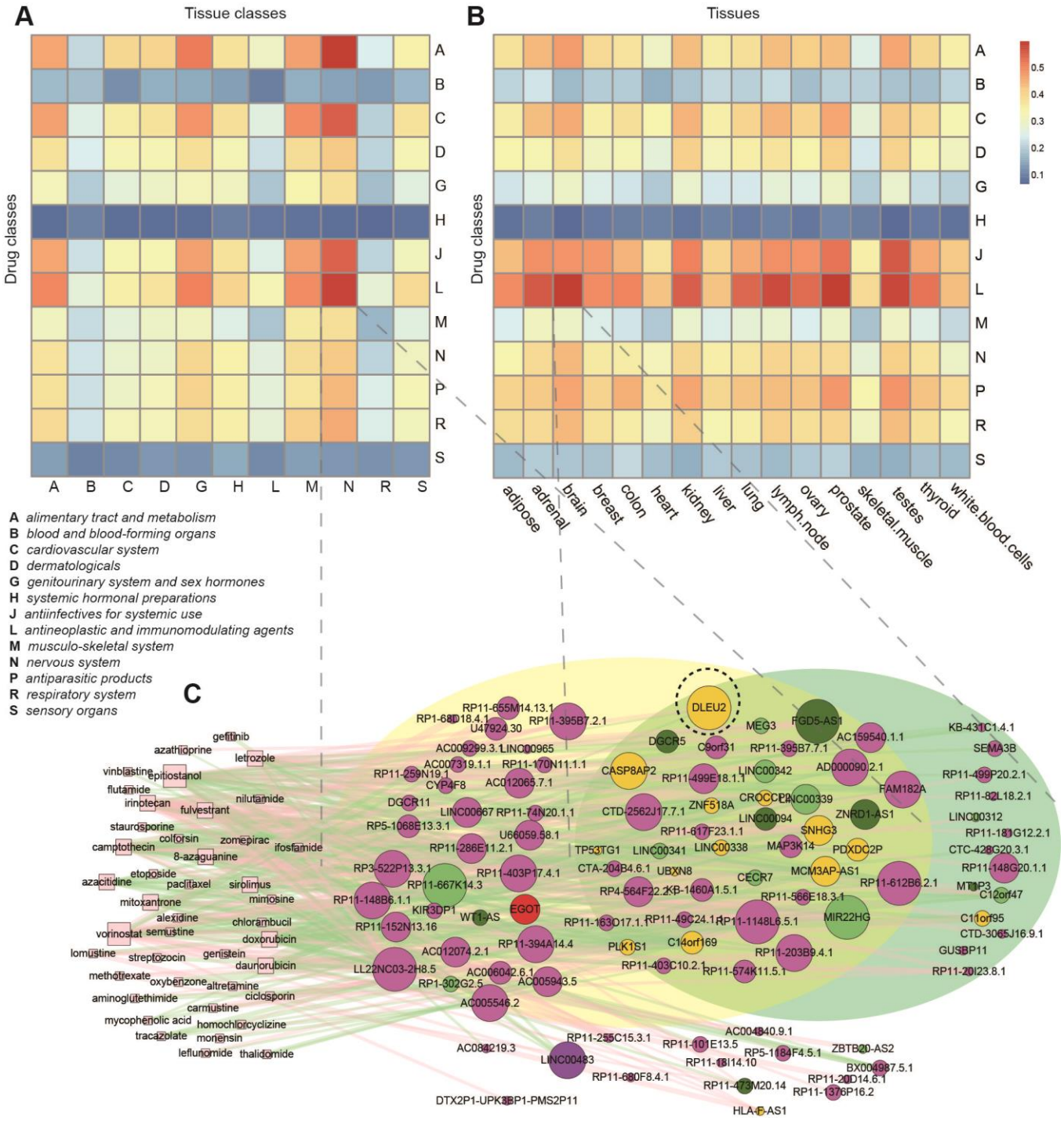


Figure 6. Tissue-specificity of drug-affected lncRNAs. (**A**) Jaccard coefficients of lncRNAs between 13 drug classes and tissues of 11 anatomical classes. (**B**) Jaccard coefficients of lncRNAs between 13 drug classes and 16 tissues. (C) Sub-network of the SMLN with drugs belonging to the (L) code and their affected lncRNAs.

DISCUSSION

LncRNAs have multiple biological function and are considered as potential therapeutic targets for diseases [40–42]. LncRNA-targeting therapeutics can modulate disease pathways that have previously been considered to be intractable [43]. However, the global relationships between drugs and their affected lncRNAs and the therapeutic potential of lncRNAs are not well-characterized. Currently, a common way to analyze lncRNA expression is RNA-seq. However, publicly available RNA-seq datasets of drug treatments are relatively limited compared with array-based expression profiles. Therefore, we extracted lncRNA expression profiles of drug treatments from LNCmap, which adopts a strict threshold in the reannotation process, and obtained similar expression values than those of RNA-seq data.

Then, we constructed a global SMLN consisting of 1206 nodes (1033 small molecules and 173 lncRNAs), and 4770 edges (2628 downregulated and 2142 upregulated). Based on the SMLN, we found that connected drug pairs (two drugs sharing the same lncRNAs) tended to have similar pharmacological properties. Especially, these connected drug pairs tended to have similar structure, indicating the possibility of structure-based combinations between drugs and lncRNAs in our SMLN. Because two drugs may bind to the same "structural motifs" in the corresponding lncRNA, future investigation of lncRNA structures may reveal critical functional RNA motifs of lncRNAs susceptible to small molecule targeting [44]. Furthermore, we investigated the functional interface between drugs and lncRNA modules. By inferring the functions of lncRNAs through their co-expressing mRNAs, we found that the functions of lncRNAs in the modular interface were related to the mode of action or side effects of the corresponding connected drugs, suggesting that lncRNAs may directly/indirectly participate in biological processes after drug administration. Finally, we investigated whether drugs tend to affect the lncRNAs expressed in their corresponding target-tissue. By integrating the tissue-specific lncRNAs and drug-affected lncRNAs, we found that drugs from all drug classes did not affect the lncRNAs expressed in the corresponding anatomical tissues. Some kinds of drugs tended to have a broader influence (e.g. antineoplastic and immunomodulating drugs), whereas some tissue-specific lncRNAs tended to be affected by multiple kinds of drugs. This might be due to the broad effects of some systemic drugs. For example, (L) drugs were used for systemic antineoplastic and immunomodulating therapy. Another reason might be that there were only 173 lncRNAs that were reannotated, leading to an insufficient drug-lncRNA map. Nevertheless, we obtained some interesting findings.

To confirm the validity of our results, we also constructed the SMLN with fold change = 1.5 and fold change = 3 respectively and repeated some of the analyses for the network. We found that the SMLN networks with different fold changes were robust (Supplementary Table 2). This result indicated that our method was robust. There were some limitations in our study. First, there were only 173 lncRNAs in the SMLN. This was due to the strict threshold in the reannotation process and the limited data on drug effects. The SMLN can be improved by integrating additional drug-affected expression profiles and RNA-seq data from different resources. Another limitation of our study was the incompleteness of the functional annotation of lncRNAs. This limitation will be alleviated, to a great extent, by the development of functional genome and RNAi research, as well as the integration of bioinformatics databases.

In summary, we have constructed and analyzed the SMLN that provided a comprehensive picture of global associations between drugs and their affected lncRNAs. A better understanding of the relation between small molecule agents and lncRNAs would not only promote the repositioning and rational clinical use of these agents but also provides new insights into lncRNA-targeting therapeutics.

MATERIALS AND METHODS

Construction of the SMLN

Drugs and affected lncRNAs were obtained from the LNCmap (http://bio-bigdata.hrbmu.edu.cn/LncMAP/). The LNCmap extracted drug-affected lncRNA expression profiles by reannotating the microarray data from the CMap database. According to the pipeline of ncFANs [12], the LNCmap developed a similar computational method to reannotate lncRNAs from expression microarray of coding genes. LNCmap reannotated 5,916 microarray profiles, with 674 instances from the Human Genome U133 Set (HG-U133A) platform and 5242 instances from the GeneChip HT Human Genome U133 Array Plate Set (HT_HG-U133A) platform. We then used the R package affy to compute expression values for all lncRNA expression profiles and obtained log2-fold change values to identify differentially expressed lncRNAs (DEL). The DELs were merged if the corresponding experiments belonged to the same drug. After the above steps, we obtained 4,770 small molecule-lncRNA relationships, including 1,005 small molecules and 173 lncRNAs, and constructed a bipartite

small molecule lncRNA network (SMLN). SMLN can be visualized by Cytoscape 3.0 [45] (Figure 2).

Generating the LLN

We generated the LLN in which lncRNAs represented nodes and two lncRNAs were connected if they shared significant numbers of small molecules. Because of the marked differences between the number of lncRNAs (173) and small molecules (1005), lncRNAs were connected to each other closely. To improve the specificity and identify the more significant lncRNA pairs, we adopted a hypergeometric test to generate the LLN.

$$p = 1 - \sum_{x=0}^{r-1} \frac{\binom{t}{x}\binom{m-t}{n-x}}{\binom{m}{n}}$$

Here, we collected m total small molecules in the SMLN, for each two lncRNAs i and j, t was the number of small molecules affected by lncRNA i, and n was the number of small molecules affected by lncRNA j, of which r was overlapped small molecules of the two small molecule sets. After calculating the P-value, we adopted the FDR-corrected q-values to reduce the false positive discovery rate. Significant lncRNA pairs ($P<0.01$, q-values<0.01) were obtained to construct the LLN.

Datasets of pharmacological properties

Indications

We collected the drug-indication associations from the study of Yildirim et al [27]. We also downloaded the drug-indication associations from Therapeutic Target Database (TTD) [46], then integrated the two datasets manually.

Drug targets

We downloaded the drug-target associations from the DrugBank database [29], which is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information. We obtained 399 small molecules in our SMLN.

Side effects

We downloaded the drug side effect dataset from a public computer-readable resource, SIDER, which is a freely available database that contains information on marketed medicines and their recorded adverse drug reactions [30]. We collected 997 drugs corresponding to 4492 side effects, including 303 small molecules in the SMLN.

Drug chemical similarity

We downloaded the SMILES files of small molecules in the SSN from the DrugBank database and Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.kegg.jp/kegg/drug/). We computed the TC scores of drug pairs using the Chemical Development Kit with default parameters [47].

Pathway enrichment

Pathway enrichment analysis was implemented based on co-expressed protein-coding genes of lncRNAs by using SubpathwayMiner tools [48]. We calculated the Pearson's correlation coefficient (PCC) between all reannotated lncRNA expression files and mRNA expression profiles of CMap. Using the setting $|PCC|>0.5$ and $p < 0.01$, we obtained the correlating mRNAs for pathway enrichment. The pathway enrichment was implemented by SubpathwayMiner with default parameters.

Tissue-specificity

We used the GSE1133 dataset and the ArrayExpress database (ERP000546) to study the tissue-specificity of drug-affected lncRNAs. We first reannotated the microarray dataset of GSE1133 and obtained 176 lncRNAs across 79 healthy tissues; then, we calculated tissue specificity scores for lncRNAs and identified tissue-specific lncRNAs (score >0.8) for each tissue [49]. According to the ATC classification of tissues and drugs, tissue-specific lncRNAs and drug-affected lncRNAs were allocated to the ATC classification separately, and we calculated the Jaccard coefficient between the tissue ATC classification and drug ATC classification to measure the similarity between lncRNAs related to different classifications of tissue and drug. We used the ArrayExpress database (ERP000546) to calculate the Jaccard coefficients of lncRNAs between 13 drug classes and 16 tissues by processing the RNA-seq data of 16 normal human individual tissues.

AUTHOR CONTRIBUTIONS

DSS conceived and designed the study, DSS and HXY collected and processed the data, HXY, YPZ and YJX performed the experiments, DSS and HXY wrote the manuscript; All (CLZ, JWH, FS, BL, XQL and KM) authors analyzed the data. DSS supervised the research and revised the manuscript. All authors reviewed the manuscript.

CONFLICTS OF INTEREST

The authors declare no potential conflicts of interest.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant Nos. 31501074, 81501868, 81803012, 61902095).

REFERENCES

1. Ritter N, Ali T, Kopitchinski N, Schuster P, Beisaw A, Hendrix DA, Schulz MH, Müller-McNicoll M, Dimmeler S, Grote P. The lncRNA Locus Handsdown Regulates Cardiac Gene Programs and Is Essential for Early Mouse Development. Dev Cell. 2019; 50:644–657.e8. https://doi.org/10.1016/j.devcel.2019.07.013 PMID:31422919

2. Butler AA, Johnston DR, Kaur S, Lubin FD. Long noncoding RNA NEAT1 mediates neuronal histone methylation and age-related memory impairment. Sci Signal. 2019; 12:eaaw9277. https://doi.org/10.1126/scisignal.aaw9277 PMID:31266852

3. Taiana E, Favasuli V, Ronchetti D, Todoerti K, Pelizzoni F, Manzoni M, Barbieri M, Fabris S, Silvestris I, Gallo Cantafio ME, Platonova N, Zuccalà V, Maltese L, et al. Long non-coding RNA NEAT1 targeting impairs the DNA repair machinery and triggers anti-tumor activity in multiple myeloma. Leukemia. 2020; 34:234–244. https://doi.org/10.1038/s41375-019-0542-5 PMID:31427718

4. Li L, Wang L, Li H, Han X, Chen S, Yang B, Hu Z, Zhu H, Cai C, Chen J, Li X, Huang J, Gu D. Characterization of LncRNA expression profile and identification of novel LncRNA biomarkers to diagnose coronary artery disease. Atherosclerosis. 2018; 275:359–67. https://doi.org/10.1016/j.atherosclerosis.2018.06.866 PMID:30015300

5. Zhang Y, Sun L, Xuan L, Pan Z, Hu X, Liu H, Bai Y, Jiao L, Li Z, Cui L, Wang X, Wang S, Yu T, et al. Long non-coding RNA CCRR controls cardiac conduction via regulating intercellular coupling. Nat Commun. 2018; 9:4176. https://doi.org/10.1038/s41467-018-06637-9 PMID:30301979

6. Yang Q, Xu E, Dai J, Liu B, Han Z, Wu J, Zhang S, Peng B, Zhang Y, Jiang Y. A novel long noncoding RNA AK001796 acts as an oncogene and is involved in cell growth inhibition by resveratrol in lung cancer. Toxicol Appl Pharmacol. 2015; 285:79–88. https://doi.org/10.1016/j.taap.2015.04.003 PMID:25888808

7. Jiang Y, Du W, Chu Q, Qin Y, Tuguzbaeva G, Wang H, Li A, Li G, Li Y, Chai L, Yue E, Sun X, Wang Z, et al. Downregulation of long non-coding RNA Kcnq1ot1: an important mechanism of arsenic trioxide-induced long QT syndrome. Cell Physiol Biochem. 2018; 45:192–202. https://doi.org/10.1159/000486357 PMID:29339628

8. Jia L, Tian Y, Chen Y, Zhang G. The silencing of LncRNA-H19 decreases chemoresistance of human glioma cells to temozolomide by suppressing epithelial-mesenchymal transition via the Wnt/β-Catenin pathway. Onco Targets Ther. 2018; 11:313–21. https://doi.org/10.2147/OTT.S154339 PMID:29391808

9. Xie C, Zhang LZ, Chen ZL, Zhong WJ, Fang JH, Zhu Y, Xiao MH, Guo ZW, Zhao N, He X, Zhuang SM. A hMTR4-PDIA3P1-miR-125/124-TRAF6 Regulatory Axis and Its Function in NF kappa B Signaling and Chemoresistance. Hepatology. 2019. [Epub ahead of print]. https://doi.org/10.1002/hep.30931 PMID:31509261

10. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006; 313:1929–35. https://doi.org/10.1126/science.1132939 PMID:17008526

11. Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, Chen Y, Liu XS. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. Nat Struct Mol Biol. 2013; 20:908–13. https://doi.org/10.1038/nsmb.2591 PMID:23728290

12. Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, Zhao G, Luo H, Bu D, Zhao H, Skogerbø G, Wu Z, Zhao Y. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. Nucleic Acids Res. 2011; 39:3864–78. https://doi.org/10.1093/nar/gkq1348 PMID:21247874

13. Yang H, Shang D, Xu Y, Zhang C, Feng L, Sun Z, Shi X, Zhang Y, Han J, Su F, Li C, Li X. The LncRNA Connectivity Map: using lncRNA signatures to connect small molecules, lncRNAs, and diseases. Sci Rep. 2017; 7:6655. https://doi.org/10.1038/s41598-017-06897-3 PMID:28751672

14. Xu J, Li CX, Li YS, Lv JY, Ma Y, Shao TT, Xu LD, Wang YY, Du L, Zhang YP, Jiang W, Li CQ, Xiao Y, Li X. MiRNA-miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features. Nucleic Acids Res. 2011; 39:825–36. https://doi.org/10.1093/nar/gkq832 PMID:20929877

15. Vanhaecke T, Papeleu P, Elaut G, Rogiers V. Trichostatin A-like hydroxamate histone deacetylase inhibitors as therapeutic agents: toxicological point of view. Curr Med Chem. 2004; 11:1629–43.

https://doi.org/10.2174/0929867043365099 PMID:15180568

16. Drummond DC, Noble CO, Kirpotin DB, Guo Z, Scott GK, Benz CC. Clinical development of histone deacetylase inhibitors as anticancer agents. Annu Rev Pharmacol Toxicol. 2005; 45:495–528. https://doi.org/10.1146/annurev.pharmtox.45.120403.095825 PMID:15822187

17. Shankar S, Srivastava RK. Histone deacetylase inhibitors: mechanisms and clinical significance in cancer: HDAC inhibitor-induced apoptosis. Adv Exp Med Biol. 2008; 615:261–98. https://doi.org/10.1007/978-1-4020-6554-5_13 PMID:18437899

18. Miller SC, Huang R, Sakamuru S, Shukla SJ, Attene-Ramos MS, Shinn P, Van Leer D, Leister W, Austin CP, Xia M. Identification of known drugs that act as inhibitors of NF-kappaB signaling and their mechanism of action. Biochem Pharmacol. 2010; 79:1272–80. https://doi.org/10.1016/j.bcp.2009.12.021 PMID:20067776

19. Boon-Unge K, Yu Q, Zou T, Zhou A, Govitrapong P, Zhou J. Emetine regulates the alternative splicing of Bcl-x through a protein phosphatase 1-dependent mechanism. Chem Biol. 2007; 14:1386–92. https://doi.org/10.1016/j.chembiol.2007.11.004 PMID:18096507

20. Grollman AP. Inhibitors of protein biosynthesis. II. Mode of action of anisomycin. J Biol Chem. 1967; 242:3226–33. PMID:6027796

21. Hori T, Kondo T, Tabuchi Y, Takasaki I, Zhao QL, Kanamori M, Yasuda T, Kimura T. Molecular mechanism of apoptosis and gene expressions in human lymphoma U937 cells treated with anisomycin. Chem Biol Interact. 2008; 172:125–40. https://doi.org/10.1016/j.cbi.2007.12.003 PMID:18241849

22. Seth AK, Misra A, Umrigar D. Topical liposomal gel of idoxuridine for the treatment of herpes simplex: pharmaceutical and clinical implications. Pharm Dev Technol. 2004; 9:277–89. https://doi.org/10.1081/PDT-200031432 PMID:15458233

23. Kasar S, Underbayev C, Yuan Y, Hanlon M, Aly S, Khan H, Chang V, Batish M, Gavrilova T, Badiane F, Degheidy H, Marti G, Raveche E. Therapeutic implications of activation of the host gene (Dleu2) promoter for miR-15a/16-1 in chronic lymphocytic leukemia. Oncogene. 2014; 33:3307–15. https://doi.org/10.1038/onc.2013.291 PMID:23995789

24. Chen CQ, Chen CS, Chen JJ, Zhou LP, Xu HL, Jin WW, Wu JB, Gao SM. Histone deacetylases inhibitor trichostatin A increases the expression of Dleu2/miR-15a/16-1 via HDAC3 in non-small cell lung cancer. Mol Cell Biochem. 2013; 383:137–48. https://doi.org/10.1007/s11010-013-1762-z PMID:23867991

25. Yumba-Mpanga A, Struck-Lewicka W, Wawrzyniak R, Markuszewski M, Roslan M, Kaliszan R, Markuszewski MJ. Metabolomic Heterogeneity of Urogenital Tract Cancers Analyzed by Complementary Chromatographic Techniques Coupled with Mass Spectrometry. Curr Med Chem. 2019; 26:216–231. https://doi.org/10.2174/0929867324666171006150326 PMID:28990506

26. Li C, Shang D, Wang Y, Li J, Han J, Wang S, Yao Q, Wang Y, Zhang Y, Zhang C, Xu Y, Jiang W, Li X. Characterizing the network of drugs and their affected metabolic subpathways. PLoS One. 2012; 7:e47326. https://doi.org/10.1371/journal.pone.0047326 PMID:23112813

27. Yildirim MA, Goh KI, Cusick ME, Barabási AL, Vidal M. Drug-target network. Nat Biotechnol. 2007; 25:1119–26. https://doi.org/10.1038/nbt1338 PMID:17921997

28. Gadow KD, Smith RM, Pinsonneault JK. Serotonin 2A receptor gene (HTR2A) regulatory variants: possible association with severity of depression symptoms in children with autism spectrum disorder. Cogn Behav Neurol. 2014; 27:107–16. https://doi.org/10.1097/WNN.0000000000000028 PMID:24968012

29. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014; 42:D1091–97. https://doi.org/10.1093/nar/gkt1068 PMID:24203711

30. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol. 2010; 6:343. https://doi.org/10.1038/msb.2009.98 PMID:20087340

31. Adamcsek B, Palla G, Farkas IJ, Derényi I, Vicsek T. CFinder: locating cliques and overlapping modules in biological networks. Bioinformatics. 2006; 22:1021–23. https://doi.org/10.1093/bioinformatics/btl039 PMID:16473872

32. Ji L, Xing W, Dong L, Gao F. Searching for phytoinsulins as cardiovascular protector in metabolic syndrome. Curr Pharm Des. 2013; 19:4848–58. https://doi.org/10.2174/1381612811319270007 PMID:23323618

33. Ito K, Caramori G, Adcock IM. Therapeutic potential of phosphatidylinositol 3-kinase inhibitors in inflammatory respiratory disease. J Pharmacol Exp Ther. 2007; 321:1–8. https://doi.org/10.1124/jpet.106.111674 PMID:17021257

34. Ozaki K, Leonard WJ. Cytokine and cytokine receptor pleiotropy and redundancy. J Biol Chem. 2002; 277:29355–58. https://doi.org/10.1074/jbc.R200003200 PMID:12072446

35. Ling H, Fabbri M, Calin GA. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. Nat Rev Drug Discov. 2013; 12:847–65. https://doi.org/10.1038/nrd4140 PMID:24172333

36. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, et al. Landscape of transcription in human cells. Nature. 2012; 489:101–08. https://doi.org/10.1038/nature11233 PMID:22955620

37. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA. 2004; 101:6062–67. https://doi.org/10.1073/pnas.0400782101 PMID:15075390

38. Barbosa-Morais NL, Irimia M, Pan Q, Xiong HY, Gueroussov S, Lee LJ, Slobodeniuc V, Kutter C, Watt S, Colak R, Kim T, Misquitta-Ali CM, Wilson MD, et al. The evolutionary landscape of alternative splicing in vertebrate species. Science. 2012; 338:1587–93. https://doi.org/10.1126/science.1230612 PMID:23258890

39. Chen M, Maloney JA, Kallop DY, Atwal JK, Tam SJ, Baer K, Kissel H, Kaminker JS, Lewcock JW, Weimer RM, Watts RJ. Spatially coordinated kinase signaling regulates local axon degeneration. J Neurosci. 2012; 32:13439–53. https://doi.org/10.1523/JNEUROSCI.2039-12.2012 PMID:23015435

40. Wang B, Zheng J, Li R, Tian Y, Lin J, Liang Y, Sun Q, Xu A, Zheng R, Liu M, Ji A, Bu J, Yuan Y. Long noncoding RNA LINC02582 acts downstream of miR-200c to promote radioresistance through CHK1 in breast cancer cells. Cell Death Dis. 2019; 10:764. https://doi.org/10.1038/s41419-019-1996-0 PMID:31601781

41. Tang J, Yu B, Li Y, Zhang W, Alvarez AA, Hu B, Cheng SY, Feng H. TGF-β-activated lncRNA LINC00115 is a critical regulator of glioma stem-like cell tumorigenicity. EMBO Rep. 2019; 20:e48170. https://doi.org/10.15252/embr.201948170 PMID:31599491

42. Xu X, Gu J, Ding X, Ge G, Zang X, Ji R, Shao M, Mao Z, Zhang Y, Zhang J, Mao F, Qian H, Xu W, et al. LINC00978 promotes the progression of hepatocellular carcinoma by regulating EZH2-mediated silencing of p21 and E-cadherin expression. Cell Death Dis. 2019; 10:752. https://doi.org/10.1038/s41419-019-1990-6 PMID:31582742

43. Wahlestedt C. Targeting long non-coding RNA to herapeutically upregulate gene expression. Nat Rev Drug Discov. 2013; 12:433–46. https://doi.org/10.1038/nrd4018 PMID:23722346

44. Li CH, Chen Y. Targeting long non-coding RNAs in cancers: progress and prospects. Int J Biochem Cell Biol. 2013; 45:1895–910. https://doi.org/10.1016/j.biocel.2013.05.030 PMID:23748105

45. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–504. https://doi.org/10.1101/gr.1239303 PMID:14597658

46. Li YH, Yu CY, Li XX, Zhang P, Tang J, Yang Q, Fu T, Zhang X, Cui X, Tu G, Zhang Y, Li S, Yang F, et al. Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. Nucleic Acids Res. 2018; 46:D1121–D1127. https://doi.org/10.1093/nar/gkx1076 PMID:29140520

47. Steinbeck C, Hoppe C, Kuhn S, Floris M, Guha R, Willighagen EL. Recent developments of the chemistry development kit (CDK) - an open-source java library for chemo- and bioinformatics. Curr Pharm Des. 2006; 12:2111–20. https://doi.org/10.2174/138161206777585274 PMID:16796559

48. Li C, Han J, Yao Q, Zou C, Xu Y, Zhang C, Shang D, Zhou L, Zou C, Sun Z, Li J, Zhang Y, Yang H, et al. Subpathway-GM: identification of metabolic subpathways via joint power of interesting genes and metabolites and their topologies within pathways. Nucleic Acids Res. 2013; 41:e101. https://doi.org/10.1093/nar/gkt161 PMID:23482392

49. Liu MX, Chen X, Chen G, Cui QH, Yan GY. A computational framework to infer human disease-associated long noncoding RNAs. PLoS One. 2014; 9:e84408. https://doi.org/10.1371/journal.pone.0084408 PMID:24392133

SUPPLEMENTARY MATERIALS

MATERIALS AND METHODS

Construction of the SMLN

Drugs and affected lncRNAs were obtained from the LNCmap. The LNCmap extracted drug-affected lncRNA expression profiles by reannotating the microarray data from the CMap database. According to the pipeline of ncFANs [1], the LNCmap developed a similar computational method to reannotate lncRNAs from expression microarray of coding genes. LNCmap reannotated 5916 microarray profiles, with 674 instances from the Human Genome U133 Set (HG-U133A) platform and 5242 instances from the GeneChip HT Human Genome U133 Array Plate Set (HT_HG-U133A) platform. We then used the R package affy to compute expression values for all lncRNA expression profiles and obtained log2-fold change values to identify differentially expressed lncRNAs (DEL). The DELs were merged if the corresponding experiments belonged to the same drug. After the above steps, we obtained 4770 small molecule-lncRNA relationships, including 1005 small molecules and 173 lncRNAs, and constructed a bipartite small-molecule lncRNA network (SMLN).

Generating the LLN

We generated the LLN in which lncRNAs represented nodes and two lncRNAs were connected if they shared significant numbers of small molecules. Because of the marked differences between the number of lncRNAs (173) and small molecules (1005), lncRNAs were connected to each other closely. To improve the specificity and identify the more significant lncRNA pairs, we adopted a hypergeometric test to generate the LLN.

$$p = 1 - \sum_{x=0}^{r-1} \frac{\binom{t}{x}\binom{m-t}{n-x}}{\binom{m}{n}}$$

Here, we collected m total small molecules in the SMLN, for each two lncRNAs i and j, t was the number of small molecules affected by lncRNA i, and n was the number of small molecules affected by lncRNA j, of which r was overlapped small molecules of the two small-molecule sets. After calculating the P-value, we adopted the FDR-corrected q-values to reduce the false positive discovery rate. Significant lncRNA pairs (P<0.01, q-values<0.01) were obtained to construct the LLN.

Datasets of pharmacological properties

Indications

We collected the drug-indication associations from the study of Yildirim et al [2]. We also downloaded the drug-indication associations from Therapeutic Target Database (TTD) [3], then integrated the two datasets manually.

Drug targets

We downloaded the drug-target associations from the DrugBank database [4], which is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information. We obtained 399 small molecules in our SMLN.

Side effects

We downloaded the drug side effect dataset from a public computer-readable resource, SIDER, which is a freely available database that contains information on marketed medicines and their recorded adverse drug reactions [5]. We collected 997 drugs corresponding to 4492 side effects, including 303 small molecules in the SMLN.

Drug chemical similarity

We downloaded the SMILES files of small molecules in the SSN from the DrugBank database and Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.kegg.jp/kegg/drug/). We computed the TC scores of drug pairs using the Chemical Development Kit with default parameters [6].

Pathway enrichment

Pathway enrichment analysis was implemented based on co-expressed protein-coding genes of lncRNAs by using SubpathwayMiner tools [7]. We calculated the Pearson correlation coefficient (PCC) between all reannotated lncRNA expression files and mRNA expression profiles of CMap. Using the setting |PCC|>0.5 and $p < 0.01$, we obtained the correlating mRNAs for pathway enrichment. The pathway enrichment was implemented by SubpathwayMiner with default parameters.

Tissue-specificity

We used the GSE1133 dataset and the ArrayExpress database (ERP000546) to study the tissue-specificity of drug-affected lncRNAs. We firstly re-annotated the microarray dataset of GSE1133 and obtained 176

lncRNAs across 79 healthy tissues; then, we calculated tissue specificity scores for lncRNAs and identified tissue-specific lncRNAs (score >0.8) for each tissue [8]. According to the ATC classification of tissues and drugs, tissue-specific lncRNAs and drug-affected lncRNAs were allocated to the ATC classification separately, and we calculated the Jaccard coefficient between the tissue ATC classification and drug ATC classification to measure the similarity between lncRNAs related to different classifications of tissue and drug. We used the ArrayExpress database (ERP000546) to calculate the Jaccard coefficients of lncRNAs between 13 drug classes and 16 tissues by processing the RNA-seq data of 16 normal human individual tissues.

The basic properties of the SMLN

The degree of small-molecule nodes spanned a wide range from 1 to 87. The highest degree node was trichostatin A (TSA), an organic compound that serves as an antifungal antibiotic and selectively inhibits class I and II mammalian histone deacetylases (HDACs) [9]. TSA can broadly alter gene expression by interfering with the removal of acetyl groups from histones [10, 11]. It is also a member of a larger class of histone deacetylase inhibitors that have a broad spectrum of epigenetic activities [10, 11]. The second highest degree small molecule node (degree=46) was emetine, an anti-malaria drug that was recently found to have broad anticancer activity in many types of malignancies including breast, colon, prostate, skin, and lymphoid tumors by inhibiting NF-κB signaling or regulating the RNA splicing of members of the Bcl-2 family [12, 13]. Although there are no specific reports about emetine and lncRNAs, it was linked to many lncRNAs, partly because of its broad anticancer effects. Interestingly, we found that other highly-connected nodes, namely anisomycin and idoxuridine (degree: 39 and 38, respectively) could inhibit protein/DNA synthesis. Anisomycin is a potent apoptosis inducer that functions by activating JNK/SAPK and inhibiting protein/DNA synthesis during translation [14, 15]. Idoxuridine, which is used as an antiviral agent, is an analog of deoxyuridine, an inhibitor of viral DNA synthesis [16]. The high connectivity may have been due to their activity related to apoptosis and the inhibition of protein/DNA synthesis.

Similar to the small molecule nodes, the lncRNA nodes also displayed evident differences in connection (range, 1–366). The lncRNA node with the highest degree was RP11-1148L6.5.1. There are no functional studies about this lncRNA. To date, few lncRNAs have been functionally annotated. Of seven lncRNAs with a degree >100, only DLEU2 (Deleted in Lymphocytic lEUkemia 2) is well studied. It encodes a pair of critical pro-apoptotic microRNAs, miR-15a/16-1, which are critical for the increased survival exhibited by chronic lymphocytic leukemia cells [17]. Chen et Al. indicated that the HDAC inhibitor TSA, the most-connected small molecule in the SMLN, could upregulate the expression of miR-15a/16-1, residing in the host tumor suppressor *DLEU2* gene [18]. Furthermore, in our SMLN, TSA could also upregulate DLEU2 (log2 fold change = 1.4), suggesting that our SMLN could identify a promising cancer therapy via targeting lncRNAs [17].

REFERENCES

1. Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, Zhao G, Luo H, Bu D, Zhao H, Skogerbø G, Wu Z, Zhao Y. Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. Nucleic Acids Res. 2011; 39:3864–78. https://doi.org/10.1093/nar/gkq1348 PMID:21247874

2. Yildirim MA, Goh KI, Cusick ME, Barabási AL, Vidal M. Drug-target network. Nat Biotechnol. 2007; 25:1119–26. https://doi.org/10.1038/nbt1338 PMID:17921997

3. Li YH, Yu CY, Li XX, Zhang P, Tang J, Yang Q, Fu T, Zhang X, Cui X, Tu G, Zhang Y, Li S, Yang F, et al. Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. Nucleic Acids Res. 2018; 46:D1121–D1127. https://doi.org/10.1093/nar/gkx1076 PMID:29140520

4. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, et al. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res. 2014; 42:D1091–97. https://doi.org/10.1093/nar/gkt1068 PMID:24203711

5. Kuhn M, Campillos M, Letunic I, Jensen LJ, Bork P. A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol. 2010; 6:343. https://doi.org/10.1038/msb.2009.98 PMID:20087340

6. Steinbeck C, Hoppe C, Kuhn S, Floris M, Guha R, Willighagen EL. Recent developments of the chemistry development kit (CDK) - an open-source java library for chemo- and bioinformatics. Curr Pharm Des. 2006; 12:2111–20. https://doi.org/10.2174/138161206777585274 PMID:16796559

7. Li C, Han J, Yao Q, Zou C, Xu Y, Zhang C, Shang D, Zhou L, Zou C, Sun Z, Li J, Zhang Y, Yang H, et al. Subpathway-GM: identification of metabolic subpathways via joint power of interesting genes and metabolites and their topologies within pathways. Nucleic Acids Res. 2013; 41:e101. https://doi.org/10.1093/nar/gkt161

PMID:23482392

8. Liu MX, Chen X, Chen G, Cui QH, Yan GY. A computational framework to infer human disease-associated long noncoding RNAs. PLoS One. 2014; 9:e84408. https://doi.org/10.1371/journal.pone.0084408 PMID:24392133

9. Vanhaecke T, Papeleu P, Elaut G, Rogiers V. Trichostatin A-like hydroxamate histone deacetylase inhibitors as therapeutic agents: toxicological point of view. Curr Med Chem. 2004; 11:1629–43. https://doi.org/10.2174/0929867043365099 PMID:15180568

10. Drummond DC, Noble CO, Kirpotin DB, Guo Z, Scott GK, Benz CC. Clinical development of histone deacetylase inhibitors as anticancer agents. Annu Rev Pharmacol Toxicol. 2005; 45:495–528. https://doi.org/10.1146/annurev.pharmtox.45.120403.095825 PMID:15822187

11. Shankar S, Srivastava RK. Histone deacetylase inhibitors: mechanisms and clinical significance in cancer: HDAC inhibitor-induced apoptosis. Adv Exp Med Biol. 2008; 615:261–98. https://doi.org/10.1007/978-1-4020-6554-5_13 PMID:18437899

12. Miller SC, Huang R, Sakamuru S, Shukla SJ, Attene-Ramos MS, Shinn P, Van Leer D, Leister W, Austin CP, Xia M. Identification of known drugs that act as inhibitors of NF-kappaB signaling and their mechanism of action. Biochem Pharmacol. 2010; 79:1272–80. https://doi.org/10.1016/j.bcp.2009.12.021 PMID:20067776

13. Boon-Unge K, Yu Q, Zou T, Zhou A, Govitrapong P, Zhou J. Emetine regulates the alternative splicing of Bcl-x through a protein phosphatase 1-dependent mechanism. Chem Biol. 2007; 14:1386–92. https://doi.org/10.1016/j.chembiol.2007.11.004 PMID:18096507

14. Grollman AP. Inhibitors of protein biosynthesis. II. Mode of action of anisomycin. J Biol Chem. 1967; 242:3226–33. PMID:6027796

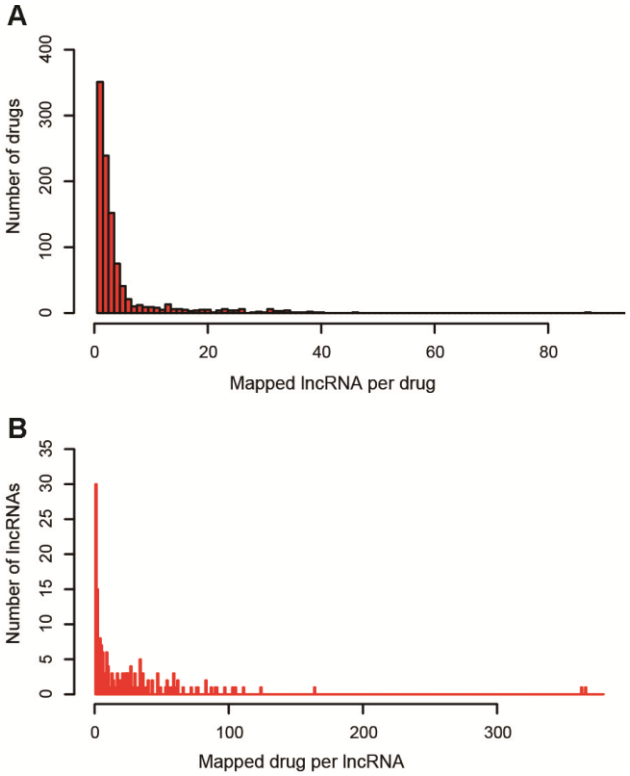
15. Hori T, Kondo T, Tabuchi Y, Takasaki I, Zhao QL, Kanamori M, Yasuda T, Kimura T. Molecular mechanism of apoptosis and gene expressions in human lymphoma U937 cells treated with anisomycin. Chem Biol Interact. 2008; 172:125–40. https://doi.org/10.1016/j.cbi.2007.12.003 PMID:18241849

16. Seth AK, Misra A, Umrigar D. Topical liposomal gel of idoxuridine for the treatment of herpes simplex: pharmaceutical and clinical implications. Pharm Dev Technol. 2004; 9:277–89. https://doi.org/10.1081/PDT-200031432 PMID:15458233

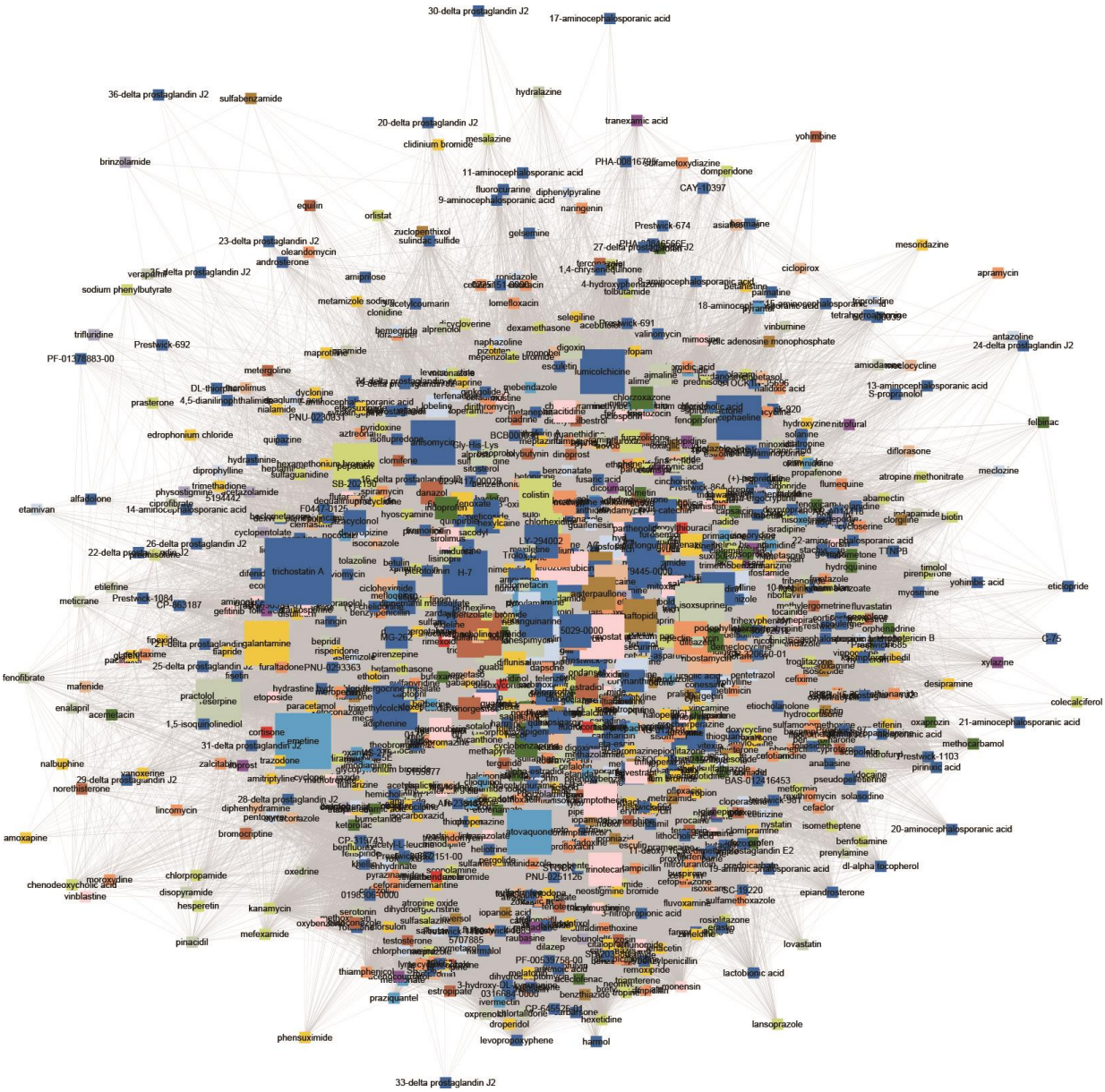
17. Kasar S, Underbayev C, Yuan Y, Hanlon M, Aly S, Khan H, Chang V, Batish M, Gavrilova T, Badiane F, Degheidy H, Marti G, Raveche E. Therapeutic implications of activation of the host gene (Dleu2) promoter for miR-15a/16-1 in chronic lymphocytic leukemia. Oncogene. 2014; 33:3307–15. https://doi.org/10.1038/onc.2013.291 PMID:23995789

18. Chen CQ, Chen CS, Chen JJ, Zhou LP, Xu HL, Jin WW, Wu JB, Gao SM. Histone deacetylases inhibitor trichostatin A increases the expression of Dleu2/miR-15a/16-1 via HDAC3 in non-small cell lung cancer. Mol Cell Biochem. 2013; 383:137–48. https://doi.org/10.1007/s11010-013-1762-z PMID:23867991

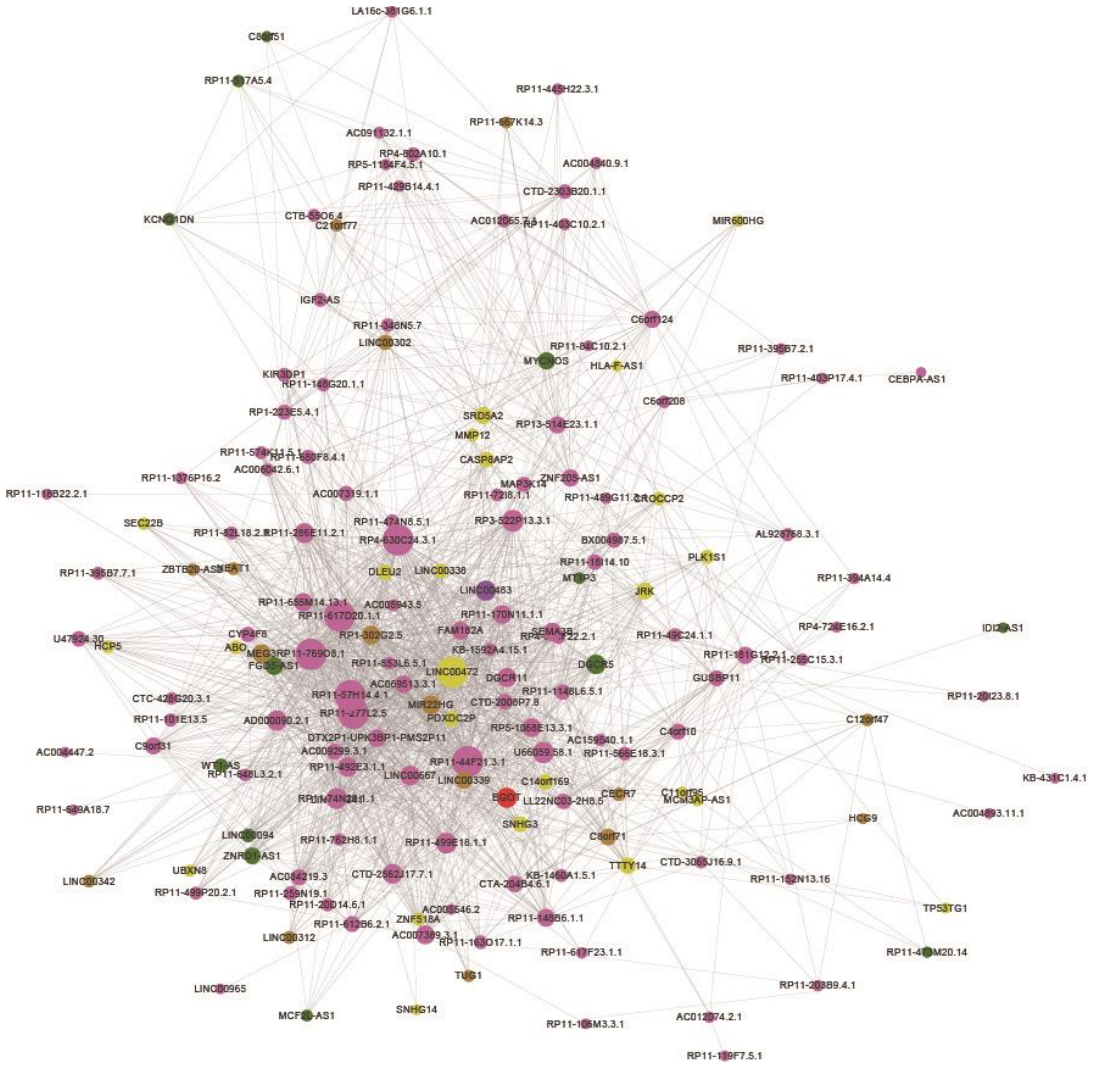
Supplementary Figures



Supplementary Figure 1. The degree distribution of (**A**) lncRNAs and (**B**) drugs of the SMLN.



Supplementary Figure 2. The small molecule-small molecule network (SSN).



Supplementary Figure 3. The lncRNA-lncRNA network (LLN).

Supplementary Tables

Supplementary Table 1. Tissue-specific lncRNAs affected by drugs.

Tissue	lncRNAs
Brain Amygdala	ENSG00000269614, ENSG00000225465
subthalamic_nucleus	ENSG00000267670
TestiIntersitial	ENSG00000267163
DRG	ENSG00000267161, ENSG00000104725, ENSG00000179935, ENSG00000235280
ciliary_ganglion	ENSG00000266897, ENSG00000240291, ENSG00000232860, ENSG00000196696, ENSG00000230223, ENSG00000242125, ENSG00000237941, ENSG00000263072
fetal liver	ENSG00000261613
Pons	ENSG00000261496
TemporalLobe	ENSG00000261087
Uterus_Corpus	ENSG00000260339
Skeletal_Muscle_Psoas	ENSG00000254488, ENSG00000236772, ENSG00000263214, ENSG00000260588, ENSG00000259539, ENSG00000246777, ENSG00000249717, ENSG00000273032, ENSG00000235994, ENSG00000267934, ENSG00000232416, ENSG00000265242, ENSG00000241881, ENSG00000231690, ENSG00000197670, ENSG00000215765, ENSG00000176728
PB_CD19BCells	ENSG00000253701
cerebellum	ENSG00000253230
atrioventricular_node	ENSG00000249790, ENSG00000229645, ENSG00000218510, ENSG00000235947, ENSG00000227719, ENSG00000260532, ENSG00000005206, ENSG00000261071, ENSG00000266171, ENSG00000236234, ENSG00000249673
Superior_Cervical_Ganglion	ENSG00000246863, ENSG00000236901, ENSG00000270742, ENSG00000257310, ENSG00000241954, ENSG00000245532, ENSG00000182165, ENSG00000182873, ENSG00000269176, ENSG00000258442, ENSG00000255224, ENSG00000232710, ENSG00000236268, ENSG00000235725, ENSG00000225733, ENSG00000234912, ENSG00000225930, ENSG00000049319, ENSG00000259417, ENSG00000236673, ENSG00000237517, ENSG00000152268, ENSG00000251023, ENSG00000261460, ENSG00000260396, ENSG00000237697, ENSG00000231074, ENSG00000229921, ENSG00000272579, ENSG00000235072, ENSG00000224945, ENSG00000259291, ENSG00000226674, ENSG00000255443, ENSG00000233237, ENSG00000233864, ENSG00000228389, ENSG00000248175, ENSG00000249532, ENSG00000229589, ENSG00000227372, ENSG00000262179, ENSG00000186526, ENSG00000261097, ENSG00000204054, ENSG00000226803, ENSG00000248161, ENSG00000229582, ENSG00000110347, ENSG00000266830, ENSG00000233718, ENSG00000204623, ENSG00000260804, ENSG00000204625, ENSG00000031544, ENSG00000125804, ENSG00000260735, ENSG00000224063, ENSG00000272599, ENSG00000232274, ENSG00000164621, ENSG00000136315, ENSG00000006062, ENSG00000259322, ENSG00000187621, ENSG00000176075, ENSG00000260400, ENSG00000272216, ENSG00000186842, ENSG00000260743, ENSG00000215117, ENSG00000233791, ENSG00000231607, ENSG00000196756, ENSG00000249859, ENSG00000249267, ENSG00000234350, ENSG00000242687, ENSG00000259849, ENSG00000253116, ENSG00000232931, ENSG00000267659, ENSG00000241295, ENSG00000225670, ENSG00000272201, ENSG00000255318, ENSG00000183242, ENSG00000132832, ENSG00000241345, ENSG00000235733, ENSG00000176734, ENSG00000177853, ENSG00000204148, ENSG00000260619, ENSG00000144596, ENSG00000236256, ENSG00000228343, ENSG00000261646, ENSG00000273311, ENSG00000257303, ENSG00000215424, ENSG00000263753, ENSG00000235437, ENSG00000259758, ENSG00000131007, ENSG00000246263, ENSG00000266904, ENSG00000256185, ENSG00000260917, ENSG00000228463, ENSG00000237775, ENSG00000237438, ENSG00000263050, ENSG00000228350, ENSG00000249604, ENSG00000264727, ENSG00000259577
Uterus	ENSG00000237125
Testi_SeminiferousTubule	ENSG00000235824, ENSG00000088970
lymph node	ENSG00000228315
Appendix	ENSG00000226334
BM_CD71EarlyErythroid	ENSG00000215908
HEART	ENSG00000213994, ENSG00000269653
WHOLEBLOOD(JJV)	ENSG00000206337
Trigeminal_Ganglion	ENSG00000198221, ENSG00000270589, ENSG00000175873, ENSG00000266968, ENSG00000262420, ENSG00000225092, ENSG00000269318, ENSG00000251632, ENSG00000259644, ENSG00000227403, ENSG00000254389, ENSG00000237250, ENSG00000263198, ENSG00000259073, ENSG00000273487, ENSG00000258517, ENSG00000270074, ENSG00000231160, ENSG00000272711, ENSG00000167117,

	ENSG00000235865
Testi_GermCell	ENSG00000197210
Liver	ENSG00000188338
fetal brain	ENSG00000188070
AdrenalCortex	ENSG00000186594, ENSG00000214548
spinal cord	ENSG00000099869
PLACENTA	ENSG00000012171

Supplementary Table 2. Robustness of SMLN with Fold Change=2 and Fold Change=1.5.

Degree rank	Drug		lncRNA	
	FC=2	FC=1.5	FC=2	FC=1.5
1	trichostatin A	trichostatin A	RP11-1148L6.5.1	LL22NC03-2H8.5
2	emetine	trichlormethiazide	LL22NC03-2H8.5	RP11-1148L6.5.1
3	pepstatin	probenecid	RP11-612B6.2.1	RP11-667K14.3
4	anisomycin	0175029-0000	RP11-667K14.3	RP11-395B7.2.1
5	idoxuridine	trimethylcolchicinic acid	FGD5-AS1	RP11-612B6.2.1
6	lumicolchicine	emetine	MIR22HG	AC005546.2
7	probenecid	galantamine	DLEU2	FAM182A
8	reserpine	mestranol	FAM182A	AC159540.1.1
9	mestranol	pyrimethamine	RP11-403P17.4.1	AC012074.2.1
10	sulfathiazole	pepstatin	CTD-2562J17.7.1	DLEU2
11	molindone	quinisocaine	RP11-148B6.1.1	RP11-148G20.1.1
12	galantamine	benperidol	RP11-395B7.2.1	RP11-148B6.1.1
13	H-7	idoxuridine	AD000090.2.1	CTD-2562J17.7.1
14	pyrimethamine	reserpine	AC005546.2	RP11-403P17.4.1
15	trimethylcolchicinic acid	lumicolchicine	RP11-394A14.4	RP11-203B9.4.1
16	trichlormethiazide	metronidazole	RP3-522P13.3.1	RP11-394A14.4
17	lanatoside C	epitiostanol	RP11-203B9.4.1	FGD5-AS1
18	mebhydrolin	canavanine	CASP8AP2	RP11-1376P16.2
19	canavanine	sulfathiazole	LINC00483	BX004987.5.1
20	vorinostat	molindone	ZNRD1-AS1	AC012065.7.1

Supplementary Dataset

Please browse Full Text version to see the data of Supplementary Datasets 1 to 4.

Supplementary Dataset 1. The edges of the SMLN

Supplementary Dataset 2. The degree distribution of SMLN

Supplementary Dataset 3. The edges of the drug-drug network

Supplementary Dataset 4. The edges of the lnc-lnc network