Briefings in Bioinformatics, 21(6), 2020, 2153–2166

doi: 10.1093/bib/bbz113
Advance Access Publication Date: 3 December 2019
Problem Solving Protocol

Identification and comprehensive characterization of lncRNAs with copy number variations and their driving transcriptional perturbed subpathways reveal functional significance for cancer

Yanjun Xu*, Tan Wu*, Feng Li*, Qun Dong*, Jingwen Wang, Desi Shang, Yingqi Xu, Chunlong Zhang, Yiying Dou, Congxue Hu, Haixiu Yang, Xuan Zheng, Yunpeng Zhang, Lihua Wang and Xia Li

Corresponding authors: Xia Li, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China. Tel.: 86-451-86615922; Fax: 86-451-86615922; E-mail: lixia@hrbmu.edu.cn; Lihua Wang, Department of Neurology, The Second Affiliated Hospital, Harbin Medical University, Harbin 150081, China. Tel.: 86-451-86605788; Fax: 86-451-86605788; E-mail: wanglh211@163.com; Yunpeng Zhang, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China. Tel.: 86-451-86615922; Fax: 86-451-86615922; E-mail: zyp19871208@126.com
*These authors contributed equally to this work.

Abstract

Numerous studies have shown that copy number variation (CNV) in lncRNA regions play critical roles in the initiation and progression of cancer. However, our knowledge about their functionalities is still limited. Here, we firstly provided a computational method to identify lncRNAs with copy number variation (lncRNAs-CNV) and their driving transcriptional perturbed subpathways by integrating multidimensional omics data of cancer. The high reliability and accuracy of our method have been demonstrated. Then, the method was applied to 14 cancer types, and a comprehensive characterization and analysis was performed. LncRNAs-CNV had high specificity in cancers, and those with high CNV level may perturb broad biological functions. Some core subpathways and cancer hallmarks widely perturbed by lncRNAs-CNV were revealed. Moreover, subpathways highlighted the functional diversity of lncRNAs-CNV in various cancers. Survival analysis indicated that functional lncRNAs-CNV could be candidate prognostic biomarkers for clinical applications, such as ST7-AS1,

Yanjun Xu is an instructor at the College of Bioinformatics Science and Technology, Harbin Medical University.
Tan Wu is an MS student at the College of Bioinformatics Science and Technology, Harbin Medical University.
Feng Li is an assistant researcher at the College of Bioinformatics Science and Technology, Harbin Medical University.
Qun Dong is an MS student at the College of Bioinformatics Science and Technology at Harbin Medical University.
Jingwen Wang is an MS student at the College of Bioinformatics Science and Technology, Harbin Medical University.
Desi Shang is an associate professor at the College of Bioinformatics Science and Technology, Harbin Medical University.
Yingqi Xu is an instructor at the College of Bioinformatics Science and Technology, Harbin Medical University.
Chunlong Zhang is an associate professor at the College of Bioinformatics Science and Technology at Harbin Medical University.
Yiying Dou is an undergraduate at the College of Bioinformatics Science and Technology, Harbin Medical University.
Congxue Hu is an MS student at the College of Bioinformatics Science and Technology, Harbin Medical University.
Haixiu Yang is an instructor at the College of Bioinformatics Science and Technology, Harbin Medical University.
Xuan Zheng is an MS student at the College of Bioinformatics Science and Technology, Harbin Medical University.
Yunpeng Zhang is an associate professor at the College of Bioinformatics Science and Technology, Harbin Medical University.
Lihua Wang is a professor and dean of the Department of Neurology, The Second Affiliated Hospital of Harbin Medical University.
Xia Li is a professor and head of the Chair in the College of Bioinformatics Science and Technology, Harbin Medical University.

Submitted: 11 July 2019; **Received (in revised form):** 5 August 2019

CDKN2B-AS1 and EGFR-AS1. In addition, cascade responses and a functional crosstalk model among lncRNAs-CNV, impacted genes, driving subpathways and cancer hallmarks were proposed for understanding the driving mechanism of lncRNAs-CNV. Finally, we developed a user-friendly web interface-LncCASE (http://bio-bigdata.hrbmu.edu.cn/LncCASE/) for exploring lncRNAs-CNV and their driving subpathways in various cancer types. Our study identified and systematically characterized lncRNAs-CNV and their driving subpathways and presented valuable resources for investigating the functionalities of non-coding variations and the mechanisms of tumorigenesis.

Key words: lncRNAs; copy number variation; subpathway; cancer

Introduction

Cancer is a multi-factorial and heterogeneous disease that involves a sequence of genetic variations that will perturb downstream molecular networks to control cancer hallmark processes, such as cell cycle, cell growth and apoptosis, and further contribute to the initiation and progression of human cancer. Perturbations of subpathway (structural sub-regions within biological pathway), an important class of molecular network, have been demonstrated to play critical roles in cancer. Identification of genetic variations perturbing subpathways will lead to a greater understanding of cancer pathogenesis.

Genetic variations are considered as driving events of cancer development and progression. In recent years, researchers mainly focused on genetic changes that occurred in protein-coding regions, identifying hundreds of driving genes [1, 2], and developing computational methods for predicting perturbed molecular networks to elucidate the effects of these genetic variations. With the development of next-generation sequencing, a large number of genetic variations in non-coding regions have been discovered. Especially, accumulated studies have demonstrated that genetic variations in long non-coding RNA (lncRNA) regions such as single-nucleotide variations, somatic mutation and copy number variations (CNVs) could contribute to tumorigenesis. For example, a single-nucleotide site variation (rs11655237) in lncRNA LINC00673 can form a binding site for miR-1231, which affects the susceptibility of pancreatic cancer [3]. Pan et al. [4] have identified a mutation in lncRNA GAS8-AS1 as a driving variant of human thyroid cancer. Northcott et al. [5] analyzed the genetic variations of 1000 medulloblastoma samples and revealed that lncRNA PVT1-related structural variants were ubiquitous in different subtypes. In particular, CNVs in lncRNA regions, some important genetic structural variants [6], have been demonstrated to play critical roles in the development and progression of human cancers. The study of Hu et al. [7] revealed that the focally amplified lncRNA FAL1 exhibits oncogenic activity, which is able to activate many cancer-related protein-coding genes (CR-PCGs) such as CDKN1A and p21, and further affect the dysregulation of cancer-related function. However, the functional interpretation for these lncRNAs-CNV on a large scale to understand their roles in tumor development and growth is a challenging task [6].

Based on the notion that non-coding variations could impact the regulation of PCGs [8], many investigators provided important data resources and generated theoretical methods to understand the function of lncRNA genetic variations and the pathogenesis of complex diseases. For example, explanation of the risk non-coding variations using genome-wide association studies was based on PCGs located nearby in the genome [6]. Quantitative trait loci methods were also general strategies for interpreting functions of non-coding variations and their linkages to diseases [9]. However, the downstream perturbed biological functions of these lncRNA genetic variations remain unclear. As cancers can be classified as different types and subtypes [10], it is crucial to investigate the commonalities and differences of functions for these lncRNA genetic variations among various cancer types. More recently, large-scale biomedical data including multidimensional molecular profiles of tumor samples from different tumor types generated by The Cancer Genome Atlas (TCGA) project, lncRNA annotation data from GENCODE project [11], biological molecular interaction networks and pathway data resources provide unprecedented opportunities to uncover the functions of these lncRNA genetic variations on a large scale and to understand pathogenesis of human tumors further.

Here, we developed a computational method to systematically identify lncRNAs-CNV and their driving subpathways by integrating multidimensional molecular profiles of 4802 tumor samples from 14 cancer types. Then, a comprehensive analysis was performed. The properties of these lncRNAs-CNV were characterized and found that lncRNAs-CNV exhibited high specificity in cancers and those with high CNV level may perturb broad biological functions. Some core subpathways widely perturbed by lncRNAs-CNV in pan-cancer were identified. An in-depth analysis of lncRNAs-CNV driving subpathway associations revealed the diverse functional characteristics of lncRNAs-CNV in cancer. Moreover, lncRNAs-CNV that are survival-related were identified as potential oncogenic drivers. Finally, we developed LncCASE, an online database to store and retrieve all lncRNAs-CNV and their driving subpathways in pan-cancer, which is available at http://bio-bigdata.hrbmu.edu.cn/LncCASE/, providing additional information that can facilitate the functional and mechanism studies of lncRNAs-CNV in human cancers.

Materials and methods

Multiple omics data sets for lncRNAs and mRNAs from TCGA

We obtained gene expression data (level 3), copy number data (level 3), mutation data (level 2) as well as clinical data of 4802 tumor patients from these 14 cancer types in TCGA Data Portal (https://tcga-data.nci.nih.gov/tcga). Besides, the expression profiles of lncRNA for each cancer type were downloaded from TANRIC [12]. The detailed sample information in each cancer type at different omics levels and the distribution of their gender, race and ethnicity were shown in Supplementary Table S1 and Supplementary Figure S1. In addition, detailed processes were shown in Supplementary Materials and Methods.

Pathways and human protein–protein interaction network data

The KGML files in KEGG database [13], containing protein–protein interaction (PPI) and biochemical reaction information of 281 pathways, were converted into undirected graphs as we previously described [14].

We obtained the PPI networks from HINT [15] (http://hint.yulab.org) and HPRD [16] (http://www.hprd.org). By combining edges in two databases, a relatively comprehensive PPI network was accomplished. The final network consists of 70,406 unique undirected interactions among 12,207 human proteins.

Collection of cancer hallmark gene sets and cancer lncRNAs

We downloaded cancer hallmark gene sets from Gene Ontology Consortium [17]. According to Plaisier et al. [18] there are 35 GO sets that could be categorized into 10 cancer hallmarks. Cancer lncRNAs were collected from LncRNADisease [19] and Lnc2Cancer [20] databases. In total, 65 cancer lncRNAs were obtained.

Construction of copy number profiles based on re-annotation strategy

We firstly downloaded lncRNA/PCG annotation data from the GENCODE database [11]. We used the copy number re-annotation method that was described previously in Akrami et al [21] to construct copy number profiles of cancer. By mapping the chromosome positions of segments to the genomic annotation data of lncRNAs/PCGs, the copy number amplitudes of the segments were designated to their overlapping lncRNAs/PCGs. Copy number amplitude of each lncRNA/PCG for each patient was determined by choosing the minimum amplitude of all overlapping segments in samples of the patient. LncRNAs/PCGs who have no overlapping segments in more than 50%/80% of patients were deleted. Finally, copy number profiles of lncRNAs across 14 cancer types were constructed.

Identification of lncRNAs-CNV driving subpathways

In order to locate lncRNAs-CNV driving subpathways within pathways, a step-by-step procedure was taken. First, signature PCGs impacted by lncRNAs-CNV were identified. Second, the driving extents of all PCGs in PPI network by lncRNAs-CNV were quantified. Third, the most susceptible subpathways impacted by CNVs of lncRNAs were located. Fourth, the statistical significances of differences between subpathway activities of patients with CNVs in lncRNA regions and patients without CNVs in lncRNA regions were evaluated. LncRNAs-CNV driving at least one subpathway in cancers were defined as functional lncRNAs-CNV. The schematic workflow is shown in Figure 1. The detailed processes for the identification are as follows.

Identification of signature PCGs impacted by lncRNAs-CNV

First, we discretized the copy number amplitudes of lncRNAs. lncRNAs were classified into two groups depending on whether they located on amplification or deletion peaks recognized by inputting segmented copy number data to GISTIC 2.0 [22]. Based on re-annotated copy number profiles, if the lncRNA locates on amplification peaks, we defined that patients whose copy number amplitudes of this lncRNA is >0.1 were assigned to group of patients with CNVs in lncRNA region and other patients were assigned to control group. In contrast, if the lncRNA locates on deletion peaks, patients whose copy number amplitudes of this lncRNA is <-0.1 were assigned to group of patients with CNVs in lncRNA region and other patients were assigned to control group. The set of this threshold referred to default parameters of GISTIC. Therefore, expression profiles were classified into the same two groups. Genes whose expressions were both differently expressed between two groups and correlated with lncRNA's copy number amplitudes, were collected as signature PCGs. A differently expressed gene was required to fulfill two needs: adjusted *P*-value (Benjamini method) calculated with DEGseq package [23] was less than 0.05 and fold change was less than 1/4 or large than 4. While another two requirements of significant correlation were set, including the value of Pearson correlation coefficient less than -0.4 or larger than 0.4 and adjusted *P*-value (Benjamini method) of correlation calculated with WGCNA package [24] less than 0.05. As described above, 14 cancer types were chosen because they contain both copy number data of tumor samples and lncRNAs, which are located on amplification or deletion peaks.

Quantifying the driving extents of PCGs by lncRNAs-CNV based on global diffusion algorithm

In this study, we used the random walk with restart algorithm [25], which simulates a random move from the seed node (s) to their directly interacted neighbors or stay at the current node (s) according to the probability transition matrix that was obtained from the network topology, to quantify the driving extents of PCGs. Next, for each lncRNA-CNV, all signature PCGs were set as seed genes for random walk in integrated PPI network. The formula was as follows:

$$p^{t+1} = \left(1 - x\right)Wp^{t} + xp^{0}$$

where W refers to the adjacency matrix of the PPI network, p^0 is initial probability vector with all seed genes are set as 1 while other genes are set as 0 and p^t is the probability vector at time t. $x \in \left(0, 1\right)$represents the restart probability. We initially set $x = 0.7$, according to a series of previous studies [26–28], some of which have demonstrated the feasibility of x was defined as 0.7.

The final activity scores of PCGs output by random walk algorithm were defined as the extents of the impact they received from the corresponding lncRNA-CNV.

Locating the most susceptible subpathways within pathway

The PCGs with their final activity scores output by random walk algorithm were mapped to 281 pathways. To magnify the difference between pathway genes, we modified the final activity scores as follows:

$$S_n = \frac{10}{-\log\ p_n}$$

where p_n is the probability of gene n at steady-state and S_n is the activity score of gene n. The activity scores of genes were mapped to pathway genes as node weights.

Meanwhile, the intensities of interactions between every two pathway genes were quantified. Here, we calculated Pearson correlations of expression between every two genes, correlation scores were normalized as follows:

$$S_e = 1 - \left|Cor_{ij}\right|$$

where Cor_{ij} refers to expression correlation coefficient between gene i and gene j and S_e is the weight of edge e (make up of gene i and gene j).

Then, PCST algorithm [29] was performed to mine subpathways that contained pathway genes with more node collection and less edge cost, which means the pathway genes located

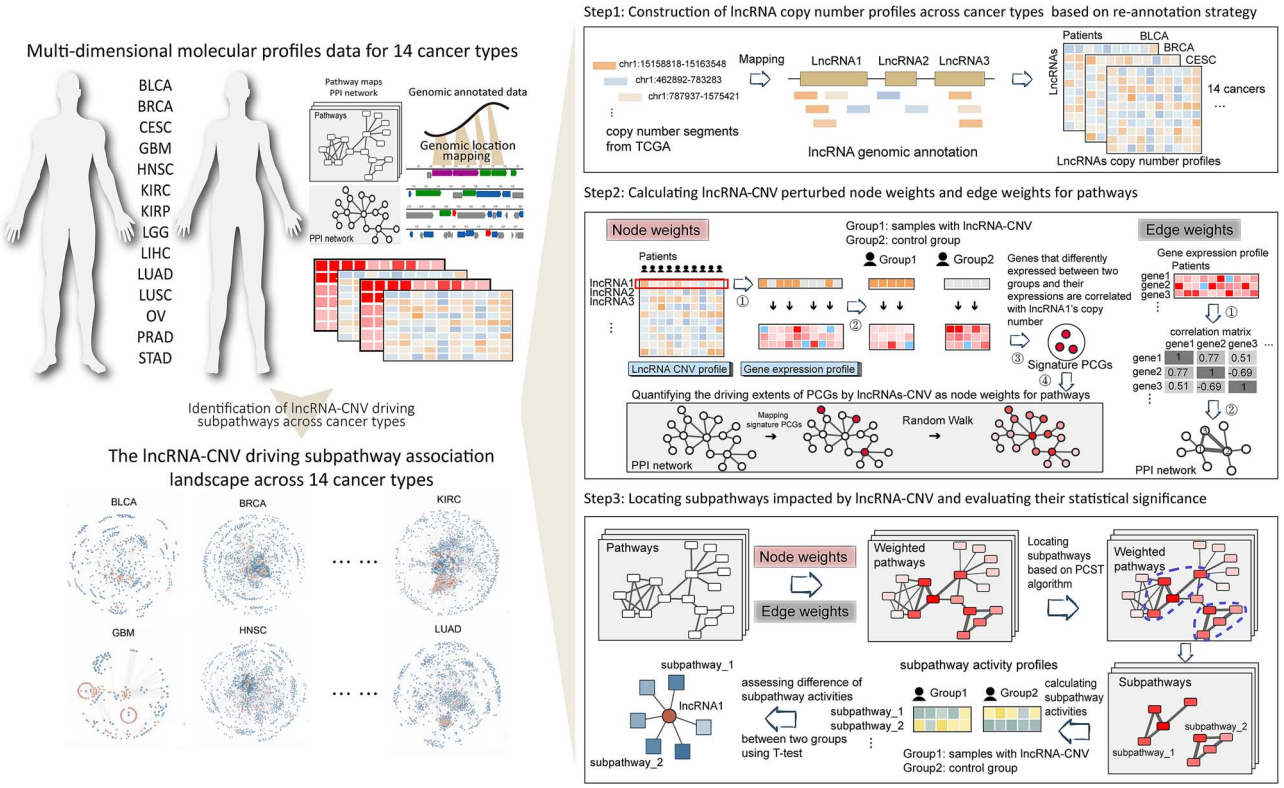

Figure 1. Schematic overview of method for identifying lncRNAs-CNV driving subpathways with transcriptional perturbations.

in subpathways need relatively higher node weights and edge weights, indicating the subpathways were greatly influenced and tightly connected regions impacted by CNVs of lncRNA. The subpathway R' = (N′, E') from overall pathway R = (N, E) was computed as follows:

$$\min_{\left(E',N'\right)connected} E' \in E, N' \in N \sum_{e\in E'} S_e - \sum_{n\in N'} S_n$$

Evaluating the statistical significance of difference of subpathway activities

Finally, for each lncRNA-CNV, we evaluated the activity scores of its driving subpathways in tumor samples and constructed the activity profiles of these subpathways. The subpathways activity score Z was defined by the formula [30]:

$$Z_{km} = \frac{\overline{G_{km}} - Gm}{\sigma_m}\sqrt{c}$$

where $\overline{G_{km}}$ is the averaged expression value of genes in subpathway k in sample m, $\overline{G_m}$ is the averaged expression value of all genes in sample m, σ_m is the standard deviation of expression of all genes in sample m, c is the number of genes in subpathway k.

T-test was applied to assess difference of subpathway activities between group of patients with CNV in lncRNA regions and control group. Subpathways with significantly differential activities (adjusted P-value of t-test by Benjamini method <0.01) were identified as driven subpathways for the corresponding lncRNA-CNV.

Construction of lncRNA-CNV driving subpathway association network

Subpathways may have overlapping genes. To reduce the redundancy of those subpathways, we merged subpathways which have a number of overlapping genes by creating a new subpathway which united all genes and structures in those subpathways. Detailed steps were seen in the Supplementary Materials. Altogether, 5184 new subpathways were obtained. The subpathways for further analysis were new subpathways after merging.

For each cancer, an lncRNA-CNV driving subpathway association network was constructed by connecting all (lncRNA-CNV)-subpathway associations identified above. Consequently, we built 14 cancer type specific lncRNA-CNV driving subpathway association networks and also a comprehensive pan-cancer network. These networks can reflect lncRNAs-CNV driving functional events occurred in the corresponding cancer types, and all lncRNAs in the association networks were functional lncRNAs-CNV.

Results

Evaluation of the method for identifying lncRNAs-CNV and their driving subpathways

To understand the function of genetic variations in lncRNAs, we developed a novel computational method to identify lncRNAs-CNV and their driving transcriptional perturbed subpathways in human cancer. First, by lacking of copy number amplitudes of lncRNAs, we introduced a way to obtain the data according to re-annotation TCGA segmented genomic copy number data. We constructed the lncRNA copy number profiles of 14 cancer types using genome mapping approach based on fragmented copy

number data from TCGA and genome annotation data of lncRNAs from GENCODE. Then, we identified signature genes and evaluated global scores, the measure of lncRNAs-CNV affecting on PCGs, based on network diffusion algorithm. Finally, the most driving local regions within each entire pathway and their significance of differences in subpathway activities were evaluated (Materials and Methods). To evaluate the accuracy of our recalculated copy number amplitudes for lncRNAs, we recalculated copy number amplitudes for PCGs using the same approach as lncRNAs. The correlations between recalculated copy number amplitudes of PCGs and standard copy number profiles generated by GISTIC 2.0 were assessed. Higher correlations were observed comparing with correlations between randomly selected copy number amplitudes for PCGs and standard copy number profiles across all 14 cancer types (Supplementary Figures S2-S3). And most of these correlations (up to 91%) were >0.8. These results suggest the excellent accuracy of the re-annotation strategy in quantifying the copy number amplitudes of lncRNAs.

We also evaluated the feasibility of parameter $x = 0.7$ for quantifying the driving extents of PCGs by lncRNAs-CNV based on random walk with restart method [25]. Firstly, we quantified the driving extents of PCGs by lncRNAs-CNV across all 14 cancer types using the parameter $x = 0.5$ and 0.9, respectively. Then, we re-identified lncRNAs-CNV driving subpathways based on the above result respectively. Finally, we compared subpathways identified for each lncRNA-CNV under $x = 0.7$ with those identified under $x = 0.5$ and 0.9 at entire pathway level, respectively. In general, these subpathways identified under different parameters exhibit a high degree of consistency (Supplementary Figures S4-S5). This indicates that parameter x has slight effects on the result. Thus, we set x as 0.7 according to the selection in the previous studies [26–28].

In order to evaluate the accuracy of our method for identifying subpathways perturbed by lncRNAs-CNV in cancer, we adopted two schemes based on PCGs and lncRNA overexpression data, respectively (Supplementary Materials and Methods). As a result, subpathways driven by genetic variations of CR-PCGs identified by our method have relatively high functional consistency with the gene sets that these CR-PCGs were annotated in GO database, as most of semantic similarity values were higher than 0.6 (Supplementary Figures S6-S7). Furthermore, we used lncRNA PVT1 overexpression dataset to evaluate the accuracy of our method in LIHC (Supplementary Materials and Methods). Semantic similarity values between four PVT1 driving subpathways in LIHC and differential expressed gene set from PVT1 overexpressed dataset were evaluated. Among these, three of four values were higher than 0.6 and significantly higher than random (Supplementary Figure S8). These results validated the high accuracy of our method in identifying subpathways that were driven by lncRNAs-CNV. This suggested our method can accurately characterize the function of lncRNAs-CNV.

The lncRNA-CNV driving subpathway association landscape across 14 cancer types

To systematically analyze and evaluate lncRNAs-CNV and their potential tumorigenic roles, we constructed a landscape which consisted of 14 (lncRNA-CNV)-subpathway association networks (Supplementary Figure S9). In total, 294 696 associations between 3912 lncRNAs-CNV and 5184 subpathways were identified in all 14 cancer types. Among these cancer-specific (lncRNA-CNV)-subpathway association networks, the number of lncRNAs ranged from 50 to 1584, and the number of subpathways ranged from 72 to 1432. The dissection of the degree distribution of these networks found that they all follow the power-law distribution and have scale-free properties (Supplementary Figure S10), which was consistent with the characteristics of most types of biological networks. In summary, these global and previously uncharacterized lncRNA-CNV driving subpathway association networks across diverse tumor types can provide insights into the function of genetic variations of lncRNAs in human cancer.

LncRNAs had high level CNV controlling broad biological functions

We totally identified 3912 functional lncRNAs-CNV driving at least one subpathways across 14 cancer types. Next, we systematically characterized these functional lncRNAs-CNV. They mainly belonged to long intergenic non-coding RNA (lincRNA) and antisense classes (Figure 2A). Dissection of the distribution of these functional lncRNAs-CNV found that they tended to be highly cancer type-specific, as up to 78.4% of these functional lncRNAs-CNV were identified in only one cancer type and only a small subset of them (0.5%) were detected in multiple cancers (>3 cancer types) (Figure 2B and Supplementary Figure S11). The majority of these lncRNAs were observed in KIRC (24.8%) or KIRP (40.5%), which were both originated from the kidney tissue (Figure 2B). In addition, copy number deletion pattern of these functional lncRNAs was widely observed in various cancer types (Figure 2B). We constructed a circular chromosome map to provide a global view of genomic location annotation of each functional lncRNAs-CNV across 14 cancers, and the numbers of cancer types that each lncRNA was implicated with were shown (Figure 2C). We found that most of these lncRNAs distributed in chr1, chr5, chr6, chr7, chr10 and chr14.

LncRNA expression has been widely explored, but an exploration of the effects of genetic variations on lncRNA expression may provide an essential framework for understanding the pathogenesis of tumors. Previous studies have shown that DNA copy number influences gene expression across a wide range of alteration patterns [31, 32]. Also, Kumar et al. [33] found that there is a strong correlation between genetic variations and expression levels of large intergenic non-coding RNAs (lincRNAs) and that is tissue-dependent. Here, we examined the association between copy number amplitudes and expression of these functional lncRNAs-CNV. About 56% lncRNAs-CNV with available expression data have significant correlation between the copy number amplitudes and their expression levels across various cancer types (Figure 2D). Specific topological characteristics could reflect functional features of these lncRNAs-CNV in human cancer. Functional lncRNAs-CNV with higher degrees were more likely to be hubs driving many subpathways and had widespread functions in human cancers. Therefore, the degrees of these functional lncRNAs-CNV were dissected, and lncRNAs with higher degrees in the lncRNA driving subpathway association networks were found to have higher levels of copy number amplitudes (Figure 2E) and lower levels of expression (Figure 2F) in most cancer types. This was consistent with previous studies that decreased gene expression may result from copy number loss [34], which was the most prevalent CNV pattern of these functional lncRNAs. This suggested that lncRNAs with high CNV levels may perturb widespread biological functions. Finally, we mapped experimentally validated cancer associated lncRNAs obtained from reliable databases (Materials and Methods) to the chromosome map, and 65 cancer lncRNAs were found to

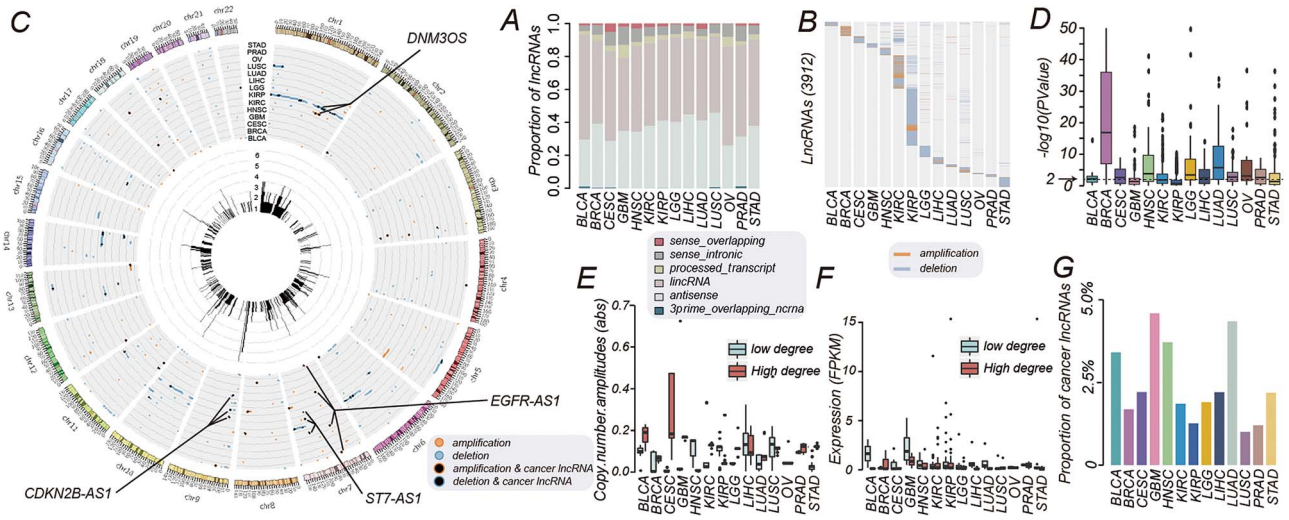

Figure 2. (**A**) Categories of functional lncRNAs-CNV in cancers. (**B**) A global map of functional lncRNAs-CNV in cancers. (**C**) Genomic locations of functional lncRNAs-CNV. Each dot refers to an lncRNA. The blue/orange dots indicate lncRNAs located in copy number deletion/amplification regions, respectively. Cancer lncRNAs are filled in black with stroke color remain unchanged. The dot tracks correspond to different cancer types. The histogram inside shows the number of cancers each lncRNA appears in. (**D**) *P*-values of correlations between expression and copy number amplitudes of functional lncRNAs-CNV. (**E**) Copy number amplitudes of lncRNAs whose degree ranked in bottom 10% (low degree) and top 10% (high degree) in cancer type-specific lncRNA-subpathway association networks. (**F**) Expressions of low degree and high degree lncRNAs in cancer type-specific networks. (**G**) Proportions of cancer lncRNAs in cancers.

be involved (Figure 2C). For example, cancer lncRNAs such as EGFR-AS1, ST7-AS1, CDKN2B-AS1 and DNM3OS are functional lncRNAs-CNV in multiple cancer types (Figure 2C]. By comparing the number of cancer lncRNAs involved in functional lncRNA-CNV set with randomly chosen lncRNAs, we found that functional lncRNAs-CNV tended to be related with cancer (Figure 2G and Supplementary Figures S12-S13). Taken together, the above results suggested that these identified lncRNAs-CNV may play critical roles in human cancer.

Cancer hallmark analysis highlighted core subpathways widely driven by lncRNAs-CNV across cancers

To investigate the functions of lncRNAs-CNV in cancer, we systematically analyzed transcriptional perturbed subpathways driven by these lncRNAs. We found that most subpathways (74%) were cancer specific, while only a small fraction of subpathways were driven by these lncRNAs in multiple cancer types (Figure 3A). Further analysis found that majority of subpathways in CESC, KIRC, OV and LUSC were cancer specific, and cancer conserved subpathways tended to be observed in BLCA, BRCA, GBM, LGG and STAD (Figure 3B).

Although the functional lncRNAs-CNV have widespread and extremely complex functions in various cancer types, cancer-related hallmarks provide a feasible way for understanding remarkable diversity of lncRNAs-CNV driving subpathways and thus uncovering functional roles of lncRNAs-CNV in cancer. Here, we first calculated semantic similarity between subpathways and cancer hallmark-related GO processes to measure their functional similarities (Supplementary Materials and Methods). As a result, 2357 of these subpathways were functionally associated with cancer hallmarks, and majority of them were associated with the hallmarks 'tissue invasion and metastasis', 'self sufficiency in growth signals', 'reprogramming energy metabolism' and 'insensitivity to antigrowth signals' (Supplementary Figure S14). This suggested that the above cancer hallmarks may be universally perturbed by lncRNAs-CNV across diverse types of cancers. We explored the degree distribution of cancer hallmark-associated subpathways in the pan-cancer network. A higher degree represents that the subpathway was driven by more functional lncRNAs-CNV. We found that subpathways associated with cancer hallmarks 'genome instability and mutation' and 'reprogramming energy metabolism' have relative high degrees in the pan-cancer network (Figure 3C). As CNVs in non-coding regions is an important class of genome instability, this observation provided further evidence that 'genome instability and mutation' was a driving factor for the development of cancer. In addition, the 'reprogrammed energy metabolism', especially glucose metabolism, was required to satisfy anabolic demands [35] for uncontrolled growth of cancer cells. It suggested that most of functions impacted by lncRNAs-CNV were related with reprogrammed energy metabolism and thus promoted cancer progression. Nevertheless, subpathways related with the same cancer hallmark exhibit varied degrees across cancers. For example, subpathways related with 'genome instability and mutation' showed high degree in BRCA, KIRP and LGG, while subpathways related with 'reprogramming energy metabolism' showed high degree in KIRC (Figure 3D).

Next, we focused on conserved subpathways, which were commonly impacted by functional lncRNAs-CNV in at least eight cancer types. In total, only 88 (2%) subpathways were conserved across various cancers, and 41 (46.6%) of them were related with cancer hallmarks (Figure 3E). Further analysis found that these 41 core subpathways, which were both conserved and related to cancer hallmarks, were originated from many oncogenic pathways such as cell cycle, mTOR signaling pathway, Notch signaling pathway, PI3K – Akt signaling pathway and Rap1 signaling pathway (Figure 3F). They were mainly related with six cancer hallmarks including 'insensitivity to antigrowth signals', 'limitless replicative potential', 'reprogramming energy metabolism', 'self sufficiency in growth signals', 'sustained angiogenesis' and

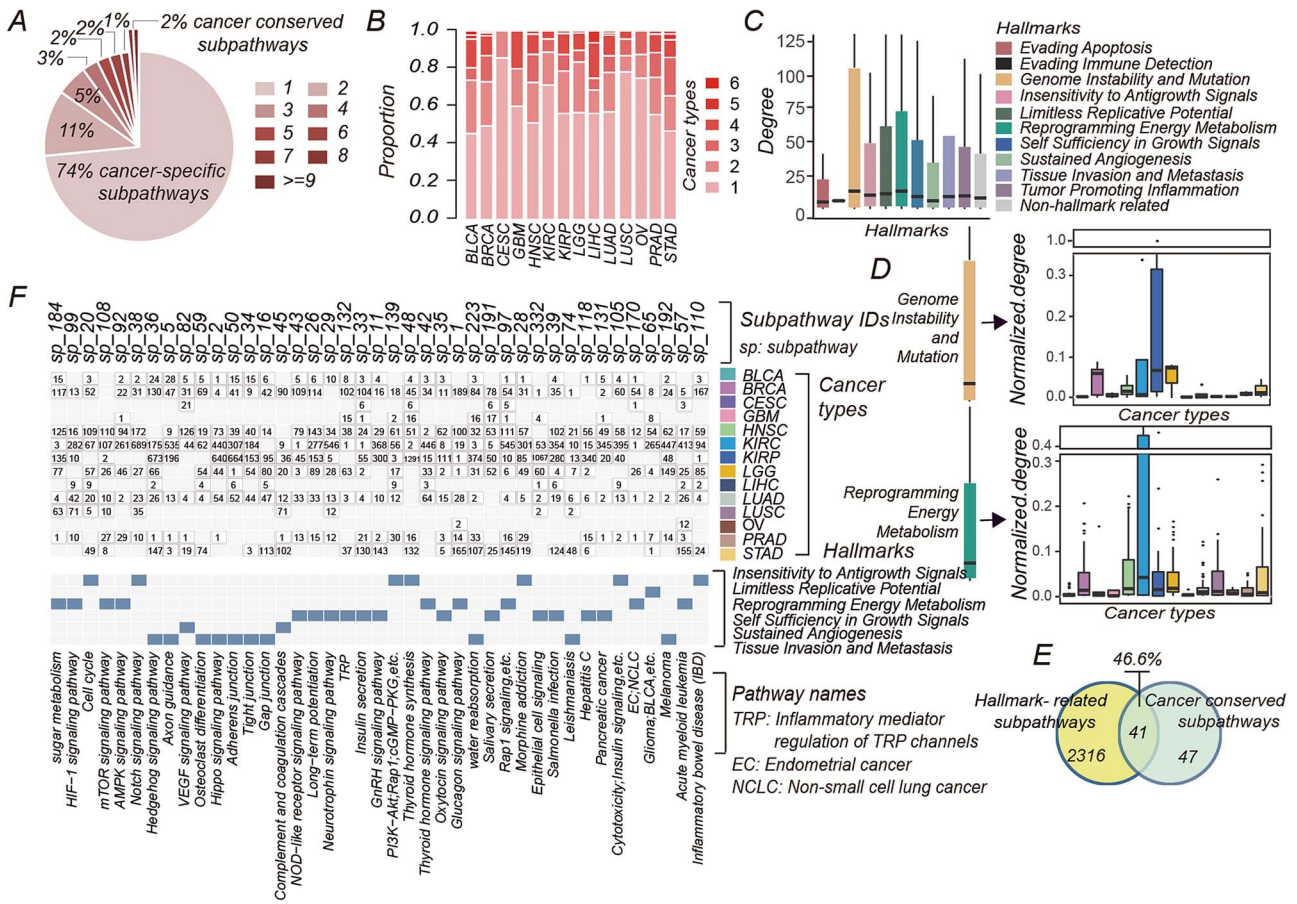

Figure 3. (**A**) Distributions of all subpathways in cancers. (**B**) Distributions of top 50% high-degree subpathways in cancers. (**C**) Degrees of cancer hallmark-related subpathways in pan-cancer lncRNA-CNV driving subpathway association network. (**D**) Normalized degrees of two cancer hallmarks-related subpathways in cancer type-specific networks. (**E**) Venn diagram of cancer hallmark-related subpathways and cancer conserved subpathways. (**F**) The 41 core subpathways with the number of lncRNAs-CNV in 14 cancer types (top), their related cancer hallmarks (middle) and corresponding pathway names (bottom).

'tissue invasion and metastasis' (Figure 3F). This indicated that many cancer specific lncRNAs-CNV may perturb common subpathway regions and thus impact these cancer hallmarks across various cancer types. In addition, our analysis also highlighted the pivot roles of these 41 core subpathways which were widely driven by lncRNAs-CNV in cancers.

Overall, the above analysis revealed that functional lncRNAs-CNV may drive dysregulation of cancer hallmarks and thus contribute to the initiation and progression of cancer.

Dissecting lncRNA-CNV driving subpathway associations revealed the functional diversity of lncRNAs-CNV across cancers

In the above analysis, we found that there were small subsets of common lncRNAs-CNV identified in more than one cancer type. An inspection of associations of lncRNAs-CNV in multiple cancer types showed that these lncRNAs rarely impacted the same subpathways in different cancers (Figure 4A). This indicated that these common lncRNAs-CNV in different cancer types may display varied functions because of the tissue-specific properties for lncRNA/gene expressions and molecular interactions. In order to reveal the functional differences of lncRNAs-CNV more precisely in cancer, we examined whether common lncRNAs-CNV could drive different subpathway regions of the same entire pathway in different cancer types. Firstly, we focused on functional lncRNAs-CNV identified in at least four cancer types and their perturbed pathways. In total, 18 lncRNAs-CNV driving 219 entire pathways that constituted 2434 lncRNA-pathway associations were dissected. We found that most of these associations were cancer specific (Figure 4B).

Next, to exemplify how these common lncRNAs-CNV impacted different functions, cancer lncRNA EGFR-AS1 that presented in the most cancer types was examined. EGFR-AS1 functioned in four tumors including BLCA, STAD, HNSC and LUAD. It was worth to note that there was hardly any (only one) common subpathway driven by EGFR-AS1 shared by different tumors (Figure 4C). In addition, most subpathways were driven by EGFR-AS1 in LUAD (Figure 4C), which indicated that CNV of EGFR-AS1 may perturb a wide range of biological functions in this cancer type. We then dissected the function of EGFR-AS1 across four cancer types. A total of 235 subpathways were driven by EGFR-AS1 totally corresponding to 134 entire pathways, 5 (the least) of which were disturbed in BLCA and 111 (the most) of which were disturbed in LUAD (Figure 4D). A total of 135 (57.4%) of these pathways were associated with six cancer hallmarks such as 'tissue invasion and metastasis' (Figure 4E). 'Tissue invasion and metastasis' was a common associated hallmark for EGFR-AS1 in four cancer types and

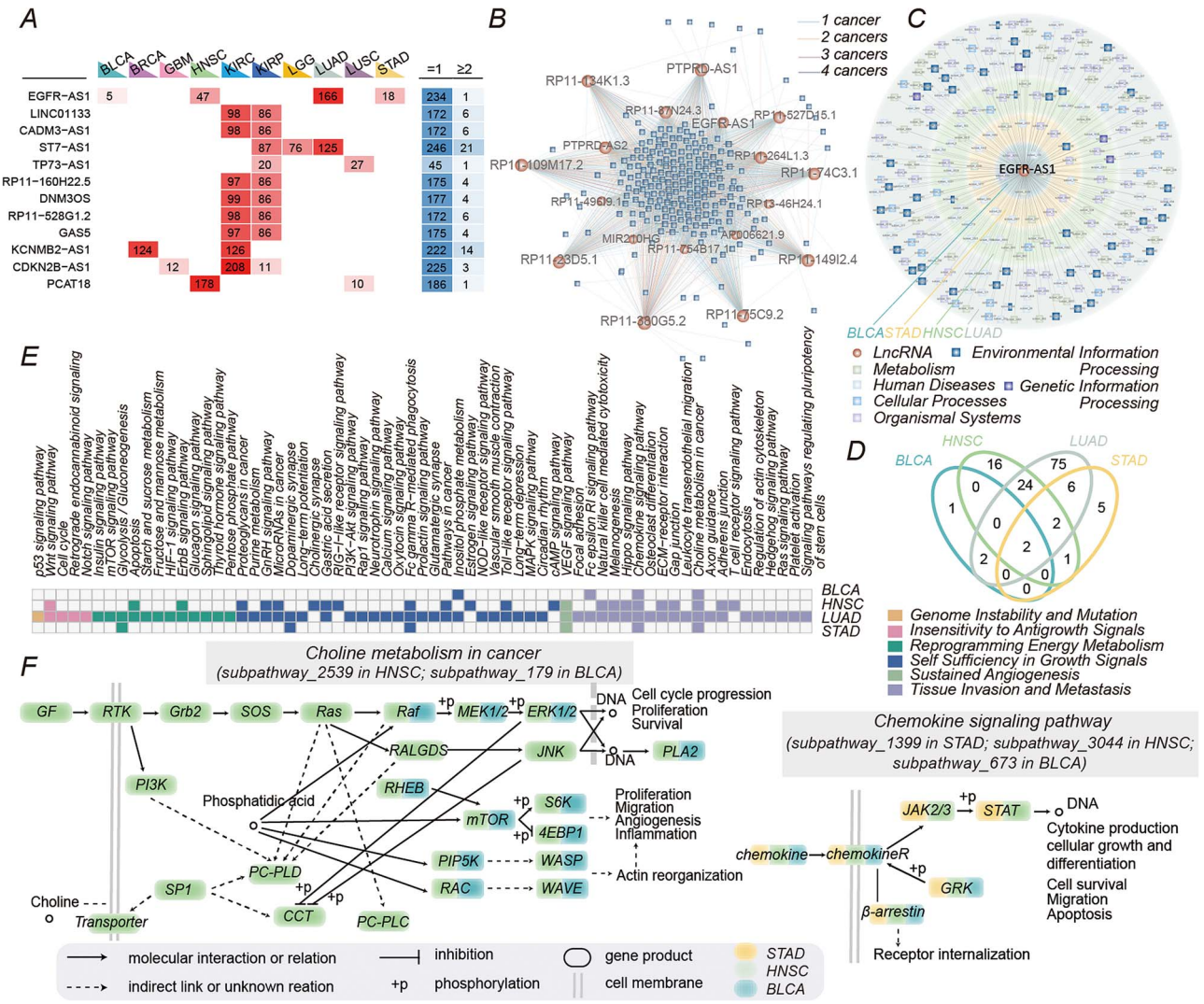

Figure 4. (**A**) The number of related subpathways of each lncRNA-CNV (left) and the number of common associations across cancers (=1 or ≥2) (right). Only cancer lncRNAs whose driving subpathways replicated in more than one cancer type were shown. (**B**) lncRNA-CNV driving an entire pathway network. The network includes functional lncRNAs-CNV and their driving subpathways corresponding pathways identified in four or more cancer types. The edge color represents the number of cancer types that the associations presented in. (**C**) EGFR-AS1-subpathway driving association network among four cancer types. The fill colors of subpathways refer to the categories of their corresponding pathways. (**D**) Venn diagram of EGFR-AS1 associated subpathways corresponding pathways among four cancer types. (**E**) The corresponding pathways of EGFR-AS1 associated subpathways and their related cancer hallmarks in four cancer types. (**F**) Subpathways structures derived from two pathways overlapped among four cancer types show in (**D**).

indicated the roles for promoting tumor development of CNV of EGFR-AS1 across cancers. EGFR-AS1 has been demonstrated to promote cell growth and metastasis in renal cancer [36]. More importantly, 'choline metabolism in cancer' and 'chemokine signaling pathway' were both commonly known cancer-related pathways that were simultaneously driven by EGFR-AS1 across all four cancer types (Figure 4D). Further dissection of the two pathways found that EGFR-AS1 perturbed different subpathways in various cancer types, and these subpathways were all related with critical cancer biological processes such as cell cycle progression, proliferation, cellular growth and differentiation, apoptosis, angiogenesis and migration (Figure 4F). In particular, two subpathways, subpathway_2539 and subpathway_179 within 'choline metabolism in cancer pathway', were driven by EGFR-AS1 in HNSC and BLCA, respectively. The subpathway (subpathway_179) disturbed in BLCA was a downstream region of that in HNSC. In addition, subpathways within 'chemokine signaling pathway' driven by EGFR-AS1 in STAD, HNSC and BLCA, respectively, were also shown (Figure 4F). The above analysis further revealed the functional diversity of lncRNAs-CNV across different cancer types and also indicated that our proposed method had the ability to depict functions of these lncRNAs-CNV at more precise level.

Functional lncRNAs-CNV could be potential biomarkers for cancer prognosis

The above analyses revealed that lncRNAs-CNV play diverse roles in different cancer types and even subtypes. This highlighted that they may serve as promising biomarkers to classify subtypes with different clinical outcomes. To assess the clinical relevance of these lncRNAs, we integrated the clinical data and then performed survival analysis for each

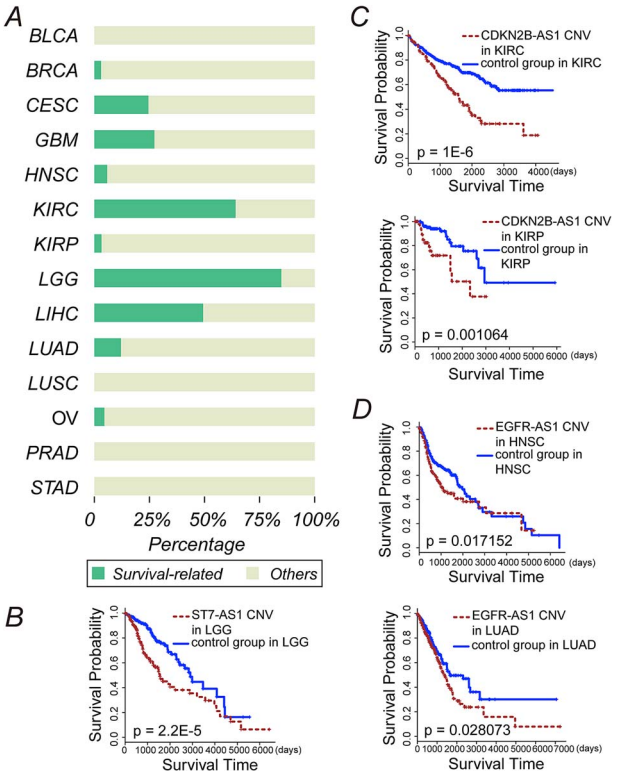

Figure 5. (**A**) The percentages of survival-related functional lncRNAs-CNV in 14 cancers. (**B–D**) K-M plots of CNV samples and control group samples for ST7-AS1, CNKN2B-AS1 and EGFR-AS1 in various cancer types, respectively.

of them across 14 cancer types (Supplementary Materials and Methods). As a result, we totally identified 1093 (28%) survival-related functional lncRNAs-CNV in all 14 cancer types, and most of them were discovered in LGG, KIRC and LIHC (Figure 5A). The proportions of survival-associated lncRNAs-CNV exhibited large difference even in cancer subtypes from the same tissue origin. Previous research showed that the copy number and expression of lncRNA FAL1 were correlated with clinical outcome in ovarian cancer [7]. Here, we also examined the expression correlation of these survival-associated lncRNAs-CNV and categorized them into six groups (Supplementary Figure S15). About 23% lncRNAs-CNV, which are significantly related to survival, have significant correlation between their expression and CNV. Detailed information in 14 cancer types, respectively, was shown in Supplementary Figure S15, indicating that some lncRNAs-CNV contribute to tumorigenesis and might affect their expressions.

Further exploration of these survival-associated lncRNAs-CNV found that some experimentally validated cancer lncRNAs were also included. For example, ST7-AS1 was identified as a survival-associated lncRNA in LGG (Figure 5B) and has been demonstrated to be differentially expressed in apoptosis glioma cells induced by agents [37]. CDKN2B-AS1 has been confirmed to play a critical role in the pathological processes of multiple cancers [38, 39]. Specifically, the CNV status of CDKN2B-AS1 could distinguish patients into two groups with different clinical outcomes in both kidney tumor subtypes including KIRC and KIRP (Figure 5C). Furthermore, another cancer lncRNA EGFR-AS1 was significantly prognostic related with the two cancer types including LUAD and HNSC, which were originated from two distinct tissues (Figure 5D). Interestingly, we also found that the three cancer lncRNAs-CNV mentioned above were consistently associated with poor prognosis of patients in the corresponding cancer types. This observation provided further evidence for conclusion that functional lncRNAs-CNV were important drivers to promote the development of cancer. Taken together, our analysis indicated that the driving roles of functional lncRNAs-CNV and their potential clinical usages as prognosis biomarkers in cancer.

(LncRNA-CNV)-gene-subpathway-cancer hallmark cascade responses and a functional crosstalk model for elucidating their driving effect in cancer

Genetic variations in lncRNA regions such as CNVs could induce the expressions alteration, then disturb corresponding subpathways and cause dysregulation of the cancer hallmark related processes. Consequently, the initiation and progression processes of tumor were activated. To explore how these functional lncRNAs-CNV contribute to the pathogenesis of cancer, a series of (lncRNA-CNV)-gene-subpathway-cancer hallmark cascade responses were proposed, and several cancer lncRNAs-CNV were further analyzed. LncRNA ST7-AS1 with CNVs driving dysregulation of downstream cancer genes such as CDK6, MET, MMP9 and CFTR and further disturbed transcriptional activities of subpathway_145 that was located in cAMP signaling pathway and associated with 'tissue invasion and metastasis' in LGG and LUAD (Figure 6A). It was worthy to note that ST7-AS1 has been demonstrated to be closely associated with human glioma [37]. The impacted genes downstream were also associated with the development of glioma, especially for invasion and metastasis. For example, it has been demonstrated that CDK6 can be regulated by lncRNAs and impact proliferation, invasion and migration of glioma cells [40]. MET is a cancer gene associated with the proliferation and invasion in glioma [41]. Up-regulation of MMP9 could promote glioma cell migration and invasion [42]. In addition, ST7-AS1 perturbed subpathway_145 that was also closely related with the biological processes of cancer including proliferation, apoptosis and cell migration (Figure 6B and Supplementary Figure S16). These analyses suggested that lncRNA-CNV ST7-AS1 is likely to play important roles in cancer progression, especially for the invasion and metastasis. Similarly, lncRNA CDKN2B-AS1 may drive subpathway_126 in GBM and KIRC (Figure 6A). Subpathway_126 located within chronic myeloid leukemia pathway and associated with reprogramming energy metabolism hallmark (Figure 6A), impacted critical processes including cell cycle progression, proliferation and survival (Figure 6C and Supplementary Figure S17). CDKN2B-AS1 has been experimentally validated to influence cell proliferation, invasion and migration of human glioma cells [43] with help of energy metabolism reorganization. These above analyses can provide novel insights to understand the functions of lncRNAs-CNV and the pathogenesis of cancer based on these lncRNAs-CNV driving subpathway associations identified by our method.

To further explore the roles of lncRNAs-CNV in cancer, we then dissected the genetic variation association pattern between lncRNAs-CNV and genes within their driving subpathways. A general analysis of genetic variation heatmap (Figure 6D) revealed that in some cases, the genetic variations of lncRNAs and genes within their driving subpathway appeared to be mutually exclusive. This indicated that the development of cancer can be driven by either the genetic variations of lncRNAs or their functionally associated

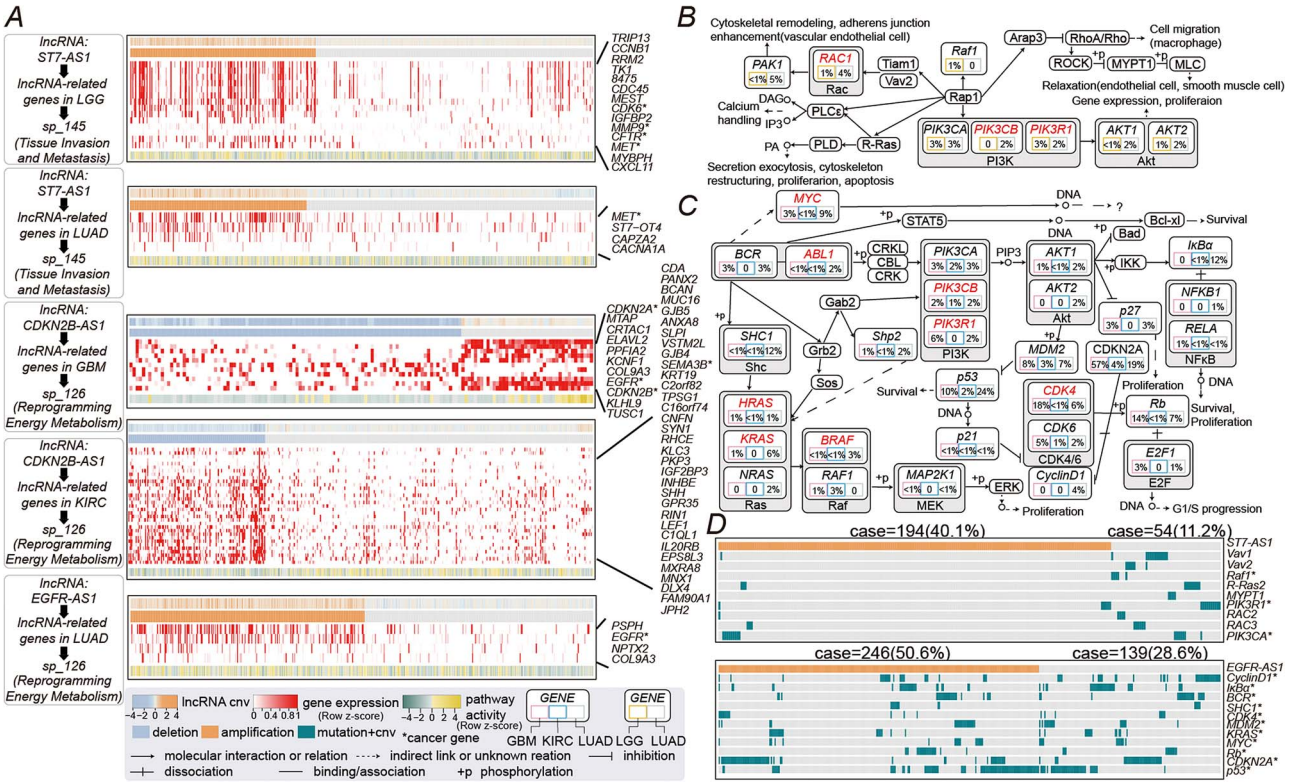

Figure 6. (**A**) Five (lncRNA-CNV)-gene-subpathway-cancer hallmark cascade responses. Heat map combinations include information about copy number of five cancer lncRNA, expression of lncRNA-related genes and subpathway activity. Columns represent patients in corresponding cancer types: the first row of each combination represents copy number amplitude of lncRNA and the second row represents discretized copy number data of lncRNA. Beneath subpathway names are their related cancer hallmark names. (**B**) Subpathway structure of subpathway_145 consistent with layout in KEGG database. CR-PCGs are in italics, with the percentages of CNVs or somatic mutations in LGG and LUAD, respectively. Genes in red are drug targets. (**C**) Subpathway structure of subpathway_126. (**D**) Columns in heat maps represent patients in corresponding cancer types: the first row of each heat map represents discretized copy number data of lncRNA, while other rows in green represent CNVs or somatic mutations of selected genes of subpathways. Patients without any genomic alterations are not shown.

genes and also supported that the functional associations between lncRNAs-CNV and their driving subpathways do exist.

As most lncRNAs-CNV drove different subpathways, we also found that there were lncRNAs (e.g., CDKN2B-AS1 and EGFR-AS1) disturbing the same subpathways in cancers (Figure 6A). This indicated that these lncRNAs-CNV may function in a cooperative pattern to drive transcriptional perturbations of subpathways and further the dysregulation of cancer hallmarks in the complex biological systems. Thus, we here proposed a cancer hallmark related functional crosstalk model for systematically understanding the driving function of lncRNAs-CNV in human cancers (Figure 7). Notably, cancer hallmark related subpathways driven by CDKN2B-AS1, LINC00942, C1QTNF9B-AS1 and DNM3OS, respectively, had functional crosstalk among the internal and external subpathway sets in cancers (Figure 7). This suggested that these four lncRNAs-CNV may function as synergistic drivers in KIRC. The above analysis indicated that our proposed cascade response and functional crosstalk model can provide insights into how intermolecular and functional relationships dictate tumorigenesis. Taken together, identification and exploration of the genetic variation lncRNAs and driving subpathways are important for understanding the function of lncRNAs, which facilitate the comprehension of pathological mechanism and development of lncRNA-based therapy of cancer.

LncCASE: a web interface for exploring lncRNAs-CNV and their driving transcriptional perturbed subpathways across cancer types

To facilitate the usage of (lncRNA-CNV)-subpathway association resource, we developed LncCASE (LncRNAs with Copy number Alteration affecting Subpathways in cancErs) (http://bio-bigdata.hrbmu.edu.cn/LncCASE/), an online database which collected all significantly dysregulated (adjusted $P < 0.05$) subpathways before merging with their associated lncRNAs-CNV across cancer types. In total, LncCASE documents 566 425 entries of associations between 4115 lncRNAs-CNV and 17 455 subpathways among 14 cancer types. Except for the information about names/IDs of cancer type, lncRNA and subpathway, each entry consists of pathway that the subpathway derived from, the number of genes included in the subpathway, the statistics for estimating difference of the subpathway activities and visualization of the subpathway structure.

LncCASE provides a user-friendly interface. The quick search enable users to filter entries with one keyword of interest, such as a lncRNA (name or ensembl ID), cancer type (full name or abbreviation) or pathway name. Also, an advanced search is provided in 'Search' page for more specific requirements. The users can input interested lncRNA, cancer type and pathway at the same time to obtain desired associations. In 'Browse' page, all lncRNAs, cancer types and pathways in LncCASE are arranged

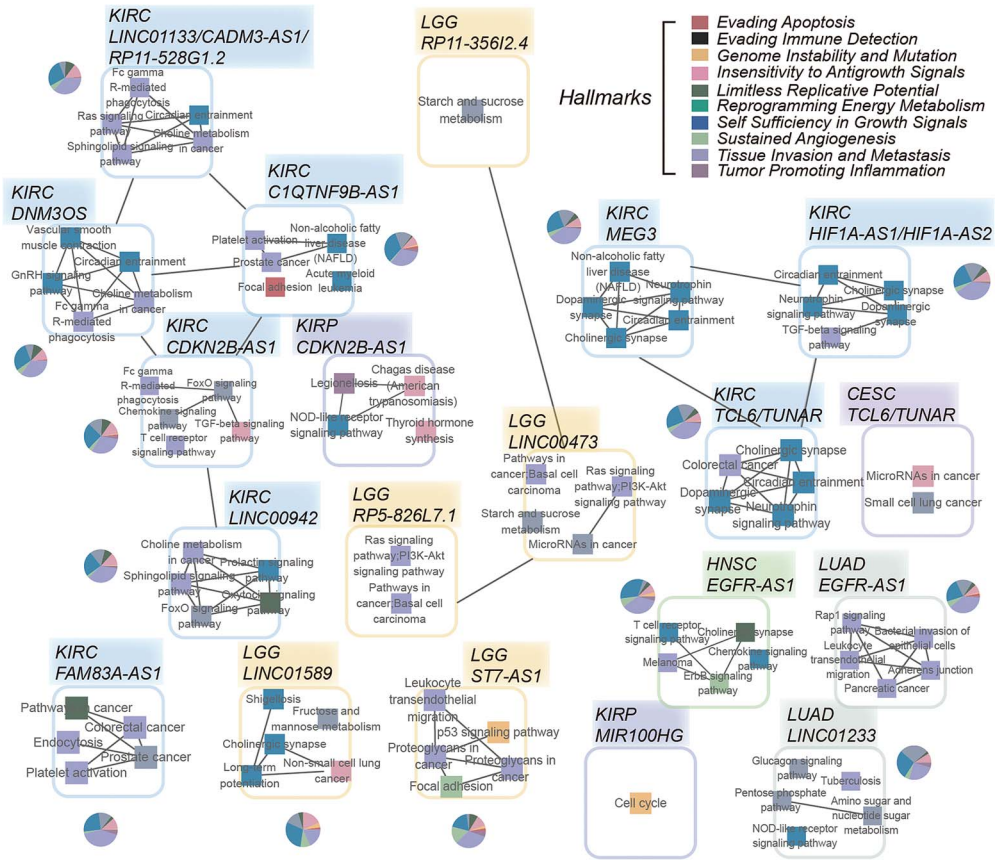

Figure 7. A cancer hallmark-related functional crosstalk network model among lncRNAs and subpathways across cancers. Only cancer lncRNAs with significant survival *P*-value ($P < 0.05$) in specific cancer type and subpathways, which have related hallmarks, are shown. Connection between lncRNAs is defined that their driving subpathway clusters have significant overlaps compared with two randomly selected subpathway clusters but with same numbers of genes respectively. Subpathways with significant overlapped genes (hypergeometric test: *P* adjust <0.01) are considered to have crosstalk. Inside square frames are top five most active subpathways impacted by lncRNA, names of subpathways in picture are their corresponding pathway names. Beside the frames are pie charts that show hallmark distributions of all lncRNA associated subpathways.

in order, respectively, for users to query. Besides, subpathways are listed under each pathway entry they derived from. The search and browse results can be freely downloaded. Furthermore, 'Help' page contains detailed guidance for users.

With the increasing amount of lncRNA and multi-dimensional molecular profiles of cancer cohorts, we will continuously update the website in terms of the coverage of lncRNAs, the types of genetic variation, the number of cancer types, etc. In addition, to further facilitate the utility of the website, we will consider to improve the subpathway visualization function and to add genomic visualization tools in the future.

Conclusion and discussion

Recently, genetic variations in lncRNAs have attracted more and more attention, as they have been demonstrated to play critical roles in the development of human cancers. However, interpreting the functions of these genetic variation lncRNAs on a large scale is still a challenge. In this study, a systematical identification and characterization of lncRNAs-CNV and the driving subpathways were performed, providing a comprehensive resource for studying the functions of genetic variation lncRNAs and the pathogenesis of human cancer.

We firstly provided a computational method to identify lncRNAs-CNV driving subpathways by integrating genomic and transcriptomic data in human cancer. The copy number level of lncRNAs was re-annotated to construct lncRNA-CNV profiles. And the high reliability and accuracy of our method in identifying lncRNAs-CNV and their driving subpathways were demonstrated. Furthermore, our method could also be applied to other types of genetic variation lncRNAs.

Then, the method was applied to 14 cancer types, and a comprehensive characterization of these lncRNAs-CNV and their driving subpathways were performed. Dissecting global properties of functional lncRNAs-CNV found that they were cancer specific. The copy number amplitudes of lncRNAs-CNV were significantly associated with their expression in cancers, which provides further evidence for the functionality of our identified lncRNAs-CNV at expression level. Dissecting the network topology properties of lncRNAs-CNV revealed that lncRNAs have high-level CNV controlling broad biological functions. Moreover, we also found that functional lncRNAs-CNV tended to be related with cancer. These results revealed lncRNAs-CNV played critical roles in cancer, which highlighted the importance of function prediction and analysis for them. The analysis of subpathways driven by lncRNAs-CNV across cancer types discovered some core subpathways widely driven by lncRNAs-CNV in different cancer types. These functional regions may help to yield potential drug target candidates. Further analyses of the associations between

lncRNAs-CNV and subpathways uncovered the functional diversity of lncRNAs-CNV across cancers and also demonstrated the advantage of subpathway for functional analysis of lncRNAs-CNV. Survival analysis of functional lncRNAs-CNV highlighted their potential for clinical usages as prognostic biomarkers and suggested their driving roles in cancer. These analyses can help to elucidate the functions of lncRNAs-CNV and will deepen our understanding of the pathogenesis of cancer.

In recent years, several methods have been proposed to help researchers for investigating the functions of lncRNAs. For example, Jiang et al. [44] predicted the functions of lncRNAs based on their co-expressed PCGs using gene set enrichment analysis method. Zhou et al. [45] developed 'LncFunNet' method to integrate ChIP-seq, CLIP-seq and RNA-seq data to predict and annotate the function of lncRNAs in mouse skeletal muscle cells. However, these strategies were mainly based on lncRNA expression. There is still a lack of effective method for the functional interpretation of lncRNAs with genetic variations. To fill this gap, we developed a novel computational approach, which mainly focused on lncRNAs-CNV. Our study has some unique aspects. First, multiple-omics data were integrated and the topological structures of pathways were also considered in our method. Second, our study focused on subpathways that can help us explain the functions of lncRNAs-CNV from a more precise and in-depth perspective. Third, we also provided a landscape of lncRNAs-CNV and their driving subpathways across 14 human cancer types. In all, we not only proposed an effective method, but also performed a systematic analysis, uncovering some important knowledge about genetic variation lncRNAs that deepened our understanding of their roles and the pathological mechanism in cancer.

We should also point out that our method can be improved in the following aspects. First, it only focused on CNVs in lncRNA regions. Adding other types of genetic variation could improve the applicability of method. LncRNA has a complex spatial structure, and the mechanism involved in expression regulation is diverse and complex. In future works, taking into account the structure and integrating other types of sequencing data such as ChIP-seq and CLIP-Seq would provide better understanding of the complex mechanism and a more comprehensive prediction of the functions of lncRNAs-CNV. Furthermore, individuals from more cancer types and races will be integrated and analyzed in our future research.

In this study, we identified the driving subpathways of each lncRNA-CNV based on the notion that lncRNA with genetic variations may alter gene expression and further disturb the activities of related pathways [a cascade response: (lncRNA-CNV)-gene-subpathway], providing explanations of the functions and driving mechanisms for lncRNAs-CNV. Furthermore, we combined all cascade responses into a cancer hallmark-related functional crosstalk framework, as the development of cancer is caused by coordinated cascade responses of multiple lncRNAs-CNV and crosstalk among their perturbations subpathways, to systematically understand the functions of lncRNAs-CNV and the pathogenesis of cancer. Finally, a free, web-accessible database called LncCASE (http://bio-bigdata.hrbmu.edu.cn/LncCASE/), which stores all lncRNAs-CNV and their driving subpathways identified in our study and also provides visualization of subpathway structures, is presented to facilitate the researches of cancer biology.

In addition, genomic imprinting is required for normal development, and the disturbance of which also play important roles in tumor [46, 47]. Experiments have confirmed that lncRNA participates in the process of imprinting genes with a variety of modes of action including promotion of chromatin compartmentalization, transcriptional occlusion and collision-based mechanisms, etc [48]. Furthermore, a well-known lncRNA H19 has been demonstrated to regulate the expression of members in an imprinted gene network including 16 co-expressed imprinted genes [49, 50]. CNV in the lncRNA region may affect the expression of lncRNA molecules and then regulate the expression of downstream genes from multiple different modes of action, and thus could participate in the process of gene imprinting. Here, we explored the imprinted gene process influenced by lncRNAs-CNV. Firstly, we obtained the curated human imprinting genes from the MetaImprint database [51] and then identified lncRNAs-CNV driving subpathways that were enriched with imprinting genes using Hypergeometric test ($P < 0.05$). Totally, 2777 lncRNAs-CNV were related with the processes of gene imprinting, including H19. We found that the CNV level of H19 was significantly correlated with its expression, and H19-CNV regulated two gene imprinting related subpathways including subpathway_95 (a merged region within inflammatory mediator regulation of TRP channels, GnRH signaling pathway, salivary secretion and bile secretion pathways) and subpathway_2879 (a region within SNARE interactions in vesicular transport pathway) in HNSC. Based on our method, H19-CNV could impact the expression of its driving signature gene TRIM21 and thus disturb these two subpathways. Further exploration of the driving cascade found that expression of H19 could regulate the expression of transcription factor E2F1 [52, 53] in cancer. Interestingly, TRIM21 is also the predicted downstream target of E2F1 from the harmonizome database [54]. Thus, we could infer that the potential mechanism for H19-CNV involved in the process of gene imprinting is that: the CNV of H19 impact its expression, then regulate the expression of E2F1 and its downstream target gene (s), and further influence the process of gene imprinting (Supplementary Figure S18). The above analysis indicates that our method could provide guidance for further disclosure of the modes of action of lncRNAs-CNV in cancer.

In summary, we presented an integrative computational method to systematically identify widespread lncRNAs-CNV and their driving subpathways with transcriptional perturbations across cancer types. The proposed method and analyses extended the existing knowledge of lncRNAs and provided a new perspective to investigate functions of non-coding genetic variations, which can help to reveal the mechanism of tumorigenesis and discovery new therapeutic targets of cancers.

Key Points

- This study provided a strategy to identify lncRNAs-CNV and their driving transcriptional perturbed subpathways based on multi-omics data of cancer and the lncRNA-CNV driving subpathway association landscape in pan-cancer was constructed.
- A comprehensive characterization and analysis of these lncRNAs-CNV revealed their high specificity in cancers and highlighted their potential for clinical usage as prognostic biomarkers.
- Some core subpathways and cancer hallmarks widely perturbed by lncRNAs-CNV were revealed.
- Dissecting lncRNA-CNV driving subpathway associations revealed the functional diversity of lncRNAs-CNV across cancers.

- Cascade responses and a functional crosstalk model were provided to understand the driving mechanism of lncRNA-CNV. And a user-friendly web resource to explore associations between lncRNAs-CNV and their driving subpathways in cancers was constructed.

Supplementary Data

Supplementary data are available online at http://bib.oxfordjournals.org/

Funding

This work was supported by the National Natural Science Foundation of China (grant nos. 61873075, 31801107, 61603116 and 31701145), the National Key R&D Program of China (2018YFC2000100) and the Fundamental Research Funds for the Provincial Universities.

References

1. Rheinbay E, Parasuraman P, Grimsby J, *et al*. Recurrent and functional regulatory mutations in breast cancer. *Nature* 2017;**547**:55–60.
2. Bailey MH, Tokheim C, Porta-Pardo E, *et al*. Comprehensive characterization of cancer driver genes and mutations. *Cell* 2018;**174**:1034–5.
3. Zheng J, Huang X, Tan W, *et al*. Pancreatic cancer risk variant in LINC00673 creates a mi R-1231 binding site and interferes with PTPN11 degradation. *Nat Genet* 2016;**48**:747–57.
4. Pan W, Zhou L, Ge M, *et al*. Whole exome sequencing identifies lnc RNA GAS8-AS1 and LPAR4 as novel papillary thyroid carcinoma driver alternations. *Hum Mol Genet* 2016;**25**:1875–84.
5. Northcott PA, Shih DJ, Peacock J, *et al*. Subgroup-specific structural variation across 1, 000 medulloblastoma genomes. *Nature* 2012;**488**:49–56.
6. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet* 2015;**24**:R102–10.
7. Hu X, Feng Y, Zhang D, *et al*. A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p 21 expression in cancer. *Cancer Cell* 2014;**26**:344–57.
8. Khurana E, Fu Y, Chakravarty D, *et al*. Role of non-coding sequence variants in cancer. *Nat Rev Genet* 2016;**17**:93–108.
9. Pai AA, Pritchard JK, Gilad Y. The genetic and mechanistic basis for variation in gene regulation. *PLoS Genet* 2015;**11**: e1004857.
10. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* 2000;**100**:57–70.
11. Harrow J, Frankish A, Gonzalez JM, *et al*. GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res* 2012;**22**:1760–74.
12. Li J, Han L, Roebuck P, *et al*. TANRIC: an interactive open platform to explore the function of lnc RNAs in cancer. *Cancer Res* 2015;**75**:3728–37.
13. Kanehisa M, Goto S, Hattori M, *et al*. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 2006;**34**:D354–7.
14. Li C, Li X, Miao Y, *et al*. Subpathway Miner: a software package for flexible identification of pathways. *Nucleic Acids Res* 2009;**37**:e131.
15. Das J, Yu H. HINT: high-quality protein interactomes and their applications in understanding human disease. *BMC Syst Biol* 2012;**6**:92.
16. Keshava Prasad TS, Goel R, Kandasamy K, *et al*. Human protein reference database—2009 update. *Nucleic Acids Res* 2009;**37**:D767–72.
17. Ashburner M, Ball CA, Blake JA, *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;**25**:25–9.
18. Plaisier CL, Pan M, Baliga NS. A mi RNA-regulatory network explains how dysregulated mi RNAs perturb oncogenic processes across diverse cancers. *Genome Res* 2012;**22**:2302–14.
19. Bao Z, Yang Z, Huang Z, *et al*. Lnc RNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res* 2019;**47**:D1034–D 1037.
20. Gao Y, Wang P, Wang Y, *et al*. Lnc 2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers. *Nucleic Acids Res* 2019;**47**:D1028–D 1033.
21. Akrami R, Jacobsen A, Hoell J, *et al*. Comprehensive analysis of long non-coding RNAs in ovarian cancer reveals global patterns and targeted DNA amplification. *PLoS One* 2013;**8**: e80306.
22. Mermel CH, Schumacher SE, Hill B, *et al*. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* 2011;**12**:R41.
23. Wang L, Feng Z, Wang X, *et al*. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 2010;**26**:136–8.
24. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;**9**:559.
25. Kohler S, Bauer S, Horn D, *et al*. Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet* 2008;**82**:949–58.
26. Lv Y, Wang S, Meng F, *et al*. Identifying novel associations between small molecules and mi RNAs based on integrated molecular networks. *Bioinformatics* 2015;**31**:3638–44.
27. Liu W, Li C, Xu Y, *et al*. Topologically inferring risk-active pathways toward precise cancer classification by directed random walk. *Bioinformatics* 2013;**29**:2169–77.
28. Li Y, Patra JC. Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network. *Bioinformatics* 2010;**26**:1219–24.
29. Bailly-Bechet M, Borgs C, Braunstein A, *et al*. Finding undetected protein associations in cell signaling by belief propagation. *Proc Natl Acad Sci U S A* 2011;**108**:882–7.
30. Levine DM, Haynor DR, Castle JC, *et al*. Pathway and gene-set activation measurement from mRNA expression data: the tissue distribution of human pathways. *Genome Biol* 2006;**7**:R93.
31. Bowtell DD. The genesis and evolution of high-grade serous ovarian cancer. *Nat Rev Cancer* 2010;**10**:803–8.
32. Pollack JR, Sorlie T, Perou CM, *et al*. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A* 2002;**99**:12963–8.
33. Kumar V, Westra HJ, Karjalainen J, *et al*. Human disease-associated genetic variation impacts large intergenic non-coding RNA expression. *PLoS Genet* 2013;**9**: e1003201.
34. Luo J, Yu Y, Mitra A, *et al*. Genome-wide copy number variant analysis in inbred chickens lines with different susceptibility to Marek's disease. *G3 (Bethesda)* 2013;**3**: 217–23.

35. Hay N. Reprogramming glucose metabolism in cancer: can it be exploited for cancer therapy? *Nat Rev Cancer* 2016;**16**:635–49.
36. Wang A, Bao Y, Wu Z, *et al*. Long noncoding RNA EGFR-AS1 promotes cell growth and metastasis via affecting HuR mediated mRNA stability of EGFR in renal cancer. *Cell Death Dis* 2019;**10**:154.
37. Liu Q, Sun S, Yu W, *et al*. Altered expression of long noncoding RNAs during genotoxic stress-induced cell death in human glioma cells. *J Neurooncol* 2015;**122**:283–92.
38. Huang Y, Xiang B, Liu Y, *et al*. Lnc RNA CDKN2B-AS1 promotes tumor growth and metastasis of human hepatocellular carcinoma by targeting let-7c-5p/NAP1L1 axis. *Cancer Lett* 2018;**437**:56–66.
39. Zhu L, Zhang Q, Li S, *et al*. Interference of the long noncoding RNA CDKN2B-AS1 upregulates mi R-181a-5p/TGFbetaI axis to restrain the metastasis and promote apoptosis and senescence of cervical cancer cells. *Cancer Med* 2019;**8**:1721–30.
40. Wu DM, Wang S, Wen X, *et al*. Long noncoding RNA nuclear enriched abundant transcript 1 impacts cell proliferation, invasion, and migration of glioma through regulating mi R-139-5p/CDK6. *J Cell Physiol* 2019;**234**:5972–87.
41. Chen L, Zhang J, Feng Y, *et al*. MiR-410 regulates MET to influence the proliferation and invasion of glioma. *Int J Biochem Cell Biol* 2012;**44**:1711–7.
42. Zhang JF, Wang P, Yan YJ, *et al*. IL33 enhances glioma cell migration and invasion by upregulation of MMP2 and MMP9 via the ST2-NF-kappa B pathway. *Oncol Rep* 2017;**38**:2033–42.
43. Dong X, Jin Z, Chen Y, *et al*. Knockdown of long non-coding RNA ANRIL inhibits proliferation, migration, and invasion but promotes apoptosis of human glioma cells by upregulation of mi R-34a. *J Cell Biochem* 2018;**119**:2708–18.
44. Jiang Q, Ma R, Wang J, *et al*. Lnc RNA2Function: a comprehensive resource for functional investigation of human lnc RNAs based on RNA-seq data. *BMC Genomics* 2015;**16**(Suppl 3):S2.
45. Zhou J, Zhang S, Wang H, *et al*. Lnc fun net: an integrated computational framework for identification of functional long noncoding RNAs in mouse skeletal muscle cells. *Nucleic Acids Res* 2017;**45**:e108.
46. Monk D, Mackay DJG, Eggermann T, *et al*. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. *Nat Rev Genet* 2019;**20**:235–48.
47. Peters J. The role of genomic imprinting in biology and disease: an expanding view. *Nat Rev Genet* 2014;**15**: 517–30.
48. Kanduri C. Long noncoding RNAs: lessons from genomic imprinting. *Biochim Biophys Acta* 2016;**1859**:102–11.
49. Monnier P, Martinet C, Pontis J, *et al*. H19 lnc RNA controls gene expression of the imprinted gene network by recruiting MBD1. *Proc Natl Acad Sci U S A* 2013;**110**:20693–8.
50. Varrault A, Gueydan C, Delalbre A, *et al*. Zac 1 regulates an imprinted gene network critically involved in the control of embryonic growth. *Dev Cell* 2006;**11**:711–22.
51. Wei Y, Su J, Liu H, *et al*. Meta imprint: an information repository of mammalian imprinted genes. *Development* 2014;**141**:2516–23.
52. He H, Wang N, Yi X, *et al*. Long non-coding RNA H19 regulates E2F1 expression by competitively sponging endogenous mi R-29a-3p in clear cell renal cell carcinoma. *Cell Biosci* 2017;**7**:65.
53. Ma L, Tian X, Wang F, *et al*. The long noncoding RNA H19 promotes cell proliferation via E2F-1 in pancreatic ductal adenocarcinoma. *Cancer Biol Ther* 2016;**17**:1051–61.
54. Rouillard AD, Gundersen GW, Fernandez NF, *et al*. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database (Oxford)* 2016;**2016**.