

# LASSO-based Cox-PH model identifies an 11-lncRNA signature for prognosis prediction in gastric cancer

YONGHONG ZHANG<sup>1\*</sup>, HUAMIN LI<sup>2\*</sup>, WENYONG ZHANG<sup>1</sup>, YA CHE<sup>3</sup>,  
WEIBING BAI<sup>4</sup> and GUANGLIN HUANG<sup>4</sup>

<sup>1</sup>Department of General Surgery, Shangluo Central Hospital, Shangluo, Shaanxi 726000; <sup>2</sup>Department of Pathology, Weinan Central Hospital, Weinan, Shaanxi 714000; <sup>3</sup>Department of Medical Oncology, Shangluo Central Hospital, Shangluo, Shaanxi 726000; <sup>4</sup>Department of General Surgery, Yulin Xingyuan Hospital, Yulin, Shaanxi 719000, P.R. China

Received December 23, 2017; Accepted September 13, 2018

DOI: 10.3892/mmr.2018.9567

**Abstract.** The present study aimed to identify a long non-coding (lnc) RNAs-based signature for prognosis assessment in gastric cancer (GC) patients. By integrating gene expression data of GC and normal samples from the National Center for Biotechnology Information Gene Expression Omnibus, the EBI ArrayExpress and The Cancer Genome Atlas (TCGA) repositories, the common RNAs in Genomic Spatial Event (GSE) 65801, GSE29998, E-MTAB-1338, and TCGA set were screened and used to construct a weighted correlation network analysis (WGCNA) network for mining GC-related modules. Consensus differentially expressed RNAs (DERs) between GC and normal samples in the four datasets were screened using the MetaDE method. From the overlapped lncRNAs shared by preserved WGCNA modules and the consensus DERs, an lncRNAs signature was obtained using L1-penalized (lasso) Cox-proportional hazard (PH) model. lncRNA-mRNA networks were constructed for these signature lncRNAs, followed by functional annotation. A total of 14,824 common mRNAs and 2,869 common lncRNAs were identified in the 4 sets and 5 GC-associated WGCNA modules were preserved across all sets. MetaDE method identified 1,121 consensus DERs. A total of 50 lncRNAs were shared by preserved WGCNA modules and the consensus DERs. Subsequently, an 11-lncRNA signature was identified by LASSO-based Cox-PH model. The lncRNAs signature-based risk score could divide patients into 2 risk groups with signifi-

cantly different overall survival and recurrence-free survival times. The predictive capability of this signature was verified in an independent set. These signature lncRNAs were implicated in several biological processes and pathways associated with the immune response, the inflammatory response and cell cycle control. The present study identified an 11-lncRNA signature that could predict the survival rate for GC.

## Introduction

Gastric cancer (GC) is the fifth leading cause of malignancy worldwide, with a 5-year survival rate of <10% (1,2). In China, it is the second most commonly diagnosed cancer in men and the third most commonly diagnosed cancer in women (3). The poor prognosis is primarily attributable to patients being frequently identified at an advanced stage and therefore difficult to cure (4). Early detection is key to improving survival rate of GC patients. Therefore, discovery of valuable molecular biomarkers is of significance for the facilitation of early diagnosis and effective prediction of prognosis and thereby contributing to improved outcomes in GC patients.

Long noncoding RNAs (lncRNAs) are defined as a group of non-protein-coding transcripts of greater than 200 nucleotides in length, which are characterized by tissue-specific expression patterns (5,6). With the number of lncRNAs being triple the number of protein-coding genes, lncRNAs are predicted to exhibit a more important role in basic, translational and clinical oncology than protein-coding genes (7). Several lncRNAs have been demonstrated in GC, including H19 (8-10), HOTAIR (11,12) and ANRIL (13). However, the association of lncRNAs with GC prognosis has not been fully elucidated. Although a recent study by Miao *et al* (14) reported a 4-lncRNA signature of prognostic value for GC patients, the signature is yielded by bioinformatics analysis of The Cancer Genome Atlas (TCGA) data only. A comprehensive analysis of gene expression data of GC patients from more databases is required for acquiring a more convincing prognostic lncRNAs signature.

In contrast with the study of Miao *et al* (14), the present study performed an integrated analysis on GC gene expression data mined in the National Center for Biotechnology Information (NCBI), Gene Expression Omnibus (GEO), EBI

---

*Correspondence to:* Dr Weibing Bai or Dr Guanglin Huang, Department of General Surgery, Yulin Xingyuan Hospital, 33 Middle Section of West Renmin Road, Yuyang, Yulin, Shaanxi 719000, P.R. China  
E-mail: wbb2012@tom.com  
E-mail: hglmhp@163.com

\*Contributed equally

**Key words:** network, mRNA, pathway, gene ontology, differentially expressed RNAs

ArrayExpress and TCGA repositories. The present study was mainly focused on revealing the critical lncRNAs involved in GC pathogenesis and the roles of the critical lncRNAs in the molecular mechanisms of GC. An 11-lncRNA signature was identified for prognostic risk assessment of GC patients using weighted correlation network analysis (WGCNA) network, the MetaDE method and a LASSO-based Cox-proportional hazard (PH) model. In addition, the prognostic significance of this signature was validated in an independent set. In order to reveal the molecular mechanisms of these critical lncRNAs, the lncRNA-mRNA interaction network was constructed for functional and pathway enrichment analysis. The results revealed that these critical lncRNAs can regulate the associated mRNAs to influence the immune response, inflammatory response and cell cycle in the pathogenesis of GC.

## Materials and methods

**Data resource and preprocessing.** Gene expression profiles for GC were searched in publicly accessible GEO at the NCBI (<http://www.ncbi.nlm.nih.gov/geo/>) and EBI ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>). Inclusion criteria were: Human gene expression data; gastric cancer specimens and paired normal specimens; total count of specimens  $\geq 50$ . Finally, Genomic Spatial Event (GSE) (15) 6580 and GSE29998 downloaded from NCBI GEO and E-MTAB-1338 from EBI ArrayExpress were selected in the present study (Table I).

Raw data (TXT) in GSE6580, GSE29998 and E-MTAB-1338 were subject to  $\log_2$  transformation by limma (version 3.34.0) software (16) (<https://bioconductor.org/packages/release/bioc/html/limma.html>). Subsequently, the data were transformed from a skewed distribution to normal distribution, followed by median normalization. Based on the platform annotation files (Table I), probe sets that were assigned with a RefSeq transcript ID and/or Ensembl gene ID were obtained, of which the probe sets labeled as 'NR' (non-coding RNA in the Refseq database) were selected. In addition, platform sequencing data was aligned with human genome (GRCh38) (17,18) using Clustal 2 (<http://www.clustal.org/clustal2/>) (19). The resulting lncRNAs and the above-mentioned lncRNAs annotated in Refseq database were combined and used in further analysis.

The present study also acquired mRNA-seq data of 384 GC samples and 26 normal controls from TCGA portal (<https://gdc-portal.nci.nih.gov/>), which did not require preprocessing. Common RNAs of the GSE6580, GSE29998, E-MTAB-1338 and TCGA sets were used for further analysis.

**WGCNA network analysis.** WGCNA (20) is a bioinformatics tool used to build a gene co-expression networks to mine network modules closely associated with diseases. Based on the common RNAs identified, WGCNA package (21) (version 1.61) in R 3.4.1 language was applied to identify GC-associated RNA modules (<https://cran.r-project.org/web/packages/WGCNA/index.html>) in the present study. The TCGA set was used as the training set, while GSE6580, GSE29998 and E-MTAB-1338 were selected as testing sets. Comparability of these 4 sets were assessed by correlation analysis of RNA expression levels. A weighted gene co-expression network was built as previously described (20). Briefly, the soft threshold power of  $\beta$  was determined using

scale-free topology criterion. Following the removal of RNAs with coefficients of variation  $< 0.1$ , the weighted adjacency matrix was then developed. A dynamic tree cut algorithm was used to mine modules with a module size  $\geq 30$  and a minimum cut height of 0.95. In addition, preservation of modules in all 4 datasets was examined using the module preservation function of the WGCNA package. In addition, functional annotation of the modules identified was investigated using the userListEnchment function of WGCNA package.

**Identification of consensus differentially expressed RNAs.** Consensus differentially expressed RNAs (DERs) between GC specimens and normal control specimens across the 4 datasets (GSE6580, GSE29998, E-MTAB-1338 and TCGA) were identified with metaDE package (22,23) (<https://cran.r-project.org/web/packages/MetaDE/>) in R language version 3.4.1. The cutoff was set at  $\tau_2=0$ ,  $Q_{pval}>0.05$ ,  $P<0.05$  and false discovery rate (FDR) $<0.05$ .  $\tau_2$  denotes the amount of heterogeneity while  $Q_{pval}$  denotes heterogeneity of a dataset. The common lncRNAs shared by the list of consensus DERs and the RNAs in the preserved WGCNA modules were selected for further analysis.

**Development of a prognostic risk scoring system for GC.** L1-penalized (lasso) characterized by simultaneous variable selection and shrinkage is a useful method for determining interpretable prediction rules in high-dimensional data (24). In order to determine an lncRNA signature for prognosis, the penalized package (24) in R language (version 3.4.1) was applied to fit a lasso Cox-PH (25) to the overlapped lncRNAs. Based on the optimal lambda value that was selected through a 1,000 cross-validations, a panel of prognostic lncRNAs was determined. An equation for calculating risk score was generated based on the expression levels of these prognostic lncRNAs and their regression coefficients from the Cox-PH model as follows:

$$\text{Risk score} = \beta_1 \text{lncRNA}_1 \times \text{exprlncRNA}_1 + \beta_2 \text{lncRNA}_2 \times \text{exprlncRNA}_2 + \dots + \beta_n \text{lncRNA}_n \times \text{exprlncRNA}_n$$

Risk score was calculated and assigned to each patient in the training set (TCGA set, Table II). With the median risk score as cutoff, all patients in the training set were split into a high-risk group and a low-risk group. Overall survival (OS) time and recurrence-free survival (RFS) time of the two risk groups were analyzed and compared by Kaplan-Meier survival analysis and the logrank test.

The robustness of the risk scoring system was validated in an independent dataset (GSE62254) (26) downloaded from NCBI GEO (platform: GPL570, Affymetrix Human Genome U133 Plus 2.0 Array). GSE62254 included the gene expression data of 300 GC tissue samples (Table II). Raw data was preprocessed using an oligo (27) package in R language (version 3.4.1). Risk score and risk groups were determined similarly for the GSE62254 dataset. Discrepancies in OS time and RFS time between the risk groups were analyzed using Kaplan-Meier survival analysis and the log rank test.

**Functional analysis of prognostic lncRNAs.** To investigate the biological function of these prognostic lncRNAs identified

Table I. Basic information of gene expression profiles from NCBI GEO, EBI ArrayExpress and TCGA.

Accession ID	Platform	Total sample	Tumor	Control
GSE65801	GPL14550 Agilent	64	32	32
GSE29998	GPL6947 Illumina	99	50	49
E-MTAB-1338	Illumina HumanHT	71	50	21
TCGA	Illumina HiSeq	420	384	36

NCBI, National Center for Biotechnology Information; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; GSE, Genomic Spatial Event.

Table II. Clinical features of TCGA dataset and GSE622254.

Clinical characteristics	TCGA (n=384)	GSE62254 (n=300)
Age (years, mean $\pm$ SD)	65.15 $\pm$ 10.61	61.94 $\pm$ 11.36
Gender (male/female/data unavailable)	243/133/8	199/101
Recurrence (yes/no/data unavailable)	78/260/46	125/157/18
Vitality (dead/alive/data unavailable)	122/238/24	135/148//17
DFS (months) (mean $\pm$ SD)	15.84 $\pm$ 17.05	33.72 $\pm$ 29.82
OS (months) (mean $\pm$ SD)	16.17 $\pm$ 16.96	50.59 $\pm$ 31.42

TCGA, The Cancer Genome Atlas; GSE, Genomic Spatial Event; SD, standard deviation; -, data unavailable; DFS, disease free survival time; OS, overall survival time.

above in GC tumorigenesis, lncRNA-mRNA networks were constructed for them based on the correlation coefficients between RNAs from WGCNA modules. Gene ontology (GO; <http://www.geneontology.org/>) function and Kyoto Encyclopedia of Genes and Genomes (KEGG; <https://www.kegg.jp/>) pathway enrichment analysis was performed for the RNAs in these lncRNA-mRNA networks by DAVID Bioinformatics Tool (28,29) (version 6.8; <https://david-d.ncifcrf.gov/>).

## Results

**RNA expression data.** Following data preprocessing, the present study identified 17,693 common RNAs in the GSE6580, GSE29998, E-MTAB-1338 and TCGA sets, including 14,824 mRNAs and 2,869 lncRNAs (Table III).

**WGCNA network and modules.** Based on these common RNAs, WGCNA was used to mine GC-associated modules, with TCGA set as the training set and GSE6580, GSE29998, E-MTAB-1338 as validation sets. The correlation of gene expression between these sets was in the range of 0.4-1 with  $P < 1 \times 10^{-200}$  (Fig. 1), indicating good comparability between the sets. For adjacencies calculation, the soft threshold power of  $\beta$  was determined to be 5 when the scale-free topology fit (scale-free  $R^2$ ) achieved 0.9 (Fig. 2).

A total of 11 modules (black, blue, brown, green, grey, magenta, pink, red, turquoise, yellow and purple) were mined with WGCNA for the TCGA dataset. In the resulting dendrogram (Fig. 3A), these modules were represented by branches in different colors. Module mining was also conducted

Table III. Numbers of mRNAs and lncRNAs in the datasets.

Accession ID	Total count	mRNA	lncRNA
GSE65801	23,081	17,056	6,025
E-MTAB-1338	18,730	15,376	3,354
GSE29998	20,586	15,376	5,210
TCGA	24,840	17,579	7,261
Common	17,693	14,824	2,869

lnc, long non-coding; GSE, Genomic Spatial Event; TCGA, The Cancer Genome Atlas.

in GSE29998, GSE6580 and E-MTAB-1338. The gene dendrograms are presented in Fig. 3B-D.

As illustrated in a gene multi-dimensional scaling (MDS) plot (Fig. 4A), RNAs of the same module were prone to cluster together, suggesting similar expression patterns of RNAs in the same module. A hierarchical clustering analysis of the 11 modules identified that the associated modules clustered together, such as the black module and the yellow module, the pink module and the purple module, the magenta module and the red module, and the grey module and the turquoise module (Fig. 4B). Not unexpectedly, these modules were also close to each other in the module MDS plot (Fig. 4C).

In addition, out of the 11 modules, black, blue, brown, turquoise and yellow modules with Z-score  $> 5$  were identified to be well preserved across the GSE6580, GSE29998, E-MTAB-1338 and TCGA sets (Table IV). Functional

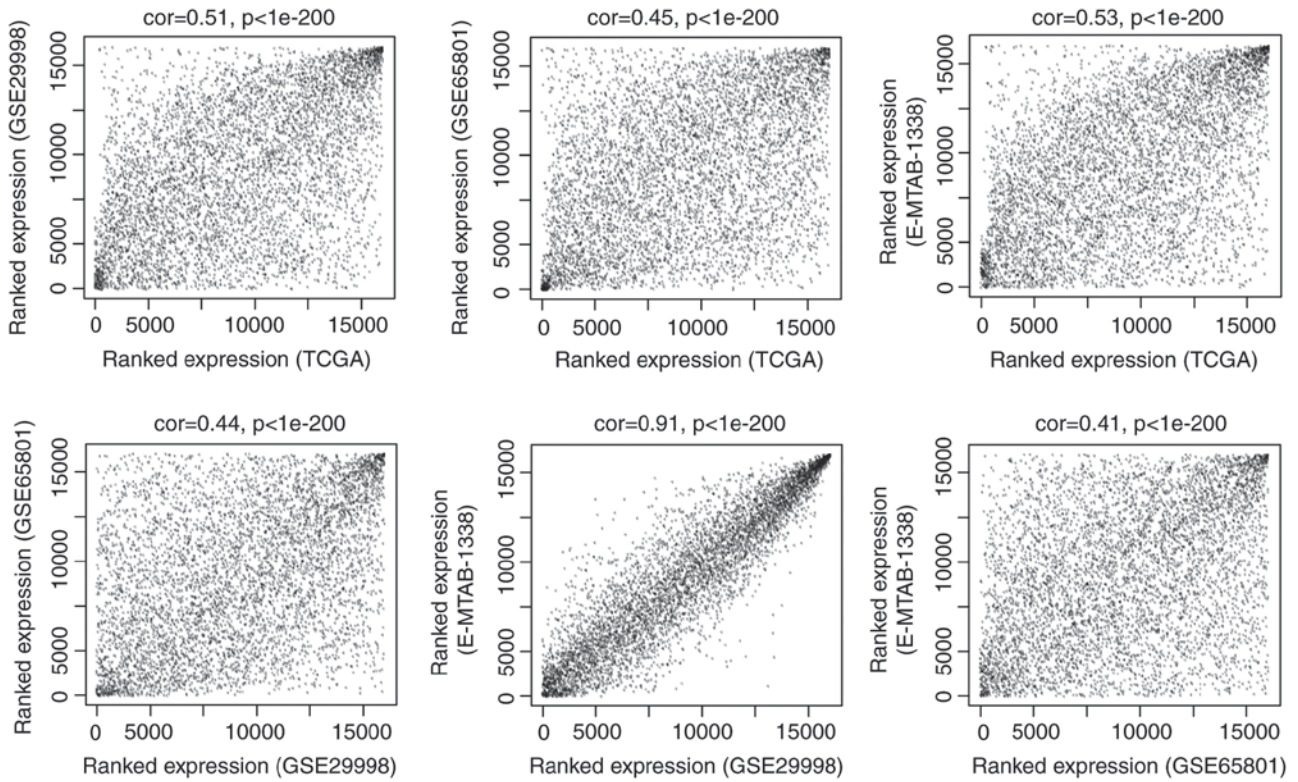


Figure 1. Analysis of comparability of the TCGA, GSE29998, GSE65801 and E-MTAB-1338 sets. Each panel presents the correlation of ranked expression of genes between 2 datasets. Cor value and P-value are calculated using the WGCNA package. TCGA, The Cancer Genome Atlas; GSE, Genomic Spatial Event; WGCNA, weighted correlation network analysis; Cor, correlation coefficient.

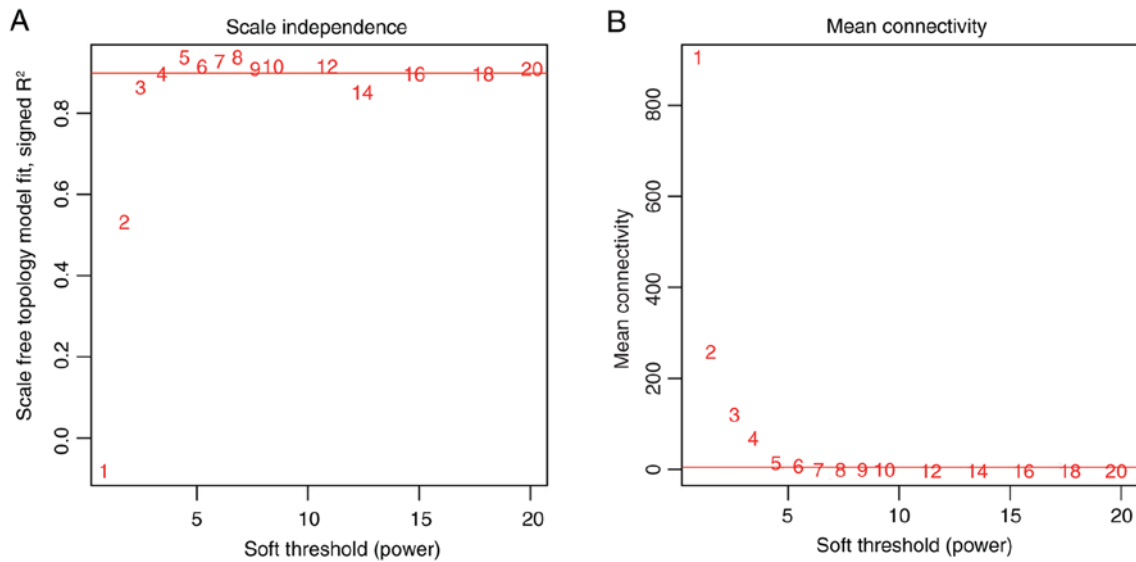


Figure 2. Net topology analysis for optimizing soft-threshold power. (A) The scale-free fit index (scale-free  $R^2$ , y-axis) as a function of the soft-threshold power (x-axis). When the scale-free topology fit reaches 0.9 (red line), the soft threshold power is 5. (B) The mean connectivity (degree, y-axis) as a function of the soft threshold power (x-axis). When the soft threshold power is 5, the mean connectivity is 2 (red line).

annotation of the 5 modules was performed using WGCNA package (Table IV). The black module was associated with digestion. The blue module was associated with immune response. The brown module was correlated with cell cycle. The turquoise module was associated with cell adhesion. The yellow module was linked to protein amino acid glycosylation (Table IV).

*Consensus DERs*. The metaDE package identified 1,121 consensus DERs in the GSE6580, GSE29998, E-MTAB-1338 and TCGA sets, of which 255 were lncRNAs. A heatmap of these consensus DERs was generated by heatmap.sig.genes function in MetaDE package (Fig. 5). Clearly, expression patterns of these consensus DERs were similar in 4 datasets. Furthermore, 288 RNAs were overlapped between the 5

Table IV. Characteristics of WGCNA network modules.

TCGA	GSE29998	GSE65801	E-MTAB-133	Color	Module size	Module preservation (Z-score)	Module characterization
D1M1	D2M1	D3M1	D4M1	Black	59	28.06	Digestion
D1M2	D2M2	D3M2	D4M2	Blue	417	31.59	Immune response
D1M3	D2M3	D3M3	D4M3	Brown	411	25.26	Cell cycle
D1M4	D2M4	D3M4	D4M4	Green	111	6.41	-
D1M5	D2M5	D3M5	D4M5	Grey	1,097	4.90	-
D1M6	D2M6	D3M6	D4M6	Nagenta	38	10.21	-
D1M7	D2M7	D3M7	D4M7	Pink	56	22.08	-
D1M8	D2M8	D3M8	D4M8	Red	78	17.64	-
D1M9	D2M9	D3M9	D4M9	Turquoise	564	29.46	Cell adhesion
D1M10	D2M10	D3M10	D4M10	Yellow	215	14.37	Protein amino acid glycosylation
D1M11	D2M11	D3M11	D4M11	Purple	35	8.30	-

WGCNA, weighted correlation network analysis; TCGA, The Cancer Genome Atlas; GSE, Genomic Spatial Event.

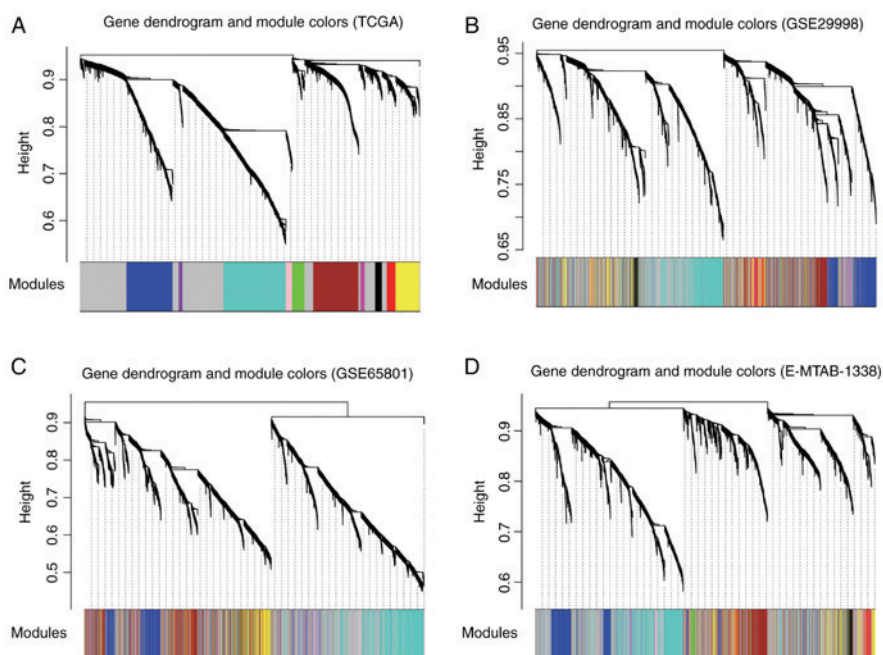


Figure 3. Clustering dendrograms of identified modules in (A) TCGA (B) GSE29998, (C) GSE65801 and (D) E-MTAB-1338 sets. Modules are labeled in different colors. TCGA, The Cancer Genome Atlas; GSE, Genomic Spatial Event.

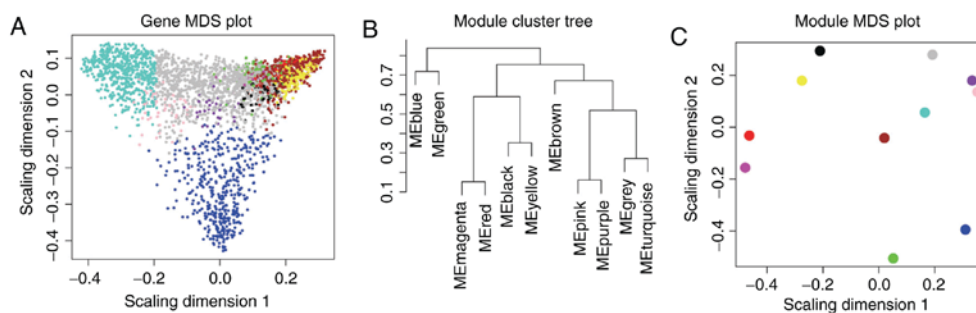


Figure 4. Module analysis. (A) MDS plot demonstrating the similarity of RNAs expression patterns between different modules. RNAs of different modules are marked in different colors. (B) Module cluster tree. (C) MDS plot exhibiting the degree of similarity between the identified modules. Modules are labeled in different colors. MDS, multi-dimensional scaling.

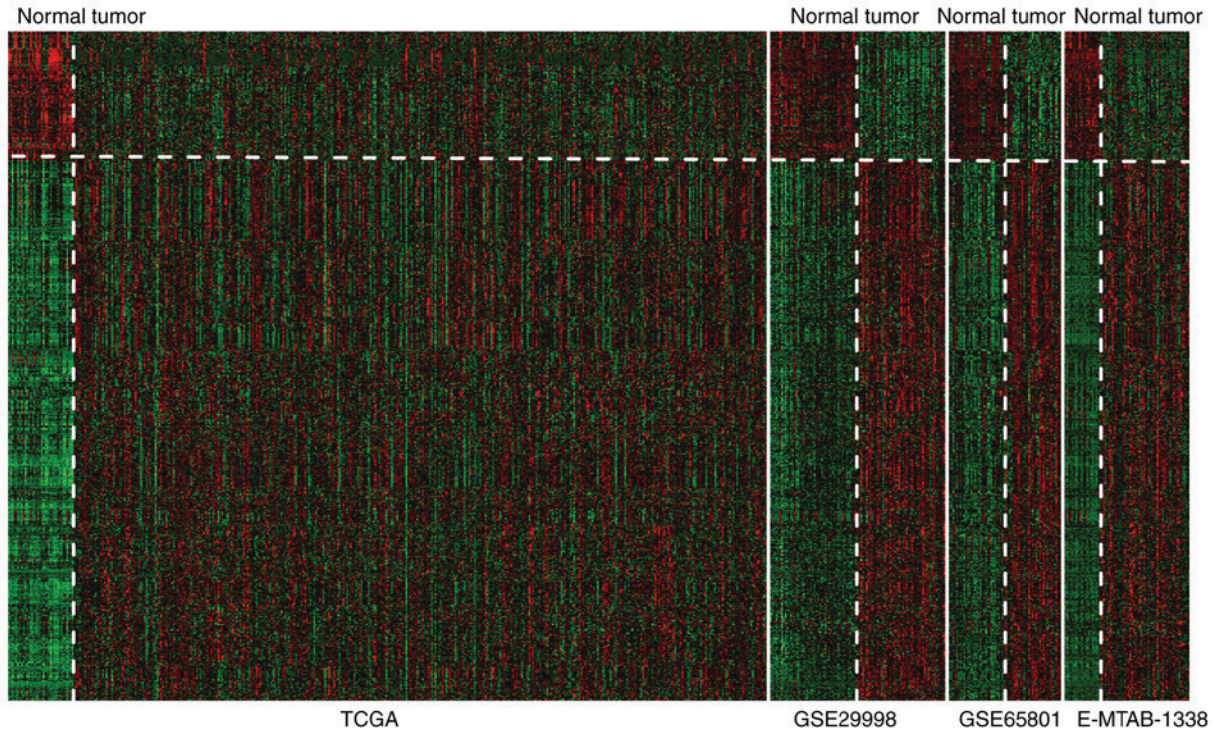


Figure 5. A heatmap of consensus RNAs identified by MetaDE. RNAs expression patterns are similar in the TCGA, GSE29998, GSE65801 and E-MTAB-1338 sets. TCGA, The Cancer Genome Atlas; GSE, Genomic Spatial Event.

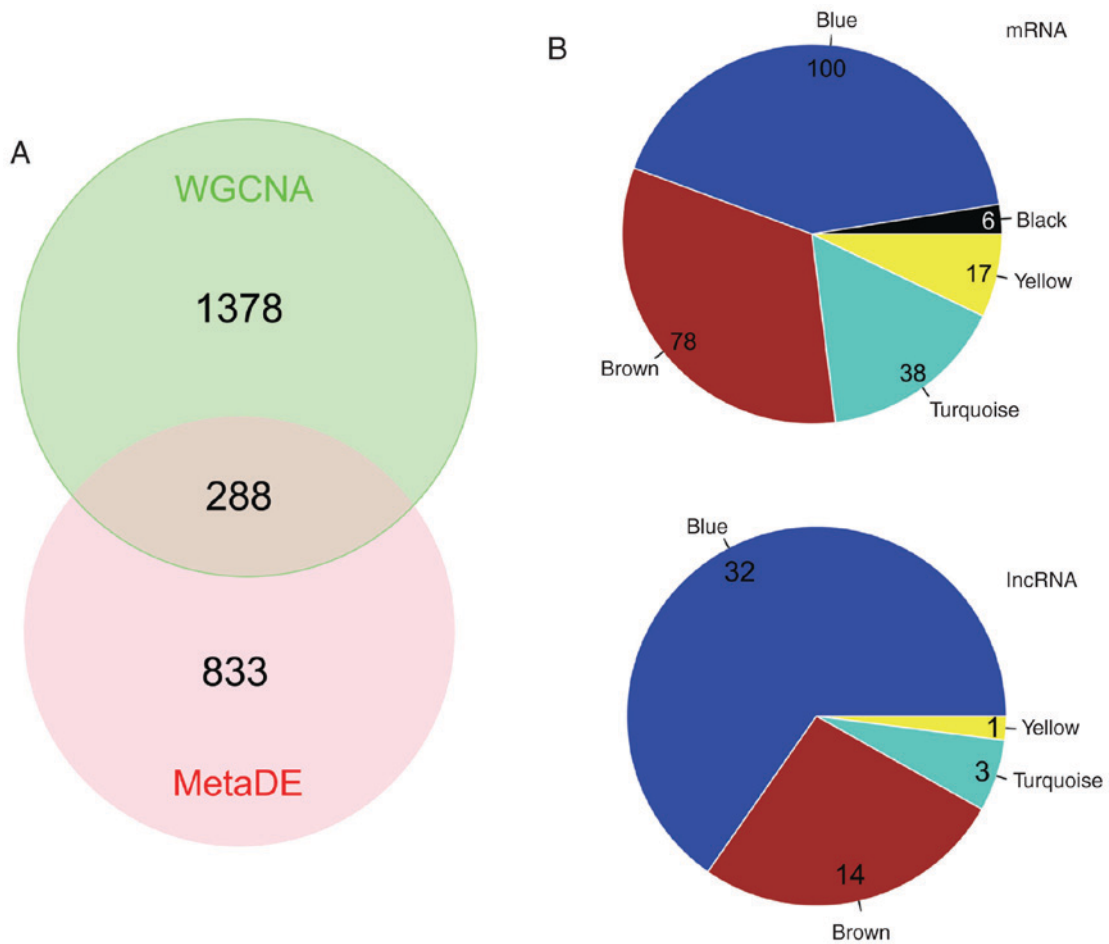


Figure 6. Analysis of overlapped RNAs. (A) Venn diagram displaying the overlapped RNAs between the preserved WGCNA modules and the consensus DERs identified by MetaDE. (B) Distribution of overlapped mRNAs (upper) and lncRNAs (lower) in the 5 preserved WGCNA modules (black, blue, brown, turquoise and yellow). Inc, long non-coding; WGCNA, weighted correlation network analysis; DERs, differentially expressed RNAs.

Table V. The 11 prognostic lncRNAs identified by LASSO-based Cox-proportion hazard model.

lncRNA	Coefficient	HR	95% CI
ARHGAP5-AS1	0.0124	1.1907	0.8259-1.7166
FLVCR1-AS1	-0.1191	0.6610	0.4916-0.8886
H19	0.9171	1.0497	0.9390-1.1735
HOTAIR	-0.4973	0.8970	0.6584-1.2222
LINC00221	1.1799	1.9190	1.2021-3.0633
MCF2L-AS1	-0.7009	0.7785	0.6053-1.0014
MUC2	-0.0902	0.9516	0.8631-1.0492
PRSS30P	0.2572	1.1254	0.8263-1.5329
SCARNA9	-0.8615	0.7383	0.5449-1.0004
TP53TG1	0.1493	1.1386	0.8808-1.4720
XIST	-0.9235	0.5469	0.1926-1.5527

lnc, long non-coding; HR, hazard ratio; CI, confidence interval.

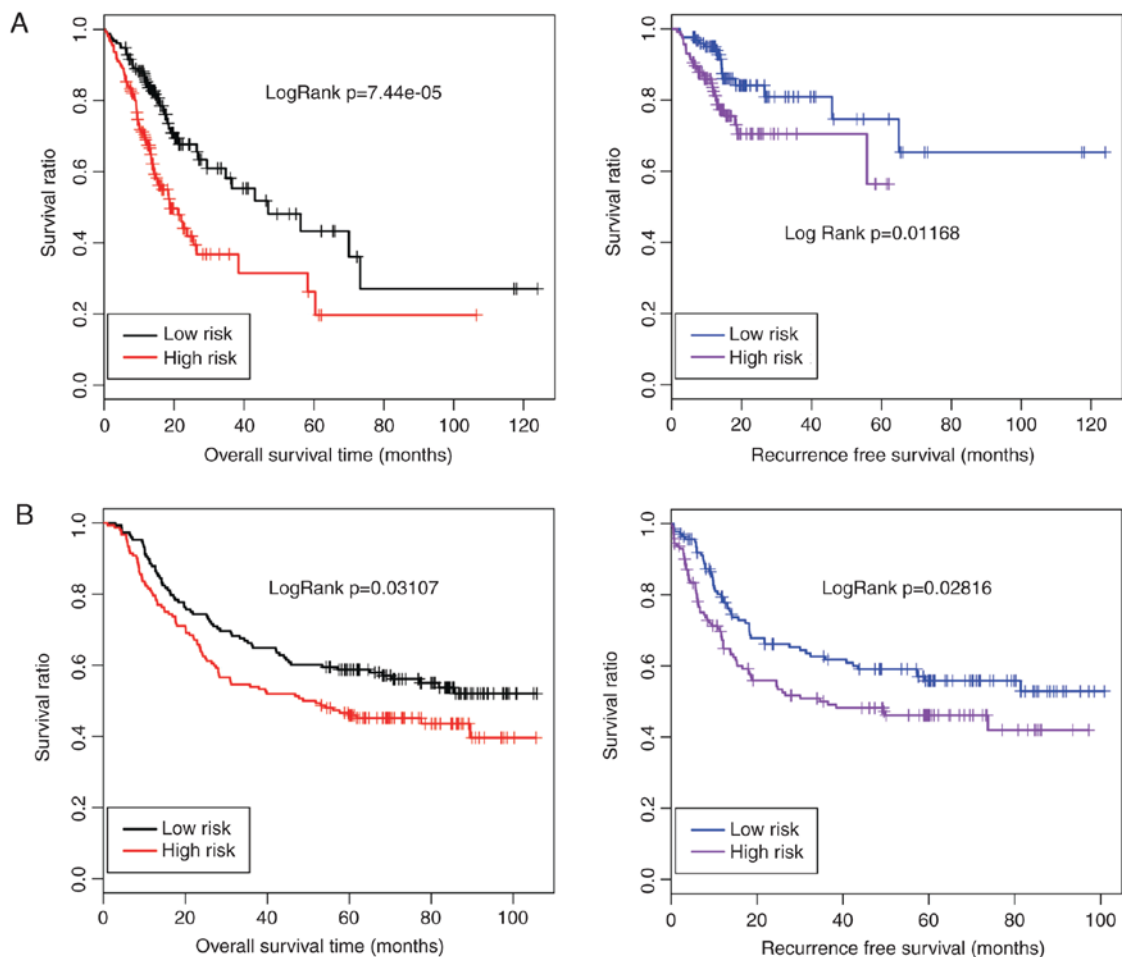


Figure 7. Kaplan-Meier curves for OS time (left) and RFS time (right) of patients in (A) TCGA and (B) GSE62254 sets. Patients of each set are divided by risk score into a high-risk group and a low-risk group. OS and RFS between two risk groups were analyzed and compared by Kaplan-Meier analysis and logRank test. TCGA, The Cancer Genome Atlas; GSE, Genomic Spatial Event; OS, overall survival; RFS, recurrence-free survival.

preserved modules and the list of consensus DERs (Fig. 6A). Among these overlapped RNAs, 50 were lncRNAs, of which 32 were included in the blue module, 14 in the brown module, 3 in the turquoise module and 1 in the yellow module (Fig. 6B).

*Development and validation of an lncRNAs-based risk scoring system.* Based on the expression of these overlapped lncRNAs in the TCGA set, the LASSO-based Cox-PH model identified an 11-lncRNA signature that was significantly asso-

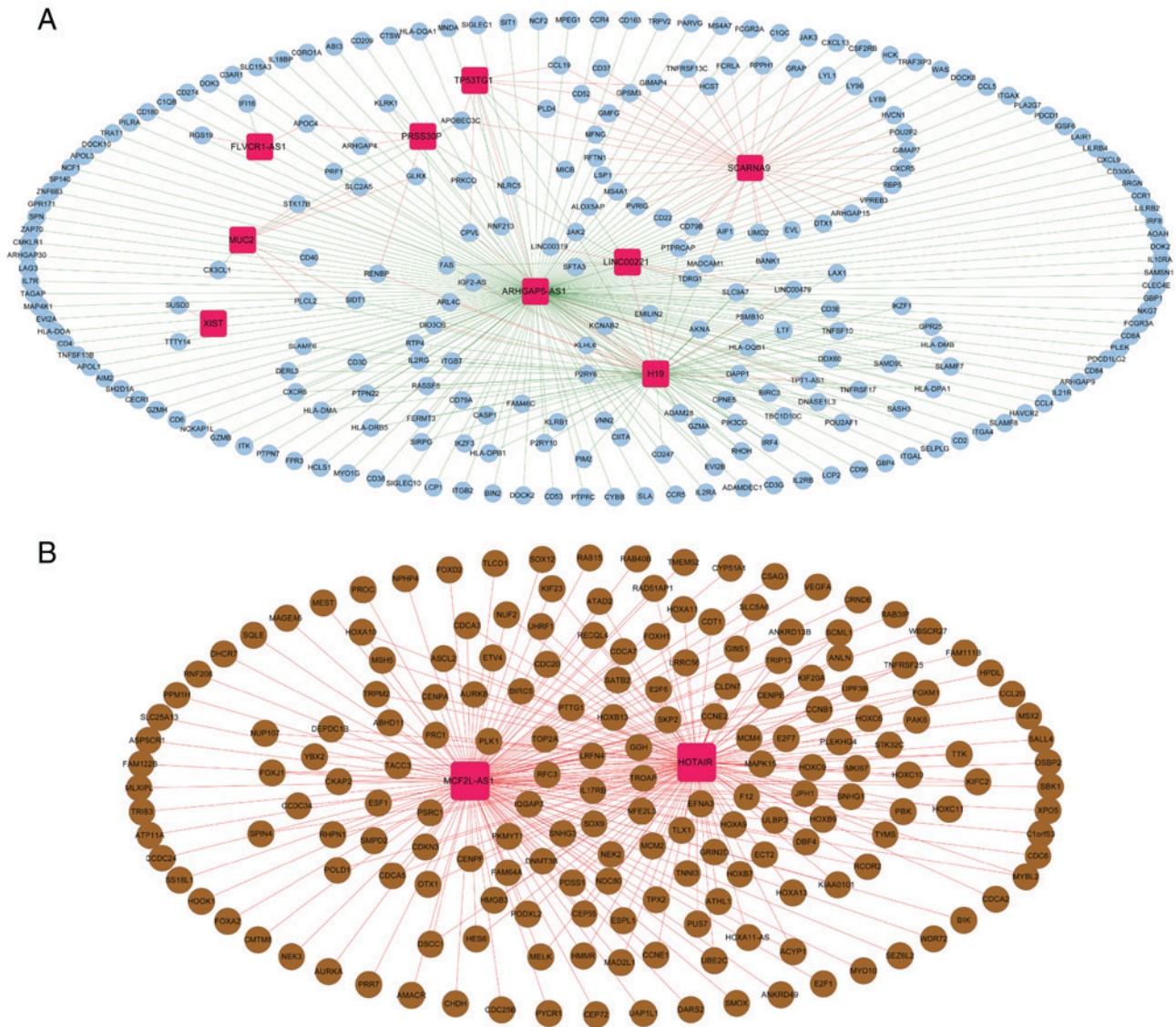


Figure 8. Constructed lncRNA-mRNA networks for prognostic lncRNAs. (A) lncRNA-mRNA network of 9 lncRNAs. The 9 lncRNAs are also contained in the WGCNA blue module. (B) lncRNA-mRNA network of 2 lncRNAs. The lncRNAs are also contained in the WGCNA brown module. Each red square module stands for an lncRNA. Each round node stands for an mRNA. A link between two nodes reveals positive (red link) or negative (green link) correlation between an lncRNA and an mRNA. Inc, long non-coding; WGCNA, weighted correlation network analysis.

ciated with survival rate based on the optimal lambda value (19.70021). This signature consisted of FLVCR1-AS1, H19, LINC00221, MUC2, RSS30P, SCARNA9, TP53TG1, XIST, ARHGAP5-AS1, HOTAIR and MCF2L-AS1 (Table V). lncRNA signature-based risk score was calculated using the following formula:

$$\text{Risk score} = 0.012437 \times \text{Exp}_{\text{ARHGAP5-AS1}} + (-0.11914) \times \text{Exp}_{\text{FLVCR1-AS1}} + 0.917082 \times \text{Exp}_{\text{H19}} + (-0.49726) \times \text{Exp}_{\text{HOTAIR}} + 1.179896 \times \text{Exp}_{\text{LINC00221}} + (-0.70093) \times \text{Exp}_{\text{MCF2L-AS1}} + (-0.09017) \times \text{Exp}_{\text{MUC2}} + 0.257189 \times \text{Exp}_{\text{PRSS30P}} + (-0.86146) \times \text{Exp}_{\text{SCARNA9}} + 0.149341 \times \text{Exp}_{\text{TP53TG1}} + (-0.92352) \times \text{Exp}_{\text{XIST}}$$

Risk score was calculated for each patient. All patients in the TCGA set were split into a high-risk group and a low-risk group with the median risk score as the cutoff. Patients in the high-risk group (n=156) demonstrated significantly shorter OS time (15.56±13.15 months vs. 21.23±19.99, logRank P=7.44x10<sup>-5</sup>)

and RFS time (15.76±11.51 months vs. 21.72±21.03, logRank P=0.0117) compared with the patients in the low-risk group (n=155, Fig. 7A). Prognostic performance of this 11-lncRNA signature-based risk scoring system was tested in an independent set (GSE62254). All 300 patients in GSE62254 were divided into a high-risk group (n=150) and a low-risk group (n=150) by risk score. Similarly, OS time (54.79±31.83 months vs. 46.40±31.83, logRank P=0.0311) and RFS time (37.45±31.08 months vs. 29.99±28.11, logRank P=0.0282) were markedly elongated in the low-risk group relative to the high-risk group (Fig. 7B).

*Function analysis of the 11-lncRNA signature.* Among the 11 signature lncRNAs, 9 lncRNAs (FLVCR1-AS1, H19, LINC00221, MUC2, RSS30P, SCARNA9, TP53TG1, XIST and ARHGAP5-AS1) were involved in the blue module, whereas another 2 lncRNAs (HOTAIR and MCF2L-AS1) were present in the brown module. Correlations between the 9 lncRNAs in the blue module and mRNAs revealed by the



Table VI. Significant GO terms and KEGG pathways for the genes in the constructed lncRNA-mRNA network of nine prognostic lncRNAs involved in the blue module.

GO category	Term	Count	Genes	FDR
Biology process	Immune response	80	MICB, CD8A, LY86, HLA-DMB, HLA-DMA, C1QC, PDCD1, CD96, SH2D1A, CLEC4E, MS4A1, LTF, FAS, FCGR3A, SPN, CIITA, LAIR1, POU2AF1, SIT1, NCF2, GZMA, NCF1, LY96, CMKLR1, TNFRSF17, WAS, HLA-DQA1, PDCD1LG2, TRAT1, CTSW, IGSF6, C1QB, LILRB2, IL18BP, CCR5, TNFSF13B, CCR4, LAX1, LILRB4, HLA-DPA1, MADCAM1, GBP4, LCP1, GBP1, LCP2, HLA-DQB1, PSMB10, ITGAL, CCR1, GPSM3, CXCL9, CX3CL1, IL7R, CCL5, CCL4, POU2F2, ZAP70, HLA-DRB5, IL2RG, CD4, HLA-DPB1, HLA-DOA, PTPRC, IL2RA, TNFRSF13C, CCL19, SLAMF7, CD180, AIM2, CORO1A, TNFSF10, CYBB, APOL1, CD300A, CXCL13, CD209, IRF8, CD274, CD79B, CD79A	5.39x10 <sup>-49</sup>
		30	KLRK1, IL7R, HLA-DMA, CD2, ZAP70, CD4, IL2RG, FAS, HLA-DOA, LAG3, SPN, PTPRC, SIT1, IL2RA, IKZF1, PLEK, CD3E, TNFRSF13C, CD40, PDCD1LG2, CD38, PRKCQ, CORO1A, SIRPG, TNFSF13B, LAX1, CD274, JAK2, IRF4, SASH3	6.22x10 <sup>-20</sup>
	Regulation of cell activation	28	KLRK1, IL7R, HLA-DMA, CD2, ZAP70, CD4, IL2RG, FAS, HLA-DOA, LAG3, SPN, PTPRC, SIT1, IL2RA, IKZF1, CD3E, TNFRSF13C, CD40, PDCD1LG2, CD38, PRKCQ, CORO1A, SIRPG, TNFSF13B, LAX1, CD274, IRF4, SASH3	1.53x10 <sup>-19</sup>
	Lymphocyte activation	31	ITGAL, MICB, CD8A, IL21R, KLRK1, PTPN22, IL7R, HLA-DMA, DOCK2, CXCR5, ZAP70, MS4A1, CD2, CD4, FAS, SPN, RHOH, PTPRC, CD3G, CD3D, IKZF1, CD3E, SLAMF7, ITGA4, CD40, WAS, LAX1, CD79A, IRF4, BANK1, LCP1	1.83x10 <sup>-19</sup>
	Positive regulation of immune system process	33	C3AR1, MICB, CD247, KLRK1, PTPN22, IL7R, C1QC, HLA-DMA, SH2D1A, CD2, ZAP70, CD4, IL2RG, LAG3, SPN, PTPRC, IL2RA, IKZF1, CD3E, TNFRSF13C, CD40, PDCD1LG2, TRAT1, CD38, PRKCQ, CIQB, CORO1A, CD37, SIRPG, TNFSF13B, LAX1, CD79A, SASH3	2.32x10 <sup>-19</sup>
	Leukocyte activation	33	ITGAL, MICB, CD8A, IL21R, KLRK1, PTPN22, CX3CL1, IL7R, HLA-DMA, DOCK2, CXCR5, ZAP70, MS4A1, CD2, CD4, FAS, SPN, RHOH, PTPRC, CD3G, CD3D, IKZF1, CD3E, SLAMF7, ITGA4, CD40, WAS, LAX1, CD79A, IRF4, BANK1, LCP1, LCP2	3.92x10 <sup>-19</sup>
	Regulation of T cell activation	25	PTPRC, SIT1, IL2RA, IKZF1, CD3E, TNFRSF13C, KLRK1, IL7R, HLA-DMA, PDCD1LG2, PRKCQ, CORO1A, SIRPG, TNFSF13B, LAX1, CD274, ZAP70, CD2, CD4, IL2RG, IRF4, HLA-DOA, SPN, LAG3, SASH3	2.21x10 <sup>-18</sup>
	Defense response	47	C3AR1, ITGAL, PRF1, AIF1, CCR1, LY86, CXCL9, ITGB2, CX3CL1, CCL5, PTPRCAP, CCL4, C1QC, SH2D1A, AOAH, LTF, SPN, CIITA, ITK, PTPRC, IL2RA, NCF2, NCF1, LY96, HCK, CCL19, CD40, SLAMF7, WAS, CD180, SPI40, TRAT1, CD163, LSP1, CD84, APOL3, SIGLEC1, C1QB, LILRB2, CORO1A, CYBB, APOL1, CCR5, CCR4, CXCL13, MND4, PLA2G7	3.44x10 <sup>-18</sup>
	Regulation of leukocyte activation	28	KLRK1, IL7R, HLA-DMA, CD2, ZAP70, CD4, IL2RG, FAS, HLA-DOA, LAG3, SPN, PTPRC, SIT1, IL2RA, IKZF1, CD3E, TNFRSF13C, CD40, PDCD1LG2, CD38, PRKCQ, CORO1A, SIRPG, TNFSF13B, LAX1, CD274, IRF4, SASH3	3.64x10 <sup>-18</sup>
	Cell activation	34	ITGAL, MICB, CD8A, IL21R, KLRK1, PTPN22, CX3CL1, IL7R, HLA-DMA, DOCK2, CXCR5, ZAP70, MS4A1, CD2, CD4, FAS, SPN, RHOH, PTPRC, CD3G, CD3D, PLEK, IKZF1, CD3E, SLAMF7, ITGA4, CD40, WAS, LAX1, CD79A, IRF4, BANK1, LCP1, LCP2	7.10x10 <sup>-18</sup>
	Positive regulation of cell activation	23	PTPRC, IL2RA, IKZF1, PLEK, CD3E, KLRK1, TNFRSF13C, CD40, IL7R, HLA-DMA, PDCD1LG2, CD38, PRKCQ, CORO1A, SIRPG, TNFSF13B, CD2, ZAP70, JAK2, CD4, IL2RG, SASH3, SPN	2.55x10 <sup>-16</sup>

Table VI. Continued.

GO category	Term	Count	Genes	FDR
Positive regulation of leukocyte activation		21	PTPRC, IL2RA, IKZF1, CD3E, KLRK1, TNFRSF13C, CD40, IL7R, HLA-DMA, PDCD1LG2, CD38, PRKCQ, CORO1A, SIRPG, TNFSF13B, CD2, ZAP70, CD4, IL2RG, SASH3, SPN	3.41x10 <sup>-14</sup>
		20	PTPRC, IL2RA, IKZF1, CD3E, KLRK1, TNFRSF13C, CD40, IL7R, HLA-DMA, PDCD1LG2, CD38, PRKCQ, CORO1A, SIRPG, TNFSF13B, ZAP70, CD4, IL2RG, SASH3, SPN	1.78x10 <sup>-13</sup>
Positive regulation of lymphocyte activation		21	ITGAL, PTPRC, MICB, CD3G, CD3D, IKZF1, CD8A, CD3E, PTPN22, IL7R, HLA-DMA, WAS, DOCK2, ZAP70, CD2, CD4, FAS, IRF4, LCP1, SPN, RHOH	1.09x10 <sup>-12</sup>
		24	PTPRC, CD3D, PLEK, IKZF1, CD8A, CD3E, HCLS1, PTPN22, ITGA4, IFI16, IL7R, HLA-DMA, DOCK2, CXCR5, CXCL13, IRF8, ZAP70, JAK2, CD4, FAS, CD79A, IRF4, SPN, RHOH	2.97x10 <sup>-09</sup>
Hemopoietic or lymphoid organ development		26	ITGAL, C3AR1, AIF1, LY86, CCR1, CXCL9, ITGB2, CCL5, C1QC, CCL4, AOH, CIITA, IL2RA, LY96, CCL19, CD40, CD180, CD163, C1QB, SIGLEC1, APOL3, CYBB, CCR5, CXCL13, CCR4, PLA2G7	6.83x10 <sup>-09</sup>
		24	PTPRC, CD3D, PLEK, IKZF1, CD8A, CD3E, HCLS1, PTPN22, ITGA4, IFI16, IL7R, HLA-DMA, DOCK2, CXCR5, CXCL13, IRF8, ZAP70, JAK2, CD4, FAS, CD79A, IRF4, SPN, RHOH	1.02x10 <sup>-08</sup>
Immune system development		22	PTPRC, CD3D, PLEK, IKZF1, CD8A, CD3E, HCLS1, PTPN22, ITGA4, IFI16, IL7R, HLA-DMA, DOCK2, IRF8, ZAP70, JAK2, CD4, FAS, CD79A, IRF4, SPN, RHOH	2.60x10 <sup>-08</sup>
		22	C3AR1, PTPRC, MICB, CD3E, CD247, KLRK1, TNFRSF13C, PTPN22, CX3CL1, CCL5, HLA-DMA, C1QC, C1QB, SH2D1A, TNFSF13B, LAX1, CCR4, ZAP70, JAK2, CD79A, SASH3, LAG3	2.60x10 <sup>-08</sup>
Response to stimulus		30	C3AR1, ITGAL, AIF1, LY86, CCR1, CXCL9, ITGB2, CCL5, C1QC, CCL4, AOH, CIITA, IL2RA, PLEK, LY96, CCL19, CD40, WAS, CD180, CD163, APOL3, PRKCQ, C1QB, SIGLEC1, CYBB, CCR5, CCR4, CXCL13, PLA2G7, JAK2	4.82x10 <sup>-07</sup>
		56	MICB, CD8A, PTPN22, CXCR5, CXCR6, SPN, LAG3, KLRB1, PIK3CG, CD3G, CD3D, LY96, CMKLR1, CD3E, GPR171, CD40, IGSF6, LILRB2, DOK2, CCR5, CCR4, LAX1, LCP2, C3AR1, ITGAL, CCR1, CD247, KLRK1, CXCL9, FPR3, ITGB2, IL7R, CCL5, P2RY6, ITGAX, ITGB7, GPR25, ZAP70, CD2, CD4, PTPRC, IL2RA, PLEK, DTX1, CCL19, RGS19, EVL, ITGA4, BIRC3, P2RY10, CD274, CD79B, JAK2, JAK3, CD79A, ADAMDEC1	4.54x10 <sup>-05</sup>
Cell surface receptor linked signal transduction		29	ITGAL, CCRI, FERMT3, ITGB2, CX3CL1, CCL5, CCL4, CD96, ITGAX, ITGB7, CD2, CD22, CD4, CD6, SELPLG, PARVG, PTPRC, PLEK, SIGLEC10, ITGA4, SLAMF7, EMILIN2, CD84, SIGLEC1, CORO1A, SIRPG, CD300A, CD209, MADCAM1	8.34x10 <sup>-04</sup>
		29	ITGAL, CCRI, FERMT3, ITGB2, CX3CL1, CCL5, CCL4, CD96, ITGAX, ITGB7, CD2, CD22, CD4, CD6, SELPLG, PARVG, PTPRC, PLEK, SIGLEC10, ITGA4, SLAMF7, EMILIN2, CD84, SIGLEC1, CORO1A, SIRPG, CD300A, CD209, MADCAM1	8.58x10 <sup>-04</sup>
Biological adhesion		26	HLA-DQB1, ITGAL, PTPRC, CD8A, ITGB2, CD40, ITGA4, HLA-DMB, HLA-DMA, PDCD1, HLA-DQA1, PDCD1LG2, SIGLEC1, ITGB7, CD274, CD2, CD22, HLA-DRB5, CD4, HLA-DPA1, MADCAM1, HLA-DPBI, HLA-DOA, CD6, SELPLG, SPN	6.37x10 <sup>-15</sup>
		12	HLA-DQA1, PRF1, HLA-DRB5, GZMB, HLA-DPA1, HLA-DPB1, FAS, CD40, HLA-DMB, HLA-DOA, HLA-DMA, HLA-DQA1	8.68x10 <sup>-08</sup>

Table VI. Continued.

GO category	Term	Count	Genes	FDR
Cytokine-cytokine receptor interaction	Graft vs.host disease	24	IL2RB, IL2RA, CCR1, IL21R, TNFRSF13C, CXCL9, TNFRSF17, CCL19, CD40, CX3CL1, IL7R, CCCL5, CCL4, TNFSF10, TNFSF13B, CXCR5, CCR5, CCR4, CXCL13, IL10RA, CXCR6, CSF2RB, IL2RG, FAS	2.80x10 <sup>-06</sup>
		11	HLA-DQB1, PRF1, HLA-DRB5, GZMB, HLA-DPA1, HLA-DPB1, FAS, HLA-DMB, HLA-DOA, HLA-DMA, HLA-DQA1	4.48x10 <sup>-06</sup>
Chemokine signaling pathway	Natural killer cell mediated cytotoxicity	19	PIK3CG, ITK, NCF1, HCK, CCR1, CXCL9, CCL19, CX3CL1, CCL5, CCL4, WAS, DOCK2, CXCR5, CCR5, CCR4, CXCL13, CXCR6, JAK2, JAK3	4.64x10 <sup>-05</sup>
		15	PIK3CG, PRF1, ITGAL, MICB, CD247, KLRK1, GZMB, ITGB2, HCST, SH2D1A, TNFSF10, ZAP70, FAS, FCGR3A, LCP2	5.69x10 <sup>-04</sup>
T cell receptor signaling pathway	Antigen processing and presentation	13	PIK3CG, ITK, PRKCQ, PTPRC, CD3G, CD8A, CD3D, CD3E, CD247, ZAP70, CD4, PDCD1, LCP2	2.20x10 <sup>-03</sup>
		11	HLA-DQB1, CIITA, CD8A, HLA-DRB5, CD4, HLA-DPA1, HLA-DPB1, HLA-DMB, HLA-DOA, HLA-DMA, HLA-DQA1	7.92x10 <sup>-03</sup>

GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; Inc, long non-coding; FDR, false discovery rate.

WGCNA were used to construct an lncRNA-mRNA network (Fig. 8A). Similarly, another lncRNA-mRNA network was built for the 2 lncRNAs (HOTAIR and MCF2L-AS1), in the brown module (Fig. 8B). The genes in the lncRNA-mRNA network that correlated with the 9 prognostic lncRNAs in the blue module were significantly associated with 23 GO biological process terms (including immune response, regulation of cell activation and regulation of lymphocyte activation) and 8 KEGG pathways (including cell adhesion molecules, allograft rejection and cytokine-cytokine receptor interaction; Table VI). The genes in the lncRNA-mRNA network that correlated with HOTAIR and MCF2L-AS1 were mainly associated with the cell cycle phase, cell cycle and mitotic cell cycle. In addition, 4 KEGG pathways were enriched for the genes in this lncRNA-mRNA network including cell cycle, DNA replication, progesterone-mediated oocyte maturation and steroid biosynthesis pathways (Table VII).

## Discussion

A growing number of studies have demonstrated that aberrantly expressed lncRNAs are implicated in GC tumorigenesis and progression (30,31). Nonetheless, the prognostic significance of lncRNAs in GC remains to be elucidated. Based on the common RNAs data and corresponding clinical information of GC patients and normal controls which were obtained through data mining in NCBI GEO, EBI ArrayExpress and TCGA, a 11-lncRNA prognostic signature was identified by a series of bioinformatics analyses featuring WGCNA, the MetaDE method and a LASSO-based Cox-PH model. Furthermore, it was identified that patients could be classified into a high-risk group and a low-risk group by the risk score based on the 11-lncRNA signature in the training set, with noticeable separations being observed in the Kaplan-Meier curves between the 2 groups. The high-risk group exhibited significantly longer OS time and PFS time compared with the low-risk group. The predictive ability of risk score was confirmed in an independent set. Therefore, the present study demonstrated that the 11-lncRNA signature has the potential for assessing survival rate of GC patients.

The 11-lncRNA signature determined in the study was comprised of FLVCR1-AS1, H19, LINC00221, MUC2, PRSS30P, SCARNA9, TP53TG1, XIST, ARHGAP5-AS1, HOTAIR and MCF2L-AS1. Among these lncRNAs, H19 is identified to be upregulated in plasma of GC patients and is proposed as a diagnostic biomarker (8). Increasing evidence also demonstrates that H19 upregulation promotes GC proliferation, migration and invasion (9,10). It has been established that MUC2 is associated with outcome of GC patients (32). lncRNA X inactive specific transcript (XIST) encoded by XIST gene acts as a regulator of X inactivation in mammals (33). Chen *et al* (34) observed upregulated XIST in GC tissue and identified that this lncRNA serves a regulatory role in GC progression via microRNA (miR)-101 and its direct target polycomb group protein enhancer of zeste homolog 2. HOTAIR transcribed from the HOXC locus is identified to be overexpressed in GC, which is a characteristic molecular alteration of GC (35). Furthermore, there is evidence that HOTAIR functions as a GC oncogene through regulating the expression of human epithelial growth factor receptor 2 by competing with miR-331-3p (12).

Table VII. Significant GO terms and KEGG pathways for the genes in the constructed lncRNA-mRNA network of two prognostic lncRNAs in the brown module.

GO category	Term	Count	Genes	FDR
Biology process	Cell cycle phase	40	E2F1, KIF23, PRC1, NEK3, NEK2, DBF4, TTK, PKMYT1, ANLN, AURKA, PTTG1, CEP55, AURKB, CCNE1, CDCA2, CDCA5, TRIP13, CDCA3, CDC6, MKI67, MSH5, TPX2, SKP2, NUF2, CENPE, CDC20, BIRC5, CENPE, NDC80, ESPL1, PBK, CDKN3, UBE2C, TACC3, CDC25B, CCNB1, MAD2L1, PLK1, POLD1, DSICC1	2.14x10 <sup>-22</sup>
	Cell cycle	50	E2F1, KIF23, CEP72, PRC1, DBF4, E2F7, TTK, PKMYT1, AURKA, PTTG1, AURKB, CDT1, CCNE2, CCNE1, CDCA2, CDCA5, CDCA3, CDC6, SKP2, TPX2, ESPL1, MCM2, PBK, TACC3, UBE2C, UHRF1, MAD2L1, DSICC1, NEK3, NEK2, FOXM1, ANLN, CEP55, CENPA, TRIP13, CKAP2, MKI67, MSH5, PSRC1, NUF2, CENPE, BIRC5, NDC80, CENPE, CDC20, CDKN3, CDC25B, CCNB1, PLK1, POLD1	2.83x10 <sup>-21</sup>
	Mitotic cell cycle	37	KIF23, E2F1, PRC1, NEK3, NEK2, DBF4, TTK, PKMYT1, ANLN, AURKA, PTTG1, CEP55, AURKB, CCNE1, CENPA, CDCA2, CDCA5, CDCA3, CDC6, TPX2, SKP2, NUF2, CENPE, CDC20, BIRC5, CENPE, NDC80, ESPL1, PBK, CDKN3, UBE2C, CDC25B, CCNB1, MAD2L1, PLK1, POLD1, DSICC1	6.89x10 <sup>-21</sup>
	Cell cycle process	42	E2F1, KIF23, CEP72, PRC1, NEK3, NEK2, DBF4, TTK, PKMYT1, ANLN, AURKA, PTTG1, AURKB, CEP55, CCNE1, CENPA, CDCA2, CDCA5, CDCA3, TRIP13, CDC6, MKI67, MSH5, TPX2, SKP2, NUF2, CENPE, CDC20, BIRC5, CENPE, NDC80, ESPL1, PBK, CDKN3, UBE2C, TACC3, CDC25B, CCNB1, MAD2L1, PLK1, POLD1, DSICC1	2.26x10 <sup>-19</sup>
	M phase	34	KIF23, PRC1, NEK3, NEK2, TTK, PKMYT1, ANLN, AURKA, PTTG1, CEP55, AURKB, CDCA2, CDCA5, TRIP13, CDCA3, CDC6, MKI67, MSH5, TPX2, NUF2, CENPE, CDC20, BIRC5, CENPE, NDC80, ESPL1, PBK, UBE2C, TACC3, CDC25B, CCNB1, MAD2L1, PLK1, DSICC1	2.68x10 <sup>-19</sup>
	Mitosis	28	KIF23, NEK3, NEK2, PKMYT1, AURKA, ANLN, CEP55, AURKB, PTTG1, CDCA2, CDCA5, CDCA3, CDC6, TPX2, NUF2, CENPE, BIRC5, CENPE, NDC80, ESPL1, CDC20, PBK, UBE2C, CDC25B, CCNB1, MAD2L1, PLK1, DSICC1	1.39x10 <sup>-17</sup>
	Nuclear division	28	KIF23, NEK3, NEK2, PKMYT1, AURKA, ANLN, CEP55, AURKB, PTTG1, CDCA2, CDCA5, CDCA3, CDC6, TPX2, NUF2, CENPE, BIRC5, CENPE, NDC80, ESPL1, CDC20, PBK, UBE2C, CDC25B, CCNB1, MAD2L1, PLK1, DSICC1	1.39x10 <sup>-17</sup>
	M phase of mitotic cell cycle	28	KIF23, NEK3, NEK2, PKMYT1, AURKA, ANLN, CEP55, AURKB, PTTG1, CDCA2, CDCA5, CDCA3, CDC6, TPX2, NUF2, CENPE, BIRC5, CENPE, NDC80, ESPL1, CDC20, PBK, UBE2C, CDC25B, CCNB1, MAD2L1, PLK1, DSICC1	2.25x10 <sup>-17</sup>
	Organelle fission	28	KIF23, NEK3, NEK2, PKMYT1, AURKA, ANLN, CEP55, AURKB, PTTG1, CDCA2, CDCA5, CDCA3, CDC6, TPX2, NUF2, CENPE, BIRC5, CENPE, NDC80, ESPL1, CDC20, PBK, UBE2C, CDC25B, CCNB1, MAD2L1, PLK1, DSICC1	4.04x10 <sup>-17</sup>
	Cell division	26	KIF23, PRC1, NEK3, NEK2, ANLN, CEP55, PTTG1, AURKB, CCNE2, CCNE1, CDCA2, CDCA5, CDCA3, CDC6, NUF2, CENPE, BIRC5, CDC20, CENPE, NDC80, ESPL1, UBE2C, CDC25B, CCNB1, MAD2L1, PLK1	3.42x10 <sup>-12</sup>
	Regulation of cell cycle	19	E2F1, CDC6, HOXA13, NEK2, SKP2, CENPE, TTK, PKMYT1, ESPL1, CENPE, ANLN, BIRC5, TACC3, UBE2C, CDKN3, CDT1, CCNE2, CCNB1, MAD2L1	4.79x10 <sup>-05</sup>
	Microtubule-based process	16	KIFC2, KIF23, CEP72, PRC1, NEK2, PSRC1, TTK, ESPL1, AURKA, NDC80, CENPE, TACC3, UBE2C, HOOK1, CENPA, KIF20A	2.40x10 <sup>-04</sup>

Table VII. Continued.

GO category	Term	Count	Genes	FDR
KEGG pathway	Pattern specification process	15	SATB2, FOXA2, FOXJ1, OTX1, HOXA11, HOXC6, FOXH1, HOXC10, HOXC9, HOXC11, HOXB7, VEGFA, HOXA10, HOXA9, HOXB9	$2.78 \times 10^{-03}$
	DNA metabolic process	20	RECQL4, GINS1, CDC6, RAD51AP1, DBF4, MSH5, CENPF, MCM2, PTTG1, MCM4, CDT1, CCNE2, TYMS, UHRF1, RFC3, POLD1, DNMT3B, TOP2A, TRIP13, DSSCC1	$5.77 \times 10^{-03}$
KEGG pathway	Cell cycle	18	E2F1, CDC6, E2F5, DBF4, SKP2, PKMYT1, TTK, CDC20, ESPL1, MCM2, PTTG1, MCM4, CDC25B, CCNE2, CCNB1, CCNE1, MAD2L1, PLK1	$1.01 \times 10^{-12}$
	DNA replication	4	RFC3, POLD1, MCM2, MCM4	$5.38 \times 10^{-03}$
	Progesterone-mediated oocyte maturation	5	CCNB1, MAD2L1, PLK1, PKMYT1, CDC25B	$1.04 \times 10^{-02}$
KEGG pathway	Steroid biosynthesis	3	CYP51A1, SQLE, DHCR7	$1.22 \times 10^{-02}$

GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; Inc, long non-coding; FDR, false discovery rate.

Investigation of lncRNA profiles in human cancer remains to be performed. Apart from H19, MUC2, XIST and HOTAIR, other prognostic lncRNAs have not been identified in GC. FLVCR1-AS1 has been reported in lung adenocarcinoma by a study based on an miR-lncRNA-mRNA network (36). TP53TG1 is a critical lncRNA responsible for correct response of p53 to DNA damage and acts as a tumor suppressor (37). There is evidence that TP53TG1 expression is elevated in human glioma tissue and TP53TG1 under glucose deprivation may promote cell proliferation and migration by influencing the expression of glucose metabolism associated genes in glioma (38). LINC00221 has been reported to be aberrantly expressed in bladder cancer (39). Li *et al* (40) noted that PRSS30P is upregulated in lung adenocarcinoma. SCARNA9 is observed to be overexpressed in breast cancer cells on exposure to cadmium (41). However, ARHGAP5-AS1 and MCF2L-AS1 are rarely studied in cancer. In future studies, the expression levels of ARHGAP5-AS1 and MCF2L-AS1 will be investigated in clinical samples of GC patients since the prognostic value of these lncRNAs was observed for GC.

Correlations between the critical lncRNAs and mRNAs revealed by the WGCNA were used to construct lncRNA-mRNA networks. In order to investigate the molecular mechanisms of the 11 prognostic lncRNAs in GC, GO function and KEGG pathway enrichment analysis were performed for the genes in the construct lncRNA-mRNA networks. The results demonstrated that the genes correlated with the 9 lncRNAs in the blue module (FLVCR1-AS1, H19, LINC00221, MUC2, RSS30P, SCARNA9, TP53TG1, XIST and ARHGAP5-AS1) were associated with the immune response, regulation of cell activation, regulation of lymphocyte activation and cytokine-cytokine receptor interaction. These results suggested that these 9 lncRNAs may serve important roles in the pathogenesis of GC by regulating their associated genes to affect the immune and inflammatory responses. The genes associated with the 2 lncRNAs (HOTAIR and MCF2L-AS1) in the brown module were revealed to be implicated in cell cycle regulation. This indicated that ARHGAP5-AS1 and MCF2L-AS1 may also be critical in the pathogenesis of GC by regulating their associated genes to influence the cell cycle. A growing body of evidence demonstrates the important roles of inflammation, immune and dysregulated cell cycle control in tumor growth and progression (42-44). Therefore, it can be concluded that the 11 critical lncRNAs may participate in the development and progression of GC by regulating their correlated genes to influence the immune response, inflammatory response and cell cycle.

Based on bioinformatics analysis of existing gene expression data from NCBI GEO, EBI ArrayExpress and TCGA, the present study identified an 11-lncRNA signature that could be used for predicting survival rate of GC patients. These 11 critical lncRNAs may participate in the pathogenesis of GC by regulating their correlated genes that are associated with the immune response, inflammatory response and cell cycle. It is hoped that the present study may contribute to an improved understanding of the pathogenesis involved with lncRNAs in GC development and progression. Validation of this 11-lncRNA signature in large cohorts of GC patients and clinical trials are also essential in further investigation.

## Acknowledgements

Not applicable.

## Funding

No funding was received.

## Availability of data and materials

The datasets analyzed during the present study are available from the corresponding author on reasonable request.

## Authors' contributions

YZ and HL performed data analyses and wrote the manuscript. WZ and YC contributed significantly to the data analyses and critical revision of the manuscript. GH and WB conceived and designed the study. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Patient consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## References

- IARC: World Cancer Report 2014. Stewart BW and Wild CP (eds). World Health Organization, Geneva, 2015.
- Orditura M, Galizia G, Sforza V, Gambardella V, Fabozzi A, Laterza MM, Andreozzi F, Ventriglia J, Savastano B, Mabilia A, *et al*: Treatment of gastric cancer. *World J Gastroenterol* 20: 1635-1649, 2014.
- Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, Jemal A, Yu XQ and He J: Cancer statistics in china, 2015. *CA Cancer J Clin* 66: 115-132, 2016.
- Sun Z, Wang ZN, Zhu Z, Xu YY, Xu Y, Huang BJ, Zhu GL and Xu HM: Evaluation of the seventh edition of american joint committee on cancer TNM staging system for gastric cancer: Results from a chinese monoinstitutional study. *Ann Surg Oncol* 19: 1918-1927, 2012.
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, *et al*: Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553-560, 2007.
- Liz J and Esteller M: lncRNAs and microRNAs with a role in cancer development. *Biochim Biophys Acta* 1859: 169-176, 2016.
- Evans JR, Feng FY and Chinnaiyan AM: The bright side of dark matter: lncRNAs in cancer. *J Clin Invest* 126: 2775-2782, 2016.
- Zhou X, Yin C, Dang Y, Ye F and Zhang G: Identification of the long non-coding RNA H19 in plasma as a novel biomarker for diagnosis of gastric cancer. *Sci Rep* 5: 11516, 2015.
- Zhou X, Ye F, Yin C, Zhuang Y, Yue G and Zhang G: The interaction between miR-141 and lncRNA-H19 in regulating cell proliferation and migration in gastric cancer. *Cell Physiol Biochem* 36: 1440-1452, 2015.
- Li H, Yu B, Li J, Su L, Yan M, Zhu Z and Liu B: Overexpression of lncRNA H19 enhances carcinogenesis and metastasis of gastric cancer. *Oncotarget* 5: 2318-2329, 2014.
- Pan W, Liu L, Wei J, Ge Y, Zhang J, Chen H, Zhou L, Yuan Q, Zhou C and Yang M: A functional lncRNA HOTAIR genetic variant contributes to gastric cancer susceptibility. *Mol Carcinog* 55: 90-96, 2016.
- Liu XH, Sun M, Nie FQ, Ge YB, Zhang EB, Yin DD, Kong R, Xia R, Lu KH, Li JH, *et al*: lnc RNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. *Mol Cancer* 13: 92, 2014.
- Zhang EB, Kong R, Yin DD, You LH, Sun M, Han L, Xu TP, Xia R, Yang JS, De W and Chen Jf: Long noncoding RNA ANRIL indicates a poor prognosis of gastric cancer and promotes tumor growth by epigenetically silencing of miR-99a/miR-449a. *Oncotarget* 5: 2276-2292, 2014.
- Miao Y, Sui J, Xu SY, Liang GY, Pu YP and Yin LH: Comprehensive analysis of a novel four-lncRNA signature as a prognostic biomarker for human gastric cancer. *Oncotarget* 8: 75007-75024, 2017.
- Danford T, Rolfe A and Gifford D: GSE: A comprehensive database system for the representation, retrieval, and analysis of microarray data. *Pac Symp Biocomput*: 539-550, 2008.
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W and Smyth GK: Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47, 2015.
- Zhou M, Guo M, He D, Wang X, Cui Y, Yang H, Hao D and Sun J: A potential signature of eight long non-coding RNAs predicts survival in patients with non-small cell lung cancer. *J Transl Med* 13: 231, 2015.
- Zhou M, Xu W, Yue X, Zhao H, Wang Z, Shi H, Cheng L and Sun J: Relapse-related long non-coding RNA signature to improve prognosis prediction of lung adenocarcinoma. *Oncotarget* 7: 29720-29738, 2016.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, *et al*: Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948, 2007.
- Zhai X, Xue Q, Liu Q, Guo Y and Chen Z: Colon cancer recurrence-associated genes revealed by WGCNA co-expression network analysis. *Mol Med Rep* 16: 6499-6505, 2017.
- Langfelder P and Horvath S: WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9: 559, 2008.
- Qi C, Hong L, Cheng Z and Yin Q: Identification of metastasis-associated genes in colorectal cancer using metaDE and survival analysis. *Oncol Lett* 11: 568-574, 2016.
- Wang X, Kang DD, Shen K, Song C, Lu S, Chang LC, Liao SG, Huo Z, Tang S, Ding Y, *et al*: An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. *Bioinformatics* 28: 2534-2536, 2012.
- Goeman JJ: L1 penalized estimation in the Cox proportional hazards model. *Biom J* 52: 70-84, 2010.
- Tibshirani R: The lasso method for variable selection in the Cox model. *Stat Med* 16: 385-395, 1997.
- Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, *et al*: Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* 21: 449-456, 2015.
- Parrish RS and Spencer HJ III: Effect of normalization on significance testing for oligonucleotide microarrays. *J Biopharm Stat* 14: 575-589, 2004.
- Huang da W, Sherman BT and Lempicki RA: Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1-13, 2009.
- Huang da W, Sherman BT and Lempicki RA: Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44-57, 2009.
- Li T, Mo X, Fu L, Xiao B and Guo J: Molecular mechanisms of long noncoding RNAs on gastric cancer. *Oncotarget* 7: 8601-8612, 2016.
- Sun M, Nie FQ, Wang ZX and De W: Involvement of lncRNA dysregulation in gastric cancer. *Histol Histopathol* 31: 33-39, 2016.
- Lee HS, Lee HK, Kim HS, Yang HK, Kim YI and Kim WH: MUC1, MUC2, MUC5AC, and MUC6 expressions in gastric carcinomas. *Cancer* 92: 1427-1434, 2001.
- Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R and Willard HF: A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349: 38-44, 1991.

34. Chen DL, Ju HQ, Lu YX, Chen LZ, Zeng ZL, Zhang DS, Luo HY, Wang F, Qiu MZ, Wang DS, *et al*: Long non-coding RNA XIST regulates gastric cancer progression by acting as a molecular sponge of miR-101 to modulate EZH2 expression. *J Exp Clin Cancer Res* 35: 142, 2016.
35. Endo H, Shiroki T, Nakagawa T, Yokoyama M, Tamai K, Yamanami H, Fujiya T, Sato I, Yamaguchi K, Tanaka N, *et al*: Enhanced expression of long non-coding RNA HOTAIR is associated with the development of gastric cancer. *PLoS One* 8: e77070, 2013.
36. Li DS, Ainiwaer JL, Sheyhiding I, Zhang Z and Zhang LW: Identification of key long non-coding RNAs as competing endogenous RNAs for miRNA-mRNA in lung adenocarcinoma. *Eur Rev Med Pharmacol Sci* 20: 2285-2295, 2016.
37. Diaz-Lagares A, Crujeiras AB, Lopez-Serra P, Soler M, Setien F, Goyal A, Sandoval J, Hashimoto Y, Martinez-Cardús A, Gomez A, *et al*: Epigenetic inactivation of the p53-induced long noncoding RNA TP53 target 1 in human cancer. *Proc Natl Acad Sci USA* 113: E7535-E7544, 2016.
38. Chen X, Gao Y, Li D, Cao Y and Hao B: LncRNA-TP53TG1 participated in the stress response under glucose deprivation in glioma. *J Cell Biochem* 118: 4897-4904, 2017.
39. Wang H, Niu L, Jiang S, Zhai J, Wang P, Kong F and Jin X: Comprehensive analysis of aberrantly expressed profiles of lncRNAs and miRNAs with associated ceRNA network in muscle-invasive bladder cancer. *Oncotarget* 7: 86174-86185, 2016.
40. Li J, Li P, Zhao W, Yang R, Chen S, Bai Y, Dun S, Chen X, Du Y, Wang Y, *et al*: Expression of long non-coding RNA DLX6-AS1 in lung adenocarcinoma. *Cancer Cell Int* 15: 48, 2015.
41. Lubovac-Pilav Z, Borràs DM, Ponce E and Louie MC: Using expression profiling to understand the effects of chronic cadmium exposure on MCF-7 breast cancer cells. *PLoS One* 8: e84646, 2013.
42. Candido J and Hagemann T: Cancer-related inflammation. *J Clin Immunol* 33 (Suppl 1): S79-S84, 2013.
43. Elinav E, Nowarski R, Thaiss CA, Hu B, Jin C and Flavell RA: Inflammation-induced cancer: Crosstalk between tumours, immune cells and microorganisms. *Nat Rev Cancer* 13: 759-771, 2013.
44. Evan GI and Vousden KH: Proliferation, cell cycle and apoptosis in cancer. *Nature* 411: 342-348, 2001.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.