SCIENCE CHINA
Life Sciences

CrossMark click for updates

• COVER ARTICLE •

October 2016 Vol.59 No.10: 981–992
doi: 10.1007/s11427-016-0054-1

Identifying and annotating human bifunctional RNAs reveals their versatile functions

Geng Chen[1], Juan Yang[2], Jiwei Chen[2], Yunjie Song[1], Ruifang Cao[2], Tieliu Shi[2*] & Leming Shi[1,3,4,5**]

[1]Center for Pharmacogenomics, School of Pharmacy and School of Life Sciences, Fudan University, Shanghai 201203, China;
[2]The Center for Bioinformatics and Computational Biology, Shanghai Key Laboratory of Regulatory Biology, the Institute of Biomedical Sciences and School of Life Sciences, East China Normal University, Shanghai 200241, China;
[3]State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 201203, China;
[4]Fudan-Zhangjiang Center for Clinical Genomics, Shanghai 201203, China;
[5]Zhanjiang Center for Translational Medicine, Shanghai 201203, China

Received March 16, 2016; accepted May 7, 2016; published online September 18, 2016

Bifunctional RNAs that possess both protein-coding and noncoding functional properties were less explored and poorly understood. Here we systematically explored the characteristics and functions of such human bifunctional RNAs by integrating tandem mass spectrometry and RNA-seq data. We first constructed a pipeline to identify and annotate bifunctional RNAs, leading to the characterization of 132 high-confidence bifunctional RNAs. Our analyses indicate that bifunctional RNAs may be involved in human embryonic development and can be functional in diverse tissues. Moreover, bifunctional RNAs could interact with multiple miRNAs and RNA-binding proteins to exert their corresponding roles. Bifunctional RNAs may also function as competing endogenous RNAs to regulate the expression of many genes by competing for common targeting miRNAs. Finally, somatic mutations of diverse carcinomas may generate harmful effect on corresponding bifunctional RNAs. Collectively, our study not only provides the pipeline for identifying and annotating bifunctional RNAs but also reveals their important gene-regulatory functions.

bifunctional RNA, noncoding RNA, RNA-seq, tandem mass spectrometry

Citation: Chen, G., Yang, J., Chen, J., Song, Y., Cao, R., Shi, T., and Shi, L. (2016). Identifying and annotating human bifunctional RNAs reveals their versatile functions. Sci China Life Sci 59, 981–992. doi: 10.1007/s11427-016-0054-1

INTRODUCTION

The human genome encodes diverse protein-coding and non-coding RNAs (ncRNAs). In general, ncRNAs do not translate into proteins and the number of ncRNAs is much larger than that of protein-coding RNAs (Derrien et al., 2012). ncRNAs can be roughly divided into small (<200 nt) and long (⩾200 nt) ncRNAs according to their lengths with an arbitrary cutoff of 200 nt. However, based on the relative genomic locations between ncRNAs and protein-coding genes, ncRNAs could be further classified into different categories including bidirectional, intronic, antisense and overlapped (Mercer et al., 2009). Recently, a growing number of studies have indicated that ncRNAs play crucial roles in chromatin remodeling, genetic imprinting, cell cycle, alternative splicing, and mRNA translation or degradation (Faghihi et al., 2008; Hung et al., 2011; Tripathi et al., 2010; Tsai et al., 2010). Moreover, long

*Corresponding author (email: tieliushi01@gmail.com)
**Corresponding author (email: lemingshi@fudan.edu.cn)

The Author(s) 2016. This article is published with open access at link.springer.com

ncRNAs (lncRNAs) are involved in the progression of many cancers and some of them have been recognized as potential biomarkers for tumor prevention, prognosis or treatment (Batista and Chang, 2013; Esteller, 2011; Hu et al., 2014).

Intriguingly, a portion of RNAs termed as bifunctional RNAs possess two distinct functions. Some bifunctional RNAs have the capacity to function as two different ncRNAs (Ender et al., 2008), whereas some others have the potential to function as ncRNAs or encode peptides/proteins (Dinger et al., 2008). Several important examples regarding how bifunctional RNAs exert their ncRNA/protein roles have been reported in *Xenopus* (Kloc et al., 2005), *Drosophila* (Jenny et al., 2006), *Escherichia coli* (Vanderpool and Gottesman, 2004; Wadler and Vanderpool, 2007) and human (Anderson et al., 2015; Chooniedass-Kothari et al., 2004; Lanz et al., 1999). Using the tandem mass spectrometry (MS/MS) and RNA-seq data of K562 and GM12878 cell lines, Bánfai et al. identified a small number of translatable lncRNAs as bifunctional RNAs (Banfai et al., 2012). A recently published paper of the human proteome also indicates that a portion of ncRNAs and pseudogenes may be translatable as seen from matched peptides (Kim et al., 2014). Furthermore, by analyzing the sequencing data of corresponding ribosome profiling, Ruiz-Orera et al. found that many lncRNAs of six distinct species were correlated with ribosomes and might be crucial for the de novo evolution of proteins (Ruiz-Orera et al., 2014). However, to date, the existence and functions of bifunctional RNAs that can function independently as ncRNA or encode functional proteins remain to be largely underexplored. Due to the uncertainty expressions of RNAs and proteins (Chen et al., 2013), it is crucial to identify and characterize those mysterious bifunctional RNAs by combing tandem mass spectrometry (MS/MS) and RNA-seq data.

In this study, by integrating diverse MS/MS and RNA-seq data, we identified 132 high-confidence bifunctional RNAs that have both noncoding and protein-coding capacity and systematically annotated their functions from distinct aspects. First, we constructed a pipeline for identifying bifunctional RNAs based on MS/MS data and reassessing their coding potential. Secondly, we interrogated the characteristics and potential functions of the encoded proteins for those bifunctional RNAs. Thirdly, we examined the expression profiles of those bifunctional RNAs in early human embryos and a large number of diverse human tissues. Fourthly, by constructing networks of ncRNA-miRNA and ncRNA-protein interactions, we determined the interactions between bifunctional RNAs and miRNAs and proteins, respectively. Fifthly, we noted that a portion of bifunctional ncRNAs might be functional as competing endogenous RNAs (ceRNAs) to influence the expression of relevant genes including multiple cancer related ones. Finally, we found that most of these bifunctional RNAs could be matched with known somatic mutations of various cancer types.

RESULTS

Identification of bifunctional RNAs based on MS/MS data and reassessment of their coding potential

To accurately discriminate bifunctional RNAs from the rest of ncRNAs, we first sought to construct a robust pipeline to identify the bona fide translatable ncRNAs, i.e. bifunctional RNAs. We parsed the MS/MS data of NCI-60 panel (including 59 cancer cell lines of 9 distinct tissue types) (Gholami et al., 2013) by employing TPP (Trans-Proteomic Pipeline) (Deutsch et al., 2010) with X!Tandem (Keller et al., 2005) as search engine (Figure 1). A protein database consisting of Ensembl v75 proteins and the in silico translated proteins of Ensembl ncRNAs (including pseudogenes) was established for MS/MS data matching (see Methods). Using more stringent criteria of protein probability >0.95 and at least one unique peptide for an identified protein, we detected 720 ncRNAs meeting such thresholds. 482 of them were reclassified as protein-coding RNAs using the effective Coding-Potential Assessment Tool (CPAT) (Wang et al., 2013) and were excluded from following analysis because they might be resulted from the mis-annotation of Ensembl protein-coding transcripts. The remaining 238 ncRNAs were still reassessed with no coding potential. We further removed 106 ncRNAs generated from protein-coding genes and genomic patch sequences, which is important to minimize the false positives of bifunctional RNA identification. Finally, 132 ncRNAs were retained as high-confidence bifunctional RNAs and none of them overlapped with those 15 ones identified by Bánfai et al (Banfai et al., 2012) (Figure 1 and Table S1 in Supporting Information). Furthermore, the possible mis-annotated ncRNAs are generally longer than those bifunctional RNAs in length (Wilcoxon rank sum test, P=8.2×10^{-8}) and contain more exons in general (Table S2 in Supporting Information) Most of the bifunctional RNAs identified in each condition were also detected in other conditions of NCI-60 cancer cell lines (Figure 2A). Particularly, 5 bifunctional RNAs (ENST00000388966, ENST00000413899, ENST00000429708, ENST00000462417 and ENST00000-513224) were identified across all those 9 different NCI-60 tissue types.

The reliability of our pipeline was tested with literature data on the identification of bifunctional RNAs. In a previous study, 15 translatable lncRNAs (bifunctional RNAs) that still remained in the updated gene set of Ensembl v75 we used in this study were identified in cell lines K562 and GM12878 (FDR ≤10% for peptide identification) (Banfai et al., 2012). Moreover, 47 lncRNAs matched with peptides they identified that still existed in Ensembl v75 were considered as the annotation error of protein-coding transcripts, we denoted them

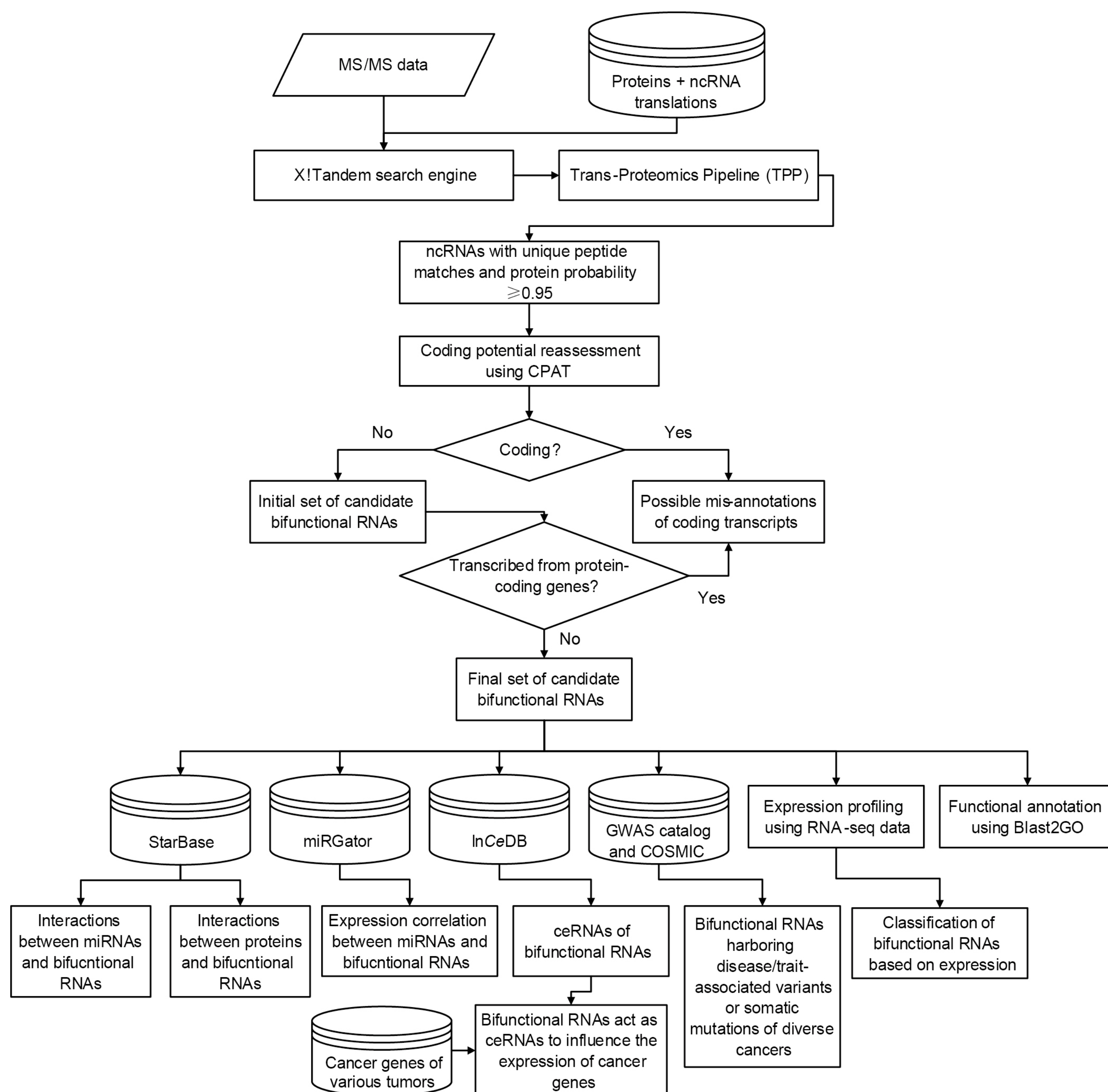

Figure 1 Pipeline for the identification and annotation of bifunctional RNAs. Bifunctional RNAs were identified by a series of steps with corresponding filters based on MS/MS data and coding capacity reassessment. The ncRNAs encoded by the protein-coding genes and genomic patches were further removed to minimize the false positives. The identified bifunctional RNAs were further characterized and annotated using RNA-seq data, network inference and related databases.

as misclassified lncRNAs. We further reassessed the coding potential of those 62 lncRNAs (15 bifunctional RNAs and 47 misclassified lncRNAs) with peptide matches using CPAT (Wang et al., 2013). Interestingly, 2 of the 15 translatable lncRNAs were reassessed as protein-coding and 11 of those 47 misclassified lncRNAs were reassessed as non-protein-coding by CPAT. Thus, assuming the prior results are correct, MS/MS data coupled with CPAT could effectively identify the bifunctional RNAs (86.67%, 13 out of 15) and misclassified lncRNAs (76.6%, 36 out of 47).

Bifunctional RNAs are shorter and encompass fewer numbers of exons than protein-coding RNAs

Bifunctional RNAs are shorter in length (Wilcoxon rank sum test, $P=4.8\times10^{-14}$) and contain fewer numbers of exons ($P<2.2\times10^{-16}$) compared to protein-coding RNAs. Those 132 bifunctional RNAs are encoded by 109 non-protein-coding genes, some of which encode more than one bifunctional

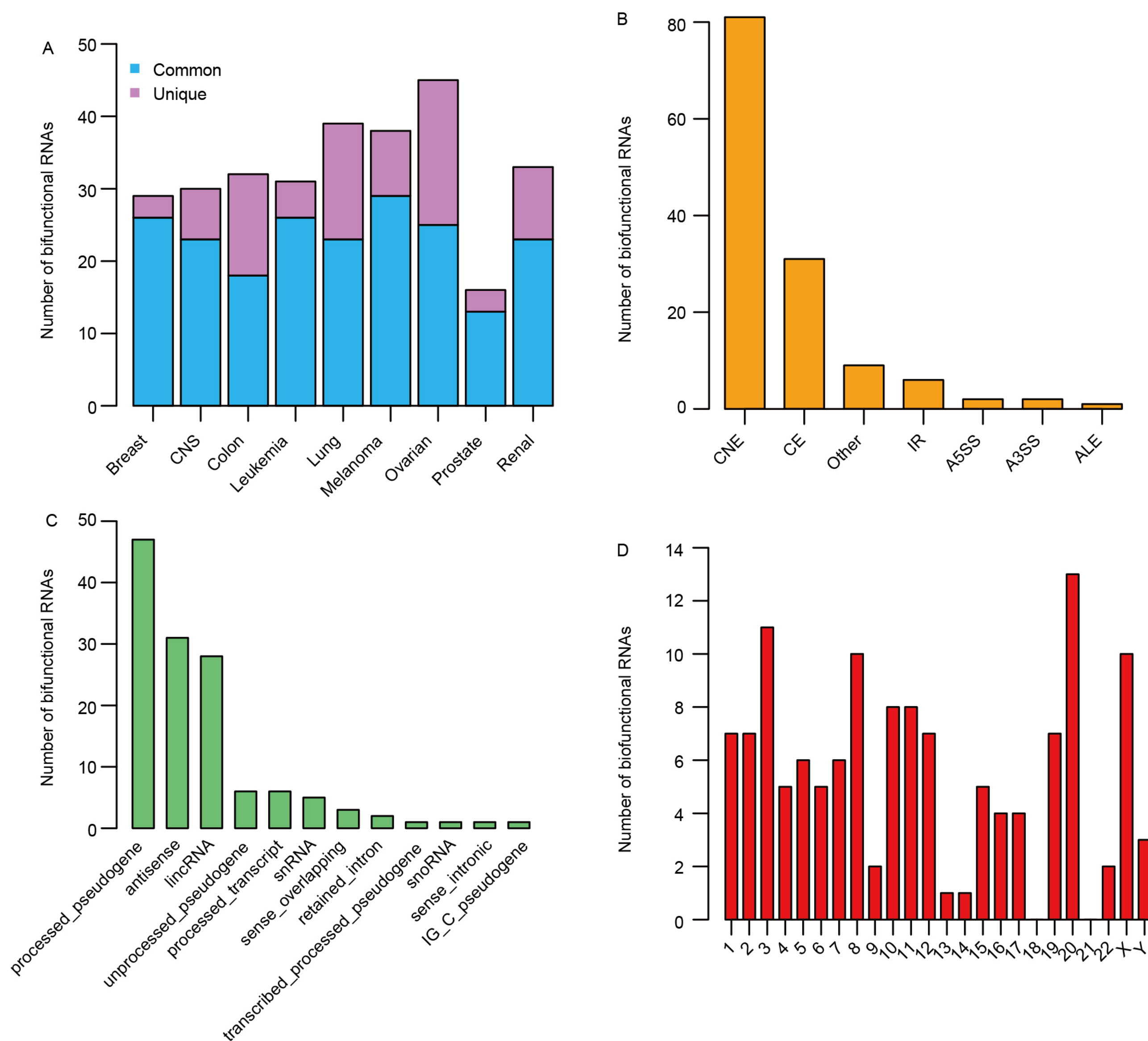

Figure 2 Statistics of bifunctional RNAs identified in the NCI-60 cancer cell lines. A, Distribution of bifunctional RNAs identified in each type of NCI-60 cancer cell lines. "Common" represents those bifunctional RNAs were identified in at least two types of NCI-60 cancer cell lines, whereas "Unique" denotes those bifunctional RNAs were only identified in corresponding type of cells. B, Alternative splicing modes of 132 bifunctional RNAs. CNE, constitutive exon; CE, cassette exon (exon skipping); Other, unclassified splicing mode; IR, intron retention; A5SS, alternative 5′ sites; A3SS, alternative 3′ sites; ALE, alternative last exon. C, The biotypes of bifunctional RNAs annotated by Ensembl. D, Number of identified bifunctional RNAs on each chromosome.

RNAs, mainly through the splice events of constitutive exons (81) and exon skipping (31) (Figure 2B). Their lengths range from 99 to 5,230 bp (median: 733.5 bp; mean: 1,073 bp), whereas the median (1,434 bp) and mean (2,036 bp) lengths of protein-coding RNAs are twice longer (Ensembl v75). Moreover, most (70.4%) of those bifunctional RNAs consist of only one (54) or two (39) exons; however, the great majority (91.8%) of protein-coding RNAs harbor more than 2 exons. Those 132 bifunctional RNAs could be classified into diverse biotypes including processed pseudogene, antisense, and lincRNA (long intergenic noncoding RNA) according to Ensembl gene annotations (Figure 2C). In addition, they scattered over 22 disparate human chromosomes (Figure 2D).

Bifunctional RNAs are partially conserved across species and their encoded proteins are functional

Most of those bifunctional RNAs harbor conserved sequences and the proteins encoded by them can be functionally important. We found that the exons of 104 out of the 132 bifunctional RNAs overlapped with (at least one-base overlap) the conserved elements across 46 vertebrates predicted by phastCons (Siepel et al., 2005), suggesting that the sequences of most of those bifunctional RNAs are partially conserved across vertebrates. Moreover, functional enrichment analysis for those bifunctional RNAs using Blast2GO suite (Gotz et al., 2008) implied that their encoded proteins are mainly involved in the biological processes of cell cycle, cell division,

organelle fission, and organelle organization.

Bifunctional RNAs are expressed in human embryos and various tissues

The expression level of genes could be an important aspect to reflect their functions in correponding cells (Gout et al., 2010; Vogel and Marcotte, 2012). To investigate the nature of expression profiles for bifunctional RNAs, we explored their expression in 2,916 samples across 53 tissue sites of the GTEx (Genotyp-Tissue Expression) project (Lonsdale et al., 2013) and 29 samples across 7 continuous early embryonic stages of human (Xue et al., 2013). The expression levels of those bifunctional RNAs in the GTEx tissues were extracted from the transcript expression matrix provided by the GTEx portal, whereas raw data of embryonic samples were used to quantify the gene/transcript expression by employing TopHat2 (Kim et al., 2013) and Cufflinks (Trapnell et al., 2010) (see Methods). We found that a small fraction of those bifunctional RNAs was expressed in early embryos (Figure 3A); however, a large portion of them was expressed across diverse tissues of the GTEx project (Figure 3B). Intriguingly, the bifunctional RNAs with relatively high expression levels in early embryos were expressed in various adult tissues as well, implying that these bifunctional RNAs possess certain basic functions in human cells. On the other hand, bifucntional RNAs only expressed in adult human tissues but not in embryos showed spatio-temporal expression characteristics and may mainly function in adult tissues. To interrogate whether bifunctional RNAs significantly changed in expression during embryonic development, we further conducted differential expression calling between adjacent embryonic stages (Trapnell et al., 2013). Interestingly, 2 bifunctional RNAs (q value <0.05), i.e., ENST00000571336 and ENST00000564694, were differentially expressed between 4-cell and 8-cell stages, suggesting that they are correlated with human embryonic development.

In addition, according to the clustering of expression profiles across different tissues, bifunctional RNAs can be clearly classified into three categories (Figure 3B): (i) those expressed with relatively high levels in most tissues (shown with orange bar); (ii) those expressed at medium levels in most tissues (shown with purple bar); and (iii) those expressed only in a fraction of tissues in a tissue-specific manner or no expression in any of those tissues (shown with red bar). Overall, the results suggest that those bifunctional RNAs may play important roles in human embryonic development as well as in maintaining the normal physiological functions for various adult tissues.

Bifunctional RNAs could interact with miRNAs, RNA-binding proteins

To gain insights into the functions of those bifunctional RNAs, we constructed the interaction networks of bifunctional RNAs with miRNAs and proteins using the miRNA-

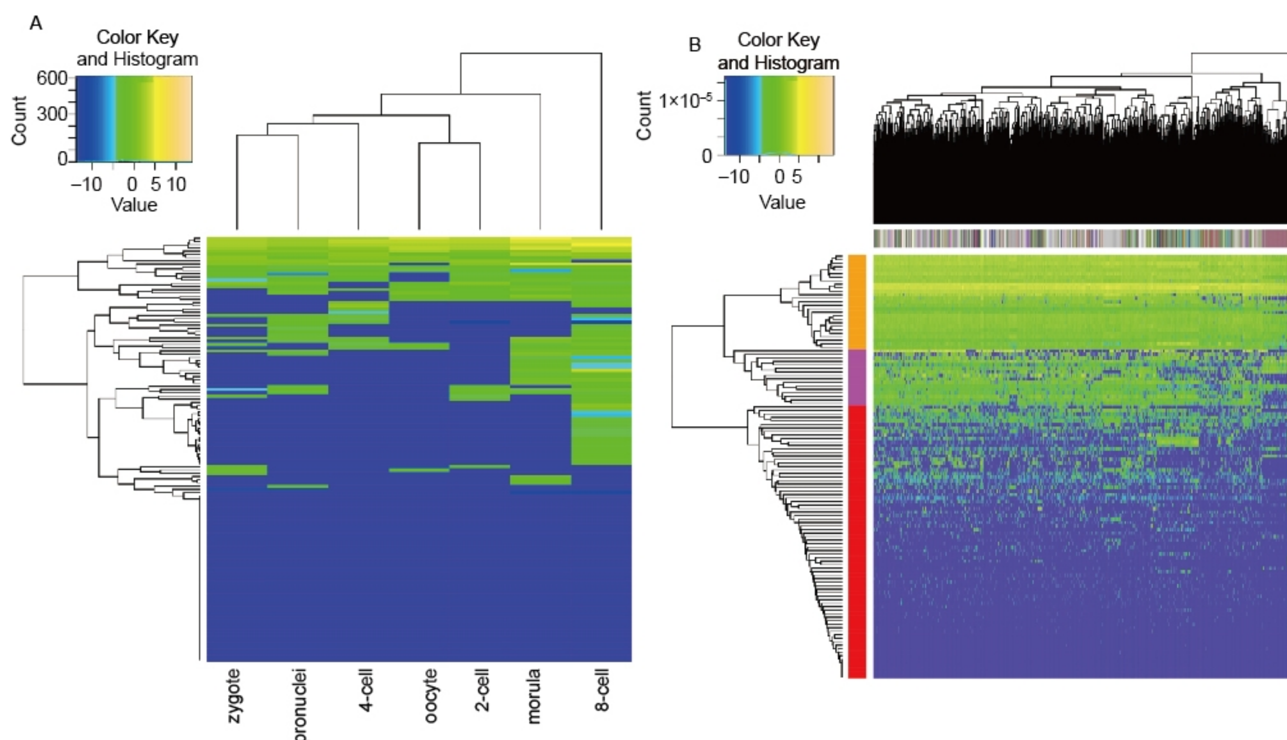

Figure 3 Expression profile of bifunctional RNAs in early human embryos and diverse tissues. A, Clustering of expression for bifunctional RNAs in 7 different embryonic stages. B, Classification of bifunctional RNAs based on their expression in 2,916 samples across 53 tissue sites of the GTEx project. The bifunctional RNAs were classified into three distinct groups with orange, purple and red bars, respectively.

ncRNA and protein-ncRNA interactions cataloged in starBase v2.0 (Li et al., 2014). The resulting network of bifunctional RNAs and miRNAs shows that 39 bifunctional RNA genes interact with 141 miRNAs through 470 interactions (Figure 4A). Furthermore, 35 out of the 39 bifunctional RNA genes interact with at least 2 and 11 with 10 or more distinct miRNAs. Specifically, the two bifunctional RNA genes ERVK3-1 and SDCBP2-AS1 form two hubs in the network that separately interact with 114 and 75 miRNAs, respectively. Notably, these two RNA genes could produce 6 or 5 bifunctional RNAs. On the other hand, 86 of those 141 miRNAs interact with 2 or more bifunctional RNAs, indicating that miRNAs and bifunctional RNAs interact mainly in a many-to-many mode (Figure 4A).

The primary interaction mode between bifunctional RNAs and RNA-binding proteins (RBPs) is many-to-many as well (Figure 4B). Briefly, 71 bifunctional RNA genes bind to 26 RBPs to yield 264 interactions. Forty-eight out of the 71 bifunctional RNA genes interact with 2 or more RBPs, while 24 RBPs bind to at least 2 bifunctional RNA genes. Surprisingly, protein UPF1 (regulator of nonsense transcripts homolog) has the greatest capacity to bind to 109 bifunctionl RNAs. Functional annotation of these RBPs using DAVID (Huang et al., 2008) indicates that they are mainly enriched in post-transcriptional regulation of gene expression, regulation of translation, gene silencing, and RNA processing (adjusted $P<0.01$). Because lncRNAs have the ability to bind to corresponding proteins to regulate gene expression by modifying the chromatin state (Mercer and Mattick, 2013), and thus, bifunctional RNAs may also have regulatory roles through binding to relevant proteins. These findings suggest that bifunctional RNAs may function similarly as lncRNAs to act as targets of diverse miRNAs and RBPs to regulate the expression of associated genes.

miRNAs could target bifunctional RNAs

To further interrogate whether miRNAs are targeted by those bifunctional RNAs in normal or cancer tissues of human, we explored the expression correlation between miRNAs and their targets in miRGator V3.0 (Cho et al., 2013). Expression correlations of miRNA-RNA cataloged in miRGator were calculated with the expression data of miRNA and RNA from the same samples, which is crucial for minimizing the biological bias. We used the genes of those bifunctional RNAs for examining their expression correlation with miRNAs, because the miRNA targets were archived at the gene level in miRGator. With the threshold of expression correlation (Pearson's correlation) $r\leq-0.3$, 7 genes (*BTF3P13*, *EPS15P1*, *GTF2H2B*, *HAS2-AS1*, *SLC38A3*, *UBXN8*, and *WASH6P*) encoding 8 bifunctional RNAs (*WASH6P* produced two) were negatively correlated with relevant miRNAs in different types of cancers or normal cells including kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), prostate cancer, leukemia, human embryonic stem cells or brain (Figure 5). When employing a more stringent cutoff of $r\leq-0.5$, bifunctional RNA genes *EPS15P1*, *GTF2H2B*, *SLC38A3*, and *UBXN8* remained to show strong inverse expression correlation with 6, 12, 109, and 135 miRNAs supported by at least one data set used in miRGator, respectively. Specifically, many of the strong negative correlations between miRNAs and bifunctional RNAs expression were only observed in a specific cancer or normal tissue, implying that bifunctional RNAs functionally are targeted by distinct miRNAs under different biological conditions. Accordingly, the results not only further demonstrate the above finding that bifunctional RNAs could function as the targets of multiple miRNAs, but also imply the important regulatory roles of bifunctional RNAs in human tumors.

Bifunctional RNAs may function as competing endogenous RNAs

Considering the fact that a portion of lncRNAs may act as competing endogenous RNAs (ceRNAs) by blocking miRNAs from binding to mRNAs (Cesana et al., 2011; Salmena et al., 2011; Sumazin et al., 2011; Tay et al., 2011), we further assessed whether the bifunctional RNAs identified in our study can function as ceRNAs. The 132 identified bifunctional RNAs were mapped to the ceRNA candidates documented in the ln*Ce*DB database (Das et al., 2014) and 26 bifunctional RNAs were matched. In summary, those 26 bifunctional RNAs can act as ceRNAs to compete with related RNAs for 963 targeting miRNAs, which may result in changes in expression levels of corresponding RNAs. Moreover, 22 out of those 26 bifunctional RNAs may function as ceRNAs to interfere the pathways of 2 or more miRNAs and 9 of them can even be targeted by over 100 miRNAs. Notably, a miRNA usually targets hundreds of genes, thus we observed that bifunctional RNAs could potentially influence the miRNA mediated regulation of expression for thousands of genes. Furthermore, 560 miRNAs could individually target at least 2 of those 26 bifunctional RNAs, suggesting that bifunctional RNAs may cross-regulate each other by competing for shared targeting miRNAs.

To investigate whether those 26 candidate ceRNAs of bifunctional RNAs have the potential to interfere the expression of human cancer related genes, we examined the target genes of miRNAs associated with ceRNAs in cancer genes of diverse tumors. First, a comprehensive human cancer gene set (2,580 in total, including oncogenes, tumor suppressor genes, driver or significantly mutated genes) was collected from databases of NCG4.0 (An et al., 2014), TSGene (Zhao et al., 2013) and two recent reports (including The Cancer Genome Atlas (TCGA)) regarding human pan-cancer (Kandoth et al., 2013; Vogelstein et al., 2013). Interestingly, the majority (2,355 out of 2,580) of these cancer genes

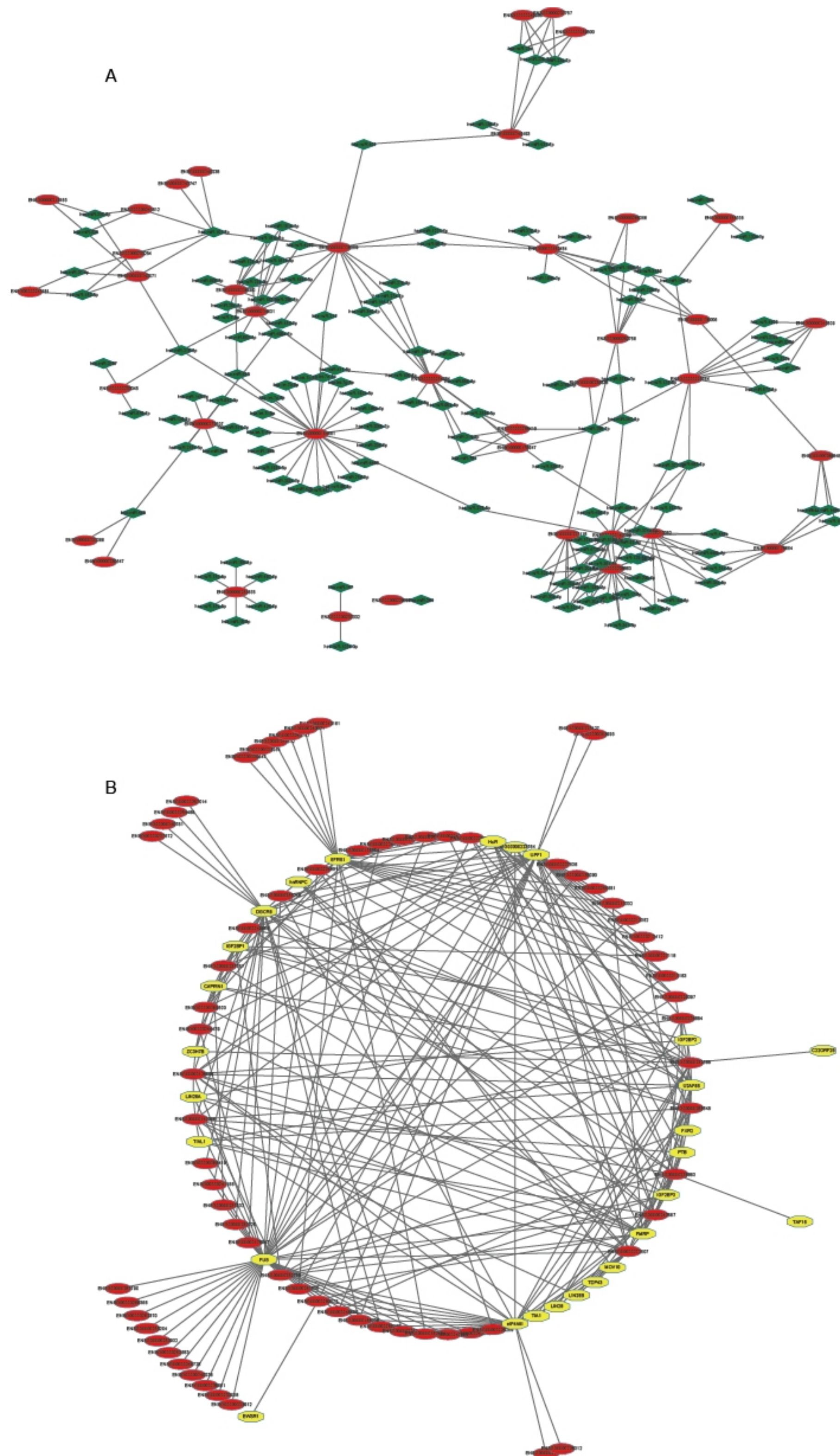

Figure 4 Interactions between bifunctional RNAs and miRNAs or RBPs. A, Interaction network between bifunctional RNAs and miRNAs. B, Interaction network between bifunctional RNAs and RBPs. Bifunctional RNAs, miRNAs and RBPs were shown in red ellipse, green diamond and yellow octagon, respectively.

associated with various cancer types are the targets of 145 miRNAs that also could potentially target 8 of those 26 candidate ceRNAs (Figure 6). These 8 bifunctional RNAs could be targeted by one or even dozens of miRNAs, while these 145 miRNAs would target one or hundreds of cancer genes. Accordingly, the results suggest that bifunctional RNAs may

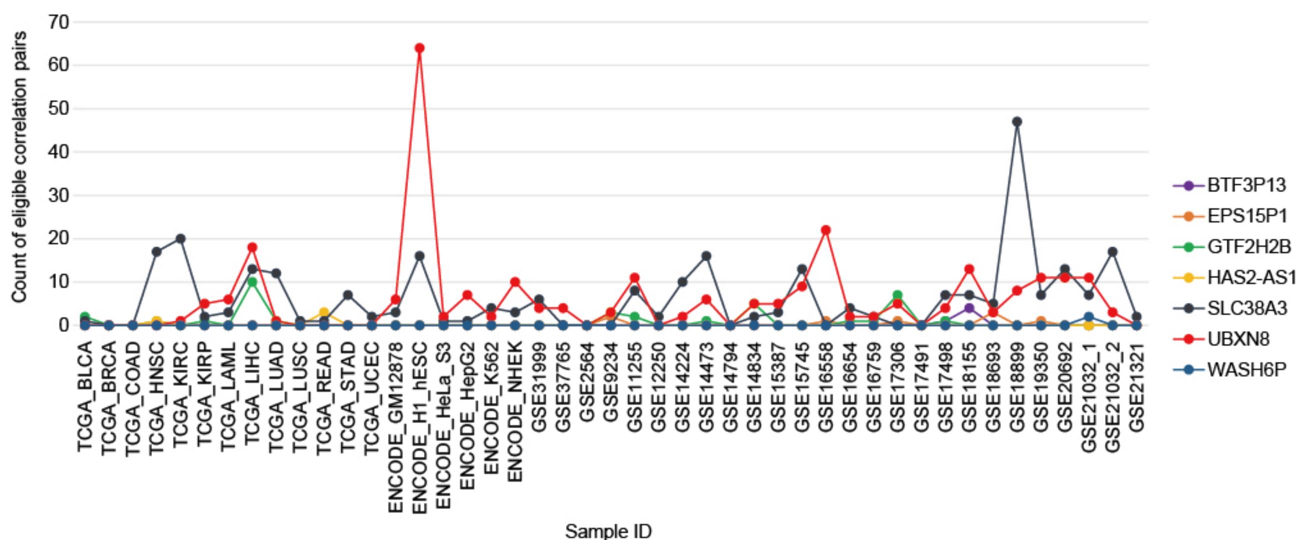

Figure 5 Count distribution of inverse expression correlation between miRNAs and bifunctional RNAs in different cancer or normal tissues. This figure shows the number of negative expression correlation ($r\leq -0.3$) between bifunctional RNAs and miRNAs in each data set obtained from miRGator V3.0 (Cho et al., 2013), which could provide the evidence that miRNAs may target bifunctional RNAs to regulate their expression.

play important roles in diverse human cancers through affecting the post-transcriptional regulation of cancer genes mediated by common targeting miRNAs.

Disease/trait-associated variants and somatic mutations influence bifunctional RNAs

To inquire whether these 132 bifunctional RNAs overlap with known disease/trait-associated variants or cancer related somatic mutations, we examined the published GWAS (genome-wide association studies) variants (Welter et al., 2014) and the somatic mutations of diverse cancers cataloged in the COSMIC database (Forbes et al., 2015). Intriguingly, 4 noncoding disease/trait-associated GWAS variants were located in the exonic or intronic regions of 6 bifunctional RNAs (Table S3 in Supporting Information). Moreover, we identified 1,801 noncoding somatic mutations of various tumor types within the exons or introns of 101 bifunctional RNAs (Table S4 in Supporting Information). The involved cancer types include the carcinomas of ovary, lung, breast, kidney, and prostate studied in the TCGA project and so on. Genetic variants in noncoding RNAs could be functionally important and might influence the expression or biogenesis of noncoding RNAs (Ward and Kellis, 2012). Thus, these identified GWAS variations and somatic mutations could generate harmful influences on relevant bifunctional RNAs in corresponding disease/trait or tumors.

DISCUSSION

MS/MS data provide great opportunities to assess the coding capacity of each transcript encode by the genome. We established a pipeline to effectively identify and annotate the human bifunctional RNAs based on MS/MS and RNA-seq data. The precision ratio is ~0.87 based on testing the identified bifunctional RNAs reported in a previous study (Banfai et al., 2012). Although we first detected 720 ncRNAs that could be matched with unique peptides, the majority of them might be resulted from mis-annotating protein-coding transcripts as noncoding and were excluded. Only 132 ncRNAs were finally assessed as high-confidence bifunctional RNAs through a series of stringent filtering processes. Our results show that a small portion of bifunctional RNAs were expressed in early human embryos; however, the majority of those bifunctional RNAs (including the ones expressed in embryos) were expressed in at least one adult tissue from the GTEx project (Lonsdale et al., 2013). Thus, most of those bifunctional RNAs would mainly function in adult tissues. Furthermore, those bifunctional RNAs could be generally classified into three groups according to their expression patterns in the huge number of samples from diverse tissue types of the GTEx project. The first group exhibits high expression in almost all the tissues and their conserved expression implies that for different tissues they might be fundamental like house-keeping genes. Bifunctional RNAs in the second group were broadly expressed in diverse tissues but with relatively lower abundance and were not expressed in a portion of the samples. Bifunctional RNAs in the third group were only lowly expressed in a fraction of samples or not expressed at all. Accordingly, the expression of bifunctional RNAs has spatio-temporal properties and could be important for embryonic development and for maintaining the normal physiological functions of diverse tissues.

Our findings indicate that bifunctional RNAs could be targeted by miRNAs or proteins to regulate the expression of corresponding genes including lots of cancer related genes. By constructing the interaction networks of bifunctional

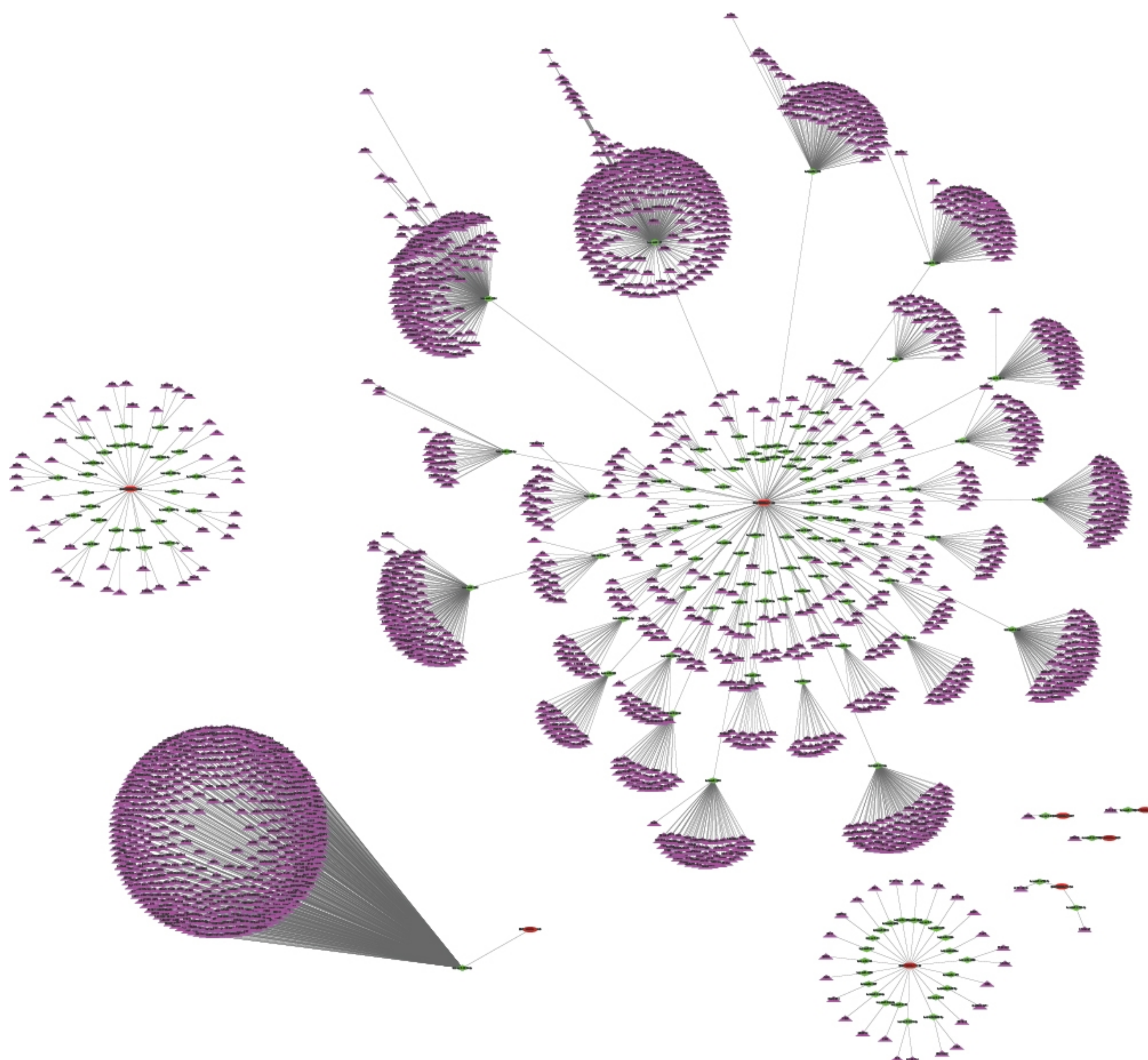

Figure 6 Interaction network among bifunctional RNAs, miRNAs and cancer genes of diverse tumors. Bifunctional RNA, miRNAs and cancer genes were denoted as red ellipse, lime diamond and purple triangle, respectively. The cancer genes associated with various tumors were obtained by integrating those in NCG4.0 (An et al., 2014) and TSGene (Zhao et al., 2013) databases as well as those reported in two recent human pan-cancer reports (Kandoth et al., 2013; Vogelstein et al., 2013).

RNAs with miRNAs and proteins, we observed that many-to-many is the main interaction mode for these two kinds of interactions. Moreover, we also found evidences for supporting the interaction between bifunctional RNAs and miRNAs in such a way that the expression of a number of miRNAs was negatively correlated with that of some bifunctional RNA genes in disparate human cancers or tissues. Besides, expression correlation between bifunctional RNAs and miRNAs extracted from miRGator (Cho et al., 2013) also suggests that bifunctional RNAs might be targeted by different miRNAs under distinct conditions. Interestingly, 26 bifuncational RNAs have the potential to act as ceRNAs to interfere the expression of corresponding genes by competing with the same targeting miRNA. We also found that bifunctional RNAs can cross-regulate their expression with each other because some of them share common targeting miRNAs. Specifically, the targeting miRNAs for 2,355 cancer related genes also target bifunctional RNAs, implying that bifunctional RNAs could function as ceRNAs to regulate the expression of cancer genes in various cancers. In addition, most of these bifunctional RNAs overlap with the somatic mutations from various cancers and those involved mutations could adversely influence the corresponding bifunctional RNAs to play certain roles in relevant cancer.

In summary, we identified 132 high-confidence bifuctional RNAs in the NCI-60 cancer cell lines and also characterized

their potential functions in detail through integrative analyses. Integrating the high-throughput data from MS/MS and RNA-seq, as well as constructing networks for interactions between bifunctional RNAs and miRNAs or proteins represent a promising strategy for identifying and annotating bifunctional RNAs. To the best of our knowledge, our work reported the first pipeline for systematically identifying and annotating bifunctional RNAs. This pipeline can be applied to further characterize more bifuncitonal RNAs in human or other species.

MATERIALS AND METHODS

Identification of bifunctional RNAs

To identify human bifunctional RNAs, MS/MS data of the NCI-60 cancer cell line panel from a previous study (Gholami et al., 2013) were downloaded and processed. X!Tandem (Keller et al., 2005) was employed as the database search engine for TPP (Trans-Proteomic Pipeline, v4.6.1) (Deutsch et al., 2010). Specifically, the MS/MS data were searched against the non-redundant protein database comprising of Ensembl proteins (version 75) and in silico translated Ensembl ncRNAs using X!Tandem-native and X!Tandem-Kscore modes with default settings. Then PeptideProphet (Keller et al., 2002) and iProphet (Shteynberg et al., 2011) were utilized to validate and integrate the peptide identification results. ProteinProphet was finally employed to identify and validate the proteins. For peptide and protein identification, the probability of 0.95 was used as the threshold.

We then reassessed the coding capacity of those potentially translatable ncRNAs using CPAT (Wang et al., 2013). The candidate ncRNAs that were reassessed as protein-coding were excluded from further analyses. We further removed those ncRNAs that were generated from protein-coding genes or genomic patches. Through above stringent filtering, the remaining identified translatable ncRNAs were considered as high-confidence candidates of bifunctional RNAs.

Expression profiling of bifunctional RNAs

We downloaded the gene/transcript expression matrix of the GTEx project from its portal (http://www.gtexportal.org/home/), which included of 2,916 samples across 53 tissue sites (Lonsdale et al., 2013). Then expressions of bifunctional RNAs were extracted from the GTEx expression matrix. To investigate the expression profile of bifunctional RNAs in early human embryos, we collected the single-cell RNA-seq data of oocyte, pronucleus, zygote, 2-cell, 4-cell, 8-cell and morula stages from Gene Expression Omnibus (GEO) (accession number: GSE44183) (Xue et al., 2013). These RNA-seq data for a total of 29 samples were separately mapped to the human genome GRCh37/hg19 using TopHat2 (Kim et al., 2013) (version 2.0.11) with parameter "–r=0" based on the experimental design of the previous study. Next, we quantified the expression of genes/transcripts in each sample by employing Cufflinks (Trapnell et al., 2010) (version 2.2.1) with parameters of "–u" and "–b" enabled for correcting read mapping. Differential expression calling between two adjacent stages was carried out using Cuffdiff2 (Trapnell et al., 2013) (version 2.2.1) with parameters of "–b" and "–u" enabled as well.

Characterization and annotation of bifunctional RNAs

To examine the conservation of the bifunctional RNAs identified in our study, we downloaded the conserved elements across 46 vertebrates predicted by phastCons (Siepel et al., 2005) from UCSC Genome Browser. In order to investigate the interactions between bifunctional RNAs and miRNA or proteins, we obtained the interactions of miRNA-ncRNA and protein-ncRNA from starBase v2.0 (Li et al., 2014) to construct corresponding networks. We also examined the expression correlatio between bifunctional RNAs and their targeting miRNAs using miRGator V3.0 (Cho et al., 2013) in different cancer and normal tissues. To interrogate whether bifunctional RNAs could act as ceRNAs to regulate the expression of relevant genes, we checked their existence in the ceRNA candidates cataloged in ln*Ce*DB database (Das et al., 2014). The cancer related genes of diverse human cancers from the databases of NCG4.0 (An et al., 2014), TSGene (Zhao et al., 2013) and two previous studies about human pan-cancer (Kandoth et al., 2013; Vogelstein et al., 2013) were collected as well. We further inspected whether bifunctional RNAs would function as ceRNA to influence the expression of cancer genes by inquiring the miRNAs that could target both cancer genes and bifunctional RNAs. The published disease/trait-associated GWAS variants were downloaded from the National Human Genome Research Institute (NHGRI) catalog (Welter et al., 2014). We also obtained the noncoding somatic mutations of various cancers from the COSMIC database (v71) (Ward and Kellis, 2012).

Author contributions *G.C., T.S. L.S. conceived and designed the study. G.C., J.Y., J.C., Y.S., R.C. and G.Z. carried out the analyses. G.C., T.S. and L.S. wrote the manuscript. All the authors read and approved the final manuscript.*

Compliance and ethics *Compliance and ethics The authors declare that they have no conflict of interest.*

Acknowledgements *We would like to thank Huan Wang for her helpful discussion in drawing some figures. This work was supported in part by the National High Technology Research and Development Program of China (2015AA020104, 2015AA020108), the China Human Proteomics Project (2014DF30030) and the National Science Foundation of China (31471239).*

An, O., Pendino, V., D'Antonio, M., Ratti, E., Gentilini, M., and Ciccarelli, F.D. (2014). NCG 4.0: the network of cancer genes in the era of massive mutational screenings of cancer genomes. Database 2014, bau015–bau015.

Anderson, D.M., Anderson, K.M., Chang, C.L., Makarewich, C.A., Nelson, B.R., McAnally, J.R., Kasaragod, P., Shelton, J.M., Liou, J.,

Bassel-Duby, R., and Olson, E.N. (2015). A micropeptide encoded by a putative long noncoding RNA regulates muscle performance. Cell 160, 595–606.

Banfai, B., Jia, H., Khatun, J., Wood, E., Risk, B., Gundling, W.E., Kundaje, A., Gunawardena, H.P., Yu, Y., Xie, L., Krajewski, K., Strahl, B.D., Chen, X., Bickel, P., Giddings, M.C., Brown, J.B., and Lipovich, L. (2012). Long noncoding RNAs are rarely translated in two human cell lines. Genome Res 22, 1646–1657.

Batista, P.J., and Chang, H.Y. (2013). Long noncoding RNAs: cellular address codes in development and disease. Cell 152, 1298–1307.

Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A., and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. Cell 147, 358–369.

Chen, G., Chen, J., Shi, C., Shi, L., Tong, W., Shi, T., and Roemer, K. (2013). Dissecting the characteristics and dynamics of human protein complexes at transcriptome cascade using RNA-Seq data. PLoS ONE 8, e66521.

Cho, S., Jang, I., Jun, Y., Yoon, S., Ko, M., Kwon, Y., Choi, I., Chang, H., Ryu, D., Lee, B., Kim, V.N., Kim, W., and Lee, S. (2013). miRGator v3.0: a microRNA portal for deep sequencing, expression profiling and mRNA targeting. Nucleic Acids Res 41, D252–D257.

Chooniedass-Kothari, S., Emberley, E., Hamedani, M.K., Troup, S., Wang, X., Czosnek, A., Hube, F., Mutawe, M., Watson, P.H., and Leygue, E. (2004). The steroid receptor RNA activator is the first functional RNA encoding a protein. FEBS Lett 566, 43–47.

Das, S., Ghosal, S., Sen, R., Chakrabarti, J., and Banfi, S. (2014). ln*Ce*DB: database of human long noncoding RNA acting as competing endogenous RNA. PLoS ONE 9, e98965.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J.B., Lipovich, L., Gonzalez, J.M., Thomas, M., Davis, C.A., Shiekhattar, R., Gingeras, T.R., Hubbard, T.J., Notredame, C., Harrow, J., and Guigo, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22, 1775–1789.

Martens, L., Hermjakob, H., Deutsch, E.W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., Eng, J.K., Martin, D.B., Nesvizhskii, A.I., and Aebersold, R. (2010). A guided tour of the Trans-Proteomic Pipeline. Proteomics 10, 1150–1159.

Dinger, M.E., Pang, K.C., Mercer, T.R., Mattick, J.S., and McEntyre, J. (2008). Differentiating protein-coding and noncoding RNA: challenges and ambiguities. PLoS Comput Biol 4, e1000176.

Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A human snoRNA with microRNA-like functions. Mol Cell 32, 519–528.

Esteller, M. (2011). Non-coding RNAs in human disease. Nat Rev Genet 12, 861–874.

Faghihi, M.A., Modarresi, F., Khalil, A.M., Wood, D.E., Sahagan, B.G., Morgan, T.E., Finch, C.E., St. Laurent III, G., Kenny, P.J., and Wahlestedt, C. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of b-secretase. Nat Med 14, 723–730.

Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., Kok, C.Y., Jia, M., De, T., Teague, J.W., Stratton, M.R., McDermott, U., and Campbell, P.J. (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res 43, D805–D811.

Gotz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., and Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36, 3420–3435.

Gout, J.F., Kahn, D., Duret, L., and Pritchard, J.K. (2010). The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. PLoS Genet 6, e1000944.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., Fleming, J., Siminoff, L., Traino, H., Mosavel, M., Barker, L., Jewell, S., Rohrer, D., Maxim, D., Filkins, D., Harbach, P., Cortadillo, E., Berghuis, B., Turner, L., Hudson, E., Feenstra, K., Sobin, L., Robb, J., Branton, P., Korzeniewski, G., Shive, C., Tabor, D., Qi, L., Groch, K., Nampally, S., Buia, S., Zimmerman, A., Smith, A., Burges, R., Robinson, K., Valentino, K., Bradbury, D., Cosentino, M., Diaz-Mayoral, N., Kennedy, M., Engel, T., Williams, P., Erickson, K., Ardlie, K., Winckler, W., Getz, G., DeLuca, D., MacArthur, D., Kellis, M., Thomson, A., Young, T., Gelfand, E., Donovan, M., Meng, Y., Grant, G., Mash, D., Marcus, Y., Basile, M., Liu, J., Zhu, J., Tu, Z., Cox, N.J., Nicolae, D.L., Gamazon, E.R., Im, H.K., Konkashbaev, A., Pritchard, J., Stevens, M., Flutre, T., Wen, X., Dermitzakis, E.T., Lappalainen, T., Guigo, R., Monlong, J., Sammeth, M., Koller, D., Battle, A., Mostafavi, S., McCarthy, M., Rivas, M., Maller, J., Rusyn, I., Nobel, A., Wright, F., Shabalin, A., Feolo, M., Sharopova, N., Sturcke, A., Paschal, J., Anderson, J.M., Wilder, E.L., Derr, L.K., Green, E.D., Struewing, J.P., Temple, G., Volpi, S., Boyer, J.T., Thomson, E.J., Guyer, M.S., Ng, C., Abdallah, A., Colantuoni, D., Insel, T.R., Koester, S.E., Little, A.R., Bender, P.K., Lehner, T., Yao, Y., Compton, C.C., Vaught, J.B., Sawyer, S., Lockhart, N.C., Demchok, J., and Moore, H.F. (2013). The Genotype-Tissue Expression (GTEx) project. Nat Genet 45, 580–585.

Hu, X., Feng, Y., Zhang, D., Zhao, S.D., Hu, Z., Greshock, J., Zhang, Y., Yang, L., Zhong, X., Wang, L.P., Jean, S., Li, C., Huang, Q., Katsaros, D., Montone, K.T., Tanyi, J.L., Lu, Y., Boyd, J., Nathanson, K.L., Li, H., Mills, G.B., and Zhang, L. (2014). A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. Cancer Cell 26, 344–357.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4, 44–57.

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., Grant, G.D., Horlings, H.M., Shah, N., Umbricht, C., Wang, P., Wang, Y., Kong, B., Langerød, A., Børresen-Dale, A.L., Kim, S.K., van de Vijver, M., Sukumar, S., Whitfield, M.L., Kellis, M., Xiong, Y., Wong, D.J., and Chang, H.Y. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet 43, 621–629.

Jenny, A., Hachet, O., Zavorszky, P., Cyrklaff, A., Weston, M.D., Johnston, D.S., Erdelyi, M., and Ephrussi, A. (2006). A translation-independent role of oskar RNA in early *Drosophila* oogenesis. Development 133, 2827–2833.

Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., Leiserson, M.D.M., Miller, C.A., Welch, J.S., Walter, M.J., Wendl, M.C., Ley, T.J., Wilson, R.K., Raphael, B.J., and Ding, L. (2013). Mutational landscape and significance across 12 major cancer types. Nature 502, 333–339.

Keller, A., Eng, J., Zhang, N., Li, X., and Aebersold, R. (2005). A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol Syst Biol 1, E1–E8.

Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 74, 5383–5392.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36.

Kim, M.S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., Thomas, J.K., Muthusamy, B., Leal-Rojas, P., Kumar, P., Sahasrabuddhe, N.A., Balakrishnan, L., Advani, J., George, B., Renuse, S., Selvan, L.D.N., Patil, A.H., Nanjappa, V., Radhakrishnan, A., Prasad, S., Subbannayya, T., Raju, R., Kumar, M., Sreenivasamurthy, S.K., Marimuthu, A., Sathe, G.J., Chavan, S., Datta, K.K., Subbannayya, Y., Sahu, A., Yelamanchi, S.D., Jayaram, S., Rajagopalan, P., Sharma, J., Murthy, K.R., Syed, N., Goel, R., Khan, A.A., Ahmad, S., Dey, G., Mudgal, K., Chatterjee, A.,

Huang, T.C., Zhong, J., Wu, X., Shaw, P.G., Freed, D., Zahari, M.S., Mukherjee, K.K., Shankar, S., Mahadevan, A., Lam, H., Mitchell, C.J., Shankar, S.K., Satishchandra, P., Schroeder, J.T., Sirdeshmukh, R., Maitra, A., Leach, S.D., Drake, C.G., Halushka, M.K., Prasad, T.S.K., Hruban, R.H., Kerr, C.L., Bader, G.D., Iacobuzio-Donahue, C.A., Gowda, H., and Pandey, A. (2014). A draft map of the human proteome. Nature 509, 575–581.

Kloc, M., Wilk, K., Vargas, D., Shirato, Y., Bilinski, S., and Etkin, L.D. (2005). Potential structural role of non-coding and coding RNAs in the organization of the cytoskeleton at the vegetal cortex of *Xenopus oocytes*. Development 132, 3445–3457.

Lanz, R.B., McKenna, N.J., Onate, S.A., Albrecht, U., Wong, J., Tsai, S.Y., Tsai, M.J., and O'Malley, B.W. (1999). A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. Cell 97, 17–27.

Li, J.H., Liu, S., Zhou, H., Qu, L.H., and Yang, J.H. (2013). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res 42, D92–D97.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nat Rev Genet 10, 155–159.

Mercer, T.R., and Mattick, J.S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. Nat Struct Mol Biol 20, 300–307.

Gholami, A.M., Hahne, H., Wu, Z., Auer, F.J., Meng, C., Wilhelm, M., and Kuster, B. (2013). Global proteome analysis of the NCI-60 cell line panel. Cell Rep 4, 609–620.

Ruiz-Orera, J., Messeguer, X., Subirana, J.A., and Alba, M.M. (2014). Long non-coding RNAs as a source of new peptides. eLife 3, 598.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA hypothesis: the rosetta stone of a hidden RNA language?. Cell 146, 353–358.

Shteynberg, D., Deutsch, E.W., Lam, H., Eng, J.K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R.L., Aebersold, R., and Nesvizhskii, A.I. (2011). iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol Cell Proteomics 10, M111.007690–M111.007690.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., Weinstock, G.M., Wilson, R.K., Gibbs, R.A., Kent, W.J., Miller, W., and Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res 15, 1034–1050.

Sumazin, P., Yang, X., Chiu, H.S., Chung, W.J., Iyer, A., Llobet-Navas, D., Rajbhandari, P., Bansal, M., Guarnieri, P., Silva, J., and Califano, A. (2011). An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. Cell 147, 370–381.

Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S.M., Ala, U., Karreth, F., Poliseno, L., Provero, P., DiCunto, F., Lieberman, J., Rigoutsos, I., and Pandolfi, P.P. (2011). Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell 147, 344–357.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2012). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat Biotechnol 31, 46–53.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28, 511–515.

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., Watt, A.T., Freier, S.M., Bennett, C.F., Sharma, A., Bubulya, P.A., Blencowe, B.J., Prasanth, S.G., and Prasanth, K.V. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell 39, 925–938.

Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Science 329, 689–693.

Vanderpool, C.K., and Gottesman, S. (2004). Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. Mol Microbiol 54, 1076–1089.

Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 13, 227–232.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer genome landscapes. Science 339, 1546–1558.

Wadler, C.S., and Vanderpool, C.K. (2007). A dual function for a bacterial small RNA: SgrS performs base pairing-dependent regulation and encodes a functional polypeptide. Proc Natl Acad Sci USA 104, 20454–20459.

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.P., and Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res 41, e74–e74.

Ward, L.D., and Kellis, M. (2012). Interpreting noncoding genetic variation in complex traits and human disease. Nat Biotechnol 30, 1095–1106.

Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., and Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42, D1001–D1006.

Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., Liu, J., Horvath, S., and Fan, G. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. Nature 500, 593–597.

Zhao, M., Sun, J., and Zhao, Z. (2013). TSGene: a web resource for tumor suppressor genes. Nucleic Acids Res 41, D970–D976.

Open Access This article is distributed under the terms of the Creative Commons Attribution License, which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

SUPPORTING INFORMATION

Table S1 Matched peptides for those132 bifunctional RNAs.

Table S2 GWAS SNPs that located in bifunctional RNA regions.

Table S3 The noncoding somatic mutations of COSMIC that located in bifunctional RNA regions.

Table S4 The noncoding somatic mutations of COSMIC that located in bifunctional RNA regions

The supporting information is available online at life.scichina.com and www.springerlink.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.