

# SCIENTIFIC REPORTS



OPEN

## Genome-wide analyses of long noncoding RNA expression profiles in lung adenocarcinoma

Zhenzi Peng<sup>1</sup>, Jun Wang<sup>1</sup>, Bin Shan<sup>1,2</sup>, Fulai Yuan<sup>1</sup>, Bin Li<sup>1</sup>, Yeping Dong<sup>1</sup>, Wei Peng<sup>1</sup>, Wenwen Shi<sup>1</sup>, Yuanda Cheng<sup>3</sup>, Yang Gao<sup>3</sup>, Chunfang Zhang<sup>3</sup> & Chaojun Duan<sup>1,3</sup>

**LncRNAs have emerged as a novel class of critical regulators of cancer. We aimed to construct a landscape of lncRNAs and their potential target genes in lung adenocarcinoma. Genome-wide expression of lncRNAs and mRNAs was determined using microarray. qRT-PCR was performed to validate the expression of the selected lncRNAs in a cohort of 42 tumor tissues and adjacent normal tissues. R and Bioconductor were used for data analysis. A total of 3045 lncRNAs were differentially expressed between the paired tumor and normal tissues (1048 up and 1997 down). Meanwhile, our data showed that the expression NONHSAT077036 was associated with N classification and clinical stage. Further, we analyzed the potential co-regulatory relationship between the lncRNAs and their potential target genes using the 'cis' and 'trans' models. In the 25 related transcription factors (TFs), our analysis of The Cancer Genome Atlas database (TCGA) found that patients with lower expression of POU2F2 and higher expression of TRIM28 had a shorter overall survival time. The POU2F2 and TRIM28 co-expressed lncRNA landscape characterized here may shed light into normal biology and lung adenocarcinoma pathogenesis, and be valuable for discovery of biomarkers.**

Lung cancer is the leading cause of death among all types of cancer<sup>1</sup>. Lung cancer has two histological types: small cell lung cancer (SCLC) and non-small lung cancer (NSCLC). The most common histological type of NSCLC is lung adenocarcinoma (ADC) that accounts for approximately 50% of NSCLC<sup>2</sup>. In general, ADC, originated from distal airways, is less frequently associated with chronic inflammation and smoking than squamous cell carcinoma (SCC)<sup>3</sup>. There is still no definitive biomarker for diagnosis and prognosis of ADC although several studies have identified a large number of genes associated with tumor initiation and progression.

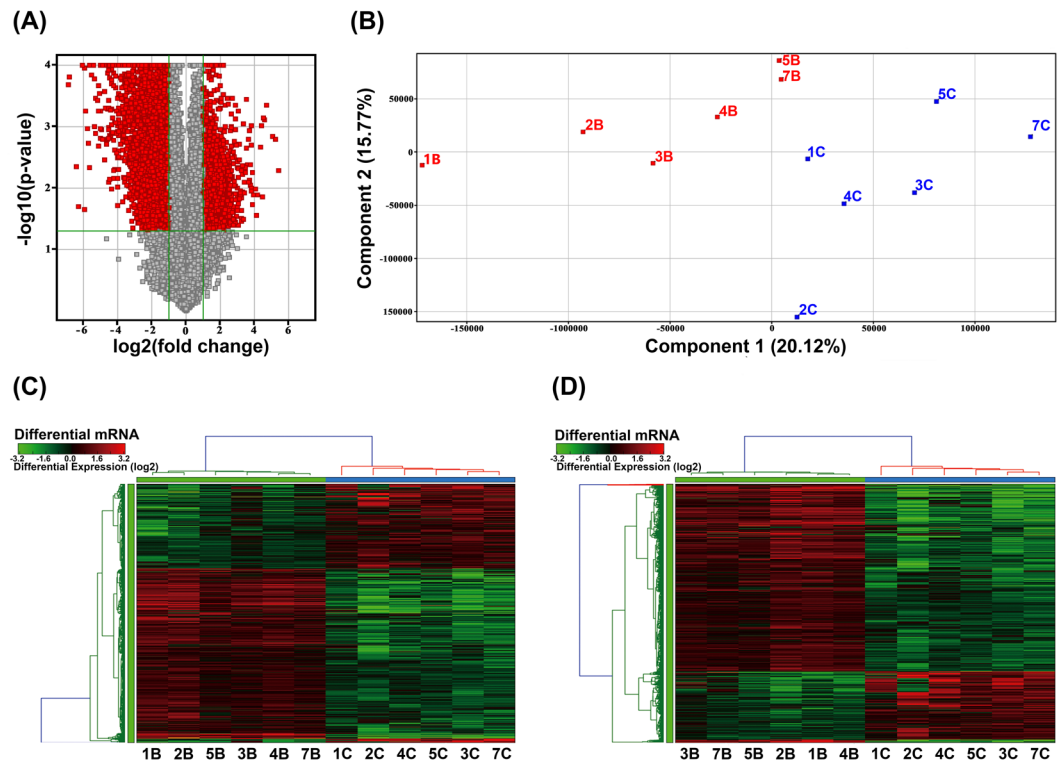
LncRNAs have emerged as important regulators of physiology and disease. However, little is known about the mechanism of their functions<sup>4</sup>. lncRNAs are noncoding RNAs that are longer than 200 nucleotides in length. Recent RNA-SEQ based transcriptome studies revealed that 68% transcripts of human are lncRNAs and approximately 80% of them are unannotated<sup>5</sup>. lncRNAs are categorized into four groups: intronic, exonic, overlapping, and intergenic according to their locations in the genome<sup>6</sup>. lncRNAs are also classified as 'cis' or 'trans' lncRNAs according to their modes of regulation of transcription. lncRNAs can act as decoys, scaffolds, sponges, and guide of the protein and RNA molecules in cells<sup>7</sup>.

To delineate genome-wide ADC-associated lncRNAs expression, we employed locus-by-locus lncRNA and mRNA microarray probes to identify the lncRNAs and mRNAs that are differentially expressed between the ADC tumor tissues and the matched adjacent normal lung tissues. Our results establish a link between lncRNAs and clinical features of ADC, namely age, gender, smoking index, differentiation, TNM stage, and clinical stage. Using the 'cis' and 'trans' mode, we reveal potential co-regulatory relationship between the lncRNAs and their target genes. We also construct a "TFs-lncRNAs" two-element network and a "TFs-lncRNAs-mRNA" three-element network in ADC using hyper geometric distribution. We also identify potential cis-regulatory target genes of the differentially expressed lncRNAs in ADC by screening the co-expressed genes located near the differentially expressed lncRNAs. We identify the TFs that potentially regulate the differentially expressed lncRNAs by combining analysis of the TCGA database and hyper geometric distribution.

<sup>1</sup>Institute of Medical Sciences, Xiangya Hospital, Central South University, Changsha, 410008, P. R. China.

<sup>2</sup>Washington State University, Elson S Floyd College of Medicine, P.O. Box 1495, Spokane, WA 99210-1495, USA.

<sup>3</sup>Department of Thoracic Surgery, Xiangya Hospital, Central South University, Changsha, 410008, P. R. China. Correspondence and requests for materials should be addressed to C.D. (email: [duancjxy@126.com](mailto:duancjxy@126.com))



**Figure 1.** LncRNAs and mRNAs expression profiles in ADC. **(A)** Volcano Plot of the differentially expressed lncRNAs in ADC tumor tissues and adjacent normal lung tissues. The red points in the plot represent differentially expressed lncRNAs with statistical significance. **(B)** Principal Components Analysis. The B group (red plots) represents adjacent normal lung tissues and C group (blue plots) represents the ADC tumor tissues. **(C)** Hierarchical Clustering shows a distinguishable lncRNA expression profile and **(D)** mRNA expression profile.

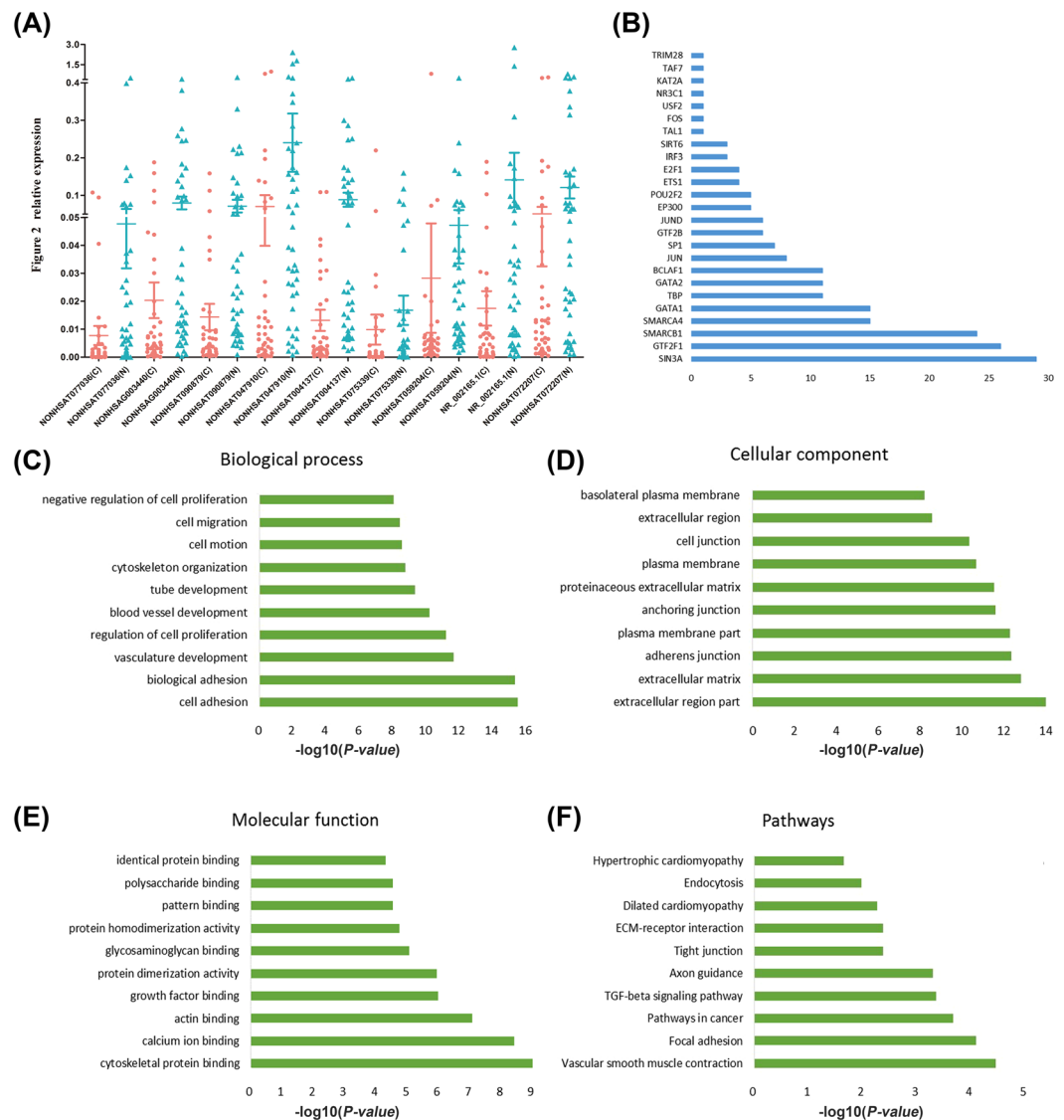
## Results

**LncRNAs and mRNAs expression profile in ADC.** To construct the lncRNA landscape in lung adenocarcinoma, we analyzed lncRNA and mRNA expression profile of 6 lung adenocarcinoma and their matched adjacent normal lung tissues using microarray. In order to avoid complicating factors derived from the heterogeneous nature of ADC clinical characteristics, we chose all female patients who have no smoking history and are between 45 and 55 year old. A volcano plot was used to provide an overview of the dysregulated lncRNAs in our microarray data sets (Fig. 1A). Principal component analysis (PCA) (Fig. 1B) and hierarchical clustering analysis (HCA) were applied to establish and cluster the lncRNA expression profile (Fig. 1C) and the mRNA expression profile (Fig. 1D). This difference distinguished the ADC group from the adjacent normal tissue group. From the microarray data, we identified 3045 lncRNAs and 2602 mRNAs that were differentially expressed between 6 paired ADC tumor tissues and adjacent normal lung tissues (Fold Change  $\geq 2.0$  or  $\leq 0.5$ ,  $p < 0.05$  and False discovery rate (FDR)  $< 0.05$ )<sup>8</sup>. Among these, 1048 lncRNAs and 694 mRNAs were upregulated in all six ADC samples, and 1997 lncRNAs and 1908 mRNAs were downregulated (see Supplementary Table S1, S2). Among them 57 upregulated lncRNAs and 329 downregulated lncRNAs were altered greater than 6.0-fold. The most upregulated lncRNAs was NONHSAT097328 (Fold change: 26.42) and the most downregulated was NR\_038190.1 (Fold change: 117.37).

**Validation of differential lncRNA expression.** To evaluate our microarray performance, we measured the expression of 9 lncRNAs in all 42 paired ADC tumor tissues and adjacent normal lung tissues using qRT-PCR (Fig. 2A). The result showed that the expression patterns of these lncRNAs were consistent with microarray data in ADC tumor tissues and adjacent normal tissues.

To further investigate an association between lncRNAs expression and clinicopathological features in ADC, the patients were categorized by age, TNM stages, and differentiation. Based on the expression in paired tumor and adjacent normal tissues, we found NONHSAT077036 expression was associated with N classification and clinical stage (Table 1  $P = 0.002$ ,  $P = 0.008$ ). Then we further investigated whether the association between NONHSAT077036 and TNM staging was specific to ADC. We measured the expression of NONHSAT077036 in 30 paired SCC tumor tissues and adjacent normal lung tissues using qRT-PCR. The data showed that NONHSAT077036 was not associated with N classification and clinical stage in SCC (Table 2  $P = 0.253$ ,  $P = 0.448$ ). This indicated that NONHSAT077036 was only associated with ADC but not with SCC.

To gain further insight into the role of NONHSAT077036 in ADC, we examined its sequence and structure across species. Generally, lncRNAs are not as conserved as protein coding genes. Therefore, it is difficult to predict



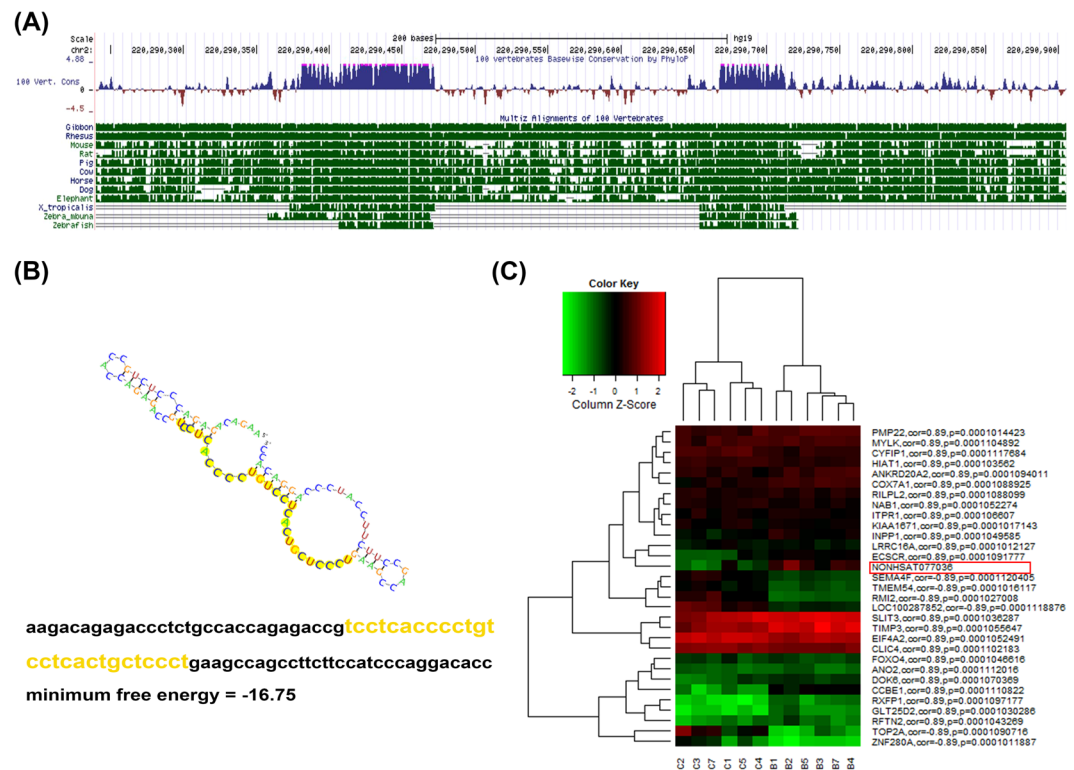
**Figure 2.** Validation of differential lncRNA expression and function. (A) Expression level of the selected lncRNAs in ADC tissues and adjacent normal tissues. qRT-PCR were used to verify differentially expressed lncRNAs ( $P \leq 0.05$ ). Gene expression was normalized to 36B4 expression. The blue points represent cancer tissues and the red points represent adjacent normal tissues. (B) Frequency distribution of lncRNAs enrichment on TFs. The X-axis is frequency distribution and Y-axis is the TFs name. (C–F) Go and Pathway analysis of lncRNAs co-expressed genes. The top 10 enriched terms were calculated as  $-\log_{10}(P\text{-value})$ .

LncRNAs No.	Age (years)	Gender	Smoking Index	Differentiation	T classification	N classification	Clinical Stage
NONHSAT077036	0.809	0.55	0.056	0.242	0.101	0.002	0.008
NONHSAG003440	0.96	0.94	0.461	0.64	0.894	0.399	0.919
NONHSAT090879	0.497	0.591	0.139	0.872	0.847	0.259	0.815
NONHSAT047910	0.114	0.048	0.117	0.682	0.705	0.607	0.823
NONHSAT004137	0.96	0.94	0.461	0.379	0.848	0.399	0.919
NONHSAT075339	0.188	0.55	0.244	0.946	0.646	0.853	0.906
NONHSAT059204	0.96	0.289	0.233	0.39	0.482	0.701	0.21
NR_002165.1	0.748	0.127	0.755	0.966	0.763	0.701	0.925
NONHSAT072207	0.957	0.824	0.102	0.436	0.285	0.091	0.269

**Table 1.** P-value of lncRNAs expression with clinicopathological features in ADC patients.

lncRNAs No.	Age (years)	Gender	Smoking Index	Differentiation	T classification	N classification	Clinical Stage
NONHSAT077036	0.569	0.612	0.688	0.139	0.574	0.253	0.448
NONHSAG003440	0.87	0.014	0.999	0.275	0.041	0.589	0.035
NONHSAT090879	0.424	0.026	0.119	0.378	0.999	0.084	0.378
NONHSAT047910	0.688	0.043	0.261	0.154	0.147	0.845	0.083
NONHSAT004137	0.146	0.645	0.039	0.495	0.466	0.157	0.347
NONHSAT075339	0.596	0.18	0.457	0.422	0.142	0.548	0.105
NONHSAT059204	0.034	0.502	0.352	0.422	0.142	0.548	0.704
NR_002165.1	0.289	0.18	0.457	0.422	0.834	0.489	0.704
NONHSAT072207	0.71	0.814	0.473	0.99	0.085	0.969	0.683

**Table 2.** *P*-value of lncRNAs expression with clinicopathological features in SCC patients.



**Figure 3.** Function prediction of lncRNAs NONHSAT077036. **(A)** Sequence conservation of NONHSAT077036 from humans to zebrafish. **(B)** Sequence similarity between NONHSAT077036 and H19. **(C)** Potential target genes of NONHSAT077036.

lncRNA function based on evolutionary conservation (Fig. 3A). However, NONHSAT077036 showed strong sequence conservation from zebrafish to human, which suggests important functions of NONHSAT077036. Intriguingly NONHSAT077036 exhibited sequence similarity to H19 (132–159, 28 bp), which is also elevated in lung cancer and promotes cancer cell proliferation<sup>9</sup> (Fig. 3B). We then predicted the target genes of NONHSAT077036 using co-expression analysis (Fig. 3C). Among these predicted target genes, many are critical regulators of tumorigenesis. For instance, there is a significant negative correlation between NONHSAT077036 and TOP2A ( $r = -0.89$ ) that is an oncogene involved in G2 checkpoint in response to DNA damage<sup>10</sup>. There is a positive correlation between NONHSAT077036 and CCBE1 ( $r = 0.89$ ) that reported to regulate extracellular matrix remodeling and migration<sup>11</sup>. These correlations suggest regulation of proliferation and invasion by NONHSAT077036 in ADC.

**Identification of potentially functional lncRNAs in ADC.** To predict potential functions of the differentially expressed lncRNAs, we calculated the correlation value of lncRNAs and mRNAs. Differentially expressed lncRNAs were divided into two subsets: the upregulated and downregulated lncRNAs. Top 200 most differentially expressed lncRNAs in each subset were selected for further analysis. lncRNA NR\_038190.1 was the most significantly downregulated lncRNAs. The analysis pertaining to it was shown as a representative result. The top 30 correlations between lncRNA NR\_038190.1 and its target genes were showed in Table 3.

Gene Symbol	p-value	Correlation	Gene Symbol	p-value	Correlation
TBX4	0.000101	0.891002	PIK3C3	0.000107	0.889654
RAPGEF4	0.000101	0.891000	PTPLAD2	0.000107	0.889621
TTC28	0.000102	0.890780	SYNGR1	0.000107	0.889601
SH2D1B	0.000102	0.890723	ZAK	0.000108	0.889386
ITGA8	0.000102	0.890706	APPBP2	0.000109	0.889201
PIP5K1B	0.000102	0.890704	SPATA13	0.000110	0.888938
C1orf145	0.000102	0.890619	CBFA2T3	0.000110	0.888911
ANXA3	0.000103	0.890460	SNX1	0.000110	0.888899
GIMAP1	0.000103	0.890407	CCNDBP1	0.000111	0.888839
FAM83A	0.000103	-0.890375	NHSL1	0.000112	0.888607
BDNF	0.000104	0.890349	FIGF	0.000112	0.888513
PLEKHH2	0.000104	0.890150	WNT7A	0.000113	0.888409
FAM162B	0.000105	0.889993	KIDINS220	0.000113	0.888351
PRKG2	0.000105	0.889954	JPH4	0.000113	0.888299
CTDSP1	0.000106	0.889779	F10	0.000115	0.887960

**Table 3.** Target genes of NR\_038190.1.

We further analyzed the enrichment of GO (<http://geneontology.org/>) and KEGG pathway (<http://www.kegg.jp/kegg/>) terms associated with the lncRNAs that were differentially expressed between ADC and normal tissues. DAVID functional annotation software (<https://david.ncifcrf.gov/home.jsp>)<sup>12</sup> was used to analyze all co-expressed genes. Top 200 upregulated lncRNA genes and 200 downregulated lncRNA genes were subjected to GO and KEGG pathway analyses. We selected the top 200 reliability prediction terms (according to the *p*-value and enrichment) for co-expressed and aberrant lncRNA genes, respectively. The top 200 terms in the GO terms were highly enriched for cell adhesion, proliferation, migration (ontology: molecular function), extracellular region part, adheren junction (ontology: cellular component) and cytoskeletal protein binding, growth factor binding (ontology: molecular function). Top 200 terms in the KEGG pathway were associated with pathways in cancer. The most significant top 10 GO terms and KEGG pathway are shown in Fig. 2C–F. Besides, we calculated the enrichment of functional terms of co-expressed genes for each differentiated lncRNAs (see Supplementary Table S3, S4).

**lncRNAs target prediction.** To explore how lncRNAs function in lung adenocarcinoma, we predicted the cis- and trans-regulated genes of the differentially expressed lncRNAs using co-expression network analysis. The co-expressed genes within 300 kb upstream or downstream from a selected lncRNA were identified as potential “cis” genes of a given lncRNA (*p*-value of correlation  $\leq 0.05$ ). We predicted the cis-regulated genes at the top of differentially expressed lncRNAs (see Supplementary Table S5). In ADC tissue and adjacent normal lung tissue controls, 56 upregulated lncRNAs had 92 ‘cis’ genes and 35 downregulated lncRNAs had 42 ‘cis’ genes. 8 upregulated lncRNAs had at least 3 ‘cis’ genes. The maximal number of cis genes assigned to a differentially expressed lncRNA was 5. The cis relationship of 6 significantly dysregulated (up.58\_NONHSAT053536, up.106\_NONHSAT024969, up.128\_NONHSAG018334, up.129\_ENST00000522875, up.190\_NONHSAT083792, up.195\_ENST00000518528) are shown in Fig. 4A–F.

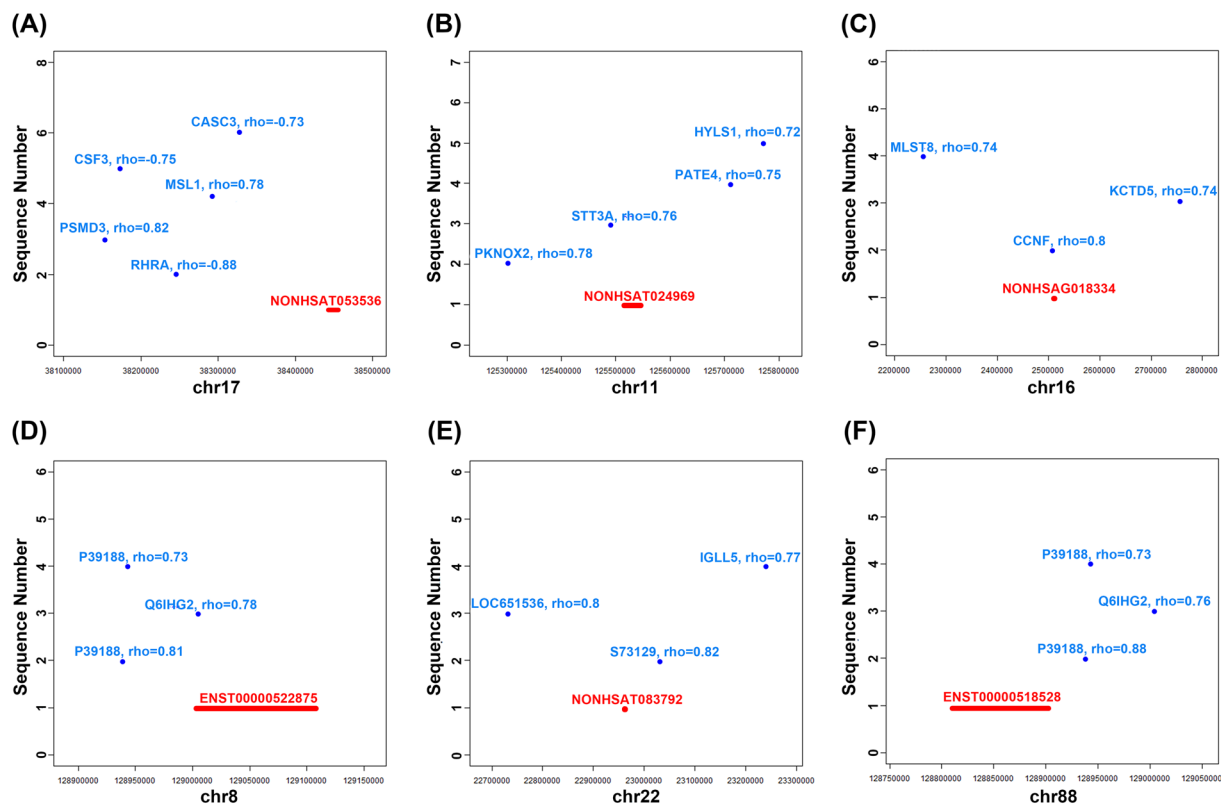
Among all these potential cis-regulated target genes, further analysis showed that the most highly related categories were the processes that affect cell growth, differentiation, and migration. The most correlated genes were related to cell cycle. Among the ‘cis’ genes, EGR1, EHF, PLCE1, PKNOX2, MLST8, FGF17, FOSB, PTPRQ, NKTR, and CAV2 are known to function in cell growth and differentiation. CDK7, CCNE, CDKN3, CDC20, TUBG1, FEN1, and GRK5 are critical regulators of cell cycle. PEA15 and CAV2 are regulators of apoptosis among ‘cis’ genes. JAM3, CD36, and RAB21 are believed to bear a critical role in adhesive processes and cell migration. These cis-regulated lncRNAs are potential regulators of ADC.

It is noteworthy that lncRNAs also act in ‘trans’ to regulate TFs mediated chromatin remodeling and transcription<sup>13</sup>. We intended to discover which TFs might interact with the differentially expressed lncRNAs using hyper geometric distribution that can calculate the overlap of TFs target genes and chromatin regulators with the co-expressed lncRNA genes. As showed in Fig. 2B, the lncRNAs were significantly regulated by 25 TFs: SIN3A, POU2F2, TRIM28, etc. Early studies have demonstrated that these TFs are master regulators of cancer<sup>14–20</sup>. We further analyzed the relationship between the top 25 lncRNA related TFs and overall survival in TCGA database. We observed that patients with lower expression of POU2F2 and higher expression of TRIM28 had shorter overall survival time (Fig. 5A,C), suggesting that POU2F2 and TRIM28 are biomarkers for poor prognosis of ADC. Based on the results of the lncRNA co-expression analysis, we generated “POU2F2-lncRNAs” and “TRIM28-lncRNAs” two-element network by Cytoscape software (Fig. 5B,D). Then we added target co-expressed genes and generated “POU2F1-lncRNAs-mRNAs” and “TRIM28-lncRNAs-mRNAs” three-element relationship tables (see Supplementary Table S6).

## Discussion

Aberrant expression lncRNAs is recognized as a hallmark feature in pathogenesis and progression of various diseases, including lung cancer. ADC is the largest histological phenotypes of lung cancer. However, the





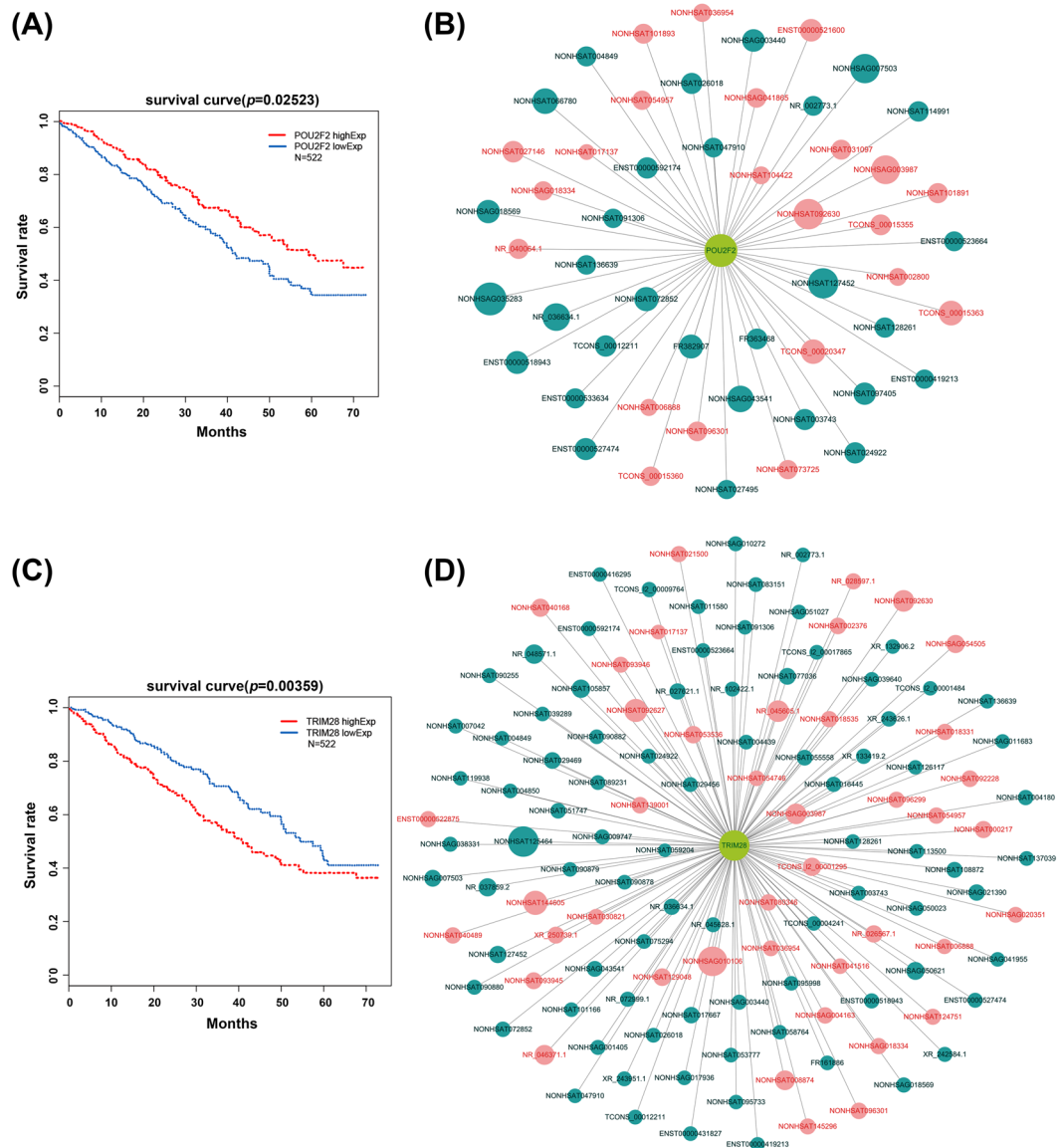
**Figure 4.** Cis-regulation genes of representative lncRNAs in the chromosome. The X-axis represents lncRNA position in chromosome, the Y-axis represents correlation coefficient of lncRNA and potential “cis” genes. The red line represents the genome width of lncRNA and blue point represents the position of potential “cis” genes.

genome-wide expression profile and classification and function of lncRNAs have not been examined in ADC. Unraveling the functions and mechanisms of these lncRNAs can substantially improve our understanding of ADC. Integration of lncRNAs and target genes profiling is also a promising approach to identify effective biomarkers of ADC. Therefore, we screened the genome-wide expression profile of lncRNAs and mRNAs in 6 paired ADC tumor tissues and adjacent normal lung tissues in this study. Microarray data were further validated by qRT-PCR in another 42 paired tissues. We predicted the function of selected lncRNAs according to co-expression genes and Gene Ontology (GO) biological process. In addition, we predicted ‘cis’ and ‘trans’ regulated modes of the 200 tops differentially expressed lncRNAs to find out how they might regulate ADC progression. Several dysregulated lncRNAs reported in the previous studies were also found in our study. The overlapping lncRNA expression profiles between our current study and the previous studies validate the significance of our study. For instance, lncRNA PVT1 is upregulated in various human cancers<sup>21,22</sup>, including lung cancer<sup>23</sup>. In our data, 13 probes were designed to measure PVT1 and all of them were upregulated. The average fold change was 5.45. More importantly, we showed that PVT1 may be regulated by 5 TFs: SIN3A, KAT2A, E2F4, E2F1 and GATA2. Another example is the tumor suppressor FENRR. FENRR inhibits cell proliferation and migration and is downregulated in cancer cell lines and cancerous tissues<sup>24</sup>. In support of this view, our results also showed a 22.57-fold decrease of FENRR expression in ADC tissues. Recent reports indicated that FENRR overexpression suppressed invasion and migration by downregulating fibronectin1 expression. Our results also showed a significant inverse correlation between all 5 FENRR probes and fibronectin1 (Correlation valued  $\geq 0.7$  or  $\leq -0.7$ ). However, functions of the novel differentially expressed lncRNAs identified in our study remain to be characterized experimentally.

It is of note that lung cancer is a highly heterogeneous disease. Our analysis showed that NONHSAT077036 expression was associated clinicopathological features only in ADC patients ( $P < 0.05$ ), but not other histologic subtypes of non-small cell lung cancer. However, additional studies are needed to verify the significance of NONHSAT077036 in ADC. Functional analysis of NONHSAT077036 needs to be carried out in cell and animal based models of ADC.

We also performed Gene Ontology and pathway analysis of co-expressed genes of 400 lncRNAs. Our data indicate that the most related categories include cell adhesion, proliferation, migration, growth factor binding, etc. These GO terms are well-established critical factors in tumorigenesis and tumor progression.

Little is known about the exact function of lncRNAs, although evidence so far indicates that lncRNAs participate in various biological processes. We analyzed the aberrantly expressed lncRNAs in ADC in “cis” and “trans” regulated mechanisms. Our data indicated that cis-regulated target genes participate in initiation and progression of ADC. For instance, EGR1 belongs to the EGR family whose activation is involved in differentiation and



**Figure 5.** TFs and lncRNAs co-expressed network. (A,C) Kaplan-Meier survival analysis of POU2F2 and TRIM28 expression of ADC patients in TCGA database. (B,D) “POU2F2-lncRNAs” and “TRIM28-lncRNAs” two-element network. Green points represent up-regulated lncRNAs. Red points represent down-regulated lncRNAs. The size of each dot is proportional to the magnitude of the change of a given lncRNA.

mitogenesis. In addition, EGR1 supports FGF-dependent angiogenesis during neovascularization and tumor growth<sup>25</sup>. Another study indicates that EGR1 upregulates the expression of lincRNA H19 in liver cancer<sup>26</sup>. In addition, EHF encodes an ETS transcription factor that is expressed in epithelial-specific manner. It has been reported that EHF is silenced by epigenetic mechanisms during NSCLC (non-small cell lung cancer) development<sup>27</sup>. EHF in ovarian cancer cells regulates cell proliferation and G1 phase checkpoint<sup>28</sup>. It is noteworthy that our bioinformatical analysis of the lncRNAs was largely based on a correlation between the expression patterns of lncRNAs and mRNAs. Further studies are needed to experimentally validate the link between lncRNAs and mRNAs identified in the current study.

It is generally accepted that lncRNAs can directly interact with gene promoters and TFs. Many of these lncRNAs recruit protein factors to enhancers and regulate the activity of enhancers<sup>29</sup>. Transcriptional processes are also controlled by lncRNAs via their interaction with primary coding transcripts. A survey of correlation between lncRNAs and TFs also helps us reveal its function. TCGA database contains a large number of clinical information about ADC patients. We carried out survival analysis of 25 related ‘trans’ mode TFs in TCGA database and found that the patients with lower expression of POU2F2 have a shorter overall survival. These findings suggest that POU2F2 may be a biomarker for a poor prognosis of ADC. To date, the role of POU2F2 is controversial. Oct-2, encoded by the gene POU2F2, is a B-cell restricted transcription factor. Emerging evidence indicates that POU2F2 is essential to the later stages of B-cell differentiation<sup>30</sup>. Mice with deletion of POU2F2 die shortly after

birth<sup>31</sup>. Recent evidence indicates that POU2F2 mediates metastasis induced by ROB01 in gastric cancer<sup>32</sup>. In our study, POU2F2 appears to be a tumor suppressor in ADC. Moreover the POU2F2-lncRNAs' two-element network modulates the expression of 53 lncRNAs. Deciphering the functions of POU2F2 in ADC needs further investigations. We identified another lncRNA related TF, TRIM28 in ADC in our study. The prevailing view is that knockdown of TRIM28 expression impairs cell proliferation in NSCLC cell lines. In addition, patients with elevated expression of TRIM28 suffered shorter tumor-specific survival<sup>33</sup>. Our data support this view because the patients with elevated expression of TRIM28 have a shorter overall survival. Moreover the TRIM28-lncRNAs' two-element network modulates the expression of 129 lncRNAs.

It is currently well accepted that molecular networks of multiple genes and pathways, instead of a single gene or a pathway, underlie pathogenesis of cancer. Network analysis has provided an efficient method to model biological processes. It should also be emphasized that dynamic feedback motifs will help us to obtain a unified view of various cellular processes<sup>34,35</sup>. Thus, it is necessary to integrate omics data (gene regulatory networks, cell signaling networks and metabolic networks) in network analysis. Network analysis is an effective approach in predicting potential lncRNA–disease associations. There are a wide range of computational models and web servers that have been developed for this purpose. Chen *et al.* introduced state-of-the-art computational and FMLNCSIM models to identify disease-related lncRNAs from experimental validation. They developed an effective computational models to construct lncRNA functional similarity and the similarity scores (Long non-coding RNAs and complex diseases: from experimental results to computational models) (FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model)<sup>36–40</sup>.

An important limitation of our work is that we have not distinguished different sub-clones. Cancer is a highly heterogeneous disease although one clone can become the dominant population in the tumor at diagnosis<sup>41</sup>. There are many distinct sub-clones that coexist in a tumor and are likely derived from different clonal backup<sup>42,43</sup>. Thus, a drug targeting only one sub-clone within a tumor may have limited effect. Tumor genome-wide analysis is needed to model cancer cells by constructing networks for individual clones.

Our findings reported in the current study warrant further investigations of the mechanisms of the differentially expressed lncRNAs to understand their clinical significance in ADC. The lncRNA profile we established in ADC will lay the foundation for a better understanding of the impacts of lncRNAs in ADC patients. The ADC-associated lncRNAs identified in our study are promising biomarkers with potential in tumor diagnosis, classification, prognosis and therapeutic evaluation.

## Materials and Methods

**Samples.** All lung adenocarcinoma patients were diagnosed at the Thoracic Surgery Department of Xiangya Hospital, Central South University. The cohort of 42 lung adenocarcinoma patients provided written informed consent in compliance with the code of ethics of the World Medical Association. The collection and usage of the clinical specimens were approved by the Xiangya Hospital Medical Research Ethics Committee. All experimental methods were performed in accordance with the relevant guidelines and regulations of Xiangya Hospital Medical Research Ethics Committee and the Scientific Research Project 201403216 (Histopathological Application). The tumor-node-metastasis was classified based on the criteria of the eighth tumor-node-metastasis (TNM) staging system. All tissues were stored at  $-80^{\circ}\text{C}$  after initial freezing in liquid nitrogen. The collection of clinical data of these patients include sex, age, smoking status, differentiation and TNM stage (see Supplementary S7).

**RNA extraction and array data production.** Sample preparation and microarray hybridization were performed by OE Biotech Corporation, Shanghai, P.R. China. Briefly, total RNA was extracted and 200ng RNA was purified from each sample (RNasey Mini Kit (Qiagen p/n 74104). The Agilent 2100 bioanalyzer and RNA LabChip<sup>®</sup> kits was used to assess RNA quality. RNA Integrity Number (RIN)  $\geq 7$  and  $28\text{S}/18\text{S} \geq 0.7$  was used for synthesizing double-stranded cDNA (see Supplementary Table S8). Then cDNA was labeled by Cy3-dCTP and hybridized to the OE Biotech Human  $4 \times 180\text{K}$  lncRNAs chip, which contained 46,506 lncRNAs probes and 30,656 mRNAs probes collected from eight databases, including Agilent ncRNA, GencodeV13, lncRNAdb, H-invDB, RefSeq, NONCODE v3.0, UCR and UCSC lncRNAs Transcripts.

**qRT-PCR validation.** qRT-PCR was used to validate our microarray data. Briefly, total RNA was extracted using Trizol reagent (Invitrogen) and then reverse-transcribed using GoScript<sup>™</sup> Reverse Transcription System (Promega) in accordance with the manufacturer's protocol. Real-time PCR was performed using All-in-One<sup>™</sup> qPCR Mix (GeneCopeia). Specific primers are listed in Table 4. Each sample was normalized by the internal control gene of 36B4. The results represent means of 3 repetitions and were quantified by the  $2^{-\Delta\Delta\text{ct}}$  method. The mRNA levels of lncRNAs between tumor and non-tumor tissues were compared using T-test ( $P < 0.05$ ) using R. The mRNA levels of lncRNAs between cancer and non-cancer tissues were compared using T-test ( $P < 0.05$ ) using R. Pearson  $\times 2$  test was used to analyze the relationship between lncRNAs expression and clinicopathologic parameters in SPSS software (version 20.0, Chicago, IL).

**Bioinformatics analysis.** Data analysis including heat map, volcano plot, PCA and survival was carried out using R by gplots, lattice, MASS, ggplot2, hash and survival packages (<https://www.R-project.org/>)<sup>44</sup>.

**Differential expression analysis.** Differentially expressed lncRNAs and mRNAs were identified using paired t-test (Fold Change  $\geq 2.0$  or  $\leq 0.5$ ,  $p < 0.05$  and FDR  $< 0.05$ ). The microarray data have been uploaded in NCBI Gene Expression Omnibus (GEO) and the GEO accession number is GSE85716. Red indicates high expression and green indicates low expression in tumor tissues.

**lncRNA co-expression analysis.** We evaluated potential co-expression between lncRNAs and mRNAs using Pearson Correlation. A positive correlation between a lncRNA and a mRNA was defined as a Pearson



LncRNAs No.	Position	Primer sequence
NONHSAT077036	Forward	TGAAGAAGTAACAAGCCTGTCT
	Reverse	TGGTCTTGATCATCACCGTCT
NONHSAG003440	Forward	GGAGGAGTGTGGAGGTCAA
	Reverse	TACATGCCTGGGTGAGCTAC
NONHSAT090879	Forward	ATATACTACAGTGCCTGTGTGTCC
	Reverse	AGCAGTTGGATGACAGAGAATAG
NONHSAT047910	Forward	CAAGTCCCAGAATCCTCCAG
	Reverse	AGGCTTACAGGAAATGTGCAG
NONHSAT004137	Forward	AGCCAGTCTAGTGGACAGAGA
	Reverse	CCTGCATTGAATAATCACAAGACCA
NONHSAT075339	Forward	GACTGGGTTTATTACCCTCTCCT
	Reverse	TAAGACTGCCTCTGCCCTTC
NONHSAT059204	Forward	GAGTGTGACCTAGCGCAGAA
	Reverse	GAGCACACCTTCCAAGCAC
NR_002165.1	Forward	ATGGCTAGAAGTGACCCAG
	Reverse	TGCCCAGCCTAGACTTCTC
FR407620	Forward	CACCTCCCTCAAACCTGTCT
	Reverse	GCCAGAATTGCTTGCCCTCAT
NONHSAT072207	Forward	TTGGGAGTGTGCATGAGGTA
	Reverse	TTTGGTTACATGTGGCAGT

**Table 4.** Primers Used for qRT-PCR Analysis of lncRNAs.

Correlation greater than 0.7 and a  $p$ -value less than 0.05. Hypergeometric cumulative distribution function was used to calculate the enrichment of co-expressed mRNAs. The False Discovery rate was determined using the method as previously described. The ontology of co-expressed genes was categorized by gene annotation and summary information obtained from DAVID database<sup>12</sup>. Annotations of the lncRNAs co-expressed mRNAs were determined using GO analysis on cellular component, molecular function, biological processes and specific pathways.

**Prediction of lncRNAs function.** We searched for an lncRNA co-expressed genes within a 300 kb window of each lncRNA in the top 200 up-regulated and downregulated lncRNAs ( $P < 0.05$ ). The co-expressed genes on both sides of an lncRNA were defined as potentially 'cis' regulated genes by a given lncRNA. To examine which genes were potentially 'trans' regulated by lncRNAs we determined which TF might interact with the lncRNAs of interest using Jemboos software. TF target gene sets were obtained from Encyclopedia of DNA Elements (ENCODE). Hypergeometric distribution was used to identify the overlap of TFs target genes and co-expressed genes of lncRNAs.  $p$ -value was used to measure the enrichment of differentially expressed genes in the term. The TF and lncRNAs relationship networks were drawn using Cytoscape software. The TFs survival analysis was performed using the RNA-Seq and survival data extracted from the TCGA database (<https://portal.gdc.cancer.gov/>).

## References

- Siegel, R. L. *et al.* Cancer Statistics, 2016. *CA: A Cancer Journal for Clinicians*. **66**, 7–30 (2016).
- Davidson, M. R. *et al.* The Pivotal Role of Pathology in the Management of Lung Cancer. *J Thorac Dis*. **5**(Suppl 5), S463–S478 (2013).
- Langer, C. J. *et al.* The Evolving Role of Histology in the Management of Advanced Non-Small-Cell Lung Cancer. *J Clin Oncol*. **28**, 5311–5320 (2010).
- Prensner, J. R. & Chinnaiyan, A. M. The Emergence of lncRNAs in Cancer Biology. *Cancer Discov*. **1**, 391–407 (2011).
- Iyer, M. K. *et al.* The Landscape of Long Noncoding RNAs in the Human Transcriptome. *Nat Genet*. **47**, 199–208 (2015).
- Derrien, T. *et al.* The GENCODE V7 Catalog of Human Long Noncoding RNAs: Analysis of their Gene Structure, Evolution, and Expression. *Genome Res*. **22**, 1775–1789 (2012).
- Shi, X. *et al.* Long Non-Coding RNAs: A New Frontier in the Study of Human Diseases. *Cancer Lett*. **339**, 159–166 (2013).
- Guttman, M. *et al.* Chromatin Signature Reveals Over a Thousand Highly Conserved Large Non-Coding RNAs in Mammals. *Nature*. **458**, 223–227 (2009).
- Matouk, *et al.* The non-coding RNAs of the H19-IGF2 imprinted loci: A focus on biological roles and therapeutic potential in Lung Cancer. *Journal of Translational Medicine*. **13**, 113 (2015).
- Al-Kuraya, K. *et al.* HER2, TOP2A, CCND1, EGFR and C-MYC oncogene amplification in colorectal cancer. *Journal of Clinical Pathology*. **60**, 768–772 (2006).
- Fiona, C. *et al.* Linkage and sequence analysis indicate that CCBE1 is mutated in recessively inherited generalised lymphatic dysplasia. *Human Genetics*. **127**, 231–241 (2010).
- DAVID Bioinformatics Resources 6.8. *Nature Protocols* 2009; **4**(1):44 & *Nucleic Acids Res*. **37**, (2009).
- Guttman, M. & Rinn, J. L. Modular Regulatory Principles of Large Non-Coding RNAs. *Nature*. **482**, 339–346 (2012).
- Suzuki, H. *et al.* Decreased Expression of the Sin3a Gene, a Candidate Tumor Suppressor Located at the Prevalent Allelic Loss Region 15Q23 in Non-Small Cell Lung Cancer. *Lung Cancer*. **59**, 24–31 (2008).
- Le Loarer, F. *et al.* Smarcb4 Inactivation Defines a Group of Undifferentiated Thoracic Malignancies Transcriptionally Related to Baf-Deficient Sarcomas. *Nat Genet*. **47**, 1200–1205 (2015).
- Agaimy, A. *et al.* Smarcb1(LN1) -Deficient Sinonasal Basaloid Carcinoma: A Novel Member of the Expanding Family of Smarcb1-Deficient Neoplasms. *Am J Surg Pathol*. **38**, 1274–1281 (2014).

17. Tessema, M. *et al.* Gata2 is Epigenetically Repressed in Human and Mouse Lung Tumors and is Not Requisite for Survival of Kras Mutant Lung Cancer. *J Thorac Oncol.* **9**, 784–793 (2014).
18. Qu, Y. *et al.* Knockdown of NF-KappaB p65 Subunit Expression Suppresses Growth of Nude Mouse Lung Tumour Cell Xenografts by Inhibition of Bcl-2 Apoptotic Pathway. *Cell Biochem Funct.* **33**, 320–325 (2015).
19. Heckman, C. A. *et al.* Oct transcription factors mediate (14;18) lymphoma cell survival by directly regulating bcl-2 expression. *Oncogene.* **25**, 888–898 (2006).
20. Chen, Y. *et al.* Inactivation of PI3-K/Akt and Reduction of Sp1 and P65 Expression Increase the Effect of Solamargine On Suppressing EP4 Expression in Human Lung Cancer Cells. *J Exp Clin Cancer Res.* **34**, 154 (2015).
21. Takahashi, Y. *et al.* Amplification of PVT-1 is Involved in Poor Prognosis Via Apoptosis Inhibition in Colorectal Cancers. *Brit J Cancer.* **110**, 164–171 (2013).
22. Kong, R. *et al.* Long Noncoding RNA PVT1 Indicates a Poor Prognosis of Gastric Cancer and Promotes Cell Proliferation through Epigenetically Regulating p15 and p16. *Mol Cancer.* **14** (2015).
23. Cui, D. *et al.* Long Non-Coding RNA PVT1 as a Novel Biomarker for Diagnosis and Prognosis of Non-Small Cell Lung Cancer. *Tumour Biol.* (2015).
24. Xu, T. P. *et al.* Decreased Expression of the Long Non-Coding RNA FENDRR is Associated with Poor Prognosis in Gastric Cancer and FENDRR Regulates Gastric Cancer Cell Metastasis by Affecting Fibronectin1 Expression. *J Hematol Oncol.* **7**, 63 (2014).
25. Fahmy, R. G. *et al.* Transcription Factor Egr-1 Supports FGF-Dependent Angiogenesis During Neovascularization and Tumor Growth. *Nat Med.* **9**, 1026–1032 (2003).
26. Li, H. *et al.* Mir675 Upregulates Long Noncoding RNA H19 through Activating EGR1 in Human Liver Cancer. *Oncotarget.* **6**, 31958–31984 (2015).
27. Song, S. *et al.* Gene Silencing Associated with SWI/SNF Complex Loss During NSCLC Development. *Molecular cancer research: MCR.* **12**, 560–570 (2014).
28. Cheng, Z. *et al.* Knockdown of EHF Inhibited the Proliferation, Invasion and Tumorigenesis of Ovarian Cancer Cells. *Mol Carcinogen.* (2015).
29. Zhenzi, P. *et al.* Functions and mechanisms of long noncoding RNAs in lung cancer. *OncoTargets and Therapy.* **9**, 4411–4424 (2016).
30. Hodson, D. J. *et al.* Regulation of Normal B-Cell Differentiation and Malignant B-Cell Survival by Oct2. *Pnatl Acad Sci.* **113**, E2039–E2046 (2016).
31. Lynn, M. C. *et al.* Oct-2, although not required for early B-cell development, is critical for later B-cell maturation and for postnatal survival. *Genes & Development.* **7**, 570–582 (1993).
32. Wang, S. *et al.* POU2F2-Oriented Network Promotes Human Gastric Cancer Metastasis. *Gut.* (2015).
33. Liu, L. *et al.* TRIM28, a New Molecular Marker Predicting Metastasis and Survival in Early-Stage Non-Small Cell Lung Cancer. *Cancer Epidemiol.* **37**, 71–78 (2013).
34. Cloutier, M. *et al.* Dynamic modeling and analysis of cancer cellular network motifs. *Integr Biol (Camb).* **3**, 724–732 (2011).
35. McGee, S. R. *et al.* Network analysis reveals a signaling regulatory Loop in the PIK3CA -mutated Breast Cancer Predicting Survival Outcome. *Genomics, Proteomics & Bioinformatics.* **15**, 121–129 (2017).
36. Chen, X. *et al.* Long non-coding RNAs and complex diseases: from experimental results to computational models. *Briefings in Bioinformatics.* **18**, 558–576 (2017).
37. Chen, X. *et al.* IRWRLDA: improved random walk with restart for lncRNA-disease association prediction. *Oncotarget.* **7**, 57919–57931 (2016).
38. Chen, X. *et al.* Predicting lncRNA-disease associations and constructing lncRNA functional similarity network based on the information of miRNA. *Sci Rep.* **5** (2015).
39. Chen, X. *et al.* Novel human lncRNA-disease association inference based on lncRNA expression profiles. *Bioinformatics.* **29**, 2617–2624 (2013).
40. Chen, X. *et al.* FMLNCSIM: fuzzy measure-based lncRNA functional similarity calculation model. *Oncotarget.* **7**, 45948–45958 (2016).
41. Shaoqiang, W. *et al.* PRKAR1A is a functional tumor suppressor inhibiting ERK/Snail/E-cadherin pathway in lung adenocarcinoma. *Sci. Rep.* **6**, 39630 (2016).
42. Wang, E. *et al.* Cancer systems biology in the genome sequencing era: Part 1, dissecting and modeling of tumor clones and their networks. *Seminars in Cancer Biology.* **23**, 279–285 (2013).
43. Wang, E. *et al.* Cancer systems biology in the genome sequencing era: Part 2, evolutionary dynamics of tumor clonal networks and drug resistance. *Seminars in Cancer Biology.* **23**, 286–292 (2013).
44. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2016).

## Acknowledgements

We are grateful for all the contributions that supported this study. This work was supported by National Natural Science Foundation of China Grants (Nos. 81171841, 81200366, 81372515, 81401901 and 81572281).

## Author Contributions

Zhenzi Peng contributed to analysis and interpretation of data and drafting of the manuscript. Zhenzi Peng performed the experiments. Jun Wang contributed to clinical samples and data. Bin Shan contributed to revise the manuscript for important intellectual content interpretation of data. Bin Li, Fulai Yuan, Yeping Dong, Weiwei Peng, Wenwen Shi, Yuanda Chen and Yang Gao contributed to acquisition of data and technical support. Chunfang Zhang contributed to the acquisition of data and revised the manuscript for important intellectual content. Chaojun Duan contributed to study concept and design, analysis and interpretation of data and drafting of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-15712-y>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017