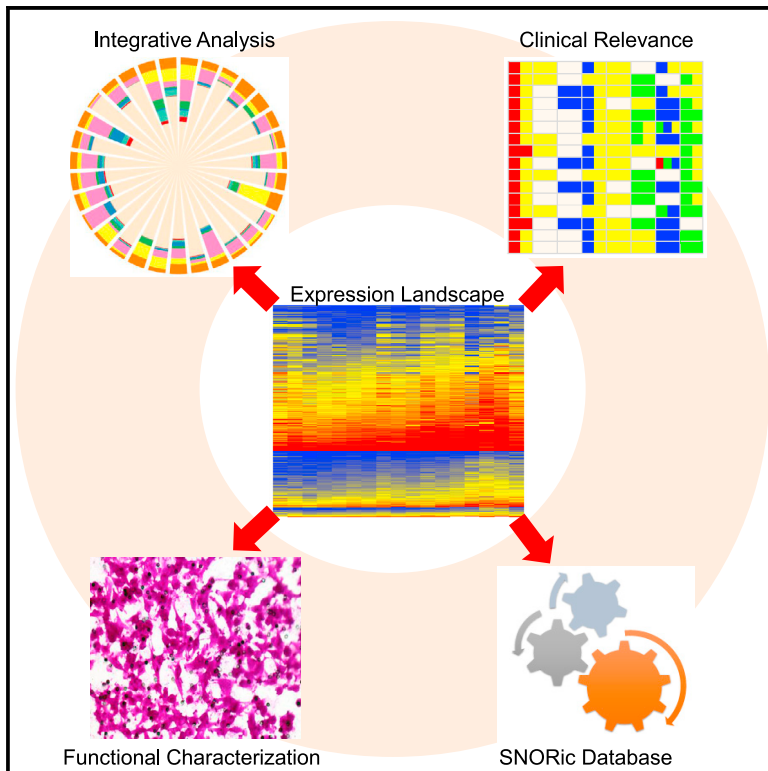


Cell Reports

A Pan-cancer Analysis of the Expression and Clinical Relevance of Small Nucleolar RNAs in Human Cancer

Graphical Abstract



Authors

Jing Gong, Yajuan Li, Chun-jie Liu, ..., An-Yuan Guo, Chunru Lin, Leng Han

Correspondence

guoay@mail.hust.edu.cn (A.-Y.G.),
clin2@mdanderson.org (C.L.),
leng.han@uth.tmc.edu (L.H.)

In Brief

Gong et al. analyze snoRNA expression landscape in >10,000 samples across 31 cancer types and perform integrative analyses. They prioritize 46 significant clinically relevant snoRNAs and characterize the functional roles of SNORD46. The data portal, SNORic, allows exploration of snoRNA expression.

Highlights

- Comprehensive expression landscape of snoRNAs generated from TCGA miRNA-seq data
- Regulation of snoRNA expression by RNPs, host genes, CNV, and DNA methylation
- Identification of 46 clinically relevant snoRNAs and characterization of SNORD46
- A data portal to explore snoRNA expression across different cancer types



A Pan-cancer Analysis of the Expression and Clinical Relevance of Small Nucleolar RNAs in Human Cancer

Jing Gong,^{1,8,9} Yajuan Li,^{2,9} Chun-jie Liu,^{1,7,9} Yu Xiang,^{1,9} Chunlai Li,² Youqiong Ye,¹ Zhao Zhang,¹ David H. Hawke,³ Peter K. Park,² Lixia Diao,⁴ John A. Putkey,¹ Liuqing Yang,^{2,5,6} An-Yuan Guo,^{7,*} Chunru Lin,^{2,5,*} and Leng Han^{1,5,10,*}

¹Department of Biochemistry and Molecular Biology, McGovern Medical School at The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

²Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

³Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁴Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁵The Graduate School of Biomedical Sciences, The University of Texas MD Anderson Cancer Center UTHealth, Houston, TX 77030, USA

⁶Center for RNA Interference and Non-Coding RNAs, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁷Department of Bioinformatics and Systems Biology, Hubei Bioinformatics and Molecular Imaging Key Laboratory, Key Laboratory of Molecular Biophysics of the Ministry of Education, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, PRC

⁸Department of Epidemiology and Biostatistics, Key Laboratory of Environmental Health of Ministry of Education, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei 430030, PRC

⁹These authors contributed equally

¹⁰Lead Contact

*Correspondence: guoay@mail.hust.edu.cn (A.-Y.G.), clin2@mdanderson.org (C.L.), leng.han@uth.tmc.edu (L.H.)
<https://doi.org/10.1016/j.celrep.2017.10.070>

SUMMARY

Increasing evidence has demonstrated that small nucleolar RNAs (snoRNAs) play important roles in tumorigenesis. We systematically investigated the expression landscape and clinical relevance of snoRNAs in >10,000 samples across 31 cancer types from The Cancer Genome Atlas. We observed overall elevated expression of snoRNAs and their ribonucleoproteins in multiple cancer types. We showed complex regulation of snoRNA expression by their host genes, copy number variation, and DNA methylation. Unsupervised clustering revealed that the snoRNA expression subtype is highly concordant with other molecular/clinical subtypes. We further identified 46 clinically relevant snoRNAs and experimentally demonstrated functional roles of SNORD46 in promoting cell proliferation, migration, and invasion. We developed a user-friendly data portal, SNORic, to benefit the research community. Our study highlights the significant roles of snoRNAs in the development and implementation of biomarkers or therapeutic targets for cancer and provides a valuable resource for cancer research.

INTRODUCTION

Small nucleolar RNAs (snoRNAs) are non-coding RNAs with 60–300 nt in length (Jorjani et al., 2016). They are primarily classified into C/D box snoRNAs and H/ACA box snoRNAs (Balakin et al., 1996; Kiss, 2002) and are associated with a set of core

proteins named ribonucleoproteins (RNPs) to form stable and functional snoRNP particles (Reichow et al., 2007). C/D snoRNAs are associated with core proteins, including *SNU13*, *NOP56*, *NOP58*, and the methyltransferase *FBL* to guide 2'-O-methylation of target molecules (Spacková et al., 2010; Szewczak et al., 2005), whereas H/ACA snoRNAs are associated with core proteins, including *NHP2*, *NOP10*, *GAR1*, and the pseudouridine synthase *DKC1* to guide pseudouridylation of target molecules (Kiss et al., 2010). Additionally, small Cajal body-specific RNAs (scaRNAs) are a specific subset of snoRNAs that accumulate in Cajal bodies and guide the post-transcriptional modification of spliceosomal RNAs (Tycowski et al., 2009).

In mammals, the majority of snoRNAs are encoded within introns of protein coding or non-coding genes, which are called “host genes” (Bachelierie et al., 2002). Host genes may affect snoRNA expression through co-transcription (Filipowicz and Pogacić, 2002). Copy number variation (CNV) is a key regulator of gene expression and has been widely investigated for protein-coding genes and non-coding genes (Zack et al., 2013). DNA methylation is a common epigenetic mechanism that regulates gene expression (Yang et al., 2014). These findings suggest that snoRNA expression may be regulated by host genes, CNV, and DNA methylation.

Emerging evidence has revealed the significance of snoRNAs in oncogenesis (Krishnan et al., 2016; Williams and Farzaneh, 2012). For example, SNORD78 (C/D box) is overexpressed in non-small-cell lung cancer and prostate cancer (Martens-Uzunova et al., 2015; Zheng et al., 2015). SNORD50A/B (C/D box), which can directly bind and inhibit K-Ras, is deleted across multiple cancer types (Siprashvili et al., 2016). Furthermore, increased SNORA42 (H/ACA box) expression is an independent prognostic factor for overall survival times among cancer patients (Okugawa et al., 2017). SNORA55 (H/ACA box) silencing in prostate cancer cell

lines significantly inhibits cell proliferation and migration (Crea et al., 2016). These findings highlight the potential roles of snoRNAs in tumorigenesis, regardless of their classification.

Despite the emerging knowledge about the role of snoRNAs in cancer, the expression landscape and clinical relevance of snoRNAs in cancer have not been investigated systematically. Taking advantage of multiple “omic” data from The Cancer Genome Atlas (TCGA), we developed a computational pipeline and characterized the snoRNA expression profiles for a large number of samples in 31 cancer types.

RESULTS

Expression Landscape of snoRNAs across Different Cancer Types

We developed a computational pipeline to comprehensively quantify snoRNA expression levels in human cancer (Figure S1A). We integrated the latest annotations from The University of California, Santa Cruz (UCSC) Genome Browser (hg19) and GENCODE (version 24) and obtained 1,524 unique snoRNA genes. Based on the snoRNA classification information from UCSC Genome Browser, Rfam, and Ensembl, we classified the snoRNAs into 1,033 C/D box snoRNAs, 438 H/ACA box snoRNAs, and 53 scaRNAs. Through mapping microRNA (miRNA) sequencing (miRNA-seq) reads to snoRNAs in 9,931 samples across 31 TCGA cancer types (Table S1) and 663 related adjacent normal samples (Figure 1A), we identified 465 detectable snoRNAs with an average number of reads per kilobase per million mapped reads (RPKM) ≥ 1 across all tumor samples (Figure 1B). Among these snoRNAs, we observed that some snoRNAs were highly expressed across all cancer types. The top 5% highly expressed snoRNAs were correlated with protein-coding genes that are enriched in several biological processes, including the metabolic process and cellular process (Figure S1B), indicating that snoRNAs may be involved in various biological processes. Furthermore, we performed cross-tumor clustering to examine whether snoRNAs have a tumor-type-specific expression profile. We observed that some snoRNAs are highly expressed in specific cancer types, e.g., kidney renal clear cell carcinoma (KIRC) and thyroid carcinoma (THCA) (Figure S1C). Among the detectable snoRNAs, there were 319 C/D box snoRNAs, 123 H/ACA box snoRNAs, and 23 scaRNAs. We observed significantly higher expression levels of C/D box snoRNAs (median $\log_2(\text{RPKM}) = 6.2$; interquartile range [IQR] = 3.1–8.3) than H/ACA box snoRNAs (median $\log_2(\text{RPKM}) = 3.4$; IQR = 2.1–4.7) and scaRNAs (median $\log_2(\text{RPKM}) = 3.3$; IQR = 2.0–5.2; Kruskal-Wallis test; $p = 1.67 \times 10^{-11}$; Figure 1C). This may be due to the shorter length distribution of C/D box snoRNAs (t test; $p = 6.59 \times 10^{-42}$; Figure S1D), and miRNA-seq might be biased toward shorter snoRNAs. We further examined the correlations between TCGA tumor purity (calculated by ESTIMATE, which uses gene expression signatures to infer the fraction of stromal and immune cells in tumor samples; Yoshihara et al., 2013) and total snoRNA expression for each cancer type. Overall, we observed weak or absent correlations (Spearman correlation $|R_s| < 0.3$), suggesting that cell origin (e.g., immune cells and stromal cells) has limited effects on snoRNA expression (Figure S2). Also, we keep in mind that the tumor purity score is used to partially estimate tumor heterogeneity (Carter et al., 2012; Yosh-

ihara et al., 2013). Further investigation of the effects of tumor heterogeneity on snoRNA expression will be necessary, e.g., by single-cell sequencing (Patel et al., 2014; Wang et al., 2014).

We examined the global expression of the total snoRNAs in 14 cancer types with paired tumor and normal samples (Figure 1D). We observed overexpression of the total snoRNAs in 13 out of 14 cancer types surveyed, including breast invasive carcinoma (BRCA) (fold change [FC] = 1.77; paired t test; $p = 1.8 \times 10^{-14}$), KIRC (FC = 1.87; $p = 1.4 \times 10^{-23}$), and prostate adenocarcinoma (PRAD) (FC = 3.39; $p = 6.9 \times 10^{-14}$). These results suggest global overexpression of total snoRNAs across multiple cancer types. We noticed a few downregulations in paired normal/tumor samples, which we believe to be largely due to individual variation in patient data, as such situations are common in cancer research (Fu et al., 2014; Liu et al., 2016).

Functional Roles of snoRNA RNP in Human Cancer

Given the overall upregulation of total snoRNA expression in multiple cancer types and snoRNA association with RNPs to exert their functions, we systematically assessed the alterations of RNPs across different cancer types. We observed that the expression levels of all RNPs, except *SNU13*, were significantly elevated in at least one cancer type (FC ≥ 1.5 and false discovery rate [FDR] < 0.05 ; Figure 2A). We showed that *FBL* is upregulated in four cancer types. Furthermore, we observed that *NOP58*, *NOP56*, and *DKC1* were overexpressed in more than five cancer types, suggesting their consistent oncogenic roles in cancer. In addition, we observed that RNPs were associated with patient prognosis and that the majority of overexpressed RNPs were associated with poor survival (Figure 2B). For example, high expression of *NOP58* was associated with poor survival in adrenocortical carcinoma (ACC) (univariate Cox test; $p = 6.8 \times 10^{-6}$; Figure 2B), high expression of *NOP56* was associated with poor survival in KIRC ($p = 2.6 \times 10^{-15}$), and high expression of *DKC1* was associated with poor survival in lower grade glioma (LGG) ($p = 2.6 \times 10^{-6}$). Interestingly, most of the significant associations were found in KIRC, kidney renal papillary cell carcinoma (KIRP), and LGG. This may be due to the variations in survival data from TCGA. For example, the follow-up time is short for BRCA samples (Cancer Genome Atlas Network, 2012).

We observed elevated expression of both total snoRNAs and RNPs across different cancer types. Therefore, we further investigated the associations between total C/D and H/ACA box snoRNAs and their corresponding RNPs. We observed an overall positive correlation for both C/D box and H/ACA box snoRNAs (Figure 2C). For example, *FBL* is highly correlated with C/D box snoRNAs in 7 cancer types, whereas *DKC1* is highly correlated with H/ACA box snoRNAs in 14 cancer types. Our results showed that RNPs were strongly correlated with snoRNAs, suggesting their co-activation and synergy in cancer. Further analysis of the protein levels of RNPs will be necessary due to the lack of protein expression levels for RNPs in TCGA.

Regulation of snoRNA Expression by Host Genes, CNV, and DNA Methylation

To understand the potential regulation of snoRNA expression, we examined the distribution of host genes for snoRNAs. There

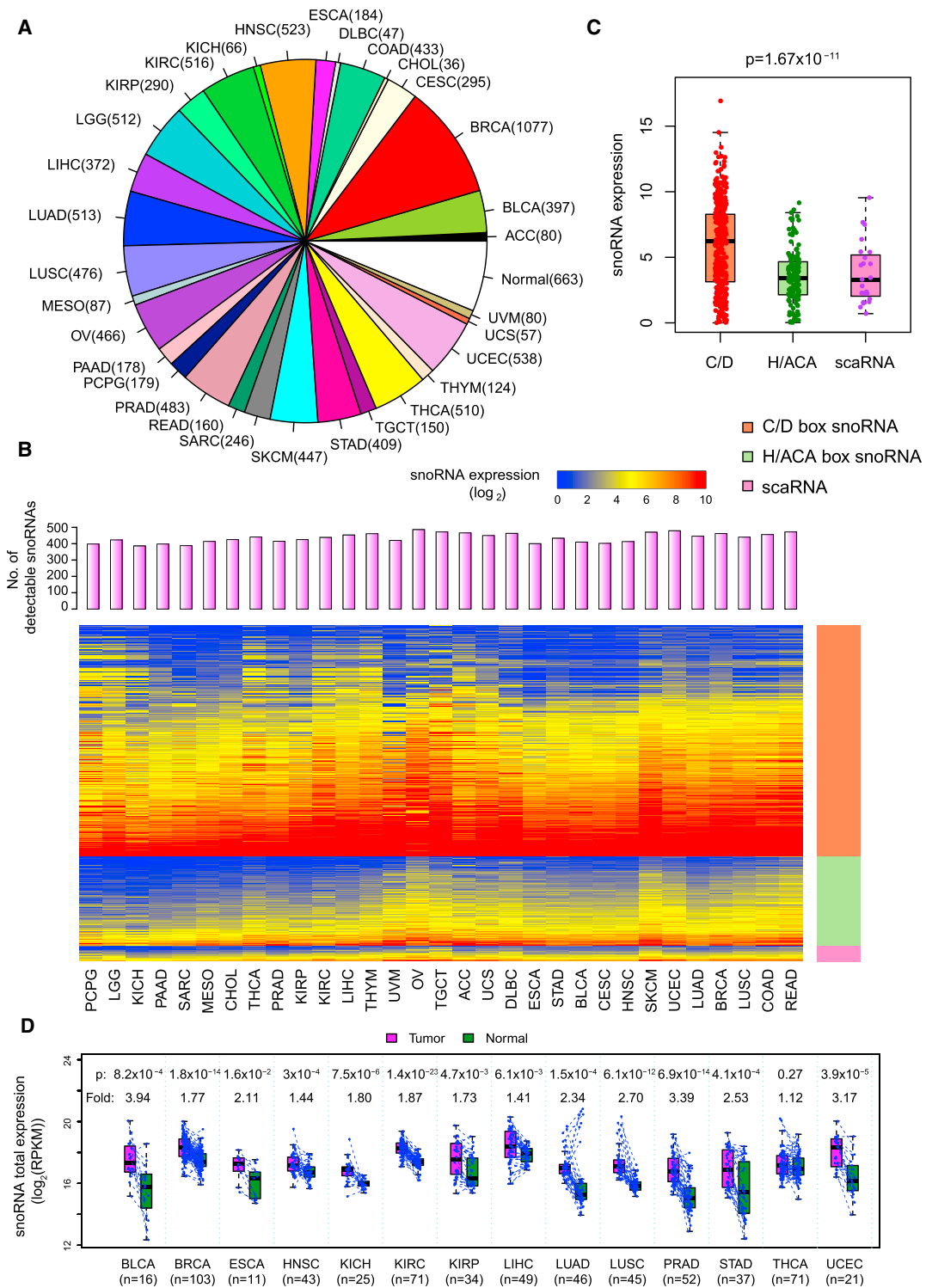


Figure 1. Overview of snoRNA Expression in Human Cancer

(A) Number of tumor and normal samples analyzed in this study.

(B) Number of detectable snoRNAs in each cancer type and the global expression profiles of snoRNAs across different cancer types. Each column represents one cancer type, and each row represents a snoRNA grouped by snoRNA type. Color represents the snoRNA expression level.

(C) Expression of different types of detectable snoRNAs.

(D) Expression of total snoRNAs between paired tumor and normal samples. The boxes show the median \pm 1 quartile, with each point representing one sample. See also [Figures S1](#) and [S2](#) and [Table S1](#).

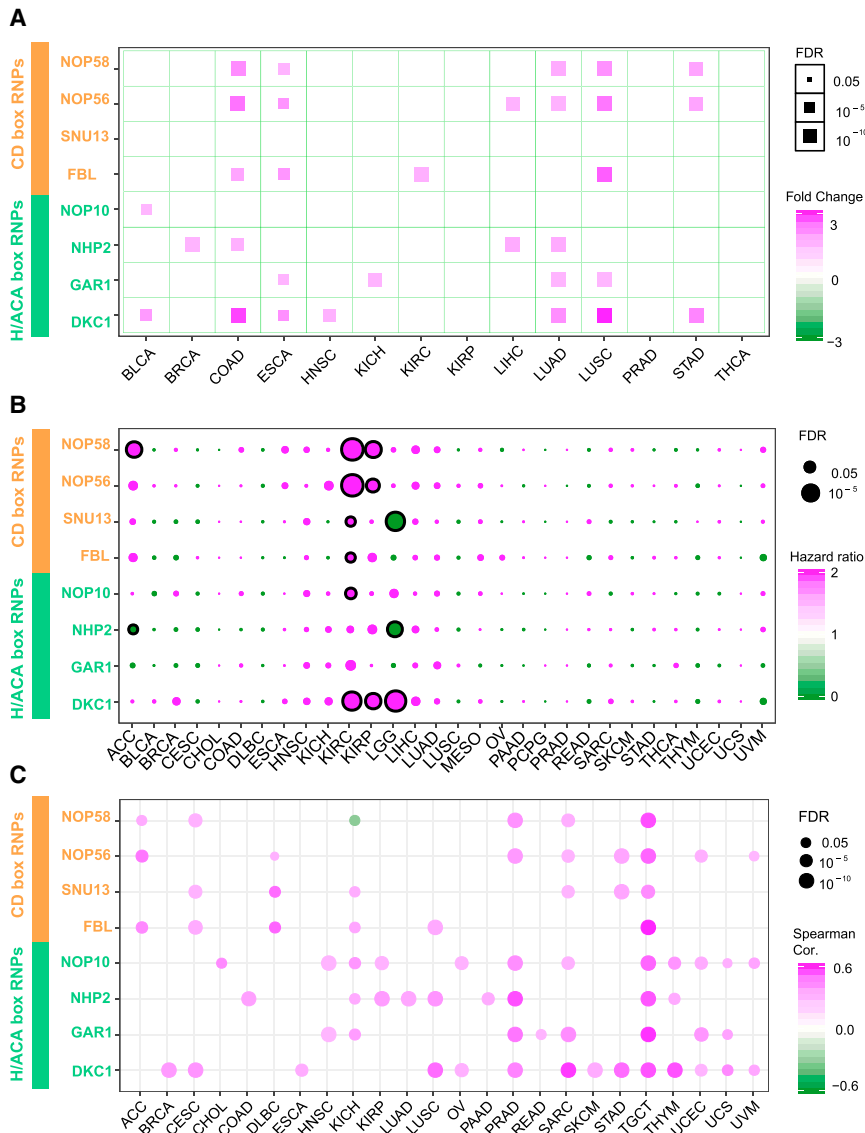


Figure 2. Expression Landscape of snoRNA RNPs

(A) Expression difference for ribonucleoproteins (RNPs) between paired tumor and normal samples. Only cancer types with more than 10 paired tumor-normal samples are shown.

(B) Prognostic values of RNPs across multiple cancers.

(C) Correlations between total C/D box, H/ACA snoRNAs, and their RNPs in tumor samples. Only cancer types with at least one significant correlation are shown.

of GAS5 (Figure S3A) that were positively correlated with GAS5 in at least 12 cancer types (Figure S3B).

CNV is an important factor in the regulation of gene expression (Gamazon and Stranger, 2015); thus, we systematically evaluated the effects of CNV on snoRNA expression. We observed a varied effect that ranged from 2/479 (0.4%) of snoRNAs in uterine corpus endometrial carcinoma (UCEC) to 147/386 (38.1%) of snoRNAs in kidney chromophobe (KICH) that were correlated with their CNV ($|Rs| \geq 0.3$ and $FDR < 0.05$; Figure 3D). For example, expression levels of SNORD56 and SNORD119 showed significantly positive correlation with their CNVs in 19 and 17 cancer types, respectively.

We evaluated the correlation of snoRNAs with DNA methylation, which is an epigenetic modification involved in regulating gene expression (Zemach et al., 2010). We observed significant correlations with DNA methylation that ranged from 1/425 (0.2%) of snoRNAs in cholangiocarcinoma (CHOL) to 191/472 (40.4%)

were 703, 151, and 670 snoRNAs in protein-coding genes, non-coding genes, and without host genes, respectively. Detectable snoRNAs were enriched in host genes (318/854 versus 147/670; chi-square test; $p = 2.93 \times 10^{-7}$; Figure 3A), and they were more enriched in non-coding genes than in protein-coding genes (78/151 versus 240/703; chi-square test; $p = 0.011$). Furthermore, snoRNAs in non-coding genes had higher expression levels than snoRNAs in protein-coding genes and snoRNAs without host genes (Kruskal-Wallis test; $p = 4.83 \times 10^{-9}$; Figure 3B). snoRNAs in non-coding genes had a peak expression of 7.3, whereas the peak expression of snoRNAs in protein-coding genes and snoRNAs without host genes were 4.2 and 2.8, respectively (Kolmogorov Smirnov test; $p = 0.001$; Figure 3C). The proportion of snoRNAs highly correlated with the host genes ranged from 6/441 (1.4%) in THCA to 113/472 (23.9%) in testicular germ cell tumors (TGCT) ($|Rs| \geq 0.3$ and $FDR < 0.05$; Figure 3D). For example, there were 10 snoRNAs located in introns

of snoRNAs in TGCT ($|Rs| \geq 0.3$ and $FDR < 0.05$; Figure 3D). For example, the methylation level of probe cg06726167, located 43 bp upstream of SNORD118, was negatively correlated with the expression of SNORD118 in 24 cancer types.

We further examined whether snoRNAs were regulated by two or more of the factors mentioned above. Of all the detectable snoRNAs in each cancer type, from 18/441 (4.1%) of snoRNAs in THCA to 268/450 (59.5%) of snoRNAs in uterine carcinosarcoma (UCS) were correlated with at least one factor. Only a small proportion of snoRNAs were co-regulated by multiple factors, ranging from 0 in THCA to 107/450 (23.8%) in UCS (Figure 3D). Interestingly, there were only a few hits/correlations in THCA, which may be due to the physiological features of THCA. We also observed larger numbers of correlations in several cancer types, including ACC, UCS, TGCT, and KICH, which may be driven by the smaller sample sizes of those datasets (Figure S3C).

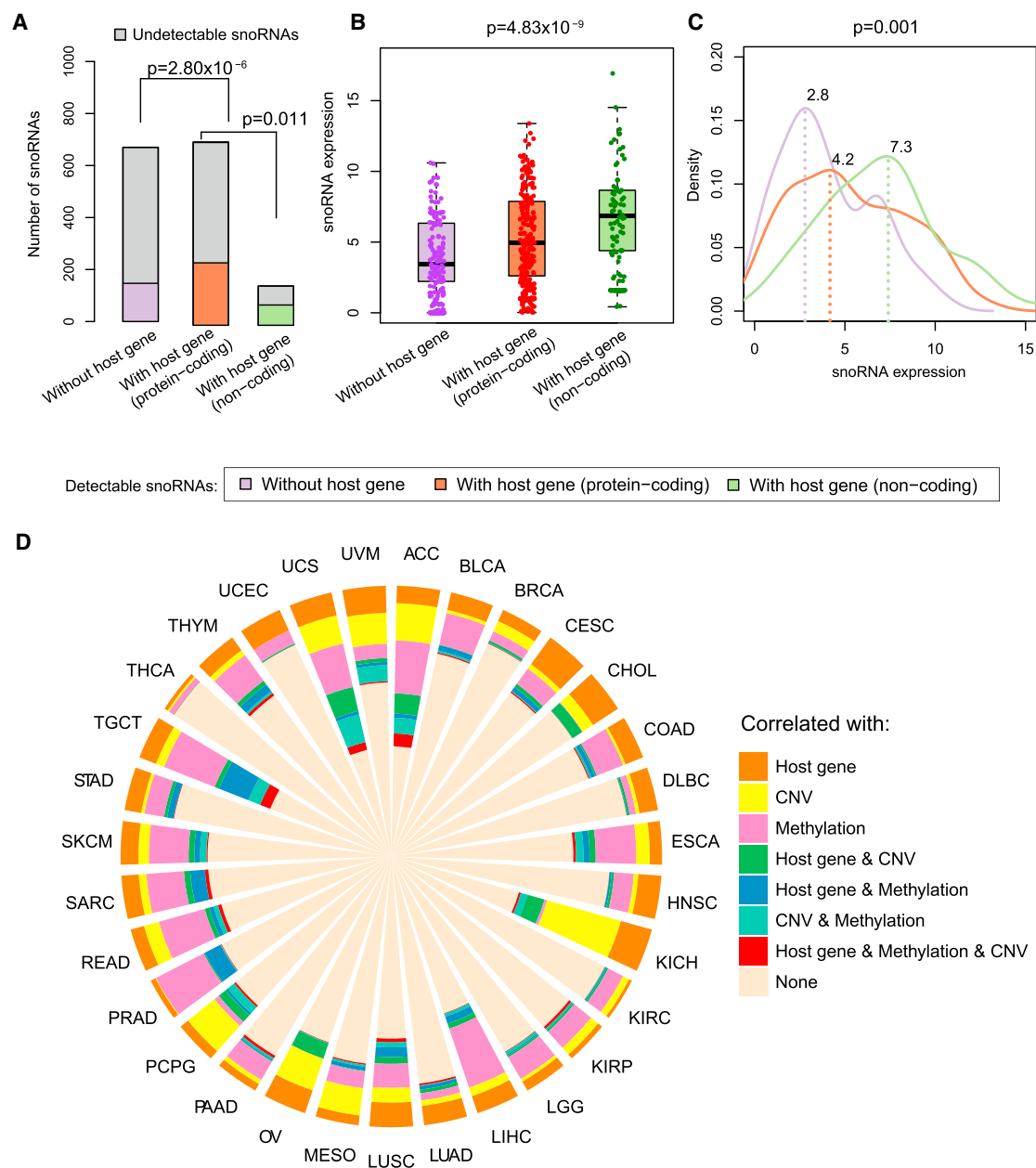


Figure 3. Regulation of snoRNA Expression

(A) Number of snoRNAs with host genes (protein-coding genes and non-coding genes) and without host genes. The three bars represent total numbers of snoRNAs with host genes (protein-coding genes and non-coding genes) and without host genes. The color bars represent detectable snoRNAs in each category, whereas gray bars represent undetectable snoRNAs in each category.

(B) Expression of snoRNAs (boxplot).

(C) Expression of snoRNAs (density plot).

(D) Number of significant correlations between snoRNA expression and their host genes, CNV, and DNA methylation.

See also [Figure S3](#).

Biomedical Significance of Subtypes Defined by snoRNA Expression

Various omic data from multiple platforms have been used to classify tumor subtypes (Han et al., 2014; Hoadley et al., 2014). To investigate the biomedical significance of subtypes defined by snoRNA expression, we selected 200 (on average,

79% C/D box snoRNAs, 20% H/ACA box snoRNAs, and 1% scaRNAs; [Figure S4A](#)) of the most variable snoRNAs to classify tumor samples for each cancer type into subtypes using ConsensusClusterPlus (Wilkerson and Hayes, 2010). Overall, we observed high concordance between snoRNA subtypes and other molecular subtypes, finding that 79 out of the 105

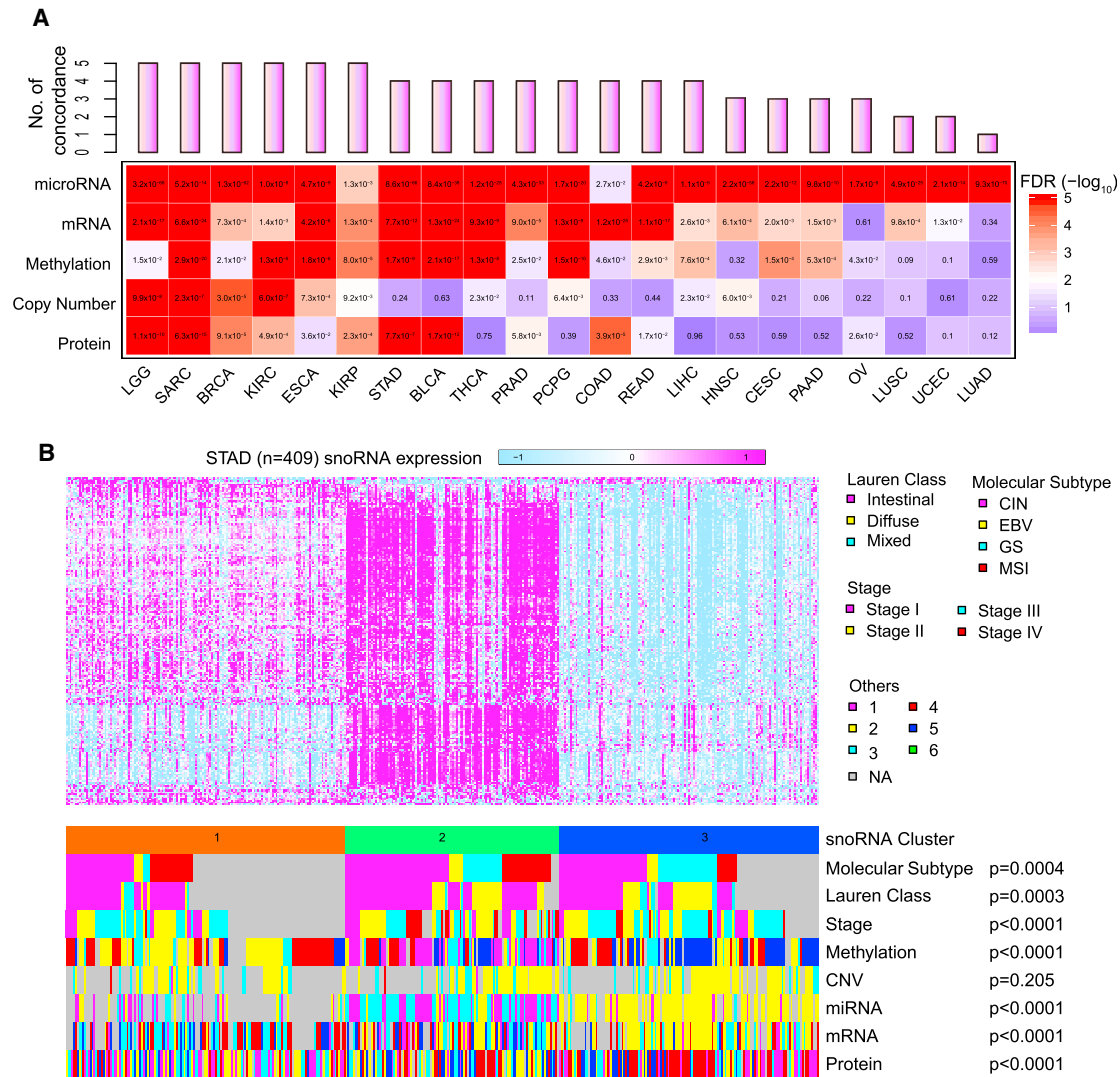


Figure 4. Concordance between snoRNA Expression Subtype and Other Subtypes

(A) Concordance between snoRNA expression subtypes (based on 200 snoRNA signature) and molecular subtypes defined by other molecular data in TCGA cancer types with more than 100 samples. Color indicates statistical significance of the chi-square test.

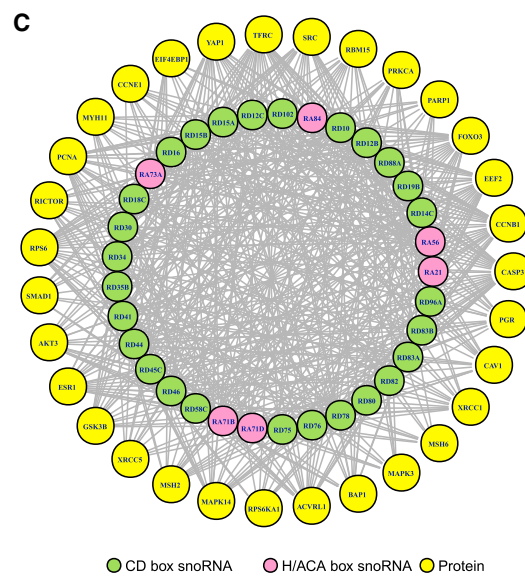
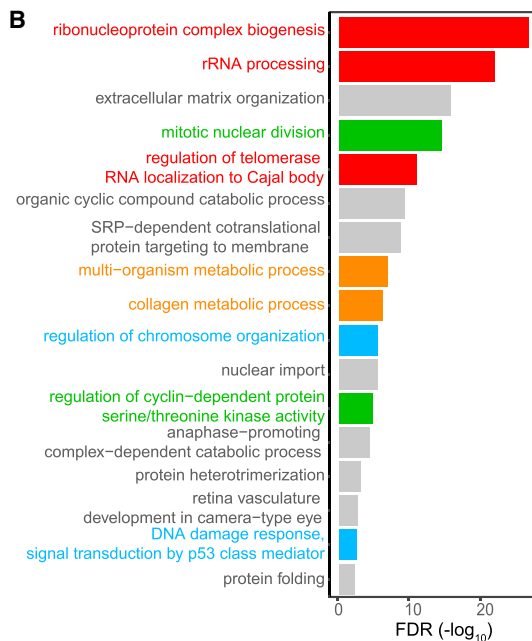
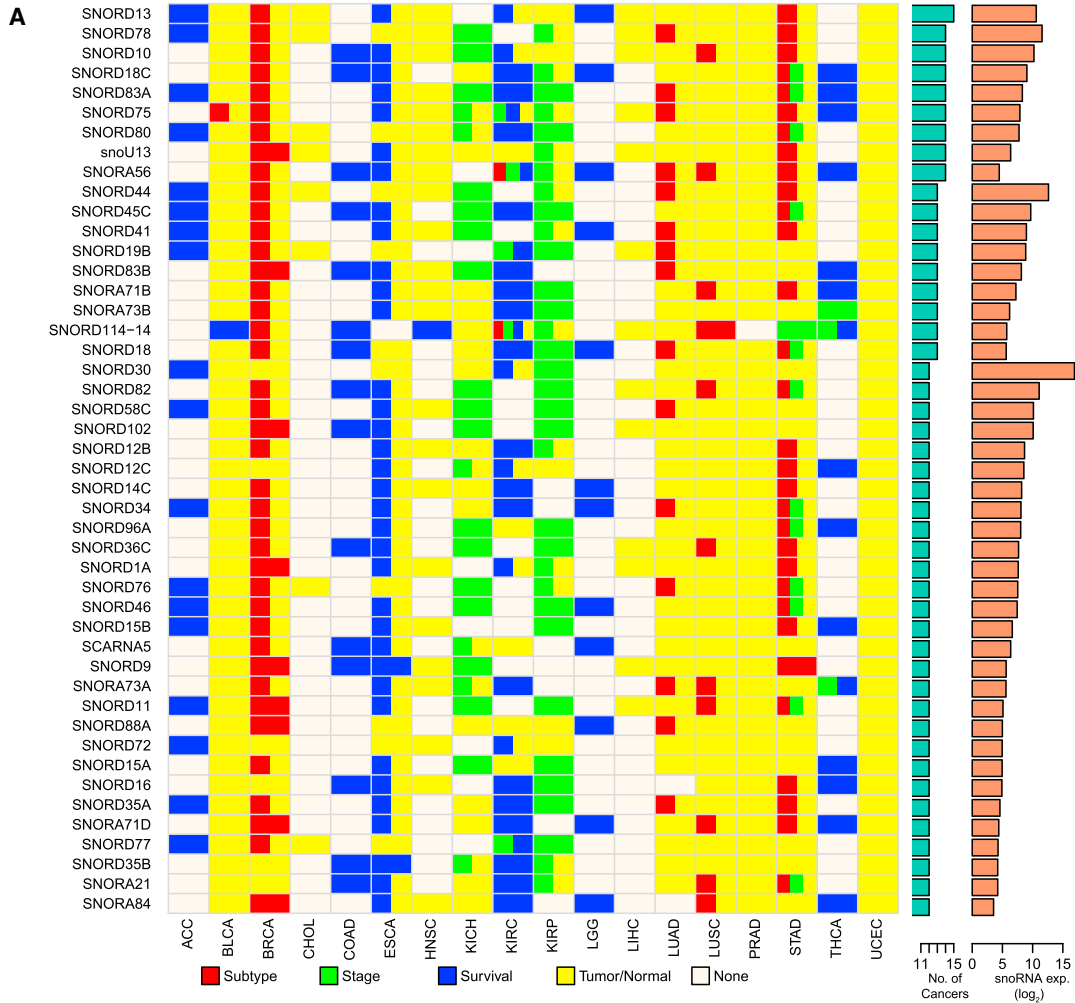
(B) SnoRNA expression subtype in stomach adenocarcinoma (STAD), including subtype 1 (orange; n = 152), subtype 2 (green; n = 116), and subtype 3 (blue; n = 141).

See also Figure S4.

chi-square tests were significant (Figure 4A), especially in LGG, sarcoma (SARC), BRCA, KIRC, esophageal carcinoma (ESCA), and KIRP, with significant concordance across all platforms. Interestingly, almost all snoRNA subtypes were concordant with miRNA subtypes. This was likely due to both miRNA and snoRNA expression levels being quantified from miRNA-seq data. The different proportions of tumor cells in the specimens from TCGA may have some impact on genomic profiling (Han et al., 2014). The snoRNA subtypes were also highly concordant with the tumor stage in five cancer types, including ACC, BRCA, KIRC, stomach adenocarcinoma (STAD), and THCA (Figure S4B). Furthermore, snoRNA subtypes demonstrated prognostic value in stratifying clinical outcomes in four cancer types

(Figure S4C), including pancreatic ductal adenocarcinoma (PAAD) (log rank test; $p = 0.019$), lung squamous cell carcinoma (LUSC) ($p = 0.039$), LGG ($p = 0.002$), and KIRC ($p = 2 \times 10^{-4}$).

Taking STAD as an example, we classified 409 STAD samples into three distinct subtypes based on snoRNA expression: subtype 1 (orange; n = 152); subtype 2 (green; n = 116); and subtype 3 (blue; n = 141; Figure 4B). The snoRNA expression subtype showed high concordance with the integrated molecular subtype defined by TCGA (Cancer Genome Atlas Research Network, 2014), which indicated that subtype 1 is enriched for samples with chromosomal instability (CIN) (magenta), whereas subtype 3 is enriched for samples genomically stable (GS) (cyan). snoRNA subtypes also showed high concordance with tumor



(legend on next page)

stage (chi-square test; $p = 1.84 \times 10^{-5}$) and other molecular subtypes, including miRNA expression ($p = 2.45 \times 10^{-67}$), mRNA expression ($p = 1.69 \times 10^{-12}$), and protein expression ($p = 2.71 \times 10^{-7}$). The concordance of certain subtypes may have clinical relevance (Hoadley et al., 2014), and our results suggest that snoRNA subtypes are an additional dimension for classifying cancer-related molecular subtypes.

Identification of Clinically Relevant snoRNAs

To comprehensively identify clinically relevant snoRNAs, we performed a series of analyses to identify snoRNAs that show differential expression in paired tumor and normal samples, among tumor subtypes, among tumor stages, or associated with patient survival time. We identified 433 snoRNAs that were differentially expressed between paired tumor-normal samples in at least one cancer type. We further identified 279 snoRNAs that were differentially expressed among established tumor subtypes, 203 snoRNAs differentially expressed among tumor stages, and 355 snoRNAs that were associated with patient survival times. Among the snoRNAs associated with patient survival times, there were 229 C/D box snoRNAs, 108 H/ACA box snoRNAs, and 18 scaRNAs. The number of clinically relevant snoRNAs ranged from 39 in ACC to 329 in UCEC. To prioritize snoRNAs through powerful pan-cancer analyses (Han et al., 2015; Wang et al., 2017), we obtained 46 snoRNAs that show clinical relevance in at least 12 cancer types (Figure 5A). For example, SNORD78, which has been reported to be overexpressed in non-small-cell lung cancer and prostate cancer (Martens-Uzunova et al., 2015; Zheng et al., 2015), showed clinical relevance in 14 cancer types, and SNORD46, which was reported as a prognostic marker (Krishnan et al., 2016), showed upregulation in 8 cancer types (Figure S5A).

To further explore the potential function of clinically relevant snoRNAs, we obtained mRNAs with expression levels that correlated with clinically relevant snoRNAs in at least 10 cancer types. These mRNAs were enriched in gene ontology (GO) terms, such as “ribonucleoprotein complex biogenesis” and “rRNA processing,” which confirmed the functional roles of snoRNAs on rRNA modification. Interestingly, we also observed enrichment in several GO terms, including “mitotic nuclear division” and “DNA damage response, signal transduction by p53 class mediator” (Figure 5B), suggesting that clinically relevant snoRNAs may influence these biological processes to affect cancer initiation and progression. We further identified several genes that were consistently correlated with clinically relevant snoRNAs in more than 10 cancer types (Figure S5B). For example, *CKS2* (CDC28 protein kinase regulatory subunit 2), which binds to

the catalytic subunit of cyclin-dependent kinases (Yu et al., 2015), was positively correlated with 15 clinically relevant snoRNAs, such as SNORA56, SNORD14C, etc.

Furthermore, we assessed the correlation between clinically relevant snoRNAs and key cancer proteins. We observed a total of 10,601 significant correlations between 46 clinically relevant snoRNAs and 226 proteins and phosphoproteins in 31 cancer types ($|Rs| \geq 0.3$ and $FDR < 0.05$). For example, *FOXO3* (forkhead box O3), a transcription factor that triggers apoptosis (Haoues et al., 2014), was negatively correlated with several snoRNAs, including SCARNA5, SNORA71B, and SNORD46. *CCNB1* (cyclin B1), which leads to cell proliferation by binding to *CDK1* (Yuan et al., 2006), was positively correlated with several snoRNAs, including SNORA71B, SNORA71D, and SNORD46 (Figure 5C). We randomly selected 46 non-clinically relevant snoRNAs with matched snoRNA type and expression levels to calculate the correlation for 1,000 simulations. We observed a significantly higher number of correlations for clinically relevant snoRNAs (permutation test; $p < 1 \times 10^{-3}$; Figure S5C). This result indicates a strong correlation between clinically relevant snoRNAs and key cancer proteins, further suggesting the significant roles of these snoRNAs in tumorigenesis.

Functional Effects of Clinically Relevant snoRNAs

To determine the functional roles of clinically relevant snoRNAs in human cancer, we prioritized snoRNAs based on the following criteria: (1) sequence length longer than 100 bp; (2) no repetitive elements in the snoRNA sequence; and (3) relatively high expression level. We selected the top 10 snoRNAs (Table S2) and identified their interacting proteins through RNA pull-down, followed by mass spectrometry (MS) analysis (Xing et al., 2014). Proteins that associated with the beads (negative control) were considered to be unspecific binding, with Mascot protein scores of 200 or less. We used the RNA oligonucleotide poly(A)₂₅ as an additional negative control because snoRNAs contain no poly(A) tails (Kim et al., 2006). Our data indicate that poly(A)₂₅ associated with a panel of RNA-binding proteins, which were considered to be general RNA-protein bindings, but not specific interactions with snoRNA sequences (Figure 6A). The 3' UTR of the androgen receptor (AR), which can bind with the ELAV-like RNA binding protein 1 (ELAV1; also known as HuR; Blaxall et al., 2000), was included as a positive control. We confirmed the association between the 3' UTR of AR and ELAV1 in our assay (Figure 6A). Our MS data shed light on the potential cellular effects of snoRNAs in regulating non-receptor tyrosine kinase (SNORD15A), small nuclear RNP complex assembly (SCARNA5, SNORD15A, and SNORD9), metabolic pathway regulation (SNORD16 and

Figure 5. Clinically Relevant snoRNAs across Multiple Cancer Types

(A) 46 clinically relevant snoRNAs identified in more than 12 cancer types. For each cancer type, red box indicates significant differential expression among tumor subtypes (fold change [FC] ≥ 2 and $FDR < 0.05$), green box indicates significant differential expression among tumor stages (FC ≥ 2 and $FDR < 0.05$), blue box indicates association with 5-year survival (FC ≥ 2 and $FDR < 0.05$), yellow box indicates significant differential expression between tumor and normal samples (FC ≥ 2 and $FDR < 0.05$), and white box represents not significant. Numbers of cancer types (aquamarine bar) and expression (orange bar) are plotted in right panel.

(B) Functional enrichment of mRNAs highly correlated with 46 clinically relevant snoRNAs in at least 10 cancer types.

(C) Network of correlations between 46 clinically relevant snoRNAs and key cancer proteins (significant correlation in ≥ 3 cancer types and nodes with degree ≥ 10).

See also Figure S5 and Table S2.

SNORD71D), and cell apoptosis (SNORD71B and SNORD71D). We further filtered out small nuclear RNP polypeptides and heterogeneous nuclear RNPs, including RUXGL, RUXG, RUXE, RUXF, and HNRPD, which are unlikely to play specific functions in cancer. We identified three snoRNAs that were associated with two interesting proteins. SNORD15A was shown to be associated with LYN (Figure 6A), which is a member of the Src family of protein tyrosine kinases that modulate various cellular outputs, including proliferation, invasion, apoptosis, and metabolic reprogramming (Choi et al., 2010; Ingle, 2012). SNORD46 and scaRNA5 were associated with PAPP1 (pappalysin-1; also known as pregnancy-associated plasma protein A), which has been recently documented as an oncogene that promotes the initiation of breast cancer (Takabatake et al., 2016). SNORD46 was reported to be a prognostic marker based on survival analysis in breast cancer, without any functional assays for further studying the functional effects (Krishnan et al., 2016). Therefore, SNORD15A and SNORD46, with higher scores for PAPP1, were chosen as candidate snoRNAs for further experiments.

To demonstrate the functional roles of snoRNAs, we knocked down snoRNAs in cancer cell lines using antisense oligonucleotide-based strategies. Locked nucleic acids (LNAs) have been used to deplete non-coding RNAs in cells and mouse models (Li et al., 2017; Lin et al., 2016), so we designed and tested the knockout efficiency of LNAs against SNORD15A and SNORD46. However, all the LNAs we tested against SNORD15 failed to exhibit detectable knockdown of SNORD15. We considered a few possible reasons for this outcome: first, the potential secondary structure might affect the binding between LNAs and snoRNAs. Although SNORD15A and SNORD46 are structurally similar (Figure S5D; predicted by RNAfold; Gruber et al., 2008), SNORD15A contains a strong base-pairing stem, which might prevent hybridization between LNAs and the following RNase H digestion (Kurreck et al., 2002). The second possibility is that the snoRNA-binding proteins may prevent the recognition of LNAs. As shown in Figure 6, SNORD15A was associated with LYN, RUXG, and RUXGL proteins; SNORD46 was associated with PAPP1, which suggests that SNORD15A might be abundantly masked by proteins, which may lead to low knockdown efficiency. Our LNAs targeting SNORD46 showed efficient knockdown of SNORD46 in the breast cancer cell line MDA-MB-231 and the lung cancer cell line A549 (Figures 6B and 6D). Interestingly, knockdown of SNORD46 led to a significant reduction in cell viability in both cell lines tested, which was not observed in cells transfected with a scrambled control (Figures 6C and 6E). In addition, depletion of SNORD46 led to impaired cell migration and invasion in both MDA-MB-231 and

A549 cell lines (Figures 6F–6I). Taken together, our data indicate that SNORD46 is associated with oncogenesis and plays important roles in maintaining the proliferation and mobility of cancer cells.

A User-Friendly Data Portal to Explore snoRNAs in Cancer

To facilitate broad access to these snoRNA data and associated data, we developed a user-friendly data portal, snoRNA in cancer (SNORic) (<http://bioinfo.life.hust.edu.cn/SNORic>). SNORic provides 4 modules: “summary”; “snoRNA-based analysis”; “gene-based snoRNAs”; and “download” (Figure 7A). The summary module shows an overview of datasets used in SNORic with a detailed description for each cancer type (e.g., number of samples, reads, and detectable snoRNAs). The snoRNA-based analysis module provides analyses for each snoRNA across different cancer types. In this module, users can obtain expression data for any annotated snoRNA and examine whether the snoRNA is clinically relevant (Figure 7B). In addition, users can examine the correlation between snoRNAs and CNV, DNA methylation, mRNAs, RNA splicing events, and protein expression. The gene-based snoRNAs module allows users to search a gene to obtain its correlated snoRNAs ($|Rs| \geq 0.3$; $FDR < 0.05$; Figure 7C). The download module allows users to download expression profiles for all snoRNAs. SNORic includes the expression levels of 1,524 snoRNAs across >10,000 samples for browsing, analyzing, visualizing, and downloading, so that the users can further test their hypotheses. We also included genes that are significantly correlated with snoRNAs at the mRNA expression level (~18,000 mRNAs), protein expression level (226 proteins), and splicing events (~41,290 splicing events) across 31 cancer types. This valuable resource will be of significant interest to the research community (Li et al., 2015).

DISCUSSION

miRNA-seq is not designed for a full snoRNA repertoire, as miRNA-seq reads are too short (15–30 bp) to distinguish snoRNA from snoRNA fragments (Krishnan et al., 2016), which may have different biological functions. Thus, it will be necessary to validate full-length snoRNAs from miRNA-seq data for further investigation. However, this method has been applied in several studies (Gao et al., 2015; Krishnan et al., 2016; Zheng et al., 2016) and is probably the most appropriate way to quantify snoRNA expression profiles from TCGA omic data. Among 1,524 annotated snoRNAs, we detected 1,520 (99.7%) snoRNAs with at least 1 read, which is far more than other studies/platforms. For

Figure 6. SNORD46 Functions as an Oncogene in MDA-MB-231 and A549 Cells

(A) Identification of protein candidates interacting with snoRNAs revealed by mass spectrometry analysis. The y axis shows the Mascot protein score for each spectrum query. Score ≥ 200 is considered a significant peptide assignment. The biotin-labeled poly(A)₂₅ and beads-only group without any biotin-labeled RNA was used as a negative control. The 3' UTR of the androgen receptor (AR) was used as positive control.

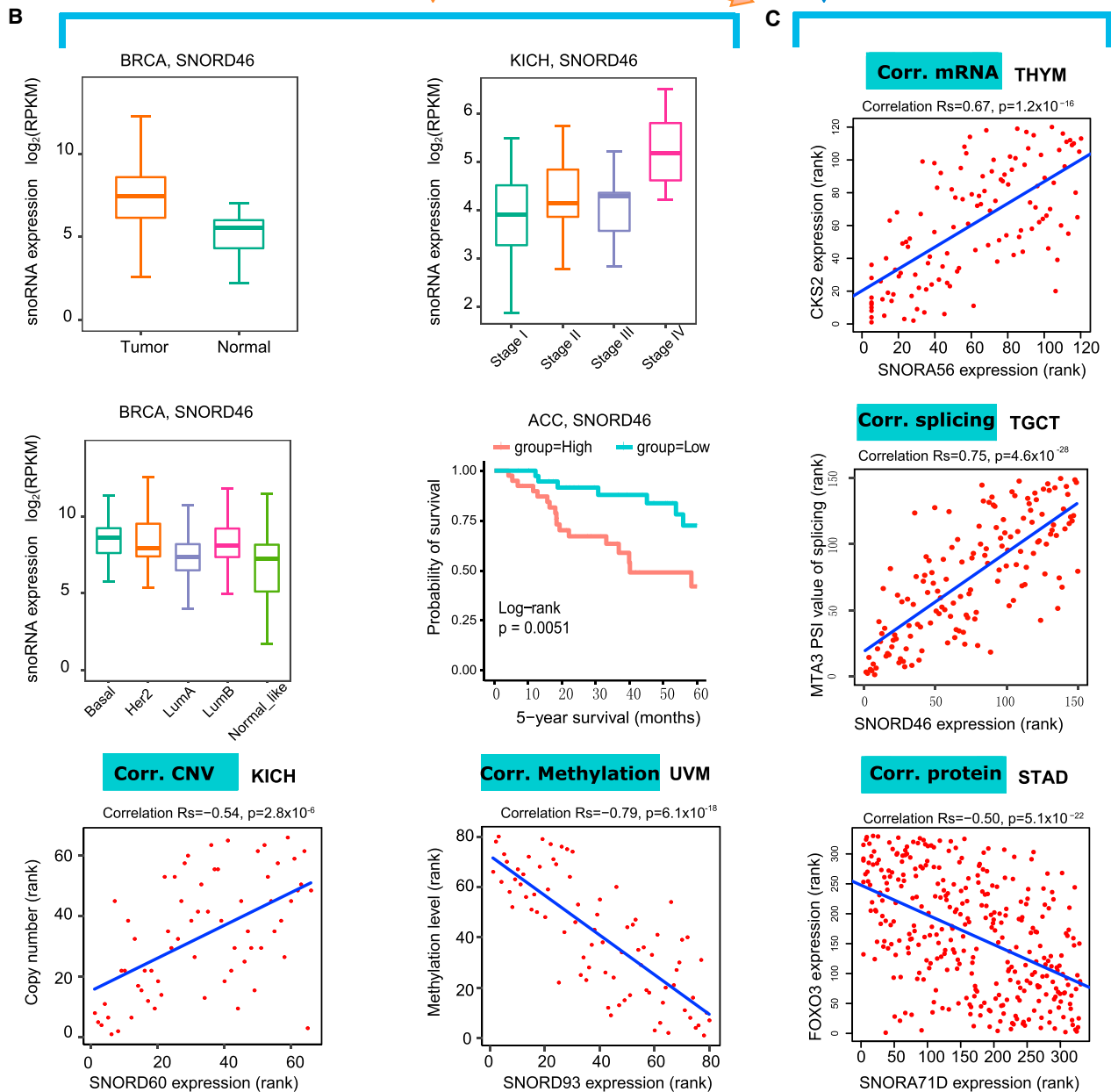
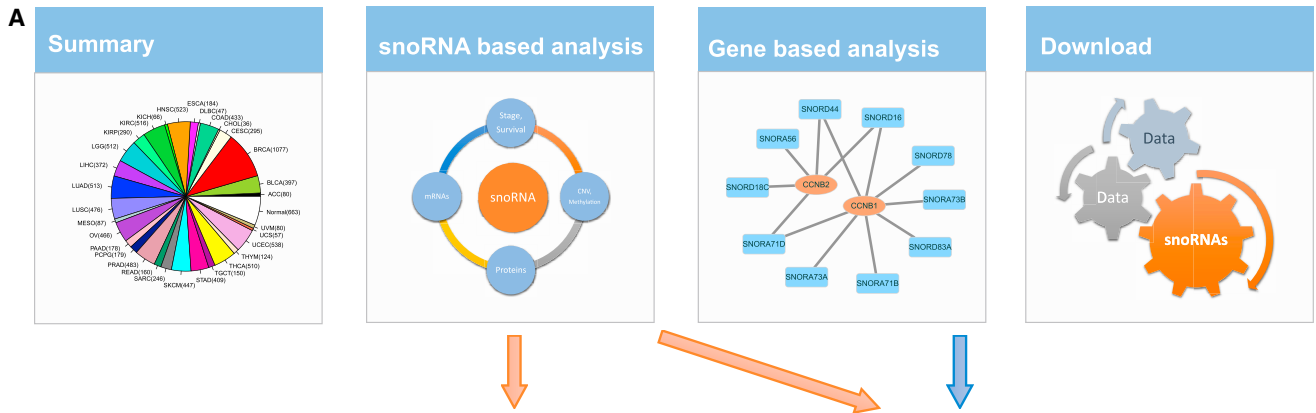
(B and D) Real-time PCR detection of snoRD46 expression with 3 biological replicates in MDA-MB-231 cells (B) and A549 (D) transfected with scrambled RNA control (Scr LNA) or snoRD46 LNAs.

(C and E) Cell proliferation assays with 5 biological replicates in MDA-MB-231 cells (C) and A549 (E) transfected with Scr LNA or snoRD46 LNAs.

(F and G) Cell migration assays with 3 biological replicates in MDA-MB-231 cells (F) and A549 (G) transfected with Scr LNA or snoRD46 LNAs.

(H and I) Cell invasion assays with 3 biological replicates in MDA-MB-231 cells (H) and A549 (I) transfected with Scr LNA or snoRD46 LNAs.

* $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$. The scale bar represents 200 μm . Data are represented as mean \pm SD. See also Figure S5.



(legend on next page)

example, snoRNAome detected 235 (21.0%) out of 1,118 annotated snoRNAs from ENCODE small RNA-seq data (Jorjani et al., 2016), whereas a recent study detected 768 (75.4%) snoRNAs out of 1,018 Ensembl-annotated snoRNAs with at least 1 read (Krishnan et al., 2016). In contrast, we only detected 19 snoRNAs from TCGA RNA-seq data with the expression level (RNA-seq by expectation maximization) ≥ 1 . The miRNA-seq-based analysis is by no means perfect, but it allows us to systematically investigate the expression landscape and clinical relevance of snoRNAs in more than 10,000 samples across 31 cancer types. Our analyses lay the groundwork for an integrated functional interpretation to illuminate the significant roles of snoRNAs.

To confirm the accuracy of our computational pipeline, we compared our results with those from an independent study, which identified 16 snoRNAs differentially expressed in lung cancer by both miRNA-seq and qRT-PCR (Gao et al., 2015). Among these 16 snoRNAs, our analysis also identified 14 of them as differentially expressed in TCGA lung cancer data. We further compared our results with several other independent sample cohorts and studies. We observed overall elevations in the expression of snoRNAs, including SNORD126, SNORA23, and SNORA42, in multiple cancer types, which are consistent with the results from published studies (Cui et al., 2017; Fang et al., 2017; Okugawa et al., 2017).

We also demonstrated the clinical relevance of snoRNA expression signatures and individual snoRNAs. Our unsupervised clustering showed that tumor subtypes defined by snoRNA expression have extensive and strong concordance with the subtypes defined by other molecular data. Furthermore, we identified 46 individual snoRNAs that have clinical relevance and which may play significant roles in tumorigenesis. We identified SNORD46, which is overexpressed in tumor samples in eight cancer types. We further experimentally demonstrated that knockdown of SNORD46 can lead to a significant reduction in cell viability and impaired cell migration and invasion in two independent cell lines, the breast cancer cell line MDA-MB-231 and the lung cancer cell line A549. Taken together, our combined computational and experimental analyses indicate that SNORD46 is an oncogene. In addition, we attempted to knock out SNORD15A and SNORD46 by CRISPR/cas9-mediated genomic editing (Platt et al., 2014). However, cells harboring designed single-guide RNAs failed to grow as single-cell colonies, suggesting the essential role of SNORD15A and SNORD46 in cancer cells.

To help the broader research community visualize and explore the expression and function of snoRNAs, we constructed an interactive and user-friendly data portal, SNORic. We expect that this data portal will facilitate the identification of biomarkers or therapeutic targets for cancer based on snoRNAs.

EXPERIMENTAL PROCEDURES

Data Collection

We downloaded the gene annotations (hg19) of 403 and 1,457 snoRNA genes from UCSC Genome Browser and GENCODE, respectively, and merged them as 1,524 unique snoRNA genes. We obtained snoRNA class information from UCSC Genome Browser, Rfam, and Ensembl. We downloaded 16,591 miRNA-seq binary alignment map (BAM) files across 31 cancer types from the Cancer Genomics Hub (CGHub) (<https://cghub.ucsc.edu>; Figure S6). We downloaded DNA methylation data, CNV data, mRNA expression, and protein expression data from TCGA data portal (<https://portal.gdc.cancer.gov/>). The annotation of RefSeq genes was downloaded from UCSC Table Browser (<http://genome.ucsc.edu/cgi-bin/hgTables>; hg19).

Characterization of the Expression Profiles of snoRNAs

We mapped the reads to snoRNA genes and quantified the expression of snoRNAs as RPKM. snoRNAs with an average RPKM ≥ 1 across samples in each cancer type were defined as detectable snoRNAs. snoRNA expression profiles were deposited in both SNORic (<http://bioinfo.life.hust.edu.cn/SNORic>) and Synapse (<https://www.synapse.org>, syn8370863).

Integrative Analysis of snoRNA Expression

We calculated the Spearman correlation between snoRNA expression and their host gene expression, their own CNV, and DNA methylation. We also calculated the Spearman correlation between the snoRNA, mRNA/proteins, and splicing. We used ConsensusClusterPlus (Wilkerson and Hayes, 2010) to classify the tumor samples and identify clinically relevant snoRNAs. Detailed information is provided in the Supplemental Experimental Procedures.

Functional Characterization of snoRNAs

We identified the interacting proteins for the top clinically relevant snoRNAs through RNA pull-down, followed by MS analysis. We then knocked down SNORD46 and examined its functional effects through cell proliferation, migration, and invasion assays. Detailed information is provided in the Supplemental Experimental Procedures.

Statistical Analysis

We used a paired Student's t test to assess the statistical difference in the snoRNA expression level between tumor and normal samples and considered $|\text{fold change}| \geq 2$ and $\text{FDR} < 0.05$ to be significant. We used the univariate Cox test to examine the association between snoRNA expression and 5-year survival times of the patient and considered $\text{FDR} < 0.05$ to be significant. The correlations between snoRNA expression and other molecular data were calculated by Spearman correlation, and coefficient $|R_s| \geq 0.3$ and $\text{FDR} < 0.05$ were considered as significant correlations. ConsensusClusterPlus (Wilkerson and Hayes, 2010) was used to classify the tumor samples into clusters based on the snoRNA expression level. GO enrichment analysis was performed by Gstats (Falcon and Gentleman, 2007).

Implementation of the SNORic Data Portal

The expression of snoRNAs in each cancer type and related data (mRNA expression, CNV, DNA methylation, and clinical information) have been organized into a set of relational MySQL tables. Django (v1.9.7), an open-source web framework based on the WSGI and Apache (v2.4.18), was used to construct the SNORic database. Correlation, differential analyses, and survival analyses were performed in R.

Figure 7. Overview of SNORic Data Portal and Example of SNORD46

- (A) Four modules in SNORic: summary; snoRNA-based analysis; gene-based analysis; and download.
- (B) Examples in snoRNA-based analysis module. The differential expression of SNORD46 in breast invasive carcinoma (BRCA) tumor and normal samples, among kidney chromophobe (KICH) tumor stages, among BRCA PAM50 subtypes, and association with patient 5-year survival in adrenocortical carcinoma (ACC). SNORD60 is significantly correlated with CNV in KICH. SNORD93 is significantly correlated with its DNA methylation in uveal melanoma (UVM).
- (C) Examples in gene-based analysis module. Users can input a gene of interest to obtain its correlated snoRNAs. The mRNA expression level of *CKS2* is significantly correlated with SNORA56 in thymoma (THYM), the splicing level of *MTA3* is significantly correlated with SNORD46 in testicular germ cell tumors (TGCT), and protein expression level of *FOXO3* is significantly correlated with SNORA71D in stomach adenocarcinoma (STAD).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and two tables and can be found with this article online at <https://doi.org/10.1016/j.celrep.2017.10.070>.

AUTHOR CONTRIBUTIONS

L.H. conceived and supervised the project. J.G., Y.X., Y.Y., Z.Z., L.D., and L.H. performed the data analyses; Y.L., C. Li, D.H.H., P.K.P., L.Y., and C. Lin performed the experiments. C. Liu, A.-Y.G., and L.H. developed the data portal; J.G., Y.X., J.A.P., C. Lin, and L.H. wrote the manuscript with input from all other authors.

ACKNOWLEDGMENTS

We gratefully acknowledge TCGA Research Network. This work was supported by the Cancer Prevention & Research Institute of Texas (RR150085 to L.H., R1218 to L.Y., and DP150093 to J.A.P.); UTHHealth Innovation for Cancer Prevention Research Training Program Post-doctoral Fellowship (RP160015); Department of Defense Breakthrough Award (BC151465 to L.Y.); NIH (R00CA166527 to L.Y., R00DK094981 and R01 CA218025 to C. Lin, and 1R01GM104290 to J.A.P.); China Scholarship Council (201606160058 to C. Liu); and National Natural Science Foundation of China (31471247 to A.-Y.G.). We thank LeeAnn Chastain and Sergey D. Egranov for manuscript proofreading.

Received: June 5, 2017

Revised: September 1, 2017

Accepted: October 18, 2017

Published: November 14, 2017

REFERENCES

- Bachellerie, J.P., Cavallé, J., and Hüttenhofer, A. (2002). The expanding snoRNA world. *Biochimie* **84**, 775–790.
- Balakin, A.G., Smith, L., and Fournier, M.J. (1996). The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86**, 823–834.
- Blaxall, B.C., Pellett, A.C., Wu, S.C., Pende, A., and Port, J.D. (2000). Purification and characterization of beta-adrenergic receptor mRNA-binding proteins. *J. Biol. Chem.* **275**, 4290–4297.
- Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70.
- Cancer Genome Atlas Research Network (2014). Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209.
- Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421.
- Choi, Y.L., Bocanegra, M., Kwon, M.J., Shin, Y.K., Nam, S.J., Yang, J.H., Kao, J., Godwin, A.K., and Pollack, J.R. (2010). LYN is a mediator of epithelial-mesenchymal transition and a target of dasatinib in breast cancer. *Cancer Res.* **70**, 2296–2306.
- Crea, F., Quagliata, L., Michael, A., Liu, H.H., Frumento, P., Azad, A.A., Xue, H., Pikor, L., Watahiki, A., Morant, R., et al. (2016). Integrated analysis of the prostate cancer small-nucleolar transcriptome reveals SNORA55 as a driver of prostate cancer progression. *Mol. Oncol.* **10**, 693–703.
- Cui, L., Nakano, K., Obchoei, S., Setoguchi, K., Matsumoto, M., Yamamoto, T., Obika, S., Shimada, K., and Hiraoka, N. (2017). Small nucleolar noncoding RNA SNORA23, up-regulated in human pancreatic ductal adenocarcinoma, regulates expression of spectrin repeat-containing nuclear envelope 2 to promote growth and metastasis of xenograft tumors in mice. *Gastroenterology* **153**, 292–306.e2.
- Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258.
- Fang, X., Yang, D., Luo, H., Wu, S., Dong, W., Xiao, J., Yuan, S., Ni, A., Zhang, K.J., Liu, X.Y., and Chu, L. (2017). SNORD126 promotes HCC and CRC cell growth by activating the PI3K-AKT pathway through FGFR2. *J. Mol. Cell Biol.* **9**, 243–255.
- Filipowicz, W., and Pogacić, V. (2002). Biogenesis of small nucleolar ribonucleoproteins. *Curr. Opin. Cell Biol.* **14**, 319–327.
- Fu, X., Meng, Z., Liang, W., Tian, Y., Wang, X., Han, W., Lou, G., Wang, X., Lou, F., Yen, Y., et al. (2014). miR-26a enhances miRNA biogenesis by targeting Lin28B and Zcchc11 to suppress tumor growth and metastasis. *Oncogene* **33**, 4296–4306.
- Gamazon, E.R., and Stranger, B.E. (2015). The impact of human copy number variation on gene expression. *Brief. Funct. Genomics* **14**, 352–357.
- Gao, L., Ma, J., Mannoor, K., Guarnera, M.A., Shetty, A., Zhan, M., Xing, L., Stass, S.A., and Jiang, F. (2015). Genome-wide small nucleolar RNA expression analysis of lung cancer by next-generation deep sequencing. *Int. J. Cancer* **136**, E623–E629.
- Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. *Nucleic Acids Res.* **36**, W70–W74.
- Han, L., Yuan, Y., Zheng, S., Yang, Y., Li, J., Edgerton, M.E., Diao, L., Xu, Y., Verhaak, R.G.W., and Liang, H. (2014). The pan-cancer analysis of pseudo-gene expression reveals biologically and clinically relevant tumour subtypes. *Nat. Commun.* **5**, 3963.
- Han, L., Diao, L., Yu, S., Xu, X., Li, J., Zhang, R., Yang, Y., Werner, H.M.J., Eterovic, A.K., Yuan, Y., et al. (2015). The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. *Cancer Cell* **28**, 515–528.
- Haoues, M., Refai, A., Mallavialle, A., Barbouche, M.R., Laabidi, N., Deckert, M., and Essafi, M. (2014). Forkhead box O3 (FOXO3) transcription factor mediates apoptosis in BCG-infected macrophages. *Cell. Microbiol.* **16**, 1378–1390.
- Hoadley, K.A., Yau, C., Wolf, D.M., Cherniack, A.D., Tamborero, D., Ng, S., Leiserson, M.D.M., Niu, B., McLellan, M.D., Uzunangelov, V., et al.; Cancer Genome Atlas Research Network (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158**, 929–944.
- Ingle, E. (2012). Functions of the Lyn tyrosine kinase in health and disease. *Cell Commun. Signal.* **10**, 21.
- Jorjani, H., Kehr, S., Jedlinski, D.J., Gumienny, R., Hertel, J., Stadler, P.F., Zavalan, M., and Gruber, A.R. (2016). An updated human snoRNAome. *Nucleic Acids Res.* **44**, 5068–5082.
- Kim, M., Vasiljeva, L., Rando, O.J., Zhelkovsky, A., Moore, C., and Buratowski, S. (2006). Distinct pathways for snoRNA and mRNA termination. *Mol. Cell* **24**, 723–734.
- Kiss, T. (2002). Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109**, 145–148.
- Kiss, T., Fayet-Lebaron, E., and Jády, B.E. (2010). Box H/ACA small ribonucleoproteins. *Mol. Cell* **37**, 597–606.
- Krishnan, P., Ghosh, S., Wang, B., Heyns, M., Graham, K., Mackey, J.R., Kovalchuk, O., and Damaraju, S. (2016). Profiling of small nucleolar RNAs by next generation sequencing: potential new players for breast cancer prognosis. *PLoS ONE* **11**, e0162622.
- Kurreck, J., Wyszko, E., Gillen, C., and Erdmann, V.A. (2002). Design of anti-sense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res.* **30**, 1911–1918.
- Li, J., Han, L., Roebuck, P., Diao, L., Liu, L., Yuan, Y., Weinstein, J.N., and Liang, H. (2015). TANRIC: an interactive open platform to explore the function of lncRNAs in cancer. *Cancer Res.* **75**, 3728–3737.
- Li, C., Wang, S., Xing, Z., Lin, A., Liang, K., Song, J., Hu, Q., Yao, J., Chen, Z., Park, P.K., et al. (2017). A ROR1-HER3-lncRNA signalling axis modulates the Hippo-YAP pathway to regulate bone metastasis. *Nat. Cell Biol.* **19**, 106–119.

- Lin, A., Li, C., Xing, Z., Hu, Q., Liang, K., Han, L., Wang, C., Hawke, D.H., Wang, S., Zhang, Y., et al. (2016). The LINK-A lncRNA activates normoxic HIF1 α signalling in triple-negative breast cancer. *Nat. Cell Biol.* **18**, 213–224.
- Liu, X., Xiao, Z.D., Han, L., Zhang, J., Lee, S.W., Wang, W., Lee, H., Zhuang, L., Chen, J., Lin, H.K., et al. (2016). LncRNA NBR2 engages a metabolic checkpoint by regulating AMPK under energy stress. *Nat. Cell Biol.* **18**, 431–442.
- Martens-Uzunova, E.S., Hoogstrate, Y., Kalsbeek, A., Pigmans, B., Vredenburg-van den Berg, M., Dits, N., Nielsen, S.J., Baker, A., Visakorpi, T., Bangma, C., and Jenster, G. (2015). C/D-box snoRNA-derived RNA production is associated with malignant transformation and metastatic progression in prostate cancer. *Oncotarget* **6**, 17430–17444.
- Okugawa, Y., Toiyama, Y., Toden, S., Mitoma, H., Nagasaka, T., Tanaka, K., Inoue, Y., Kusunoki, M., Boland, C.R., and Goel, A. (2017). Clinical significance of SNORA42 as an oncogene and a prognostic biomarker in colorectal cancer. *Gut* **66**, 107–117.
- Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401.
- Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455.
- Reichow, S.L., Hamma, T., Ferré-D'Amaré, A.R., and Varani, G. (2007). The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res.* **35**, 1452–1464.
- Siprashvili, Z., Webster, D.E., Johnston, D., Shenoy, R.M., Ungewickell, A.J., Bhaduri, A., Flockhart, R., Zarnegar, B.J., Che, Y., Meschi, F., et al. (2016). The noncoding RNAs SNORD50A and SNORD50B bind K-Ras and are recurrently deleted in human cancer. *Nat. Genet.* **48**, 53–58.
- Spacková, N., Réblová, K., and Sponer, J. (2010). Structural dynamics of the box C/D RNA kink-turn and its complex with proteins: the role of the A-minor O interaction, long-residency water bridges, and structural ion-binding sites revealed by molecular simulations. *J. Phys. Chem. B* **114**, 10581–10593.
- Szewczak, L.B., Gabrielsen, J.S., Degregorio, S.J., Strobel, S.A., and Steitz, J.A. (2005). Molecular basis for RNA kink-turn recognition by the h15.5K small RNP protein. *RNA* **11**, 1407–1419.
- Takabatake, Y., Oxvig, C., Nagi, C., Adelson, K., Jaffer, S., Schmidt, H., Keely, P.J., Eliceiri, K.W., Mandeli, J., and Germain, D. (2016). Lactation opposes pappalysin-1-driven pregnancy-associated breast cancer. *EMBO Mol. Med.* **8**, 388–406.
- Tycowski, K.T., Shu, M.D., Kukoyi, A., and Steitz, J.A. (2009). A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. *Mol. Cell* **34**, 47–57.
- Wang, Y., Waters, J., Leung, M.L., Unruh, A., Roh, W., Shi, X., Chen, K., Scheet, P., Vattathil, S., Liang, H., et al. (2014). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* **512**, 155–160.
- Wang, Y., Xu, X., Yu, S., Jeong, K.J., Zhou, Z., Han, L., Tsang, Y.H., Li, J., Chen, H., Mangala, L.S., et al. (2017). Systematic characterization of A-to-I RNA editing hotspots in microRNAs across human cancers. *Genome Res.* **27**, 1112–1125.
- Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573.
- Williams, G.T., and Farzaneh, F. (2012). Are snoRNAs and snoRNA host genes new players in cancer? *Nat. Rev. Cancer* **12**, 84–88.
- Xing, Z., Lin, A., Li, C., Liang, K., Wang, S., Liu, Y., Park, P.K., Qin, L., Wei, Y., Hawke, D.H., et al. (2014). lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell* **159**, 1110–1125.
- Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., and Liang, G. (2014). Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **26**, 577–590.
- Yoshihara, K., Shahmoradgol, M., Martínez, E., Vegesna, R., Kim, H., Torres-García, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612.
- Yu, M.H., Luo, Y., Qin, S.L., Wang, Z.S., Mu, Y.F., and Zhong, M. (2015). Up-regulated CKS2 promotes tumor progression and predicts a poor prognosis in human colorectal cancer. *Am. J. Cancer Res.* **5**, 2708–2718.
- Yuan, J., Krämer, A., Matthess, Y., Yan, R., Spänkuch, B., Gätje, R., Knecht, R., Kaufmann, M., and Strebhardt, K. (2006). Stable gene silencing of cyclin B1 in tumor cells increases susceptibility to taxol and leads to growth arrest in vivo. *Oncogene* **25**, 1753–1762.
- Zack, T.I., Schumacher, S.E., Carter, S.L., Cherniack, A.D., Saksena, G., Tabak, B., Lawrence, M.S., Zhsng, C.Z., Wala, J., Mermel, C.H., et al. (2013). Pan-cancer patterns of somatic copy number alteration. *Nat. Genet.* **45**, 1134–1140.
- Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916–919.
- Zheng, D., Zhang, J., Ni, J., Luo, J., Wang, J., Tang, L., Zhang, L., Wang, L., Xu, J., Su, B., et al. (2015). Small nucleolar RNA 78 promotes the tumorigenesis in non-small cell lung cancer. *J. Exp. Clin. Cancer Res.* **34**, 49.
- Zheng, L.L., Li, J.H., Wu, J., Sun, W.J., Liu, S., Wang, Z.L., Zhou, H., Yang, J.H., and Qu, L.H. (2016). deepBase v2.0: identification, expression, evolution and function of small RNAs, LncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Res.* **44** (D1), D196–D202.