

Identification of LncRNA Prognostic Biomarkers Associated with Copy Number Variants in Gastric Cancer

Anqiang Wang

Peking University Cancer Hospital & Institute

Jiahui Chen

Peking University Cancer Hospital & Institute

Kai Zhou

Peking University Cancer Hospital & Institute

Qifei He

Peking University Cancer Hospital & Institute

Ke Ji

Peking University Cancer Hospital & Institute

Xin Ji

Peking University Cancer Hospital & Institute

Xiaojiang Wu

Peking University Cancer Hospital & Institute

Ji Zhang

Peking University Cancer Hospital & Institute

Zhaode Bu (✉ buzhaode@cjcrn.org)

Peking University Cancer Hospital & Institute

Jiafu Ji (✉ jjiafu@hsc.pku.edu.cn)

Peking University Cancer Hospital & Institute <https://orcid.org/0000-0001-6878-5543>

Primary research

Keywords: gastric cancer, CNV, lncRNA, TCGA, prognostic

Posted Date: June 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-32433/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Backgrounds: Gastric cancer is one of the most common gastrointestinal carcinomas worldwide, with a poor prognosis. Prognosis prediction is very important in the treatment of gastric cancer. This study aimed to explore the prognostic value of lncRNA deregulated by copy number variants (CNVs) in gastric cancer.

Methods: Multi-omics cluster analysis was performed to identify subtypes on the prognosis associated coding genes, CNVs and methylation sites. We conducted survival analyses on median expression level for all identified lncRNAs. Finally, we constructed and validated a prognostic model on lncRNAs with data of public and our center.

Results: As a result, we identified six subtypes of gastric cancer with different prognosis ($P=0.00446$) and a total of 83 disease associated lncRNAs. We finally obtained five prognostic lncRNA biomarkers. Survival analyses showed that the high expression of all identified lncRNAs positive associated with worse prognosis. We also found that the prognostic model on five identified lncRNAs could predict survival with high 5-years AUC 0.69. And the differences of survival between high risk and low risk groups were significant in both of our and public database. In the multivariable analyses, we found that the prognostic model was an independent prognostic factor ($p = 0.0104$).

Conclusions: We concluded that the prognostic model on five identified lncRNAs was closely related to the overall survival of gastric cancer and may serve as promising prognostic biomarkers of gastric cancer.

Introduction

Gastric cancer is one of the most common gastrointestinal malignant tumors, with an estimated 951600 cases and 713100 deaths being attributed to this cancer in 2012[1]. The regional variations observed regarding stomach cancer are substantial. Eastern Asia had the highest incidence of gastric cancer, with an incidence of 35.4 per 100000[1, 2]. In China, gastric cancer ranks second in the incidence of cancer following lung cancer and ranks third in cancer deaths [3, 4]. Approximately 4104000 gastric cancer cases and 293800 gastric cancer deaths were estimated to occur in China in 2014[4, 5]. Although the prognosis of gastric cancer has improved substantially, the mortality of this cancer is still high, which is partly attributed to the advanced stages of diagnosed cancer and the high proportion of recurrence, including lymph node metastasis, distant metastasis, and peritoneal metastasis[6, 7]. Therefore, early detection of patients with aggressive biological characteristics and recurrent cancer undergoing radical surgery becomes significant for improving patients' survival with gastric cancer.

Long noncoding RNAs (lncRNAs) are noncoding RNAs longer than 200 nucleotides. Increasing evidence has pointed towards lncRNAs as regulators of several human diseases, including cancer[8, 9]. Many cancer-related lncRNAs have been reported to play important roles in multiple steps of carcinogenesis, including cell proliferation, cellular signaling, angiogenesis and metastasis[10, 11]. HOX transcript

antisense RNA (HOTAIR) is reported to be overexpressed in gastric cancer and associated with lymph node metastasis and vessel invasion in gastric cancer by promoting epithelial to mesenchymal transition (EMT)[12, 13]. As the first imprinted lncRNA to be reported, H19 is upregulated in gastric cancer tissues compared with normal tissues and associated with proliferation, migration and invasion of gastric cancer cells[14, 15]. The upregulated H19 in plasma makes it a potential diagnostic biomarker for gastric cancer[15]. Growth arrest-specific transcript 5 (GAS5) arrests tumor growth by regulating apoptosis and the cell cycle[16]. GAS5 has been reported to be correlated with tumor stage and lymph node metastasis[17]. Maternally expressed gene 3 (MEG3) is a tumor suppressor gene downregulated in gastric cancer. MEG3 has been demonstrated to be associated with deep tumor invasion, metastasis and poor prognosis, which make it a potential prognostic biomarker for gastric cancer[18, 19]. Many lncRNAs, such as long intergenic noncoding RNA00152 and urothelial carcinoma-associated 1, have also been reported to be related to biological processes and the diagnosis and prognosis of gastric cancer.

Copy number variations (CNVs) refer to genomic structural variations with gene amplification, gain, loss and deletion, which have been regarded to play a significant role in carcinogenesis of gastric cancer[20, 21]. Researchers have found numerous CNVs on many cancer-associated genes, such as CTNNB1, MYC and CDKN2A, located on different chromosomes, including 3p22, 4q25, 11p13, 1p36 and 9p21[22, 23]. The PIK3CA gene is frequently amplified in gastric cancer and is involved in multiple steps of tumorigenesis, including cancer cell proliferation and apoptosis[24]. The CNVs of this gene could predict the prognosis of gastric cancer regardless of tumor stage[25, 26]. As a well-known suppressor gene, adenomatous polyposis coli (APC) exhibits frequent deletions in gastric cancer[27, 28]. APC was regarded as a potential prognostic biomarker due to its close relationship with lymph node invasion and metastasis[29]. Epidermal growth factor receptor (EGFR) CNVs were also reported to be associated with an increased risk of invasion and metastasis, thereby resulting in worse prognosis[30]. The CNVs of other well-known genes, such as MET, HER-2 and TP53, were also correlated with the prognosis of gastric cancer[31, 32]. As mentioned above, most studies focused on the effects of CNVs on protein-coding genes; however, an increasing number of researchers have found that the expression of lncRNAs and miRNAs were also regulated by CNVs[33, 34]. It was reported that an estimated one-third of aberrant lncRNA expression could be attributed to CNVs[33]. Considering the role of lncRNAs in carcinogenesis, the lncRNAs deregulated by CNVs warrant further study to determine their role in the prognosis of gastric cancer.

Therefore, we mined the CNVs and lncRNAs of gastric cancer from The Cancer Genome Atlas (TCGA) and identified prognostic biomarkers of lncRNA regulated by CNVs.

Methods

Sequenced data collection

We downloaded methylation, RNA sequencing, CNV, and mutation data of gastric cancer and their corresponding follow-up information from the database TCGA Genomic Data Commons (GDC). First, we

obtained all fragments per kilobase of transcript per million fragments mapped (FPKM) and counted data from TCGA GDC. Then, we transferred the FPKM data to Transcripts Per Million (TPM) data. LncRNA, sense_intronic, sense_overlapping, antisense, processed_transcript, and 3prime_overlapping_ncRNA were classified as lncRNA. We then obtained the FPKM expression profile of lncRNA and protein-coding genes. We downloaded the expression data of all samples sequenced for methylation with the HumanMethylation450 Beadchip from TCGA GDC. CpG sites with cross-reactivity in the genome according to the cross-reactive sites from the discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microassay were excluded[35]. The CpGs and single nucleotide sites were also removed.

Subtype identification and differential analyses

First, we conducted COX regression analyses between protein-coding genes, CNV and methylation and the survival of patients with gastric cancer. Second, we analyzed a total of 337 samples with CNV, methylation and RNA sequencing data. Multi-omics cluster analysis was performed to identify subtypes using the iClusterPlus R package on the prognosis-associated coding genes, CNVs and methylation sites. iClusterPlus is developed based on unsupervised cluster analysis, which can generate tumor classification by capturing patterns from multiple genomic data. Before iClusterPlus analysis, we first selected and optimized the necessary parameters. Firstly, we repeatedly divided the samples into different training and verification sets to determine the optimal number of clusters k . In order to visualize the results, we plotted the percent of explained variation vs. the number of clusters. The optimal k value is the point where the curve starts to flatten out. We then use the Bayesian information criterion to select the optimal combination of the sparse model and the penalty parameter, or λ . Finally, we combine the optimal clustering number (k) and penalty parameter (λ) to run the iClusterPlus analysis. Furthermore, we analyzed the differential lncRNA and protein-coding genes between tumor and normal samples in different subtypes. Foldchanges greater than 2 and $FDR < 0.05$ were used as cutoff values. A comparison was performed between identified lncRNAs and 232 lncRNAs closely related to disease from the databases LncRNADisease and Lnc2Cancer[36]. To evaluate the differential expression of lncRNAs in each subtype of cancer, we conducted gene set enrichment analysis (GESA) according to the absolute value of lncRNA fold change.

Weighted gene coexpression network analysis

We constructed coexpression modules of differential protein-coding genes and lncRNAs using WGCNA. We first transferred the FPKM data to TPM data and extracted the expression profiles of lncRNAs and protein-coding genes. Hierarchical clustering analysis was performed to identify outlier samples. We obtained 405 samples excluding outlier samples with distances greater than 8000. The Pearson correlation coefficient was used to calculate the distance between each gene and lncRNA. We screened coexpression modules by constructing a weighted coexpression network with 3 as its soft threshold

using the WGCNA R package. We transferred the expression matrix to the adjacent matrix and then to the topology matrix (TOM), on which we conducted average-linkage hierarchical clustering analysis according to the hybrid dynamic shear standard. Each coexpression module was set to include more than 30 genes. Then, we calculated eigengenes of each module after they were determined. After that step, we conducted clustering analysis on modules, and the adjacent modules emerged as a new one. The parameters are height 0.25, deepsplit 2 and minModulesize 30. We then made statistics of protein-coding genes and lncRNAs in each module. We further performed gene ontology enrichment (GO) analysis on significantly enriched modules and analyzed the crosstalk of GO terms.

CNV-related lncRNA biomarker selection

We analyzed the CNVs of 442 gastric cancer samples from TCGA using GISTIC 2.0 software. The CNV profile of lncRNAs was extracted for further analysis. We defined a copy number more than one as copy number amplification and a copy number less than one as copy number deletion[37, 38] Then, we performed statistical analyses on the proportions of copy number amplifications and deletions for each lncRNA. To explore the relationship between lncRNA expression and CNV, we identified lncRNAs with more than 10 percent CNVs from each sample for further analysis. To systematically identify lncRNA prognostic markers, we analyzed the CNVs of differential lncRNAs in each subtype of gastric cancer. We selected lncRNAs with CNVs greater than 0.1 percent and differential expression within identified subtypes. We screened samples with expression levels greater than 0 for each lncRNA and divided samples according to the median expression level. Then, the survival analyses were performed with a threshold P value less than 0.05 as significance. We performed prognostic identification efficiency by constructing a receiver operating characteristic (ROC) curve for all identified lncRNA prognostic biomarkers. Then, Pearson correlation coefficient analyses for the expression level and CNV of identified lncRNAs were conducted. The area under the curve (AUC) greater than 0.6 was chosen for further analysis We retained lncRNAs positively associated with CNVs and that with a Pearson correlation coefficient greater than 0.1. We classified samples into high and low expression groups based on the median expression level of each sample for all included lncRNAs. Survival analyses were conducted between both groups. We conducted multivariable cox analyses and constructed prognostic model on identified lncRNA. The independence of prognostic model was also analyzed using multivariable analyses. The validation was performed using GSE62254 data from the GPL570 platform and in our center.

Results

Six genomic subtypes of gastric cancer were identified

We included 1886 protein-coding genes, 3176 CNV and 9256 CpG sites by univariate COX regression analyses between protein-coding genes, CNV and methylation and the survival of patients with gastric cancer (Fig. 1, Figure S1A). Before iClusterPlus analysis, we determined that the optimal value of k was 5,

and the number of clusters was $k + 1$, that is, 6 clusters (Figure S1B). In order to build the final model, we selected the 95th percentile as the threshold to select the most discriminative features, and only features larger than this threshold could be finally expressed in the 6 clusters. Based on the multi-omics cluster analysis on the prognosis-associated coding genes, CNV and methylation sites, we acquired 6 subtypes (Supplementary Table 1). We found a substantial difference in prognosis among the 6 subtypes with statistical significance ($P = 0.00446$) (Fig. 2A). Meanwhile, we extracted the top 10 mutated genes from each subtype, and a total of 28 genes were obtained, which indicated substantial overlaps of highly frequent mutated genes within all subtypes (Fig. 2B).

Differential analyses of lncRNAs and protein-coding genes in different subtypes

A total of 2507 differential lncRNA and 3453 protein-coding genes were obtained from all subtypes (Supplementary Table 2). The minimum quantities and the largest number of lncRNA and protein-coding genes were found in C3 and C2 subtypes, respectively (Supplementary Table 2). The differential lncRNAs were presented for all subtypes (Fig. 3A-G). We found that lncRNA downregulation was greater than upregulation in the C2, C3 and C4 subtypes (Fig. 3B-D). In contrast, the opposite results were detected in the C6 subtype (Fig. 3F). In the C1 and C5 subtypes, a similar number of lncRNAs with differential regulation were found (Fig. 3A, 3E). We found that the differential coding genes were more than those of lncRNAs (Fig. 3H). A total of 83 disease-associated lncRNAs were obtained by comparing the differential lncRNAs identified from all subtypes with disease-associated lncRNAs from the database ($P < 0.0001$) (Fig. 3I).

To evaluate the differentiation of lncRNAs in each subtype of cancer, we conducted gene set enrichment analysis (GSEA). We found that the differential lncRNAs were abundant in the gene sets with substantial fold changes (Figure S2A-G). We analyzed the overlapping differential lncRNAs among six subtypes of gastric cancer, and substantial common differential lncRNAs were identified (Figure S2H).

Weighted gene coexpression network analysis (WGCNA) of subtype-associated differential protein-coding genes and lncRNAs

Hierarchical clustering analysis was performed to identify outlier samples. We excluded outlier samples with a distance greater than 8000 (Figure S3A). In our study, the coexpression network conforms to the scale-free network by choosing 3 as its soft threshold (Fig. 3B-C). Finally, we acquired 24 gene modules (Fig. 3D), in which the gray module represented genes not clustered in other modules. We then made statistics of protein-coding genes and lncRNAs in each module (Supplementary Table 3). The black magenta and purple modules are enriched for lncRNA (Figure S3E). We further performed gene ontology enrichment analysis on these three modules and analyzed the crosstalk of GO terms. A total of 843 GO

terms were enriched, and few instances of crosstalk were found (Fig. 4A). The results indicated that the three modules may have different functions. The top 20 GO terms from the black module are associated with many metabolic processes, including fatty acid metabolism and xenobiotic metabolic process (Fig. 4B). The Magenta module was associated with extracellular structure organization and extracellular matrix organization (Fig. 4C). The purple module was mainly related to pattern specification process and embryonic limb morphogenesis (Fig. 4D). All the results suggested the important role of lncRNA in carcinogenesis of gastric cancer.

Conjoint analyses of CNVs and lncRNAs in the TCGA database

To explore the function of lncRNAs associated with CNVs on carcinogenesis, we conjointly analyzed the CNVs and lncRNAs of 442 gastric cancer samples from TCGA using GISTIC 2.0 software. We found that the proportion of copy number amplifications was greater than that of deletions (Figure S4A). The most frequent deletions were identified on chromosome 8, and the largest copy number amplifications were distributed on chromosomes 5, 12 and 19 (Figure S4A). We further analyzed the distribution of correlation between lncRNA expression profiles and CNVs. We found a positive correlation between them rather than a random distribution (Figure S4B). In focal CNV peaks of the genome, we identified more copy number deletions than amplifications of lncRNA genes, which suggested a close relationship between the lncRNA gene copy number deletions and gastric cancer (Figure S4C-D).

To study the relationship between lncRNA expression and CNV, we identified lncRNAs with more than 10 percent CNVs from each sample for further analysis. A total of 13 lncRNAs were selected. We further analyzed the differential expression between samples with lncRNA CNVs and normal samples. We found that the expression levels of 10 lncRNAs were higher in samples with copy number amplifications than in normal samples (Figure S5). However, Linc00861 showed more expression in normal samples than in samples with copy number amplifications (Figure S5). The results showed that the lncRNA could be regulated by CNVs.

Identification and validation of CNV-related lncRNA biomarkers

A total of 187 subtypes specific differential CNV-related lncRNAs were identified. Survival analyses were conducted, and we selected 19 prognostic lncRNA biomarkers with statistical significance (Supplementary Table 4). To evaluate the prognostic differentiation efficiency, we constructed a ROC curve for all 19 lncRNA prognostic biomarkers. Twelve lncRNAs with an area under the curve (AUC) greater than 0.6 were included for further analyses (Fig. 5). We finally obtained five prognostic lncRNA biomarkers after Pearson correlation coefficient analyses for the expression level and CNV of 12 lncRNAs (Table 1). Survival analyses were performed on the high and low expression groups of the five lncRNAs.

The results supported that all five lncRNAs could effectively predict the prognosis of patients with gastric cancer (Figure S6A-E). To validate the effects of the five CNV-related lncRNAs on prognosis, we conducted analyses using GSE62254 data from the GPL570 platform. However, only three lncRNAs could be annotated. Finally, survival analyses supported the prognostic efficiency of ENSG00000246859, ENSG00000237187 and ENSG00000245105 (Figure S7A-C).

Table 1
Independent prognostic lncRNAs associated with CNVs

LncRNA	p.value	HR	Low95%	High95%	Symbol	AUC	R
ENSG00000237187	0.020	0.670	0.477	0.941	NR2F1-AS1	0.613	0.156
ENSG00000246859	0.003	0.597	0.423	0.844	STARD4-AS1	0.620	0.130
ENSG00000253405	0.037	0.627	0.404	0.975	EVX1-AS	0.714	0.116
ENSG00000253554	0.039	0.633	0.409	0.980	RNA1414	0.691	0.152
ENSG00000245105	0.004	0.607	0.432	0.855	A2M-AS1	0.612	0.112

Notes: HR=hazard ratio; AUC=area under curve;CNVs=copy number variants

We conducted multivariable cox analyses and constructed prognostic model on the five identified lncRNA. $RiskScore_5 = 0.11041842 * \exp^{NR2F1AS1} + 0.17242272 * \exp^{STARD4-AS1} - 0.03026703 * \exp^{EVX1-AS} + 0.02301935 * \exp^{LOC102724623} - 0.05347689 * \exp^{A2M-AS1}$. We found that the RiskScore model could predict survival with high 5-years AUC 0.69 (Fig. 6A). Meanwhile, the difference of survival between high risk and low risk groups was significant (P = 0.01012) (Fig. 6B). We also performed validation of prognostic model in our center. The predictive value for survival is similar (Fig. 6C). The difference of survival was also observed significantly (P = 0.033) (Fig. 6D).

To further identify the independence of RiskScore model in clinic, we systematically analyzed the clinical characteristics and RiskScore model. Multivariable analyses showed that risk score was independent prognostic factors (HR = 1.583 95%CI = 1.114–2.249, p = 0.0104) (Table 2).

Table 2
univariate and multivariable analyses of prognostic factors

Variables	Univariate analysis			Multivariable analysis		
	HR	95%CI of HR	P value	HR	95%CI of HR	P value
TCGA cohort						
5-lncRNA risk score						
Risk score (High/Low)	1.554	1.107–2.181	0.0107	1.583	1.114–2.249	0.0104
Age	1.022	1.005–1.039	0.012	1.03	1.012–1.048	0.0008
Gender (Male/Female)	1.273	0.887–1.825	0.19	1.204	0.835–1.734	0.3193
AJCC PT	1.681	1.097–2.576	0.017	1.788	1.178–2.713	0.0063
AJCC PN	1.559	1.116–2.179	0.009	1.591	1.071–2.363	0.0215
AJCC PM	1.727	1.038–2.871	0.035	2.162	1.280–3.65	0.0039
AJCC STAGE	1.757	1.222–2.527	0.0023	0.816	0.575–1.158	0.2555
Grade	1.303	0.918–1.848	0.138	1.21	0.836–1.748	0.3114
Notes: HR: hazard ratio, CI: confident interval, TCGA: The Cancer Genome Atlas						

Discussion

With the development of next-generation sequencing and target drugs for cancer, molecular classification has become considerably more important for the detection and treatment of cancer. In 2014, TCGA molecularly divided gastric cancer into four types, including tumors with Epstein-Barr virus, microsatellite unstable tumors, genomically stable tumors and tumors with chromosomal instability[39]. The role of this molecular classification for gastric cancer therapy was partly supported by the high response rate for programmed cell death 1 (PD1)-targeted therapy of EB virus tumors and microsatellite instability-high tumors[40]. However, the classification is not sufficiently comprehensive due to a lack of transcription data, and it could not effectively predict the survival of gastric cancer. In our study, we proposed a new molecular classification method for DNA methylation, CNVs and coding genes. We also described 6 clusters in combination with the clinical characteristics of patients and found that C2 and C5 showed good prognosis due to a higher proportion of patients in early stage, while C1 and C6 showed poor prognosis due to a higher proportion of patients in advanced stage. In the molecular typing comparison of TCGA, we found that the EBV positive and MSI rates were higher in C2 and C4. Interestingly, the gene mutation rate of patients was also higher in these two clusters. It is well known that EBV positive and MSI high patients have good efficacy in current anti-PD-L1 treatment of gastric cancer, and high tumor mutation burden is also considered as a potential biomarker. Therefore, patients in these two clusters

may be potential beneficiaries of immunotherapy. Considering the complexity of our molecular classification, we identified a few prognostic lncRNAs according to the different subtypes, which could be used to distinguish gastric cancer with different risks.

lncRNA has been used to predict the survival of many types of cancer, including gastric cancer, HCC and prostate cancer[41–44]. Genomic CNVs play important roles in the carcinogenesis and development of cancer. Approximately one-third of deregulated lncRNAs are associated with their CNVs. However, the function and effects of CNV-related lncRNAs on cancer have not been thoroughly elucidated to date. Therefore, we conducted conjoint analyses on lncRNAs and their CNVs.

In our research, we identified five CNV-related lncRNA prognostic biomarkers. We found that gastric cancer patients with high expression of lncRNA NR2F1 antisense RNA1 (NR2F1-AS1, ENSG00000237187) tend to have a worse prognosis. Previous research showed that NR2F1-AS1 knockdown could reduce hepatocellular carcinoma (HCC) cell invasion, migration and drug resistance[45]. Similar to NR2F1-AS1, the other four lncRNAs were also observed to have a negative association between their expression and prognosis in patients with gastric cancer. STARD4-antisense RNA1 (STARD4-AS1, ENSG00000246859) is rarely studied, and its role in carcinogenesis is unknown[46]. EVX1 antisense RNA (EVN1-AS, ENSG00000253405) is reported to be expressed during embryonic body differentiation[47]; however, its association with cancer has not been determined. Additionally, the roles of long intergenic non-protein-coding RNA1414 (LINC01414, ENSG00000253554) and A2M antisense RNA1 (A2M-AS1, ENSG00000245105) in carcinogenesis have not been determined. However, most of the identified CNV-related lncRNAs were not further studied for their role in tumorigenesis and the development of cancer in addition to NR2F1-AS1. Our survival analyses showed that the high expression of these lncRNAs was associated with poor prognosis, supporting the importance of these lncRNAs in gastric cancer. Therefore, the mechanisms of lncRNA regulation in gastric cancer merit further study.

Our research first conducted analyses of lncRNAs and their CNVs and identified five novel lncRNA prognostic biomarkers. The deep bioinformatics analyses on multidimensional genomic data make our results convincing. However, the large-scale multi-omics data are not very sufficient; therefore, our results may not be very reliable. Furthermore, although studies validating identified lncRNAs have been performed in another database, we could not perform validation in patients at our center.

Conclusions

In summary, we identified five innovative CNV-related lncRNAs, including NR2F1-AS1, STARD4-AS1, EVN1-AS, LINC01414 and A2M-AS1. The prognostic model on five identified lncRNAs was closely related to the overall survival of gastric cancer and may serve as promising prognostic biomarkers of gastric cancer.

List Of Abbreviations

GC: gastric cancer, CNV: copy number variants, lncRNA: long non-coding RNA, TCGA: The Cancer Genome Atlas,

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The study data is available at TCGA (<https://portal.gdc.cancer.gov>).

Competing interests

The authors declare that they have no competing interests.

Funding

This research was supported by the Peking University Clinical Scientist Program (BMU2019LCKXJ011), National Science Foundation for Young Scientists of China (No. 81802735), Beijing Youth Talent Plan (No. QML20191101), 'San Ming' Project of Shenzhen city, China (No. SZSM201612051), Science Foundation of Peking University Cancer Hospital (No: 2020-11) and Beijing Municipal Administration of Hospitals Clinical Medicine Development of Special Funding Support (ZYLX201701).

Authors' contributions

AW designed the study and wrote the manuscript. JC, KZ and QH conducted data analyses. XJ, KJ, XW and JZ reviewed and edited this manuscript. ZB and JJ co-ordinated and provided financial support for this work.

Acknowledgement

Not applicable

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A: **Global cancer statistics, 2012.** *CA: a cancer journal for clinicians* 2015, **65**(2):87-108.
2. Torre LA, Siegel RL, Ward EM, Jemal A: **Global Cancer Incidence and Mortality Rates and Trends—An Update.** *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2016, **25**(1):16-27.
3. Chen W, Zheng R, Zhang S, Zeng H, Zuo T, Xia C, Yang Z, He J: **Cancer incidence and mortality in China in 2013: an analysis based on urbanization level.** *Chinese journal of cancer research = Chung-kuo yen cheng yen chiu* 2017, **29**(1):1-10.
4. Chen W, Sun K, Zheng R, Zeng H, Zhang S, Xia C, Yang Z, Li H, Zou X, He J: **Cancer incidence and mortality in China, 2014.** *Chinese journal of cancer research = Chung-kuo yen cheng yen chiu* 2018, **30**(1):1-12.
5. Yang L, Zheng R, Wang N, Yuan Y, Liu S, Li H, Zhang S, Zeng H, Chen W: **Incidence and mortality of stomach cancer in China, 2014.** *Chinese journal of cancer research = Chung-kuo yen cheng yen chiu* 2018, **30**(3):291-298.
6. Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H: **Gastric cancer.** *Lancet (London, England)* 2016, **388**(10060):2654-2664.
7. Karimi P, Islami F, Anandasabapathy S, Freedman ND, Kamangar F: **Gastric cancer: descriptive epidemiology, risk factors, screening, and prevention.** *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2014, **23**(5):700-713.
8. Beermann J, Piccoli MT, Viereck J, Thum T: **Non-coding RNAs in Development and Disease: Background, Mechanisms, and Therapeutic Approaches.** *Physiological reviews* 2016, **96**(4):1297-1325.
9. Yang Y, Junjie P, Sanjun C, Ma Y: **Long non-coding RNAs in Colorectal Cancer: Progression and Future Directions.** *Journal of Cancer* 2017, **8**(16):3212-3225.
10. Kondo Y, Shinjo K, Katsushima K: **Long non-coding RNAs as an epigenetic regulator in human cancers.** *Cancer science* 2017, **108**(10):1927-1933.
11. Ling H, Fabbri M, Calin GA: **MicroRNAs and other non-coding RNAs as targets for anticancer drug development.** *Nature reviews Drug discovery* 2013, **12**(11):847-865.
12. 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* 2013, **8**(10):e77070.
13. Chen WM, Chen WD, Jiang XM, Jia XF, Wang HM, Zhang QJ, Shu YQ, Zhao HB: **HOX transcript antisense intergenic RNA represses E-cadherin expression by binding to EZH2 in gastric cancer.** *World journal of gastroenterology* 2017, **23**(33):6100-6110.

14. Gao T, He B, Pan Y, Xu Y, Li R, Deng Q, Sun H, Wang S: **Long non-coding RNA 91H contributes to the occurrence and progression of esophageal squamous cell carcinoma by inhibiting IGF2 expression.** *Molecular carcinogenesis* 2015, **54**(5):359-367.
15. Hashad D, Elbanna A, Ibrahim A, Khedr G: **Evaluation of the Role of Circulating Long Non-Coding RNA H19 as a Promising Novel Biomarker in Plasma of Patients with Gastric Cancer.** *Journal of clinical laboratory analysis* 2016, **30**(6):1100-1105.
16. Pickard MR, Mourtada-Maarabouni M, Williams GT: **Long non-coding RNA GAS5 regulates apoptosis in prostate cancer cell lines.** *Biochimica et biophysica acta* 2013, **1832**(10):1613-1623.
17. Guo X, Deng K, Wang H, Xia J, Shan T, Liang Z, Yao L, Jin S: **GAS5 Inhibits Gastric Cancer Cell Proliferation Partly by Modulating CDK6.** *Oncology research and treatment* 2015, **38**(7-8):362-366.
18. Yoon JH, Abdelmohsen K, Gorospe M: **Functional interactions among microRNAs and long noncoding RNAs.** *Seminars in cell & developmental biology* 2014, **34**:9-14.
19. Peng W, Si S, Zhang Q, Li C, Zhao F, Wang F, Yu J, Ma R: **Long non-coding RNA MEG3 functions as a competing endogenous RNA to regulate gastric cancer progression.** *Journal of experimental & clinical cancer research : CR* 2015, **34**:79.
20. Liang L, Fang JY, Xu J: **Gastric cancer and gene copy number variation: emerging cancer drivers for targeted therapy.** *Oncogene* 2016, **35**(12):1475-1482.
21. Nakamura Y: **DNA variations in human and medical genetics: 25 years of my experience.** *Journal of human genetics* 2009, **54**(1):1-8.
22. Deng N, Goh LK, Wang H, Das K, Tao J, Tan IB, Zhang S, Lee M, Wu J, Lim KH *et al*: **A comprehensive survey of genomic alterations in gastric cancer reveals systematic patterns of molecular exclusivity and co-occurrence among distinct therapeutic targets.** *Gut* 2012, **61**(5):673-684.
23. Zhang D, Wang Z, Luo Y, Xu Y, Liu Y, Yang W, Zhang X: **Analysis of DNA copy number aberrations by multiple ligation-dependent probe amplification on 50 intestinal type gastric cancers.** *Journal of surgical oncology* 2011, **103**(2):124-132.
24. Samuels Y, Wang Z, Bardelli A, Silliman N, Ptak J, Szabo S, Yan H, Gazdar A, Powell SM, Riggins GJ *et al*: **High frequency of mutations of the PIK3CA gene in human cancers.** *Science (New York, NY)* 2004, **304**(5670):554.
25. Shi J, Yao D, Liu W, Wang N, Lv H, Zhang G, Ji M, Xu L, He N, Shi B *et al*: **Highly frequent PIK3CA amplification is associated with poor prognosis in gastric cancer.** *BMC cancer* 2012, **12**:50.
26. Takahashi N, Yamada Y, Taniguchi H, Fukahori M, Sasaki Y, Shoji H, Honma Y, Iwasa S, Takashima A, Kato K *et al*: **Clinicopathological features and prognostic roles of KRAS, BRAF, PIK3CA and NRAS mutations in advanced gastric cancer.** *BMC research notes* 2014, **7**:271.
27. Fang Z, Xiong Y, Li J, Liu L, Zhang W, Zhang C, Wan J: **APC gene deletions in gastric adenocarcinomas in a Chinese population: a correlation with tumour progression.** *Clinical & translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico* 2012, **14**(1):60-65.

28. Chang VY, Federman N, Martinez-Agosto J, Tatishchev SF, Nelson SF: **Whole exome sequencing of pediatric gastric adenocarcinoma reveals an atypical presentation of Li-Fraumeni syndrome.** *Pediatric blood & cancer* 2013, **60**(4):570-574.
29. Buffart TE, Carvalho B, van Grieken NC, van Wieringen WN, Tijssen M, Kranenbarg EM, Verheul HM, Grabsch HI, Ylstra B, van de Velde CJ *et al*: **Losses of chromosome 5q and 14q are associated with favorable clinical outcome of patients with gastric cancer.** *The oncologist* 2012, **17**(5):653-662.
30. Higaki E, Kuwata T, Nagatsuma AK, Nishida Y, Kinoshita T, Aizawa M, Nitta H, Nagino M, Ochiai A: **Gene copy number gain of EGFR is a poor prognostic biomarker in gastric cancer: evaluation of 855 patients with bright-field dual in situ hybridization (DISH) method.** *Gastric cancer : official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* 2016, **19**(1):63-73.
31. Teng L, Lu J: **cMET as a potential therapeutic target in gastric cancer (Review).** *International journal of molecular medicine* 2013, **32**(6):1247-1254.
32. Karaman A, Kabalar ME, Binici DN, Ozturk C, Pirim I: **Genetic alterations in gastric precancerous lesions.** *Genetic counseling (Geneva, Switzerland)* 2010, **21**(4):439-450.
33. Hu Y, Wang J, Qian J, Kong X, Tang J, Wang Y, Chen H, Hong J, Zou W, Chen Y *et al*: **Long noncoding RNA GAPLINC regulates CD44-dependent cell invasiveness and associates with poor prognosis of gastric cancer.** *Cancer research* 2014, **74**(23):6890-6902.
34. Hu L, Wu Y, Tan D, Meng H, Wang K, Bai Y, Yang K: **Up-regulation of long noncoding RNA MALAT1 contributes to proliferation and metastasis in esophageal squamous cell carcinoma.** *Journal of experimental & clinical cancer research : CR* 2015, **34**:7.
35. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R: **Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray.** *Epigenetics* 2013, **8**(2):203-209.
36. Gao Y, Wang P, Wang Y, Ma X, Zhi H, Zhou D, Li X, Fang Y, Shen W, Xu Y *et al*: **Lnc2Cancer v2.0: updated database of experimentally supported long non-coding RNAs in human cancers.** *Nucleic acids research* 2019, **47**(D1):D1028-d1033.
37. **Comprehensive molecular characterization of clear cell renal cell carcinoma.** *Nature* 2013, **499**(7456):43-49.
38. **Comprehensive genomic characterization defines human glioblastoma genes and core pathways.** *Nature* 2008, **455**(7216):1061-1068.
39. **Comprehensive molecular characterization of gastric adenocarcinoma.** *Nature* 2014, **513**(7517):202-209.
40. Kim ST, Cristescu R, Bass AJ, Kim KM, Odegaard JI, Kim K, Liu XQ, Sher X, Jung H, Lee M *et al*: **Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer.** *Nature medicine* 2018, **24**(9):1449-1458.
41. Sun B, Dang Y, Zhang F, Li K, Ouyang X, Wang K, Huang Q: **Long noncoding RNA RP1163G9.1 is downregulated in gastric adenocarcinoma and is associated with a poor prognosis.** *Oncology reports*

2019.

42. Tan C, Cao J, Chen L, Xi X, Wang S, Zhu Y, Yang L, Ma L, Wang D, Yin J *et al*: **Noncoding RNAs Serve as Diagnosis and Prognosis Biomarkers for Hepatocellular Carcinoma**. *Clinical chemistry* 2019.
43. Prensner JR, Zhao S, Erho N, Schipper M, Iyer MK, Dhanasekaran SM, Magi-Galluzzi C, Mehra R, Sahu A, Siddiqui J *et al*: **RNA biomarkers associated with metastatic progression in prostate cancer: a multi-institutional high-throughput analysis of SchLAP1**. *The Lancet Oncology* 2014, **15**(13):1469-1480.
44. Veltri WR: **Non-coding RNAs as biomarkers for metastatic prostate cancer**. *The Lancet Oncology* 2014, **15**(13):1412-1413.
45. Huang H, Chen J, Ding CM, Jin X, Jia ZM, Peng J: **LncRNA NR2F1-AS1 regulates hepatocellular carcinoma oxaliplatin resistance by targeting ABCC1 via miR-363**. *Journal of cellular and molecular medicine* 2018, **22**(6):3238-3245.
46. Mo XB, Wu LF, Zhu XW, Xia W, Wang L, He P, Bing PF, Lu X, Zhang YH, Deng FY *et al*: **Identification and evaluation of lncRNA and mRNA integrative modules in human peripheral blood mononuclear cells**. *Epigenomics* 2017, **9**(7):943-954.
47. Bell CC, Amaral PP, Kalsbeek A, Magor GW, Gillinder KR, Tangermann P, di Lisio L, Cheetham SW, Gruhl F, Frith J *et al*: **The Evx1/Evx1as gene locus regulates anterior-posterior patterning during gastrulation**. *Scientific reports* 2016, **6**:26657.

Figures

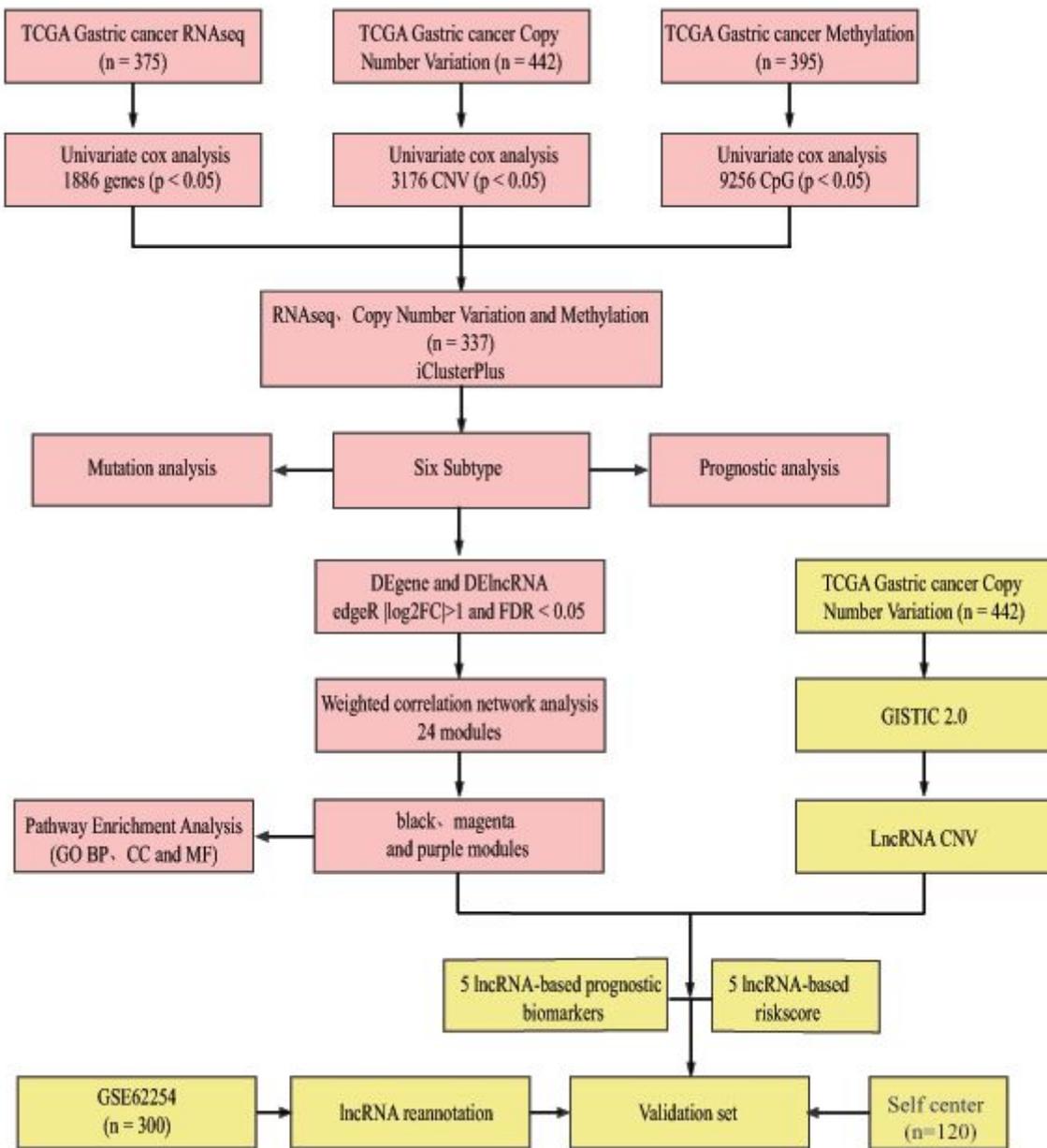


Figure 1

Flowchart of cluster establishment

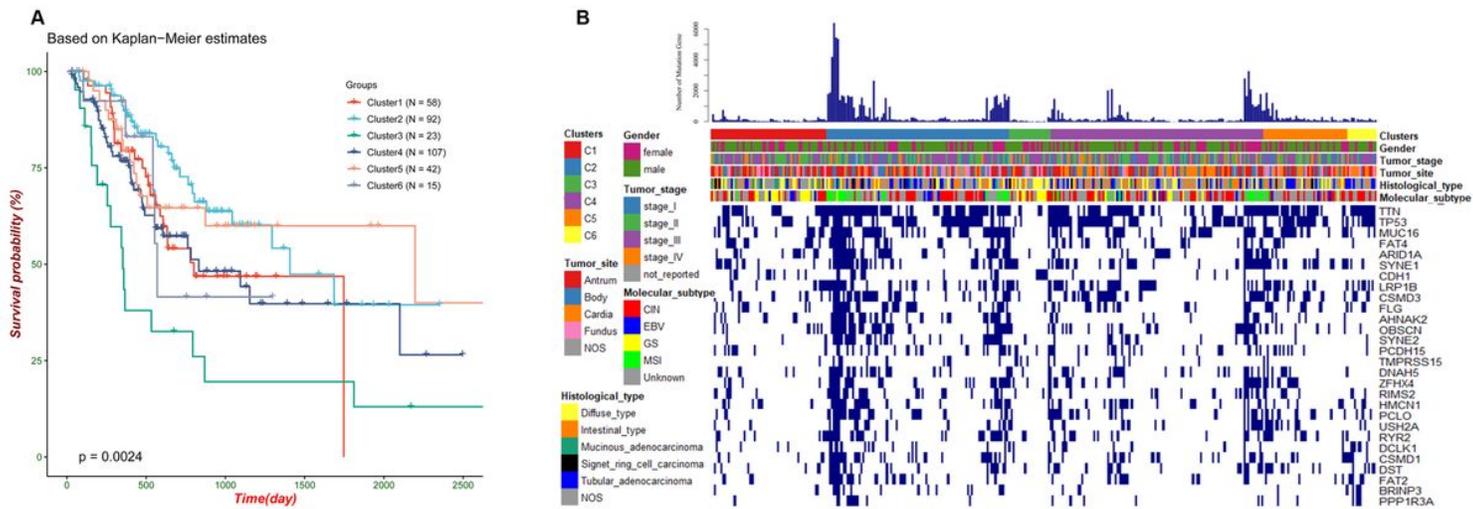


Figure 2

Clinicopathological features and genomic profiles of six subtypes of gastric cancer. A: the survival analysis of different subtypes of gastric cancer. B: The clinical data and mutation landscape of top ten genes mutations in each subtype of gastric cancer.

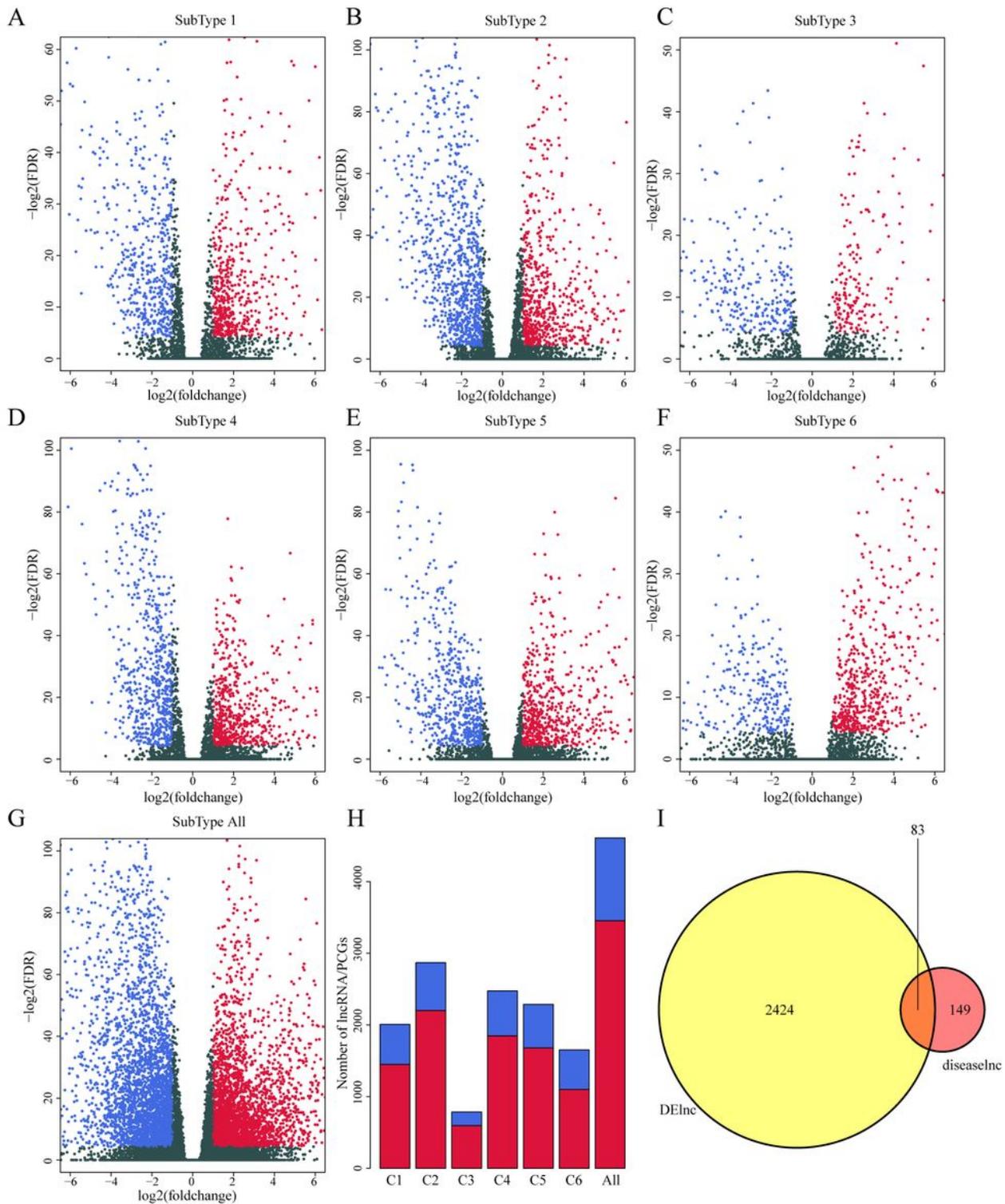


Figure 3

The differential lncRNA and coding genes in different subtypes of gastric cancer. A-F: the volcano plot of six subtypes of cancer. Red dot and blue dot indicate upregulated and downregulated differential lncRNAs. G: differential lncRNAs in all samples. H: the differential lncRNAs and protein coding genes (PCGs) in different subtypes of cancer. Blue and red columns represent differential lncRNAs and PCGs,

respectively. I: Venn diagram shows the relationship between differential IncRNAs and disease associated IncRNAs.

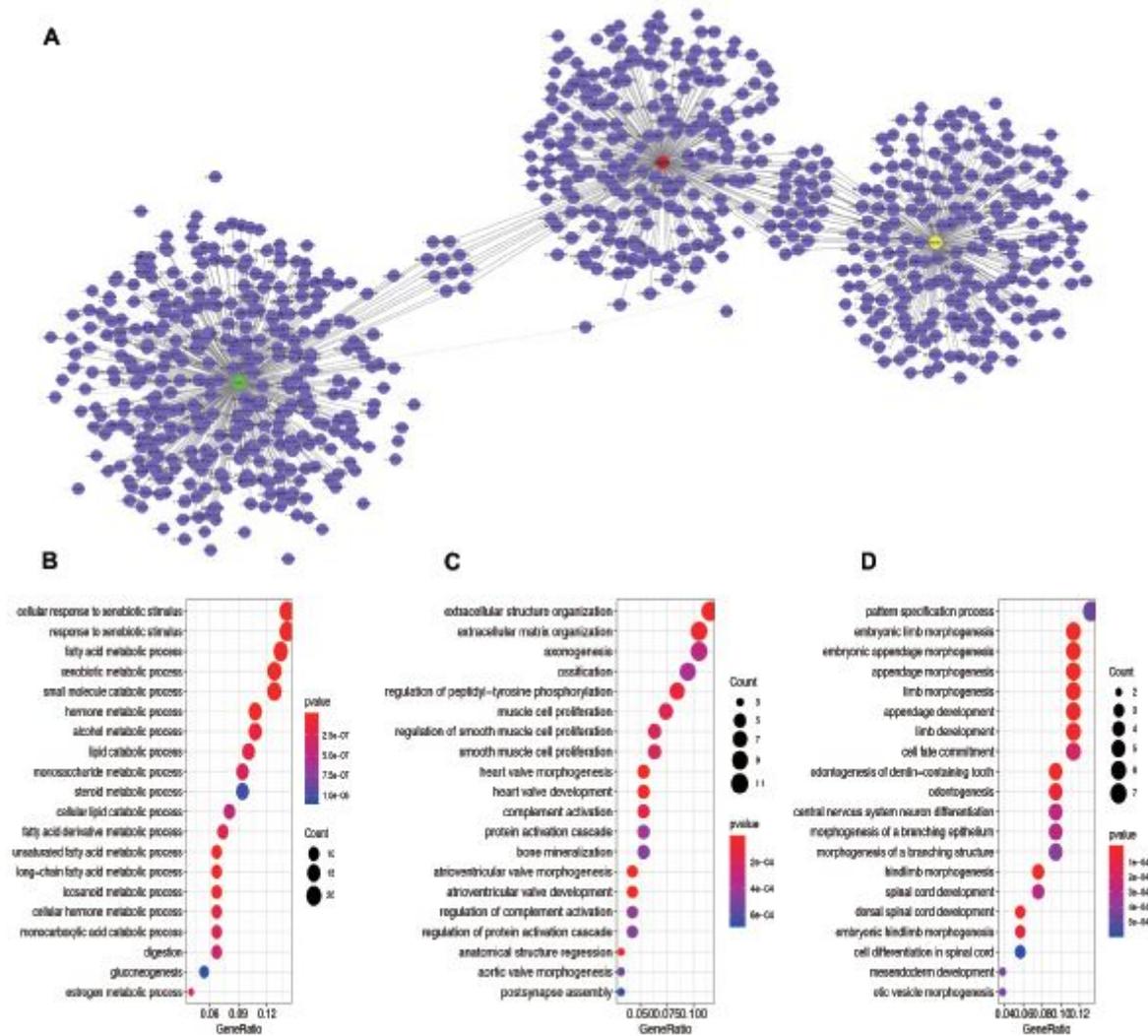


Figure 4

GO enrichment analyses of three modules with substantial enrichment of lncRNA. A: network relationship within three modules enrichment. B-D: GO enrichment of black, magenta and purple modules.

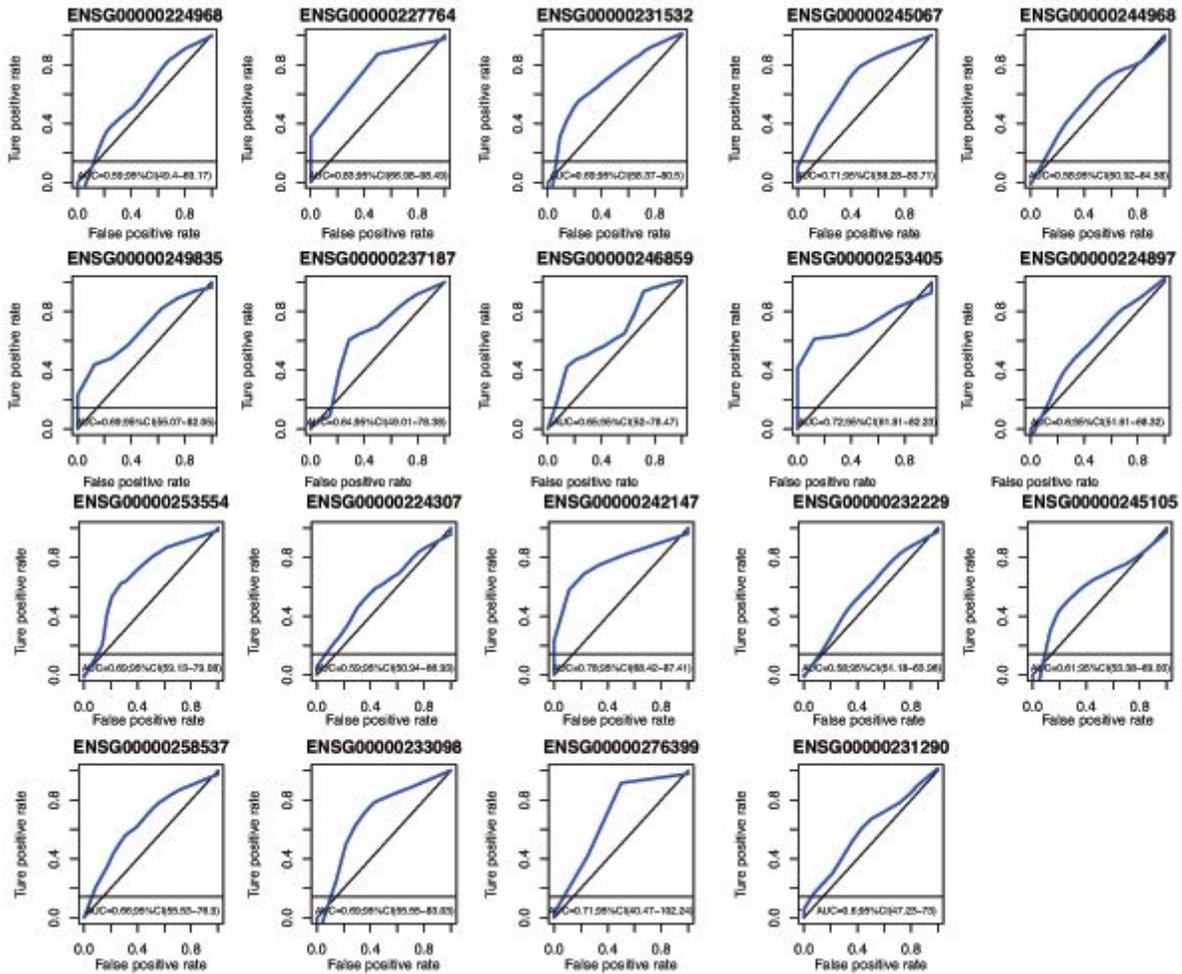


Figure 5

Prognostic identification efficiency of 19 prognostic lncRNAs. The ROC curve of prognostic lncRNA

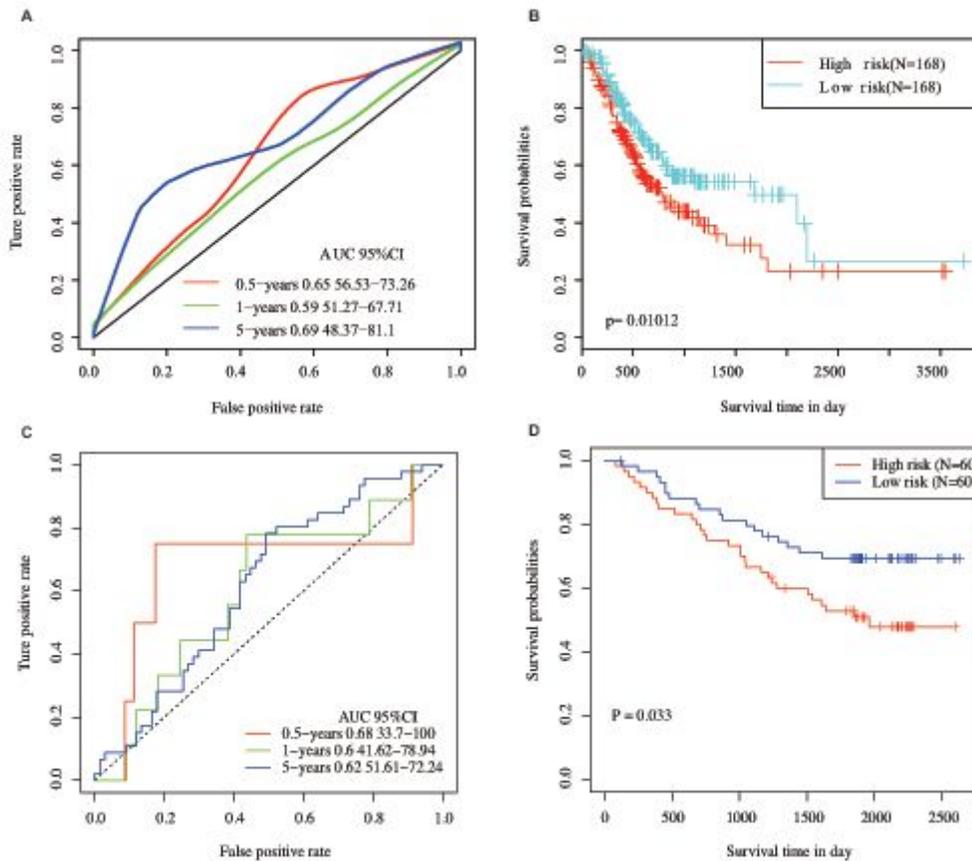


Figure 6

Survival prediction and survival analyses of RiskScore models. A: The ROC curve of relapse prediction for 1-years and 5-years of lncRNA-based RiskScore models. B: Kaplan Meier curve between high and low riskscore groups. C: The ROC curve of relapse prediction for 1-years and 5-years of lncRNA-based RiskScore models. In our cohort. D: Kaplan Meier curve between high and low riskscore groups in our cohort.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterials.docx](#)