

Integrative analysis reveals clinical phenotypes and oncogenic potentials of long non-coding RNAs across 15 cancer types

Ze-Lin Wang^{1,2,*}, Bin Li^{1,2,*}, Stephen R. Piccolo^{3,4}, Xiao-Qin Zhang⁵, Jun-Hao Li^{1,2}, Hui Zhou^{1,2}, Jian-Hua Yang^{1,2}, Liang-Hu Qu^{1,2}

¹Key Laboratory of Gene Engineering of the Ministry of Education, Sun Yat-sen University, Guangzhou, China

²State Key Laboratory for Biocontrol, Sun Yat-sen University, Guangzhou, China

³Department of Biology, Brigham Young University, Provo, Utah, USA

⁴Department of Biomedical Informatics, University of Utah, Salt Lake City, Utah, USA

⁵Department of Surgery, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong

*These authors have contributed equally to this work

Correspondence to: Jian-Hua Yang, **e-mail:** yangjh7@mail.sysu.edu.cn
Liang-Hu Qu, **e-mail:** lssqlh@mail.sysu.edu.cn

Keywords: lncRNA, expression profile, prognostification, somatic copy number alteration

Received: October 22, 2015

Accepted: April 11, 2016

Published: April 27, 2016

ABSTRACT

Long non-coding RNAs (lncRNAs) have been shown to contribute to tumorigenesis. However, surprisingly little is known about the comprehensive clinical and genomic characterization of lncRNAs across human cancer. In this study, we conducted comprehensive analyses for the expression profile, clinical outcomes, somatic copy number alterations (SCNAs) profile of lncRNAs in ~7000 clinical samples from 15 different cancer types. We identified significantly differentially expressed lncRNAs between tumor and normal tissues from each cancer. Notably, we characterized 47 lncRNAs which were extensively dysregulated in at least 10 cancer types, suggesting a conserved function in cancer development. We also analyzed the associations between lncRNA expressions and patient survival, and identified sets of lncRNAs that possessed significant prognostic values in specific cancer types. Our combined analysis of SCNA data and expression data uncovered 116 dysregulated lncRNAs are strikingly genomic altered across 15 cancer types, indicating their oncogenic potentials. Our study may lay the groundwork for future functional studies of lncRNAs and help facilitate the discovery of novel clinical biomarkers.

INTRODUCTION

More than 80% of mammalian genome sequences are transcribed, but a major proportion of the transcripts do not code for functional proteins [1–3]; these transcripts are termed non-coding RNAs (ncRNAs). These ncRNAs can generally be divided into two classes based on their length: short ncRNAs (<200 nt) and long ncRNAs (lncRNAs, >200 nt). Short ncRNAs include miRNAs, which have been extensively studied in the context of cancer during the past decade [4–7]. By contrast, the functional role of lncRNAs in cancers remains largely unknown, although their number far exceeds that of miRNAs. An increasing number of studies have shown that lncRNAs may play important regulatory roles in gene expression at various levels, including at the epigenetic, transcriptional and

post-transcriptional levels, and therefore they may affect the development and progression of cancer [8–11].

Recent advances in high-throughput technologies (e.g., RNA-Seq and microarray) have provided comprehensive ways to characterize the genomic profiles of lncRNAs. Indeed, the application of these approaches has revealed that the aberrant expression of specific lncRNAs may also act as biomarkers for cancer diagnosis and prognosis in glioma [12, 13], prostate cancer [14] and liver cancer [15]. In addition, through the integration of bioinformatics analyses of lncRNA SCNAs and expression profiles, a few lncRNAs with oncogenic activity have also been reported in human cancers. These include the PVT1 amplification in combination with MYC [16], FAL1 amplification and its oncogenic role in ovarian cancer [17], and PCAN-R1/2 amplification in the proliferation

of prostate cancer cells [18]. However, interpreting the potential roles of thousands of lncRNAs in tumorigenesis across diverse cancer types remains a daunting challenge.

It has been widely accepted that cancer is fundamentally a genetic disease. Cancers of disparate organs may share common molecular features, whereas, conversely, cancers from the same organ can also be markedly distinct. For example, p53 mutations are common in serous ovarian, serous endometrial and basal-like breast cancer, all of which share a global transcriptional signature that involves the activation of similar oncogenic pathways [19, 20]. Such examples illustrate the importance of performing comprehensive molecular profiling analyses across multiple tumor types [21–23]. Such analyses may help to identify the commonalities and/or differences in molecular aberrations across cancer types and to assess therapeutic implications. A prime example is a pan-cancer analysis of SCNA patterns across 11 cancer types, which identified common patterns of SCNAs across cancer types and provided insight into the mechanisms of generation and functional consequences of cancer-related SCNAs [24]. We and others have used TCGA expression data to characterize cancer-related lncRNAs and their interaction maps [25–29] in multiple cancer types. As an increasing amount of cancer genomics data have become available, there remains a need to integrate various cancer genomics and clinical datasets to reveal the clinical phenotypes and driver potentials of lncRNAs in a pan-cancer manner.

To determine the landscape of lncRNA signatures across various cancer types, in the present study, we

performed a large-scale lncRNA expression and SCNA profiling analysis in ~7000 clinical specimens from 15 different cancer types. We performed systemic comparison analysis of lncRNA profiles within and across cancer types and identified hundreds of dysregulated lncRNAs in each cancer and characterized 47 extensively dysregulated lncRNAs linked to at least 10 cancer types. We further confirmed our results by experimental q-PCR. We also identified many survival-related lncRNAs in each cancer which were independent of age and gender by multivariate Cox regression analysis. Finally, we analyzed the copy number variation of lncRNAs and suggested their oncogenic potentials.

RESULTS

Global lncRNA expression profiles across 15 cancer types

We first identified global expression signatures based on 985 lncRNAs that we observed across 6910 specimens, including matched normal tissues from 15 cancer types by performing unsupervised clustering analysis (Figure 1A; Supplementary Table 1). This pan-cancer clustering analysis revealed a highly tissue-specific lncRNA expression pattern for multiple cancer types (Figure 1B), which is in agreement with previous studies of normal human tissues [30, 31]. Out of the 15 cancer types that were analyzed, samples from 11 of the following cancer

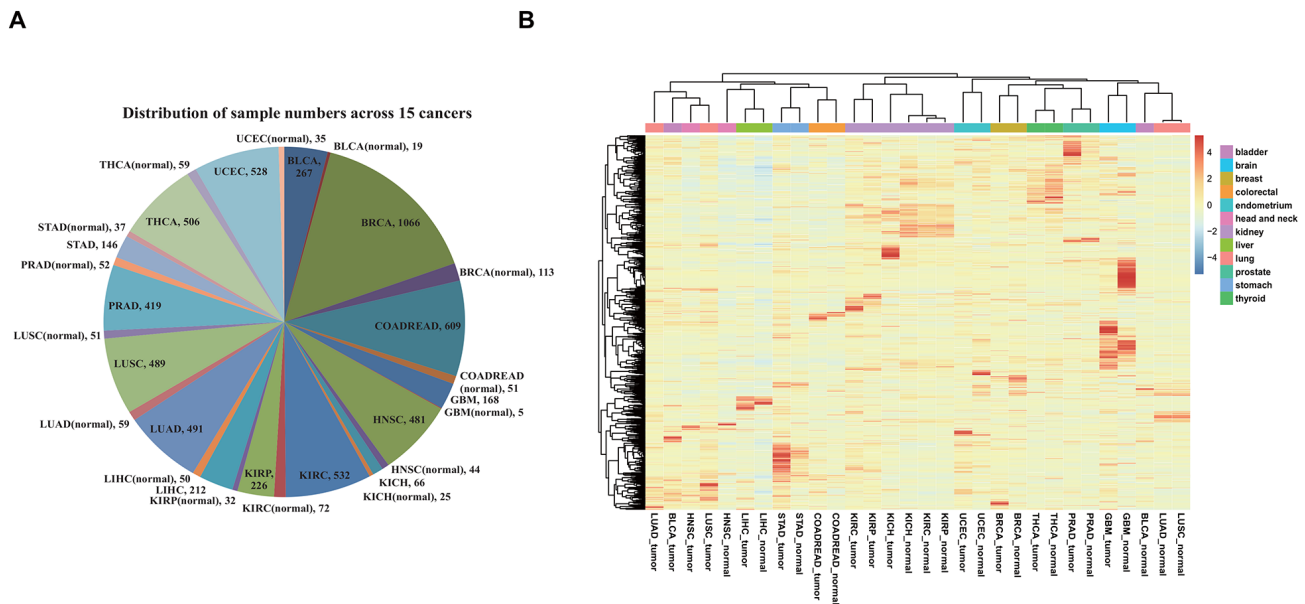


Figure 1: Overview of all specimens and lncRNAs. **A.** Distributions of all specimens across 15 cancer types. **B.** Global expression profiles of 985 lncRNAs across cancer types. The tree displays their average expression values. The mean was computed from all specimens derived from the same type of normal tissue or tumor. Clustering was performed using the average and correlation metrics in the *heatmap* function.

types tended to cluster together, irrespective of disease status (tumor or normal): breast cancer (BRCA), colon and rectum adenocarcinoma (COADREAD), glioblastoma multiforme (GBM), kidney carcinoma (KICH, KIRC, and KIRP), liver hepatocellular carcinoma (LIHC), prostate adenocarcinoma (PRAD), stomach adenocarcinoma (STAD), thyroid cancer (THCA), and uterine corpus endometrial carcinoma (UCEC). However, bladder urothelial carcinoma (BLCA), head and neck squamous cell carcinoma (HNSC), and lung carcinoma (LUAD and LUSC) presented significantly different signatures between tumor tissues and matched normal tissues, which suggests that these lncRNAs have potential to distinguish between these tissue types. Moreover, we discovered that many different tumor types—including BLCA, HNSC, LUAD and LUSC—showed similar signatures to each other, which suggests that a common expression of some lncRNAs may be ubiquitous in cancers.

Identification of dysregulated lncRNAs within each individual cancer type

We subsequently identified lncRNAs that are differentially expressed between tumors and normal tissues within each of the 15 cancer types analyzed. Using an $FDR < 0.05$ and a fold change > 2 as the threshold, we identified significantly dysregulated lncRNAs for each cancer type (Figure 2A; Supplementary Table 2). Of these, we identified 145 significantly dysregulated lncRNAs in STAD (with a minimum number), and 369 lncRNAs in

KICH (with a maximum number). Interestingly, across 15 cancer types, we discovered that more lncRNAs tended to be down-regulated than up-regulated (median 16.3% vs 7.3%). Our results included some well-known cancer-associated lncRNAs such as HOTAIR [32], PCA3 [33], PCAT1 [34], and CRNDE [35]. These lncRNAs demonstrated a similar pattern of dysregulation as previous studies for the specific cancer types in our study, which suggests the robustness of our approach. We also identified many novel dysregulated lncRNAs. For example, ANKRD34C-AS1 was found to be markedly down-regulated in GBM (~38-fold, $FDR = 4.62e-15$), whereas UNC5B-AS1 was up-regulated in THCA (~17-fold, $FDR = 3.52e-22$), and TTC21B-AS1 was up-regulated in KIRC (~92-fold, $FDR = 2.11e-30$).

Identification of commonly dysregulated lncRNAs across multiple cancer types

We further cross-compared the dysregulated lncRNAs that were identified from each cancer type. This cross-comparison identified 651 (of 811) lncRNAs that were dysregulated across at least two cancer types (Supplementary Table 2), which indicates a more common dysregulation pattern among multiple cancer types. For example, the lncRNAs LOC100128593 and PGM5-AS1 showed the most pervasive down-regulation in 13 cancer types; several well-characterized lncRNAs such as HOTAIR [32, 36], H19 [37–39] and PVT1 [16, 40, 41], also showed dysregulation in at least nine different cancer

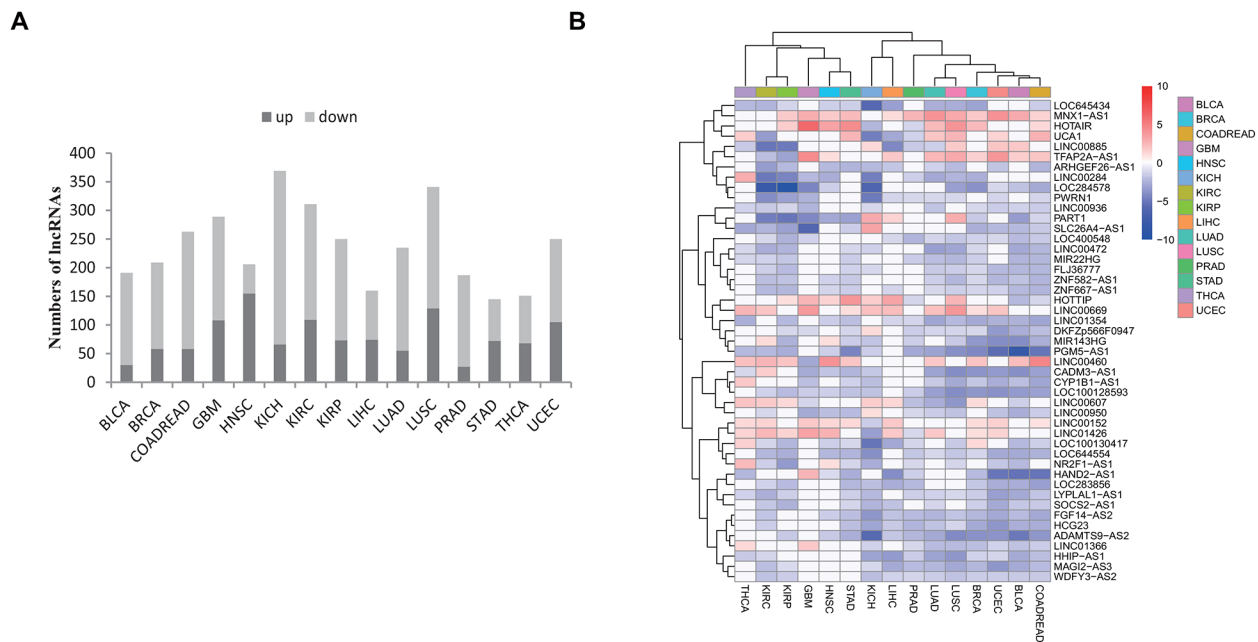


Figure 2: Identification of dysregulated lncRNAs in cancer. **A.** The bar plot shows the numbers of dysregulated lncRNAs in each cancer type. The dark color represents up-regulation whereas the light color represents down-regulation. **B.** 47 lncRNAs show more pervasive patterns of dysregulation in $\geq 3/10$ tumor types. The tree displays their fold change levels after \log_2 transformation. Clustering was performed using the average and correlation metrics in the *heatmap* function.

types. An expression signature consisting of 47 commonly dysregulated lncRNAs in over 2/3 of the cancer types analyzed (≥ 10) is depicted in Figure 2B. It is interesting that most of these lncRNAs were down-regulated as opposed to up-regulated across the 15 cancer types. The identification of commonly dysregulated lncRNAs across multiple cancer types indicates that these lncRNAs are possibly involved in the common and fundamental pathways of human tumorigenesis.

Experimental validation of lncRNA dysregulation by q-PCR

To confirm the alterations we observed for the above-mentioned lncRNAs, which were identified from the RNA-Seq data, we conducted quantitative real-time PCR (q-PCR). Based on the availability of cancer cell lines as well as corresponding normal controls (see Materials and Methods), we performed q-PCR validation for COAD/READ. We randomly selected three lncRNAs that were significantly up-regulated in tumors compared

with normal tissues according to the above RNA-Seq data analysis for experimental validation (SNHG15, MAFG-AS1 and SLCO4A1-AS1) (Figure 3A). In agreement with these findings, the q-PCR results confirmed the changes in expression patterns for the three lncRNAs in all eight CRC cell lines (compared with normal colon cells; Figure 3B). This suggests the reliability of our RNA-Seq analysis.

Evaluation of the prognostic power of lncRNAs

We assessed the prognostic significance of lncRNAs by multivariate Cox regression analysis with gender and age as covariates. With a threshold of $p < 0.05$, we identified survival-related lncRNAs in each cancer type (ranging from 32 to 310 in number; Figure 4A; Supplementary Table 3). Of these, 32 lncRNAs (out of 985) were significant prognostic indicators for PRAD, and 310 lncRNAs were prognostic indicators for KIRC (Supplementary Figure 1A and 1B). Further analysis revealed that 77.4% of these prognostic lncRNAs appeared in at least two different cancers, which suggests

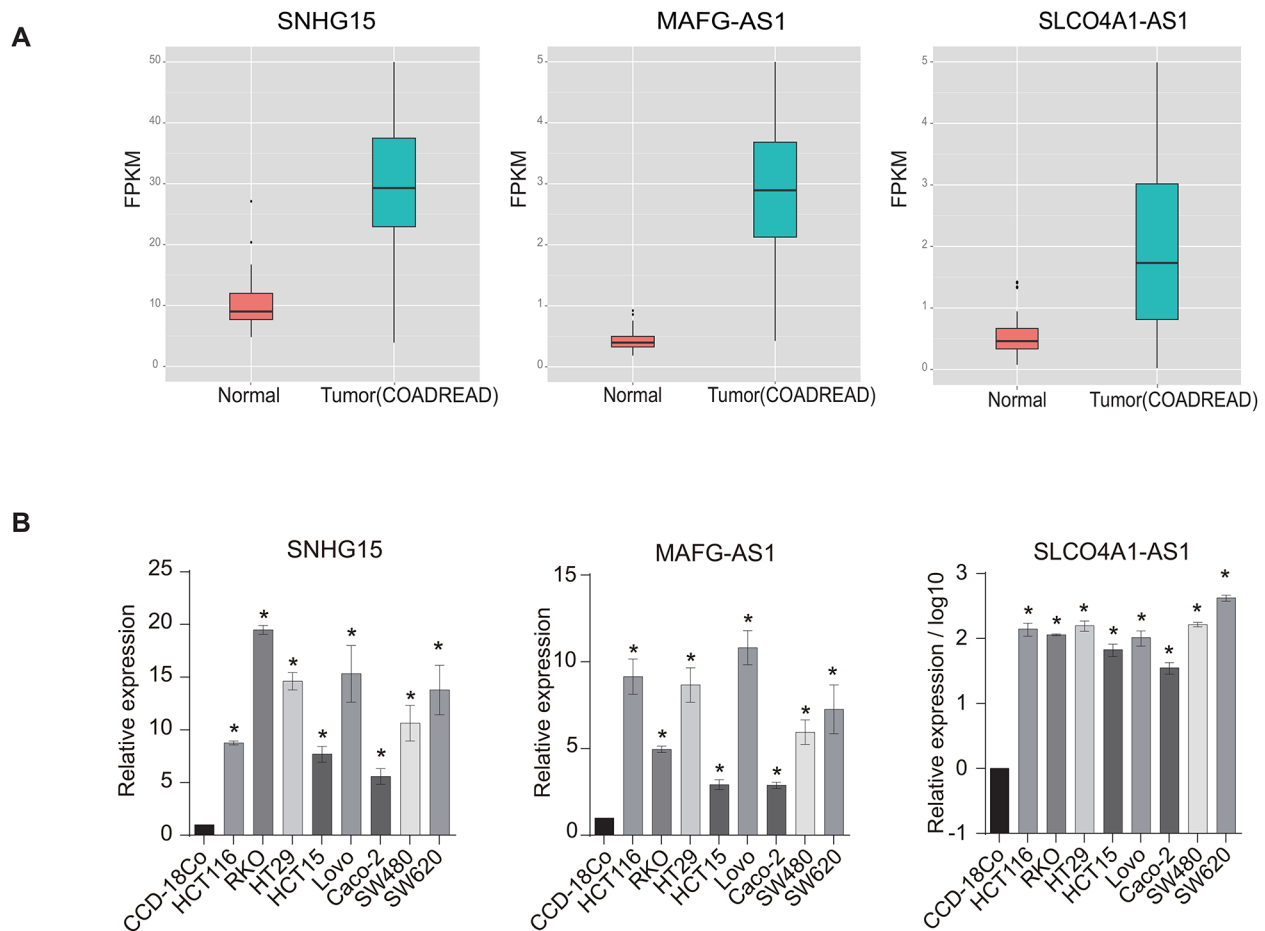


Figure 3: Experimental validation of dysregulated lncRNAs in colorectal cancer (CRC) cell lines. **A.** The box plot shows three lncRNAs that are significantly up-regulated in CRC relative to normal tissues according to RNA-Seq data. **B.** Three randomly selected up-regulated lncRNAs were validated by q-PCR across eight CRC cell lines ($p < 0.05$).

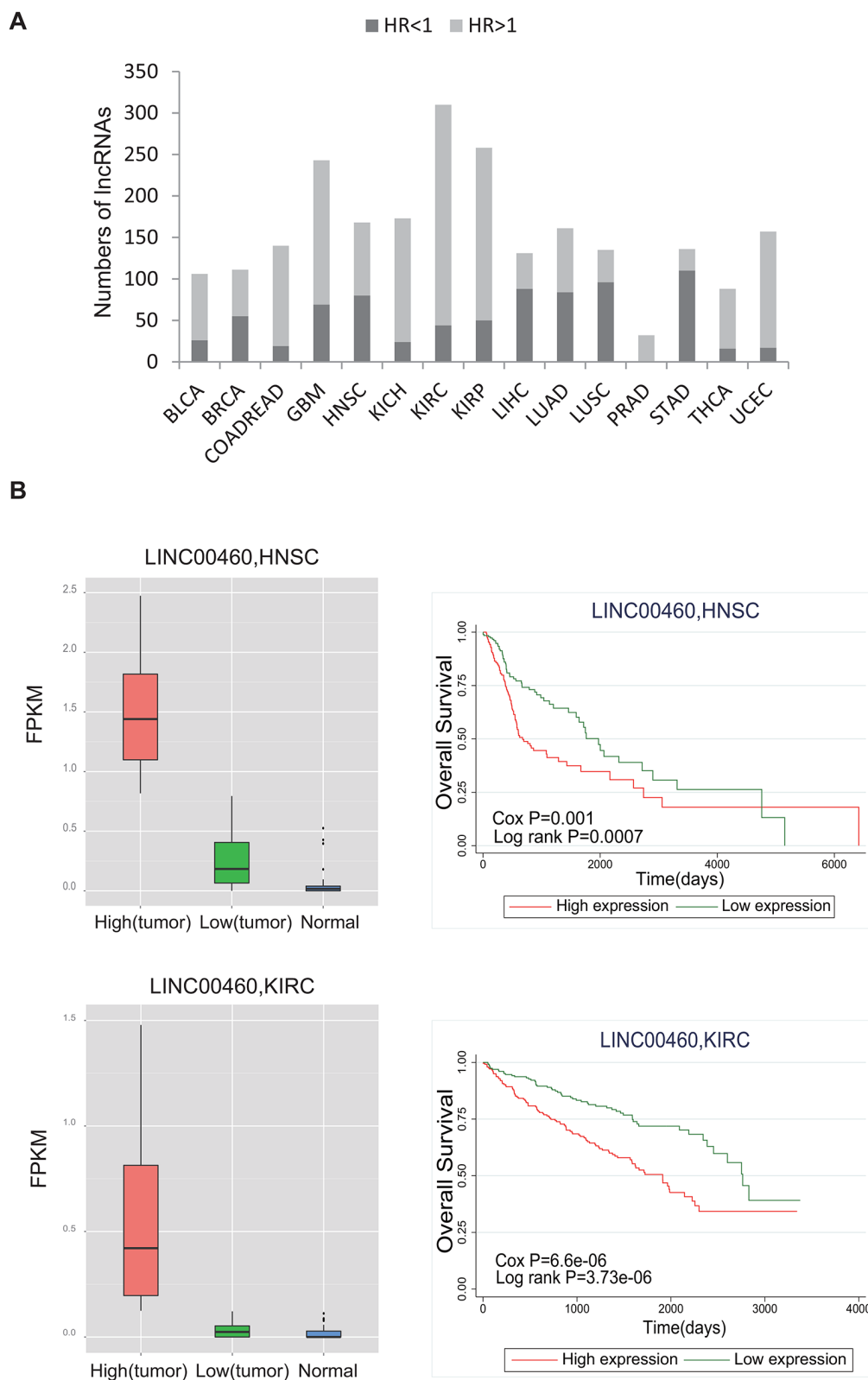


Figure 4: Evaluation of the prognostic power of lncRNAs. **A.** The bar plot shows the number of significantly prognostic lncRNAs in each cancer type. The dark color represents HR<1, and the light color represents HR>1. **B.** The box plot shows differential expression of LINC00460 between tumor and normal tissue for HNSC and KIRC. Survival curves demonstrate that higher expression of LINC00460 is associated with worse survival. LINC00460 expression values were dichotomized into low and high expression groups using the within-cohort median expression value as a cutoff. Red curves represent the high expression group, and green curves represent the low expression group. Survival curves were plotted using Kaplan–Meier methods, and statistical significance was assessed using log-rank tests.

their potential power as general prognostic biomarkers. For example, LOC90768 was found to be significantly associated with overall patient survival in nine different cancer types. Moreover, lncRNA expression tended to be associated with an increased risk of cancer (hazard ratio (HR)>1; Figure 4A). For instance, for LINC00460, which is up-regulated in HNSC and KIRC relative to normal tissues, higher expression is significantly correlated with worse survival (Figure 4B).

Somatic copy number analysis reveals oncogenic potentials of lncRNAs

It has been suggested that genes with causal roles in oncogenesis are often located in the SCNAs that are frequently altered across tumors [16–18, 24, 42]. Thus, we mapped lncRNA genes to regions of recurrent SCNA regions identified from 15 cancer types. Among 985 lncRNAs, we found 85.7% of the lncRNAs were located in recurrent SCNA regions and approximately half of them presented significantly positive correlation with copy number (FDR<0.05; Figure 5A; Supplementary Table 4). This demonstrated SCNA may be an important mechanism for disrupting expression of lncRNAs. Notably, most of these lncRNAs were located in deleted regions rather than in amplified regions (median 3.8% vs 1.1%). Moreover, it was interesting that many of these lncRNAs were extensively expressed in multiple cancer types (Figure 5B). For example, PVT1, a lncRNA that was consistently amplified and expressed in four different cancer types, is located at the chromosome site 8q24. This region is commonly amplified in multiple human cancers [16, 40, 41]. Another example is CHKB-AS1, the deletion of which was consistently present in 10 different cancer types in our study. Strikingly, 116 dysregulated lncRNAs were markedly genomic altered across 15 cancer types (Figure 5C; Supplementary Table 5), suggesting their oncogenic potentials. For example, PVT1 was significantly amplified and up-regulated in GBM, KIRC and LIHC; PCAT1 was significantly amplified and up-regulated in GBM and PRAD.

DISCUSSION

The dysregulated expression patterns of lncRNAs, as well as their biological relevance, have been reported in an increasing number of cancer types [32, 43–46]. However, most of these studies have been restricted to single tumor types and have used a modest number of samples. In the present study, we integrated multiple types of data, including expression profiling data, SCNA data and clinical information for ~7000 clinical samples from 15 cancer types and then conducted a large-scale pan-cancer analysis. We identified differently expressed lncRNAs within and cross-cancer. We also evaluated the prognostic power of lncRNAs and identified a group of

lncRNAs as general biomarkers that may simultaneously predict the prognosis of multiple cancers. Moreover, we also suggested potential oncogenic lncRNAs via an integrative analysis of the association between the expression of lncRNAs and SCNAs.

We found that lncRNAs tend to be down-regulated rather than up-regulated in cancers compared to normal tissues, suggesting a major tumor-suppressive potential. We reported many novel, markedly dysregulated lncRNAs that have not been reported by previous studies. For example, expression levels for TTC21B-AS1 are specifically increased ~92-fold in KIRC compared with normal tissues. These cancer-specific and dysregulated lncRNAs may act as potential novel clinical biomarkers. Cross-cancer analysis identified a group of lncRNAs that was commonly dysregulated among multiple cancer types. For example, PVT1 was found to be up-regulated in nine different cancer types. PVT1 has been reported to be an oncogene in colorectal cancer (CRC), the knockdown of which could inhibit cell proliferation and the invasion capabilities of CRC cells via TGF-beta signaling and apoptotic pathways [47]. The lncRNA HOTAIR was extensively up-regulated in eight different cancer types. HOTAIR was originally found to be up-regulated in breast and colon cancers and was correlated with poor prognosis based on its interaction with PRC2 and its epigenetic regulation of metastasis-related genes [32, 36]. The common dysregulation of these lncRNAs in multiple cancer types suggests that these lncRNAs may be involved in the fundamental pathways that are important to the initiation and progression of various cancers. These lncRNAs may therefore lead to the discovery of new drug targets [48].

Given the widespread change in lncRNA expression across different cancer types, is there a general cancer-related mechanism that effects lncRNA biosynthesis? Indeed, aberrant expression of some molecules similarly tend to the same cause across different cancer types. For example, oncogenic PVT1 are overexpressed across multiple cancer types because of amplification [16]; TERT presented increased expression across multiple cancer types because of recurrent somatic mutations in the promoter [49]; MEG3 are underexpressed in tumors mainly due to promoter silencing by hypermethylation [50]. However, expression of other molecules can be controlled by multiple factors. An example, HOTAIR, overexpressed in multiple cancer types because of effects of c-Myc, TGF-β or other regulatory factors [51]. Thus, in this instance, it is hard to say there is a general cancer-related mechanism that effects lncRNA biosynthesis. If a lncRNA, we argue, is certainly need for cancer development, it will always be active via multiple selective mechanisms such as copy number variation or mutation.

We also assessed the associations between the expression of lncRNAs and patient survival and identified a group of lncRNAs with significant prognostic value.

provide a wider perspective on the prognostic power of these lncRNAs.

In addition, through integrating SCNAs and expression data, we revealed sets of lncRNAs that were strikingly genomic altered across 15 cancer types. Meanwhile, their expressions were significantly positive correlation with gene copy number. Strikingly, 116 of them show significant dysregulation in tumor relative to normal tissues. Previous studies have identified suchlike lncRNAs as drivers such as FAL1 [17] and PCAN-R1/2 [18]. These data may be valuable for investing the function of lncRNAs and lay the groundwork for future experiment studies.

Our study has unique advantages over related studies [29, 52]. First, more cancer types were included in this study, which was two times as many as previous studies. Thus, we revealed many commonly dysregulated lncRNAs among cancer types instead of the few reported by previous studies. This greatly increased our outstanding for the dysregulation of lncRNAs across cancer types. Second, we used the expression level generated by the Rsubread pipeline that produced more consistent expression levels across replicate samples than the TCGA pipeline [53]. In addition, we applied limma package for differential expression analysis, this method has been demonstrated as the best practice among various of software packages in RNA-seq studies [54]. Third, our results illuminated the prognostic landscape of lncRNAs across cancer types, and suggested potential prognostic biomarkers for further investigation and clinical translation.

In conclusion, we performed a large-scale pan-cancer analysis of lncRNAs via an integration of the matched expression profiles, the SCNA profiles, and the clinical information for ~7000 specimens across 15 different cancer types. We identified a group of clinically relevant lncRNAs that may act as potential drivers and biomarkers for cancer typing, prognosis and targeted therapy. Our study provided a landscape of lncRNAs in cancer and may accelerate the pace of experimental or clinical studies of lncRNAs in human cancer.

MATERIALS AND METHODS

RNA-Seq and the collection and processing of clinical data

For the 15 cancer types we analyzed, summarized expression values and clinical information (Supplementary Table 6) for all available tumors and for at least 5 normal samples per cancer type were downloaded from NCBI's Gene Expression Omnibus (GEO) via accession number GSE62944 [53]. These data include expression values for lncRNA and protein-coding genes. In creating this data set, raw RNA-Sequencing reads and the corresponding clinical information for a total of 6910 specimens from

15 cancer types had been downloaded via <https://cghub.ucsc.edu> and the TCGA Data Portal [55]. The reads had been processed and normalized using the Rsubread package (version 1.14.2) [56] and aligned to the UCSC hg19 reference genome. The featureCounts function was used to summarize the gene expression values as integers. These summarized gene values were then normalized to FPKM values. To accurately annotate lncRNA, we used our lncSeeker pipeline [57] to filter the lncRNAs downloaded from Refseq database. The filter steps were briefly described as follow: (1) remove the transcripts with length less than 200 nt. (2) Transcripts were excluded from further consideration if they overlap with non-lincRNA annotation, such as protein-coding genes, pseudogenes and other small RNAs. (3) Coding potential filter. Transcripts with ORF <100 aa, ORF coverage less than 30% and txCdsPredict score <800 were classified as lncRNAs. (4) Known protein domains filter. Transcript was mapped to unitRef90 database with BlastX and transcripts with E-value < 1E-30 were removed. Moreover, Transcripts with a Pfam hit showing both the full sequence E-value < 1E-5 and the single best domain E-value < 1E-5 were removed. In total, 985 known lncRNAs were measured in our study (Supplementary Table 7).

Data analysis and statistical methods

Unsupervised clustering was performed using average and correlation options in the “pheatmap” software package in R (version 3.2.0). Significant differences in the expression of lncRNAs between any two comparison groups were defined using the “limma” software package in the same R version. A false discovery rate (FDR) value of less than 0.05 and a fold change (FC) of greater than 2 were defined as statistically significant. The correlation between lncRNA expression and the overall survival of the patients was assessed by both Cox regression analysis and the Kaplan-Meier estimation method. In the Cox regression analysis, the lncRNA was evaluated as a continuous variable with age and gender as additional covariables. For the Kaplan-Meier estimates, we defined the high-expression groups and low-expression groups using the median lncRNA expression value as a cut-off point. Two-group survival curves were assessed for significance using a log-rank test.

Somatic copy number analysis of lncRNAs

SCNA regions identified using the GISTIC 2.0 [58] software were downloaded from the Firebrowse website (<https://confluence.broadinstitute.org/display/GDAC/Dashboard-Analyses>) created by the Broad Institute. We identified “wide peaks” of the focally amplified or deleted regions defined by GISTIC2.0; the lncRNAs within these regions were identified using BEDTools [59]. The correlation between copy number values and the

corresponding gene expression levels was estimated using R (version 3.2.0, Pearson correlation), and p-values were adjusted using an FDR correction.

q-PCR validation of the dysregulated lncRNAs in cell lines

Human colon carcinoma cell lines HCT116, RKO, HT29, HCT15, Lovo, Caco-2, SW480 and SW620 were obtained from the Cell Bank of the Type Culture Collection of the Chinese Academy of Sciences (Shanghai, China) and were cultured according to the manufacturer's instructions. The CCD-18Co normal human colon cell line was a gift from Dr Yanyun Zhang (The Institute of Health Sciences, SIBS, CAS / SJTUSM, Shanghai, China) and was cultured in MEM medium (Gibco) supplemented with 0.11 g/L Sodium Pyruvate (Sigma), 10% FBS and 1% penicillin-streptomycin (Gibco) at 37°C in a humidified atmosphere with 95% air and 5% CO₂.

Total cellular RNA was extracted using TRIzol (Invitrogen) according to the manufacturer's instructions. cDNA was then synthesized using a PrimeScript™ RT reagent Kit with gDNA Eraser (TaKaRa). q-PCR was performed with SYBR Premix Ex Taq II (Ti RNaseH Plus) (TaKaRa) in a Roche LightCycler480 System with three biological replicates. The comparative Ct Method ($\Delta\Delta$ CT Method) was used to quantify the relative gene expression; β -actin was used for normalization. The gene-specific primers are listed in Supplementary Table 8.

ACKNOWLEDGMENTS

We thank The Cancer Genome Atlas (TCGA) project and its contributors for providing this valuable public data set.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

GRANT SUPPORT

This research is supported by the Ministry of Science and Technology of China, National Basic Research Program (No. 2011CB811300); the National Natural Science Foundation of China (No. 30900820, 31230042, 31370791, 31471223, 91440110); funds from Guangdong Province (No. S2012010010510, S2013010012457); The project of Science and Technology New Star in ZhuJiang Guangzhou city (No. 2012J2200025); Fundamental Research Funds for the Central Universities (No. 2011330003161070, 14lgjc18); China Postdoctoral Science Foundation (No.200902348); the project of Science and Technology of Guangzhou (201504010022). This research is supported in part by the Guangdong Province Key Laboratory of Computational

Science and the Guangdong Province Computational Science Innovative Research Team.

Author Contributions

JHY, LHQ and ZLW conceived the project and conception. ZLW, SRP, JHL, XQZ and JHY performed the computational and statistical analysis. BL performed the experimental validation. ZLW, JHY, BL, SRP, XQZ, JHL, LHQ, and HZ wrote the manuscript. LHQ and JHY supervised the project. All authors read and approved the final manuscript.

REFERENCES

1. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhu X, Rinn JL, Tongprasit W, Samanta M, Weissman S, Gerstein M, Snyder M. Global identification of human transcribed sequences with genome tiling arrays. *Science*. 2004; 306:2242-2246.
2. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest AR, et al. The transcriptional landscape of the mammalian genome. *Science*. 2005; 309:1559-1563.
3. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, et al. Landscape of transcription in human cells. *Nature*. 2012; 489:101-108.
4. Zhang L, Huang J, Yang N, Greshock J, Megraw MS, Giannakakis A, Liang S, Naylor TL, Barchetti A, Ward MR, Yao G, Medina A, O'Brien-Jenkins A, Katsaros D, Hatzigeorgiou A, Gimotty PA, et al. microRNAs exhibit high frequency genomic alterations in human cancer. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:9136-9141.
5. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:2257-2261.
6. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR. MicroRNA expression profiles classify human cancers. *Nature*. 2005; 435:834-838.
7. Croce CM. Causes and consequences of microRNA dysregulation in cancer. *Nature reviews Genetics*. 2009; 10:704-714.
8. Bonasio R, Shiekhattar R. Regulation of transcription by long noncoding RNAs. *Annual review of genetics*. 2014; 48:433-455.

9. Lee JT. Epigenetic regulation by long noncoding RNAs. *Science*. 2012; 338:1435-1439.
10. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nature reviews Genetics*. 2009; 10:155-159.
11. Ulitsky I, Bartel DP. lincRNAs: genomics, evolution, and mechanisms. *Cell*. 2013; 154:26-46.
12. Zhang X, Sun S, Pu JK, Tsang AC, Lee D, Man VO, Lui WM, Wong ST, Leung GK. Long non-coding RNA expression profiles predict clinical phenotypes in glioma. *Neurobiology of disease*. 2012; 48:1-8.
13. Zhang XQ, Sun S, Lam KF, Kiang KM, Pu JK, Ho AS, Lui WM, Fung CF, Wong TS, Leung GK. A long non-coding RNA signature in glioblastoma multiforme predicts survival. *Neurobiology of disease*. 2013; 58:123-131.
14. Crea F, Watahiki A, Quagliata L, Xue H, Pikor L, Parolia A, Wang Y, Lin D, Lam WL, Farrar WL, Isogai T, Morant R, Castori-Eppenberger S, Chi KN, Wang Y, Helgason CD. Identification of a long non-coding RNA as a novel biomarker and potential therapeutic target for metastatic prostate cancer. *Oncotarget*. 2014; 5:764-774. doi: 10.18632/oncotarget.1769.
15. Yuan JH, Yang F, Wang F, Ma JZ, Guo YJ, Tao QF, Liu F, Pan W, Wang TT, Zhou CC, Wang SB, Wang YZ, Yang Y, Yang N, Zhou WP, Yang GS, et al. A long noncoding RNA activated by TGF-beta promotes the invasion-metastasis cascade in hepatocellular carcinoma. *Cancer cell*. 2014; 25:666-681.
16. Tseng YY, Moriarity BS, Gong W, Akiyama R, Tiwari A, Kawakami H, Ronning P, Reuland B, Guenther K, Beadnell TC, Essig J, Otto GM, O'Sullivan MG, Largaespada DA, Schwertfeger KL, Marahrens Y, et al. PVT1 dependence in cancer with MYC copy-number increase. *Nature*. 2014; 512:82-86.
17. Hu X, Feng Y, Zhang D, Zhao SD, Hu Z, Greshock J, Zhang Y, Yang L, Zhong X, Wang LP, Jean S, Li C, Huang Q, Katsaros D, Montone KT, Tanyi JL, et al. A functional genomic approach identifies FAL1 as an oncogenic long noncoding RNA that associates with BMI1 and represses p21 expression in cancer. *Cancer cell*. 2014; 26:344-357.
18. Du Z, Fei T, Verhaak RG, Su Z, Zhang Y, Brown M, Chen Y, Liu XS. Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nature structural & molecular biology*. 2013; 20:908-913.
19. Network TCGA, Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, Fulton LL, Dooling DJ, Ding L. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012; 490:61-70.
20. Network TCGA, Getz G, Gabriel SB, Cibulskis K, Lander E, Sivachenko A, Sougnez C, Kandoth C, Dooling D, Fulton R. Integrated genomic characterization of endometrial carcinoma. *Nature*. 2013; 497:67-73.
21. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics*. 2013; 45:1113-1120.
22. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. *Nature genetics*. 2013; 45:1127-1133.
23. Omberg L, Ellrott K, Yuan Y, Kandoth C, Wong C, Kellen MR, Friend SH, Stuart J, Liang H, Margolin AA. Enabling transparent and collaborative computational analysis of 12 tumor types within The Cancer Genome Atlas. *Nature genetics*. 2013; 45:1121-1126.
24. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, Lawrence MS, Zhang CZ, Wala J, Mermel CH, Sougnez C, Gabriel SB, Hernandez B, Shen H, Laird PW, Getz G, et al. Pan-cancer patterns of somatic copy number alteration. *Nature genetics*. 2013; 45:1134-1140.
25. Li JH, Liu S, Zheng LL, Wu J, Sun WJ, Wang ZL, Zhou H, Qu LH, Yang JH. Discovery of Protein-lncRNA Interactions by Integrating Large-Scale CLIP-Seq and RNA-Seq Datasets. *Frontiers in bioengineering and biotechnology*. 2014; 2:88.
26. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic acids research*. 2014; 42:D92-97.
27. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, Barrette TR, Prensner JR, Evans JR, Zhao S, Poliakov A, Cao X, Dhanasekaran SM, Wu YM, Robinson DR, Beer DG, et al. The landscape of long noncoding RNAs in the human transcriptome. *Nature genetics*. 2015; 47:199-208.
28. White NM, Cabanski CR, Silva-Fisher JM, Dang HX, Govindan R, Maher CA. Transcriptome sequencing reveals altered long intergenic non-coding RNAs in lung cancer. *Genome biology*. 2014; 15:429.
29. Cabanski CR, White NM, Dang HX, Silva-Fisher JM, Rauck CE, Cicka D, Maher CA. Pan-cancer transcriptome analysis reveals long noncoding RNAs with conserved function. *RNA biology*. 2015; 12:628-642.
30. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J, Veeravalli L, Ruan X, Ruan Y, Lassmann T, Carninci P, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research*. 2012; 22:1775-1789.
31. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development*. 2011; 25:1915-1927.
32. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, Tsai MC, Hung T, Argani P, Rinn JL, Wang Y, Brzoska

- P, Kong B, Li R, West RB, van de Vijver MJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010; 464:1071-1076.
33. Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, Debruyne FM, Ru N, Isaacs WB. DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer research*. 1999; 59:5975-5979.
 34. Prensner JR, Iyer MK, Balbin OA, Dhanasekaran SM, Cao Q, Brenner JC, Laxman B, Asangani IA, Grasso CS, Kominsky HD, Cao X, Jing X, Wang X, Siddiqui J, Wei JT, Robinson D, et al. Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nature biotechnology*. 2011; 29:742-749.
 35. Graham LD, Pedersen SK, Brown GS, Ho T, Kassir Z, Moynihan AT, Vizgoft EK, Dunne R, Pimlott L, Young GP, Lapointe LC, Molloy PL. Colorectal Neoplasia Differentially Expressed (CRNDE), a Novel Gene with Elevated Expression in Colorectal Adenomas and Adenocarcinomas. *Genes cancer*. 2011; 2:829-840. doi: 10.1177/1947601911431081.
 36. Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, Tanaka F, Shibata K, Suzuki A, Komune S, Miyano S, Mori M. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer research*. 2011; 71:6320-6326.
 37. Bartolomei MS, Zemel S, Tilghman SM. Parental imprinting of the mouse H19 gene. *Nature*. 1991; 351:153-155.
 38. Barsyte-Lovejoy D, Lau SK, Boutros PC, Khosravi F, Jurisica I, Andrulis IL, Tsao MS, Penn LZ. The c-Myc oncogene directly induces the H19 noncoding RNA by allele-specific binding to potentiate tumorigenesis. *Cancer research*. 2006; 66:5330-5337.
 39. Pantoja C, de Los Rios L, Matheu A, Antequera F, Serrano M. Inactivation of imprinted genes induced by cellular stress and tumorigenesis. *Cancer research*. 2005; 65:26-33.
 40. Guan Y, Kuo WL, Stilwell JL, Takano H, Lapuk AV, Fridlyand J, Mao JH, Yu M, Miller MA, Santos JL, Kalloger SE, Carlson JW, Ginzinger DG, Celniker SE, Mills GB, Huntsman DG, et al. Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. *Clinical cancer research*. 2007; 13:5745-5755.
 41. Wang F, Yuan JH, Wang SB, Yang F, Yuan SX, Ye C, Yang N, Zhou WP, Li WL, Li W, Sun SH. Oncofetal long noncoding RNA PVT1 promotes proliferation and stem cell-like property of hepatocellular carcinoma cells by stabilizing NOP2. *Hepatology*. 2014; 60:1278-1290.
 42. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho YJ, Haery L, Greulich H, et al. The landscape of somatic copy-number alteration across human cancers. *Nature*. 2010; 463:899-905.
 43. Pandey GK, Mitra S, Subhash S, Hertwig F, Kanduri M, Mishra K, Fransson S, Ganeshram A, Mondal T, Bandaru S, Ostensson M, Akyurek LM, Abrahamsson J, Pfeifer S, Larsson E, Shi L, et al. The risk-associated long noncoding RNA NBAT-1 controls neuroblastoma progression by regulating cell proliferation and neuronal differentiation. *Cancer cell*. 2014; 26:722-737.
 44. Yang X, Song JH, Cheng Y, Wu W, Bhagat T, Yu Y, Abraham JM, Ibrahim S, Ravich W, Roland BC, Khashab M, Singh VK, Shin EJ, Yang X, Verma AK, Meltzer SJ, et al. Long non-coding RNA HNF1A-AS1 regulates proliferation and migration in oesophageal adenocarcinoma cells. *Gut*. 2014; 63:881-890.
 45. Srikantan V, Zou Z, Petrovics G, Xu L, Augustus M, Davis L, Livezey JR, Connell T, Sesterhenn IA, Yoshino K, Buzard GS, Mostofi FK, McLeod DG, Moul JW, Srivastava S. PCGEM1, a prostate-specific gene, is overexpressed in prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97:12216-12221.
 46. Xing Z, Lin A, Li C, Liang K, Wang S, Liu Y, Park PK, Qin L, Wei Y, Hawke DH, Hung MC, Lin C, Yang L. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell*. 2014; 159:1110-1125.
 47. Takahashi Y, Sawada G, Kurashige J, Uchi R, Matsumura T, Ueo H, Takano Y, Eguchi H, Sudo T, Sugimachi K, Yamamoto H, Doki Y, Mori M, Mimori K. Amplification of PVT-1 is involved in poor prognosis via apoptosis inhibition in colorectal cancers. *British journal of cancer*. 2014; 110:164-171.
 48. Malek E, Jagannathan S, Driscoll JJ. Correlation of long non-coding RNA expression with metastasis, drug resistance and clinical outcome in cancer. *Oncotarget*. 2014; 5:8027-8038. doi: 10.18632/oncotarget.2469.
 49. Vinagre J, Almeida A, Populo H, Batista R, Lyra J, Pinto V, Coelho R, Celestino R, Prazeres H, Lima L, Melo M, da Rocha AG, Preto A, Castro P, Castro L, Pardal F, et al. Frequency of TERT promoter mutations in human cancers. *Nature communications*. 2013; 4:2185.
 50. Zhou Y, Zhang X, Klibanski A. MEG3 noncoding RNA: a tumor suppressor. *Journal of molecular endocrinology*. 2012; 48:R45-53.
 51. Hajjari M, Salavaty A. HOTAIR: an oncogenic long non-coding RNA in different cancers. *Cancer biology & medicine*. 2015; 12:1-9.
 52. Yan X, Hu Z, Feng Y, Hu X, Yuan J, Zhao SD, Zhang Y, Yang L, Shan W, He Q, Fan L, Kandalaft LE, Tanyi JL, Li C, Yuan CX, Zhang D, et al. Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. *Cancer cell*. 2015; 28:529-540.
 53. Rahman M, Jackson LK, Johnson WE, Li DY, Bild AH, Piccolo SR. Alternative preprocessing of RNA-Sequencing

- data in The Cancer Genome Atlas leads to improved analysis results. *Bioinformatics*. 2015; 31:3666-3672.
54. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. *Briefings in bioinformatics*. 2015; 16:59-70.
 55. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008; 455:1061-1068.
 56. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research*. 2013; 41:e108.
 57. Zheng LL, Li JH, Wu J, Sun WJ, Liu S, Wang ZL, Zhou H, Yang JH, Qu LH. deepBase v2.0: identification, expression, evolution and function of small RNAs, lncRNAs and circular RNAs from deep-sequencing data. *Nucleic acids research*. 2016; 44:D196-202.
 58. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhi R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome biology*. 2011; 12:R41.
 59. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26:841-842.