Long non-coding RNA GAPLINC regulates CD44-dependent cell invasiveness and associates with poor prognosis of gastric cancer

Ye Hu 1, Jilin Wang 1, Jin Qian 1, Xuan Kong 1, Jieting Tang 1, Yingchao Wang 1, Haoyan Chen 1, Jie Hong 1, Weiping Zou 2, Yingxuan Chen 1, Jie Xu 1,*, Jing-Yuan Fang 1,*

1 State Key Laboratory for Oncogenes and Related Genes; Key Laboratory of Gastroenterology & Hepatology, Ministry of Health; Division of Gastroenterology and Hepatology, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine; Shanghai Cancer Institute, Shanghai Institute of Digestive Disease, 145 Middle Shandong Rd, Shanghai 200001, China
2 Department of Surgery, University of Michigan, Ann Arbor, MI 48109, USA

*Address correspondence and reprint requests to:
Jing-Yuan Fang, or Jie Xu. Renji Hospital, School of Medicine, Shanghai Jiao Tong University, 145 Middle Shandong Rd, Shanghai 200001, China. Tel: 0086-21-63200874, Fax: 0086-21-63266027, E-mail: jingyuanfang@yahoo.com, or jiexu@yahoo.com.

CONFLICT OF INTEREST
All authors declare that they have no conflict of interests.
ABSTRACT

It is increasingly evident that long non-coding RNAs (lncRNA) have causative roles in carcinogenesis. In this study, we report findings implicating a novel lncRNA in gastric cancer (GC), termed GAPLINC, based on the use of global microarray and in situ hybridization (ISH) analyses to identify aberrantly expressed lncRNA in human GC specimens. GAPLINC is a 924-bp long lncRNA that is highly expressed in GC tissues. GAPLINC suppression and with gene expression profiling in GC cells revealed alterations in cell migration pathways, with CD44 expression the most highly correlated. Manipulating GAPLINC expression altered CD44 mRNA abundance and the effects of GAPLINC on cell migration and proliferation were neutralized by suppressing CD44 expression. Mechanistic investigations revealed that GAPLINC regulates CD44 as a molecular decoy for miR211-3p, a microRNA that targets both CD44 and GAPLINC. Tissue ISH analysis suggested that GAPLINC overexpression defines a subgroup of GC patients with very poor survival. Taken together, our results identify a non-coding regulatory pathway for the CD44 oncogene, shedding new light on the basis for GC cell invasiveness.
INTRODUCTION

Gastric cancer (GC) is the second leading cause of cancer mortality in the world, and the majority of GC patients are diagnosed at an advanced stage and died in 24 months after operation because of recurrence and metastasis (1,2). To improve GC early diagnosis and targeted therapy, an in-depth understanding of molecular underpinnings of the disease is required (3-5). It is of clinical importance to identify genes that contribute to GC development and present predictive value for diagnosis or prognosis (6-8). Currently, most reported potential biomarkers for GC are protein-coding genes (PCGs), including the novel somatic gene targets (ARID1A, FAT4, MLL and KMT2C) revealed by large-scale cancer genomic studies. Despite extensive efforts to develop PCG-based biomarkers, only modest successes have been obtained in biomarker-assisted GC diagnosis and treatment (9,10).

Recent integrative genomic studies have revealed that the human genome encodes over 10,000 long noncoding RNAs (lncRNAs) with limited or no protein-coding capacity (11-14). Though a small number of lncRNAs have been functionally characterized (15-18), a large number of members in the class remain functionally uncharacterized (19,20). Growing evidence suggests that cancer lncRNAs, similar to protein-coding genes, may mediate oncogenic or tumor-suppressing effects and may be a new class of cancer biomarkers and therapeutic targets (21-23). One such lncRNA is HOTAIR, which is expressed from the developmental HOXC locus and associates with chromatin modifications in cooperation with the Polycomb complex PRC2 (24). Overexpression of HOTAIR is a powerful predictor of the tumour progression and overall survival in patients with diverse cancers, including diffuse type of GC (25). Recent studies also suggest that lncRNAs can also act as decoys for microRNAs, and an example of this mechanism is represented by the tumour suppressor gene PTEN and its pseudogene PTENP1. The PTENP1 3’UTR was found to increase PTEN expression by binding to microRNAs that downregulate PTEN expression. Despite the above findings, our current knowledge about the expression patterns and functional roles of lncRNAs in gastric cancer are still limited. In previous studies, efforts have been made to re-annotate the probes of Affymetrix microarrays that match lncRNA sequences in several cancers, including GC (26). However, this method covers less than 65% of lncRNA genes and thus should not be considered as unbiased analysis.
In this study we identify deregulated IncRNAs in gastric cancer that are associated with copy number variations (CNVs) or oncogenic transcription factors (TFs). Notably, a long intergenic non-coding RNA GAPLINC displayed considerable predictive effects, when applied alone or combined with others, in the diagnosis and prognosis of gastric cancer. GAPLINC expression strongly correlated with CD44 in GC tissues, and the pro-malignant functions of GAPLINC could be neutralized by suppression of CD44. We provide both in vitro and in vivo data to demonstrate that GAPLINC forms a molecular decoy for miR211-3p, which targets CD44 for degradation. By these efforts, we aim to a model for GAPLINC-mediated cell migration and proliferation in gastric cancer.

**MATERIALS AND METHODS**

**Clinical and histological evaluation of human tissues**

The human specimens in this study were sanctioned by the local ethics committee at the Shanghai Jiao-Tong University School of Medicine Renji Hospital. None of the patients received preoperative treatment, including chemotherapy or radiotherapy. The para-tumorous tissues were taken at a distance of 2-3 cm from the tumor and nontumorous samples were taken at a distance of at least 5 cm from the tumor, and all tissues were examined histologically. The biopsies of chronic gastritis were obtained from outpatients during endoscopic procedure.

**Cell culture**

The human MGC803, SGC7901, HCT116 and H1299 cells were obtained from the Cell Bank of the Chinese Academy of Sciences (Shanghai, China) where they were characterized by mycoplasma detection, DNA-Fingerprinting, isozyme detection and cell vitality detection. These cell lines were immediately expanded and frozen such that they could be restarted every 3 to 4 months from a frozen vial of the same batch of cells. Cells were cultured at 37 °C in an atmosphere of 5% CO2 in RPMI-1640 medium (Invitrogen) supplemented with 10% fetal bovine serum, penicillin and streptomycin (Thermo Scientific) in 25-ml culture flasks.

**Microarray study on IncRNA expression in cancer tissues**

Briefly, samples (ten gastric cancer tissues and ten corresponding nontumor tissues)
were used to synthesize double-stranded complementary DNA (cDNA), which was then labeled and hybridized to the 8×60 K LncRNA Expression Microarray (Arraystar, Rockville, MD). The LncRNA expression microarray used in this study mainly classifies its probes as the following subtypes: (a) Enhancer LncRNAs: Contains profiling data of all LncRNAs with enhancer-like function [19]. (b) Rinn lincRNAs: Contains profiling data of all lincRNAs based on John Rinn's papers [20], [21]. (c) HOX cluster: Contains profiling data of all probes in the four HOX loci, targeting 407 discrete transcribed regions, LncRNAs and coding transcripts [22]. (d) LincRNAs nearby coding gene: Contains the differentially expressed lincRNAs and nearby coding gene pairs (distance <300 kb). (e) Enhancer LncRNAs nearby coding gene: Contains the differentially expressed enhancer-like LncRNAs and their nearby coding genes (distance <300 kb). After having washed the slides, the arrays were scanned by the Agilent Scanner G2505C. Agilent Feature Extraction software (version 11.0.1.1) was used to analyze acquired array images. Quantile normalization and subsequent data processing were performed using the GeneSpring GX v11.5.1 software package (Agilent Technologies). Data are available via GEO GSE50710 (Arraystar LncRNA array, 20 samples).

Microarray for detection of GAPLINC-associated signaling

Total RNA from the human gastric cancer MGC803 cells with GAPLINC stably knockdown and control MGC803 cells were isolated and quantified. The RNA integrity was assessed by standard denaturing agarose gel electrophoresis. The expression profiles were determined using Affymetrix Human Genome U133Plus 2.0 arrays. Quantile normalization and subsequent data processing were performed using the Affymetrix®Microarray Suite 5.0 statistical algorithm. Data are available via GEO GSE51651 (Affymetrix HGU133-P2 array, 6 samples).

Sample classification model based on microarray data

We used a widely applied approach to predict gastric cancer from gene expression profiling, based on an enhancement of the simple nearest prototype (centroid) classifier (27). The PAM algorithm shrinks the prototypes and hence obtains a classifier that is often more accurate than competing methods. The method of “nearest shrunken centroids” identifies subsets of genes that best characterize each class. The shrinkage consists of moving the centroid towards zero by a threshold,
which is determined according to the prediction error of the model. As the threshold increases, the number of genes left in the model decreases. To guide the choice of threshold, PAM does K-fold cross-validation for a range of threshold values. It chooses the highest threshold (i.e., the least genes), given the same prediction error. (Additional Materials and Methods in Supplementary Materials)

RESULTS

Deregulation of lncRNAs is associated with recurrent CNVs in GC

In order to obtain the transcriptional profiles for both lncRNAs and mRNAs in GC, paired gastric cancer tissues and normal tissues (n=20) were analyzed using ArrayStar lncRNA microarray. When the criteria \( P<0.05 \) & Fold Change >1.5 was adopted, we found similar numbers of lncRNAs being significantly upregulated (n=659) or downregulated (n=709) in GC (plotted in Figure 1A, detailed gene information shown in Supplementary Table 1). As the alteration of gene expression is often associated with copy number variation (CNV) in cancers, we tested whether CNV is also prevalent in deregulated lncRNAs in GC. To this end, the genomic regions with CNV in gastric cancers were obtained from the Cancer Genome Atlas (TCGA) dataset, and then re-assigned to lncRNA gene loci using the CNTools algorithm. In all 659 upregulated lncRNAs, 215 (32.6%) were mapped to genomic loci with gained CNVs in gastric cancers (Figure 1A, data in Supplementary Table 2), suggesting that gained CNV may contribute to upregulation of lncRNAs in GC.

Preferential binding of transcription factors to upregulated lncRNAs

Since transcriptional factors (TFs) play central roles in controlling the initiation of gene expression, thus we tried to search for TFs that might be linked to lncRNA deregulation. Recent advances in chromatin immunoprecipitation sequencing (ChIP-seq) provide unbiased and comprehensive knowledge of TF binding patterns throughout the genome, and the growing archive of ChIP-seq data have included many important DNA-binding proteins. In order to search for TFs that may contribute to the upregulation pattern of lncRNAs in GC, we set to analyze preferential binding of TFs to the promoters of upregulated lncRNAs. To this end, we obtained ChIP-seq
data of 97 transcription factors (TFs) from the ENCODE project and Gene Expression Omnibus (GEO) database, and analyzed their binding to promoters of differentially expressed lncRNAs (criteria: <5kb upstream of transcription starting site). Interestingly, a few TFs (namely mutant p53, STAT1, and BCL3) seemed to bind preferential to the promoters of upregulated lncRNAs (Figure 1A and Supplementary Figure 1A, detailed binding lncRNAs listed in Supplementary Table 3). According to the mRNA levels revealed by microarray, these TFs were also upregulated in GC (Supplementary Figure 1B). Experimental validation using qPCR suggested that mutant p53 and STAT1 could indeed upregulate the bound lncRNAs, while BCL3 displayed more varied effects. We further tested the effect of mutant p53 on one of the bound lncRNA uc002kmd.1, and found that ectopic expression of mutant p53 R248W could indeed promote the expression of uc002kmd.1 in MGC803 gastric cancer cells, HCT116 colorectal cancer and H1299 lung cancer (p53-null) cells (Supplementary Figure 2A). In addition, gastric cancer tissues carrying missense p53 mutations also expressed higher levels of uc002kmd.1 (Supplementary Figure 2B). The uc002kmd.1 promoter contains mutant p53-binding motifs that have been revealed previously (28), and chromatin immunoprecipitation (ChIP) assay confirmed that mutant p53 R248W could bind to uc002kmd.1 promoter in vivo (Supplementary Figure 2C&D). In fact, mutation of p53 is one of the most common steps in gastric carcinogenesis (29), and the involvement of STAT1 proto-oncogene in GC have also been suggested by multiple studies (30,31). These findings suggest that cancer-related TFs may participate in modulating the expression patterns of lncRNAs in GC, analyzing ChIP-seq data seems to be useful for interpreting the potential effects of TFs in lncRNA expression.

**LncRNA-based GC sample prediction**

The protein-coding gene expression profiles have been thoroughly investigated for their abilities to diagnose cancers or discriminate between cancer types, but the efficacy of lncRNAs for such purposes has rarely been reported. Here we applied a widely-used “nearest shrunken centroid method” to classify GC and normal tissues according to their lncRNA or mRNAs expression profiles. Interestingly, lncRNAs displayed equal predictive power as mRNAs on discriminating cancerous and normal tissue (lowest error rate=0.196 for both sets) (Figure 1B). The trained prediction...
signature included 9 lncRNAs, wherein the most upregulated was uc002kmd.1 (Entrez gene ID: AX721193) (**Figure 1C&D**), a long intergenic noncoding RNA (lincRNA) sitting on the shorter arm of chromosome 18 (924-bp long). We used real-time quantitative PCR (RT-qPCR) to quantify the level of uc001kmd.1 in 48 normal gastric mucosa and paired gastric cancer mucosa, and confirmed the significant upregulation of uc002kmd.1 in GC ($P<0.0001$, **Figure 1E**). Further, receiver operating characteristic (ROC) curves was determined to evaluate the sensitivity and specificity of uc002kmd.1 expression in predicting gastric cancer tissues from normal tissues. Notably, uc002kmd.1 displayed considerable predictive significance, with an area under curve (AUC) of 0.714 (**Figure 1F**). Given the cancer-predictive value of this RNA, it is hereafter referred to as GAPLINC (Gastric Adenocarcinoma Predictive Long Intergenic Non-Coding RNA).

**GAPLINC upregulation associates with shorter survival of GC patients**

To test whether GAPLINC expression is correlated with poor prognosis of GC, the expression level of GAPLINC was evaluated by in situ hybridization (ISH) in 90 GC patients with different clinicopathological features (**Figure 2A&B**). GAPLINC level was higher in gastric cancer tissues compare with paired normal gastric tissue based on ISH (**Figure 2C**). The gastric cancer patients were then stratified according to GAPLINC expression level (median split) and compared for different clinicopathological features (age, sex, tumor size, lymph node status, distant metastasis and survival time). The average tumor size in the GAPLINC-high expression group was significantly larger than that in the GAPLINC-low expression group (**Figure 2D**, Mann-Whitney test, $P=0.0097$). Moreover, the occurrence of severe lymph node invasion was more frequent in the GAPLINC-high expression group (Chi-square test, $P=0.0319$, **Figure 2E**). In addition, high expression of GAPLINC associated with shorter patient survival (**Figure 2F**, $P<0.01$, Mantel-Cox test), and the association was stronger than a protein marker that we reported previously (synbindin, $P=0.0468$ Mantel-Cox test) (10). Receiver operating characteristic (ROC) curves was determined to evaluate the sensitivity and specificity of the survival prediction based on the lncRNA ISH intensity and AJCC stages (**Figure 2G**). Interestingly, the area under curve (AUC) for GAPLINC-based prediction was higher than AJCC-based prediction (0.758 vs 0.682), and combine both indexes could further improve the survival prediction (AUC=0.794).
The American Joint Committee on Cancer (AJCC) TNM staging system has been widely accepted as a powerful predictor of treatment response and survival in GC, thus it is of interest to test whether the prognostic value of the GAPLINC is independent of AJCC stage. Multivariable Cox regression analysis adjusting AJCC stage and other factors confirmed the association between GAPLINC expression and shorter survival (hazard ratio=1.539, 95% confidence interval=1.219 to 1.944, P<0.01, Supplementary Table 4).

**GAPLINC is required for efficient proliferation and invasion of GC cells**

To test if GAPLINC is required for maintenance of malignant phenotypes of GC cells, specific siRNAs were used to knockdown GAPLINC expression in two GC cell strains MGC-803 and SGC-901, which express higher level of GAPLINC (Supplementary Figure 2E). Transwell assay revealed a substantial decrease in the number of cells that penetrated the porous filter, suggesting impaired invasion ability for both cell lines (Figure 3A-C). Meanwhile, cDNA-mediated ectopic expression of GAPLINC significantly increased the invasiveness of both cell lines (Figure 3A-C). The CCK8-based viability assay detected significant decrease in the proliferation of MGC-803 and SGC-901 cells after knockdown of GAPLINC, while overexpression of GAPLINC dramatically promoted the cells proliferation (Figure 3D&E, Supplementary Figure 3). Flow cytometry assay indicated that suppression of GAPLINC induced the increase of cell apoptosis by PE-conjugated annexin V staining and FACS (Figure 3E-H). These findings suggest that GAPLINC may not only be a potential marker, but also play a driving role in GC development.

**GAPLINC regulates cell invasion by controlling CD44 expression**

In order to probe the GAPLINC-associated pathway on an unbiased basis, we investigated the gene expression profiles of gastric cancer cells that were suppressed for GAPLINC expression (schematic shown in Figure 4A). To this end, the MGC-803 cells were treated with specific small interfering RNAs (siRNAs) for GAPLINC, and the levels of all mRNAs were measured by Affymetrix Human Genome U133 Plus 2 microarrays (triple repeats for each condition, data accessible via GEO #GSE51651). The GAPLINC-associated pathways were determined by gene set enrichment analysis.
(GSEA), which determines whether different pathways (sets of genes) show statistically significant differences between two biological states (32). Among the significantly affected pathways, “regulation of cell migration” was assigned with the highest enrichment score (Figure 4B&C). By analyzing the correlation between GAPLINC and mRNAs in the lncRNA microarray dataset (containing 10 tumors and 10 normal tissues), we found CD44 with the highest correlation coefficient in this pathway (Pearson correlation R=0.810, P<0.0001, Figure 4D). We used reverse transcription quantitative-PCR (RT-qPCR) to measure the expression of GAPLINC and CD44 in 15 gastric cancer tissues, 15 normal tissues, and 27 chronic gastritis tissues (Figure 4E). The result confirmed the strong correlation between GAPLINC and CD44 (Pearson correlation R=0.827, P<0.0001).

Next, we manipulated GAPLINC expression in gastric cancer cells and monitored its effect on CD44 expression. Knockdown of GAPLINC by specific siRNAs substantially decreased CD44 expression, whereas ectopic expression of GAPLINC caused elevation in CD44 mRNA level (Figure 4F&G). Consistently, Western Blot revealed that knockdown/overexpression of GAPLINC respectively decreased/increased CD44 protein level in gastric cancer cells (Figure 4H). By TransWell assay, we found that the pro-invasion behavior of GAPLINC is neutralized by knockdown of CD44 (Figure 4I&J). Consistently, the suppression effect of GAPLINC knockdown on cell invasion could be rescued by ectopic expression of CD44 (Supplementary Figure 4). These findings suggest that GAPLINC confers CD44-dependent effects in gastric cancer cells.

**GAPLINC regulates CD44 expression by competing for miR-211-3p**

To probe the mechanism for GAPLINC-regulated CD44 expression, we firstly tested whether GAPLINC could modulate the transactivation of CD44 mRNA. The promoter of CD44 was cloned upstream of a luciferase reporter gene, and the resultant construct was co-transfected with GAPLINC in gastric cancer cells. As a result, GAPLINC did not affect the transactivation of CD44 promoter (Figure 5A), suggesting that GAPLINC may regulate CD44 mRNA after it is transcribed.

Since many lncRNAs function as natural “micro RNA sponges” that protect mRNAs by competing for their targeting microRNAs, it is worthy to test if GAPLINC may play such a role. Interestingly, a micro RNA, namely miR-211-3p, was predicted
to target both CD44 and GAPLINC, with remarkable binding energy estimated by the widely-used RNAup algorithm as in Figure 5B (33). To validate the effects of miR-211-3p, we cloned the 3’ UTR of CD44, mutant 3’ UTR of CD44 and GAPLINC downstream of a luciferase gene, and co-transfected these reporters with miR-211-3p mimics in gastric cancer cells. As expected, miR-211-3p significantly decreased the luciferase signals of both reporters (Figure 5C&D). However, miR-211-3p had no effect on mutant 3’UTR of CD44 (Figure 5E), suggesting the specificity of miR-211-3p binding site on CD44 3’UTR. Further, treatment of gastric cancer cells by miR-211-3p mimics significantly decreased CD44 and GAPLINC RNA levels (Figure 6A&B), conforming that miR-211-3p could target GAPLINC and CD44. The effect of miR-211-3p on CD44 was also confirmed by Western Blot (Figure 6C).

Further, we cloned 3’UTR of CD44 into a luciferase reporter and cotrasfected the construct with GAPLINC siRNA or control siRNA. As a result, knockdown of GAPLINC significantly reduced the luciferase intensity, suggesting that GAPLINC is required for the abundant expression of CD44 (Figure 6D). Interestingly, the pro-CD44 effect of GAPLINC could be neutralized by miR-211-3p in MGC803 and SGC7901 cells (Figure 6E-G), suggesting that GAPLINC confers miR-211-3p-dependent effects on CD44 expression. Finally, we studied the relationships between miR-211-3p and its target RNAs (both CD44 and GAPLINC) in tumor tissue samples. The expression level of miR-211-3p negatively correlated with CD44 and GAPLINC RNA levels (Pearson R=-0.33 and -0.25, respectively), which strongly supported our notion that miR-211-3p negatively regulates CD44 and GAPLINC in gastric cancers (Figure 6H&I).

GAPLINC correlates with CD44 activation in gastric cancer tissues

To confirm the impact of GAPLINC on CD44 in vivo, we generated xenograft models by implanting MGC-803 GAPLINC-KD (stably knockdown GAPLINC) or the control MGC803 vector cells into nude mice. As a result, suppression of GAPLINC produced a marked decrease in the rate of xenograft subcutaneous tumor growth (Figure 7A-C, Supplementary Table 5). Intriguingly, knockdown of GAPLINC dramatically decreased the level of CD44 (Figure 7D&E) and increase the level of miR-211-3p in xenograft tumors (Figure 7F). These data confirmed the effect of GAPLINC on CD44 activation and strongly support our notion that GAPLINC contributes to the malignant phenotype by activating CD44.
DISCUSSIONS

Despite recent progresses in discovering cancer-related lncRNAs, few lncRNAs have been characterized for their exact roles in gastric cancer (GC). Our study has provided a landscape of lncRNA deregulation mapped with associated CNVs and TFs, and this may facilitate further exploration of functional lncRNAs in GC. The genes with causal roles in gastric cancer are often located in the genomic regions with CNVs (34). We found that approximately one third of all upregulated lncRNA genes sit in recurrently amplified regions in GC, which suggests a considerable contribution of genomic-level alteration to the aberrant expression of lncRNAs in GC. In addition to CNVs, many transcription factors have been found associated with the aberrant regulation of genes in gastric cancer (10). One outstanding example is mutant p53, which loses the wild-type transactivity but adopts a novel gain-of-function (GOF) transcriptome to promote cancer development (35). We found that mutant p53 can induce the expression of GAPLINC, suggesting that mutant p53 GOF transcriptome may involve not only protein-coding genes, but also lncRNA genes. Overall, the combination of lncRNA expression, CNV recurrence and TF binding might be beneficial for discovering cancer-related lncRNAs.

We have presented a proof-of-principle study for the prediction of cancer/normal tissues using lncRNA expression profiles. Previous efforts have thoroughly explored the feasibility of using mRNA expression profiles for predicting cancer or discriminating between cancer subtypes, but lncRNA-based predictive biomarkers are rarely reported. This might be due to the assumption that mRNAs are more likely to be functional and thus may reflect the status of various biological pathways. However, our results suggest that lncRNA expression signatures are at least comparable with mRNAs in sample stratification. In fact, lncRNAs have an obvious merit of their relatively simple functional layout as transcriptional levels. In many cancers, the functions of protein-coding genes may also be affected by mutations and protein translation/modifications. Taking the TP53 gene as an example, the upregulation of p53 mRNA or protein in cancer often associates with somatic mutation of TP53 (occurring in 50% cancers), rather than functional activation of this tumor suppressive pathway (36). In the form of measured RNA expression (by PCR or hybridization based assays), lncRNAs may better reflect the biological status of cancer
cells than protein-coding genes, and thus should be further studied as predictive biomarkers.

Importantly, GAPLINC was the most upregulated gene in the predictive model containing 9 lncRNAs, and its expression alone could also predict GC with considerable accuracy. GAPLINC expression is required for the efficient invasion and proliferation of GC cells, suggesting the functional involvement of GAPLINC in GC. By suppression of GAPLINC and microarray study, we identified a robust correlation between GAPLINC and CD44, a well-characterized gene involved in cancer proliferation, migration and angiogenesis. Both in vitro and in vivo data demonstrated that GAPLINC regulates CD44 expression by competing for miR-211-3p, which targets CD44 for degradation (model illustrated in Figure 7G). We have also demonstrated by xenograft models that targeting of GAPLINC could suppress tumor progression in vivo. These findings reveal a regulatory network involving different forms of RNAs that is imbalanced in gastric cancers, and suggest GAPLINC as a potential biomarker and therapeutic target for this deadly disease.

In conclusion, our study identifies GAPLINC as a potential biomarker in gastric cancer and reveals its pivotal regulatory effect on the expression of CD44 oncogene.

ACKNOWLEDGEMENT
We thank Wenxin Qin and Nan Shen for their helpful comments on our manuscript.

FUNDING
This project was supported by grants from National Natural Science Foundation of China (30971330, 31371420, 81320108024, 81000861, 81322036, 81372232 and 81272383); the Program for Innovative Research Team of Shanghai Municipal Education Commission; Shanghai “Oriental Scholars” project (2013XJ); Shanghai Science and Technology Commission “Pujiang Project” (13PJ1405900) and Shanghai Natural Science Foundation (12ZR1417900). The sponsors of this study had no role in the collection of the data, the analysis and interpretation of the data, the decision to submit the manuscript for publication, or the writing of the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

REFERENCES


FIGURE CAPTIONS

Figure 1. Deregulated lncRNAs predicts gastric cancer from normal tissues

(A) An overview of deregulated lncRNAs in gastric cancer (GC) mapped with recurrent copy number variations (CNVs) and cancer-related transcriptional factor (TFs). The histogram in blue shows all deregulated lncRNAs in GC (criteria: \( P < 0.05 \) and fold change >1.5, gene names labeled in outermost layer). The scatter plot shows CNVs of genes encoding lncRNAs (outer layer for amplification and inner layer for deletion). The upregulated lncRNAs that associate with gained CNV are labeled in red above the histogram. The links in the innermost layer indicate preferential binding of three TFs (mutant p53 and STAT1) to the promoters of upregulated lncRNAs in GC.

(B) Misclassification error curves of predictive models using mRNAs (upper panel) and lncRNAs (lower panel) in the cross validation process. The error rates for classification of cancerous and normal tissues have been respectively plotted in red and in green. The lower x-axis show the threshold for shrinking the centroids (parameter for classification algorithm), and the upper x-axis indicate the number of genes left in the model (corresponding to each threshold value). Both prediction models reached the lowest error rates of 0.196.

(C) Differential expression of lncRNAs included in the predictive model for GC. Genes were ranked by the differences in their average expression levels in normal (plotted in green) and cancer (in red) tissues. GAPLINC was recognized as the most upregulated lncRNA included in the predictive model.

(D) Scatter plot showing the expression levels of predictive lncRNAs in normal (circles in green) and cancer (in red) samples.

(E) Relative expression level of GAPLINC using real-time PCR in 48 paired normal gastric tissues and gastric cancer tissues, which indicated significantly higher expression level of GAPLINC in gastric cancer tissues by paired ttest. \( P < 0.0001 \)

(F) ROC curve for prediction of gastric cancer using RT-qPCR based GAPLINC expression level. The area under curve (AUC) was 0.714, with 95% confidence interval and P value indicated.
**Figure 2. GAPLINC expression correlates with poor outcome of gastric cancers.**

(A) In situ hybridization (ISH) of GAPLINC in normal gastric mucosa or gastric cancer tissues. Paraffin-embedded tissue sections were stained using specific probe for GAPLINC in purple-blue.

(B) In situ hybridization of U6 spliceosomal RNA in gastric cancer tissues or normal gastric mucosa as control. Paraffin-embedded tissue sections were stained using specific probe for U6 in purple-blue.

(C) Statistical analysis of GAPLINC expression in 90 paired normal and cancerous gastric tissues. The y-axis indicates staining intensity of GAPLINC. The expression level of GAPLINC was significantly higher in cancerous tissues (P<0.0001, paired t-test).

(D) Statistical analysis on the size of gastric cancers in GAPLINC-low and -high expression groups. The average tumor size in two groups was compared using the Mann-Whitney test (P<0.0001).

(E) Relevance of GAPLINC expression to clinicopathological features of gastric cancers. The patients were classified into two groups according to GAPLINC expression levels. The numbers of patients, distant metastasis, average tumor volume, and severe invasion into lymph nodes (3/5 affected) are displayed in each group. The P-values indicating statistical significance of difference between the two groups are also shown in the table.

(F) Survival of patients in GAPLINC-low expression group and -high expression group. The survival time of patients after surgery was compared between groups using the Mantel-Cox test, which indicated significantly longer survival of patients in the GAPLINC-low expression group (P<0.0001).

(G) ROC analysis of in situ hybridization based GAPLINC expression level for survival prediction of gastric cancer patients.

**Figure 3. GAPLINC is required for efficient proliferation and invasion of GC cells.**

(A,B,C) GAPLINC regulates invasion of gastric cancer cells. The expression of GAPLINC was suppressed by specific siRNAs or upregulated by cDNA vector in human gastric
cancer MGC-803 (A) and SGC-7901 cells (B). Transwell assay was used to determine the invasion of gastric cancer cells. The crystal violet-stained images show the number of cells penetrated the porous membrane, and zoomed sections (surrounded by dashed lines) are shown in the lower panels. Statistical result based on 3 independent experiments is indicated in (C).

(D,E) Knockdown of GAPLINC inhibited proliferation of gastric cancer cells, while ectopic expression of GAPLINC accelerates the growth of gastric cancer cells. The MGC-803(D) and SGC7901 cells (E) were transfected with GAPLINC siRNAs/cDNA or control siRNA/cDNA, and cell growth rates were determined by CCK-8 viability assay. The x-axis shows the time after transfection, and the y-axis indicates the readout of CCK-8 assay (absorbance at 450nm).

(F,G,H) Apoptosis of MGC803 and SGC7901 cells induced by GAPLINC knockdown. Flow cytometric assay based on Phycoerythrin(PE)-conjugated annexin V staining showed increased apoptosis of MGC803 and SGC7901 cells treated by GAPLINC siRNA. Representative FACS images are shown in (F) and (G), and statistics based on 3 independent experiments is shown in (H).

Figure 4. GAPLINC promotes cell migration by regulating CD44.

(A) Schematic flowchart showing the process of microarray study on GAPLINC-associated pathways. Gastric cancer cells were treated by siRNAs for GAPLINC or control siRNAs, and the mRNA expression profiles were determined by microarray. The combination of gene set enrichment analysis (GSEA) and gene expression correlation study identified CD44 in the cell migration pathway as potential regulatory target of GAPLINC.

(B) Knockdown of GAPLINC caused alteration in multiple pathways, and “regulation of cell migration” pathway was found with the highest significance. This is based on the following principle: when most genes in a defined pathway (gene set) are affected, a higher enrichment score is assigned to that pathway.

(C) Enrichment plot of the cell migration pathway in the GSEA analysis. All genes were ranked by their changes in association with GAPLINC knockdown (filled gray
curve in the lower panel), and the positions of migration-associated genes in the list are labeled with vertical lines (middle panel). Most genes in the migration pathway belonged to the upregulated set (heat map red and blue). The P-value for GSEA analysis is provided in the enrichment plot (upper panel).

**D** Correlation between the expression of GAPLINC and CD44 as revealed by microarray study (P<0.0001, Pearson correlation).

**E** The levels of GAPLINC and CD44 were measured by RT-qPCR, and their correlation was determined by Pearson correlation analysis (P<0.0001, R=0.827).

**F,G** Modulating GAPLINC expression significantly affected CD44 mRNA expression. Gastric cancer cells were stably transfected with either GAPLINC shRNA/control shRNA (F), or GAPLINC cDNA/control vector (G), followed by detection of CD44 mRNA using RT-qPCR. Knockdown of GAPLINC decreased CD44 expression (MGC803, P<0.01; SGC7901, P<0.001, Student t-test), while ectopic expression of GAPLINC increased CD44 mRNA level (MGC803, P<0.01; SGC7901, P<0.001, Student t-test).

**H** The expression of GAPLINC was suppressed or enhanced as described above, and the expression level of CD44 protein was determined by Western Blot. The level of α-tubulin was also detected as loading control.

**I,J** The pro-invasion effect of GAPLINC is neutralized by suppressing CD44 expression. Gastric cancer cells were transfected with GAPLINC cDNA in presence or absence of siRNAs for CD44, and the invasion ability of cells was determined by Transwell assay. Statistical result based on 3 independent experiments is indicated in (I), and representative Transwell cell staining images are shown in (J). The magnified sections are shown in lower panels.

**Figure 5. Dual roles of miR-211-3p in regulating GAPLINC and CD44.**

**A** GAPLINC displays no effect on the transactivity of CD44 promoter. The promoter of CD44 was cloned upstream of a luciferase reporter gene, and the resultant construct was co-transfected with GAPLINC in gastric cancer cells.

**B** The RNAup algorithm predicted potential binding of miR-211-3p to GAPLINC and
to CD44, with considerable sequence complementary in the indicated regions (binding energy indicated below).

(C,D) Both GAPLINC and CD44 are targeted by miR-211-3p. The 3’UTR of CD44 mRNA and GAPLINC were respectively inserted downstream of a luciferase gene. The reporter vector was co-transfected with a renilla luciferase vector (for normalization) to gastric cancer cells, which were treated by miR-211-3p mimics or control mimics. The luciferase signals (firefly/renilla) of both CD44 3’UTR (C) and GAPLINC reporter (D) genes were significantly decreased when cells were treated with miR-211-3p (P-values indicated, Student t-test).

(E) Mutant 3’UTR of CD44 is not affected by miR-211-3p. The 3’UTR of CD44 was mutated on the predicted binding site that is shown in (B), and was tested in the luciferase assay as described above. The result show that miR-211-3p did not alter the luciferase signal of the mutant CD44 3’UTR (P-values indicated, Student t-test).

Figure 6. GAPLINC regulates CD44 via competing for miR-211-3p.

(A,B) MGC803 and SGC7901 Cells were treated with miR-211-3p or control mimics, followed by detection of GAPLINC and CD44 levels using RT-qPCR. Treatment by miR-211-3p significantly decreased the levels of CD44 mRNA and GAPLINC in both MGC803 (A) and SGC7901 cells (B).

(C) The expression of miR-211-3p was enhanced as described above, and the expression level of CD44 protein was determined by Western Blot. The level of α-tubulin was also detected as loading control.

(D) GAPLINC is required for the stability of CD44 3’UTR. The 3’UTR sequence of CD44 mRNA was inserted downstream of a luciferase gene. The reporter vector was co-transfected with a renilla luciferase vector (for normalization) to gastric cancer cells, which were treated by GAPLINC siRNA or control siRNA. The luciferase signal (firefly/renilla) of reporter gene was significantly decreased when cells were treated with GAPLINC siRNA (P-values indicated, Student t-test).

(E,F) Gastric cancer cells were transfected with GAPLINC cDNA in presence or absence of miR-211-3p mimics, and level of cDNA was determined by real time PCR in both MGC803 (E) and SGC7901 cells (F).
(G) Cells were treated as described above, and CD44 protein level was determined by Western Blot. The α-tubulin protein was detected as loading control.

(H, I) The correlation between miR-211-3p and CD44/GAPLIC in 15 gastric cancer tissues, 15 normal tissues, and 27 chronic gastritis tissues. Expression levels of miR-211-3p, CD44 and GAPLINC were determined by RT-qPCR, and Pearson correlation was used to analyze the relationships between miR-211-3p and CD44 (H), and between miR-211-3p and GAPLIC (I). P-values have been indicated in the plot.

Figure 7. Targeting GAPLINC decreased CD44 expression and tumor growth in vivo.

(A, B) Knockdown of GAPLINC inhibited growth of xenograft in nude mice. Human gastric cancer MGC-803 cells with GAPLINC stable knockdown or negative control were injected subcutaneously into the left hip to establish xenograft model. At the last time point (17 days after first injection), tumors in both groups were measured both in situ (a) and after resection (b) are show for both groups.

(C) Statistical analysis of tumor volume in MGC-803 GAPLINC-KD (knockdown) and MGC-803 vector (control) groups (values also shown in Supplementary Table 5). Knockdown of GAPLINC significantly delayed tumor progression as compared to the control group.

(D) RT-qPCR of GAPLINC and CD44 in xenografts treated with GAPLINC stable knockdown clone and control clone. It can be observed that knockdown of GAPLINC dramatically decreased the level of CD44.

(E) Western Blot analysis of xenografts in MGC-803 GAPLINC-KD and MGC-803 vector groups suggest decreased expression of CD44 protein upon stable knockdown of GAPLINC.

(F) RT-qPCR of miR-211-3p in xenografts in MGC-803 GAPLINC-KD and MGC-803 vector groups. It can be observed that knockdown of GAPLINC dramatically increased the level of miR-211-3p (P value indicated, student t-test).

(G) Schematic representation showing GAPLINC-mediated cell proliferation and invasion in gastric cancer. The micro RNA 211-3p has dual target specificity for both GAPLINC and CD44, which leads to degradation of the bound RNA. The high expression of GAPLINC competes to bind miR-211-3p, which reduces the degradation
of CD44 mRNA and leads to increased translation of the functional CD44 protein. The CD44 oncprotein has well-established roles in promoting cancer proliferation, migration and angiogenesis, which mediates the oncogenic effects of GAPLINC in gastric cancer.
Figure 5

A. pGL3-CD44-promoter

B. GAPLINC and CD44 (1787-1803) interactions:

C. pGL3-CD44-3'UTR

D. pGL3-GAPLINC

E. pGL3-CD44-MUT-3'UTR

Relative FUR/L ratio comparisons for MGC803 and SGC7901 cell lines under different conditions.
Long non-coding RNA GAPLINC regulates CD44-dependent cell invasiveness and associates with poor prognosis of gastric cancer

Ye Hu, Jilin Wang, Jin Qian, et al.

Cancer Res  Published OnlineFirst October 2, 2014.

Updated version 
Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-14-0686

Supplementary Material
Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2014/10/03/0008-5472.CAN-14-0686.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.