Computational and Structural Biotechnology Journal 21 (2023) 535–549

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Integrative network analysis reveals subtype-specific long non-coding RNA regulatory mechanisms in head and neck squamous cell carcinoma

Jiang Li [a,b], Tan Wu [a,b], Kai Song [a], Lina Zhu [a], Yijuan Wang [a], Tingting Chen [a], Xin Wang [a,b,c,*]

[a] Department of Surgery, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China
[b] Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China
[c] Shenzhen Research Institute, City University of Hong Kong, Hong Kong SAR, China

ARTICLE INFO

Article history:
Received 23 August 2022
Accepted 19 December 2022
Available online 21 December 2022

Keywords:
Long non-coding RNAs
Head and neck cancer
Network inference
Subtype-specific
Competitive endogenous RNA

ABSTRACT

Head and neck squamous cell carcinoma (HNSC) is one of most common malignancies with high mortality worldwide. Importantly, the molecular heterogeneity of HNSC complicates the clinical diagnosis and treatment, leading to poor overall survival outcomes. To dissect the complex heterogeneity, recent studies have reported multiple molecular subtyping systems. For instance, HNSC can be subdivided to four distinct molecular subtypes: atypical, basal, classical, and mesenchymal, of which the mesenchymal subtype is characterized by upregulated epithelial-mesenchymal transition (EMT) and associated with poorer survival outcomes. Despite a wealth of studies into the complex molecular heterogeneity, the regulatory mechanism specific to this aggressive subtype remain largely unclear. Herein, we developed a network-based bioinformatics framework that integrates lncRNA and mRNA expression profiles to elucidate the subtype-specific regulatory mechanisms. Applying the framework to HNSC, we identified a clinically relevant lncRNA *LNCOG* as a key master regulator mediating EMT underlying the mesenchymal subtype. Five genes with strong prognostic values, namely *ANXA5, ITGA5, CCBE1, P4HA2,* and *EPHX3*, were predicted to be the putative targets of *LNCOG* and subsequently validated in other independent datasets. By integrative analysis of the miRNA expression profiles, we found that *LNCOG* may act as a ceRNA to sponge miR-148a-3p thereby upregulating *ITGA5* to promote HNSC progression. Furthermore, our drug sensitivity analysis demonstrated that the five putative targets of *LNCOG* were also predictive of the sensitivities of multiple FDA-approved drugs. In summary, our bioinformatics framework facilitates the dissection of cancer subtype-specific lncRNA regulatory mechanisms, providing potential novel biomarkers for more optimized treatment of HNSC.

1. Introduction

Head and neck squamous cell carcinoma (HNSC), including cancers developed in the oral cavity, salivary glands, and oropharynx, accounts for roughly 5% of all cancer types and is the sixth leading cause of cancer-related deaths worldwide [1,2]. The last decades have witnessed tremendous improvements in surgery, chemotherapies and radiation therapies for this malignancy, leading to an increase of 5-year survival rate to ~65% in the United States [3]. However, most patients are still diagnosed at advanced stages, missing the critical treatment timing and making it hard to perform surgical operations due to the requirements of inalienable life-essential behaviours like speaking and swallowing [4]. Even after tumour removal and clinical treatments, patients are 40–60% likely to suffer from the recurrence of HNSC [5]. Understanding the molecular mechanisms underlying HNSC progression and metastasis is of great

Abbreviations: HNSC, head and neck squamous cell carcinoma; EMT, epithelial-mesenchymal transition; lncRNA, long non-coding RNA; miRNA, microRNA; ceRNA, the competitive endogenous RNA; TPM, transcripts per million; FPKM, fragments per kilobase million; UCSC, the University of California Santa Cruz; TCGA, The Cancer Genome Atlas; ICGC, The International Cancer Genome Consortium; GEO, Gene Expression Omnibus; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; CTRP, The Cancer Therapeutics Response Portal; BH, Benjamini-Hochberg; GSEA, gene set enrichment analysis; AUC, area under the curve; ROC, receiver operating characteristic curve; LASSO, least absolute shrinkage and selection operator; HR, hazard ratio; CI, confidence interval; DEG, differentially expressed gene; OS, overall survival; DFS, disease-free survival; DEX, dexamethasone

* Correspondence to: Department of Surgery, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong SAR, China.
E-mail address: xwang@surgery.cuhk.edu.hk (X. Wang).

https://doi.org/10.1016/j.csbj.2022.12.030

urgency for the development of more precise diagnostic tools and therapeutic treatments.

Similar to other major malignancies, HNSCs are highly heterogeneous disease entities. To dissect the molecular heterogeneity of the disease, several subtyping systems have been developed [6–8]. A representative study was performed by The Cancer Genome Atlas Network (TCGA), which comprehensively characterized four distinct transcriptome-based subtypes: atypical, basal, classical, and mesenchymal, with distinct biological features and clinical outcomes [9]. Importantly, the mesenchymal subtype was characterized by upregulated epithelial-mesenchymal transition (EMT), higher invasiveness and higher risk of lymph node metastasis [10,11]. Moreover, survival analysis of these four subtypes confirmed that patients classified to the mesenchymal subtype was associated with the worst survival. However, the regulatory mechanism underlying the mesenchymal subtype remains elusive, preventing the development of targeted agents for this aggressive disease.

Recent studies have revealed the transcriptional and post-transcriptional importance of non-coding regions of the human genome, highlighting that non-coding RNAs are also key players in human cancers [12,13]. Long non-coding RNAs (lncRNAs), classified as transcripts over 200 bps with little coding potential, have been discovered using high-throughput sequencing [14]. Considered to be one of the fundamental regulators at transcriptional, post-transcriptional and epigenetic levels, lncRNAs have been reported for diverse regulatory functions and implicated for potential clinical values in cancers [14,15]. Four major lncRNA archetypes, namely signal, decoy, guide and scaffold, have been proposed to comprehensively summarize the molecular functions of lncRNAs [16]. First, as lncRNA transcriptions could occur at a specific time due to the transcriptional control, they may function as a molecular 'signal' to regulate downstream genes, thus influencing the transcriptional levels of targets [16]. Second, as a 'decoy' molecule, lncRNA could bind transcription regulators, by preventing them from locating into the destined position to intervene certain cancer-associated pathways [17]. Third, 'Guide' lncRNAs often interact with transcription factors, leading them to specific DNA regions and thus regulate gene translation [18]. Fourth, when lncRNAs exert their 'scaffold' property, they function as a platform to facilitate chromatin modifying complexes and may affect transcription depending on the biological nature of involved proteins [19].

Among the diverse molecular functions, growing interests have been focused on the role of lncRNAs as the competing endogenous RNA (ceRNA) to sponge miRNAs, thereby repressing the expression of target genes in various malignancies [20–22]. Of note, quite a few lncRNAs have been reported to play critical roles in HNSC progression and metastasis. For example, the overexpression of lncRNA *HOXC13-AS* could promote tumour migration and cell proliferation of HNSC by sponging miR-378 g, thus inducing the expression of *HOXC13* [23]. LncRNA *PTCSC3*, functioning as a ceRNA for miR-574–5p, was experimentally validated as a tumour suppressor in HNSC cancer cells [24]. LncRNA *MALAT1* induced EMT pathway, resulting in the tumour growth and invasiveness of HNSC [25]. Furthermore, two other lncRNAs, *TUG1* and *UCA1*, contributed to the tumour progression of HNSC by interacting with the Wnt/beta-catenin pathway [26,27].

To date, several computational methods have been reported to identify potential cancer-related and tumour-specific ceRNAs using gene expression profiles. Do et al. developed a pipeline called 'Cancerin' for the identification ceRNA modules enriched in cancer-associated biological processes by incorporating copy number alteration, DNA methylation, and transcription factor profiles between tumour and normal samples [28]. Paci et al. applied a partial correlation method to analyse lncRNA, miRNA and mRNA expression levels and identified miRNA-mediated sponge interaction networks in breast cancer and normal samples, respectively, highlighting rewiring of the ceRNA program from normal to cancerous states [29]. However, despite the rich ceRNA literature, there is no computational approach to infer lncRNA regulatory networks with the identification of master regulatory lncRNAs and their ceRNA candidates on a genome-wide scale for specific cancer subtypes.

In this study, we first employed a multi-dimensional network approach to dissect the lncRNA-mRNA regulatory mechanism underlying the mesenchymal subtype of HNSC, followed by the identification of master regulatory lncRNAs mediating the EMT pathway. Next, we predicted putative target genes of the key master regulator *LNCOG* and validated their clinical associations in multiple independent datasets. Subsequently, we comprehensively performed ceRNA analysis by integrative analysis of lncRNA, miRNA, and mRNA expression profiles. Our results showed that *LNCOG* might promote EMT pathway via sponging miR-148a-3p, thus antagonizing its function to repress *ITGA5* protein translation. Finally, using comprehensive drug sensitivity analysis, we found that the five putative targets of *LNCOG* were also predictive of the sensitivities of multiple drugs, of which some are FDA-approved.

2. Materials and methods

2.1. Data collection and pre-processing

A flow chart was produced to summarize the study design (Fig. 1). TCGA-HNSC gene expression profiles and corresponding clinical information were downloaded from the University of California Santa Cruz (UCSC) Xena data portal (https://xenabrowser.net/datapages). Gene expression profiles were converted from fragments per kilobase million (FPKM) to transcripts per million (TPM), followed by log2-transformation and filtering of genes with low expression. Human lncRNA annotation was downloaded from the GENCODE database (https://www.gencodegenes.org/). A total of 275 patients with subtype classification labels were downloaded from TCGA for downstream analysis, out of which 75 belong to the mesenchymal subtype. The stromal content of patients in the TCGA-HNSC cohort was obtained from the ESTIMATE database (https://bioinformatics.mdanderson.org/estimate). Independent validation datasets were downloaded from The International Cancer Genome Consortium (ICGC) [30] and Gene Expression Omnibus (GSE41613, GSE27020, and GSE65858). The detailed patient demographics for the datasets were shown in Table 1.

2.2. LncRNA regulatory network inference and master regulator identification

We performed regulatory network inference to dissect the relationship between lncRNAs and potential targets by integrative analysis of lncRNA and mRNA expression profiles. To this end, differential analysis was performed between the mesenchymal subtype and non-mesenchymal subtypes on the TCGA-HNSC dataset using the R package 'limma' [31]. Upregulated lncRNAs in the mesenchymal subtype (log2 fold change > 0.5, Benjamini-Hochberg adjusted [BH-adjusted] $P < 0.05$) were prioritized as potential regulators, and differentially expressed genes (DEGs) between the mesenchymal and non-mesenchymal tumours (|log2 fold change| > 0.5, BH-adjusted $P < 0.05$) were identified as potential targets. The lncRNA and mRNA expression profiles were Z-normalized and merged for

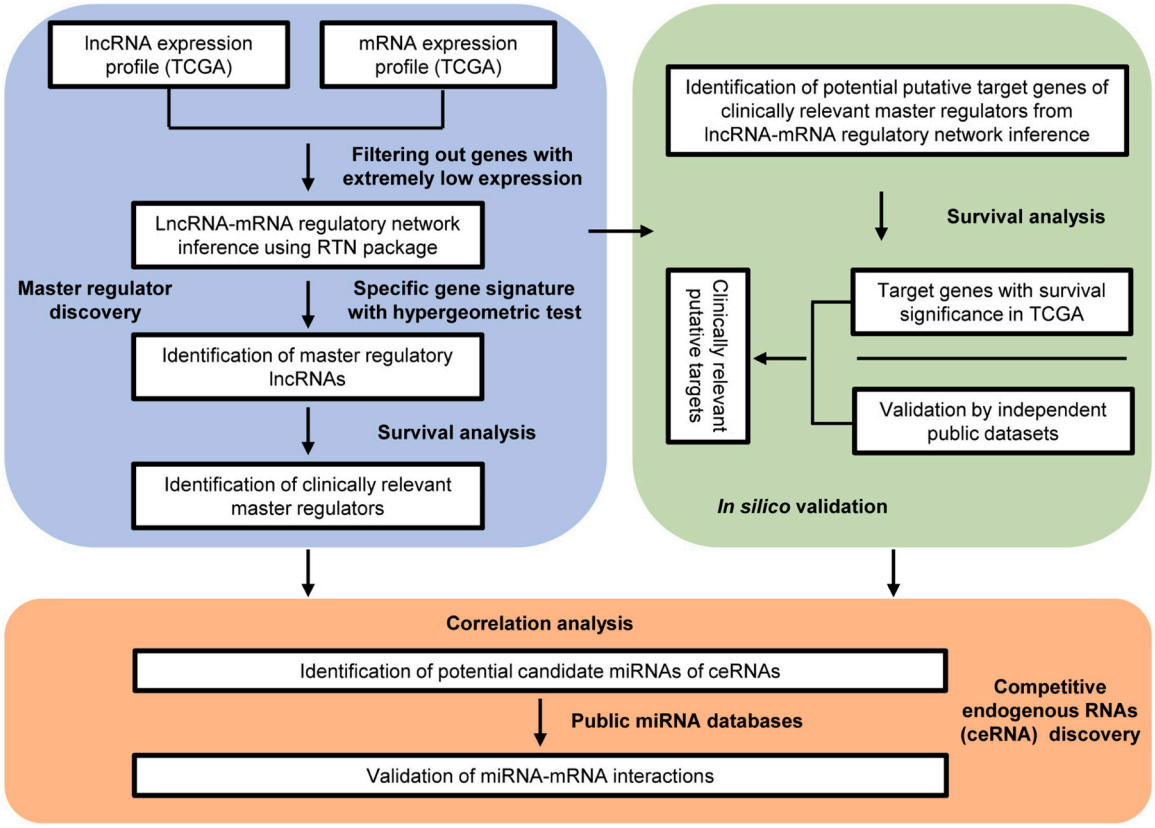


Fig. 1. A schematic workflow illustrating the overall study design of our bioinformatics framework. A lncRNA-mRNA regulatory network was constructed on the TCGA-HNSC dataset, followed by master regulator analysis and ceRNA analysis. Multiple independent datasets were employed to validate the clinical relevance of the putative target mRNAs.

Table 1
Patient characteristics of in the discovery and validation datasets.

	TCGA-HNSC (n = 275) No. (%)	ICGC-ORCA (n = 40) No. (%)	GSE65858 (n = 270) No. (%)	GSE41613 (n = 97) No. (%)	GSE27020 (n = 109) No. (%)
Age					
< 60	110 (0.40)	32 (0.80)	153 (0.57)	50 (0.52)	35 (0.32)
> = 60	165 (0.60)	8 (0.20)	117 (0.43)	47 (0.48)	74 (0.68)
Gender					
Male	200 (0.73)	34 (0.85)	223 (0.83)	66 (0.68)	–
Female	75 (0.27)	6 (0.15)	47 (0.17)	31 (0.32)	–
Stage T					
1	13 (0.05)	0 (0.00)	35 (0.13)	–	–
2	84 (0.31)	1 (0.03)	80 (0.30)	–	–
3	82 (0.30)	3 (0.08)	58 (0.21)	–	–
4	96 (0.35)	36 (0.90)	97 (0.36)	–	–
Stage N					
0	136 (0.49)	10 (0.25)	94 (0.35)	–	–
1	49 (0.18)	18 (0.45)	32 (0.12)	–	–
2	83 (0.30)	11 (0.28)	132 (0.49)	–	–
3	4 (0.01)	0 (0.00)	12 (0.04)	–	–
Others	3 (0.01)	1 (0.03)	0 (0.00)	–	–
Stage M					
0	273 (0.99)	–	263 (0.97)	–	–
1	2 (0.01)	–	7 (0.03)	–	–
Stage TNM					
I	9 (0.03)	–	18 (0.07)	–	–
II	57 (0.21)	–	37 (0.18)	–	–
III	65 (0.24)	–	37 (0.18)	–	–
IV	144 (0.52)	–	178 (0.66)	–	–
Subtype					
Atypical	67 (0.24)	–	73 (0.27)	–	–
Basal	85 (0.31)	–	84 (0.31)	–	–
Classical	48 (0.18)	–	30 (0.11)	–	–
Mesenchymal	75 (0.27)	–	83 (0.31)	–	–

regulatory network inference using R package 'RTN' [32], with permutation for 1000 times, bootstrap resampling and weak interaction filtering. Subsequently, master regulatory analysis was performed based on a hypergeometric test for overrepresentation of the EMT signature genes [33] in the predicted regulon of each lncRNA. The regulatory network was visualized using R package 'RedeR' [34].

2.3. Functional annotation analysis

We performed gene set enrichment analysis (GSEA) using R package 'HTSanalyzeR2' [35] with 1000 permutations. Log2 fold changes between gene expression profiles of the mesenchymal and non-mesenchymal tumour samples were used as input for each dataset, and C2 gene sets were downloaded from MSigDB [36]. For the analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and GO gene sets, we utilized functions enrichKEGG and enrichGO of the R package 'clusterProfiler' [37]. Overrepresentation of the mesenchymal subtype-specific DEGs and potential putative target genes regulated by identified master regulators in ten cancer hallmark gene sets [38] was quantified by one-sided hypergeometric tests and ranked by the statistical significance ($-\log_{10}(P)$).

2.4. ceRNA analysis

The mature miRNA expression profiles were downloaded using R package 'TCGABiolinks' from TCGA [39], followed by log2-transformation. Candidate miRNAs were selected if they were both negatively correlated with the master regulatory lncRNA and target genes (Pearson correlation coefficients or PCC < 0 & BH-adjusted P < 0.05). Experimentally validated miRNA-mRNA interactions were downloaded from miRTarBase database [40]. The predicted miRNAs with experimental validations were prioritized as the sponged candidates of the master regulator lncRNA. Univariate Cox regression analysis was performed on the selected miRNAs to evaluate the association with survival.

2.5. Drug sensitivity analysis

Drug activity data involving 481 approved and investigational drugs and corresponding gene expression profiles of The Cancer Therapeutics Response Portal (CTRP) were downloaded from the CellMinerCDB database [41]. More specifically, the CTRP drug activity data and corresponding gene expression levels of 32 HNSC cell lines were downloaded, and drug sensitivity analysis was performed with the following procedures. We first removed drugs without drug activity scores in over half of the HNSC cell lines and imputed the remaining by R package 'impute'. To investigate the predictive power of the lncRNA target genes, cell lines were stratified by the median value of the drug activity scores, and the area under the curve (AUC) was calculated based on the target gene expression. Moreover, a univariate linear regression model was constructed to predict drug activity scores based on gene expression levels, and ANOVA tests were performed to assess the overall model significance. A gene was significantly associated with the sensitivity of cell lines to a certain drug if ANOVA P < 0.05 and AUC > = 0.75 or significantly associated with the resistance if ANOVA P < 0.05 and AUC < = 0.25. To further investigate whether the putative targets of *LNCOG* are predictive of drug activities, we performed multivariate regression with LASSO regularization involving all the DEGs between the mesenchymal and non-mesenchymal subtypes for each drug. Genes with non-zero regression coefficients were selected and determined as contributing factors to the drug sensitivity.

2.6. Statistical analysis

Statistical analyses were performed using R (version 4.0.3). Kaplan–Meier curves were generated to investigate the prognostic values of the identified master regulators and the predicted target genes, with the significance assessed by log-rank tests. Univariate and multivariate Cox proportional hazard regression analyses were performed by R package 'survival' to calculate hazard ratios (HRs) with 95% confidence intervals (CIs) based on the gene expression levels. Survival plots were visualized by R package 'survminer'. LASSO regression analysis was performed by R package 'glmnet' to prioritize mesenchymal subtype-specific contributing genes to drug activity. Two-sided Student's *t*-tests were used to assess the differences between two groups. Statistical significance was denoted by $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, ns Not Significant, and a P < 0.05 was considered statistically significant.

3. Results

3.1. Integrative analysis of lncRNA and mRNA expression profiles established a regulatory network underlying the mesenchymal subtype of HNSC

Using the TCGA-HNSC dataset with corresponding subtype labels, we first performed biological and clinical characterizations of the mesenchymal subtype of HNSC. Consistent with previous conclusions, the mesenchymal subtype showed upregulated EMT signature genes, higher stromal content, and worse overall survival (Fig. S1A-**C**). Pathway analysis based on cancer hallmarks found that the differentially expressed genes (DEGs) between the mesenchymal and non-mesenchymal subtypes (|log2 fold change| > 0.5, BH-adjusted P < 0.05) were strongly associated with 'tumor promoting inflammation', 'angiogenesis', 'activating invasion' and 'sustaining proliferative signaling' (P < 0.05) (Fig. S1D). Consistently, using the abovementioned DEGs, functional annotations based on hypergeometric tests on KEGG pathways and GO biological processes found that the mesenchymal subtype specific DEGs were overrepresented in EMT-related pathways such as 'focal adhesion' and 'extracellular matrix organization' (Fig. S2A-**B**). Gene set enrichment analysis confirmed the significant enrichment of focal adhesion and EMT signature genes in the mesenchymal subtype (Fig. S2C-**D**). Together, our results confirmed previous characterizations of the mesenchymal subtype of HNSC, providing a strong rationale for us to further elucidate the regulatory mechanism underlying this aggressive subtype.

To interrogate the subtype-specific regulatory mechanism, we inferred a lncRNA-mRNA regulatory network by integrative analysis of the mRNA and lncRNA expression profiles in the TCGA-HNSC dataset (detailed method in Section 2.2). Briefly, 39 lncRNAs upregulated in the mesenchymal subtype (log2 fold change > 0.5, BH-adjusted P < 0.05) (Fig. 2**A**) were prioritized as potential regulators, while 2569 differentially expressed genes between the mesenchymal and non-mesenchymal tumours (|log2 fold change| > 0.5, BH-adjusted P < 0.05) were selected as the potential functional targets of the lncRNAs. As a result, we obtained a lncRNA-mRNA regulatory network with 2618 nodes and 6569 edges using RTN package [32], with 1000 times permutation, bootstrapping analysis and filtering out weak interactions (Fig. 2**B**). Both activation and repression relationships between lncRNAs and their putative targets could be observed in the network, suggesting the diverse regulatory roles of lncRNAs in HNSC.

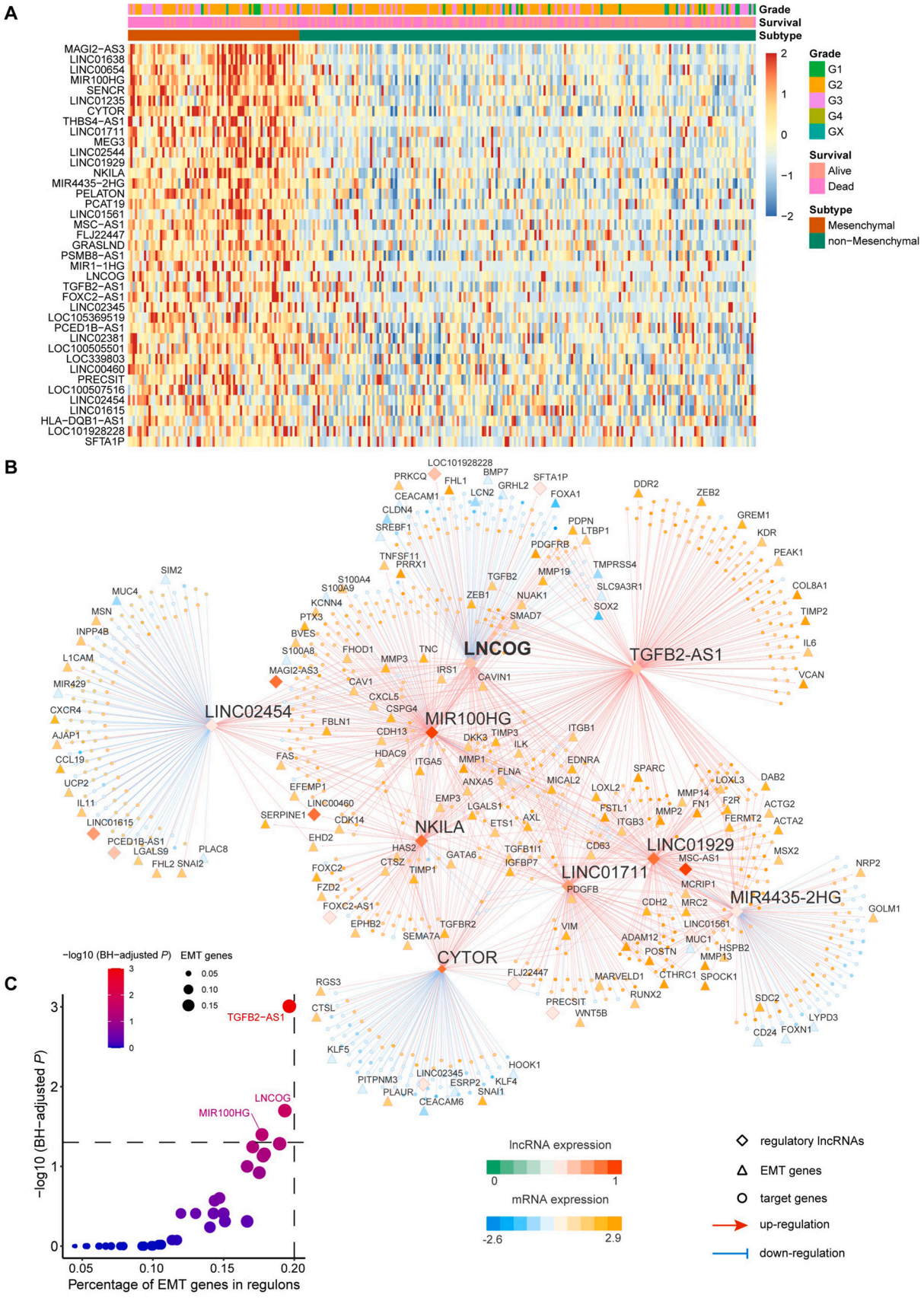


Fig. 2. Differentially expressed lncRNAs and mRNAs were integrated to infer a lncRNA regulatory network underlying the mesenchymal subtype of HNSC. (A) Heatmap of the expression of lncRNAs upregulated in the mesenchymal subtype in the TCGA-HNSC dataset. **(B)** A lncRNA-mRNA regulatory network of the prioritized lncRNAs and putative target mRNAs. **C.** Prioritization of master regulatory lncRNAs was based on statistical significance derived from hypergeometric tests for overrepresentation of a lncRNA's regulon for EMT signature genes and the proportion of EMT genes regulated by a lncRNA.

Table 2
Master regulator analysis results.

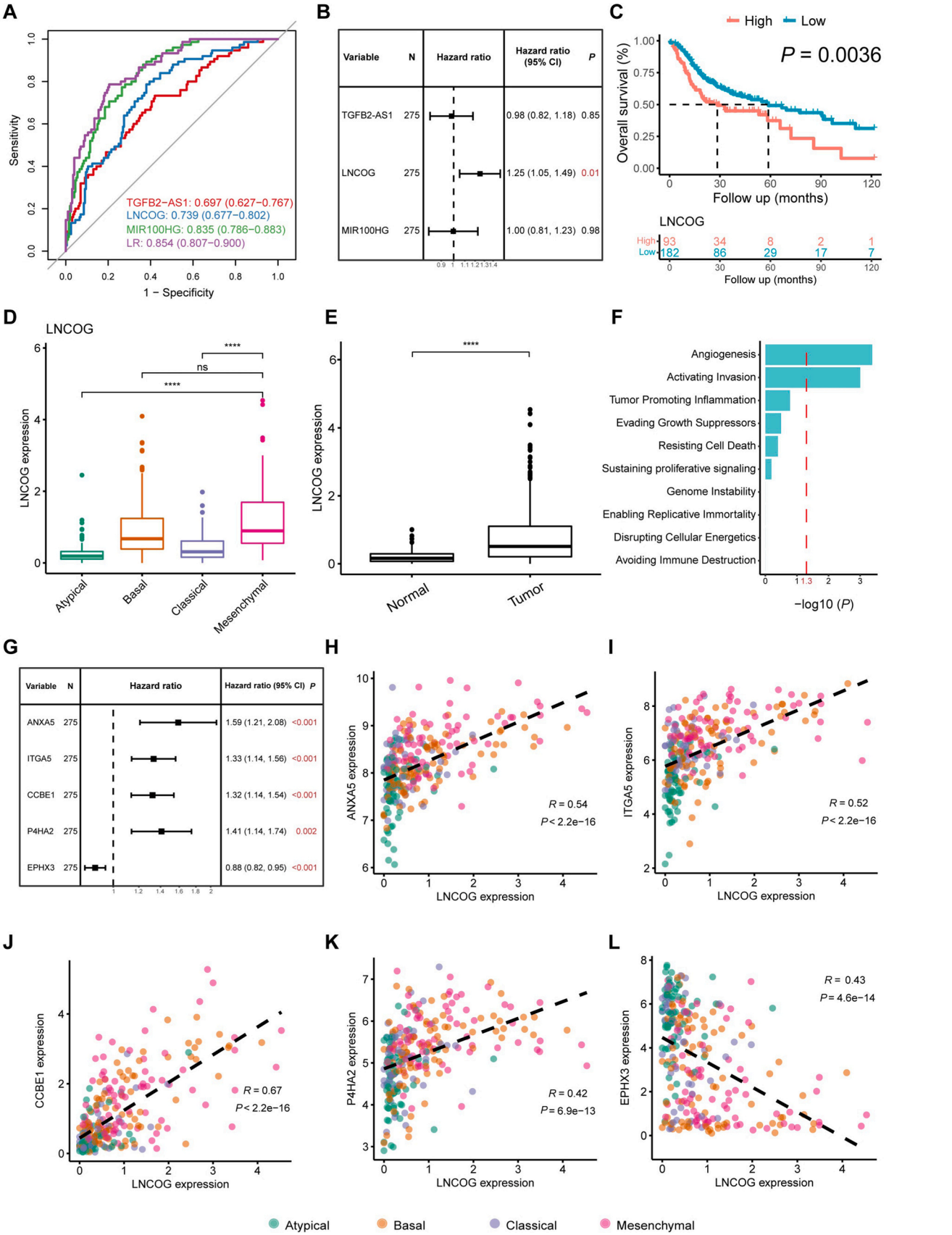
Regulon	Regulon size	Expected hits	Observed hits	p-value	BH-adjusted p-value
TGFB2-AS1	229	24.85	45	2.60E-05	9.80E-04
LNCOG	150	16.28	29	1.10E-03	2.00E-02
MIR100HG	175	18.99	31	3.20E-03	4.00E-02
LINC01929	116	12.59	22	5.50E-03	5.20E-02
LINC02454	170	18.45	29	7.40E-03	5.70E-02
LINC01711	123	13.35	22	1.10E-02	7.00E-02
MIR4435–2HG	118	12.8	21	1.40E-02	7.40E-02
CYTOR	138	14.97	23	2.10E-02	1.00E-01
NKILA	97	10.53	17	2.90E-02	1.20E-01
MSC-AS1	170	18.45	25	6.50E-02	2.50E-01
LINC01561	181	19.64	26	7.70E-02	2.70E-01
LINC02345	133	14.43	19	1.20E-01	3.90E-01
MAGI2-AS3	751	81.49	90	1.30E-01	3.90E-01
LINC01638	80	8.68	12	1.50E-01	3.90E-01
LOC101928228	238	25.83	31	1.50E-01	3.90E-01
FOXC2-AS1	53	5.75	8	2.10E-01	4.90E-01
GRASLND	30	3.26	5	2.20E-01	4.90E-01
LINC02544	57	6.19	8	2.70E-01	5.80E-01
LOC100505501	94	10.2	11	4.40E-01	8.40E-01
PRECSIT	149	16.17	17	4.50E-01	8.40E-01
LINC02381	123	13.35	14	4.70E-01	8.40E-01
LOC339803	97	10.53	11	4.90E-01	8.40E-01
PCAT19	85	9.22	9	5.90E-01	9.60E-01
LINC01615	48	5.21	5	6.10E-01	9.60E-01
LINC01235	80	8.68	8	6.50E-01	9.90E-01
LINC00460	111	12.04	11	6.70E-01	9.90E-01
SFTA1P	39	4.23	3	8.10E-01	1.00E+00
SENCR	331	35.92	31	8.50E-01	1.00E+00
PELATON	325	35.27	30	8.70E-01	1.00E+00
FLJ22447	90	9.77	6	9.40E-01	1.00E+00
HLA-DQB1-AS1	56	6.08	3	9.50E-01	1.00E+00
THBS4-AS1	145	15.73	9	9.80E-01	1.00E+00
LOC100507516	241	26.15	17	9.90E-01	1.00E+00
MIR1–1HG	244	26.48	11	1.00E+00	1.00E+00
PSMB8-AS1	191	20.73	10	1.00E+00	1.00E+00
LINC00654	308	33.42	19	1.00E+00	1.00E+00
PCED1B-AS1	632	68.58	50	1.00E+00	1.00E+00
LOC105369519	159	17.25	7	1.00E+00	1.00E+00

3.2. LNCOG is a master regulator of epithelial-mesenchymal transition and a strong prognosticator of HNSC patients

To identify lncRNAs regulating EMT in the mesenchymal subtype of HNSC, we performed master regulator analysis with hypergeometric tests for overrepresentation of a lncRNA's predicted regulons, i.e., a set of target genes, for EMT signature genes (Table 2). Three lncRNAs were prioritized to be potential master regulators of the EMT pathway, namely *TGFB2-AS1*, *LNCOG,* and *MIR100HG* (Fig. 2**C**). The expression levels of the three lncRNAs were predictive of HNSCs from the mesenchymal status, and a multivariate logistic regression model built by integrating the three lncRNAs also demonstrated promising diagnostic power, suggesting their potential to be used as diagnostic biomarkers for the prediction of the mesenchymal subtype of HNSC (Fig. 3**A**). We subsequently investigated the prognostic relevance of these lncRNAs and found that higher expression of *LNCOG* was significantly associated with worse overall survival (OS) based on univariate Cox regression analysis ($P = 0.01$, hazard ratio (HR) = 1.25 [95% CI, 1.05 – 1.49]) (Fig. 3**B**) and Kaplan-Meier analysis ($P = 0.0036$, log-rank test, Fig. 3**C**). Compared to non-mesenchymal tumours, we observed significantly higher expression of *LNCOG* in the mesenchymal samples (Fig. 3**D**). As expected, the expression of *LNCOG* was also significantly upregulated in HNSC tumours compared to normal samples (Fig. 3**E**). Interestingly, no significant difference was observed in the expression levels of *LNCOG* between tumour stages (Fig. S3), suggesting that *LNCOG* could be used as a diagnostic biomarker of HNSC that is pathologically independent. Functional analysis performed on the cancer hallmarks suggested that the putative targets of *LNCOG* were significantly enriched in 'angiogenesis' and 'activating invasion' (Fig. 3**F**), which again demonstrated the functional importance in regulating this aggressive subtype of HNSC.

3.3. Prediction and validation of putative target genes of LNCOG as prognosticators for HNSC

To further dissect the downstream regulatory mechanism, we next sought to identify putative target genes of *LNCOG* that are prognostic and investigate their potential functional associations with *LNCOG*. Univariate Cox regression analysis was performed to evaluate the association of the putative target genes of *LNCOG* with overall survival. Our results showed that five target genes (*ANXA5*,



(caption on next page)

Fig. 3. The prioritized master regulator LNCOG was significantly associated with overall survival and could act as a prognosticator for HNSC patients of the mesenchymal subtype. **(A)** ROC curves compare the predictive power of the expression levels of the three prioritized lncRNAs and the combined multivariate logistic regression model. **(B)** Univariate Cox regression analysis of the prioritized master regulatory lncRNAs with OS. **(C)** Kaplan–Meier curves compare the OS of *LNCOG* in low and high expression subgroups (stratified by the average expression level of *LNCOG*). **(D)** The master regulator *LNCOG* was significantly upregulated in the mesenchymal subtype of HNSC. **(E)** *LNCOG* was significantly upregulated in HNSC compared to normal samples. **(F)** Functional annotation based on hypergeometric tests for overrepresentation of the putative target genes of *LNCOG* in the ten cancer hallmark gene sets. **(G)** Univariate Cox regression analysis of the prioritized putative target genes with overall survival. **(H-L)** The expression of the five putative target genes were significantly positively or negatively correlated with the expression of *LNCOG*.

ITGA5, *CCBE1*, *P4HA2,* and *EPHX3*) showed significant prognostic values (BH-adjusted $P < 0.05$, Fig. 3**G**). Except for *EPHX3*, the expression of the other four genes were all positively correlated with the expression of *LNCOG* ($P < 0.05$, Pearson's correlation tests) (Fig. 3**H-L**). Our subsequent Kaplan-Meier survival analysis further demonstrated the prognostic values of these putative target genes, confirming that *ANAX5*, *CCBE1*, *ITGA5* and *P4HA2* are risk factors and *EPHX3* is a protective factor of overall survival (Fig. 4**A**). Moreover, these putative target genes were significantly up- or down-regulated in the mesenchymal HNSCs (Fig. 4**B**), suggesting their strong subtype specificities.

To validate the clinical associations of these putative targets of *LNCOG*, we collected four independent datasets from ICGC (ICGC-ORCA) and GEO (GSE27020, GSE41613 and GSE65858). In the ICGC-ORCA ($n = 40$) and GSE27020 ($n = 109$) datasets, patients with higher expression levels of *ITGA5*, *ANXA5* and *P4HA2* showed worse OS and disease-free survival (DFS) (Fig. 5**A-B and** Fig. S4A-**B**); In GSE65858 ($n = 270$), higher expression of *ITGA5* correlated with worse OS, whereas higher expression of *EPHX3* was associated with more favourable OS (Fig. 5**C and** Fig. S4C). In GSE41613 ($n = 97$), patients with higher expression of *ITGA5* and *ANXA5* displayed worse OS, demonstrating the same survival trend as observed in the other validation datasets (Fig. 5**D and** Fig. S4D). Moreover, consistent with the TCGA-HNSC dataset, we found significant differential expression of the five putative targets between the mesenchymal and non-mesenchymal subtypes in GSE65858 (Fig. S5). Altogether, our survival analysis demonstrated the potential to employ the five target genes of *LNCOG* for HNSC prognosis, which warrants further validation in larger clinical cohorts.

3.4. LNCOG may function as a ceRNA by sponging miR-148a-3p in HNSC

We postulated that *LNCOG* may function as a ceRNA to regulate the expression of target genes by sponging miRNAs. To test the hypothesis, we attempted to identify potential candidate miRNAs using two criteria: (1) the expression of a candidate miRNA should be both negatively correlated with the expression of *LNCOG* and a target gene; (2) the interactions between the candidate miRNAs and five putative target genes should be previously reported (e.g., in miRTarBase [40]). Using the strategy, we found 87 miRNAs with inverse correlation with the expression of *LNCOG* and *ITGA5*, and 231 miRNAs with reported interactions with *ITGA5* **(**Fig. 6**A)**. Among the five candidate miRNAs commonly found in the intersection, miR-148a-3p turned out to be the clinically relevant candidate based on univariate Cox regression analysis ($P = 0.047$, HR = 0.85 [95% CI, 0.72 – 1.00]) **(**Fig. 6**B)**, with its low expression associated with poorer survival ($P = 0.017$, log-rank test) and significant negative correlation with the expression of *LNCOG* ($P = 1.72E-04$) and its target gene *ITGA5* ($P = 3.60E-06$) (Fig. 6**C-E and** Table 3). Indeed, we found that the 3′ UTR of *ITGA5* contains a binding site of miR-148a-3p **(**Fig. 6**F)**. In addition, the miRNA downstream analysis of *ANXA5* and *P4HA2* was not conducted due to the clinical insignificance in the survival analysis of candidate miRNAs (Fig. S6A-**B**). Together, our computational analysis suggested that *LNCOG* might function as a ceRNA to regulate downstream genes such as *ITGA5* via sponging miR-148a-3p.

3.5. Drug sensitivity analysis confirmed the predictive values of the putative targets of LNCOG

To explore the potential predictive values of the putative target genes of *LNCOG*, we performed drug sensitivity analysis (detailed method in Section 2.5) by integrating 481 drug activity scores and corresponding gene expression profiles in 32 HNSC cell lines. After filtering out drugs with activity scores in less than half of the cell lines, we assessed for each drug the predictive value of each gene's expression by the area under ROC (AUC), with cell lines stratified by the median drug activity to a more sensitive and a more resistant subgroup. Interestingly, we found that each of the five target genes of *LNCOG* was predictive of the sensitivities of several drugs (Table 4), of which some are FDA-approved. For instance, the expression of *ANXA5* showed predictive value of the sensitivity of trametinib, which is a FDA-approved inhibitor of *MEK1* and *MEK2* for the treatment of anaplastic thyroid cancer (AUC = 0.76 and ANOVA $P = 0.0226$) (Fig. 7**A)**. Methotrexate is also a FDA-approved drug, which is an inhibitor of dihydrofolate reductase leading to the suppression of inflammation and prevention of cell division [42]. We found that the sensitivity of methotrexate could be predicted by the high expression of *CCBE1* (AUC = 0.77 and ANOVA $P = 0.0395$) (Fig. 7**B)**. Interestingly, the drug sensitivity of an effective severe acute respiratory syndrome coronavirus 2 inhibitor, masitinib [43], could be significantly predicted by the high expression of *EPHX3* (AUC = 0.80 and ANOVA $P = 0.0319$) (Fig. 7**C**). Cell lines with higher expression of *ITGA5* tended to be more resistent to an FDA-approved drug belinostat (AUC = 0.77 and ANOVA $P = 0.0206$) (Fig. 7**D**). Furthermore, the low expression of *P4HA2* showed strong predictive performance of the sensitivity of BRD-K80183349, which is a HDAC inhibitor (AUC = 0.86 and ANOVA $P = 3.03E-05$) (Fig. 7**E**). To further confirm the predictive values of the putative targets of *LNCOG*, we performed multivariate regression with LASSO regularization involving all the DEGs between the mesenchymal and non-mesenchymal subtypes for each drug. Consistently, we found that the expression of *ITGA5* was positively associated with the sensitivity of apicidin and Merck60, *CCBE1* with ruxolitinib, and *P4HA2* with BRD-K80183349 (Table 5), suggesting the robustness of their predictive values. Taken together, our lncRNA-mRNA regulatory network analysis revealed putative target genes of the master regulator lncRNA, with strong potential to be developed as prognostic and predictive biomarkers for HNSC.

4. Discussion

HNSCs are highly heterogeneous diseases, which can be subdivided into multiple molecularly distinct subtypes [6–8,11]. Despite the extensive studies about the heterogeneity of this malignancy,

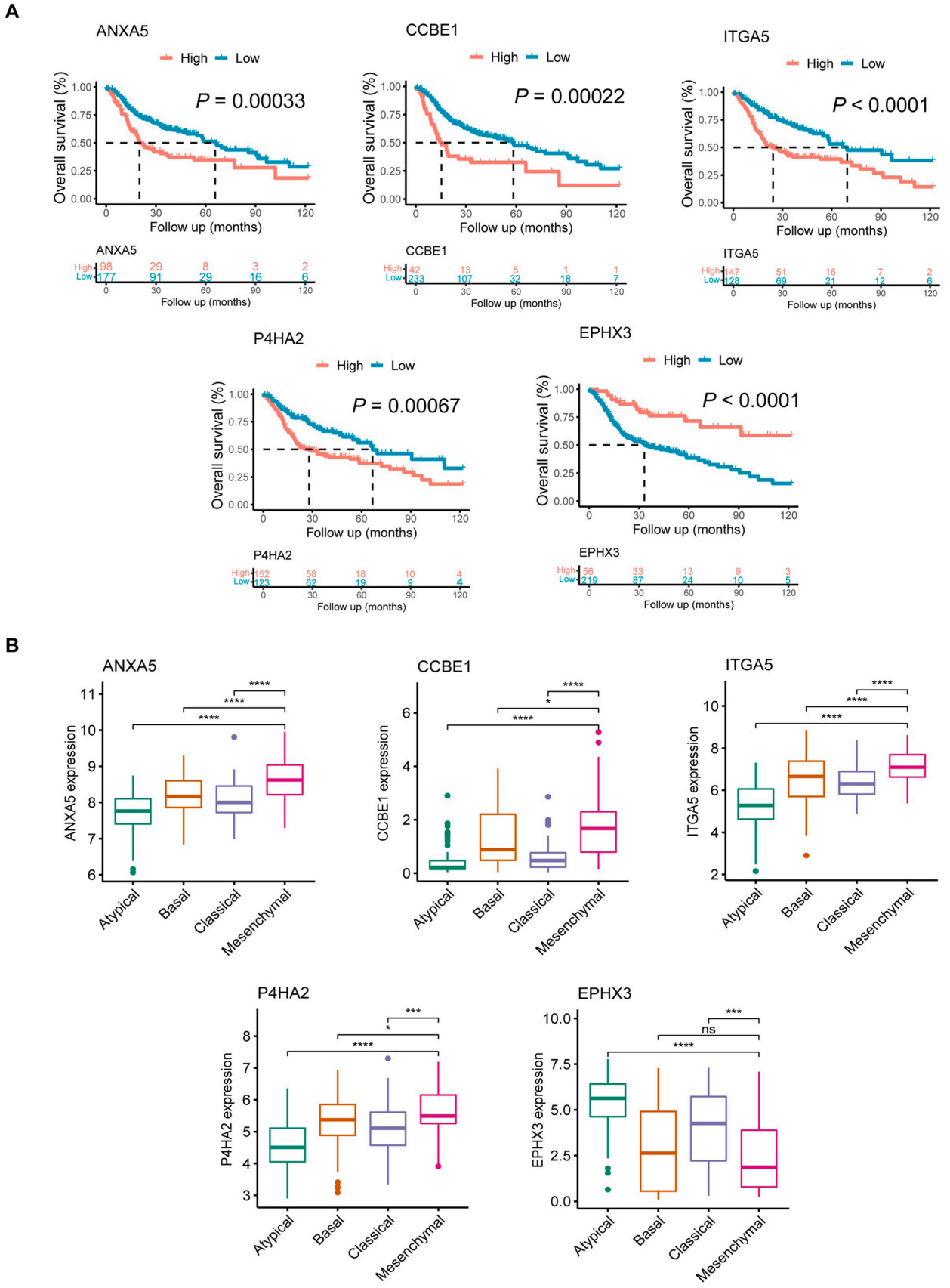


Fig. 4. Five putative target genes of *LNCOG* were significantly associated with OS and were differentially expressed between mesenchymal and non-mesenchymal HNSC. **(A)** Kaplan–Meier curves illustrate the prognostic relevance of the five putative target genes of *LNCOG*. In the Kaplan–Meier analysis for each gene, the patients were stratified by the optimal cutoff on the expression level of the gene. **(B)** Boxplots show that the five putative target genes of *LNCOG* were significantly up- or down-regulated in the mesenchymal HNSC patients, respectively.

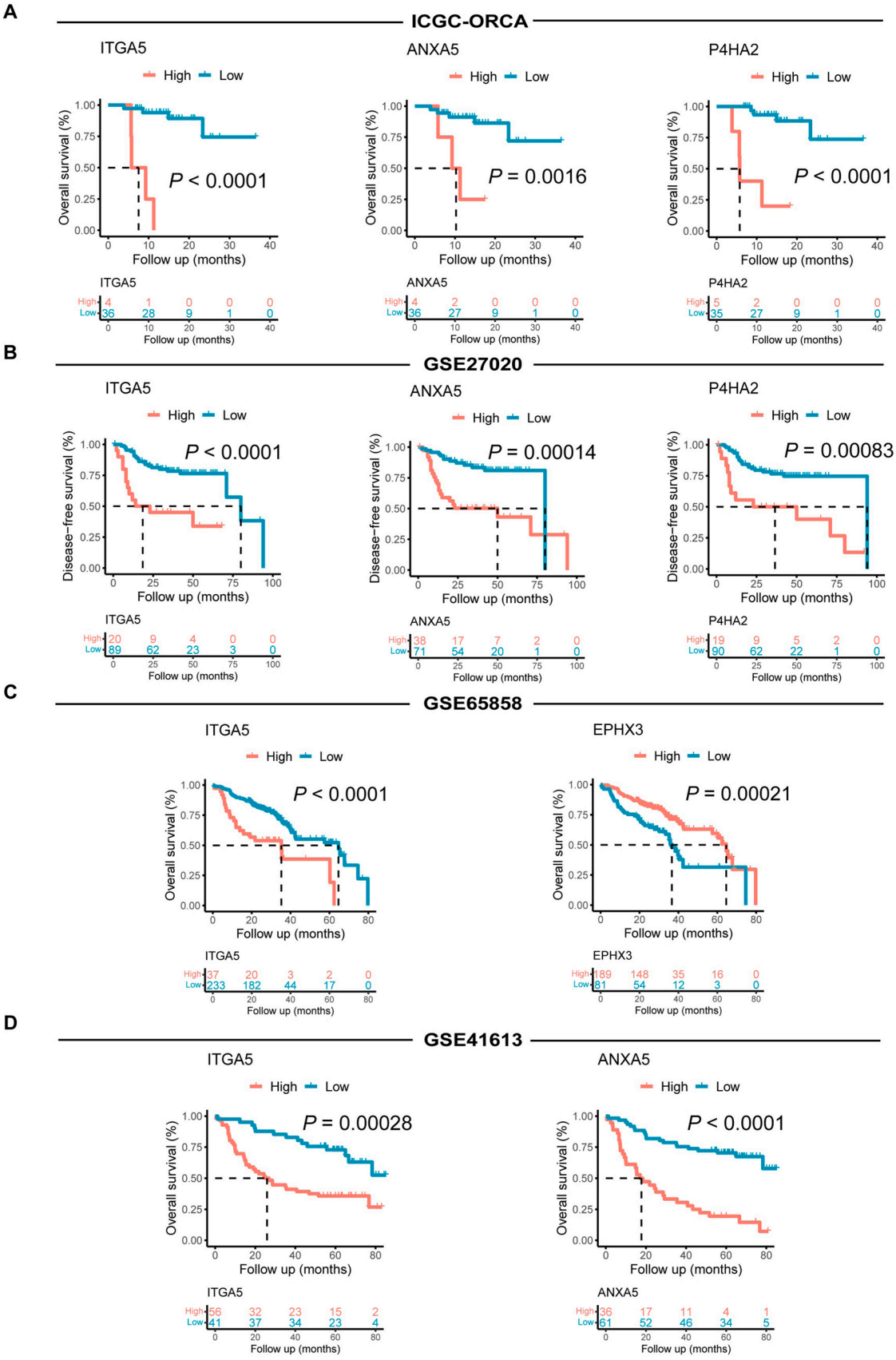


Fig. 5. The prognostic values of the five putative target genes were confirmed in independent validation datasets. **(A)** Kaplan–Meier plots show that *ITGA5*, *ANAX5* and *P4HA2* were significant risk factors of OS in the ICGC-ORCA dataset. **(B)** Kaplan–Meier plots show that *ITGA5*, *ANAX5* and *P4HA2* were risk factors of DFS in GSE27020 dataset. **(C)** Kaplan–Meier plots illustrate that *EPHX3* was a protective factor of OS whereas *ITGA5* was a risk factor in the GSE65858 dataset. **(D)** Kaplan–Meier plots illustrate that *ITGA5* and *ANAX5* were both risk factors of OS in the GSE41613 dataset.

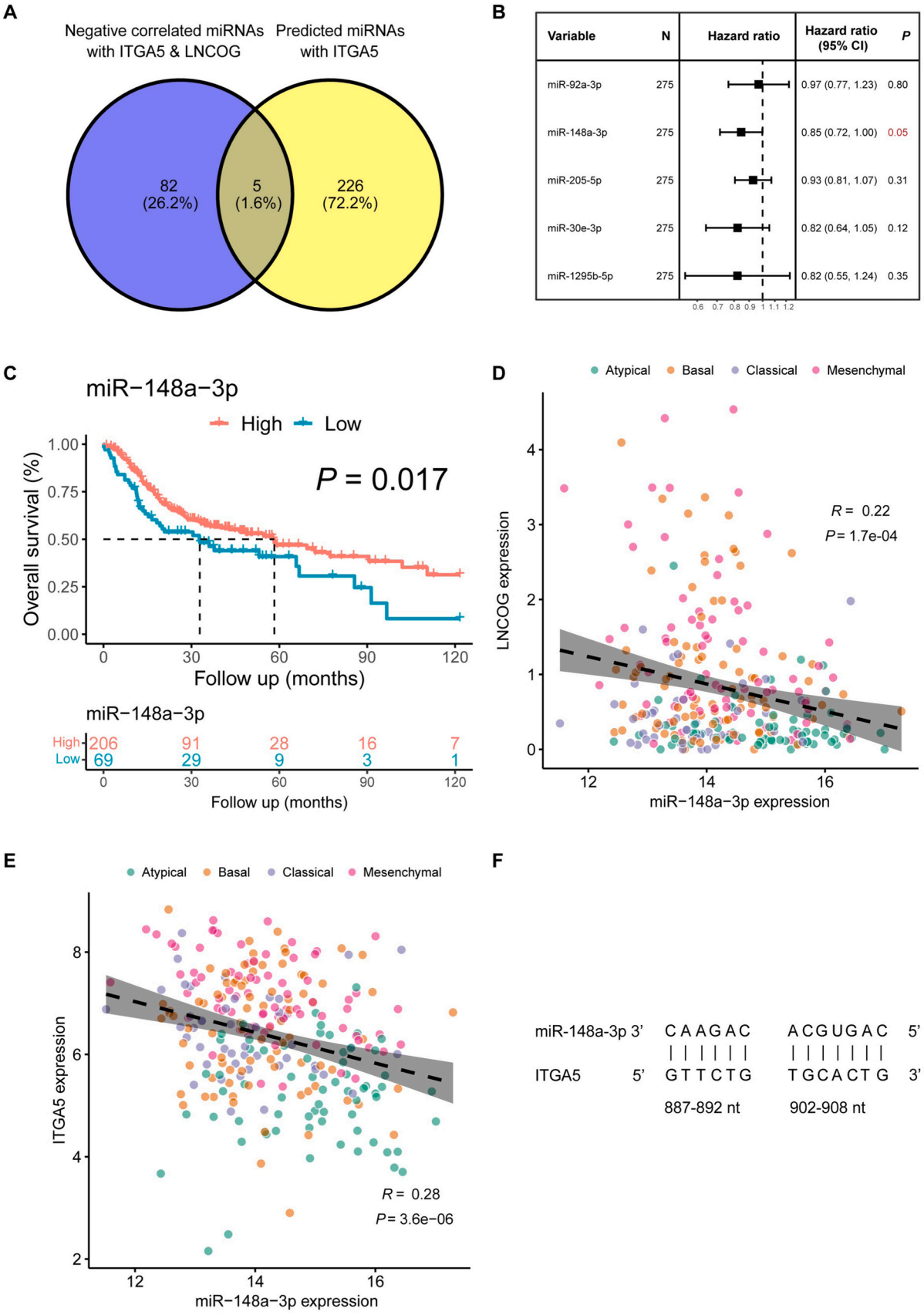


Fig. 6. CeRNA analysis identified miR-148a-3p as a potential target miRNA sponged by *LNCOG*. **(A)** Venn diagram illustrates the prioritization of miRNAs potentially sponged by *LNCOG* based on the overlap between the miRNAs with significant inverse correlation with *ITGA5* and *LNCOG* and the miRNAs with evident interactions based on miRTarBase database of *ITGA5*. **(B)** Univariate Cox regression analysis of the five prioritized miRNAs with OS. **(C)** Kaplan–Meier curves compare the OS of patients with low and high expression of miR-148a-3p (stratified by the optimal cutoff on the expression level of miR-148a-3p). **(D)** The expression of miR-148a-3p was significantly negatively correlated with the expression of *LNCOG*. **(E)** The expression of miR-148a-3p was significantly negatively correlated with its putative target *ITGA5*. **(F)** The predicted binding sequences of miR-148a-3p with its target gene *ITGA5*.

Table 3
Pearson correlation analysis between the master regulator LNCOG, target gene ITGA5 and candidate miRNAs.

miRNA	Pearson correlation coefficient with *LNCOG*	p-value with *LNCOG*	BH-adjusted p-value of *LNCOG*	Pearson correlation coefficient with *ITGA5*	p-value with *ITGA5*	BH-adjusted p-value with *ITGA5*	Hazard ratio	Univariate Cox p-value
miR-92a-3p	-0.17	5.53E-03	4.28E-02	-0.26	9.02E-06	1.55E-04	0.97	7.96E-01
miR-148a-3p	-0.22	1.72E-04	3.37E-03	-0.28	3.60E-06	7.19E-05	0.85	4.70E-02
miR-205–5p	-0.21	3.88E-04	6.43E-03	-0.27	7.34E-06	1.31E-04	0.93	3.07E-01
miR-30e-3p	-0.17	3.82E-03	3.27E-02	-0.17	5.62E-03	3.79E-02	0.82	1.21E-01
miR-1295b-5p	-0.21	5.93E-04	9.12E-03	-0.23	1.54E-04	1.79E-03	0.82	3.47E-01

Table 4
Drug sensitivity analysis results.

Gene name	Drug name	ANOVA p-value	AUC	Clinical status
CCBE1	brefeldin A	3.79E-04	0.87	–
ANXA5	PX-12	3.53E-02	0.75	–
ANXA5	lovastatin	3.85E-03	0.77	–
ITGA5	Merck60	6.95E-04	0.82	–
P4HA2	GANT-61	1.69E-02	0.84	–
ANXA5	trametinib	2.26E-02	0.76	FDA approved
ITGA5	belinostat	2.06E-02	0.77	FDA approved
P4HA2	BIX-01294	2.46E-02	0.77	–
CCBE1	ruxolitinib	1.97E-03	0.80	–
CCBE1	methotrexate	3.95E-02	0.77	FDA approved
ITGA5	apicidin	7.44E-05	0.77	–
P4HA2	apicidin	3.99E-02	0.78	–
ANXA5	BRD-K64610608	8.80E-03	0.80	–
EPHX3	masitinib	3.19E-02	0.80	Clinical trial
ITGA5	BRD-K80183349	1.19E-04	0.89	–
P4HA2	BRD-K80183349	3.03E-05	0.86	–
ITGA5	SMER-3	2.02E-02	0.75	–

few studies have been focused on the regulatory mechanism underlying specific subtypes, preventing the identification of novel drug targets and the design of novel subtype-specific therapeutics. Recent studies, including ours [44], have demonstrated strong specificities of lncRNAs in regulating particular cancer types [45–48], but there is still a lack of a computational framework for genome-wide analysis of lncRNA regulatory network.

Herein, we proposed an efficient bioinformatics framework by integrating lncRNA and mRNA expression profiles for regulatory network inference, followed by master regulator analysis with respect to a specific biological process. To exemplify the usefulness in HNSC, we inferred a lncRNA-mRNA regulatory network specifically for the mesenchymal subtype and identified three master regulatory lncRNAs mediating EMT. Focusing on the prognostic master regulator *LNCOG*, we predicted five putative target genes showing strong clinical relevance in multiple independent HNSC datasets. Subsequently, we identified ceRNA candidates by integrative analysis of lncRNA, miRNA and mRNA expression profiles. Our comprehensive analysis yielded novel insights into the subtype-specific regulatory mechanism and provided potential novel prognosticators and therapeutic strategies for HNSC.

Using the lncRNA regulatory network, five putative targets (*ANXA5*, *ITGA5*, *CCBE1*, *P4HA2*, and *EPHX3*) of *LNCOG* with prognostic values were identified and validated in multiple independent cohorts, highlighting the reliability of our network-based approach. Indeed, the cancer-related biological functions of these putative targets have been implicated in the literature. *ITGA5*, involved in PI3K-Akt signalling pathway critical for cell proliferation, differentiation, and apoptosis [49,50], was identified as a novel therapeutic target in the oral squamous carcinoma [51] and osteolytic lesions [52], as well as an anti-stroma target in pancreatic cancer to promote chemotherapy efficacy [53]. Another target gene, *ANXA5*, was found to be associated with dexamethasone (DEX) resistance in B-cell acute lymphoblastic leukaemia, and the inhibition of *ANXA5* may overcome the drug resistance [54]. *CCBE1*, whose overexpression was shown to promote tumour lymphangiogenesis and lymphatic metastasis in colorectal cancer [55], could be targeted by miR-330–3p, resulting in invasion and metastasis of breast cancer [56]. *P4HA2* was discovered to possess the capacity of uncoupling extracellular matrix signals in lung adenocarcinoma [57] and its inhibition could supresses breast cancer progression and metastasis [58]. Furthermore, the expression and DNA methylation of *EPHX3* were both found to be prognostic in two independent studies based on gene expression and DNA methylation profiles, respectively [59,60], which consistently demonstrate the reliability of our results.

Among the diverse regulatory mechanisms of lncRNAs, growing interests were focussed on the potential for them to act as ceRNAs, which are RNAs that share miRNA recognition elements (MREs). Therefore, lncRNAs may function as a decoy to regulate mRNA expression indirectly by competing for shared miRNAs [61]. Herein, by integrative analysis of multi-omics profiles (lncRNA, mRNA and miRNA expression profiles) in the TCGA-HNSC dataset, we proposed a ceRNA-based regulatory model, where *LNCOG* may sponge miR-148a-3p to upregulate *ITGA5* in the mesenchymal subtype of HNSC. Since miR-148a-3p was reported to be the tumour suppressor in several cancer types [62], the downregulation of miR-148a-3p in the mesenchymal subtype explains the poorer survival of patients classified to this specific subtype. Together, our integrative bioinformatic analysis revealed a potential novel mechanism for *LNCOG* to function via the lncRNA-miRNA-mRNA axis to regulate EMT in the mesenchymal subtype of HNSC, which warrants further experimental validations.

In addition to the strong prognostic utilities, we finally asked whether the target genes of *LNCOG* could also be employed as potential predictive biomarkers for more optimized therapeutics for HNSC patients. To this end, we performed drug sensitivity analysis by integrating the gene expression profiles and drug sensitivity data for 32 HNSC cell lines. As a result, we found that these putative target genes were indeed predictive of the sensitivities of multiple drugs, of which some were FDA-approved drugs such as trametinib, methotrexate, and belinostat. Therefore, we speculated that these putative target genes of *LNCOG* might be used also for the prediction of therapeutic response, especially for trametinib, which has been approved for the treatment of anaplastic thyroid cancer together with dabrafenib. Together, our comprehensive drug sensitivity analysis demonstrated the promising potentials of the predicted putative target genes to be used as prognostic and predictive biomarkers, which warrants further clinical validations.

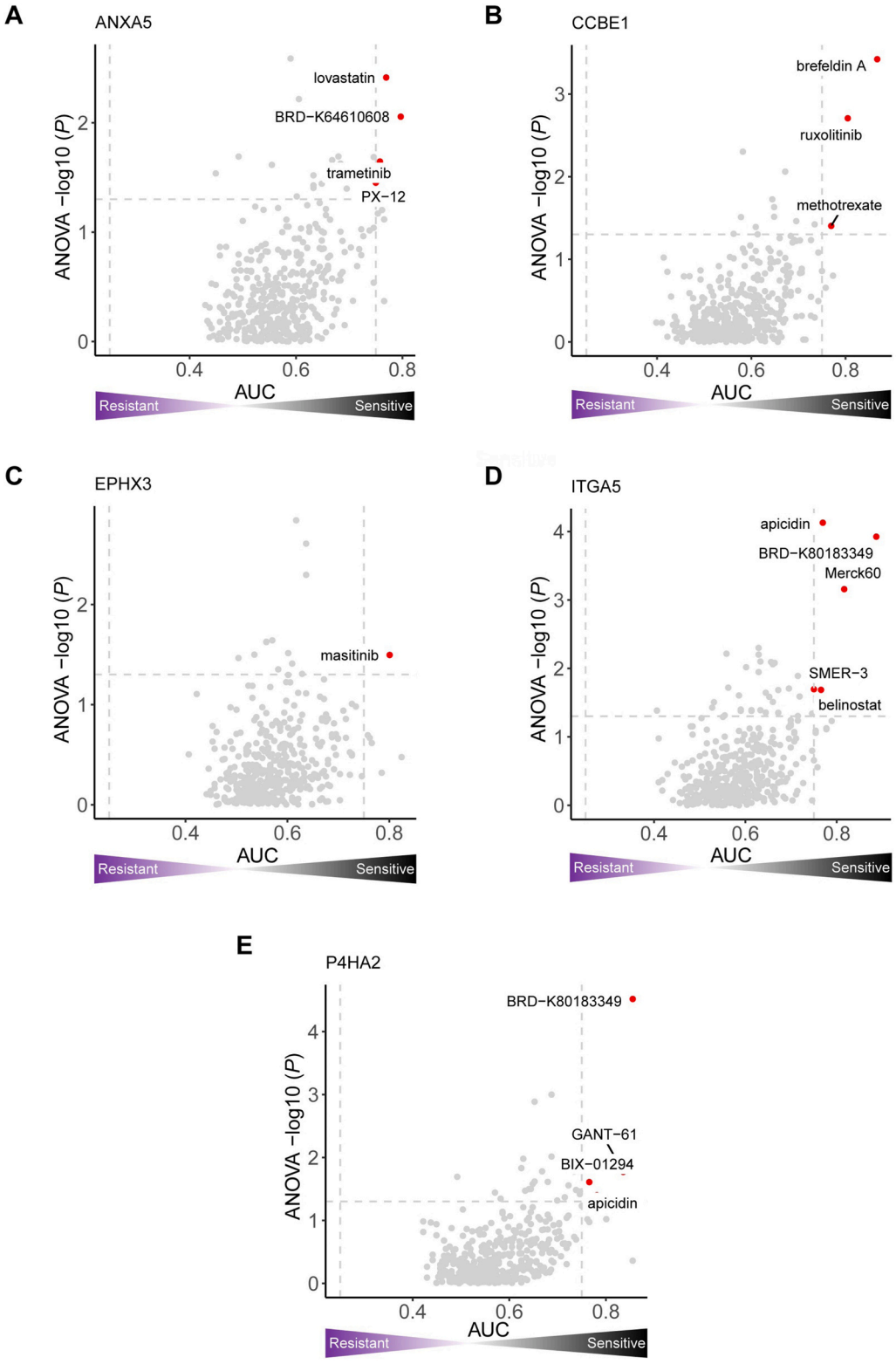


Fig. 7. Drug sensitivity analysis evaluating the predictive values of the five putative target genes of *LNCOG*. **(A-E)** Drug sensitivity analysis of the five putative target genes, **(A)** *ANXA5*, **(B)** *CCBE1*, **(C)** *EPHX3*, **(D)** *ITGA5*, and **(E)** *P4HA2*, with statistical significance evaluated by ANOVA, respectively. A gene was significantly associated with the sensitivity of cell lines to a certain drug if ANOVA $P < 0.05$ and AUC > = 0.75 or significantly associated with the resistance if ANOVA $P < 0.05$ and AUC < = 0.25.

Table 5
Multivariate regression with LASSO regularization confirmed the predictive values of ITGA5, CCBE1, EPHX3 and P4HA2.

Gene name	Drug name	LASSO coefficient	Clinical status
ITGA5	Merck60	0.03	–
CCBE1	ruxolitinib	0.03	–
ITGA5	apicidin	0.01	–
EPHX3	BRD-A86708339	0.06	–
P4HA2	BRD-K80183349	0.20	–
P4HA2	BRD-K66532283	0.08	–

5. Conclusions

In this study, we proposed an integrative bioinformatics framework for lncRNA regulatory network inference, followed by the identification of master regulators for specific cancer subtypes and the prediction of putative mRNA targets. The effectiveness of the framework was demonstrated by our comprehensive study in HNSC, which gained novel insights into the subtype-specific lncRNA regulatory mechanism as a potential ceRNA. Our survival and drug sensitivity analysis also demonstrated the versatile prognostic and/or predictive values of *LNCOG* and its putative targets, pending further clinical validations.

Funding

The present work was supported by a grant from Shenzhen Science, Technology and Innovation Commission (Project No. 基2020N368), a grant from Guangdong Basic and Applied Basic Research Foundation (Project No. 2019B030302012), a startup fund (Project No. 4937084), direct grant (2021.077) and Faculty Postdoctoral Fellowship Scheme 2021/22 (FPFS/21–22/32) from the Chinese University of Hong Kong, grants from the Research Grants Council (Project No. 11103718, 11103619, 11103921, 14111522) of the Hong Kong Special Administrative Region, China, awarded to Xin Wang.

CRediT authorship contribution statement

Jiang Li: Conceptualization, Software, Methodology, Investigation, Formal analysis, Visualization, Data Curation, Writing – original draft. **Tan Wu:** Visualization, Writing - review & editing. **Kai Song:** Writing - review & editing. **Lina Zhu:** Writing – review & editing. **Yijuan Wang**: Writing - review & editing. **Tingting Chen:** Writing - review & editing. **Xin Wang:** Conceptualization, Writing - review & editing, Funding acquisition, Project administration, Supervision.

Data availability

The data analysed in this study are publicly available in TCGA, GEO, ICGC, and CellMinerCDB databases.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2022.12.030.

References

[1] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394–424.
[2] Ringash J, Bernstein LJ, Devins G, Dunphy C, Giuliani M, Martino R, et al. Head and neck cancer survivorship: learning the needs, meeting the needs. Semin Radiat Oncol 2018;28:64–74.
[3] Horton JD, Knochelmann HM, Day TA, Paulos CM, Neskey DM. Immune evasion by head and neck cancer: foundations for combination therapy. Trends Cancer 2019;5:208–32.
[4] Hashim D, Genden E, Posner M, Hashibe M, Boffetta P. Head and neck cancer prevention: from primary prevention to impact of clinicians on reducing burden. Ann Oncol 2019;30:744–56.
[5] Posner MR, Hershock DM, Blajman CR, Mickiewicz E, Winquist E, Gorbounova V, et al. Cisplatin and fluorouracil alone or with docetaxel in head and neck cancer. N Engl J Med 2007;357:1705–15.
[6] Keck MK, Zuo Z, Khattri A, Stricker TP, Brown CD, Imanguli M, et al. Integrative analysis of head and neck cancer identifies two biologically distinct HPV and three non-HPV subtypes. Clin Cancer Res 2015;21:870–81.
[7] Chen YP, Wang YQ, Lv JW, Li YQ, Chua MLK, Le QT, et al. Identification and validation of novel microenvironment-based immune molecular subgroups of head and neck squamous cell carcinoma: implications for immunotherapy. Ann Oncol 2019;30:68–75.
[8] Chung CH, Parker JS, Karaca G, Wu JY, Funkhouser WK, Moore D, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell 2004;5:489–500.
[9] Cancer Genome Atlas. Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. Nature 2015;517:576–82.
[10] Zevallos JP, Mazul AL, Walter V, Hayes DN. Gene expression subtype predicts nodal metastasis and survival in human papillomavirus-negative head and neck cancer. Laryngoscope 2019;129:154–61.
[11] Walter V, Yin X, Wilkerson MD, Cabanski CR, Zhao N, Du Y, et al. Molecular subtypes in head and neck cancer exhibit distinct patterns of chromosomal gain and loss of canonical cancer genes. PLoS One 2013;8:e56823.
[12] Hulshoff MS, Del Monte-Nieto G, Kovacic J, Krenning G. Non-coding RNA in endothelial-to-mesenchymal transition. Cardiovasc Res 2019;115:1716–31.
[13] Klingenberg M, Matsuda A, Diederichs S, Patel T. Non-coding RNA in hepatocellular carcinoma: mechanisms, biomarkers and therapeutic targets. J Hepatol 2017;67:603–18.
[14] Bhan A, Soleimani M, Mandal SS. Long noncoding RNA and cancer: a new paradigm. Cancer Res 2017;77:3965–81.
[15] Gao N, Li Y, Li J, Gao Z, Yang Z, Li Y, et al. Long non-coding RNAs: the regulatory mechanisms, research strategies, and future directions in cancers. Front Oncol 2020;10:598817.
[16] Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. Mol Cell 2011;43:904–14.
[17] Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. Nat Genet 2011;43:621–9.
[18] Lee JT. Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. Nat Rev Mol Cell Biol 2011;12:815–26.
[19] Zhang E, Han L, Yin D, He X, Hong L, Si X, et al. H3K27 acetylation activated-long non-coding RNA CCAT1 affects cell proliferation and migration by regulating SPRY4 and HOXB13 expression in esophageal squamous cell carcinoma. Nucleic Acids Res 2017;45:3086–101.
[20] Lv J, Fan HX, Zhao XP, Lv P, Fan JY, Zhang Y, et al. Long non-coding RNA Unigene56159 promotes epithelial mesenchymal transition by acting as a ceRNA of miR-140-5p in hepatocellular carcinoma cells. Cancer Lett 2016;382:166–75.
[21] Sun M, Nie FQ, Wang YF, Zhang ZH, Hou JK, He DD, et al. LncRNA HOXA11-AS promotes proliferation and invasion of gastric cancer by scaffolding the chromatin modification factors PRC2, LSD1, and DNMT1. Cancer Res 2016;76:6299–310.
[22] Zhong ME, Chen YY, Zhang GN, Xu L, Ge W, Wu B. LncRNA H19 regulates PI3K-Akt signal pathway by functioning as a ceRNA and predicts poor prognosis in colorectal cancer: integrative analysis of dysregulated ncRNA-associated ceRNA network. Cancer Cell Int 2019:19.
[23] Li W, Zhu Q, Zhang S, Liu L, Zhang H, Zhu D. HOXC13-AS accelerates cell proliferation and migration in oral squamous cell carcinoma via miR-378g/HOXC13 axis. Oral Oncol 2020;111:104946.
[24] Fan M, Li X, Jiang W, Huang Y, Li J, Wang Z. A long non-coding RNA, PTCSC3, as a tumor suppressor and a target of miRNAs in thyroid cancer cells. Exp Ther Med 2013;5:1143–6.
[25] Zhou X, Liu S, Cai G, Kong L, Zhang T, Ren Y, et al. Long non coding RNA MALAT1 promotes tumor growth and metastasis by inducing epithelial-mesenchymal transition in oral squamous cell carcinoma. Sci Rep 2015;5:15972.
[26] Liang S, Zhang S, Wang P, Yang C, Shang C, Yang J, et al. LncRNA, TUG1 regulates the oral squamous cell carcinoma progression possibly via interacting with Wnt/beta-catenin signaling. Gene 2017;608:49–57.
[27] Yang YT, Wang YF, Lai JY, Shen SY, Wang F, Kong J, et al. Long non-coding RNA UCA1 contributes to the progression of oral squamous cell carcinoma by regulating the WNT/beta-catenin signaling pathway. Cancer Sci 2016;107:1581–9.
[28] Do D, Bozdag S. Cancerin: a computational pipeline to infer cancer-associated ceRNA interaction networks. PLoS Comput Biol 2018;14:e1006318.

[29] Paci P, Colombo T, Farina L. Computational analysis identifies a sponge interaction network between long non-coding RNAs and messenger RNAs in human breast cancer. BMC Syst Biol 2014;8:83.
[30] International Cancer Genome. Consortium, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, et al. International network of cancer genome projects. Nature 2010;464:993–8.
[31] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
[32] Fletcher MN, Castro MA, Wang X, de Santiago I, O'Reilly M, Chin SF, et al. Master regulators of FGFR2 signalling and breast cancer risk. Nat Commun 2013;4:2464.
[33] Zhao M, Kong L, Liu Y, Qu H. dbEMT: an epithelial-mesenchymal transition associated gene resource. Sci Rep 2015;5:11459.
[34] Castro MA, Wang X, Fletcher MN, Meyer KB, Markowetz F. RedeR: R/Bioconductor package for representing modular structures, nested networks and multiple levels of hierarchical associations. Genome Biol 2012;13:R29.
[35] Wang X, Terfve C, Rose JC, Markowetz F. HTSanalyzeR: an R/Bioconductor package for integrated network analysis of high-throughput screens. Bioinformatics 2011;27:879–80.
[36] Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst 2015;1:417–25.
[37] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284–7.
[38] Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell 2013;155:934–47.
[39] Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res 2016;44:e71.
[40] Huang HY, Lin YC, Li J, Huang KY, Shrestha S, Hong HC, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. Nucleic Acids Res 2020;48:D148–54.
[41] Luna A, Elloumi F, Varma S, Wang Y, Rajapakse VN, Aladjem MI, et al. CellMiner Cross-Database (CellMinerCDB) version 1.2: exploration of patient-derived cancer cell line pharmacogenomics. Nucleic Acids Res 2021;49:D1083–93.
[42] Inoue K, Yuasa H. Molecular basis for pharmacokinetics and pharmacodynamics of methotrexate in rheumatoid arthritis therapy. Drug Metab Pharmacokinet 2014;29:12–9.
[43] Drayman N, DeMarco JK, Jones KA, Azizi S-A, Froggatt HM, Tan K, et al. Masitinib is a broad coronavirus 3CL inhibitor that blocks replication of SARS-CoV-2. Science 2021;373:931–6.
[44] Zhu Wang. Integrative network analysis identified master regulatory long noncoding RNAs underlying the squamous subtype of pancreatic ductal adenocarcinoma. 2020 IEEE Int Conf Bioinform Biomed 2020;0:2936–42.
[45] Zhao Z, Guo Y, Liu Y, Sun L, Chen B, Wang C, et al. Individualized lncRNA differential expression profile reveals heterogeneity of breast cancer. Oncogene 2021;40:4604–14.
[46] Liu Y-R, Jiang Y-Z, Xu X-E, Yu K-D, Jin X, Hu X, et al. Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer. Breast Cancer Res 2016;18:33.
[47] Chen Y, Cheng WY, Shi H, Huang S, Chen H, Liu D, et al. Classifying gastric cancer using FLORA reveals clinically relevant molecular subtypes and highlights LINC01614 as a biomarker for patient prognosis. Oncogene 2021;40:2898–909.
[48] Li Q, Yu Q, Ji J, Wang P, Li D. Comparison and analysis of lncRNA-mediated ceRNA regulation in different molecular subtypes of glioblastoma. Mol Omics 2019;15:406–19.
[49] Lim W, Yang C, Bazer FW, Song G. Luteolin inhibits proliferation and induces apoptosis of human placental choriocarcinoma cells by blocking the PI3K/AKT pathway and regulating sterol regulatory element binding protein activity. Biol Reprod 2016;95:82.
[50] Hu R, Wang MQ, Niu WB, Wang YJ, Liu YY, Liu LY, et al. SKA3 promotes cell proliferation and migration in cervical cancer by activating the PI3K/Akt signaling pathway. Cancer Cell Int 2018;18:183.
[51] Deng Y, Wan Q, Yan W. Integrin alpha5/ITGA5 promotes the proliferation, migration, invasion and progression of oral squamous carcinoma by epithelial-mesenchymal transition. Cancer Manag Res 2019;11:9609–20.
[52] Pantano F, Croset M, Driouch K, Bednarz-Knoll N, Iuliani M, Ribelli G, et al. Integrin alpha5 in human breast cancer is a mediator of bone metastasis and a therapeutic target for the treatment of osteolytic lesions. Oncogene 2021;40:1284–99.
[53] Kuninty PR, Bansal R, De Geus SWL, Mardhian DF, Schnittert J, van Baarlen J, et al. ITGA5 inhibition in pancreatic stellate cells attenuates desmoplasia and potentiates efficacy of chemotherapy in pancreatic cancer. Sci Adv 2019;5. eaax2770.
[54] Yang J, Liu P, Ma D, Zhao P, Zhang Y, Lu Y, et al. Glucocorticoid resistance induced by ANXA5 overexpression in B-cell acute lymphoblastic leukemia. Pediatr Hematol Oncol 2021;38:36–48.
[55] Song J, Chen W, Cui X, Huang Z, Wen D, Yang Y, et al. CCBE1 promotes tumor lymphangiogenesis and is negatively regulated by TGFβ signaling in colorectal cancer. Theranostics 2020;10:2327–41.
[56] Mesci A, Huang X, Taeb S, Jahangiri S, Kim Y, Fokas E, et al. Targeting of CCBE1 by miR-330-3p in human breast cancer promotes metastasis. Br J Cancer 2017;116:1350–7.
[57] Pankova D, Jiang Y, Chatzifrangkeskou M, Vendrell I, Buzzelli J, Ryan A, et al. RASSF1A controls tissue stiffness and cancer stem-like cells in lung adenocarcinoma. EMBO J 2019;38:e100532.
[58] Xiong G, Deng L, Zhu J, Rychahou PG, Xu R. Prolyl-4-hydroxylase alpha subunit 2 promotes breast cancer progression and metastasis by regulating collagen deposition. BMC Cancer 2014;14:1.
[59] Liu B, Su Q, Ma J, Chen C, Wang L, Che F, et al. Prognostic value of eight-gene signature in head and neck squamous carcinoma. Front Oncol 2021;11:657002.
[60] Bai G, Song J, Yuan Y, Chen Z, Tian Y, Yin X, et al. Systematic analysis of differentially methylated expressed genes and site-speci fi c methylation as potential prognostic markers in head and neck cancer. J Cell Physiol 2019;234:22687–702.
[61] Dastsooz H, Alizadeh A, Habibzadeh P, Nariman A, Hosseini A, Mansoori Y, et al. LncRNA-miRNA-mRNA networks of gastrointestinal cancers representing common and specific LncRNAs and mRNAs. Front Genet 2021;12:791919.
[62] Kim H, Ko Y, Park H, Zhang H, Jeong Y, Kim Y, et al. MicroRNA-148a/b-3p regulates angiogenesis by targeting neuropilin-1 in endothelial cells. Exp Mol Med 2019:51.