Molecular Therapy
Original Article

lncRNA GCAWKR Promotes Gastric Cancer Development by Scaffolding the Chromatin Modification Factors WDR5 and KAT2A

Mingzhe Ma,[1,2,8] Yan Zhang,[3,8] Mingzhe Weng,[4,8] Ye Hu,[5] Yi Xuan,[1,2] YiRen Hu,[6] and Kun Lv[7]

[1]Department of Gastric Surgery, Fudan University Shanghai Cancer Center, Shanghai, China; [2]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China; [3]Department of Gastroenterology, Yijishan Hospital, The First Affiliated Hospital of Wannan Medical College, Wuhu, Anhui, China; [4]Department of General Surgery, Xinhua Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China; [5]State Key Laboratory for Oncogenes and Related Genes, Division of Gastroenterology and Hepatology, Ren Ji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; [6]Department of General Surgery, Wenzhou No. 3 Clinical Institute of Wenzhou Medical University, Wenzhou People's Hospital, Wenzhou, Zhejiang, China; [7]Central Laboratory of Yijishan Hospital, The First Affiliated Hospital of Wannan Medical College, Wuhu, Anhui, China

Long noncoding RNAs (lncRNAs) have been demonstrated to play a role in carcinogenesis, but their mechanisms of function remain elusive. We explored the mechanisms of the oncogenic role of *GCAWKR* in gastric cancer (GC) using human tissues and cell lines. The *in situ* hybridization analysis was utilized to determine *GCAWKR* levels in samples from 42 GC patients and real-time qPCR in tissues from 123 patients. The *GCAWKR* levels were modulated in GC cell lines, and relevant biological and molecular analyses were performed. Levels of the *GCAWKR* were upregulated in GC tissues compared with normal tissues and associated with tumor size, lymph node metastasis, TNM stage, and patient outcomes. *GCAWKR* affected cell proliferation and cell invasion in multiple GC models. Mechanistically, *GCAWKR* bound WDR5 and KAT2A and acted as a molecular scaffold of WDR5/KAT2A complexes, modulating the affinity for WDR5/KAT2A complexes in the target gene's promoter region. Thus, our data defined a mechanism of lncRNA-mediated carcinogenesis in GC, suggesting new therapeutic targets in GC.

INTRODUCTION

Gastric cancer (GC) is the fourth most common cancer by incidence and the third leading cause of cancer death worldwide.[1,2] Although great advances have been made in the diagnosis and management of GC, the prognosis of GC patients remains dismal, especially for patients at advanced stages. Tumor recurrence, metastasis, and therapy resistance are the main causes of cancer-related death.[3] Although a growing number of studies have tried to clarify the pathophysiological mechanisms of GC progression and the development of early GC diagnostic and prognostic biomarkers, the improvement in the prognosis of GC patients at advanced stages is considered unsatisfactory.[4] Therefore, more efforts are required for the development of novel biomarkers and targets for GC diagnosis and therapy.

Long noncoding RNAs (lncRNAs) are a class of noncoding RNAs longer than 200 nucleotides in length and with no protein-coding potential.[5] Accumulating evidence has demonstrated that lncRNAs play an important role in a wide array of cellular processes, including carcinogenesis.[6–8] lncRNAs have been reported to regulate genes at multiple levels, including transcriptional modulation by recruiting chromatin-modifying complexes[9] and post-transcriptional regulation by interacting with microRNAs (miRNAs),[10] mRNAs,[11] or proteins.[12] Although there have been up to 548,640 lncRNAs that have been annotated according to NONCODE V5,[13] only a small portion of them have been functionally characterized.

In the present study, we analyzed our previous genome-wide lncRNA profiling data in GC tissues and paired normal tissues.[14] We characterized a novel lncRNA (ENST00000431060). The biological functions of the gastric-cancer-associated WDR5 and KAT2A binding lncRNA (lncRNA-*GCAWKR*) was determined in the *in vitro* and *in vivo* models.

RESULTS

lncRNA *GCAWKR* Was Upregulated in GC Tissues

The lncRNA and mRNA profiling data in 10 GC and paired normal tissues that we previously reported[14] can be accessed via Gene Expression Omnibus (GEO): GSE50710. We analyzed the microarray data and used stringent filtering criteria (fold change >2, $p < 0.05$, raw

Received 20 June 2018; accepted 5 September 2018;
https://doi.org/10.1016/j.ymthe.2018.09.002.

[8]These authors contributed equally to this work.

Correspondence: YiRen Hu, Department of General Surgery, Wenzhou No. 3 Clinical Institute of Wenzhou Medical University, Wenzhou People's Hospital, No. 57 Canghou Street, Wenzhou 325000, China.
E-mail: yirenhu@hotmail.com
Correspondence: Kun Lv, Central Laboratory of Yijishan Hospital, The First Affiliated Hospital of Wannan Medical College, 2 West Zheshan Road, Wuhu, 241001 Anhui, China.
E-mail: lvkun315@126.com
Correspondence: Mingzhe Ma, Department of Gastric Surgery, Fudan University Shanghai Cancer Center, 270# Dongan Road, Shanghai 200032, China.
E-mail: mmz666@163.com

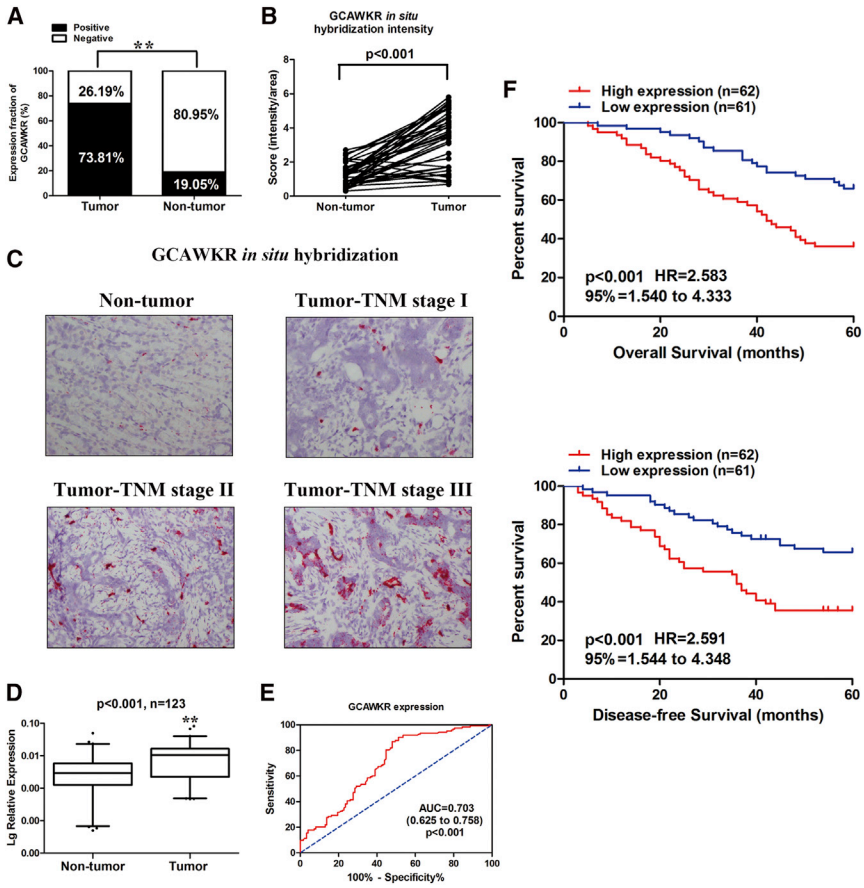


Figure 1. *GCAWKR* Upregulation in GC Tissues

(A) *GCAWKR* expression in GC tissues and adjacent nontumor tissues was determined in ISH assays (cohort 1, n = 42). **p < 0.01. (B) Statistical analysis of *GCAWKR* expression in 42 paired normal and cancerous gastric tissues. The y axis indicates staining intensity of *GCAWKR*. The expression level of *GCAWKR* was significantly higher in cancerous tissues (p < 0.0001, paired t test). (C) ISH of *GCAWKR* in normal gastric mucosa or gastric cancer tissues. Paraffin-embedded tissue sections were stained using specific probe for *GCAWKR*. The expression level of *GCAWKR* was higher in tissues with advanced TNM stage. (D) *GCAWKR* expression was analyzed by qRT-PCR in GC samples and adjacent nontumor tissues (cohort 2, n = 123). *GCAWKR* expression level was normalized to that of β-actin. Horizontal lines in the boxplots represent the medians, the boxes represent the interquartile range, and the whiskers represent the 2.5th and 97.5th percentiles. The significant differences between samples were analyzed using the Wilcoxon signed-rank test. **p < 0.01. (E) ROC curve for prediction of gastric cancer using qRT-PCR-based *GCAWKR* expression level. The AUC was 0.703, with 95% CI and p value indicated. (F) Kaplan-Meier survival analysis of overall survival (OS) and disease-free survival (DFS) in GC patients (p < 0.001 for both OS and DFS) based on *GCAWKR* expression.

signal intensity >1,000). Bioinformatics analysis found that the level of lncRNA-*GCAWKR* was higher in GC tissues than that in normal tissues. *GCAWKR* (linc-ALDH1A3-1, LNCipedia annotation; NONHSAT051130, NONCODE v4; ENST00000431060, Ensembl release 64, Sep 2011) was on human chromosome 15:101390036–101404487. *GCAWKR* consists of five exons and spanned about 3,900 bp. The 5′ and 3′ rapid amplification of cDNA ends (RACE)-PCR was performed to identify the 5′ and 3′ ends of *GCAWKR* (Figure S1). Then, we evaluated the potential coding capability of *GCAWKR*; the results were as follows: although seven short open reading frames (ORF1, 3, 8, 9, 11, 12, and 13) with more than 200 nt were predicted using ORF Finder from the National Center for Biotechnology Information (Figure S2A); none of their AUGs showed the Kozak consensus and no homologous protein sequences were found using a BLAST search; The online protein-coding potential assessment softwares (coding potential calculator; coding potential score, −0.870037; PhyloCSF, codon substitution frequency scores, −15.3106) (Figure S2B) and *in vitro* translation assays (Figure S2C) also confirmed *GCAWKR* has no coding capability. We separated the cytoplasm and nuclear fractions of GC cell lines (SGC7901 and BGC823 cells) and performed PCR analysis. *GCAWKR* was mainly located in the nucleus of GC cells (Figure S2D). We found that the expression of *GCAWKR* was upregulated in GC cell lines compared to normal human gastric epithelial cell line (GES-1) (Figure S3A). The expression level of *GCAWKR* was comparable to the well-known oncogenic lncRNA *MALAT1* (Figure S3B). The predicted secondary structures image with ViennaRNA is presented in Figure S3C.

As human *GCAWKR* consists of three transcripts (*GCAWKR* 1–3; Figure S4A), we examined the expression levels of all *GCAWKR* transcripts in GC cell lines. Only the *GCAWKR*-1 was highly expressed of GC cell lines and GC tissues, while *GCAWKR*-2 and −3 were barely detected in GC cell lines, GC tissues, and adjacent normal tissues (Figures S4B and S4C).

lncRNA *GCAWKR* Correlated with GC Progression

qRT-PCR analysis demonstrated that the expression level of *GCAWKR* increased from normal gastric tissue to intestinal metaplasia (IM), to dysplasia, and to GC (Figure S5A). The expression level of *GCAWKR* was upregulated in GC tissues compared to adjacent normal tissues in another GC dataset from the GEO database (Figure S5B). There was no difference in the expression level of *GCAWKR* between *Helicobacter pylori*-positive and -negative GC tissues (Figure S5C). *In situ* hybridization (ISH) of 42 paraffin-embedded GC cancer tissues and adjacent normal tissues (Table S3, cohort 1) confirmed that *GCAWKR* was upregulated in GC tissues (Figures 1A and 1B) and higher expression of *GCAWKR* was more

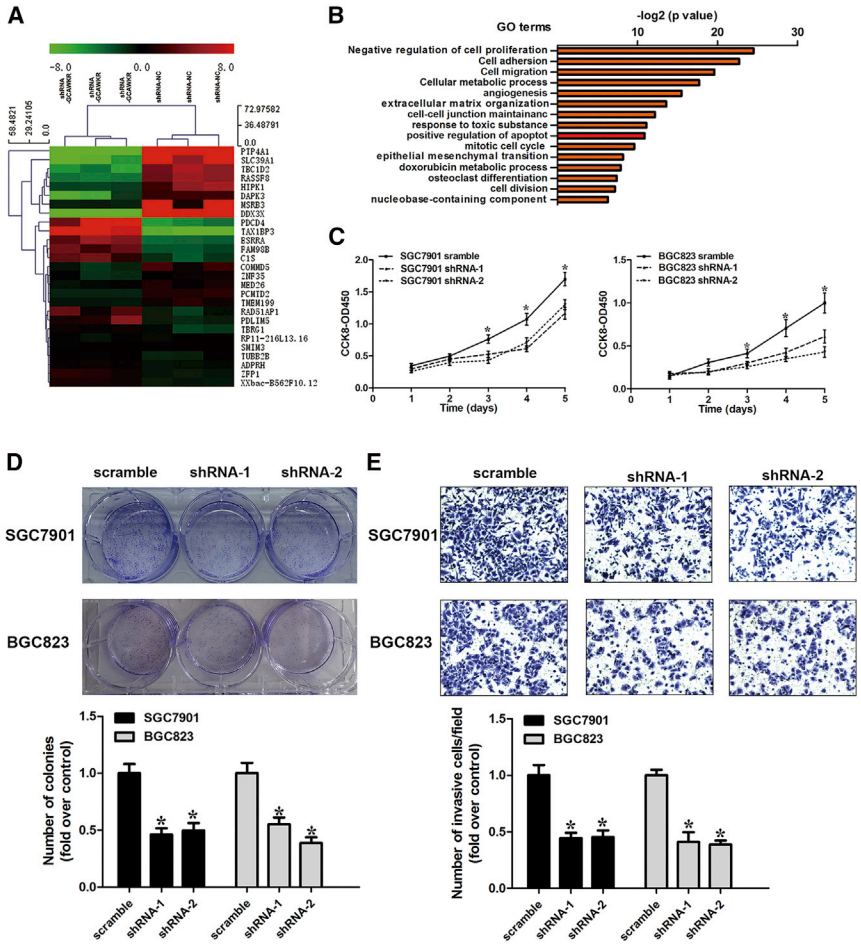


Figure 2. Knockdown of *GCAWKR* Inhibited Cell Proliferation, Colony Formation, and Invasion in GC Cells

(A) Hierarchical clustering analysis of the top 27 genes that were differentially expressed (>2-fold; $p < 0.05$) in sh-NC-treated cells and shRNA-*GCAWKR*-treated cells, with three repeats. (B) Gene ontology analysis for all genes with altered expressions. (C) The cell growth rates were determined with CCK-8 proliferation assays. *GCAWKR* knockdown in SGC7901 and BGC823 cells significantly inhibited cell proliferation. (D) Colony formation assays were used to determine the cell colony formation ability of shRNA-*GCAWKR*-transfected SGC7901 and BGC823 cells. (E) Effects of *GCAWKR* knockdown on cell invasion in the presence of the anti-proliferative drug MMC (5 μM) treatment were determined using the transwell assay. *$p < 0.05$.

frequent in tissues with advanced tumor-node-metastasis (TNM) stage (Figure 1C). As shown in Table S3, a positive correlation among *GCAWKR* expression and tumor size ($p = 0.023$), advanced American Joint Committee on Cancer (AJCC) TNM stage ($p = 0.028$), lymph node status ($p = 0.002$), and gross type ($p = 0.048$) was found. To further investigate the role of *GCAWKR* in GC, we determined the transcript levels of *GCAWKR* in 123 pairs of GC tissues and adjacent normal tissues. We found that *GCAWKR* was upregulated in 78.86% (97/123) GC tissues compared to matched adjacent normal tissues (Figure 1D). Furthermore, we utilized the receiver operating characteristic (ROC) curves to assess the predictive power of *GCAWKR* expression in differentiating GC tissues from normal tissues. It is worth noting that *GCAWKR* manifested itself with considerable predictive significance, with an area under the curve (AUC) of 0.703 (95% confidence interval [CI], 0.625–0.758; $p < 0.001$; Figure 1E). The median *GCAWKR* expression level was used as the cutoff value. Low *GCAWKR* expression levels in 61 patients (with an average ΔCt expression value of 8.842, ranking from 6.560 to 11.093 when compared with β-actin) was classified as values < median ratio. High *GCAWKR* expression in 62 patients (with an average ΔCt expression value of 5.675, ranking from 3.649 to 6.545) was classified as values ≥ median ratio. The high-*GCAWKR*-expression group demonstrated a larger tumor size ($p = 0.012$) and advanced lymph node status ($p = 0.016$), depth of invasion ($p = 0.045$), and TNM stage ($p = 0.031$) (Table S4). High *GCAWKR* expression was associated with a poorer overall survival (OS) ($p < 0.001$; Figure 1F) and disease-free survival (DFS) ($p < 0.001$; Figure 1F). Univariate and multivariate Cox proportional hazards analysis showed that *GCAWKR*, depth of invasion, and TNM stage were independent factors for OS in GC patients (Table S5). *GCAWKR* and TNM stage were identified to be independent prognostic factors for DFS in GC (Table S6).

GCAWKR Promotes GC Progression, and PTP4A1 Is a Downstream Target of *GCAWKR*

To explore the biological function of *GCAWKR* on the progression of GC, we next performed loss-of-function and gain-of-function studies in GC cells. After detecting *GCAWKR* expression in five GC cell lines (Figure S3A), we selected SGC7901 and BGC823 cells lines for manipulation of *GCAWKR* expression (Figures S6A–S6D), as they had moderate expression levels of *GCAWKR*. To elucidate the potential role involved in the oncogenic function of *GCAWKR*, an RNA transcriptome-sequencing analysis was performed in SGC7901 cells that were transfected with *GCAWKR* short hairpin RNA (shRNA) or control shRNA (Figure 2A; data are available via GEO: GSE112586). Top altered cellular pathways were the negative regulation of cell proliferation, cell adhesion and cell migration (Figure 2B). These genes (proliferation-related genes, CDK1, CDC25B, CCND3, and GDF13; migration-related genes, KLF6, MMP9, MAPK4, and HOXB13) were selectively confirmed by qRT-PCR analysis (Figures S6E

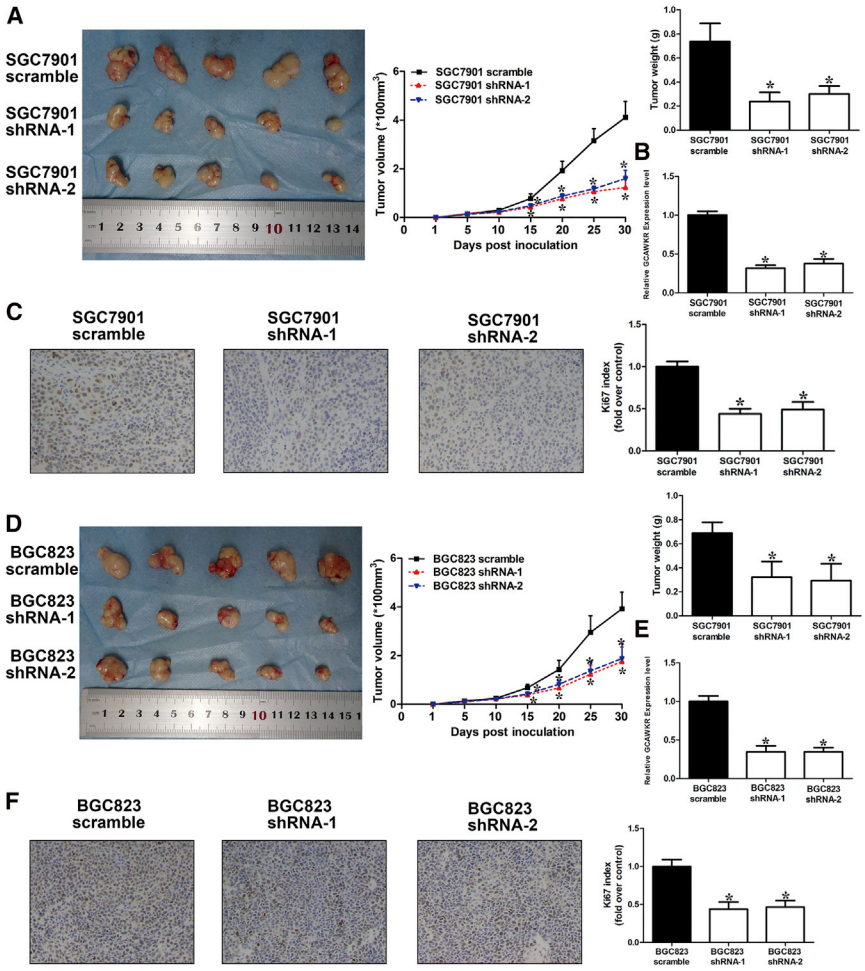


Figure 3. Knockdown of *GCAWKR* Inhibited Tumor Growth *In Vivo*

(A and D) Top left, representative images of tumors formed in nude mice injected subcutaneously with *GCAWKR*-silencing SGC7901 cells (A) or BGC823 cells (D) (n = 5 per group). Top middle, tumor growth curves. Top right, tumor weights. (B and E) *GCAWKR* levels were determined in tumors from mice with qRT-PCR in SGC7901 (B) and BGC823 cells (E). (C and F) Representative images of IHC staining of Ki67 in SGC7901 (C) and BGC823 cells (F) (original magnification, ×200). *p < 0.05.

and S6F). These data suggest that *GCAWKR* may play a role in the development of GC.

Cell-counting kit-8 and colony-formation assays demonstrated that knockdown of *GCAWKR* suppressed the proliferative capacity in SGC7901 and BGC823 cells (Figures 2C and 2D). Conversely, overexpression of *GCAWKR* promoted cell proliferation *in vitro* (Figures S7A and S7B). *GCAWKR*-knockdown GC cells had a significantly higher percentage of Annexin V-positive cells than did cells expressing a scrambled shRNA (Figure S7C). Furthermore, knockdown of *GCAWKR* increased the sensitivity to chemotherapeutics (Figures S7D and S7E). Then, we performed the transwell assays to investigate the effect of *GCAWKR* on GC cell invasive ability. To prevent cell number changes that could potentially affect the outcome of the assays, the transwell assays were performed in the presence of mitomycin C (MMC), which is an anti-proliferative drug. The results demonstrated that the invasive ability in SGC7901 and BGC823 cells was decreased when *GCAWKR* expression was knocked down (Figure 2E), while *GCAWKR*-overexpressing SGC7901 and BGC823 cells exhibited higher invasive ability (Figure S8A).

The oncogenic function of *GCAWKR* was further confirmed in the xenograft mouse tumor models. Knockdown of *GCAWKR* significantly suppressed tumor growth, as illustrated by decreased tumor growth rates, tumor weights, and Ki-67 indexes from the *GCAWKR*-downregulated xenografts compared with that of tumors from the control xenografts (Figures 3A–3F), while *GCAWKR* overexpression promoted tumor growth (Figures S8B–S8E). These data suggest that *GCAWKR* promoted GC cell progression.

Among the differentially expressed genes (SGC7901 cells that were transfected with shRNA-GCAWKR versus shRNA-negative control [NC]), PTP4A1 (protein tyrosine phosphatase type IVA, member 1), located in 6q129, is associated with poor clinical prognosis and could promote cell growth and invasion in cancer.[15–17] Furthermore, PTP4A1 was significantly overexpressed in 100% of gastric carcinomas[18] (Figure S9A) and a variety of cancer tissues (Figure S9B). Immunohistochemical staining analysis confirmed that PTP4A1 was upregulated in GC tissues compared to normal tissues (Figure S9C). A higher expression level of PTP4A1 is correlated with a significantly poorer disease-specific survival (DSS; $p < 0.05$) and DFS ($p < 0.05$) according to data from the KMPlot database (http://www.kmplot.com) (Figures S10A and S10B). A positive correlation between *GCAWKR* and PTP4A1 transcript levels was found in 35 GC patients (cohort 2, $r^2 = 0.461$, $p < 0.001$; Figure 4A). Therefore, we chose PTP4A1 for further investigation. qRT-PCR and western blot analysis showed that *GCAWKR* silencing markedly decreased the expression levels of PTP4A1 (Figure 4B), while *GCAWKR* overexpression increased the expression levels of PTP4A1 in SGC7901 and BGC823 cells (Figure 4C). To determine the regulatory effect of *GCAWKR* on the transcriptional activity of PTP4A1, we cloned the 5′ flanking DNA fragment (~2 kbp) of PTP4A1 promoter region and inserted it into pGL3 enhancer plasmid. Luciferase activity of pGL3 enhancer plasmid was significantly decreased with *GCAWKR* silencing in SGC7901 and BGC823 cells (Figure 4D). It suggests that *GCAWKR* might promote the transcription of PTP4A1 in GC.

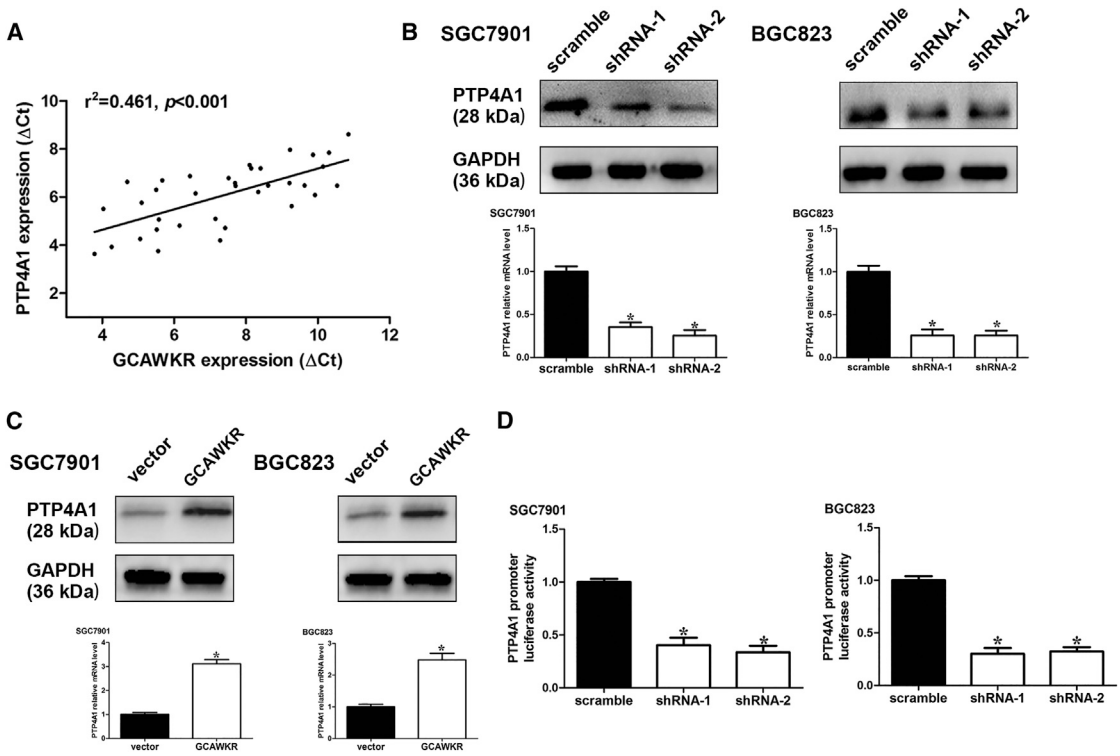


Figure 4. *GCAWKR* Correlates With and Regulates PTP4A1

(A) The correlation between *GCAWKR* transcript level and PTP4A1 mRNA level was measured in 35 GC tissues. The ΔCt values (normalized to β-actin) were subjected to Spearman rank-correlation analysis. (B) qRT-PCR and western blot assays were performed in SGC7901 and BGC823 cells after transfection of *GCAWKR* shRNAs. n = 3, nonparametric Mann-Whitney test. Error bars in the bar graphs represent SD. (C) The PTP4A1 expression was detected with real-time PCR and western blot in SGC7901 and BGC823 cells after transfection of lentivirus harboring the full-length human *GCAWKR* sequence or the empty vector. (D) Luciferase reporter vector was generated by inserting the promoter region (−2,000 to 0 bp) of the PTP4A1 gene. The reporter vectors were then cotransfected into SGC7901 and BGC823 cells with *GCAWKR* or control shRNAs. Cells were harvested for luciferase activity assay. Results shown are the mean ± SD of triplicate determination from three independent experiments. *p < 0.05.

GCAWKR Functions as a Scaffold for WDR5 and KAT2A in the Nucleus and Thus Epigenetically Upregulates PTP4A1

Previous reports have demonstrated that a part of lncRNAs function via scaffolding chromatin modifier to facilitate epigenetically modulation of gene expression.[19] As *GCAWKR* was mainly located in the nucleus, the interaction probabilities of *GCAWKR* and a panel of RNA binding proteins, including HuR, Ago, STAU1, TDP-43, EZH2, SUZ12, SUV39H1, LSD1, SIRT1, DNMT1, CoREST, DNMT3a, DNMT3b, SETDB1, WDR5, HAT1(KAT1), GCN5(KAT2A), PCAF(KAT2B), CBP(KAT3A), and p300(KAT3B) were predicted via the online RNA-Protein Interaction Prediction Program (Figure S11A). We found that *GCAWKR* potentially interacts with WDR5 and KAT2A (more stringent filtering criteria, reading frame (RF), and support vector machine [SVM] score > 0.5). RNA pull-down assay (Figure 5A) confirmed that *GCAWKR* specifically binds to WDR5 and KAT2A. WDR5, a key component of SET/MLL (SET-domain/mixed-lineage leukemia) histone-methyltransferase complexes, have an essential role in histone H3 Lys 4 (H3K4) trimethylation and subsequent transcriptional activation.[20] KAT2A (GCN5), a histone acetyltransferase, modulates histone H3K9 acetylation levels and gene transactivation.[21] Reports have shown that lncRNAs could function as scaffold for RNA-binding proteins.[22,23] As WDR5 and KAT2A exert oncoepigenic effects in tumorigenesis,[24–26] we speculated that *GCAWKR* might coordinately interact with WDR5 and KAT2A. We found that the binding activity mapped to 3′ end segment (nucleotides 3,000–3,893) of *GCAWKR* (Figure 5B). The 3′ end segment formed a stable stem-loop structure according to RNA folding analyses[27] (Figure S11B), which may provide the spatial conformation for the interaction. Furthermore, the effect that knockdown of *GCAWKR* had on the protein levels of WDR5 and KAT2A was not significant (Figure S11C). To further consolidate the interaction, we analyzed the interaction between *GCAWKR* and WDR5/KAT2A in the nucleus, and the radioimmunoprecipitation (RIP) assays showed that *GCAWKR* specifically binds to WDR5 and KAT2A (Figure 5C). Significant enrichment of *GCAWKR* was observed in the anti-WDR5 and anti-KAT2A immunoprecipitates compared with GAPDH (Figure S11D). These data suggests that *GCAWKR* might serve as a scaffold and specifically binds with WDR5 and KAT2A in GC. In addition, coimmunoprecipitation (coIP) assays demonstrated that WDR5 retrieved KAT2A and KAT2A retrieved WDR5 (Figure 5D), which suggests that WDR5 and KAT2A physically bind with each other. *GCAWKR* silencing

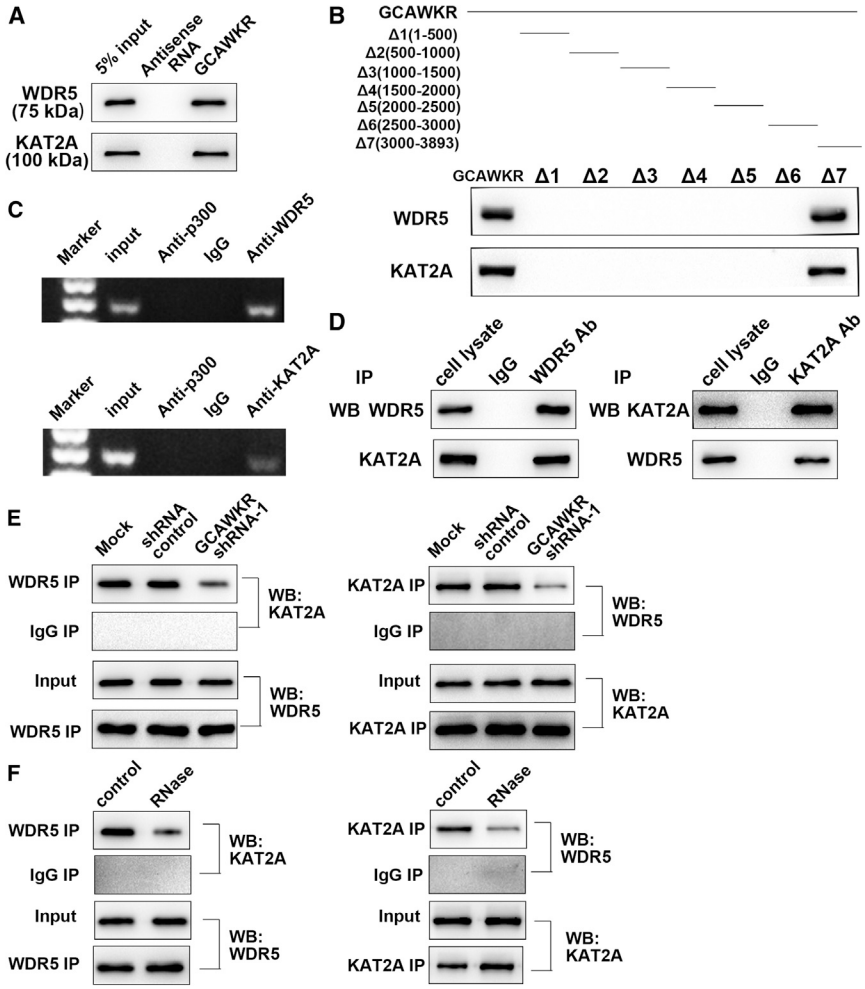


Figure 5. *GCAWKR* Interacts with WDR5 and KAT2A

(A) Biotinylated GCAWKR or antisense RNA was incubated with cell extracts of SGC7901 cells, targeted with streptavidin beads, and washed, and the associated proteins were resolved on a gel. Western blot analysis detected the specific association of WDR5 or KAT2A and GCAWKR (n = 3). (B) Biotinylated RNAs corresponding to different fragments of GCAWKR were incubated with SGC7901 cell lysates, and associated proteins were resolved electrophoretically. Western blot analysis of the specific association of WDR5 or KAT2A and GCAWKR (n = 3). (C) RIP experiments were performed using the WDR5, KAT2A, or p300 antibody for immunoprecipitation and a primer to detect GCAWKR. RIP enrichment was determined relative to the input controls (n = 3). (D) Coimmunoprecipitation detected the interaction between WDR5 and KAT2A in SGC7901 cells. The 20% of input and WDR5 or KAT2A immunoprecipitates were separated by SDS-PAGE. The interaction between WDR5 and KAT2A was confirmed by western blot (n = 3). (E and F) Immunoprecipitation assays were used to determine the interaction between WDR5 and KAT2A after transfection of GCAWKR shRNA (E) or RNase treatment (F). n = 3. For RNase treatment, cells were treated with RNase A (100 μg/mL) at 37°C for 1 hr.

(Figure 5E) or RNAase treatment (Figure 5F) greatly suppressed the interaction between WDR5 and KAT2A. Our data illustrated that *GCAWKR* mediated the interaction between WDR5 and KAT2A.

To determine whether histone acetylation is involved in PTP4A1 transcription, we found that the treatment of Trichostatin A (an inhibitor of class I and II histone deacetylases) induced a significant upregulation of PTP4A1 in SGC7901 and BGC823 cells (Figure S12A). To determine which group of histone acetyltransferases (HATs) (HATs can be classified into five groups: EP300/CREBBP, basal transcription factors, MYST, GNAT, and nuclear receptor cofactors)[21] might be involved in the transcription of PTP4A1, we treated GC cells with three pharmacological HAT inhibitors: CPTH2 (an inhibitor of KAT2A network), curcumin (an inhibitor of EP300/CREBBP), and anacardic acid (an inhibitor of both EP300 and PCAF). We found that CPTH2 abolished the increase in luciferase activity of pGL3 enhancer plasmid of PTP4A1 induced by *GCAWKR* overexpression in BGC823 cells (Figure S12B), which suggests that KAT2A might play a role in this process. To further consolidate the hypothesis, we performed western blot analysis and found that WDR5 or KAT2A knockdown (Figure S12C) obviously attenuated the *GCAWKR*-induced PTP4A1 upregulation in SGC7901 (Figure 6A) and BGC823 cells (Figure S12D). It indicates that WDR5 and KAT2A are involved in the coregulation of PTP4A1. Furthermore, WDR5, KAT2A, or *GCAWKR* silencing obviously inhibited the luciferase activity of pGL3 enhancer plasmid of PTP4A1 (Figure 6B). The chromatin IP (ChIP) assay confirmed the binding of WDR5 and KAT2A to the promoter region of PTP4A1 (Figure 6C). *GCAWKR* silencing markedly suppressed the binding ability of WDR5 and KAT2A (Figure 6D). In addition, we observed the significant enrichment of H3K9ac and H3K4me3 at the promoter region of PTP4A1 (Figure 6E). GCAWKR silencing induced obvious decreased H3K9ac and H3K4me3 levels at the promoter region of PTP4A1 (Figure 6F). These results suggest that *GCAWKR* serves as a molecular scaffold for WDR5 and KAT2A and activate the transcription of PTP4A1.

GCAWKR Promotes GC Development by Upregulating PTP4A1 Expression

Next, we would like to investigate whether PTP4A1 modulates the biological function of *GCAWKR* in GC. *GCAWKR* overexpression significantly promoted the GC cell proliferation (Figure 7A), colony formation (Figure 7B), and invasion (Figure 7C) in GC cells, and PTP4A1 knockdown (Figure S13A) attenuated the oncogenic effect induced by the *GCAWKR* overexpression. Furthermore, the growth of tumors from *GCAWKR*-overexpressing xenografts was blocked by the downregulation of PTP4A1 (Figure 7D). On the contrary,

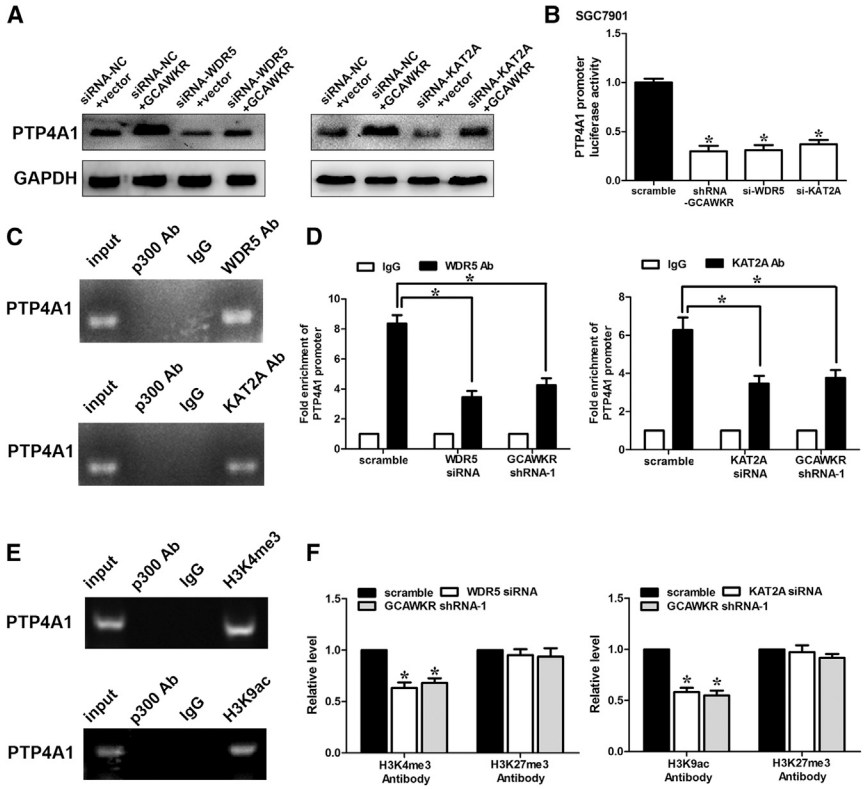


Figure 6. *GCAWKR* Functions as a Link between WDR5/KAT2A and PTP4A1

(A) Western blot assay was performed to determine PTP4A1 protein level after transfection of WDR5/KAT2A siRNA, lentivirus harboring the full-length human *GCAWKR* sequence in SGC7901 cells (n = 3). (B) Luciferase reporter vector was generated by inserting the promoter region (−2,000 to 0 bp) of the PTP4A1 gene. The reporter vectors were then cotransfected into SGC7901 cells with *GCAWKR* or WDR5/KAT2A siRNA. Cells were harvested for luciferase activity assay. Results shown are the mean ± SD of triplicate determination from three independent experiments. *p < 0.05. (C) The PTP4A1 DNA was detected in the chromatin sample immunoprecipitated from SGC7901 cells using an antibody against WDR5, KAT2A, or p300. n = 3. (D) Real-time PCR of the ChIP samples shows the binding efficacy of WDR5 or KAT2A to the PTP4A1 gene promoter after transfection of WDR5/KAT2A siRNA or shRNA-*GCAWKR* in SGC7901 cells. Data are shown as the mean ± SD from three independent repeats. (E) The PTP4A1 DNA was detected in the chromatin sample immunoprecipitated from SGC7901 cells using an antibody against H3K4me3 or H3K9ac. (F) Real-time PCR of the ChIP samples shows the binding efficacy of H3K4me3, H3K9ac, or H3K27me3 to the PTP4A1 gene promoter after transfection of WDR5/KAT2A siRNA or shRNA-*GCAWKR* in SGC7901 cells. Data are shown as the mean ± SD from three independent repeats.

PTP4A1 overexpression (Figure S13B) markedly ameliorated the tumor-suppressive effect of *GCAWKR* silencing (Figures S13C and S13D). Thus, *GCAWKR* promotes GC development may partially depend on the upregulation of PTP4A1 expression.

DISCUSSION

The research on the underlying molecular mechanism of GC progression has identified some protein-coding genes[28–30] and miRNAs[31,32] as potential biomarkers and therapeutic targets. However, there has been a poor overlap between these biomarkers and prognostic indicators identified.[33–35] Thus, we suppose that it might to be feasible to establish a more accurate prognostic gene signature with various kinds of transcripts. In light of the notion that lincRNAs are abundantly expressed in mammalian cells and play a vital role in cell development and human disease, we hypothesize that lncRNAs might be potential biomarkers and prognostic indicators in GC. Up till now, several lncRNAs, including *GAPLINC*[14] and HOXA11-AS,[23] have been demonstrated to be involved in the development of GC. *GAPLINC* overexpression is associated with the poor survival of GC patients, and *GAPLINC* regulates CD44 as a molecular decoy for miR211-3p.[14] Patients with high HOXA11-AS expression had a shorter survival and poorer prognosis.[23] HOXA11-AS promotes proliferation and invasion of GC by scaffolding the chromatin modification factors PRC2, LSD1, and DNMT1.[23] FOXD2-AS1 was upregulated markedly in GC and positively correlated with a poor prognosis.[36] FOXD2-AS1 promoted GC tumorigenesis partly through EZH2 and LSD1 mediated EphB3 downregulation. Yet, only a very small portion of lncRNAs have been functionally characterized. lncRNAs were reported to bind to EZH2, LSD1, or DNMT.[23,36] However, most were focused on the histone-methyltransferase complexes; there were few studies on histone acetyltransferase. Yet, histone acetylation and methylation are both vital parts of histone modification. The finding of this study was consistent with our previous study.[14] It further highlights the importance of histone acetyltransferase and the universality of the pattern that histone acetylation and methylation might coordinately regulate the expression of target gene.

In this study, through the genomic analysis of our previous profiling data,[14] we identified a novel lncRNA (*GCAWKR*) that was markedly upregulated in GC tissues compared to adjacent normal tissues in two cohorts of patients with ISH and qRT-PCR analysis. High expression levels of *GCAWKR* was correlated with some clinicopathological factors, including larger tumor size, the presence of lymph node metastasis, advanced TNM stage, more frequent recurrence, and cancer-related mortality. These data suggest that *GCAWKR* could be a potential prognostic indicator of GC.

The gene ontology pathways enrichment found that *GCAWKR* had an impact on cell proliferation and cell migration. The bioinformatics analyses were functionally verified in *in vitro* and *in vivo* experimental models with loss-of-function and gain-of-function approaches. The molecular mechanisms of lncRNAs function can be diversified, and

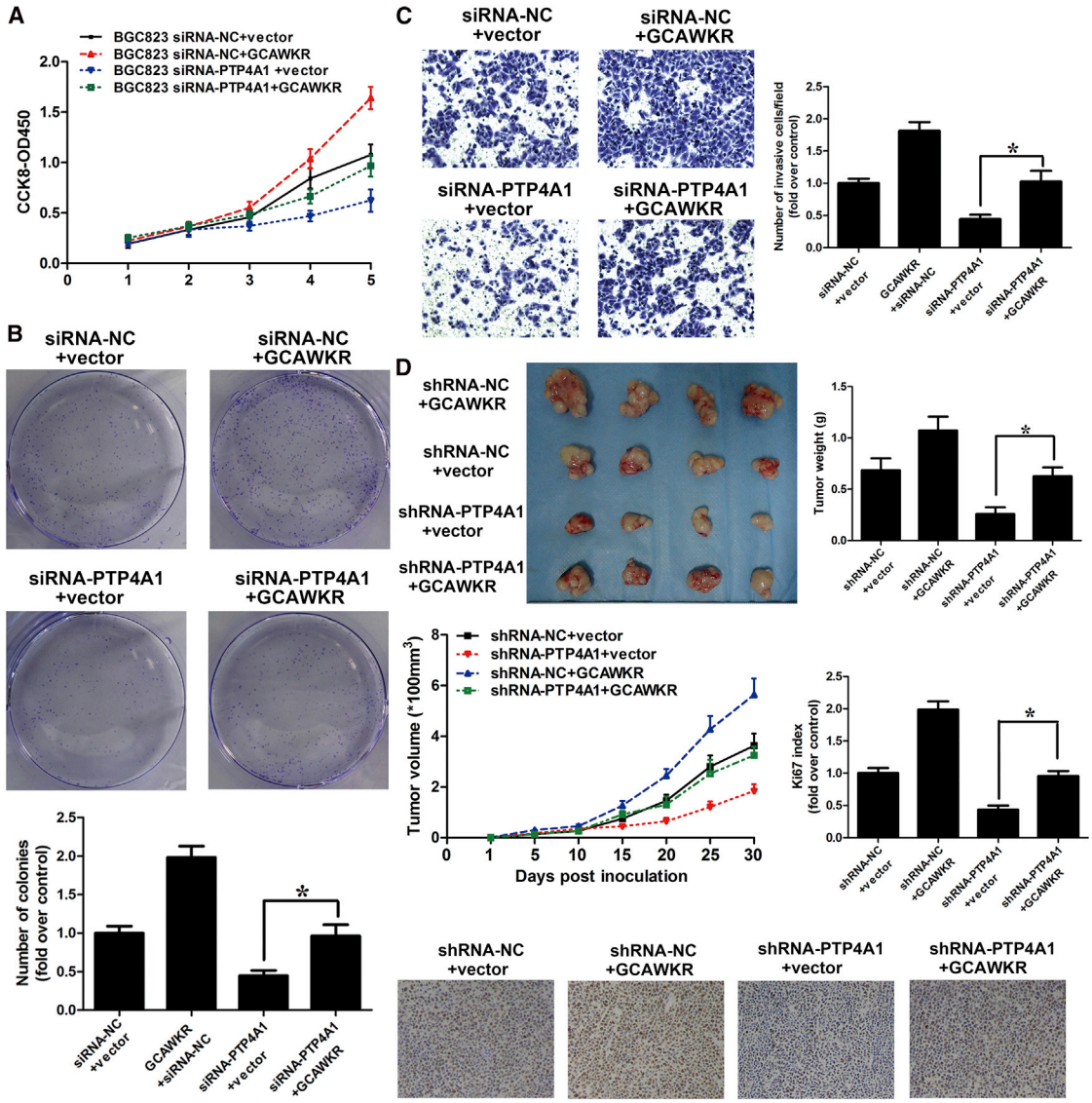


Figure 7. *GCAWKR* Regulates GC Development by Mediating PTP4A1 Expression

(A) CCK-8 assays showed that cell proliferation was inhibited in *GCAWKR*-overexpressing BGC823 cells after the cells were transfected with siRNA-PTP4A1. (B) Colony-formation assays showed that colony-formation ability was suppressed in *GCAWKR*-overexpressing BGC823 cells after the cells were transfected with siRNA-PTP4A1. (C) Effects of PTP4A1 knockdown on cell invasion in *GCAWKR*-overexpressing BGC823 cells in the presence of the anti-proliferative drug mitomycin C (MMC, 5 μM) treatment were determined using the transwell assay. *$p < 0.05$. (D) Top left, representative images of tumors formed in nude mice injected subcutaneously with BGC823 cells transfected with desired vector. Top right, tumor weights. Middle left, tumor growth curves. Bottom, representative images of IHC staining of Ki67 (original magnification, ×200). *$p < 0.05$.

lncRNAs located in the nucleus can guide and recruit chromatin-modifying complexes or transcriptional factors to the genomic loci, resulting in the activation or inactivation of target genes.[12] Here, we found that *GCAWKR* can function as a scaffold for WDR5 and KAT2A to regulate PTP4A1 expression transcriptionally. We verified the hypothesis with three lines of solid experimental evidence: (1) *GCAWKR* physically interacts with WDR5 and KAT2A; (2) *GCAWKR* knockdown or RNase treatment significantly suppressed the interaction between WDR5 and KAT2A; (3) *GCAWKR* knockdown inhibited the recruitment of WDR5/KAT2A complexes to the promoter regions of target genes. In general, *GCAWKR* was capable of modulating the interaction between WDR5 and KAT2A, mediating the histone modification of target genes.

When exploring the underlying mechanisms of *GCAWKR*'s oncogenic role in GC, we found that PTP4A1 might be involved in this process with RNA sequencing. PTP4A1 was reported to be upregulated in GC and promotes tumor development.[18] Consistent with

the previous reports, we found that *GCAWKR* modulates the expression of PTP4A1, and PTP4A1 is the downstream effector of *GCAWKR*-mediated oncogenic effects. Also, RNA sequencing in *GCAWKR*-silencing cells demonstrated that some other genes that are involved in regulation of cell proliferation and migration could also be affected by *GCAWKR*, such as CDK1, CDC25B, CCND3, KLF6, and HOXB13. Thus, we do not rule out the possibility that other genes might take part in *GCAWKR*-mediated biological functions.

Our data suggest that *GCAWKR* is an oncogenic lncRNA that promotes the tumorigenesis of GC by functioning as a scaffold and recruiting histone-modifying complexes to target genes. The study suggests that *GCAWKR* might be the potential therapeutic target.

MATERIALS AND METHODS

Patient Samples

The study was under the censorship of the Clinical Research Committee of the Fudan University Shanghai Cancer Center (FUSCC). Written informed consent was obtained from all patients. Two cohorts of patients involving 165 GC tissues and paired normal tissues collected during 2008 and 2012 were obtained from the biobank of FUSCC. The clinicopathological features of the patients are in Tables S1 and S2. These data do not contain identity-related information. None of these patients had received preoperative chemo- or radio-therapy. All patients were staged according to the criteria of the WHO Classification of Tumors of the Digestive System, 2010 edition.

Bioinformatics Analysis

lncRNA profiling data are available via GEO: GSE50710. RNA sequencing data can be accessed via GSE112586. Detailed information can be found in the Supplemental Materials and Methods.

Cell Culture

The human gastric epithelial cell line GES-1 and five GC cell lines (SGC7901, BGC823, MKN45, HGC27, and AGS) were used in this study. These cell lines were purchased from the Health Science Research Resources Bank on September 2015. The cell lines were characterized by DNA fingerprinting, cell vitality detection, isozyme detection, and mycoplasma detection. The last cell characterization was performed on November 2017. Cells were cultured in DMEM (Gibco BRL) supplemented with 10% fetal bovine serum (FBS), 100 U/mL penicillin, 100 μg/mL streptomycin (Thermo Scientific), and 8 mg/L antibiotic tylosin tartrate against mycoplasma (Sigma-Aldrich, St. Louis, Missouri, USA) at 37°C in an atmosphere of 5% CO_2. The passage numbers for GES-1, SGC7901, BGC823, MKN45, HGC27, and AGS were 25, 30, 32, 37, 16, and 19, respectively.

RNA and Protein Extraction, Real-Time qPCR Analysis, and Western Blot Analysis

RNA and protein extraction, real-time qPCR analysis, and western blot analysis were performed.[12]

Subcellular Fractionation Analysis and 5′ and 3′ RACE Analysis

Subcellular fractionation analysis and 5′ and 3′ RACE analysis were performed as we described previously.[12]

Cell Viability, Colony Formation, and Transwell Assay

Cell viability was determined with the Cell Counting Kit 8 (CCK-8, Donjindo).[12] For colony formation assay, cells were seeded in 6-well plates at a density of 800 cells per well and incubated for 10 days. Colonies were fixed with ethanol for 20 min and stained with 0.1% crystal violet for 30 min. Transwell assays were utilized to assess cell-invasive ability with the Transwell system (Corning) precoated with Matrigel in the presence of anti-proliferative drug MMC.

RNA Pull-Down Assay, RNA IP, Coimmunoprecipitation, and ChIP

RNA pull-down assay and RNA IP were utilized to confirm the interaction between the lncRNA and the specific protein. Coimmunoprecipitation assay was used to verify the direct interaction between proteins. ChIP assay was employed to determine the binding of the chromatin-modifying complexes to the promoter regions of genes.

Plasmid Construction and Cell Transfection

Plasmid construction and cell transfections were performed as we described previously.[12]

ISH of *GCAWKR* and *In Vivo* Experiments

ISH was performed to evaluate the expression of *GCAWKR*. To further illustrate the role of *GCAWKR* in tumor growth, the *in vivo* tumor growth assay was performed.

4∼5-week-old male BALB/c nude mice were maintained under pathogen-free conditions in the Experimental Animal Centre of Xinhua Hospital. All animal experiments of laboratory animals were performed by the Guide for the Care and Use published by the US NIH (NIH publication number 85-23, revised 1996). Nude mice were injected subcutaneously into unilateral flank areas with GC cells (1×10^6) transfected with desired vector to establish tumors (n = 5 per group). Tumor growth was measured every 5 days for 30 days. Then, tumors were removed and weighed. The length (L) and width (W) were determined using calipers, and the tumor volumes were calculated using the following equation: $(L^*W^2)/2$.

Statistical Analysis

All statistical analyses were performed using SPSS software (version 17.0, Chicago, IL). The difference between two or multiple groups was compared with Student's t test or one-way ANOVA test. The difference in *GCAWKR* levels in paired human samples was analyzed with the Wilcoxon test. The nonparametric Mann-Whitney-Wilcoxon test was utilized to assess the relationship between *GCAWKR* expression levels and clinicopathological factors. Spearman rank correlation test was used to test the correlation between *GCAWKR* and PTP4A1 expression levels. The survival curve was plotted with the Kaplan-Meier method, and the significance was calculated with the

log-rank test. The effect of clinicopathological factors on survival was determined with univariate and multivariate Cox proportional hazards models. All data are presented as the mean ± SD from three independent repeats. A two-sided p value less than 0.05 was considered to be statistically significant.

SUPPLEMENTAL INFORMATION

Supplemental Information includes eight tables, thirteen figures, and Supplemental Materials and Methods and can be found with this article online at https://doi.org/10.1016/j.ymthe.2018.09.002.

AUTHOR CONTRIBUTIONS

M.M., Y.Z., M.W., Y.H., and Y.X. acquired data, drafted the manuscript, and critically revised the manuscript for important intellectual content; M.M., M.W., and Y.Z. performed *in vitro* and *in vivo* assays; M.M. and Y.X. analyzed and interpreted data and statistical analysis; Y.R.H. and K.L. provided technical or material support; M.M., Y.R.H., and K.L. approved the final version of the manuscript.

CONFLICTS OF INTEREST

The authors have no conflicts of interest.

ACKNOWLEDGMENTS

This work was supported by the National Nature Science Foundation of China (grants 81602046 to M.Z., 81702388 to Y.R.H., and 81772180 and 81472017 to K.L.), the Key University Science Research Project of Anhui Province (KJ2016A738 to Y.Z.), the Wenzhou Science and Technology Bureau (grant Y20160426 to Y.R.H.), and the Shanghai Training and Support Program for Young Physician (to M.Z.).

REFERENCES

1. Siegel, R.L., Miller, K.D., and Jemal, A. (2015). Cancer statistics, 2015. CA Cancer J. Clin. *65*, 5–29.
2. Chen, W., Zheng, R., Baade, P.D., Zhang, S., Zeng, H., Bray, F., Jemal, A., Yu, X.Q., and He, J. (2016). Cancer statistics in China, 2015. CA Cancer J. Clin. *66*, 115–132.
3. Dassen, A.E., Dikken, J.L., van de Velde, C.J., Wouters, M.W., Bosscha, K., and Lemmens, V.E. (2013). Changes in treatment patterns and their influence on long-term survival in patients with stages I-III gastric cancer in The Netherlands. Int. J. Cancer *133*, 1859–1866.
4. Elimova, E., Wadhwa, R., Shiozaki, H., Sudo, K., Estrella, J.S., Badgwell, B.D., Das, P., Matamoros, A., Jr., Song, S., and Ajani, J.A. (2015). Molecular biomarkers in gastric cancer. J. Natl. Compr. Canc. Netw. *13*, e19–e29.
5. Ponting, C.P., Oliver, P.L., and Reik, W. (2009). Evolution and functions of long noncoding RNAs. Cell *136*, 629–641.
6. Qu, L., Ding, J., Chen, C., Wu, Z.J., Liu, B., Gao, Y., Chen, W., Liu, F., Sun, W., Li, X.F., et al. (2016). Exosome-Transmitted lncARSR Promotes Sunitinib Resistance in Renal Cancer by Acting as a Competing Endogenous RNA. Cancer Cell *29*, 653–668.
7. Wang, P., Xue, Y., Han, Y., Lin, L., Wu, C., Xu, S., Jiang, Z., Xu, J., Liu, Q., and Cao, X. (2014). The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. Science *344*, 310–313.
8. Yang, F., Zhang, H., Mei, Y., and Wu, M. (2014). Reciprocal regulation of HIF-1α and lincRNA-p21 modulates the Warburg effect. Mol. Cell *53*, 88–100.
9. Zhou, L., Sun, K., Zhao, Y., Zhang, S., Wang, X., Li, Y., Lu, L., Chen, X., Chen, F., Bao, X., et al. (2015). Linc-YY1 promotes myogenic differentiation and muscle regeneration through an interaction with the transcription factor YY1. Nat. Commun. *6*, 10026.
10. Yuan, J.H., Yang, F., Wang, F., Ma, J.Z., Guo, Y.J., Tao, Q.F., Liu, F., Pan, W., Wang, T.T., Zhou, C.C., et al. (2014). A long noncoding RNA activated by TGF-β promotes the invasion-metastasis cascade in hepatocellular carcinoma. Cancer Cell *25*, 666–681.
11. Cao, L., Zhang, P., Li, J., and Wu, M. (2017). *LAST*, a c-Myc-inducible long noncoding RNA, cooperates with CNBP to promote *CCND1* mRNA stability in human cells. eLife *6*, e30433.
12. Ma, M.Z., Zhang, Y., Weng, M.Z., Wang, S.H., Hu, Y., Hou, Z.Y., Qin, Y.Y., Gong, W., Zhang, Y.J., Kong, X., et al. (2016). Long Noncoding RNA GCASPC, a Target of miR-17-3p, Negatively Regulates Pyruvate Carboxylase-Dependent Cell Proliferation in Gallbladder Cancer. Cancer Res. *76*, 5361–5371.
13. Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., Zhao, L., Li, X., Teng, X., Sun, X., et al. (2018). NONCODEV5: a comprehensive annotation database for long non-coding RNAs. Nucleic Acids Res. *46* (D1), D308–D314.
14. Sun, T.T., He, J., Liang, Q., Ren, L.L., Yan, T.T., Yu, T.C., Tang, J.Y., Bao, Y.J., Hu, Y., Lin, Y., et al. (2016). LncRNA GClnc1 Promotes Gastric Carcinogenesis and May Act as a Modular Scaffold of WDR5 and KAT2A Complexes to Specify the Histone Modification Pattern. Cancer Discov. *6*, 784–801.
15. Zhou, C., Liu, G., Wang, L., Lu, Y., Yuan, L., Zheng, L., Chen, F., Peng, F., and Li, X. (2013). MiR-339-5p regulates the growth, colony formation and metastasis of colorectal cancer cells by targeting PRL-1. PLoS ONE *8*, e63142.
16. Stephens, B., Han, H., Hostetter, G., Demeure, M.J., and Von Hoff, D.D. (2008). Small interfering RNA-mediated knockdown of PRL phosphatases results in altered Akt phosphorylation and reduced clonogenicity of pancreatic cancer cells. Mol. Cancer Ther. *7*, 202–210.
17. Zhang, J.X., Mai, S.J., Huang, X.X., Wang, F.W., Liao, Y.J., Lin, M.C., Kung, H.F., Zeng, Y.X., and Xie, D. (2014). MiR-29c mediates epithelial-to-mesenchymal transition in human colorectal carcinoma metastasis via PTP4A and GNA13 regulation of β-catenin signaling. Ann. Oncol. *25*, 2196–2204.
18. Dumaual, C.M., Sandusky, G.E., Soo, H.W., Werner, S.R., Crowell, P.L., and Randall, S.K. (2012). Tissue-specific alterations of PRL-1 and PRL-2 expression in cancer. Am. J. Transl. Res. *4*, 83–101.
19. Marchese, F.P., and Huarte, M. (2014). Long non-coding RNAs and chromatin modifiers: their place in the epigenetic code. Epigenetics *9*, 21–26.
20. Grebien, F., Vedadi, M., Getlik, M., Giambruno, R., Grover, A., Avellino, R., Skucha, A., Vittori, S., Kuznetsova, E., Smil, D., et al. (2015). Pharmacological targeting of the Wdr5-MLL interaction in C/EBPα N-terminal leukemia. Nat. Chem. Biol. *11*, 571–578.
21. Trievel, R.C., Rojas, J.R., Sterner, D.E., Venkataramani, R.N., Wang, L., Zhou, J., Allis, C.D., Berger, S.L., and Marmorstein, R. (1999). Crystal structure and mechanism of histone acetylation of the yeast GCN5 transcriptional coactivator. Proc. Natl. Acad. Sci. USA *96*, 8931–8936.
22. Li, Y., Wang, Z., Shi, H., Li, H., Li, L., Fang, R., Cai, X., Liu, B., Zhang, X., and Ye, L. (2016). HBXIP and LSD1 Scaffolded by lncRNA Hotair Mediate Transcriptional Activation by c-Myc. Cancer Res. *76*, 293–304.
23. Sun, M., Nie, F., Wang, Y., Zhang, Z., Hou, J., He, D., Xie, M., Xu, L., De, W., Wang, Z., and Wang, J. (2016). LncRNA HOXA11-AS Promotes Proliferation and Invasion of Gastric Cancer by Scaffolding the Chromatin Modification Factors PRC2, LSD1, and DNMT1. Cancer Res. *76*, 6299–6310.
24. Thomas, L.R., Foshage, A.M., Weissmiller, A.M., and Tansey, W.P. (2015). The MYC-WDR5 Nexus and Cancer. Cancer Res. *75*, 4012–4015.
25. Kim, J.Y., Banerjee, T., Vinckevicius, A., Luo, Q., Parker, J.B., Baker, M.R., Radhakrishnan, I., Wei, J.J., Barish, G.D., and Chakravarti, D. (2014). A role for WDR5 in integrating threonine 11 phosphorylation to lysine 4 methylation on histone H3 during androgen signaling and in prostate cancer. Mol. Cell *54*, 613–625.
26. Chen, L., Wei, T., Si, X., Wang, Q., Li, Y., Leng, Y., Deng, A., Chen, J., Wang, G., Zhu, S., and Kang, J. (2013). Lysine acetyltransferase GCN5 potentiates the growth of non-small cell lung cancer via promotion of E2F1, cyclin D1, and cyclin E1 expression. J. Biol. Chem. *288*, 14510–14521.
27. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. Nucleic Acids Res. *36* (*Suppl 2*), W70–W74.

28. Zhou, Z., Ji, Z., Wang, Y., Li, J., Cao, H., Zhu, H.H., and Gao, W.Q. (2014). TRIM59 is up-regulated in gastric tumors, promoting ubiquitination and degradation of p53. Gastroenterology *147*, 1043–1054.

29. Chen, L., Min, L., Wang, X., Zhao, J., Chen, H., Qin, J., Chen, W., Shen, Z., Tang, Z., Gan, Q., et al. (2015). Loss of RACK1 Promotes Metastasis of Gastric Cancer by Inducing a miR-302c/IL8 Signaling Loop. Cancer Res. *75*, 3832–3841.

30. Wang, K., Liang, Q., Li, X., Tsoi, H., Zhang, J., Wang, H., Go, M.Y., Chiu, P.W., Ng, E.K., Sung, J.J., and Yu, J. (2016). MDGA2 is a novel tumour suppressor cooperating with DMAP1 in gastric cancer and is associated with disease outcome. Gut *65*, 1619–1631.

31. Shen, J., Xiao, Z., Wu, W.K., Wang, M.H., To, K.F., Chen, Y., Yang, W., Li, M.S., Shin, V.Y., Tong, J.H., et al. (2015). Epigenetic silencing of miR-490-3p reactivates the chromatin remodeler SMARCD1 to promote Helicobacter pylori-induced gastric carcinogenesis. Cancer Res. *75*, 754–765.

32. Sousa, J.F., Nam, K.T., Petersen, C.P., Lee, H.J., Yang, H.K., Kim, W.H., and Goldenring, J.R. (2016). miR-30-HNF4γ and miR-194-NR2F2 regulatory networks contribute to the upregulation of metaplasia markers in the stomach. Gut *65*, 914–924.

33. Ueda, T., Volinia, S., Okumura, H., Shimizu, M., Taccioli, C., Rossi, S., Alder, H., Liu, C.G., Oue, N., Yasui, W., et al. (2010). Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis. Lancet Oncol. *11*, 136–146.

34. Li, X., Wu, W.K., Xing, R., Wong, S.H., Liu, Y., Fang, X., Zhang, Y., Wang, M., Wang, J., Li, L., et al. (2016). Distinct Subtypes of Gastric Cancer Defined by Molecular Characterization Include Novel Mutational Signatures with Prognostic Capability. Cancer Res. *76*, 1724–1732.

35. Busuttil, R.A., George, J., Tothill, R.W., Ioculano, K., Kowalczyk, A., Mitchell, C., Lade, S., Tan, P., Haviv, I., and Boussioutas, A. (2014). A signature predicting poor prognosis in gastric and ovarian cancer represents a coordinated macrophage and stromal response. Clin. Cancer Res. *20*, 2761–2772.

36. Xu, T.P., Wang, W.Y., Ma, P., Shuai, Y., Zhao, K., Wang, Y.F., Li, W., Xia, R., Chen, W.M., Zhang, E.B., and Shu, Y.Q. (2018). Upregulation of the long noncoding RNA FOXD2-AS1 promotes carcinogenesis by epigenetically silencing EphB3 through EZH2 and LSD1, and predicts poor prognosis in gastric cancer. Oncogene *37*, 5020–5036.

YMTHE, Volume *26*

Supplemental Information

lncRNA GCAWKR Promotes Gastric Cancer Development by Scaffolding the Chromatin Modification Factors WDR5 and KAT2A

Mingzhe Ma, Yan Zhang, Mingzhe Weng, Ye Hu, Yi Xuan, YiRen Hu, and Kun Lv

Supplemental Table 1. Clinicopathological characterisitcs of 42 GC patients (Cohort 1)

Characterisitcs	Number of patients (%)
Patients	42
Gender	
Male	28 (66.67%)
Female	14 (33.33%)
Age (years)	36 to 76, mean 58.26
Tumor size (cm)	0.5 to 9, mean 2.88
Lymph node metastasis (N stage)	
N0	20 (47.62%)
N1	4 (9.52%)
N2	13 (30.95%)
N3	5 (11.90%)
Depth of invasion (T stage)	
T1	12 (28.57%)
T2	9 (21.43%)
T3	0 (0.00%)
T4	21 (50.00%)
TNM stage	
I	16 (38.10%)
II	8 (19.05%)
III	18 (42.86%)
Histology	
well and morderately	20 (47.62%)
Poorly and others	22 (52.38%)
Perineural Invasion	
Negative	29 (69.05%)
Positive	13 (30.95%)
Lymphovascular invasion	
Negative	29 (69.05%)
Positive	13 (30.95%)
Gross type	
EGC	6 (14.29%)
Borrmann type I	1 (2.38%)
Borrmann type II	7 （16.67%）
Borrmann type III	26 （61.90%）
Borrmann type IV	2 （4.76%）

Abbreviations: EGC, early gastric cancer;TNM, tumor-nodes-metastasis,based on the American Joint Committee On Cancer/International Union Against Cancer Staging Manual(8th editon, 2016).

Supplemental Table 2. Clinicopathological characterisitcs of 123 GC patients (Cohort 2)

Characterisitcs	Number of patients (%)
Patients	123
Gender	
Male	81 (65.85%)
Female	42 (34.14%)
Age (years)	31 to 82, mean 56.84
Tumor size (cm)	0.4 to 12, mean 3.93
Lymph node metastasis (N stage)	
N0	47 (38.21%)
N1	20 (16.26%)
N2	28 (22.76%)
N3	28 (22.76%)
Depth of invasion (T stage)	
T1	28 (22.76%)
T2	23 (18.70%)
T3	1 (0.81%)
T4	71 (57.72%)
TNM stage	
I	35 (28.46%)
II	25 (20.33%)
III	63 (51.22%)
Histology	
well and morderately	60 (48.78%)
Poorly and others	63 (51.22%)
Perineural Invasion	
Negative	76 (61.79%)
Positive	47 (38.21%)
Lymphovascular invasion	
Negative	78 (63.41%)
Positive	45 (36.59%)
Gross type	
EGC	17 (13.82%)
Borrmann type I	6 (4.87%)
Borrmann type II	22 （17.89%）
Borrmann type III	73 （59.35%）
Borrmann type IV	6（4.88%）

Abbreviations: EGC, early gastric cancer;TNM, tumor-nodes-metastasis,based on the American Joint Committee On Cancer/International Union Against Cancer Staging Manual(8th editon, 2016).

Supplemental Table 3. Correlation between GCAWKR expression and GC clinicopathological characteristics in 42 patients (Cohort 1)

Characterisitcs	GCAWKR expression levels		p value
	low expression	high expression	
Gender			
Male	15	13	1.000
Female	8	6	
Age (years)			
⩽59	11	10	1.000
>59	12	9	
Tumor size (cm)			
⩽3	19	9	0.023*
>3	4	10	
Lymph node metastasis (N stage)			
Negative (N0)	16	4	0.002*
Positive (N1-3)	7	15	
Depth of invasion (T stage)			
T1-T2	15	6	0.061
T3-T4	8	13	
TNM stage			
I-II	17	7	0.028*
III	6	12	
Histology			
well and morderately	13	7	0.232
Poorly and others	10	12	
Perineural Invasion			
Negative	17	12	0.516
Positive	6	7	
Lymphovascular invasion			
Negative	16	13	1.000
Positive	7	6	
Gross type			
EGC and Borrmann type I-II	11	3	0.048*
Borrmann type III-IV	12	16	

Note: Differences among variable were assessed by chi-square test. *, the values had statistical significant differences.Abbreviations: EGC, early gastric cancer; TNM, tumor-nodes-metastasis,based on the American Joint Committee On Cancer/International Union Against Cancer Staging Manual (8th editon, 2016).

Supplemental Table 4. Correlation between GCAWKR expression and GC clinicopathological characteristics in 123 patients (Cohort 2)

Characterisitcs	GCAWKR expression levels		p value
	low expression	high expression	
Gender			
Male	41	40	0.850
Female	20	22	
Age (years)			
≤59	36	35	0.856
>59	25	27	
Tumor size (cm)			
≤3	36	22	0.012*
>3	25	40	
Lymph node metastasis (N stage)			
Negative (N0)	30	17	0.016*
Positive (N1-3)	31	45	
Depth of invasion (T stage)			
T1-T2	31	20	0.045*
T3-T4	30	42	
TNM stage			
I-II	36	24	0.031*
III	25	38	
Histology			
well and morderately	35	25	0.072
Poorly and others	26	37	
Perineural Invasion			
Negative	37	39	0.854
Positive	24	23	
Lymphovascular invasion			
Negative	43	35	0.135
Positive	18	27	
Gross type			
EGC and Borrmann type I-II	27	18	0.094
Borrmann type III-IV	34	44	

Note:Differences among variable were assessed by chi-square test. *, the values had statistical significant differences.Abbreviations: EGC, early gastric cancer; TNM, tumor-nodes-metastasis,based on the American Joint Committee On Cancer/International Union Against Cancer Staging Manual (8th editon, 2016).

Supplemental Table 5. Univariate and multivariate analyses of OS in 123 GC patients by Cox regression analysis

variable	Univariate analysis			Multivariate analysis		
	Hazard ratio	CI (95%)	p value	Hazard ratio	CI (95%)	p value
Gender (male/Female)	1.231	0.767-1.759	0.217			
Age (years,≤59/>59)	1.152	0.893-1.653	0.158			
Tumor size (cm,>3/≤3)	1.729	1.189-2.416	0.035*			
Lymph node metastasis (N stage) (N1-3/N0)	1.946	1.374-2.788	0.016*			
Depth of invasion (T3,4/T1,2)	2.352	1.482-5.148	0.042*	1.752	1.259-2.385	0.014*
TNM stage (III/I-II)	3.254	2.078-4.364	<0.001*	3.396	1.956-5.785	<0.001*
Pathological Differentiation (Poor/Well&Moderate)	1.553	1.161-2.391	0.045*			
Perineural Invasion (Yes/No)	1.389	0.989-1.873	0.095			
Lymphovascular invasion (Yes/No)	1.753	1.186-2.516	0.009*			
Gross type (Borrmann type III-IV/EGC and I-II)	1.058	0.994-1.585	0.143			
GCAWKR (High/Low)	2.583	1.540-4.333	<0.001*	2.015	1.378-2.854	0.003*

Abbreviations: OS, overall survival; CI, confidence interval; HR,hazard ratio;TNM, tumor-nodes-metastasis,based on the American Joint Committee On Cancer/International Union Against Cancer Staging Manual(7th editon, 2009). * p values <0.05 were considered statistically significant.

Supplemental Table 6. Univariate and multivariate analyses of DFS in 123 GC patients by Cox regression analysis

variable	Univariate analysis			Multivariate analysis		
	Hazard ratio	CI (95%)	p value	Hazard ratio	CI (95%)	p value
Gender (male/Female)	1.323	0.961-2.028	0.106			
Age (years,≤59/>59)	1.215	0.832-1.761	0.264			
Tumor size (cm,>3/≤3)	1.456	1.153-2.189	0.021*			
Lymph node metastasis (N stage) (N1-3/N0)	4.612	1.983-7.162	<0.001*			
Depth of invasion (T3,4/T1,2)	1.039	1.459-8.264	0.004*			
TNM stage (III/I-II)	3.657	2.252-6.193	<0.001*	3.518	2.182-5.916	<0.001*
Pathological Differentiation (Poor/Well&Moderate)	1.613	1.030-2.532	0.041*			
Perineural Invasion (Yes/No)	1.571	1.247-2.385	0.024*			
Lymphovascular invasion (Yes/No)	1.694	1.076-2.427	0.019*			
Gross type (Borrmann type III-IV/EGC and I-II)	1.416	0.735-1.669	0.159			
GCAWKR (High/Low)	1.703	1.152-2.439	0.017*	1.728	1.221-2.537	0.003*

Abbreviations: OS, overall survival; CI, confidence interval; HR,hazard ratio;TNM, tumor-nodes-metastasis,based on the American Joint Committee On Cancer/International Union Against Cancer Staging Manual(7th editon, 2009). * p values <0.05 were considered statistically significant.

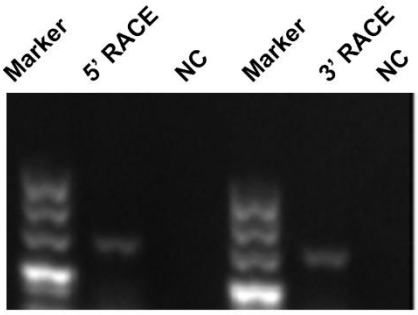
Supplemental Table 7. Primer sequence used in this study.

PCR Primers	
Gene	sequence
U6-F	5'- CTCGCTTCGGCAGCACA -3'
U6-R	5'- AACGCTTCACGAATTTGCGT -3'
β-Actin-F	5'-GGGAAATCGTGCGTGACATTAAG-3'
β-Actin -R	5'-TGTGTTGGCGTACAGGTCTTTG-3'
GAPDH-F	5'-CTTAGCACCCCTGGCCAAG-3'
GAPDH-R	5'-GATGTTCTGGAGAGCCCCG-3'
GCAWKR-F	5'- GACACCAGCAAGGGTGAGAT-3'
GCAWKR-R	5'- TTGAGCCTGAGCCAAGGTAT-3'
PTP4A1-F	5'- ACCAATGCGACCTTAAACAAA -3'
PTP4A1-R	5'- AATCTGGTTGGATGGTGGTG -3'
WDR5-F	5'- AATATCCGATGTAGCCTGGTC-3'
WDR5-R	5'-TTGGACTGGGGATTGAAGTTG -3'
KAT2A-F	5'- ATGACCAGCTCGCAGACCTAC -3'
KAT2A-F	5'- TTGACCACCGAACCCATGGAG -3'
CDK1-F	5'-CCAAGCAAGGGTTTGACATC-3'
CDK1-R	5'-GTGTGCCGGTGTCTACTTCA-3'
CDC25B-F	5'-TAAGGCGAAGATCAACATGG-3'
CDC25B-R	5'-TTACCAATGTCCCCAAGAGC-3'
CCND3-F	5'-CAAATGTGTGCAGAAGGAGGT-3'
CCND3-R	5'-GAAGCGGTCCAGGTAGTTCA-3'
GDF13-F	5'-CAAGGTGAGGAAGCGGAGGAA-3'
GDF13-R	5'- ACCCCGTTCTCACTGTGTCCA-3'
MMP9-F	5'-ACGCAGACATCGTCATCCAGT-3'
MMP9-R	5'- GGACCACAACTCGTCATCGTC-3'
MAPK4-F	5'- AGCTGCTATCCACATCAGACA-3'
MAPK4-R	5'- CGGCAGCATTAATCACAGGAG-3'
RIP-F	5'-ACGATATTCTTTCCCCCACC-3'
RIP-R	5'-AACTCAACTTCCACCTGCG-3'
RIP-GAPDH-F	5'-GCCCATCTACGAGGGGTAT-3'
RIP-GAPDH-R	5'-AATGTACAGCACGATTTCC-3'
Vector construction	
sh-GCAWKR-1-F	5'-CCGGGCGATGTAATGCCTTCTAAGACTCGAGTCTTAGAAGGCATTACATCGCTTTTTG-3'
sh-GCAWKR-1-R	5'- AATTCAAAAAGCGATGTAATGCCTTCTAAGACTCGAGTCTTAGAAGGCATTACATCGC -3'
sh-GCAWKR-2-F	5'- CCGGGCTCATGAGAGAAGGAAATATCTCGAGATATTTCCTTCTCTCATGAGCTTTTTG-3'
sh-GCAWKR-2-R	5'-AATTCAAAAAGCTCATGAGAGAAGGAAATATCTCGAGATATTTCCTTCTCTCATGAGC-3'
sh-GCAWKR-3-F	5'-CCGGGCATTCACTGAGTCACCAAAGCTCGAGCTTTGGTGACTCAGTGAATGCTTTTTG-3'
sh-GCAWKR-3-R	5'-AATTCAAAAAGCATTCACTGAGTCACCAAAGCTCGAGCTTTGGTGACTCAGTGAATGC-3'
sh-GCAWKR-4-F	5'-CCGGGGTGCAGGAGTTAGGACAAAGCTCGAGCTTTGTCCTAACTCCTGCACCTTTTTG-3'
sh-GCAWKR-4-R	5'- AATTCAAAAAGGTGCAGGAGTTAGGACAAAGCTCGAGCTTTGTCCTAACTCCTGCACC-3'
sh-NC-F	5'-CCGGCAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTGTTTTTG-3'

sh-NC-R	5’- AATTCAAAAACAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTG-3’
PTP4A1-F	5'- CCAGCTCCTGTGGAAGTCAC -3'
PTP4A1-R	5'- CCATCATCAAAAGGCCAATC -3'
siRNA-PTP4A1-1	5’- GTTTAAGGTCGCATTGGTTGG -3’
siRNA-PTP4A1-2	5’- CGTCCTGGGCAGAGTGAA -3’
siRNA-WDR5-1	5’-GCUGGGAAUAUCCGAUGUATT-3’
siRNA-WDR5-2	5’-GCUCAGAGGAUAACCUUGUTT-3’
SiRNA-WDR5-3	5’-CCCAGUCCAACCUUAUUGUTT-3’
WDR5-F	5’- ATGGCGACGGAGGAGAAGAAGC -3’
WDR5-R	5’- TTAGCAGTCACTCTTCCACAGTTTA -3’
KAT2A-F	5’-ATGAGTGTGGATCCAGCTTGTCCC-3’
KAT2A-R	5’- TCACACGTCTTCAGGTTGCATGTT-3’
shRNA-PTP4A1	5’-CCGGGCCCCCAGCCTATCACCTAAGAGACAACTGGACCAGAG-3’
PTP4A1-promoter-F	5’-TCTGCCCATGTCGGGGCT-3’
PTP4A1-promoter-R	5’-TGCCTCCAAAAGGGCCTCC-3’

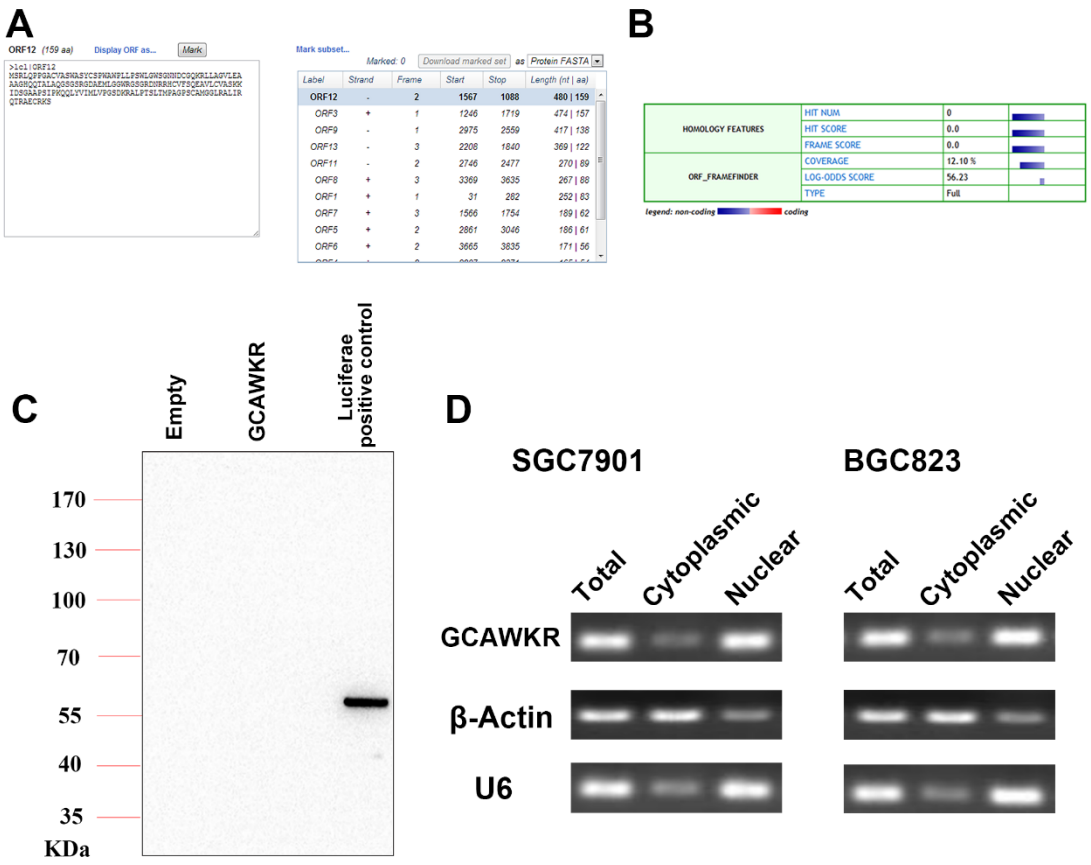
Supplemental Table 8. Information on antibodies used for correlation analysis.

Antibody	WB	IHC	CHIP	Specificity	Company (catalog number)
PTP4A1	1:1000	1:200	/	Rabbit monoclonal	Abcam (ab121185)
WDR5	1:1000	/	1:200	Rabbit polyclonal	Abcam (ab56916)
KAT2A	1:1000	/	1:200	Rabbit polyclonal	Abcam(18381)
Ki67	/	1:300	/	Rabbit polyclonal	Abcam (ab15580)
GAPDH	1:2000	/	/	Rabbit polyclonal	Abcam (8245)
H3K4me3	/	/	1:300	Rabbit polyclonal	Abcam (8580)
H3K9ac	/	/	1:300	Rabbit polyclonal	Abcam (10812)

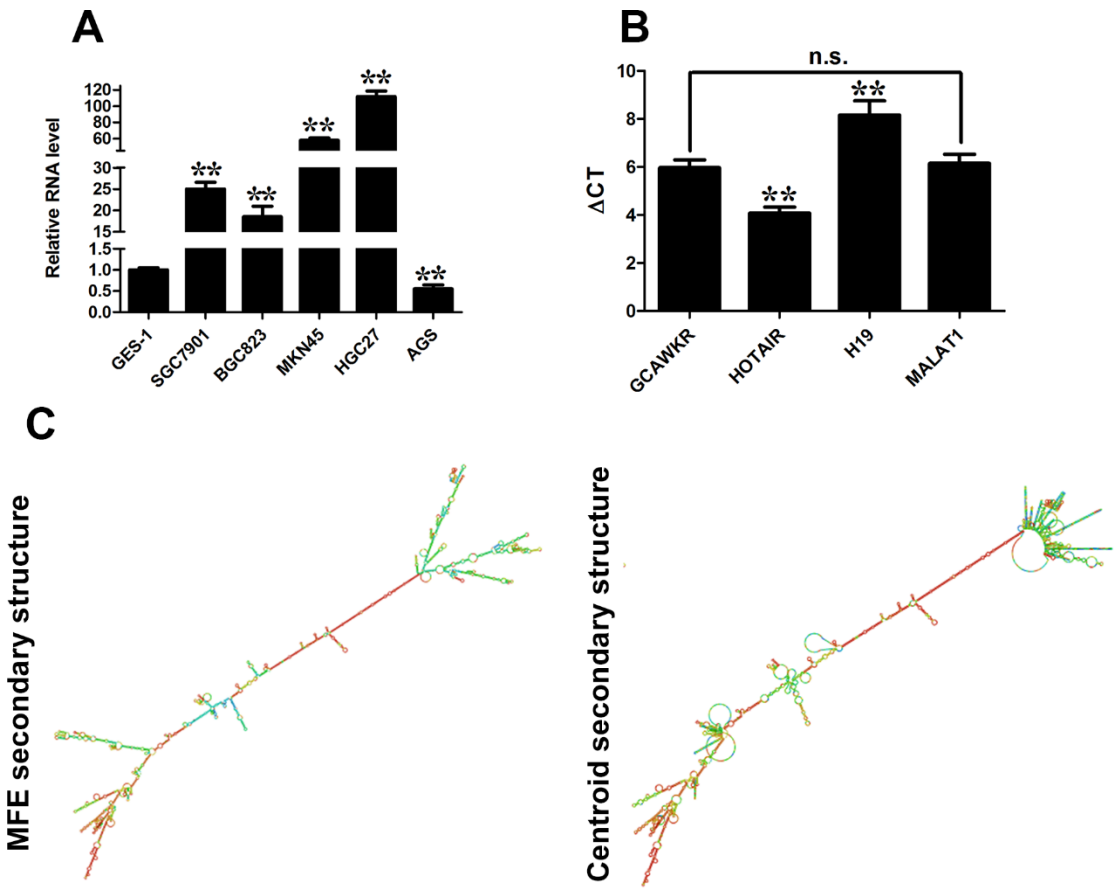


5'-GATTCTGTCGCGAACCCTCTCACGGCAGCTATGAGACTGCTGCGCTCCCCAGACGAGGAAACCTGGTCTCCCCGGAGCCAGGT GGA GTC T
CGCTC TGTCGCCCAGGCTGGAGTGCGGTGGCGCGATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCA GCCT
CCCGAGTAGCTGGGACTACAGAGTCTTGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCGCGATCTCGGCTCACTACAACCTCTGCCTCCCAGGT
TCAAGTGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGCTCTGGGACAAACCAGCTTCTATGTCATGAGCACACTCAGGCAGCTC
ATTGGAGAGATCTGCATGGTAAAGCAATGAGGCCTCCTGCCAGCAGCCATGTGAGTGGCATCTTGGAGCCAGACCCTCCAGCCCCAGTCAAG
CCTTCGGATGCCTGCAGCCCTGGCTGACATTTAACCACAACTTCATGAGAGATTCTGAGCCAGAATCACCCAGCTAAGCTGCTCCCGGGTTG
CCAGATCAATCCACAGTTCTCCAGGGACCCATTGTGGTTTGTGCTTCCTCTGTGCCTCCTGCTCCTCATGCACCCCAGCCCTCCCAG
TCCCCTCAACGGTCCTTCTGGCCAGGCCCATAGCCCCCTCTCTGCCCCACAGTCCTCAGGCTCCAGCTTTCTCGTTAGGAAGTGTTAGAAAGTT
GCTGGAGTGAGAAATGAGCCTGTGAAACTCGTTGTCCTAATTAAACTTCCACAAGGAGAGTGACAAGAGGCGACGGGGAGGAAACGGTGCC
ACCACCAAGGTCTCAGCCTGGGCCCAGGTCGAGGAAGGCCTTTGTCCTGCAGAACAAAAGCCACTGTTTGCACGTGTTAAACTTCATCAAAC
TTCAACAGACCCAGGTGCAAAACCCACATCCCAAATGCCATCCGGGCTTATGCCACCTTCCCGAGCATGGACCGTCCCTAGCCGTCAACCCT
AATCAGGGCCATGCCTACATCAAGCTGCACTCCCTCCACAGACAGCTTCTGCTCTGAGATGCATTGTTCCTATGACTTCTACACTCCGCCC
TTGTCTGACGAATTAGGGCACGGAGCCCCCCCATGGCACAGCTGGGCCCTGCTGCATTGTTAATGAGGTAGGAAGGGCGCGTTTGTCTGAG
CCTGGGACTAACATAATCACATAAAGTTGTTGCTTTGGAATAGATGGAGCTGCCCCAGAGTCGATTTTCTTGGAAGCAACGCAAAGCACAGC
TTCCTGGGAGAAGACACAATGCCTCCTATTGTCTCTGCCACTTCCCCTCCAGCCTCCCAGCATTTCTGCGTCTCCTCGGGAGCCACTGCCCTGT
GCAAGAGCTGTTTGCTGATGCCCAGCCGCGGCTTCCAAAACTCCCGCGAGAAGTCGTTTCTGTCCACAGTCATTGTTTCCAGACCAGCCGAGC
CAACTGGGGAGGAGAGCCAAGCCCAGGGAGAGCAGTAAGAGGCCCAGCTTGCAACACAGGCCCCCGGCGGCTGGAGTCGGCTCATGGGTT
GCTGGGGGGCAGCTGGTGCCAGCCCTGCTAGACATGTCCAGGGTCGGGGCAGAGGCCAGGGTGGCTCTCATCCACAACGAGGCTCCAGGAA
GAGGCTCCAGAGGGCCCAGGTCCCTGGAGGAAGACAGAGGGTGGATTATCCAGTGAGTCCTCAGATTGCTCCAGGGAGACTCTACCCCTAGG
CCCTCATCCTACCTAGGACTGAGGCAGCCATCTGGGTCTCTGCTTCCCTAACCACTTCCTCCTCCACCCGACCATCACCTAGTCTCAAGGACTT
GGGCAGCTTCTGGAAGTGGCTGAGACTGGAAGAATGTTGTCACACCTACCAGATATCCCTACGTGTATCTCTACCACCCCCACTACATATTCT
GAAGAACCAAGGTGAGCTGGGGGAATCCCCAGCCCAGCAGGGCAAGGCCTGGGCAGGAATGGAGTGAGGAAGTGAGAAGTGGGGATCTTG
GGGGTGGGCGGGCTGTGGGGCTGGGCTGTGCTGCACACTGGGCAGGAGGCAGGAGATGGCAATGGAAGCGACAGCTCCCATCTTCCCCGG
CTGCATACGTGAGGGATTCTGCTGCATTCACTGAGTCACCAAAGGACACTGGAAGGCCATCAGGATTTTGGGTGACCAAGACATGGCTCCC
CTCTTACTGCCTGTGGGGAGACAGGCATGTGAACACCCAAGTCTAACCCAAGCTGAGTTGACTGAGGGCCATTTAAAGATTCCACAGAGGCT
GAGAGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAGGCAGGTGGATCACTTGAGGTCAGGAGCTCGAGACCAGCCTGGCCA
ACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGTGCACACCAGTAGTCCCAGCTACTCAGGAGGCTGAGGC
AGGAGAATCGCTGGAACCCAGGAGGCAAAGGTTGCAGTGAGCCAGGATCACACCACAGTACTCCAGCCTAGGTGACAGAGCGAGACTCCGT
CTCGAAAATAAACAAATAAAGAAAGGTTCCAAGGATGGTCTGGAGGACAAACAAAGGAGGAGATTTGGTGCAGGAGTTAGGACAAAGACTC
CTGGGACAGCAGCCCTCACTTGGGGCCATAGTGAATGAATGGGGGCCTTTTGCAGAAGGGAAATGGAGGACATGTGTGGTGAAGGCTTTGCT
GAAGCATCTCTTATAAAAGGGGCAAAAGCACTTGCTCTCCTCCCAGCTCTGGCTCAGCCTCTCTCTTCAAGTCCCTGGCATCCCTGGAAAA
CCATGATTTTTCCATTCTGCAGCCAGCCCAGGCACCTCCTCCCACCCCCGCACACTGCAGTCTGGTGTGTGTGGTGGGCACACATAGAACACT
TGCTGGTCTTGAAAATGAAGCCATCTTTTGGGGCCCCTAGAAATCAAAGAGACCCACTTAGCCAAGATATCACTCTTGGACCACAGCAAGGG
TGAGATCACTTGATAGGTATTTGCCGTGTGTTTCTTTTAACTATATAAAAACAAAAGGTGGATGTGCAAACCCAACCAAGAATACCTT
GGCTCAGGCTCAACTAGATGAAACTTTTAAAAATGATTTAATGTTATGTTAGACTGTTGCATATTAAAGCGATGTAATGCCTTCTAAGAAACG
AGAGTCAGATCCTGAAAGCGATGACTTGGGTAAATGTAAGGAGTTATAACTCACTTTTCAAGAGAGCAACTATAAACATTTTGCCCAATATG
TACAGACCTCTAATGACAGTGTGCAAAGACTCTTTCGTTCCCAGATTATGAGGCTGCCACTTTCTCCTCCCCTTCCTGAGCCGCACTTTTCTAA
ACAGCAACTGTCGGGGCAGGGGGATGGGAAGCTCATGAGAGAAGGAAATATTTGGGACCGTGTCGGCCAGTGGAATGGGGCTTGTGGTCAC
CGTGTTACAGAGTCAAATCACTTCCCAGTAAACACAGTTCCGATGGCTCTCTGGCCAGGGTTGCCGGGGACAAGCAGGGAGTATATGTG
TCGGGTACGTTGCGTCACAAGCTGTGCAGCTGTGTAGGCCTCTCTGTCCAGAGGCATTAATGGAAAATGATTGGGCACCCAGCCCACAGCAG
ATGCGCTGAGGAATCCAGCCAGGGGGAGCTGGGATGCAGTGGGAGAGGGGACACGCAGGGCCTGGCTGAGAAGCCTCGCGAGGTAACCAC
AGGCGGAAATCTGAAAAATGAAGACGTTCCCACCAGTTCCAGAAAACAGGGCTAACTTAAAATGCAGCCCCTTTCCTTGAAGTGGGCTTGGC
AGGCCCCAGGTGTTGAAAGGC-3'

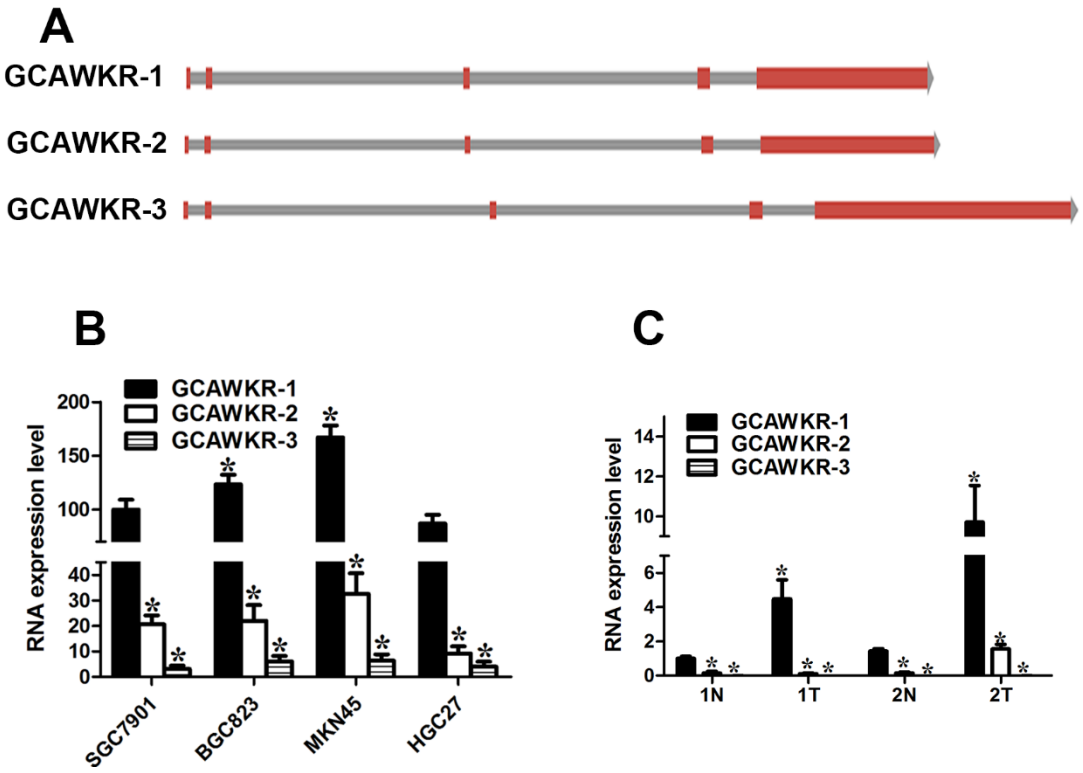
Supplementary Figure1. Left: PCR products from the 5'RACE and 3'RACE procedure was shown by agarose gel electrophoresis. Right: Nucleotide sequence of the full-length human *GCAWKR*.



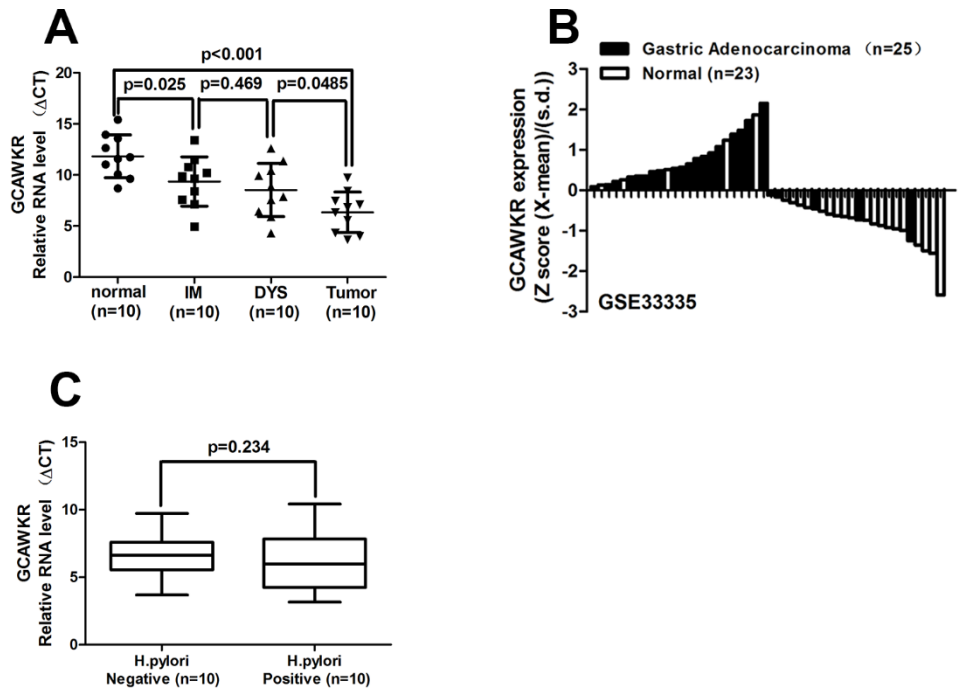
Supplementary Figure 2. *GCAWKR* has no coding capability. (A) Putative proteins possibly encoded by *GCAWKR* as predicted by the ORF Finder. (B) The transcript's noncoding nature was suggested by negative score with CPC. The transcript's noncoding nature was suggested by negative score with PhyloCSF (-15.3106, meaning that *GCAWKR* is $10^{15.3106}$ times more likely to be a noncoding sequence than a coding one). (C) In the translation assay, a 5406-bp genomic region containing *GCAWKR* was cloned into a pcDNA vector and expressed using the TnT Quick Coupled Transcription/Translation System (Promega). The absence of a specific band indicated that *GCAWKR* is a transcript with no protein-coding capacity. Luciferase in vitro translation served as positive control. (D) *GCAWKR* was mainly located in the nuclear as shown by RT-PCR.



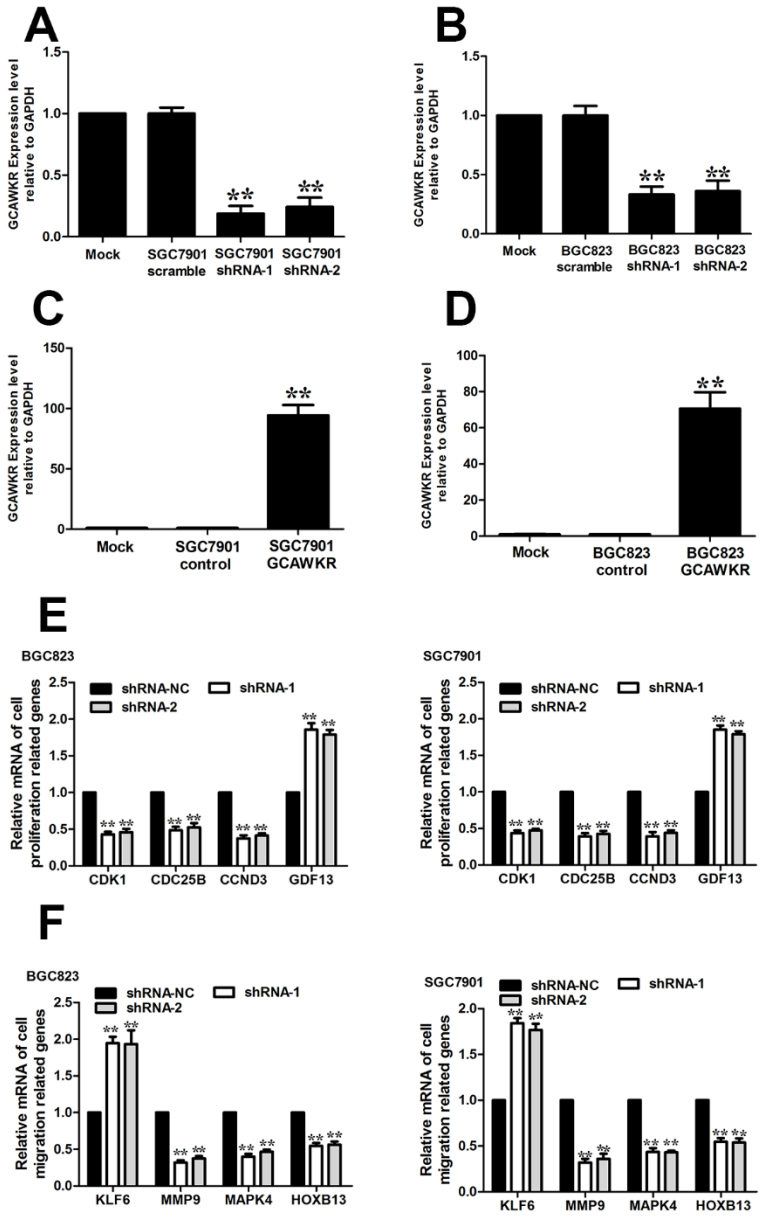
Supplemental Figure 3. (A) The relative expression of *GCAWKR* detected in normal gastric epithelium cell line GSE-1 and GC cell lines SGC7901, BGC823, MKN45, HGC27 and AGS using qRT-PCR. Error bars show standard deviation. ** $p < 0.01$, significantly different from GES-1 cells. (B) The expression of *GCAWKR* in BGC823 cells is comparable to *MALAT1*. ΔCt values were used to measure gene expression, which was normalized by β-Actin expression. Error bars show standard deviation. **$p < 0.01$, significantly different from expression of *GCAWKR*. (C,D) Prediction of *GCAWKR* structure based on minimum free energy (MFE) and partition function. Color scale indicates the confidence for the prediction for each base with shades of red indicating strong confidence.



Supplemental Figure 4. (A) Schematic representation of three transcripts of *GCAWKR*. (B) The relative expression of *GCAWKR* detected in GC cell lines SGC7901, BGC823, MKN45 and HGC27 using qRT-PCR. Error bars show standard deviation. *p < 0.05, significantly different from expression of *GCAWKR-1* in SGC7901 cells. (C) The relative expression of *GCAWKR* detected in GC tissues and paired normal tissues. Error bars show standard deviation. *p < 0.05, significantly different from expression of *GCAWKR-1* in 1N.



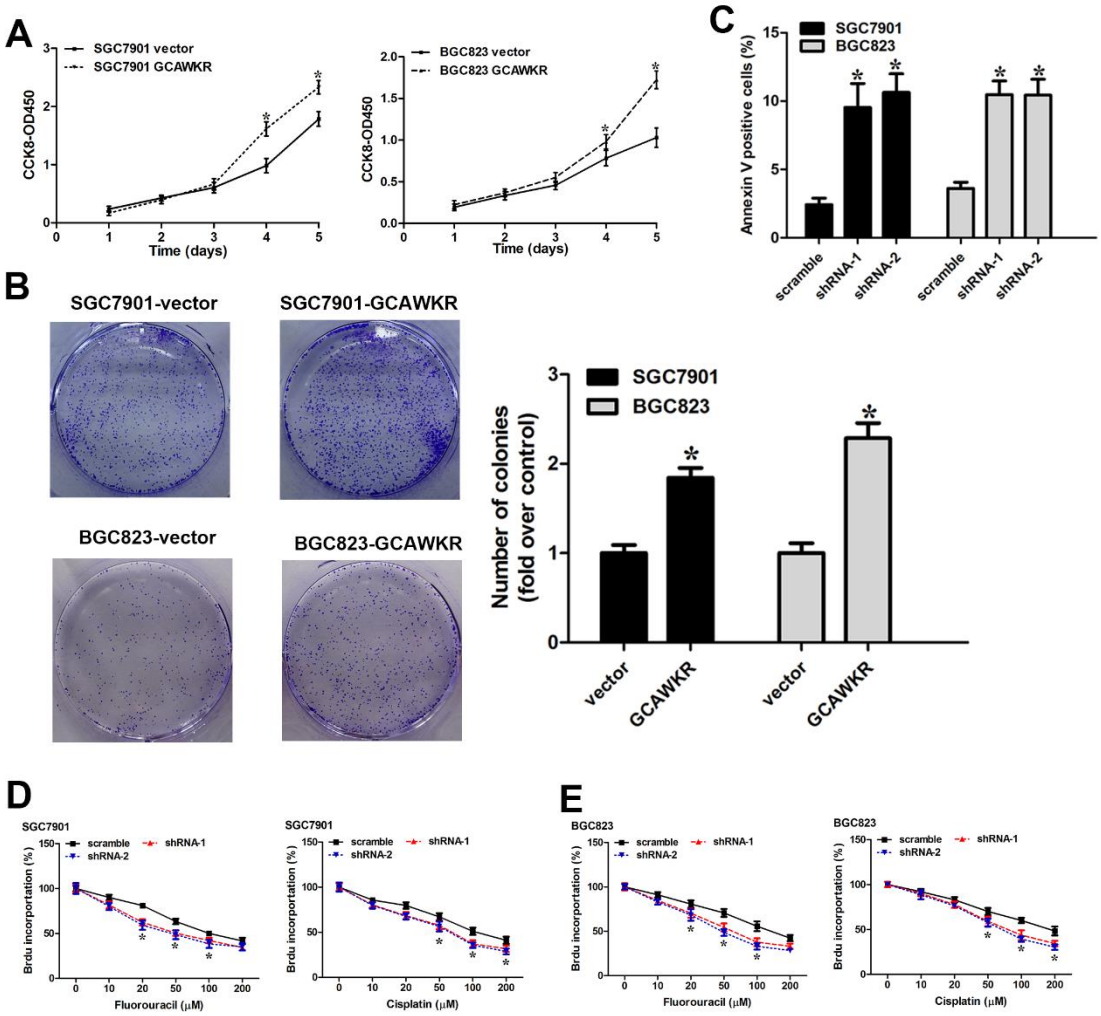
Supplemental Figure 5. (A) Statistical analysis of *GCAWKR* expression is shown in normal, intestinal metaplasia (IM), dysplasia, and cancerous gastric tissues. non-parametric Mann-Whitney test. (B) The expression of *GCAWKR* was validated in an independent gastric cancer patient cohort. (C) *GCAWKR* in H. pylori+ and H. pylori- tumours tissues. $p > 0.05$.



Supplemental Figure 6. Gene Ontology (GO) analysis in gastric cancer cells.

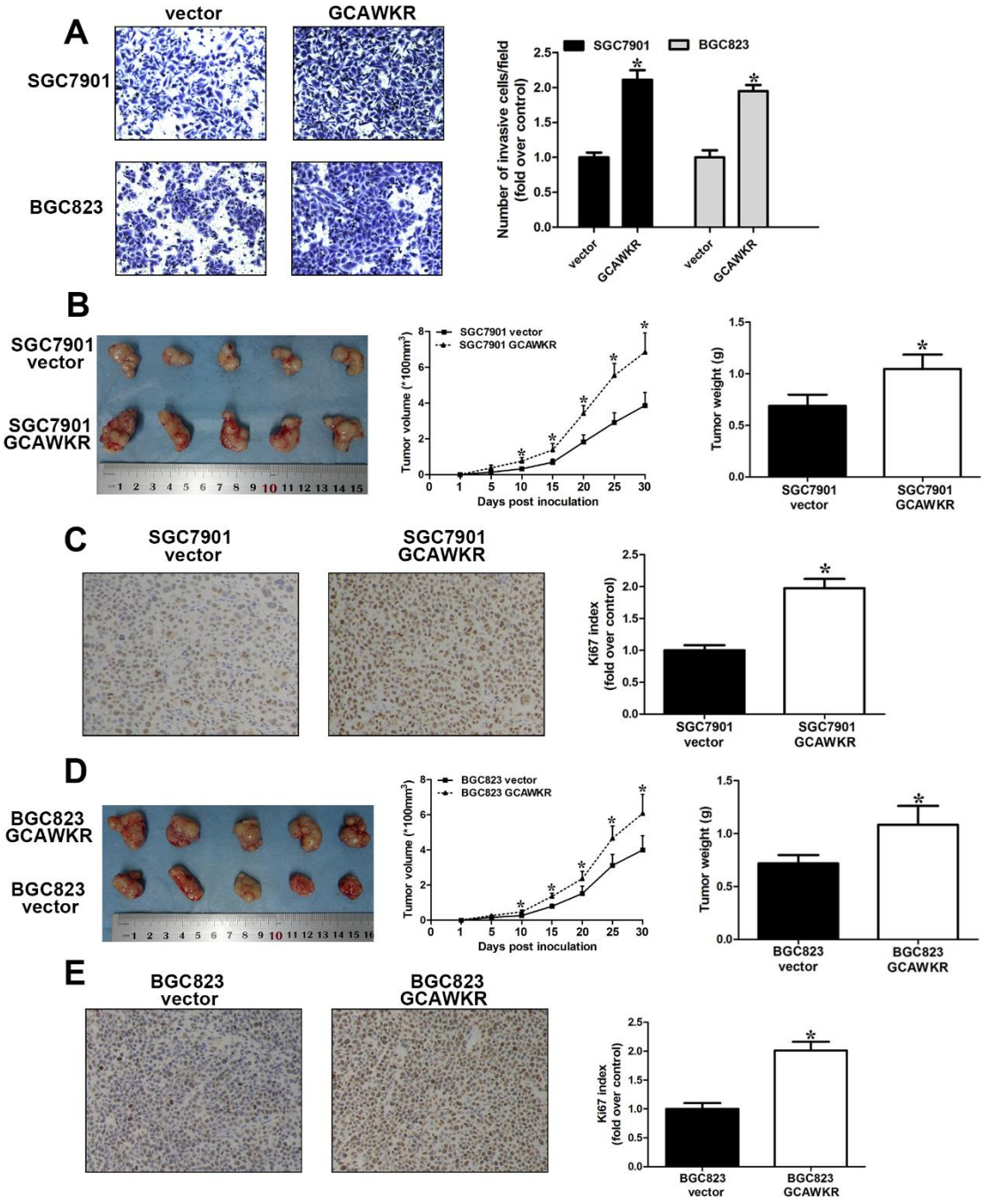
(A,B) *GCAWKR* expression was detected in SGC7901 (A) and BGC823 (B) cells by qRT-PCR after transduction of lentiviruses encoding *GCAWKR* shRNA-1 or shRNA-2 or a scrambled shRNA. **, $p < 0.01$. (C) *GCAWKR* expression was detected in SGC7901 (C) and BGC823 (D) cells by qRT-PCR after transfection of lentivirus harboring the full-length human *GCAWKR* sequence or the empty vector. **, $p < 0.01$.

The mRNA levels of the cell proliferation (E) and migration (F) -related genes were measured in BCG823 and SGC7901 gastric cells after transfection of *GCAWKR* shRNAs or control shRNA. n = 3. Error bars in the bar graphs represent SD. **, $p < 0.01$.



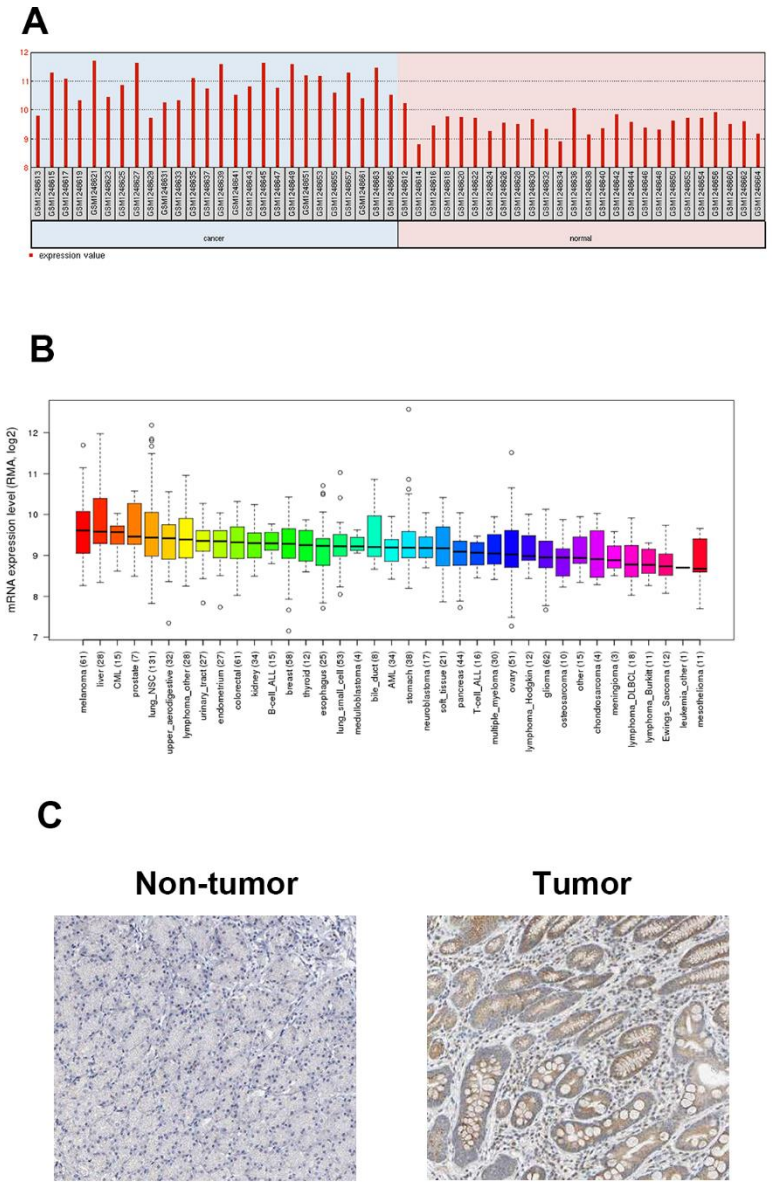
Supplemental Figure 7. (A) The cell growth rates were determined with CCK-8 proliferation assays. *GCAWKR* overexpression in SGC7901 and BGC823 cells significantly inhibited cell proliferation. (B) Colony formation assays were used to determine the cell colony formation ability of SGC7901 and BGC823 cells transfected of lentivirus harboring the full-length human *GCAWKR* sequence or the empty vector. (C) Annexin V/PI staining and flow cytometry analysis assessing apoptosis in SGC7901 and BGC823 cells after *GCAWKR* shRNAs transfection. Data represent the mean ± S.D. from three independent experiments. (D) Dose-response curve of a

representative experiment shows the relative fluorouracil or cisplatin sensitivity determined by BrdU incorporation. SGC7901 cells were treated with fluorouracil or cisplatin after transfection with scramble or *GCAWKR* shRNAs. n = 3. (E) Dose-response curve of a representative experiment shows the relative fluorouracil or cisplatin sensitivity determined by BrdU incorporation. BGC823 cells were treated with fluorouracil or cisplatin after transfection with scramble or *GCAWKR* shRNAs. n = 3.

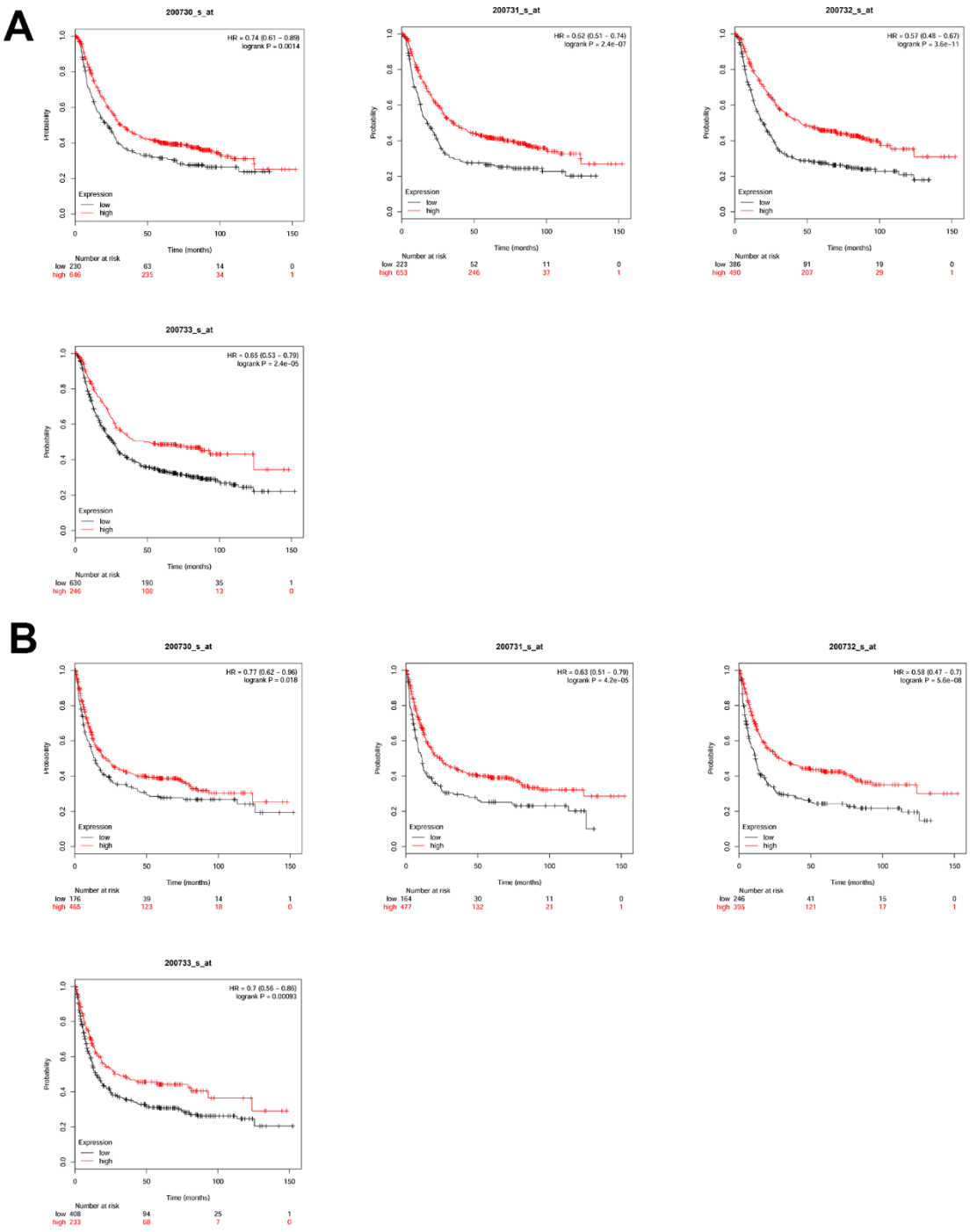


Supplemental Figure 8. (A) Effects of *GCAWKR* overexpression on cell invasion in the presence of the anti-proliferative drug mitomycin C (MMC, 5 μM) treatment were determined using the transwell assay. *, $p < 0.05$. (B-E) Top left, representative images

of tumours formed in nude mice injected subcutaneously with *GCAWKR*-overexpressing SGC7901 cells (B) or BGC823 cells (D). Top middle: tumour growth curves. Top right: tumour weights. (C,E) Lower: representative images of IHC staining of Ki67 (original magnification, ×200).

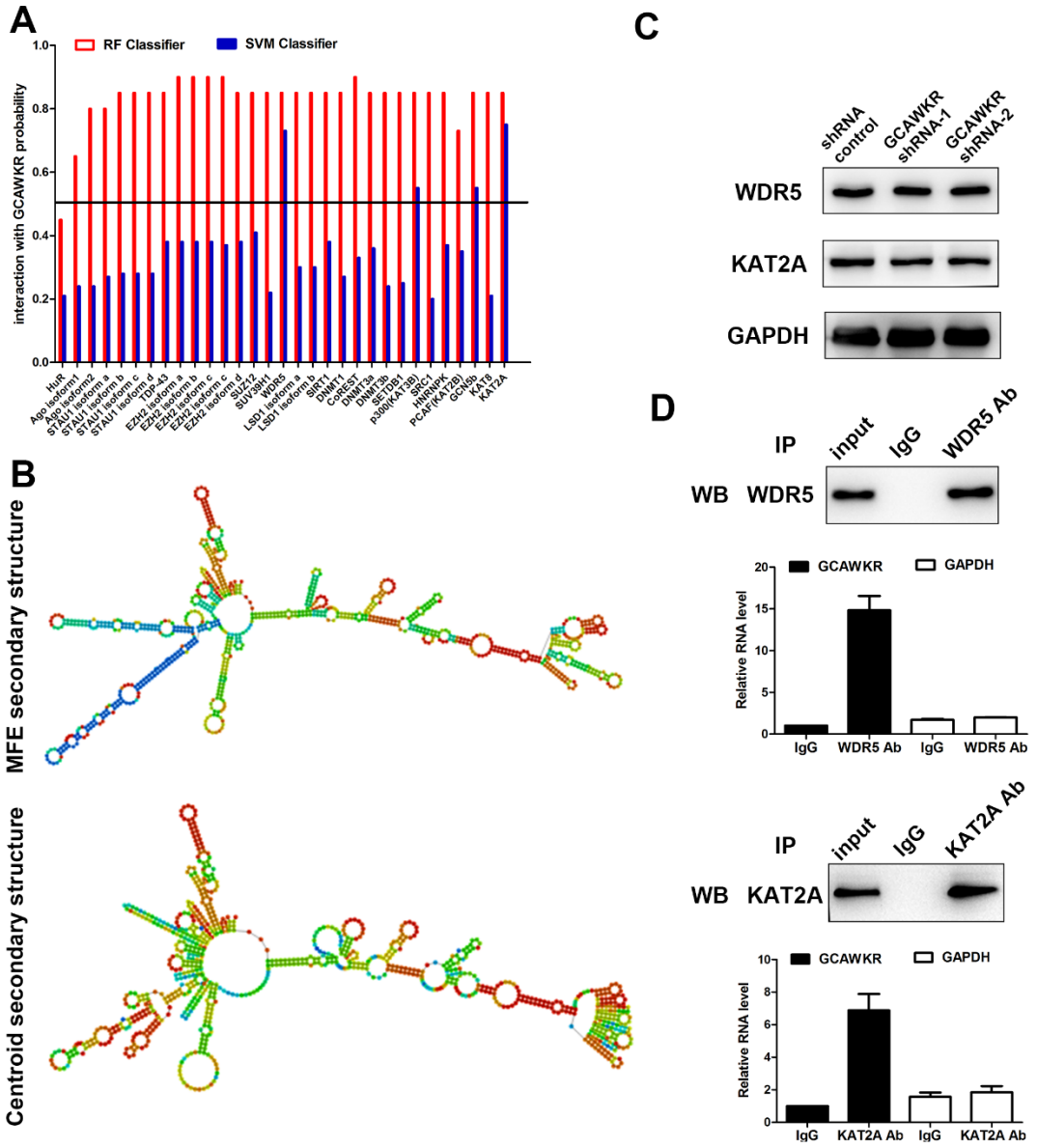


Supplemental Figure 9. Analysis of PTP4A1 expression levels in GC tissues. (A) Analysis of PTP4A1 expression levels in gastric cancer tissues and normal tissues from GEO dataset GSE51575. (B) PTP4A1 expression levels vary across different cancer types in the TCGA database. (C) Immunohistochemical staining analysis confirmed the PTP4A1 was upregulated in GC tissues compared to normal tissues.



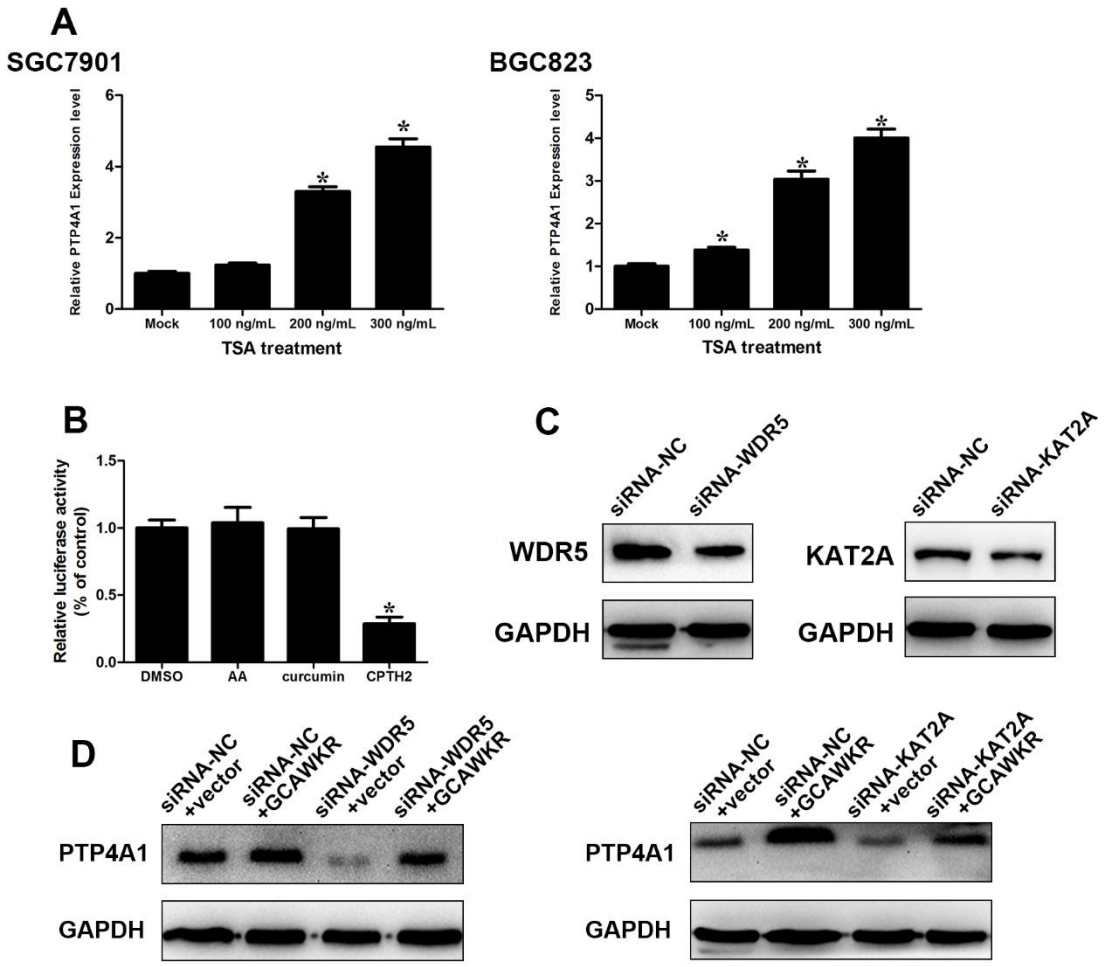
Supplemental Figure 10. (A) A higher expression level of PTP4A1 is correlated with a significantly poorer DSS (p<0.05) according to data from the KMPlot database. (B) A

higher expression level of PTP4A1 is correlated with a significantly poorer DFS (p<0.05) according to data from the KMPlot database. The HRs and p values were calculated with log-rank tests.



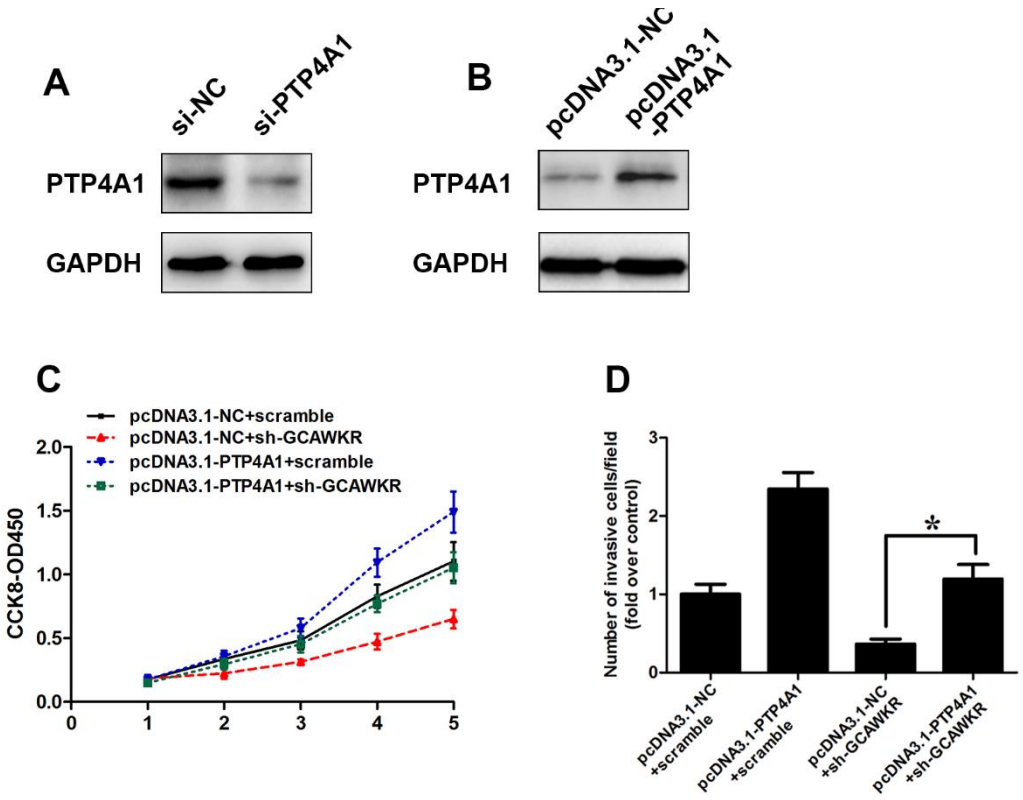
Supplemental Figure 11. (A) The interaction of *GCAWKR* and potential RNA binding proteins was predicted on http://pridb.gdcb.iastate.edu/RPISeq/. (B) 3'-end segment of *GCAWKR* (3000~3893nt) is predicted to have a stable stem-loop structure. Prediction of *GCAWKR* structure of a 3000~3893nt region based on minimum free energy (MFE) and partition function. Color scale indicates the confidence for the prediction for each

base with shades of red indicating strong confidence. (http://rna.tbi.univie.ac.at/). (C) Western blot was performed to detect WDR5 and KAT2A expression after transfection of *GCAWKR* shRNAs in BGC823 GC cells. n = 3. (D) Nuclear lysates of BGC823 cells were immunoprecipitated with anti-WDR5, anti-KAT2A antibody or control IgG. Aliquots of Nuclear lysates (20% of input) and the WDR5 and KAT2A immunoprecipitates were separated by SDS-PAGE and the specific immunoprecipitation of WDR5 or KAT2A was confirmed by the Western Blot assay. The complexes were analyzed for the presence of *GCAWKR* or GAPDH by qRT-PCR. Signals were normalized to β-Actin mRNA. Results were mean ± S.D. from three independent experiments.



Supplemental Figure 12. (A) PTP4A1 expression was measured by qRT-PCR in SGC7901 and BGC823 cells after the treatment of Trichostatin A for 24 h. (B) Cells were transfected with the PTP4A1-promoter 2kb-pGL3 and Renilla plasmids; after 18 h, the cells were treated with the indicated HAT inhibitors. After 6 h of treatment, a dual luciferase assay was performed. Statistical significance was determined using the Student's t-test (*, $p < 0.05$). DMSO, dimethyl sulfoxide; AA, anacardic acid; CPTH2, cyclopentylidene-[4-(4-chlorophenyl) thiazol-2-yl)] hydrazone. (C) Western blot was performed to detect WDR5 and KAT2A expression after transfection of WDR5 or

KAT2A siRNA in BGC823 GC cells. n = 3. (D) Western blot assay was performed to determine PTP4A1 protein level after transfection of siRNA-NC+vector, lentivirus harboring the full-length human *GCAWKR* sequence, WDR5/KAT2A siRNA+vector, WDR5/KAT2A siRNA+lentivirus harboring the full-length human *GCAWKR* sequence in BGC823 cells (n=3).



Supplemental Figure 13. (A) Western blot assay was performed to determine PTP4A1 protein level after transfection of siRNA-PTP4A1 or negative control in BGC823 cells. (B) Western blot assay was performed to determine PTP4A1 protein level after transfection of pcDNA3.1-PTP4A1 or negative control in BGC823 cells. (C) CCK-8 assays showed that cell proliferation was relieved in *GCAWKR*-silencing BGC823 cells after the cells were transfected with pcDNA3.1-PTP4A1. (D) Effects of PTP4A1 overexpression on cell invasion in *GCAWKR*-knockdown BGC823 cells in the presence of the anti-proliferative drug mitomycin C (MMC, 5 μM) treatment were determined using the transwell assay. *, $p < 0.05$.

Patients Samples

Each sample were stored in RNA later at −80°C prior to RNA isolation and qRT-PCR analysis. Upon removal of the surgical specimen, research personnel performed a gross evaluation of the specimen and selected the gallbladder tissues that appeared to be cancerous. Paired noncancerous tissues were isolated from at least 3 cm away from the tumor border. All patients in this study met the following inclusion criteria: Resected nodules were identified as GC by pathological examination; no anticancer treatments were given before surgery; complete resection of all tumor nodules was verified by the cut surface being free of cancer by pathological examination; and complete clinical-pathologic and follow-up data were available. Patients that died of noncancer-related diseases or accidents were excluded. Gallbladder cancer patients were staged according to the tumor node metastasis (TNM) staging system (the 8th edition) of the American Joint Committee on Cancer (AJCC) staging system.

Follow-up investigation include clinical check-up, laboratory parameters including CA199, CEA, CA72-4 and CA125, radiological assessment (chest X-ray, a magnetic resonance imaging or Contrast-enhanced CT scan of the upper abdomen every three months during the first two years and every six months afterwards). The recurrence was diagnosed comprehensively based on the results of radiological and histopathological examinations. Once recurrent tumors were confirmed, further treatment was implemented based on the tumor's diameter, number and location. The recurrence-free survival (RFS) was calculated from the date of tumor resection until the detection of

tumor recurrence, death, or the last observation. The overall survival (OS) was defined as the length of time between the surgery and death or the last follow-up examination.

Human gastric mucosal tissues (normal tissues, tissues diagnosed with IM or DYS) were obtained from patients by gastroscope inspection in Renji Hospital with writtern informed consent. None of the patients had taken nonsteroidal anti-inflammatory drugs, H_2 receptor antagonists, proton pump inhibitors or antimicrobials 4 weeks before the study. The different extent of inflammation in these tissues was examined according to the updated Sydney system.

Cell culture

These cell lines were immediately expanded and frozen so that they could be restarted every 3 to 4 months from a frozen vial of the same batch of cells. All cell lines have been passaged for fewer than 6 months in our laboratory after resuscitation.

ROC curve analysis

The upper 95% reference interval of *GCAWKR* value in controls was set as the threshold to code the expression level of the corresponding factor for each sample as 0 and 1 in patients. A risk score function (RSF) to predict GC was defined according to a linear combination of the expression level for the two factors. For example, the RSF for sample *i* using information from the two factors was: RSFi=Σ2j−1Wj.sij. In the above equation, *sij* is the risk score for factor j on sample *i*, and *Wj* is the weight of the risk

score of factor *j*. The risk score of the two factors was calculated using the weight by the regression coefficient that was derived from the univariate logistic regression analysis of each factor. Samples were ranked according to their RSF and then divided into a high-risk group and a low-risk group. Frequency tables and ROC curves were then used to evaluate the diagnostic effects of the profiling and to find the appropriate cutoff point. The cluster analysis was based on the RSF results. Statistical analysis was performed using STATA 9.2 (Stata Corp., College Station, TX, USA) and presented with the GraphPad Prism 5.0 software (La Jolla, CA, USA). Results were considered statistically significant at $P<0.05$.

ISH staining evaluation

The ISH-staining regions were reviewed and scored by 2 pathologists, the score standard for the staining intensity was 0 (negative), 1 (weak), 2-3 (medium) and 4-5 (strong); and 0 (0%), 1 (1-20%), 2 (21-40%), 3 (41-60%), 4 (61-80%) and 5 (81-100%) for the staining extent. The total scores contain the intensity and extent scores ranging from 0 to 5. We defined total scores of $\geqslant 4$ as the high-expression group (positive group). This scoring method was simple, reproducible. Results were highly concordant between the two independent pathologists. The sequence of probe was (5'-CACAGAGTATGCTTATTTGTCAAAGTAGAATGATACACCC-3').

5'and 3'rapid amplification of cDNA ends (RACE) analysis

We used the 5'- and 3'-RACE analyses to determine the transcriptional initiation and

termination site of GCAWKR using a SMARTer RACE cDNA Amplification Kit (Clontech, Palo Alto, CA, USA), according to the manufacturer's instructions. Polymerase chain reaction (PCR) of the internal region was performed when starting points of 5' and 3' RACE had an unamplified gap. RACE PCR products were separated on a 1.5% agarose gel. Gel products were extracted with the Gel and PCR Clean-Up System (Promega, A9282), cloned into the pGEM-TVector Systems I (Progema, A3600) and sequenced bidirectionally using the M13 forward and reverse primers by Sanger sequencing at Invitrogen. At least five colonies were sequenced for every RACE PCR product that was gel purified. The gene-specific primers used for the PCR of the RACE analysis were given at Supplementary Table S7.

Assessment of lncRNA protein-coding potential

We determined whether this transcript has protein-coding potential using an in vitro translation assay and a combination of online softwares. For the in vitro translation assay, full-length GCAWKR (amplified using primer GCAWKR F2 and R2) was cloned into a pcDNA vector and expressed using the TnT Quick Coupled Transcription/Translation System (Promega). The absence of a specific band indicated that GCAWKR is a transcript with no protein-coding capacity. Luciferase in vitro translation served as positive control. The online softwares include ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), PhyloCSF (https://github.com/mlin/PhyloCSF/wiki) and Coding Potential Calculator (CPC; http://cpc.cbi.pku.edu.cn/).

Subcellular fraction

To determine the cellular localization of *GCAWKR*, cytosolic and nuclear fractions were isolated and collected with the Nuclear/cytoplasmic Isolation Kit (Biovision, USA) according to the manufacturer's instructions. After that, total RNA were extracted from the collections of both cytoplasm and nucleus and cDNA was synthesized for the evaluation of *GCAWKR*. Briefly, we lysed SGC7901 and BGC823 cells (provided by the Cell Bank of Chinese Academy of Science) with a buffer containing 10mM Tris-HCl (pH=7.4), 100mM NaCl, 2.5mM $MgCl_2$, and 40mg/ml digitonin for 10min. The resulting lysates centrifuged with 2,060×g for 10min at 4°C. The supernatant was used for the cytosolic fraction. Subsequently, the pellets were washed and incubated with RIPA buffer at 4°C for 10 min. After centrifugation at 4°C for 10min at 2,060×g, the nuclear fraction was collected. RNAs extracted from each of the fractions were subjected to following RT-PCR analysis of the levels of nuclear control transcript (U6), cytoplasmic control transcript (β-Actin) and *GCAWKR*.

RNA preparation and qRT-PCR

Total RNA from tissues and cells was extracted using Trizol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The quality of total RNA was detected at an A260/A280 ratio using 1% agarose gel electrophoresis. The GoScript Reverse Transcription System (Promega, Madison, Wis) was used to generate combinational DNA. The cDNA template was amplified by real-time RT-PCR using

the SYBR Premix Dimmer Eraser kit (TaKaRa). Gene expression in each sample was normalized to β-Actin expression. Real-time RT-PCR reactions were performed by the ABI7500 system (Applied Biosystems, Carlsbad, CA, USA). The real-time PCRs were performed in triplicate. The relative expression fold change of mRNAs was calculated by the $2^{-\Delta\Delta Ct}$ method. Primers used in this study are listed in Supplementary Table S7.

Western blot analysis

The harvested cells were centrifuged at 6,000 rpm for 1 min. The total cellular proteins were prepared using RIPA cell lysis buffer (Cell Signaling Technology) supplemented with protease inhibitors. The lysates were then collected and subjected to ultrasonication and centrifugation. The supernatants were collected, and protein content was determined by Bradford assay. Equal amounts (30-50 μg) of proteins were applied to an 8-12% SDS-polyacrylamide separating gel and transferred to a PVDF Immobilon-P membrane (Millipore). The membrane was blocked with 5% skim milk in TBST and then probed with indicated primary antibodies with gentle shaking at 4°C overnight. The membranes were washed with TBST (3×10 min), incubated in secondary antibodies conjugated to horseradish peroxidase at room temperature for 1 hour. Antibody-bound proteins were detected by ECL (enhanced chemiluminescence) Western Blotting Substrate (Pierce, Rockford, IL). The band intensity of western blotting and the normalization were analyzed using the Image J program (National Institutes of Health, Bethesda, MD). The antibodies used were listed in Supplementary Table S8.

Cell viability assay

Cell viability was assessed by the Cell Counting Kit 8 (CCK-8, Donjindo). Briefly, control and treated GC cells were seeded into 96-well plates at an initial density of 10000 cells per 100μl. After 24, 48, 72, and 96 hours of cultivation, CCK-8 solution (10μl per 100μl of medium in each well) was added to each well and incubated for 2 h. The absorbance was measured by scanning with a microplate reader (MRX; Dynex Technologies, West Sussex, United Kingdom) at 450 nm. Data were expressed as the percentage of viable cells as follows: relative viability (%)=($A450_{treated}$-$A450_{blank}$)/($A450_{control}$ - $A450_{blank}$)×100%. Six replicates for each group and the experiment were repeated at least 3 times.

Transwell assay

The ~~lower~~ upper chambers were pre-coated with 100 μl Matrigel (BD Bioscience, USA) for 30 min, and 24 h after transfection, cells were seeded in the upper chamber at a density of 3.0×10^4/well in 200 μl serum-free medium with mitomycin C (MMC, 5 μM). Medium containing 20% fetal bovine serum medium was applied to the lower chamber as a chemo-attractant. After a 24 h incubation at 37°C, cells that invaded through the Matrigel and adhered to the lower surface of the filter were fixed with ethanol, stained with crystal violet (Life, USA), photographed at 40 ×, and counted in 10 different fields to determine the average number of cells at 200× under a microscope (BX51, Olympus, Japan).

Drug Sensitivity Assays to 5'-Fluorouracil and Cisplatin

Cell proliferation was assessed by the BrdUrd incorporation assay (Roche Molecular Biochemicals). Briefly, control and shRNA-transfected cells, which were seeded onto 96-well plates at an initial density of 5×10^3 cells per well, were treated with different doses of 5'-fluorouracil and cisplatin, and BrdUrd labeling solution (10 μL/well) was added to the cells at specified time points. After 2-hour incubation, culture medium was removed and the cells were fixed. Then DNA was denatured by adding FixDenat (200 μL/well), and anti–BrdUrd-POD working solution (100 μL/well) was added to the cells and incubated for 90 minutes. The immune complexes were detected by the subsequent substrate reaction. The reaction product was quantified by measuring the absorbance at 370 nm (reference wavelength: approximately 492 nm).

RNA sequencing

We transiently transfected 5.0×10^6 SGC7901 cells with shRNA, and the total RNA samples were collected by TRIzol reagent. Before the RNA libraries were constructed, rRNAs in the RNA samples were eliminated using the RiboMinus Eukaryote kit (Qiagen, Valencia, CA, USA). Next, strand-specific RNA-seq libraries were prepared using the NEBNext Ultra Directional RNA Library Prep kit (New England Biolabs, Beverly, MA, USA), according to the manufacturer's instructions. Briefly, ribosome-depleted RNA samples were fragmented and prepared for first- and second-strand cDNA synthesis with random hexamer primers. The prepared cDNA fragments were

treated with End-It DNA End Repair kit to repair the ends, an A was added at the 3'-end by the Klenow fragment, and finally, the fragments were ligated with adaptor sequences. The ligated cDNA products treated with uracil DNA glycosylase to remove the dUTP-labelled second-strand cDNA. The purified libraries were subjected to quality control on a Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) and sequenced using a HiSeq 3000 (Illumina, San Diego, CA, USA) on a 150-bp paired-end run. For the data processing, the raw sequencing reads were aligned to human reference genome (hg19) using the splice-aware aligner HISAT246. Read counts for each gene were normalized into FPKM (Fragments Per Kilobase of transcript per Million mapped reads) values. The cutoff of differential gene expression was FDR <0.05, normalized by the respective shRNA-NC control.

RNA Pull-Down Assay

Synthesized Biotinylated lncRNAs were refolded in NEB enzyme buffer with RNase-out (Invitrogen, USA) at a final concentration of 200 ng/μL. The diluted RNAs were incubated at 60 °C for 10 min and slowly cooled to 4 °C. Aliquots of 2 μg of folded RNAs were used for pull-down experiments. To prepare cell lysates, cells were harvested into 5 mL of buffer A [10 mM Tris·HCl, pH 7.0, 1.5 mM MgCl2, 10 mM KCl, 0.5 mM DTT, 1 mM PMSF, and protease inhibitor mixture (Roche Molecular Biochemicals, Mannheim, Germany)]. Cells were lysed by the addition of 0.25% Nonidet P-40 and incubated for 10 min at 4 °C. The lysates were centrifuged at 2,500 × g for 15 min, and the supernatant was removed. Pellets containing the nuclear

fractions were resuspended in 3 mL of buffer C (25 mM Tris·HCl, pH 7.0, 0.5% Nonidet P-40, 150 mM KCl, 0.5 mM DTT, and protease inhibitor mixture) and mechanically sheared by homogenizing for 15–20 strokes. Samples were cleared by centrifuging at 15,000 × g for 10 min. Protein concentrations in the nuclear lysates were measured by the DC assay (Bio-Rad, USA). For the pull-down incubations, nuclear lysates containing 1 mg of protein were precleared with streptavidin beads and then incubated with 2 μg of synthesized biotinylated RNA and 40 μl of streptavidin beads for 2 hours at 4 °C. Beads were collected by centrifugation and washed with buffer C three times. RNA-associated proteins were eluted and resolved by SDS/PAGE.

RNA Immunoprecipitation (RIP)

RIP experiments were performed using a Magna RIP RNA-Binding Protein Immunoprecipitation Kit (Millipore) according to the manufacturer's instructions. The pyruvate carboxylase antibodies were used for RIP (Abcam). The coprecipitated RNAs were detected by reverse transcription PCR and quantitative PCR. The primer sequences are listed in Supplemtary Table S7. Total RNAs (input controls) and isotype controls were assayed simultaneously to demonstrate that the detected signals were the result of RNAs specifically binding to WDR5 or KAT2A (n=3 for each experiment).

Chromatin immunoprecipitation

GC cells were serum-starved overnight. Chromatin was cross-linked with 1% formaldehyde for 10 min. After cell lysis, the chromatin was sonicated with a Bioruptor (Diagenode) in a cold room using the following parameters: H- high setting, pulse

interval- 30 sec ON and 45 sec OFF, cycle time- 15 min each. Change ice in water bath chamber after each cycle. After about 9 cycles, a DNA smear with an average size of 500 bp was obtained. After centrifugation, the supernatants were subjected to immunoprecipitation overnight with antibodies at 4°C, or with isotype rabbit IgG at 4°C overnight. Chromatin-antibody complexes were isolated using Protein A/G PLUS Agarose (Santa Cruz). The crosslinks for the enriched and the input DNA were reversed and the DNA was cleaned by RNase A (0.2 mg/mL) and proteinase K (2 mg/mL) before phenol/chloroform-purification. PCR was employed to analyzed the specific sequences from immunoprecipitated and input DNA. The results are representative of at least three independent experiments.

Coimmunoprecipitation

Briefly, both input and IP samples were analyzed by Western blot using various antibodies at the indicated dilutions: anti-WDR5 antibody (1:1,000), anti-KAT2A antibody (1:1,000), and normal rabbit IgG.

Plasmid construction, lentiviral construction, and cell transfections

For *GCAWKR* overexpression, the full-length cDNA of human *GCAWKR* was synthesized by GeneWiz (Beijing, China) and subcloned to pGC-LV vectors (Genechem Company, Shanghai, China) before sequenced. To produce lentivirus containing *GCAWKR* gene, HEK-293FT cells were co-transfected with the resulting vector described above, pHelper 1.0 and pHelper 2.0 (Genechem Company, Shanghai, China) using Lipofectamine 2000 according to the manufacturer's guidelines. Infectious lentiviruses were harvested at 48h post transfection and filtered through 0.45μm PVDF filters, designated LV-*GCAWKR*. For negative control of LV-*GCAWKR*,

we used empty vectors containing the green fluorescent protein as the negative control and designated “LV-Control”. Recombinant lentiviruses were concentrated 100-fold by ultracentrifugation (2 h at 50,000 g). The virus-containing pellet was dissolved in DMEM, aliquoted and stored at -80 °C. GC cells were infected with concentrated virus at a multiplicity of infection of 60 or 100 in the presence of 8 μg/ml polybrene (Sigma-Aldrich, St. Louis, MO). The supernatant was replaced with complete culture media after 24 h. The expression of *GCAWKR* in infected cells was confirmed by RT-PCR 96 h after infection.

Two pairs of cDNA oligonucleotides were designed and synthesized by GenePharma (Shanghai, China) to suppress the *GCAWKR* expression. The design of the shRNAs was assisted by the use of web-based software provided by Invitrogen (http://rnaidesigner.invitrogen.com/rnaiexpress/). Blast searches were performed using the National Center for Biotechnology Information expressed sequence tag database to ensure that the shRNA construct only targeted human *GCAWKR* expression. After annealing, double-strand oligos were inserted to the liner vector BLOCK-iT™ Pol II miR RNAi Expression Vector (Invitrogen, Catalog no. K4936-00). The resulting plasmids were sequenced, designated as *GCAWKR* shRNA-1, *GCAWKR* shRNA-2 and shRNA control. The final double-strand oligo DNAs for *GCAWKR* and negative control are presented in Supplementary Table 3.

The cDNA encoding PTP4A1 was PCR-amplified by the Pfu Ultra II Fusion HS DNA Polymerase (Stratagene, Agilent Technologies) and subcloned into the Hind III and

EcoR I sites of pcDNA3.1 vector (Invitrogen), named pcDNA3.1-PTP4A1.

Cells were grown on six-well plates to 60% confluency, 50 nM of WDR5/KAT2A siRNA (GenePharma, Shanghai, China), was transfected into GC cells using Lipofectamine 2000 Reagent (Life Technologies, USA) according to the manufacturer's instructions. The oligonucleotide sequences are presented in Supplementary Table 3. Cells that were transfected with the transfection agent either without siRNA (Mock) or with scramble-control siRNA (Negative control, NC) were used as controls.

For the deletion-mapping experiments, *GCAWKR* fragment (1-500, 500-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, and 3000-3893) were inserted into PBSK vector to generate PBSK-*GCAWKR*, PBSKΔ1, PBSKΔ2, PBSKΔ3, PBSKΔ4, PBSKΔ5, PBSKΔ6 and PBSKΔ7 vectors.

Luciferase reporter assay

Cells were transfected with the pGL3-based constructs containing the PTP4A1 promoter plus the Renilla luciferase plasmid (pRL-TK). Then, the cells were harvested after 48 hours for firefly/Renilla luciferase assays using the Dual-Luciferase Reporter Assay System (Promega). Luciferase activities were normalized to the cotransfected pRL-TK plasmid (mean ± SD).