Identification of novel multi-omics expression landscapes and meta-analysis of landscape-based competitive endogenous RNA networks in ALDH+ lung adenocarcinoma stem cells

Wei Yang[1,2,3,4,#]**, Yong Liang**[5,#]**, Yuanyuan Zheng**[1,2,3,4]**, Haitao Luo**[1,2,3,4]**, Xiaofei Yang**[1,2,3,4*] **and Furong Li**[1,2,3,4*]

1 Translational Medicine Collaborative Innovation Center, The Second Clinical Medical College (Shenzhen People's Hospital), Jinan University, Shenzhen 518020, China.
2 The First Affiliated Hospital (Shenzhen People's Hospital), Southern University of Science and Technology, Shenzhen 518020, China.
3 Guangdong Engineering Technology Research Center of Stem cell and Cell therapy, Shenzhen 518020, China.
4 Shenzhen key Laboratory of Stem Cell Research and Clinical Transformation, Shenzhen 518020, China
5 Shenzhen Xbiome Biotech Co., Ltd, Shenzhen 518057, China
* Correspondence: frli62@163.com; sophiayang81@hotmail.com
\# Wei Yang and Yong Liang contributed equally to this manuscript.

Abstract: ALDH+ H1975 lung adenocarcinoma stem cells (LSCs) are a rare cell population identified in lung adenocarcinoma (LUAD). LSCs can self-renew, drive tumor initiation, growth, metastasis, and recurrence, and are also the predominant cause of poor prognosis due to their intrinsic resistance to drugs and chemotherapy. Consequently, LSCs are a promising target for LUAD therapy. Noncoding RNAs (ncRNAs), including microRNAs (miRNAs), long noncoding RNAs (lncRNAs), and circular RNAs (circRNAs), exert many significant regulatory functions in the pathogenesis of human cancers, showing the necessity for a comprehensive understanding of the mechanisms that underlie lung carcinogenesis. Nonetheless, research on many known transcripts and messenger RNAs (mRNAs) has already generated new information. Unknown biomarkers in ncRNAs and systematic and comprehensive interrelation with unknown ncRNAs and mRNAs may provide further insights into the biology of LUAD. Herein, a set of novel ncRNAs that include miRNAs, lncRNAs, circRNAs, were identified and differentially expressed patterns of ncRNAs and mRNAs in LSCs and ALDH- H1975 LUAD tumor cells (LTCs) were obtained using stringent bioinformatics pipelines. Through a meta-analysis of the identified landscapes, novel competitive endogenous RNA (ceRNA) networks were constructed to reveal the potential molecular mechanisms that regulate the hallmarks of LSCs and LTCs. This study presents a summary of novel ncRNAs and the fundamental roles of differentially expressed ncRNAs implicated in the activity of LSCs and LTCs. In addition, the study also provides a comprehensive resource for the future identification of diagnostic, therapeutic, and prognostic biomarkers in LUAD.

Keywords: LUAD; cancer stem cell; novel ncRNAs; ceRNA networks; molecular pathways

Note：A preprint has previously been published[1].

1. Introduction

Lung cancer is the most common respiratory tumor and is the leading cause of cancer deaths worldwide, with ~18% mortality rate and ~22.4% incident ratio among all cancer deaths[2,3]. Non-small cell lung cancer (NSCLC) is the most common form of lung cancer, accounting for ~90% of all cases, and lung adenocarcinoma (LUAD) is its most common histological subtype[4], with a dismal prognosis and 15% of patients surviving 5 years after diagnosis[5].

Cancer stem cells (CSCs), a subset of cancer cells with stem-like characteristics, play a critical role in tumor heterogeneity and are involved in tumor initiation, growth, metastasis, recurrence, and drug resistance[6]. Increasing research has demonstrated that the leading cause of poor prognosis for LUAD[7] is the presence of LUAD stem cells[8], which could initiate and sustain primary and metastatic cancer relapse and growth[9]. Previous evidence[10] and our most recent studies[11,12] showed that ALDH+ H1975 LUAD cells (LSCs) display strong CSC-like properties[13] and act as LSCs to influence the formation of a tumor niche, tumor initiation, and recurrence[7]. Our most recent study revealed that LSCs have the capacity for self-renewal and differentiation, which facilitates the generation of tumor cells and promotes tumor growth[11], recurrence, and metastasis. Therefore, LSCs are a promising target for interfering with the therapeutic mechanism of LUAD formation and LUAD therapy[7]. Elucidation of the differences and underlying molecular pathways that characterize LSCs[14] is important to identify and evaluate potential diagnostic, therapeutic, and prognostic biomarkers of LUAD[15].

LncRNAs, circRNAs, and miRNAs are three subtypes of non-coding RNAs, which account for a huge proportion of the human transcriptome[16]. The diverse types of these ncRNAs are marked by sequence length, spatial architecture, and regulatory functions in various pathophysiological processes in cancer. Separately, miRNAs[17] are the most studied single-stranded ncRNAs, with a length of ~20 nucleotides[18]. Through binding to specific 3'-untranslated regions (UTRs) of their target mRNAs or RNA, miRNAs could silence the expression of multiple targeted mRNAs and RNA simultaneously[19]. Regulatory modulation of one-to-many miRNAs influences multiple oncogenic and tumor repressive pathways[20,21]; for example, overexpression of has-miR-17-92 in lung cancer enhances cell proliferation[22], while has-let-7 in lung cancer is associated with the overall survival of patients[23]. LncRNAs are over 200 nucleotides in length[24] and have the ability to regulate gene expression rather than having protein-encoding capability[25]. Extensive studies have shown that lncRNAs may influence cancer progression by regulating mRNAs and miRNAs[26]. CircRNAs are a novel subtype of non-coding RNAs. Although circRNAs function as lncRNAs[27], their uniquely circular structure provides them with greater stability compared with linear lncRNAs[28]. CircRNAs play a vital role in various types of cancer[29]. Furthermore, there is evidence that mRNA transcripts are regulated or repressed by lncRNAs/circRNAs and miRNAs, individually or in combination [30]. A novel mechanism in the regulatory network[31] is the identification of the interaction between mRNAs, lncRNAs/circRNAs, and miRNAs, which has been defined as the ceRNA network. More importantly, lncRNAs/circRNAs upregulate transcript expression by binding miRNA sites, for example, lncRNA HRCR acts as a miR-233 sponge to prevent cardiac hypertrophy[32], lncRNA CHRF functions as an endogenous sponge of miR-489 to limit miR-489 expression[33], and lncRNA APF works as a sponge of miR-188-3p to prevent decomposition of ATG7[34].

Previous research focused on the differences between LUAD cells and the adjacent normal cells based on known transcripts and genes to explore the cause and treatment

options of tumorigenesis, tumor metastasis, relapse, and drug resistance. However, identification of unknown ncRNA sets and the associated molecular mechanism between LSCs and LUAD tumor cells (LTCs) is necessary to provide insight into the biological properties of LUAD.

In this study, the identification of unknown biomarkers and a meta-analysis of unknown ncRNAs and mRNAs between LSCs and LTCs were conducted to facilitate in the elucidation of the molecular events associated with tumorigenesis, metastasis, relapse, and drug resistance from the perspective of LSCs.

2. Materials and Methods

2.1 RNA isolation and miRNA sequencing

Two ALDH + LSC and two ALDH- LTC samples were obtained from our previous study[11]. A total of ~3–5×10⁶ cells were used for the isolation of RNA. Total RNA of cellular samples was extracted using TRIzol™ reagent (Invitrogen, Carlsbad, CA, USA) based on the manufacturer's protocols. At least 3 μg RNA per sample was used as a substrate for the subsequent analyses. Sequencing libraries were generated using the NEB-Next® Multiplex Small RNA Library Prep Set for Illumina® (NEB, USA), according to the manufacturer's recommendations. Index codes were added to the attribute sequences of each sample. Briefly, the NEB 3′ sequence replication (SR) adaptor was directly and specifically linked to the 3′ end of miRNA, small interfering RNA (siRNA), and piwi-interacting (piRNA). After the 3′ ligation reaction, the SR random primer (RT) was hybridized to the excess of the 3′ SR adaptor, which was transformed from a single-stranded DNA adaptor into a double-stranded DNA molecule. This important step prevented adaptor-dimer formation. Thus, dsDNAs were not ligation-mediated substrates for T4 RNA Ligase 1 and therefore did not ligate to the 5′ SR adaptor in the subsequent ligation step. The 5′ end adapters were ligated to the 5′ ends of miRNA, siRNA, and piRNA. First-strand cDNA was then synthesized using M-MuLV reverse transcriptase (RNase H-). PCR amplification was performed using LongAmp® Taq 2X Master Mix, SR Primer for Illumina, and index primer (X).

The PCR products were purified on an 8% polyacrylamide gel (100 V, 80 min). DNA fragments corresponding to 140–160 bp (the length of small non-coding RNA plus the 3′ and 5′ adaptors) were recovered and dissolved in 8 μL elution buffer. Library quality was evaluated on the Agilent Bioanalyzer 2100 system using DNA High Sensitivity Chips. Clustering of the index-coded samples was performed on a cBot Cluster Generation System using the TruSeq SR Cluster Kit v3-cBot-HS (Illumia) according to the manufacturer's instructions. After cluster generation, library preparations were sequenced on an Illumina Hiseq 2500/2000 platform, and 50-bp single-end reads were generated.

2.2 RNA raw data quality control and reads mapping statistics

Two LSC samples and two LTC samples were obtained from our previous study[11]. Raw reads of miRNAs and lncRNAs/circRNAs/mRNAs from the Illumina Hiseq platform were conducted using ReSeqTools[35] to remove reads with poly-N (ratio of N greater than 10%)/adapter sequence and low-quality reads (reads with more than 50% bases having quality Q-value ≤5). Raw reads of miRNAs were then processed to filter reads containing poly A/T/G/C regions. Clean reads from each miRNA sample were mapped to the human reference genome (version GRCh38) using Bowtie[36], with no mismatches. Clean reads of lncRNAs/circRNAs/mRNAs were aligned with the same human reference genome mentioned above using HISAT2 with '-rna-standness RF' and other default parameters[37]. The mapping results were sorted, analyzed, and indexed using ReSeqTools[38] and Samtools[39].

2.3 Identification of miRNAs

Based on the aligned small RNA reads, miRbase 20.0[40] was used as a reference, and miRDeep2/srna-tools-cli[41] were used to identify potential miRNAs and to define the secondary structures. Based on RepeatMasker[42], the Rfam database[43] mapped reads originating from protein-coding genes, repeat sequences, rRNA, tRNA, snRNA, and snoRNA[44] were removed. Based on the characteristics of the hairpin structure of the miRNA precursor, the Dicer cleavage site, and the minimum information of the small RNA reads not annotated in the above steps, miREvo[45] and miRDeep2[41] were integrated to predict novel miRNAs, and the base bias was counted at the first position and at all positions of all identified miRNAs. For alignment and annotation, the diversity RNA was counted with the following priority rule: known miRNA > rRNA > tRNA > snRNA > snoRNA > repeat > mRNA > novel miRNA, to ensure each unique small RNA to be matched to only one annotation. MiRNA expression levels were estimated as Transcripts Per Million (TPM)[46] using StringTie[47], whose equation was normalized to the expression level:

$$TPMi = (Ni/Li)*1{,}000{,}000/sum\ (Ni/Li + \ldots\ldots + Nm/Lm)$$

In the above formula, Ni indicates the mapped reads in exon i. Li indicates the length of exon i. Nm indicates the mapped reads in exon m. Lm indicates the length of exon m.

2.4 Identification and quantification of lncRNAs/circRNAs and mRNAs

The total mapped reads for each sample were assembled by StringTie[47] using a reference-based approach with an optional de novo assembly step, and a comprehensive annotation file with full-length transcripts and potential novel ncRNAs was generated. Transcripts with low confidence (exon numbers <2), length ≤200 nt, repeat annotations, and low expression (FPKM<0.05) were removed by the filtering step in Figure 2c. Quantification of known ncRNAs and mRNAs was performed by StringTie for each sample. Four tools, including CNCI (Coding-Non-Coding-Index)[48], CPC (Coding Potential Calculator)[49], Pfam-scan[50], and PhyloCSF (phylogenetic codon substitution frequency)[51] with default parameters, were used to predict the coding potential of transcripts, which were filtered and became candidate sets of novel lncRNAs. Cufflinks were used to calculate Fragments Per Kilobase of exon model per Million mapped fragments (FPKM)[52] of both the lncRNAs and the coding genes in each sample. FPKMs were defined as the total mapped exon Fragments/(mapped reads (millions)*exon length (KB).

The BAM files mentioned above were used to identify circRNAs. The detection and identification of circRNAs was based on our previous studies[11]. TMP was used to quantify the circle RNAs.

2.5 Differential expression analysis

After quantification of digital transcript expression, a differential expression analysis of lncRNAs and mRNAs between the LSC and LTC groups was performed using Cuffdiff[53], which was based on the negative binomial distribution model. mRNA gene transcripts with a P-adjust < 0.05 and |log2 foldchange| ≥ 2 were assigned as differentially expressed (DE) genes.

DE analysis of miRNA and circRNAs between two conditions was performed using the DESeq2 R package[54]. *P*-values were adjusted using the Benjamini & Hochberg method to control the false discovery rate[55]. Corrected *P*-value < 0.05 and |log2 foldchange| ≥ 2 were established as the threshold for significant DE.

2.6 Kyoto Encyclopedia of Genes and Genomes analysis and Gene Ontology analysis on differentially expressed mRNAs

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of DE mRNA was conducted using DAVID 6.8[56] with default parameters to ascertain the potential functions of genes that participate in cancer progression of LSCs and LTCs, separately. Gene Ontology (GO) analysis was performed based on Metascape 3.5 websites[57], which organized genes into hierarchical categories and uncovered the gene regulatory network using the database of the most recent biological processes. The *P*-value was adjusted using the Benjamini & Hochberg method to control the false discovery rate and to identify significant KEGG pathways and GO terms.

2.7 Analysis of the CeRNA network and survival analysis

LncRNA-mRNA co-location networks were predicted and performed based on the parameters of the upstream and downstream 100 kb distance of DE lncRNAs.

Prediction of miRNA target gene was performed on miRanda[58], PITA[59], and RNAhybrid[60], separately. The overlapping results of miRNAs and mRNAs pairs using the three software algorithms provided the final pairings. Concurrently, miRNA target prediction of lncRNAs/circRNAs was performed by miRanda.

The ceRNA networks were constructed using Cytoscape 3.9.0[61] and the hub modules in the ceRNA networks were identified using MCODE 2.0.0[62], a Cytoscape 3.9.0 plug-in, with default parameters. Each molecule of those hub modules was identified using the UALCAN website and the lnCAR website to obtain known hub molecules.

According to the location of the novel molecules and using the gene symbols of known molecules, survival analysis was performed using the UALCAN website[63] and the lnCAR website[64] with default parameters.

3. Results

3.1. Identification of novel and known miRNAs and differentially expressed miRNAs in LSCs and LTCs

Two LSC samples (ALDH+ H1975 LUAD tumor cells (LSC1-m) and ALDH+ H1975 LUAD tumor cells (LSC2-m)) and two LTC samples (ALDH- H1975 LUAD tumor cell (LTC1-m) and ALDH- H1975 LUAD tumor cells (LTC2-m)) were obtained as described in our previous study[11]. Raw MiRNA data were obtained using the Illumina Hiseq 2500/2000 platform and 50-bp single-end reads were generated. Sequencing reads of small RNA tags (~18–30 nt) extracted from raw reads were 16.62 Mb (LTC1-m), 17.93 Mb (LTC2-m), 15.58 Mb (LSC1-m), and 11.82 Mb (LSC2-m), respectively. The correlation between these four samples is shown in Figure 1b. MiRNAs expression landscape analysis on sequencing reads was performed to dissect the distribution of ncRNAs, to predict novel miRNAs, also known as miRNAs and DE miRNAs. Of these, at least 91.61% of the total small RNA tags were perfectly mapped to the Hg38.94 genome in the four tumor samples. Interestingly, more than 55.74% of the small RNA tags in LTC2-m were mapped to known miRNAs, and only 0.01% of those small RNA tags were predicted to be novel miRNAs. In contrast, the percentage of known miRNAs in LSC2-m was only 32.5%, while the percentage of novel miRNAs was only 0.02%. The remaining small RNA tags were assigned to rRNA, tRNA, snoRNA, and exons. (Supplementary Table S1.1). The most abundant class of small RNA tags was miRNAs, followed by other (Figure 1a). Overall, 100 new miRNAs and 1379 known miRNAs were obtained (Figure 1c). The base bias at the first position (based on 22 nt) and all the positions of the identified miRNAs (~18–30 nt) were similar in the new miRNAs and the known miRNAs (Figure 1d; Figure S1). These predicted novel miRNAs and known miRNAs (Figure 1e; Table S1.3) were used for the subsequent analysis. The overview of the study flow is shown in Figure 1 and Supplementary Figure S1.

A total of 954 common miRNAs were obtained in the two groups (LSCs contained LSC1-m/LSC2-m and the LTC group consisted of LTC1-m/LTC2-m). A total of 208 specific

miRNAs were expressed in the LSC group, and 196 specific miRNAs were expressed in the LTC group (Figure 1f). Of these, 53 upregulated miRNAs and 27 downregulated miR-NAs were identified (Figure 1g).
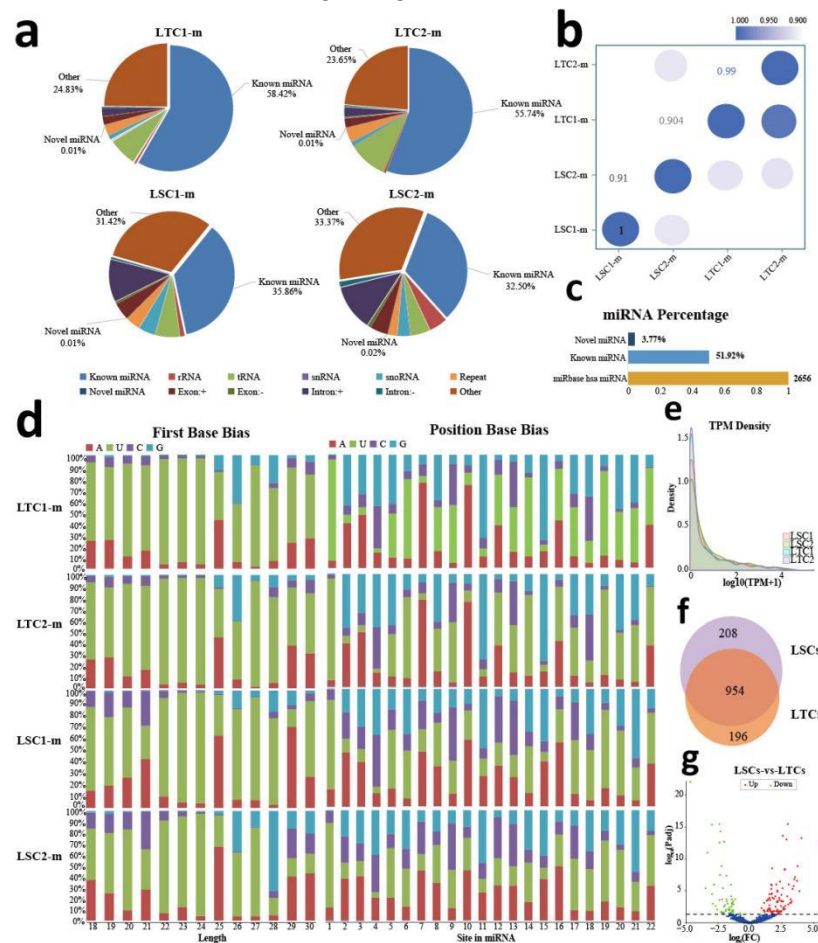
Figure 1. Distribution of novel and known miRNA and DE miRNAs. (a) small RNA (sRNA) category ratio; (b) Sample correlation; (c) MiRNA percentage from the miRbase database of known miR-NAs; (d) first base bias and position base bias on known miRNA; (e) TMP density of all miRNAs; (f) Venn diagram on predicted miRNA in LSCs and LTCs; (g) volcano plot of DE miRNAs

3.2. Identification of novel and known lncRNAs/circRNAs and DE lncRNAs/circRNAs, mRNA identified in LSCs and LTCs

Raw data relative to lncRNAs/circRNAs/mRNAs were retrieved from our previous study[11]. Reads were sequenced using the Illumina Hiseq2500 platform as 125 bp pair-end reads. The correlation between the four samples evaluated is shown in Figure 2a. Approximately 79.3 Mb pair-end reads were obtained, 93.45% of which could be matched to the Hg38.94 genome. Subsequently, the distribution of known transcripts was identified using HTseq software. Approximately 79.64% of the reads were in protein-coding genes and 4.12% were known lncRNAs for a total of 3091 known lncRNAs (Figure 2b, h). Meanwhile, 188,945 new transcripts were assembled using HISAT2 and StringTie. Based on the assembled transcripts and known transcripts, a total of 6487 novel lncRNAs were identified using a specific bioinformatics pipeline A (Figure 2c, h; Methods 2.4). The most abundant size class of known lncRNAs consisted of antisense lncRNAs (37%) and lincRNAs (31%), and the novel lncRNAs were intronic lncRNAs (77%) (Figure 2f). In addition, from these short reads, 681 known circRNAs and 173 novel circRNAs (Figure 2h) were identified using the in-silico pipeline B (Figure 2d; Methods 2.4). Most of the circRNAs were located in exon regions (Figure 2i). Of these, 630 circRNAs were identified (Figure 2j) in

two groups. A total of 155 circRNAs were only expressed in LTCs, while 68 circRNAs were only expressed in LSCs.
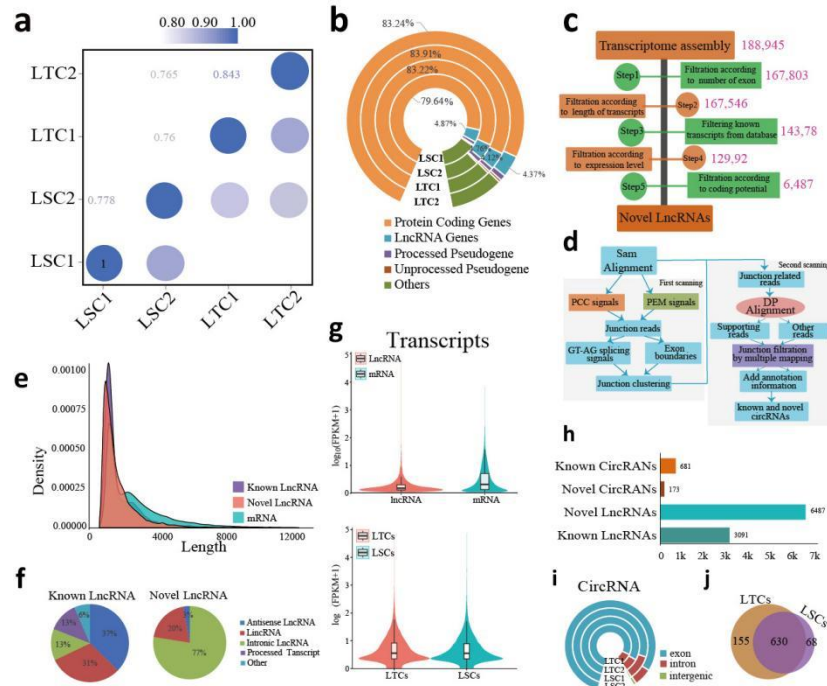
Figure 2. Distribution of novel and known lncRNAs/circRNAs and DE lncRNAs/circRNAs, mRNAs. (a) Samples correlation; (b) Transcript classes; (c) in-silico pipeline used for the identification of novel lncRNAs. The purple number next to the bar is the remaining count of lncRNAs after filtering each step; (d) Bioinformatics pipeline used for the identification of circRNAs; (e) FPKM density of all transcripts; (f) Composition of known and novel lncRNAs ; (g) Violin plot of transcripts between lncRNA and mRNA, and group LTCs and group LSCs; (h) Counts of lncRNA and circRNA; (i) Composition of circRNAs; (j)Venn diagram showing predicted circRNAs in LTCs and LSCs, respectively.

3.3. Differential expression analysis and annotation analysis

Quantitative analysis on mRNAs/lncRNAs was performed using Cufflinks software (Figure 2e, g) and Cuffdiff was used for DE analysis on lncRNAs, yielding 16 upregulated lncRNAs, 18 downregulated lncRNAs, 378 upregulated mRNAs, and 364 downregulated mRNAs (Padj<0.05 and |log2 (Fold Change)|≥2). Meanwhile, DE analysis was conducted on circRNAs using DESeq2 software, and 54 upregulated circRNAs and 60 downregulated circRNAs (Padj<0.05 and |log2 (Fold Change)|≥2) were obtained.

KEGG analysis was performed to identify the potential molecular pathways that regulate LSCs and LTCs. Upregulated gene sets and downregulated gene sets were analyzed separately. The results of KEGG analysis identified common pathways (Figure 3a) in LTCs and LSCs, which showed the commonality between LSCs and LTCs, such as Pathways in cancer, the MAPK signaling pathway, and the PI3K-Akt signaling pathway. KEGG analysis of upregulated gene sets (Figure 3c) included pathways, such as Steroid biosynthesis, Fatty acid metabolism, Antigen processing and presentation, Fatty acid degradation, Inflammatory bowel disease (IBD), Biosynthesis of antibiotics, HTLV-I infection[65],and Drug metabolism-other enzymes, which may be related to immune evasion[66] and poor prognosis[65] in LUAD. Other pathways (Figure 3b), which were enriched by upregulated gene sets included Cell cycle, Osteoclast differentiation, Progesterone-mediated oocyte maturation[67], and Hippo signaling pathway, and these may be involved in cancer proliferation and cancer cell heterogeneity[68].

With screening criteria using the *P*-value < 0.01, GO analysis was conducted on up-regulated DE mRNAs and downregulated DE mRNAs, separately, then hierarchical clustering on GO analysis was performed. Nine clusters of GO terms were found based on downregulated gene sets and 12 clusters of GO terms were found based on upregulated gene sets (Figure 3d, e). The nine clusters included cell fate commitment, positive regulation of programmed cell death, intrinsic apoptotic signaling pathway, cell junction organization, actin cytoskeleton organization, regulation of cell-substrate adhesion[69], positive regulation of cell migration, and regulation of I-kappaB kinase/NF-kappaB signaling[70], which may play a vital role in cell migration, cell apoptosis, and cell communication as described in previous studies[71]. Clusters of GO terms based on upregulated gene sets were enriched in the mitotic cell cycle, DNA replication initiation, telomere maintenance, regulation of cell morphogenesis, homeostasis of the number of cells, cell morphogenesis involved in differentiation, cell junction organization, positive regulation of alpha-beta T cell activation, steroid biosynthetic process, which may be linked to cancer growth, cancer proliferation, metastasis, and immune evasion.
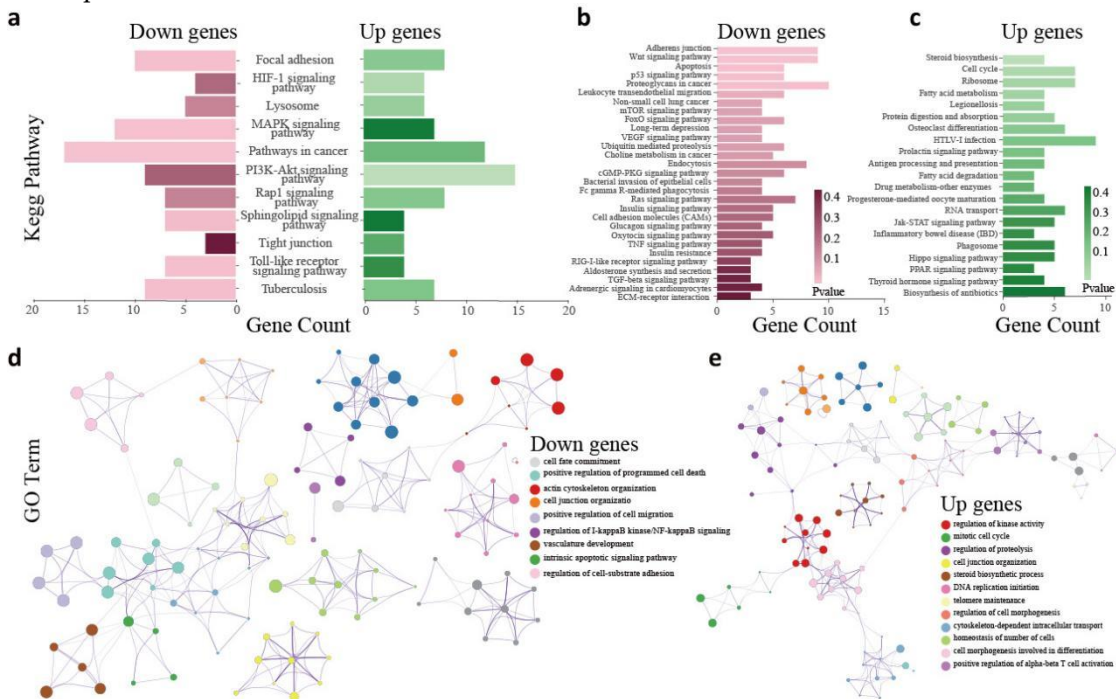
Figure 3. KEGG pathway analysis and GO cluster analysis. (a) Common pathway found in both upregulated gene sets and downregulated gene sets; (b) specific Kegg pathways in downregulated gene sets; (c) specific KEGG pathways in upregulated gene sets; (d)clusters of GO terms in downregulated gene sets; (e) clusters of GO terms in downregulated gene sets.

3.4. Target prediction and integrative analysis of competitive endogenous RNA (ceRNA) networks

Of the 742 DE mRNAs and 80 DE miRNAs, the target mRNAs of miRNAs were predicted using miRana, PITA, and RNAhybrid, separately. The overlapping results in the three software algorithms represented the final results. Of the 32 DE lncRNAs, 115 DE circRNAs, and 80 DE miRNAs, miRanda was used to search for the target lncRNAs/circRNAs of miRNAs . In addition, target mRNAs of lncRNAs were searched by scanning coding genes 10k/100k upstream and downstream of lncRNAs, and 12 lncRNA-mRNA pairs were obtained. Finally, ceRNA networks were constructed with 420 pairs, which consisted of 90 DE mRNAs, 80 DE miRNAs and 32 DE lncRNAs/96 DE circRNAs (Table S5.1:ceRNA-pairs; Table S5.2: ceRNA-node-information**)**. The 80 DE miRNAs included 78

known miRNAs and 2 novel miRNAs; the 32 DE lncRNAs included 8 known lncRNAs and 24 novel lncRNAs; and the 96 DE circRNAs included 73 known circRNAs and 32 novel circRNAs (Table S5.2:ceRNA-nodes-information). The 420 pairs included 363 pairs comprising known molecules. The remaining pairs mostly involved novel molecules. Based on the StarBase v3.0[72] website, 280 pairs were identified, which contained most of the known ncRNAs and mRNAs that have been described previously. Interestingly, based on the ceRNA networks constructed using Cytoscape, we determined that the network could be divided into two subpopulations (Figure 4) yielding two sub-networks that satisfied the ceRNA theory. That is, the right sub-network in Figure 4 included downregulated miRNAs, upregulated mRNAs, and lncRNA/circRNAs, while the sub-network on the left one consisted of upregulated miRNAs, downregulated mRNAs, and lncRNA/circRNAs.
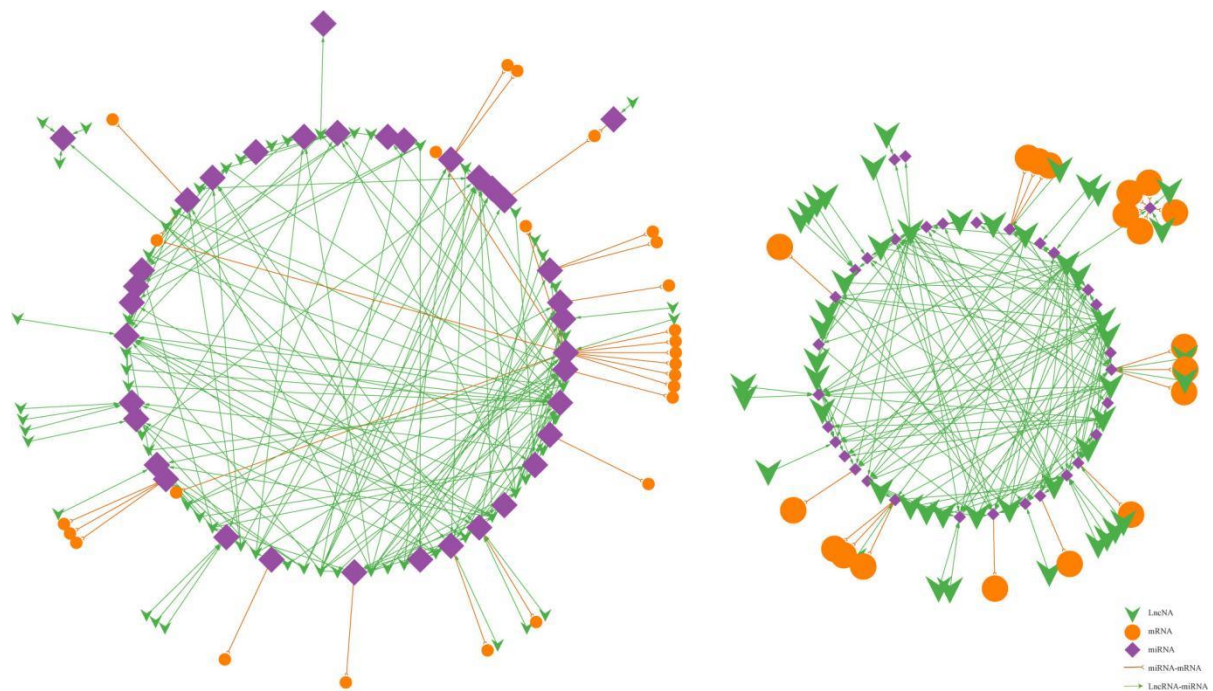
Figure 4. Constructed ceRNA networks. The left sub-network consisted of upregulated miRNAs, downregulated mRNAs, and downregulated lncRNA/circRNAs; the right sub-network consisted of downregulated miRNAs, upregulated mRNAs, and upregulated lncRNA/circRNAs. Green arrow, lncRNA/circRNA; orange circle, mRNA; purple rhombus, miRNAs; combination of brown line and semicircle, inhibition or degradation, respectively; light green line or arrow, promotion. A large or small shape (green arrow/orange circle/purple rhombus) indicates upregulation or downregulation, respectively.

To reveal different pathways based on DE mRNAs in the ceRNA network of LSCs and LTs, GO analysis was performed on 33 upregulated genes and 54 downregulated genes. The results are shown in Figure 5a, b. The GO terms of downregulated gene sets clustered in cell junction, localization of cell, cell motility, cell migration, locomotion, intracellular signal transduction, endocytosis, extracellular matrix terms. These enriched GO terms were similar to those of upregulated gene sets, including cell adhesion, positive regulation of signaling transduction, vesicle-mediated transport, lysosome, regulation of secretion, cell migration, locomotion, localization of cell, and cell motility. The remaining GO terms in the upregulated gene sets were cell growth, cell differentiation, cell junction assembly, cell morphogenesis, and cell proliferation.

Next, the hub modules in ceRNA networks were identified using the Cytoscape plug-in: MCODE with default parameters. Four main hub modules were obtained as shown in Figure 5c.
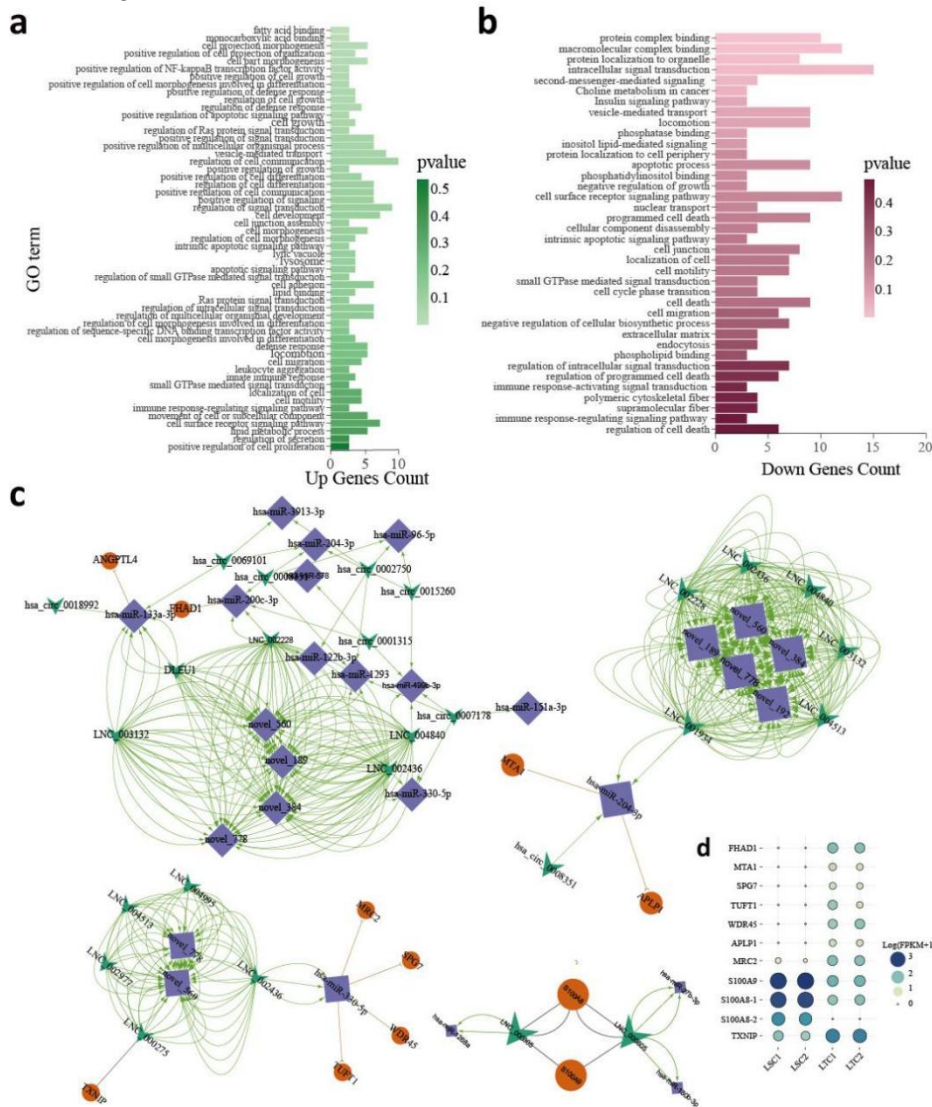
Figure 5. GO analysis and MCODE analysis. (a) GO analysis of upregulated gene sets comprising the ceRNA networks; (b) GO analysis of downregulated gene sets comprising the ceRNA networks.; (c) Four main sub-networks; (d) Heatmap of 11 transcripts in the four sub-networks.

A clustering heatmap was developed for all 11 mRNAs in the four modules (Figure 5d). Among these genes, two S100A8 transcripts and one S100A9 transcript were significantly upregulated and all were regulated by the same two novel lncRNAs (LNC_000305 and LNC_000605). We then listed all genes/RNAs in the four hub sub-networks based on the results of the Cytoscape plug-in MCODE. According to the location of the novel molecules and using the gene symbols of known molecules, a survival analysis was performed including all the molecules obtained in the four hub modules in lung adenocarcinoma patients using the UALCAN website and the lnCAR website. Then, based on the survival analysis from the UALCAN website and the lnCAR website, the survival curves of a novel lncRNA and the mRNAs, LNC_00236 and TUFT1, which belonged to the same sub-module were constructed (Figure 5c; Figure 6a–c). The rest of the molecules from the same sub-module are unknown, according to the UALCAN website and the lnCAR website. The three curves revealed that expression of TUFT1 and LNC_00236 was associated

with the overall survival of patients with LUAD. In addition, LNC_00236 was highly ex-
pressed in 40 patients with LUAD and may represent a potential prognostic biomarker.
Furthermore, the remaining novel lncRNAs may represent a potential resource for future
research into prognostic biomarkers (Figure 6d).
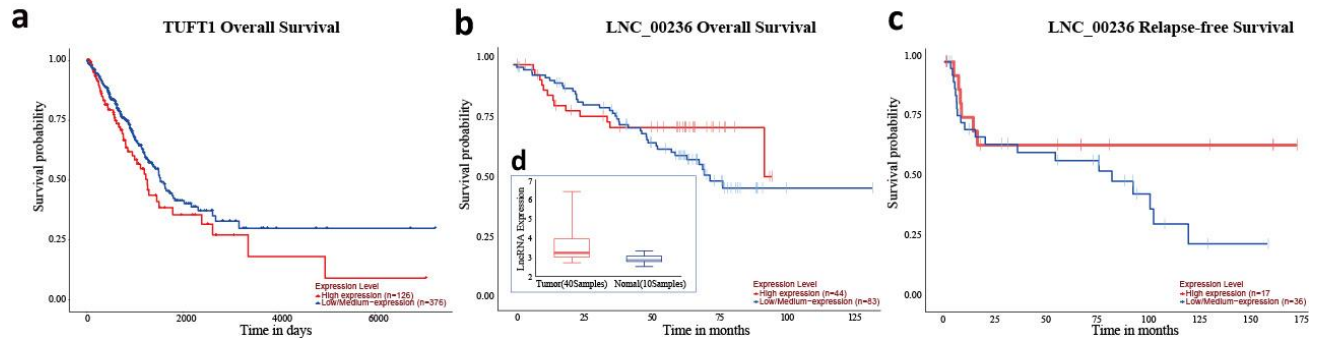
Figure 6. Survival analysis of patients with LUAD. Overall survival analysis stratifying patients by
median expression of (a) TUFT1; and (b) LNC_00236; (c) Relapse-free survival analysis of patients
stratified by LNC_00236 expression; (d)LNC_00236 expression atlas in tumor and normal tissue.

4. Conclusions

Our study identified novel miRNAs, lncRNAs, and circRNAs, which represent an
essential and crucial resource in understanding the underlying biology of LUAD. Specif-
ically, the study provided a detailed dissection of the constituents and functional proper-
ties of differentially expressed miRNAs, mRNAs, lncRNAs, and circRNAs between
ALDH+ H1975 LSCs and ALDH- H1975 LTCs. Based on these data, an interaction network
of differentially expressed miRNAs, mRNAs, and lncRNAs/circRNAs was constructed to
investigate the potential novel regulatory mechanism. The findings of this meta-analysis
provide comprehensive and systematic information on the underlying hallmarks distin-
guishing LSCs and LTCs, and the pathogenesis of LUAD from the perspective of lung
CSCs. This study also provides a rationale for further experimental validation of these
regulatory mechanisms and their role in the prognosis of LUAD and as therapeutic bi-
omarkers of LUAD.

5. Discussion

CSCs, as a rare cell population of cancers, display a broad spectrum of functional
heterogeneity[12]. They exhibit similar properties as normal stem cells, such as self-re-
newal, asymmetric cell division, and evasion of apoptosis[14], which enable them to initi-
ate cancers, promote cancer growth, metastasis, and relapse[6]. Experimental evidence
from previous studies[10] and our most recent study[11] indicated that ALDH+ H1975
LSCs are the key CSCs that proliferate extensively and form new tumors, while the re-
maining cancer cells lack this ability. The activities of LSCs are diverse[16]. For example,
exosomes from LSCs are associated with cell proliferation, migration, adhesion, and cell-
cell communication[11]. Furthermore, several pathways are associated with cell cycle ki-
netics, DNA repair, and mRNA expression of multi-drug-resistance genes in CSCs that
promote oncogenic resistance to chemotherapy and drugs[75]. In addition, steroid bio-
synthesis of T cells induced from the microenvironment in CSCs and resident cancer cells
may lead to immune evasion[66]. The above functions of LSCs may be broadly regulated
by ncRNAs, such as miRNAs[21], lncRNAs, and circRNAs[27]. Furthermore, identifica-
tion of known and unknown biomarkers of ncRNAs and their systematic and comprehen-
sive interrelation with unknown ncRNAs and mRNAs may also be necessary to provide

insights into the function of LSCs. Herein, we provide, for the first time, the ncRNA expression profile, including known and unknown miRNAs, lncRNAs, and circRNAs of LSCs and LTCs.

The gene expression profile of LSCs reflects the properties of LSCs that drive tumorigenesis. The regulatory mechanisms of mRNAs are associated with ncRNAs[27]. Therefore, a meta-analysis on ncRNAs and their relative regulated mRNAs is needed to enhance understanding of the underlying regulatory mechanisms of LSCs. In the current study, the detailed constituents and functional properties of LSCs were dissected from ceRNAs, GO enrichment, and KEGG pathway analyses. GO enrichment based on all DE mRNAs confirmed that (a) clusters of GO terms based on upregulated gene sets were closely related to processes such as DNA repair, the mitotic cell cycle, cell morphogenesis involved in differentiation, cell junction, positive regulation of alpha-beta T cell activation, and steroid biosynthesis. These pathways were highly correlated with the function of LSCs in oncogenic resistance to chemotherapy and drugs[75] and immune evasion[66]. In addition, (b) clusters of GO terms based on downregulated gene sets were closely related to the commitment of cell fate, the positive regulation of programmed cell death, and cell adhesion. These pathways may play a vital role in cell migration and cell apoptosis. Likewise, LSCs and LTCs are cancer cells. Thus, they have the traits of cancer cells and may share common cancer-related signaling pathways, which is comparable with our results in the KEGG pathway analyses based on all DE mRNAs.

In ceRNAs and relative GO enrichment analysis, two predominant properties were uncovered. First, the GO terms of downregulated gene sets are clustered in cell-cell communication, such as cell junction, localization of the cell, cell motility, intracellular signal transduction, endocytosis, and the extracellular matrix, which are mechanisms that may be linked to cell communication, the tumor microenvironment, exosomes, and cancer metastatsis[73]. These GO terms were similar to our previous study, that is, communication between cells, such as LSCs-LTCs and LTCs-LTCs, and they may depend on exosomes secreted by neighboring cells, which ultimately lead to altered cell functions, such as cell migration and proliferation of cancer cells. Second, the GO terms of upregulated gene sets also clustered in fatty acid-binding, positive regulation of the defense response, Ras protein signal transduction, the defense response, and the innate immune response. These GO terms may influence the immune resistance of LSCs, such as immune evasion, and may eventually be associated with a poor prognosis, as suggested above. The remaining GO terms in the upregulated gene sets involved pathways such as cell growth, cell differentiation, cell junction assembly, cell morphogenesis, and cell proliferation, which are closely related to the stemness of LSCs[73]. The ncRNAs associated with these terms and pathways may play a vital role in regulating the relative expression of mRNAs. These findings will provide a guide for future experimental verification of key ncRNAs and the identification of efficient biomarkers for cancer therapy and diagnosis. The three mRNAs (two S100A8 transcripts and one S100A9 transcript) from MCODE analysis were associated with several regulatory functions, such as regulation of defense response, regulation of NF-kappaB transcription factor (TF) activity, and the Toll-like receptor signaling pathway. Furthermore, S100A9 was correlated with a poor prognosis in LUAD[11].

In summary, the novel ncRNA transcripts and the overall profile-based meta-analysis of our study provide a comprehensive description of the biological molecular composition and properties of LUAD CSCs. We have dissected what is known and unknown about LSCs to provide a therapeutic perspective on these findings.

Supplementary Materials: The supporting information includes Figure S1: First base bias and position base bias of novel miRNAs; Table S1: Detailed information of miRNAs; Table S2: Novel lncRNAs; Table S3: Novel miRNAs; Table S4: Novel circRNAs; Table S5: Information on ceRNA-based molecular interactions and pairs.

Author Contributions: Conceptualization, Wei Yang and Yong Liang; Data curation, Wei Yang and Yong Liang; Formal analysis, Wei Yang; Funding acquisition, Xiao Yang and Fu Li; Investigation, Xiao Yang; Methodology, Wei Yang; Project administration, Wei Yang and Xiao Yang; Resources, Yuan Zheng and Hai Luo; Software, Wei Yang and Yong Liang; Supervision, Xiao Yang and Fu Li; Visualization, Wei Yang; Writing – original draft, Wei Yang; Writing – review & editing, Wei Yang, Yong Liang and Fu Li.

Funding: This work was funded by the Science and Technology Project of Shenzhen [grant numbers: ZDSYS20190902093203727, JCYJ20170413092711058, GJHZ20170310090257380] and by the China Postdoctoral Science Foundation [grant number: 2019M66336].

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are openly available in the Sequence Read Archive of NCBI at https://dataview.ncbi.nlm.nih.gov/object/PRJNA675697, reference number **PRJNA675697**, and this contains the sequencing reads of all replicates in the fastq format.

Acknowledgments: Not applicable.

Conflicts of Interest: Author Yong Liang is employed by the Shenzhen Xbiome Biotech Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

ncRNAs, Non-Coding RNAs
CSCs, Cancer Stem Cells
miRNAs, Micro RNAs
lncRNAs, Long Non-Coding RNAs
circRNAs, Circular RNAs
mRNAs, Messenger RNAs
LSCs, Lung Adenocarcinoma ALDH+ H1975 Stem Cells
LTCs, Lung Adenocarcinoma ALDH- H1975 Tumor Cells
LUAD, Lung Adenocarcinoma
ceRNA, Competitive Endogenous RNA
NSCLC, Non-Small Cell Lung Cancer
UTR, Untranslated Region
TMP, Transcripts Per Million
FPKM, Fragments Per Kilobase of exon model per Million mapped fragments
PCC, Pair Chiastic Clipping
PEM, Pair End Mapping
DP, Dynamic Programming Alignment
KEGG, Kyoto Encyclopedia of Gene and Genome
GO, Gene Ontology
Padj, Adjusted *P*-value

Appendix B (ORCID identifiers)

0000-0003-2138-5563 (Wei Yang)
0000-0002-3702-3113 (Yong Liang)
0000-0003-3671-7786 (Haitao Luo)
0000-0001-6255-5181 (Xiaofei Yang)
0000-0002-0606-8861 (Furong Li)

Uncategorized References

1. Yang, W.; Liang, Y.; Zheng, Y.Y.; Luo, H.T.; Yang, X.F.; Li, F.R. Identification of Novel Multi-Omics Expression Landscapes and Meta-Analysis of Landscape-Based Competitive Endogenous RNA Networks in ALDH+ Lung Adenocarcinoma Stem Cells. **2022**.

2. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians* **2021**, *71*, 209-249.

3. Ferlay, J.; Colombet, M.; Soerjomataram, I.; Parkin, D.M.; Piñeros, M.; Znaor, A.; Bray, F. Cancer statistics for the year 2020: An overview. *International Journal of Cancer* **2021**.

4. Tang, X.; Cheng, L.; Li, G.; Yan, Y.-M.; Su, F.; Huang, D.-L.; Zhang, S.; Liu, Z.; Qian, M.; Li, J. A small-molecule compound D6 overcomes EGFR-T790M-mediated resistance in non-small cell lung cancer. *Communications biology* **2021**, *4*, 1-13.

5. de Bruin, E.C.; McGranahan, N.; Mitter, R.; Salm, M.; Wedge, D.C.; Yates, L.; Jamal-Hanjani, M.; Shafi, S.; Murugaesu, N.; Rowan, A.J. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. *Science* **2014**, *346*, 251-256.

6. Nassar, D.; Blanpain, C. Cancer Stem Cells: Basic Concepts and Therapeutic Implications. *Annu Rev Pathol* **2016**, *11*, 47-76, doi:10.1146/annurev-pathol-012615-044438.

7. Morrison, R.; Schleicher, S.M.; Sun, Y.; Niermann, K.J.; Kim, S.; Spratt, D.E.; Chung, C.H.; Lu, B. Targeting the mechanisms of resistance to chemotherapy and radiotherapy with the cancer stem cell hypothesis. *J Oncol* **2011**, *2011*, 941876, doi:10.1155/2011/941876.

8. Chen, W.-J.; Ho, C.-C.; Chang, Y.-L.; Chen, H.-Y.; Lin, C.-A.; Ling, T.-Y.; Yu, S.-L.; Yuan, S.-S.; Chen, Y.-J.L.; Lin, C.-Y. Cancer-associated fibroblasts regulate the plasticity of lung cancer stemness via paracrine signalling. *Nature communications* **2014**, *5*, 1-17.

9. Mitra, S.; Stemke-Hale, K.; Mills, G.B.; Claerhout, S. Interactions between tumor cells and microenvironment in breast cancer: a new opportunity for targeted therapy. *Cancer science* **2012**, *103*, 400-407.

10. Zhang, W.C.; Shyh-Chang, N.; Yang, H.; Rai, A.; Umashankar, S.; Ma, S.; Soh, B.S.; Sun, L.L.; Tai, B.C.; Nga, M.E. Glycine decarboxylase activity drives non-small cell lung cancer tumor-initiating cells and tumorigenesis. *Cell* **2012**, *148*, 259-272.

11. Luo, H.-T.; Zheng, Y.-Y.; Tang, J.; Shao, L.-J.; Mao, Y.-H.; Yang, W.; Yang, X.-F.; Li, Y.; Tian, R.-J.; Li, F.-R. Dissecting the multi-omics atlas of the exosomes released by human lung adenocarcinoma stem-like cells. *NPJ Genomic Medicine* **2021**, *6*, 1-14.

12. Eramo, A.; Haas, T.; De Maria, R. Lung cancer stem cells: tools and targets to fight lung cancer. *Oncogene* **2010**, *29*, 4625-4635.

13. Ginestier, C.; Hur, M.H.; Charafe-Jauffret, E.; Monville, F.; Dutcher, J.; Brown, M.; Jacquemier, J.; Viens, P.; Kleer, C.G.; Liu, S. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell stem cell* **2007**, *1*, 555-567.

14. Reya, T.; Morrison, S.J.; Clarke, M.F.; Weissman, I.L. Stem cells, cancer, and cancer stem cells. *nature* **2001**, *414*, 105-111.

15. Eramo, A.; Haas, T.L.; De Maria, R. Lung cancer stem cells: tools and targets to fight lung cancer. *Oncogene* **2010**, *29*, 4625-4635, doi:10.1038/onc.2010.207.

16. He, X.; Qi, Y.; Zhang, X.; Liu, X.; Li, X.; Li, S.; Wu, Y.; Zhang, Q. Current landscape of tumor-derived exosomal ncRNAs in glioma progression, detection, and drug resistance. *Cell death & disease* **2021**, *12*, 1-11.

17. Dalmay, T.; Edwards, D. MicroRNAs and the hallmarks of cancer. *Oncogene* **2006**, *25*, 6170-6175.

18. Correia de Sousa, M.; Gjorgjieva, M.; Dolicka, D.; Sobolewski, C.; Foti, M. Deciphering miRNAs' Action through miRNA Editing. *Int J Mol Sci* **2019**, *20*, doi:10.3390/ijms20246249.

19. Esquela-Kerscher, A.; Slack, F.J. Oncomirs—microRNAs with a role in cancer. *Nature reviews cancer* **2006**, *6*, 259-269.

20. Tai, M.C.; Yanagisawa, K.; Nakatochi, M.; Hotta, N.; Hosono, Y.; Kawaguchi, K.; Naito, M.; Taniguchi, H.; Wakai, K.; Yokoi, K. Blood-borne miRNA profile-based diagnostic classifier for lung adenocarcinoma. *Scientific reports* **2016**, *6*, 1-9.

21. Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *cell* **2004**, *116*, 281-297.

22. Hayashita, Y.; Osada, H.; Tatematsu, Y.; Yamada, H.; Yanagisawa, K.; Tomida, S.; Yatabe, Y.; Kawahara, K.; Sekido, Y.; Takahashi, T. A polycistronic microRNA cluster, miR-17-92, is overexpressed in human lung cancers and enhances cell proliferation. *Cancer research* **2005**, *65*, 9628-9632.

23. Yanaihara, N.; Caplen, N.; Bowman, E.; Seike, M.; Kumamoto, K.; Yi, M.; Stephens, R.M.; Okamoto, A.; Yokota, J.; Tanaka, T. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. *Cancer cell* **2006**, *9*, 189-198.

24. Kopp, F.; Mendell, J.T. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **2018**, *172*, 393-407.

25. Nie, J.-H.; Li, T.-X.; Zhang, X.-Q.; Liu, J. Roles of non-coding RNAs in normal human brain development, brain tumor, and neuropsychiatric disorders. *Non-coding RNA* **2019**, *5*, 36.

26. Deng, X.; Xiong, W.; Jiang, X.; Zhang, S.; Li, Z.; Zhou, Y.; Xiang, B.; Zhou, M.; Li, X.; Li, G. LncRNA LINC00472 regulates cell stiffness and inhibits the migration and invasion of lung adenocarcinoma by binding to YBX1. *Cell death & disease* **2020**, *11*, 1-13.

27. Wu, P.; Mo, Y.; Peng, M.; Tang, T.; Zhong, Y.; Deng, X.; Xiong, F.; Guo, C.; Wu, X.; Li, Y. Emerging role of tumor-related functional peptides encoded by lncRNA and circRNA. *Molecular cancer* **2020**, *19*, 1-14.

28. Li, X.; Yang, L.; Chen, L.-L. The biogenesis, functions, and challenges of circular RNAs. *Molecular cell* **2018**, *71*, 428-442.

29. Wang, Y.; Liu, J.; Ma, J.; Sun, T.; Zhou, Q.; Wang, W.; Wang, G.; Wu, P.; Wang, H.; Jiang, L. Exosomal circR-NAs: biogenesis, effect and application in human diseases. *Molecular cancer* **2019**, *18*, 1-10.

30. Wang, J.; Zhao, X.; Wang, Y.; Ren, F.; Sun, D.; Yan, Y.; Kong, X.; Bu, J.; Liu, M.; Xu, S. circRNA-002178 act as a ceRNA to promote PDL1/PD1 expression in lung adenocarcinoma. *Cell death & disease* **2020**, *11*, 1-11.

31. Chen, W.; Li, W.; Liu, Z.; Ma, G.; Deng, Y.; Li, X.; Wang, Z. Comprehensive analysis of competitive endogenous RNA associated with immune infiltration in lung adenocarcinoma. *Scientific Reports* **2021**, *11*, 1-14.

32. Wang, K.; Long, B.; Liu, F.; Wang, J.-X.; Liu, C.-Y.; Zhao, B.; Zhou, L.-Y.; Sun, T.; Wang, M.; Yu, T. A circular RNA protects the heart from pathological hypertrophy and heart failure by targeting miR-223. *European heart journal* **2016**, *37*, 2602-2611.

33. Wang, K.; Liu, F.; Zhou, L.-Y.; Long, B.; Yuan, S.-M.; Wang, Y.; Liu, C.-Y.; Sun, T.; Zhang, X.-J.; Li, P.-F. The long noncoding RNA CHRF regulates cardiac hypertrophy by targeting miR-489. *Circulation research* **2014**, *114*, 1377-1388.

34. Wang, K.; Liu, C.-Y.; Zhou, L.-Y.; Wang, J.-X.; Wang, M.; Zhao, B.; Zhao, W.-K.; Xu, S.-J.; Fan, L.-H.; Zhang, X.-J. APF lncRNA regulates autophagy and myocardial infarction by targeting miR-188-3p. *Nature communications* **2015**, *6*, 1-11.

35. He, W.; Zhao, S.; Liu, X.; Dong, S.; Lv, J.; Liu, D.; Wang, J.; Meng, Z. ReSeqTools: an integrated toolkit for large-scale next-generation sequencing based resequencing analysis. *Genet Mol Res* **2013**, *12*, 6275-6283, doi:10.4238/2013.December.4.15.

36. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **2009**, *10*, 1-10.

37. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **2019**, *37*, 907-915.

38. He, W.; Zhao, S.; Liu, X.; Dong, S.; Lv, J.; Liu, D.; Wang, J.; Meng, Z. ReSeqTools: an integrated toolkit for large-scale next-generation sequencing based resequencing analysis. *Genet Mol Res* **2013**, *12*, 6275-6283.

39. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078-2079.

40. Kozomara, A.; Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research* **2014**, *42*, D68-D73.

41. Friedländer, M.R.; Mackowiak, S.D.; Li, N.; Chen, W.; Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic acids research* **2012**, *40*, 37-52.

42. Saha, S.; Bridges, S.; Magbanua, Z.V.; Peterson, D.G. Computational approaches and tools used in identification of dispersed repetitive DNA sequences. *Tropical Plant Biology* **2008**, *1*, 85-96.

43. Kalvari, I.; Nawrocki, E.P.; Ontiveros-Palacios, N.; Argasinska, J.; Lamkiewicz, K.; Marz, M.; Griffiths-Jones, S.; Toffano-Nioche, C.; Gautheret, D.; Weinberg, Z. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research* **2021**, *49*, D192-D200.

44. Saha, S.; Bridges, S.; Magbanua, Z.V.; Peterson, D.G. Empirical comparison of ab initio repeat finding programs. *Nucleic acids research* **2008**, *36*, 2284-2294.

45. Wen, M.; Shen, Y.; Shi, S.; Tang, T. miREvo: an integrative microRNA evolutionary analysis platform for next-generation sequencing experiments. *BMC bioinformatics* **2012**, *13*, 1-10.

46. Li, B.; Ruotti, V.; Stewart, R.M.; Thomson, J.A.; Dewey, C.N. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics* **2010**, *26*, 493-500.

47. Pertea, M.; Kim, D.; Pertea, G.M.; Leek, J.T.; Salzberg, S.L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols* **2016**, *11*, 1650-1667.

48. Sun, L.; Luo, H.; Bu, D.; Zhao, G.; Yu, K.; Zhang, C.; Liu, Y.; Chen, R.; Zhao, Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research* **2013**, *41*, e166-e166, doi:10.1093/nar/gkt646.

49. Kang, Y.-J.; Yang, D.-C.; Kong, L.; Hou, M.; Meng, Y.-Q.; Wei, L.; Gao, G. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic acids research* **2017**, *45*, W12-W16.

50. Punta, M.; Coggill, P.C.; Eberhardt, R.Y.; Mistry, J.; Tate, J.; Boursnell, C.; Pang, N.; Forslund, K.; Ceric, G.; Clements, J. The Pfam protein families database. *Nucleic acids research* **2012**, *40*, D290-D301.

51. Lin, M.F.; Jungreis, I.; Kellis, M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **2011**, *27*, i275-i282.

52. Love, M.I.; Hogenesch, J.B.; Irizarry, R.A. Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nature biotechnology* **2016**, *34*, 1287-1291.

53. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; Van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **2010**, *28*, 511-515.

54. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **2014**, *15*, 1-21.

55. Storey, J.D. The positive false discovery rate: a Bayesian interpretation and the q-value. *The Annals of Statistics* **2003**, *31*, 2013-2035.

56. Huang, D.W.; Sherman, B.T.; Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **2009**, *37*, 1-13.

57. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A.H.; Tanaseichuk, O.; Benner, C.; Chanda, S.K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature communications* **2019**, *10*, 1-10.

58. Enright, A.; John, B.; Gaul, U.; Tuschl, T.; Sander, C.; Marks, D. MicroRNA targets in Drosophila. *Genome biology* **2003**, *4*, 1-27.

59. Kertesz, M.; Iovino, N.; Unnerstall, U.; Gaul, U.; Segal, E. The role of site accessibility in microRNA target recognition. *Nature genetics* **2007**, *39*, 1278-1284.

60. Krüger, J.; Rehmsmeier, M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic acids research* **2006**, *34*, W451-W454.

61. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **2003**, *13*, 2498-2504.

62. Bader, G.D.; Hogue, C.W. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics* **2003**, *4*, 1-27.

63. Chandrashekar, D.S.; Bashel, B.; Balasubramanya, S.A.H.; Creighton, C.J.; Ponce-Rodriguez, I.; Chakravarthi, B.V.; Varambally, S. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. *Neoplasia* **2017**, *19*, 649-658.

64. Zheng, Y.; Xu, Q.; Liu, M.; Hu, H.; Xie, Y.; Zuo, Z.; Ren, J. lnCAR: a comprehensive resource for lncRNAs from cancer arrays. *Cancer research* **2019**, *79*, 2076-2083.

65. Manns, A.; Wilks, R.J.; Murphy, E.L.; Haynes, G.; Figueroa, J.P.; Barnett, M.; Hanchard, B.; Blattner, W.A. A prospective study of transmission by transfusion of HTLV-I and risk factors associated with seroconversion. *International journal of cancer* **1992**, *51*, 886-891.

66. Mahata, B.; Pramanik, J.; van der Weyden, L.; Polanski, K.; Kar, G.; Riedel, A.; Chen, X.; Fonseca, N.A.; Kundu, K.; Campos, L.S. Tumors induce de novo steroid biosynthesis in T cells to evade immunity. *Nature communications* **2020**, *11*, 1-15.

67. Luo, Y.; Wu, Y.; Peng, Y.; Liu, X.; Bie, J.; Li, S. Systematic analysis to identify a key role of CDK1 in mediating gene interaction networks in cervical cancer development. *Irish Journal of Medical Science (1971-)* **2016**, *185*, 231-239.

68. Huang, J.; Wu, S.; Barrera, J.; Matthews, K.; Pan, D. The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the Drosophila Homolog of YAP. *Cell* **2005**, *122*, 421-434.

69. Knights, A.J.; Funnell, A.P.; Crossley, M.; Pearson, R.C. Holding tight: cell junctions and cancer spread. *Trends in cancer research* **2012**, *8*, 61.

70. Staudt, L.M. Oncogenic activation of NF-kappaB. *Cold Spring Harb Perspect Biol* **2010**, *2*, a000109, doi:10.1101/cshperspect.a000109.

71. Adli, M.; Baldwin, A.S. IKK-i/IKKε controls constitutive, cancer cell-associated NF-κB activity via regulation of Ser-536 p65/RelA phosphorylation. *Journal of Biological Chemistry* **2006**, *281*, 26976-26984.

72.   Li, J.-H.; Liu, S.; Zhou, H.; Qu, L.-H.; Yang, J.-H. 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-D97.

73.   El-Sabban, M.; Pauli, B. Role of cell adhesion molecules in endothelial cell/tumor cell communication and tumor cell extravasation. In *Progress in Cell Research*; Elsevier: 1995; Volume 4, pp. 87-92.

74.   Hanahan, D.; Weinberg, R.A. Hallmarks of cancer: the next generation. *cell* **2011**, *144*, 646-674.

75.   Perona, R.; López-Ayllón, B.D.; de Castro Carpeño, J.; Belda-Iniesta, C. A role for cancer stem cells in drug resistance and metastasis in non-small-cell lung cancer. *Clinical and Translational Oncology* **2011**, *13*, 289-293.