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.

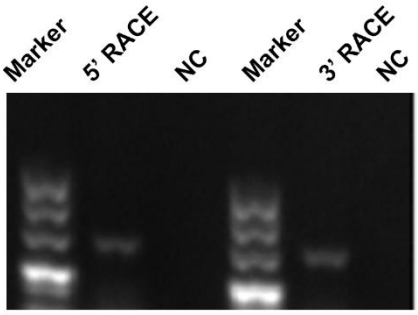
Supplemental Table 7. Primer sequence used in this study.

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’

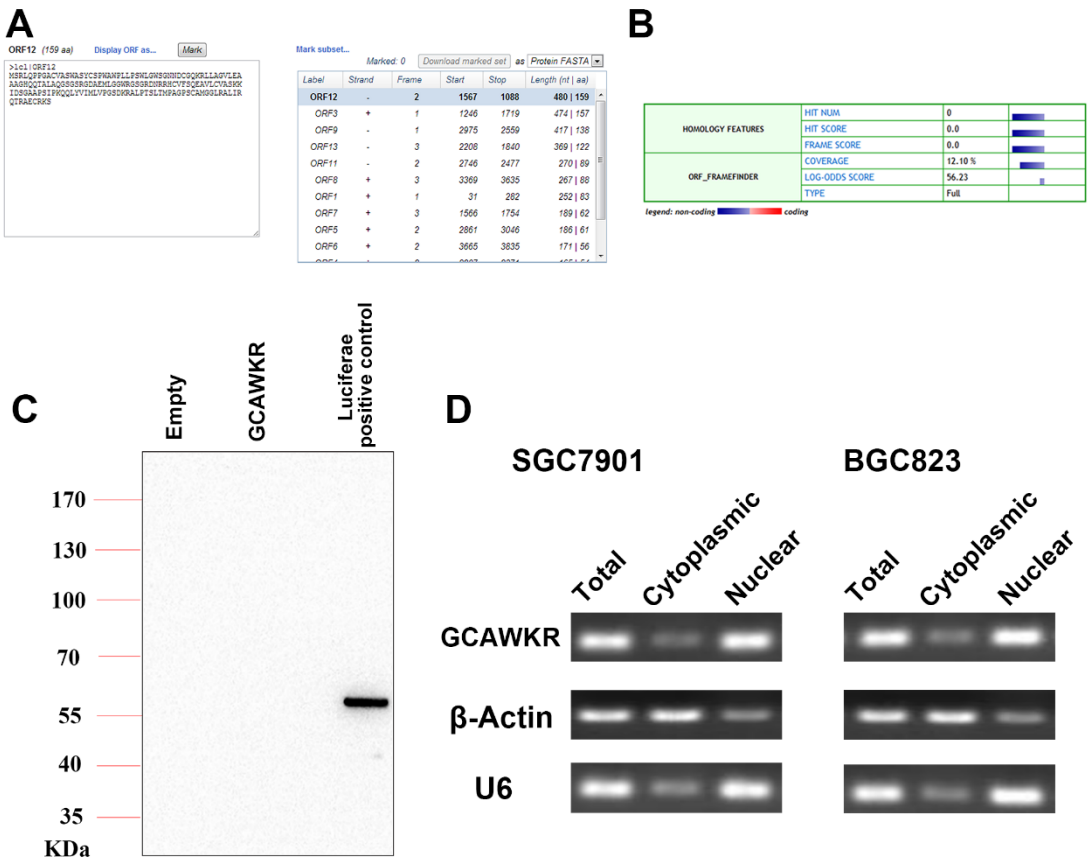
Supplemental Table 8. Information on antibodies used for correlation analysis.

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)

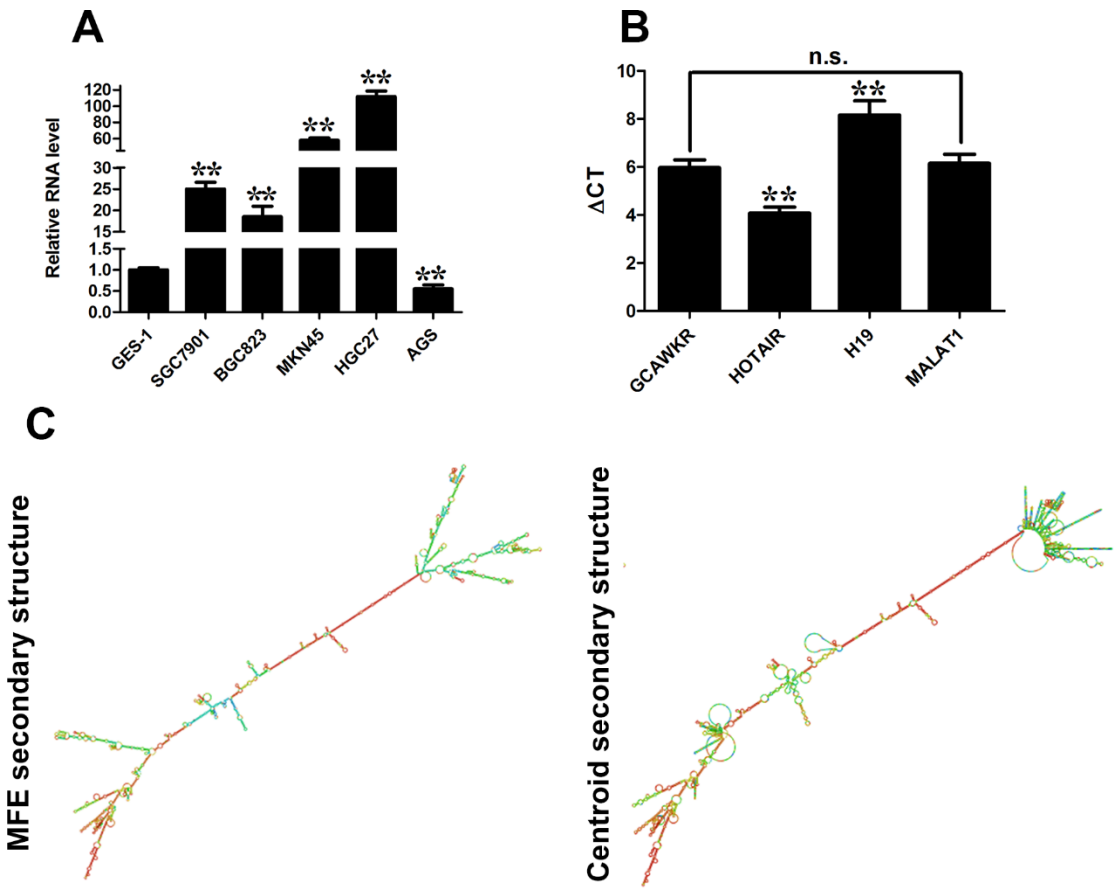


5'-GATTCTGTCGCGAACCCTCTCACGGCAGCTATGAGACTGCTGCGCTCCCCAGACGAGGAAACCTGGTCTCCCCGGAGCCAGGT GGA GTC T
CGCTC TGTCGCCCAGGCTGGAGTGCGGTGGCGCGATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCA GCCT
CCCGAGTAGCTGGGACTACAGAGTCTTGCTCTGTCGCCCAGGCTGGAGTGCAGTGGCGCGATCTCGGCTCACTACAACCTCTGCCTCCCAGGT
TCAAGTGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGCTCTGGGACAAACCAGCTTCTATGTCATGAGCACACTCAGGCAGCTC
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'

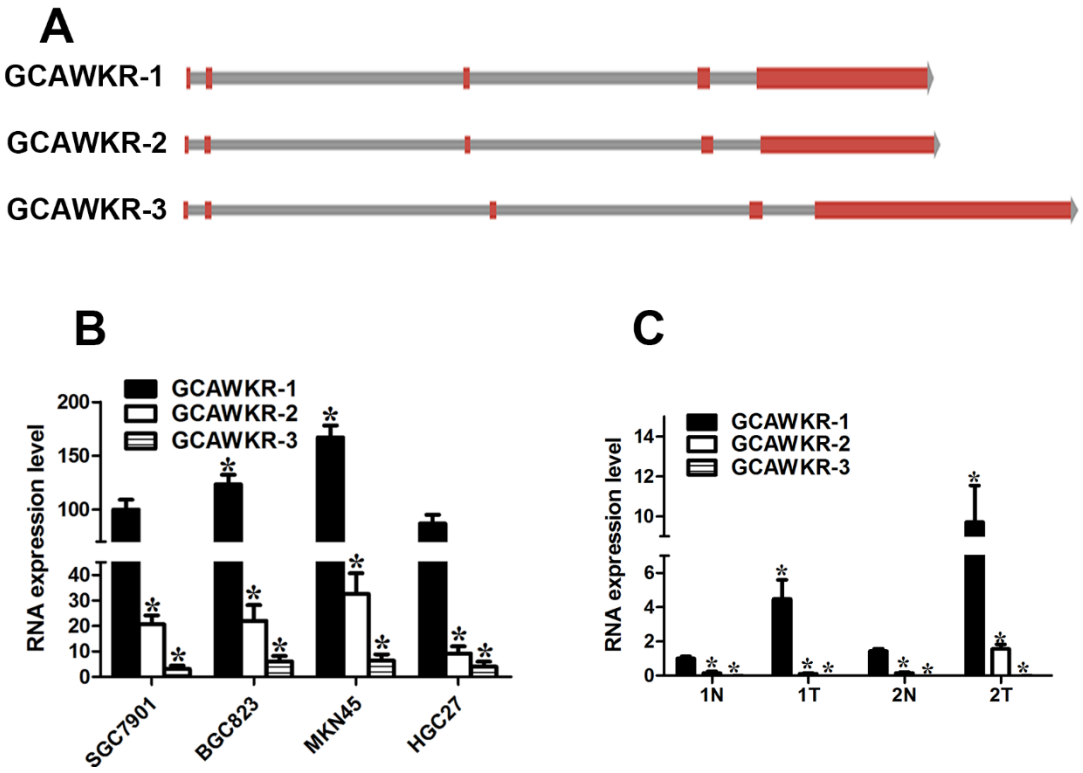
Supplementary Figure1. Left: PCR products from the 5'RACE and 3'RACE procedure was shown by agarose gel electrophoresis. Right: Nucleotide sequence of the full-length human *GCAWKR*.



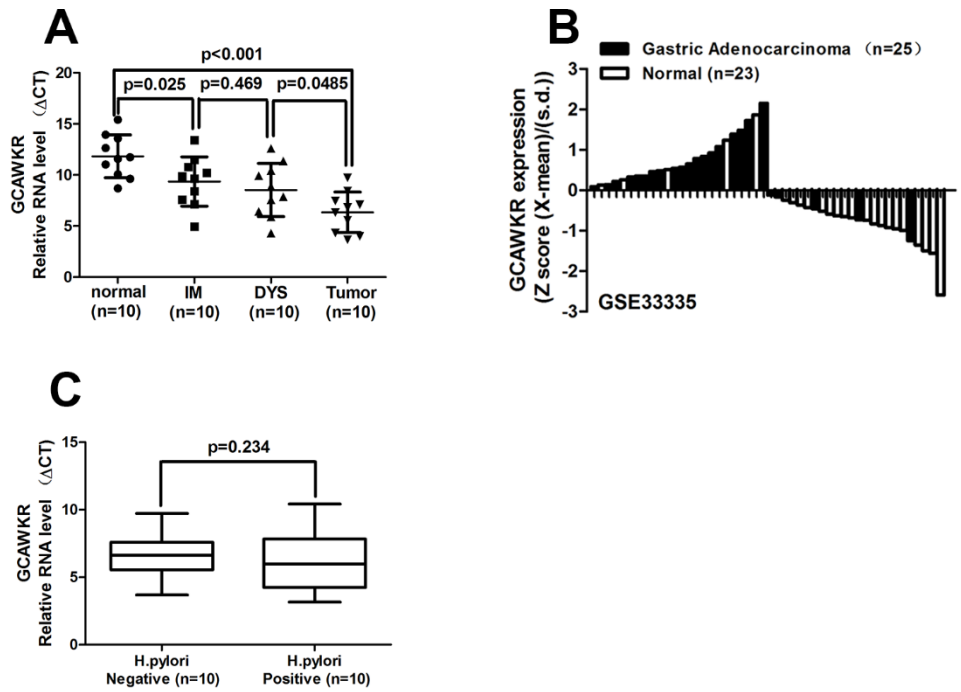
Supplementary Figure 2. *GCAWKR* has no coding capability. (A) Putative proteins possibly encoded by *GCAWKR* as predicted by the ORF Finder. (B) The transcript's noncoding nature was suggested by negative score with CPC. The transcript's noncoding nature was suggested by negative score with PhyloCSF (-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.



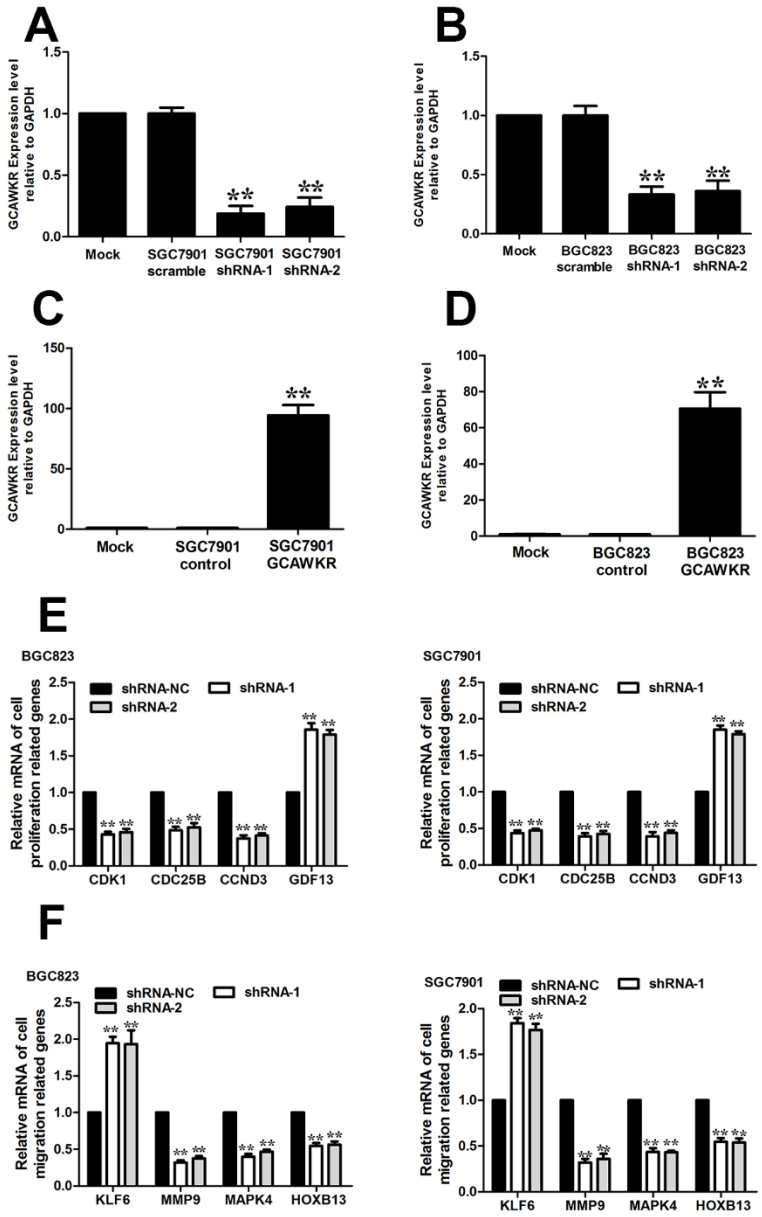
Supplemental Figure 3. (A) The relative expression of *GCAWKR* detected in normal gastric epithelium cell line GSE-1 and GC cell lines SGC7901, BGC823, MKN45, HGC27 and AGS using qRT-PCR. Error bars show standard deviation. ** $p < 0.01$, significantly different from GES-1 cells. (B) The expression of *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.



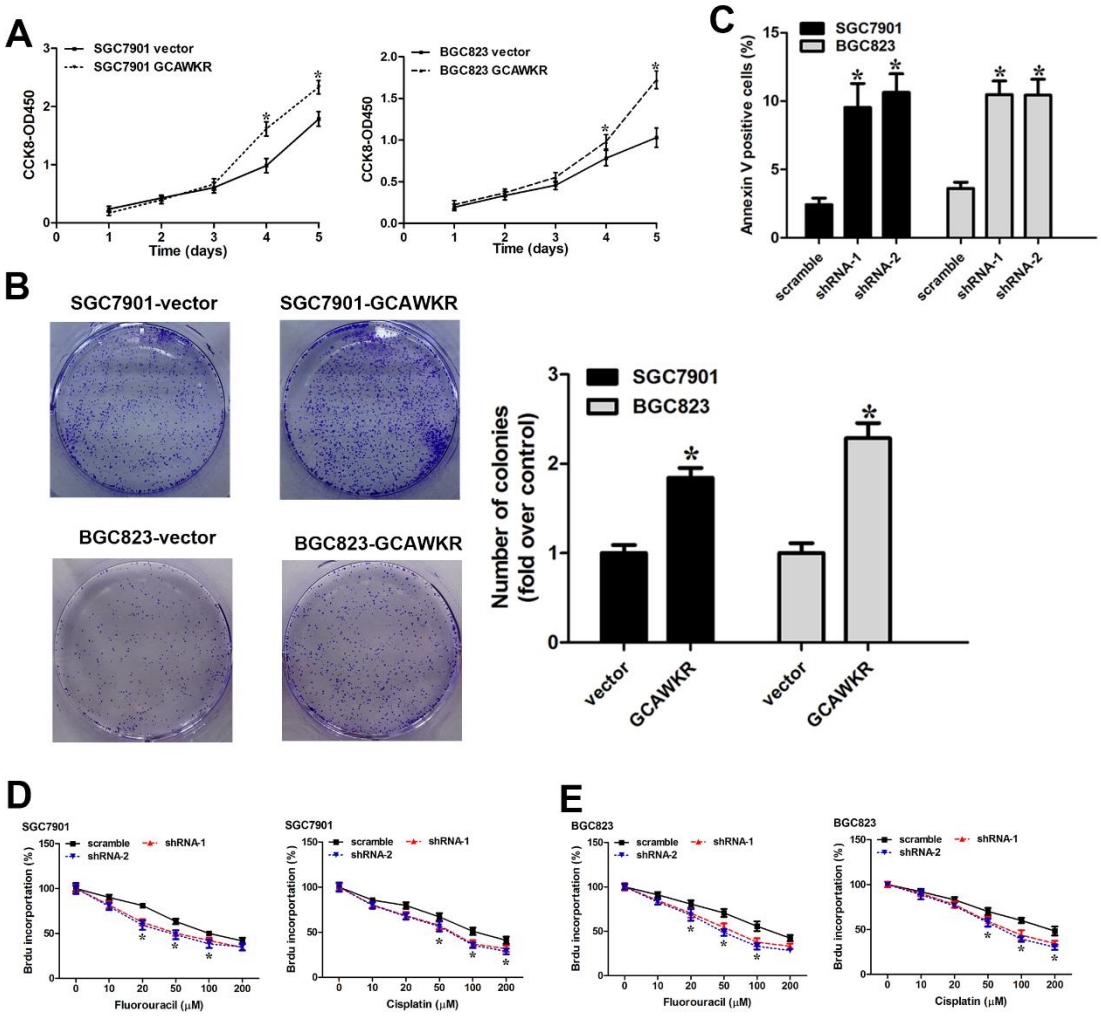
Supplemental Figure 5. (A) Statistical analysis of *GCAWKR* expression is shown in normal, intestinal metaplasia (IM), dysplasia, and cancerous gastric tissues. non-parametric Mann-Whitney test. (B) The expression of *GCAWKR* was validated in an independent gastric cancer patient cohort. (C) *GCAWKR*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



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

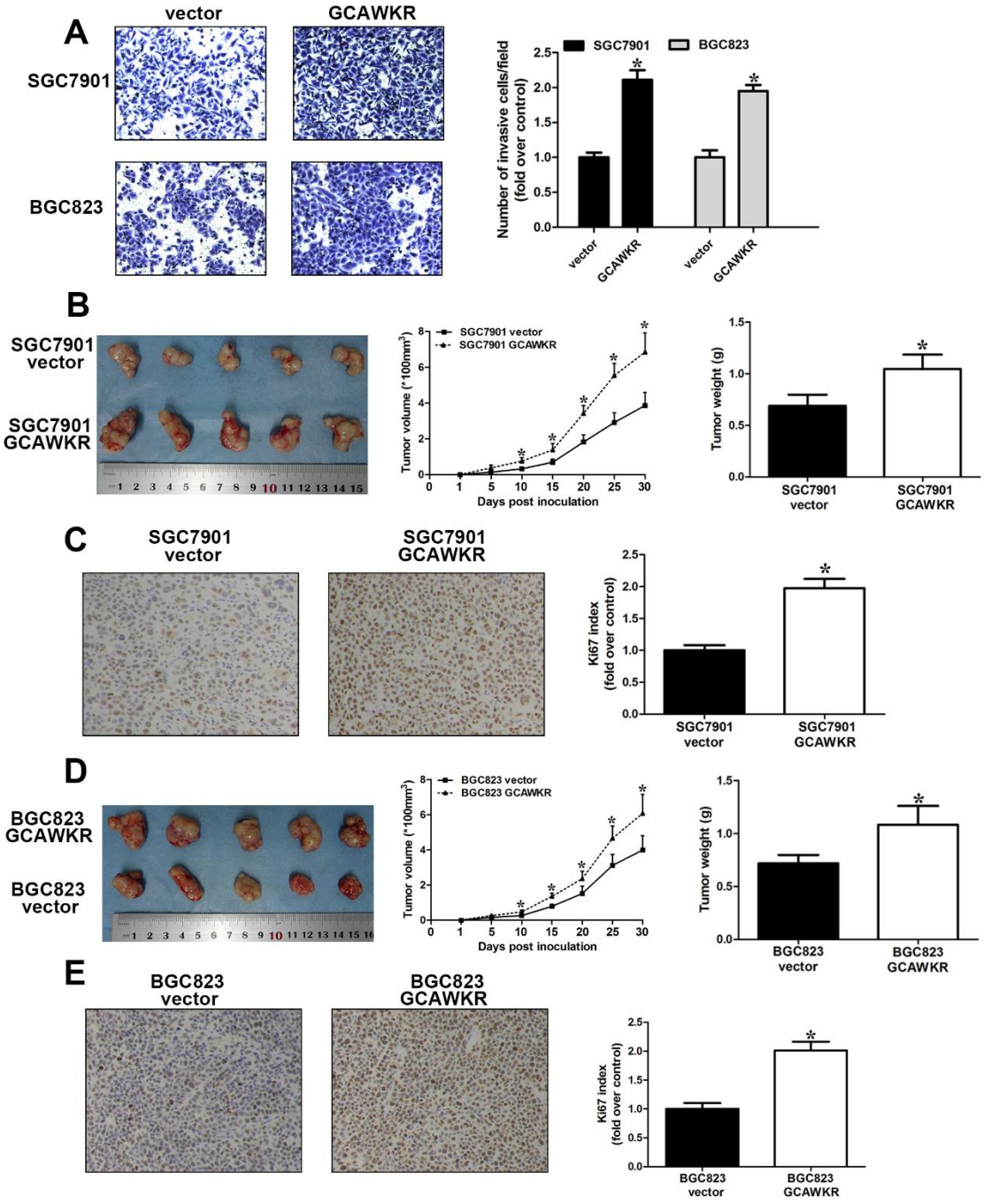
(A,B) *GCAWKR* expression was detected in SGC7901 (A) and BGC823 (B) cells by qRT-PCR after transduction of lentiviruses encoding *GCAWKR* shRNA-1 or shRNA-2 or a scrambled shRNA. **, $p < 0.01$. (C) *GCAWKR* expression was detected in SGC7901 (C) and BGC823 (D) cells by qRT-PCR after transfection of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



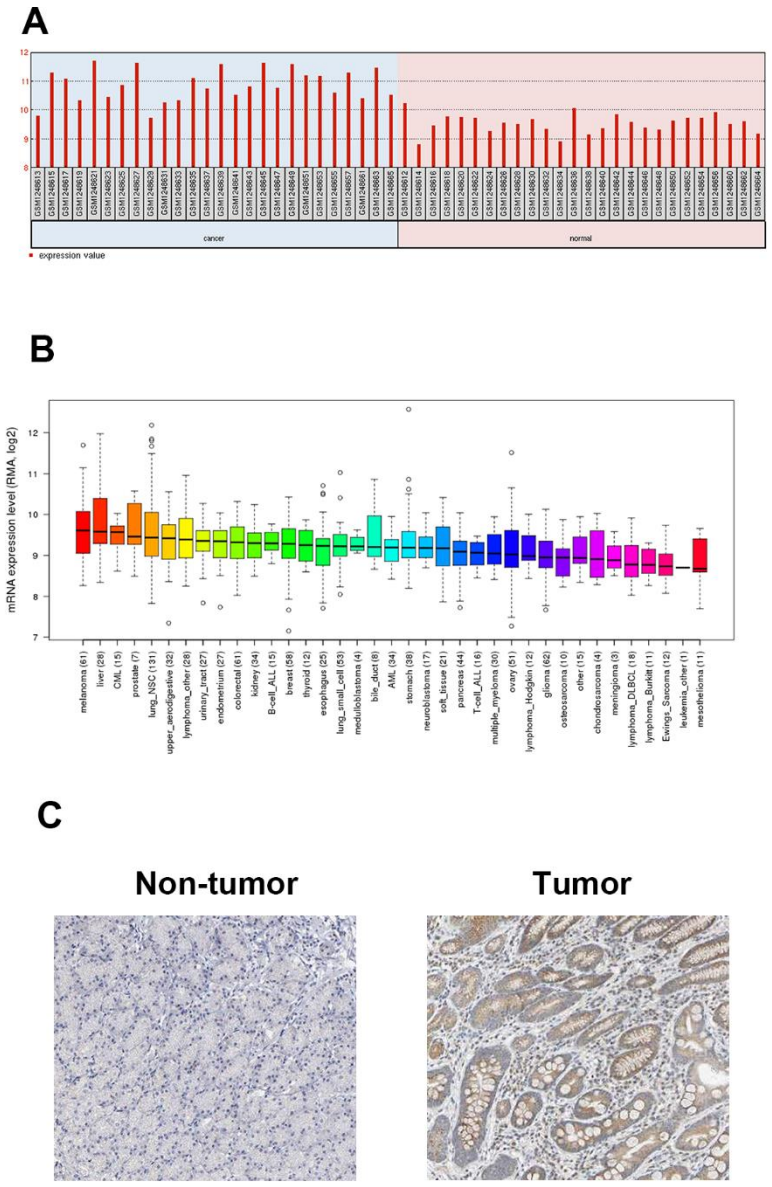
Supplemental Figure 7. (A) The cell growth rates were determined with CCK-8 proliferation assays. *GCAWKR* overexpression in SGC7901 and BGC823 cells significantly inhibited cell proliferation. (B) Colony formation assays were used to determine the cell colony formation ability of SGC7901 and BGC823 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.

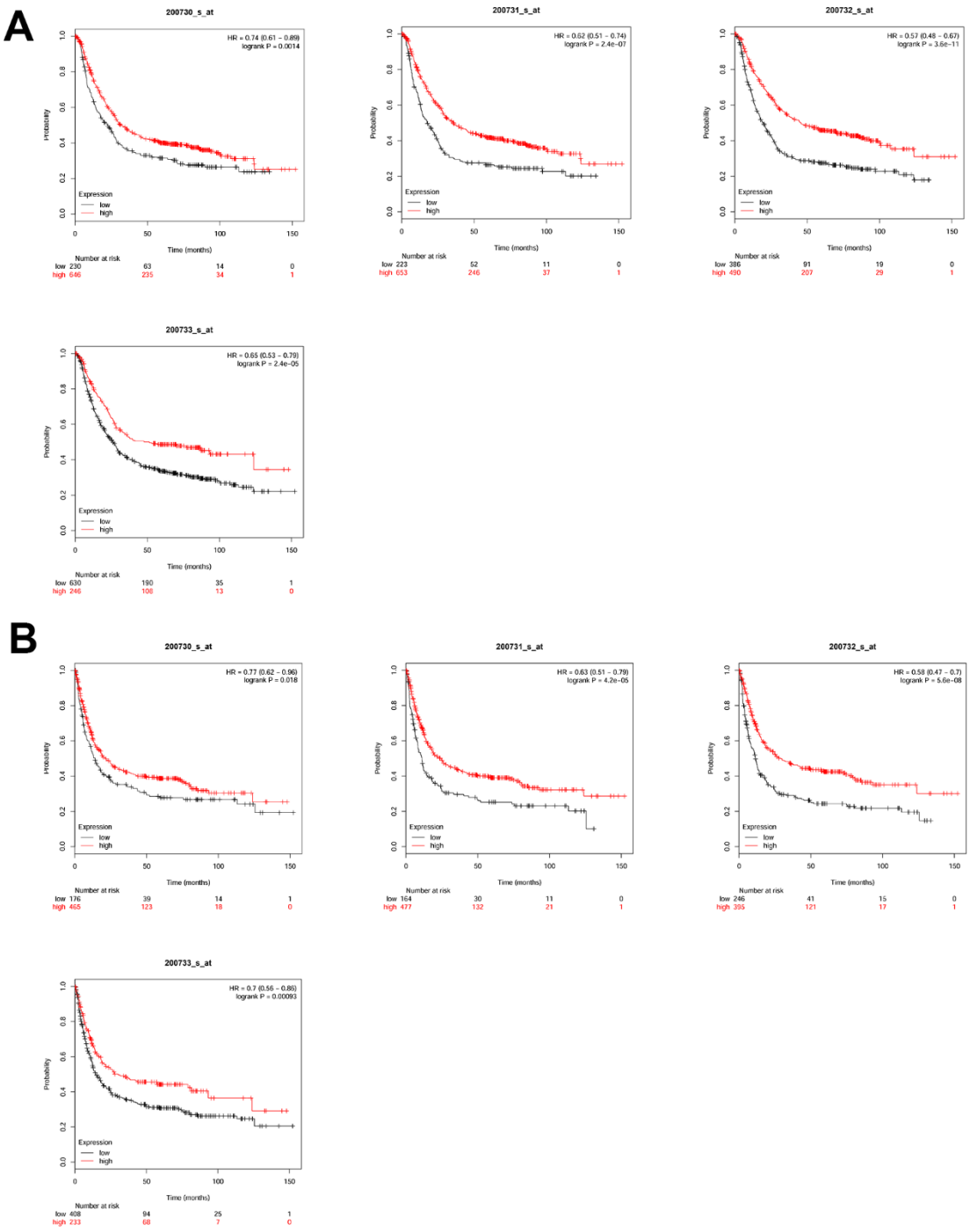


Supplemental Figure 8. (A) Effects of *GCAWKR* overexpression on cell invasion in the presence of the anti-proliferative drug mitomycin C (MMC, 5 μM) treatment were determined using the transwell assay. *, $p < 0.05$. (B-E) Top left, representative images

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).

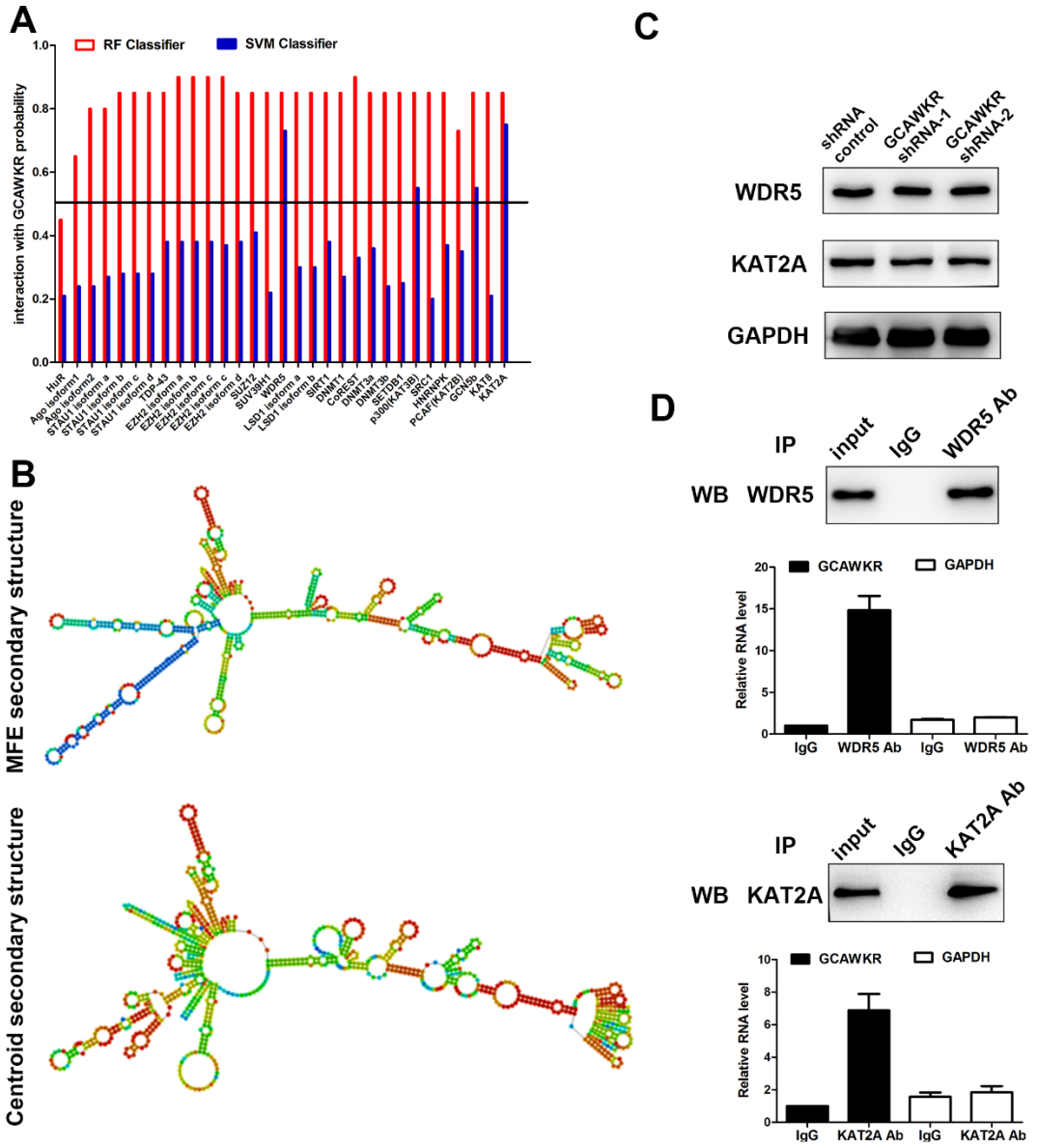


Supplemental Figure 9. Analysis of PTP4A1 expression levels in GC tissues. (A) Analysis of PTP4A1 expression levels in gastric cancer tissues and normal tissues from GEO dataset GSE51575. (B) PTP4A1 expression levels vary across different cancer types in the TCGA database. (C) Immunohistochemical staining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



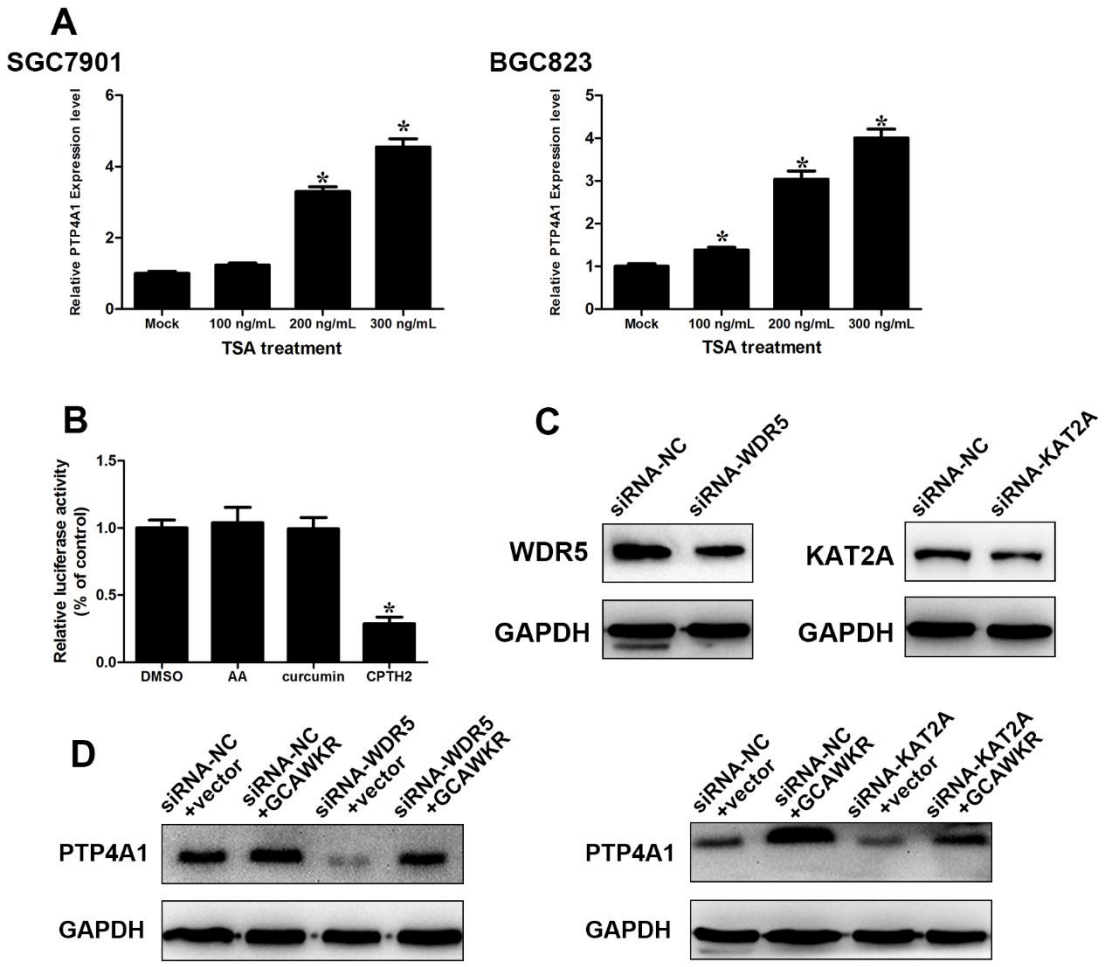
Supplemental Figure 10. (A) A higher expression level of PTP4A1 is correlated with a significantly poorer DSS (p<0.05) according to data from the KMPlot database. (B) A

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.



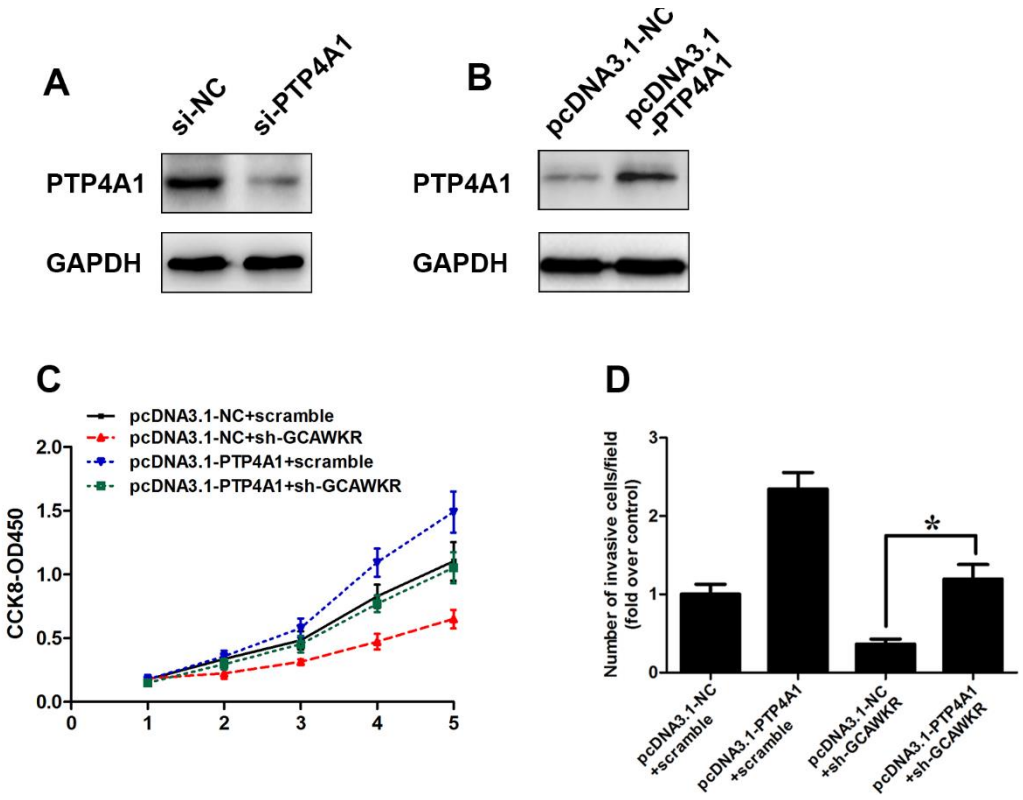
Supplemental Figure 11. (A) The interaction of *GCAWKR* and potential RNA binding proteins was predicted on http://pridb.gdcb.iastate.edu/RPISeq/. (B) 3'-end segment of *GCAWKR* (3000~3893nt) is predicted to have a stable stem-loop structure. Prediction of *GCAWKR* structure of a 3000~3893nt region based 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).