

A novel lncRNA GClnc1 promotes gastric carcinogenesis and may act as a modular scaffold of WDR5 and KAT2A complexes to specify the histone modification pattern

Tian-Tian Sun^{1,2†}, Jie He^{3†}, Qian Liang¹, Lin-Lin Ren¹, Ting-Ting Yan¹, Ta-Chung Yu¹, Jia-Yin Tang¹, Yu-Jie Bao¹, Ye Hu¹, Yanwei Lin^{1,2}, Danfeng Sun^{1,2}, Ying-Xuan Chen¹, Jie Hong¹, Haoyan Chen¹, Weiping Zou² and Jing-Yuan Fang¹

¹State Key Laboratory for Oncogenes and Related Genes, Key Laboratory of Gastroenterology & Hepatology, Ministry of Health, Division of Gastroenterology and Hepatology, Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai Cancer Institute, Shanghai Institute of Digestive Disease, 145 Middle Shandong Road, Shanghai 200001, China.

²Department of Surgery, University of Michigan School of Medicine, Ann Arbor, MI 48109, USA.

³Department of Gastroenterology & Guangzhou Key Laboratory of Digestive Disease, Guangzhou Digestive Disease Center, Guangzhou First People's Hospital, Guangzhou Medical University, Guangzhou, China

Corresponding Authors: Jing-Yuan Fang, Jie Hong and Haoyan Chen, Division of Gastroenterology and Hepatology, Renji Hospital, Shanghai Jiao-Tong University School of Medicine, 145 Middle Shandong Road, Shanghai 200001, China. Phone: 86-21-53882450; Email: jingyuanfang@sjtu.edu.cn, jiehong97@shsmu.edu.cn or haoyanchen@shsmu.edu.cn; Or Weiping Zou, Department of Surgery, University of Michigan School of Medicine, Ann Arbor, MI, USA, 48109. Phone: 001-734-763-6402; Email: wzou@med.umich.edu

Running title: lncRNA GClnc1 modifies histone complex

Authorship note:

†The two authors contributed equally to this work

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Transcript Profiling: The RNA sequence and CHIP sequence data have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE63765.

Abstract: Long non-coding RNAs (lncRNAs) play a role in carcinogenesis. However, the function of lncRNAs in human gastric cancer remains largely unknown. In this study, we identified a novel lncRNA, GClncl, which was upregulated and associated with tumorigenesis, tumor size, metastasis, and poor prognosis in gastric cancer. GClncl affected gastric cancer cell proliferation, invasiveness, and metastasis in multiple gastric cancer models. Mechanistically, GClncl bound WDR5 (a key component of histone methyltransferase complex) and KAT2A histone acetyltransferase, acted as a modular scaffold of WDR5 and KAT2A complexes, coordinated their localization, and specified the histone modification pattern on the target genes, including *SOD2*, and consequently altered gastric cancer cell biology. Thus, GClncl is mechanistically, functionally, and clinically oncogenic in gastric cancer. Targeting GClncl and its pathway may be meaningful for treating patients with gastric cancer.

Significance: This report documents a novel lncRNA, GClncl, which may act as a scaffold to recruit the WDR5 and KAT2A complex, and modify the transcription of target genes. This study reveals that GClncl is an oncogenic lncRNA in human gastric cancer.

Key words: Long non-coding RNA, GClncl, SOD2, gastric cancer, WDR5, KAT2A, carcinogenesis

Introduction

Gastric cancer (GC) is the fourth most highly diagnosed type of cancer and the third most common cause of cancer-related death worldwide(1, 2). Patients with advanced gastric cancer have a poor prognosis(3, 4). Pathological classification is used to assess prognosis and inform the treatment of gastric cancer. Massive efforts have been made to develop the noninvasive biomarkers to detect early cancer and/or reflect an individual's cancer risk, which is essential to reducing GC mortality(5). However, there has been little success in improving the disease-free survival rate of patients. Since the pathological mechanisms of gastric cancer progression are not fully understood, more research is needed to discover and develop effective biomarkers and targets for gastric cancer diagnosis and treatment.

Long non-coding RNAs (lncRNAs) are a class of noncoding transcripts greater than 200 nucleotides in length. Recent studies have revealed that lncRNAs may affect cancer progression(6, 7). For example, lncRNAs, HOTAIR, MEG3, MALAT-1, H19 and GAPLINC may play a role in carcinogenesis(8-12). The role of lncRNAs and the underlying mechanisms in gastric cancer has been reported before(11-16). However, more specific mechanisms of lncRNAs in gastric cancer initiation and development need to be further teased out.

In the current study, we examined genome-wide lncRNA expression profiles in gastric cancer and paired adjacent tissue, from which we identified and characterized a novel lncRNA (BC041951), which was associated with the prognosis in gastric cancer patients, and therefore designated it as “gastric cancer-associated lncRNA 1 (GCInc1)”. The biological roles of GCInc1 in gastric carcinogenesis were genetically assessed in several *in vitro* and *in vivo* models. Mechanistically, mass spectrometry combined with integrative analysis revealed that GCInc1 not only modified genomic

binding of WDR5 and KAT2A, but also specified the pattern of histone modifications on target genes. Furthermore, GCInc1 upregulated the transcription of the superoxide dismutase 2 mitochondrial (SOD2), one of the target genes of the WDR5/KAT2A complex, by acting as a scaffold to recruit the WDR5 and KAT2A complex to the *SOD2* promoter and increasing the H3K4 trimethylation and H3K9 acetylation levels in the *SOD2* promoter region. Thus, our study has identified a novel lncRNA, GCInc1, with a biological, mechanistic, and clinical impact on human gastric cancer.

Results

LncRNA candidate BC041951 is clinically relevant in gastric cancer

The Arraystar Human LncRNA Microarray V2.0 (8×60K, Arraystar) was used to profile the lncRNA expression in 10 paired gastric cancer tissues and paired adjacent tissues. In total, 33,045 lncRNAs and 30,215 coding transcripts were collected from the most authoritative databases such as RefSeq, UCSC Known genes dataset (Known Genes 4), Ensembl 37.59, H-invDB 7.0 and RNAdb 2.0 as described in the manufacturer's instructions, and the raw data can be accessed via GSE50710 (12). In order to study these microarray data and rediscover potential biomarkers not yet mined completely, more stringent filtering criteria (raw signal intensity > 1500, fold change > 2, $P < 0.01$, **Supplementary Fig. S1A**) were used in our investigation. We found that eight candidate lncRNAs were significantly increased in gastric cancer tissues compared to adjacent tissues (**Table S1**). To validate microarray data, we analyzed the eight candidate lncRNAs expression in 165 cases of fresh gastric cancer and adjacent tissues (Cohort 1, **Table S2**). Real-time PCR revealed that four of the eight lncRNAs were significantly increased and two of the eight lncRNAs are significantly decreased in cancer versus adjacent tissues of cohort 1 (**Supplementary**

Fig. S1B). These data indicate that a set of lncRNAs is aberrantly expressed in gastric cancer tissues.

We next analyzed the correlation between the four lncRNA candidates with the clinical outcome in cohort 1. The Kaplan-Meier analyses showed that lncRNA candidates, AK094163, ENST00000430239, and NR_024373 have no predictive value for the clinical outcome of gastric cancer patients while high expression of lncRNA candidate, BC041951, was significantly associated with a poor prognosis in these patients (**Fig. 1A and Supplementary Fig. S1C**). Furthermore, real-time PCR showed that only BC041951 expression gradually increased from normal gastric tissue to intestinal metaplasia (IM), to dysplasia, and to gastric cancer (**Fig. 1B and Supplementary Fig. S1D**). The expression of BC041951 was significantly increased in gastric cancer tissues, compared with normal tissues in patients enrolled in Renji Hospital (**Supplementary Fig. S1E**) and in an external gastric cancer dataset from the GEO database repository (**Supplementary Fig. S1F**). In laser capture microdissection (LCM) assays, BC041951 was mainly expressed and was significantly increased in the epithelial cells of gastric cancer tissues, compared with that in normal gastric epithelial cells (**Supplementary Fig. S1G**) and tumor stromal cells (**Supplementary Fig. S1H**). Thus, an increase in BC041951 may be an early event in the multistep progression of gastric carcinogenesis.

To evaluate the pathological and clinical value of the four increased lncRNA candidates with the American Joint Committee on Cancer (AJCC) stage for prognosis, we computed their accuracy by prediction error curves. Prediction error over time was calculated using the Brier Score. The prediction error was the lowest using the combination of BC041951 and AJCC stage in cohort 1 (**Fig. 1C**), but not with the other three lncRNA candidates (**Supplementary Fig. S1I**). Further ROC analysis

showed that the area under curve (AUC) of the combination of BC041951-based prediction and AJCC-based model (0.732) was higher than the AJCC-based model alone (0.695) (**Supplementary Fig. S1J**). The data indicate that the combination of BC041951 and AJCC stage is more precise in predicting clinical outcome than AJCC stage alone. Therefore, we focused our research on BC041951, and henceforth named this lncRNA candidate as “gastric cancer-associated lncRNA1 (GClnc1)”.

We next evaluated and compared GClnc1 expression with different clinicopathological features in cohort 1. We found that the GClnc1 expression positively correlated with pathological differentiation, vascular invasion, tumor size, and AJCC stage (**Fig. 1D**). Univariate and multivariate regression analyses of cohort 1 demonstrated that GClnc1 expression was an independent predictor of gastric cancer aggressiveness with significant hazard ratios for predicting clinical outcome. Its predictive value was comparable to that of the AJCC stage (**Supplementary Fig. S1K, L and Fig. 1E**).

To further validate the pathological and clinical significance of GClnc1 expression in gastric cancer, we detected and compared GClnc1 expression by *in situ* (ISH) hybridization in an additional 105 paraffin-embedded gastric cancer and adjacent tissues (cohort 2) (**Table S3**). GClnc1 expression was higher in gastric cancer tissues than adjacent tissues (**Fig. 1F**). Consistent with the results in cohort 1, high levels of GClnc1 expression were significantly associated with poor survival (**Fig. 1G**) in univariate and multivariate regression analyses (**Supplementary Fig. S1M, N and Fig. 1H**). Since *H. pylori* infections may be detected in 78% of all cases of gastric cancer patients(17), we measured and found similar levels of GClnc1 expression in *H. pylori*-positive and -negative tumor tissues (**Supplementary Figure 10**).

In order to determine whether GClnc1 is a novel lncRNA, we first analyzed its

sequence. Northern blot revealed that the size of GClnc1 was ~2kb in length in gastric cancer cell lines and gastric cancer tissues (**Supplementary Fig. S1P, Q**). The 5' rapid amplification of cDNA ends (RACE)-PCR (**Supplementary Fig. S1R**) and the ChIP assay with anti-RNA Pol II (**Supplementary Fig. S1S**) were performed to identify the 5' ends and the transcription start site (TSS) of GClnc1. Sequencing of PCR products revealed the boundary between the universal anchor primer and GClnc1 (**Supplementary Fig. S1R**). ChIP assay demonstrated that RNA Pol II bound to the predicted bind site of the 5' ends of GClnc1 (**Supplementary Fig. S1S**). Thus, GClnc1 is a RNA Pol II transcript. Analysis of the GClnc1 sequences by ORF Finder from the National Center for Biotechnology Information failed to predict a protein (**Supplementary Fig. S1T**). GClnc1 did not contain a valid Kozak sequence. It was identified to be an lncRNA rather than a protein-coding transcript by CNCI software(18). We performed a codon substitution frequency analysis using PhyloCSF(19). GClnc1 had a very low codon substitution frequency score (-838.1) (**Supplementary Fig. S1U**), indicating that it was unlikely to encode any protein product. We confirmed that GClnc1 was a non-coding RNA by *in vitro* translation analysis (**Supplementary Fig. S1V**). We also have separated the nuclear and cytoplasm fractions of BGC823 and MKN45 gastric cancer cells and performed real-time PCR. We found that GClnc1 was mainly located in the nucleus of BGC823 and MKN45 cancer cells (**Supplementary Fig. S1W**). Collectively, GClnc1 is a novel lncRNA and highly expressed in gastric cancer tissues.

GClnc1 is an oncogenic lncRNA in gastric cancer

To elucidate whether GClnc1 plays a role in gastric cancer tumorigenesis, a RNA-seq analysis was performed to compare the gene expression profiles of GClnc1 shRNA

and control shRNA transfectants. A total of 2050 downregulated genes and 1894 upregulated genes (≥ 2 -fold) were detected (raw data accessible via GEO number: GSE63765) after knockdown of *GCLnc1* in gastric cancer cells (**Table S4**). Gene set enrichment analysis (GSEA) revealed that the gene sets related to Benporath_Proliferation (cell proliferation), Provenzani_Metastasis_Up (metastasis), Kim_Gastric_Cancer_Chemosensitivity (upregulated genes in chemo-resistant gastric cancer cells) and Vecchi_Gastric_Cancer_Early_Up (gastric cancer-specific signature) negatively correlated with *GCLnc1* downregulation in gastric cancer cells (**Fig. 2A-D**). The top-scoring genes recurring in the four gene sets included key cancer genes, *c-MYC* and *CDK1*. Real-time PCR confirmed that alteration of *GCLnc1* expression dramatically affected the key tumorigenesis gene signatures (**Supplementary Fig. S2A**). To further gain insight into the biological pathways involved in gastric cancer pathogenesis, based on the median of *GCLnc1* expression levels, we performed the GSEA analysis in an independent public dataset from Gene Expression Omnibus (GSE27342) (**Supplementary Fig. S2B**). Enrichment plots of GSEA showed that the gene signatures of cell proliferation, cell cycle, cancer-related pathways, metastasis, and chemotherapy resistance pathways were enriched in patients with high *GCLnc1* expression, but not in patients with low *GCLnc1* expression (**Fig. 2E**). These data suggest that *GCLnc1* may be an important modulator in gastric tumorigenesis.

To functionally validate the pathway findings, we transfected *GCLnc1* siRNAs into the gastric cancer cell lines, BGC823 and MKN45. The two cell lines expressed higher levels of *GCLnc1* (**Supplementary Fig. S3A**) and exhibited higher proliferation (**Supplementary Fig. S3B**) and invasion capacities (**Supplementary Fig. S3C**) compared to normal GES-1 gastric cells (**Supplementary Fig. S3A-C**). The results showed that knockdown of *GCLnc1*, but not lncRNA ENST00000430239

(ENST430239), significantly impaired gastric cancer cell proliferation and colony formation in BGC823 cells and MKN45 cells (**Fig. 3A** and **Supplementary Fig. S3D-F**). Knockdown of GClnc1 dramatically reduced BGC823 and MKN45 tumor growth (**Fig. 3B, C** and **Supplementary Fig. S3G, H**) and tumor weight (**Fig. 3D** and **Supplementary Fig. S3I**) in xenograft mouse tumor models. In support of the pro-tumor role of GClnc1, Ki67 staining revealed that downregulation of GClnc1 decreased tumor cell proliferation *in vivo* (**Supplementary Fig. S3J, K**). Furthermore, knockdown of GClnc1 sensitized BGC823 and MKN45 cells to Fluorouracil and Cisplatin treatment (**Fig. 3E, F** and **Supplementary Fig. S3L, M**). The data suggest that GClnc1 may be an oncogenic lncRNA in gastric cancer and control gastric cancer cell proliferation and chemo-resistance.

Next, we examined the effects of GClnc1 on gastric cancer cell invasion and metastasis. In the invasion assay, we showed that downregulation of GClnc1, but not lncRNA ENST430239, significantly reduced the invasion ability in gastric cancer cells (**Fig. 3G** and **Supplementary Fig. S3N**). In a gastric cancer metastatic model, the mice inoculated with GClnc1 shRNA-expressing tumor cells had a longer overall survival time than the mice that received control shRNA-expressing tumor cells or Phosphate Buffered Solutions (PBS) (**Fig. 3H** and **Supplementary Fig. S3O**). There were fewer metastatic foci in the lungs of nude mice at 13 weeks after injection of GClnc1 shRNA adenovirus, when compared with control groups (**Fig. 3I** and **Supplementary Fig. S3P**). In the gain-of-function assays, overexpression of GClnc1 increased cell proliferation and invasion ability of GES-1 and MGC803 cells *in vitro* and *in vivo* (**Supplementary Fig. S3Q-V**). The data strongly suggest that GClnc1 may promote gastric cancer progression by regulating gastric cancer cell proliferation and metastasis, and increasing chemotherapy resistance.

GClnc1 interacts with WDR5 and KAT2A epigenetic modification complex

LncRNAs may function by physically interacting with transcriptional factors, histone regulators, and other cellular factors (20-23). To explore the mechanism of GClnc1-mediated carcinogenesis in gastric cancer, we sought to identify intracellular GClnc1-binding factors using an unbiased approach. Biotinylated GClnc1 or antisense GClnc1 RNA (negative control), was incubated with total protein extracts from BGC823 cells and pulled down with streptavidin (**Fig. 4A**). The associated proteins were analyzed by SDS/PAGE and silver staining. There were two specific bands in the GClnc1 pull-down samples (**Fig. 4B**). We excised and analyzed the two bands by mass spectrometry. We identified 94 and 101 potential binding proteins in band one and two, respectively (**Table S5**). Among these binding proteins, we selected 10 proteins, which may likely participate in transcription regulation(24-32), for further binding validation. The 10 genes included eukaryotic translation initiation factor 3 subunit I (EIF3I), Zinc finger Ran-binding domain-containing protein 2 (ZRANB2), eukaryotic translation initiation factor 3 subunit G (EIF3G), WD repeat-containing protein 5(WDR5), methyl-CpG-binding domain protein 3(MBD3), AP-1 complex subunit gamma-1 (AP1G1), nucleolar transcription factor 1(UBTF), histone acetyltransferase KAT2A (KAT2A), eukaryotic translation initiation factor 3 subunit B (EIF3B), and transcription initiation factor TFIID subunit 5(TAF5) proteins. Western blot showed that only WDR5 and KAT2A bound specifically to GClnc1 (**Fig. 4C and Supplementary Fig. S4A**). The data suggest that the two proteins may form a complex to interact with GClnc1. WDR5 is a core subunit of the human mixed-lineage leukemia (MLL) and SET domain containing 1A (SET1, hCOMPASS) histone H3 Lys4 (H3K4) methyltransferase complexes, and an “effector” of H3K4

methylation in gene transactivation(33, 34). KAT2A is a histone acetyltransferase (HAT) that functions primarily as a transcriptional activator, via modulating histone H3 Lys9 (H3K9) acetylation levels in the gene promoter region(35). To identify the WDR5/KAT2A-interacting region of GCIn1, we constructed and biotinylated four fragments of GCIn1 (1–500 Δ 1, 501–1000 Δ 2, 1001–1500 Δ 3, 1501–2155 Δ 4), and used them in the pull-down assay with BGC823 cell lysates. We found that the 5' fragment of GCIn1 mediated the interaction with WDR5 and KAT2A (**Fig. 4D**). To substantiate the observation, anti-WDR5 and anti-KAT2A antibodies were used to immunoprecipitate endogenous WDR5 and KAT2A from nuclear extracts of BGC823 cells. To this end, RNAs bound to WDR5 and KAT2A were extracted and analyzed. PCR data revealed that GCIn1 directly bound with WDR5 and KAT2A in gastric cancer cells (**Fig. 4E**). We also detected ~5.5 fold and ~4 fold enrichments of GCIn1, but not GAPDH RNA, in the anti-WDR5 and anti-KAT2A immunoprecipitates, respectively, compared with the IgG control (**Fig. 4 F, G**). Thus, GCIn1 may specifically bind with WDR5 and KAT2A in gastric cancer cells. Furthermore, we observed that WDR5 and KAT2A interacted with each other in gastric cancer cells (**Fig. 4H**). Downregulation of GCIn1 had no effect on the expression of WDR5 and KAT2A (**Supplementary Fig. S4B**). However, knockdown of GCIn1 or RNAase treatment abolished the interaction between WDR5 and KAT2A (**Fig. 4I**). Thus, GCIn1 is important for WDR5 and KAT2A complex interaction.

GCIn1 coordinates the localization of WDR5 and KAT2A genome-wide

To address whether GCIn1 modulates WDR5 and KAT2A genomic binding genome-wide, we performed ChIP coupled with high-throughput sequencing (ChIP-seq) for WDR5 and KAT2A in BGC823 cells. WDR5 and KAT2A ChIP-Seq

data revealed 5549 and 6485 called peaks, respectively, in BGC823 cells transfected with control shRNA. GCIn1 shRNA virus transfection caused reduced WDR5 and KAT2A occupancies in these two histone modulator binding DNA regions (**Fig. 5A, B**). Further analysis showed that WDR5 and KAT2A occupied 966 and 734 gene promoters, respectively (**Fig. 5C**). Nearly 20% of KAT2A occupied genes (147 genes) were also occupied by WDR5, revealing a significant overlap ($P < 0.01$, hypergeometric distribution). Knockdown of GCIn1 led to a concordant decrease of WDR5 and KAT2A occupancies in 80% of the WDR5 and KAT2A co-occupied gene promoter regions (**Supplementary Fig. S5A, B and Table S6**). The promoter regions of these genes correspondingly decreased the levels of H3K4me3 and H3K9Ac, the respective histone methylation and acetylation products of the WDR5 and KAT2A complex (**Supplementary Fig. S5C**). To further verify these findings, we performed the ChIP for WDR5, KAT2A, H3K4me3, and H3K9Ac in BGC823 and MKN45 cells transfected with control shRNA or GCIn1 shRNA virus. We randomly selected 10 % of the overlapped genes (147 genes) to experimentally validate the effect of GCIn1-mediated WDR5/KAT2A complex genomic binding. We found that knockdown of GCIn1 resulted in a decrease in the binding efficiency of WDR5 and KAT2A and the levels of H3K4me3/H3K9Ac in ~80% of the selected binding genes both in BGC823 and MKN45 cells (**Supplementary Fig. S5D, E**). Moreover, overexpression of GCIn1 dramatically increased the binding efficiency of WDR5 and KAT2A and the levels of H3K4me3/H3K9Ac of the same measured genes in normal GES-1 cells (**Supplementary Fig. S5F**).

Next, we used the RNA-seq data from BGC823 cells transfected with control or GCIn1 shRNA virus (GSE63765) to characterize the relationship between WDR5/KAT2A complex binding and GCIn1-mediated changes in the gene

expression. After identifying a gene signature with highly significant changes in expression (**Table S6D**), we intersected this signature with the ChIP-seq data. We observed that a substantial subset of genes with relative decrease in WDR5/KAT2A complex genomic binding were dysregulated after knockdown of GCIn1 (**Fig. 5D and Table S6E**). Integrative GSEA of the RNA seq and WDR5/KAT2A ChIP-seq data demonstrated significant enrichment for genes that were repressed when GCIn1 was downregulated (**Fig. 5E**). Thus, knockdown of GCIn1 impairs the WDR5 and KAT2A complex formation. This suggests that GCIn1 may act as a scaffold for this complex, thereby controlling its ability to regulate histone modification, gene expression, and biological activity.

GCIn1 promotes gastric cancer progression via SOD2

According to the ChIP-seq and RNA-seq data, superoxide dismutase 2 mitochondrial (SOD2) is one of the target genes of the WDR5/KAT2A complex (**Table S6E**). Furthermore, we found the strongest association between *SOD2* and GCIn1 expression (**Table S7 and Fig. 6A**) among 12 genes. The positive correlation between *SOD2* and GCIn1 was also validated in the 165 cases of gastric cancer in our cohort 1 (**Fig. 6B and Supplementary Fig. S6A**).

GCIn1 is located in chromosome 6 (chr6:159679116-159681261) with a 2155nt length. A 195nt length DNA fragment of GCIn1 at the 3' terminal overlaps with the last exon of the two transcript variants of the *SOD2* gene (NM_001024465, NM_001024466). The other part of GCIn1 is completely overlapped with the last intron of the *SOD2* gene (**Fig. 6C**). We next designed the sense-specific PCR primers (**Supplementary Fig. S6B**) to detect whether GCIn1 and *SOD2* were different

transcripts. If *GClnc1* and *SOD2* were different transcripts, the PCR products of primer 1 + primer 2 should be approximately 150 bp (part of the *GClnc1* fragment) and the PCR products of primer 4 + primer 2 should be approximately 200 bp (part of the *SOD2* fragment). In addition, there should be no PCR product to be detected by primer 3 + primer 4. If *GClnc1* and *SOD2* were the same transcripts, in addition to the 150 bp and 200 bp PCR products, there should be an additional PCR product (around 3260 bp) to be detected by primer 3 and primer 4. Our results showed that the 150 bp and 200 bp PCR products were detectable; however, the 3260 bp PCR product was not detectable (**Supplementary Fig. S6C**). Furthermore, Northern blot analysis confirmed the sense-specific RT PCR data (**Supplementary Fig. S6D**). Thus, *GClnc1* and *SOD2* are different transcripts. MicroRNA is not found in the *GClnc1* coding region. We next explored the mechanism by which *GClnc1* regulates the expression of *SOD2*. Real-time PCR and Western blot data showed that *SOD2* expression was significantly decreased in BGC823 cells transfected with two different *GClnc1* siRNAs, but not with lncRNA ENST430239 (**Fig. 6D**). Similar results were observed in MKN45 cells (**Supplementary Fig. S6E**). Real-time PCR and Western blot data (**Supplementary Fig. S6F**) showed that *SOD2* expression was significantly increased in *GClnc1* overexpressing GES-1 cells. We next examined the effect of *GClnc1* on the transcriptional activity of the *SOD2* gene. Luciferase assay revealed that knock down of *GClnc1*, not lncRNA ENST430239, impaired the transcriptional level of the *SOD2* promoter in BGC823 cells (**Fig. 6E**) and in MKN45 cells (**Supplementary Fig. S6G**). The data suggest that *GClnc1* may positively regulate *SOD2* transcription in gastric cancer cells. It has been reported that high levels of *SOD2* expression correlated with poor clinical outcome in gastric cancer(36), and *SOD2* may regulate cancer cell proliferation and invasion(37, 38). In addition, gastric cancer cell proliferation and

invasion ability were compared after knockdown GClnc1 and SOD2. We found that downregulation of SOD2 and GClnc1 similarly impaired cell proliferation and invasion (**Supplementary Fig. S6H**).

We next hypothesized that SOD2 mediated the biological function of GClnc1 in gastric cancer. To test this hypothesis, we transfected SOD2 siRNA into gastric cancer cells and examined its effects on cancer cell biological function. Knockdown of SOD2 expression significantly reduced gastric cancer cell proliferation induced by GClnc-1 in BGC823 cells (**Fig. 6F**) and MKN45 cells (**Supplementary Fig. S6I**). We next used the CRISPR-Cas9 system to knock out the *SOD2* gene in BGC823 and MKN45 cells. We obtained similar results in these cells (**Supplementary Fig. S6J**). Furthermore, inhibition of SOD2 expression significantly blocked tumor growth in GClnc1-overexpressing xenograft mice (**Supplementary Fig. S6K-M**). Downregulation of SOD2 also markedly blocked the GClnc1-induced cell invasion in BGC823 cells (**Fig. 6G and Supplementary Fig. S6N**) and MKN45 cells (**Supplementary Fig. S6O, P**). Thus, SOD2 mediates the regulatory function of GClnc1 in gastric cancer cells.

GClnc1 is a molecular link among WDR5/KAT2A and SOD2

H3K4 methylation and H3K9 acetylation may regulate the transcription of the *SOD2* gene promoter(39). Real-time PCR and Western blot showed that knockdown of WDR5 or KAT2A significantly blocked GClnc1-induced SOD2 upregulation in BGC823 cells (**Fig. 7A and Supplementary Fig. S7A**). The data suggest that WDR5, KAT2A, and GClnc1 may cooperate to regulate *SOD2* gene transactivation via histone modification. The ChIP assay indicated that WDR5 and KAT2A directly

bound to the promoter region of *SOD2* (**Fig. 7B**), and knockdown of WDR5, KAT2A, or GCInc1 significantly decreased the binding efficiency of WDR5 and KAT2A, respectively (**Fig. 7C**). The data indicate that GCInc1 may play an important role in the binding progression of WDR5 or KAT2A to the *SOD2* gene promoter region. Furthermore, we observed that H3K4 trimethylation and H3K9 acetylation existed in the same promoter region of *SOD2* (**Fig. 7D**). Knockdown of GCInc1 significantly decreased the H3K9 acetylation and H3K4 methylation levels in the *SOD2* promoter region, but not the H3K27 trimethylation level (**Fig. 7E**). Overexpression of GCInc1 dramatically increased the H3K9 acetylation and H3K4 methylation levels at the promoter region of the *SOD2* gene (**Supplementary Fig. S7B**). The ChIRP assay revealed that GCInc1 directly bound to the *SOD2* promoter region (**Fig. 7F**). The combined analyses strongly suggest that GCInc1 may act as a scaffold lncRNA to recruit WDR5 and KAT2A to the promoter of the *SOD2* gene and transactivates the transcription of the *SOD2* gene.

Discussion

Multiple oncogenic pathways may contribute to gastric cancer carcinogenesis(40, 41), however, the potential involvement of lncRNA(s) is poorly defined in human gastric cancer. Through a combination of genomic, biochemical, and cell biological analyses, we have demonstrated that GClnc1 is a novel oncogenic lncRNA in gastric cancer. GSEA analyses have demonstrated that cell proliferation, metastasis, and chemo-resistance pathways in cancer are significantly enriched in response to GClnc1 alteration in the gastric cancer patients' datasets. The bioinformatics analyses have been functionally validated in several *in vitro* and *in vivo* experimental models. In cultured GC cells and xenograft mouse models, downregulation of GClnc1 markedly suppresses cell growth, metastasis, and increases gastric cancer cell chemosensitivity. The data consistently point to the notion that high GClnc1 expression is a decisive factor of controlling human gastric cancer aggressiveness.

lncRNAs may participate in chromatin remodeling complexes(42-46), however, the underlying molecular mechanisms remain unknown. Our mass spectrometry and RNA pull-down data have demonstrated that GClnc1 directly interacts with the key component of histone methyltransferase complex, WDR5, and the histone acetylase, KAT2A. Furthermore, GClnc1 can link WDR5 and KAT2A by acting as a potential modular scaffold in GC cells. This notion is supported by 3 lines of experimental evidence. (i) GClnc1 directly binds with WDR5 and KAT2A via its 5' domain; (ii) Genetic deficiency of GClnc1 or RNase treatment of the immunoprecipitation products abrogates the interaction between WDR5 and KAT2A; (iii) Knockdown of GClnc1 significantly disrupts the binding of the WDR5/KAT2A complex to the promoter regions of WDR5/KAT2A co-occupied genes. In support of our observation,

lncRNAs, HOTAIR, and HOTTIP, assist epigenetic complexes including PRC2 and MLL, facilitate their genomic binding and enhance their functions(20, 21, 46). In short, GCInc1 is the lncRNA capable of modulating the interaction of the WDR5 and KAT2A complex and further altering the pattern of histone modification and transactivating target genes in gastric cancer cells (**Fig. 7G**).

When we explored the mechanisms by which GCInc1 contributes to gastric carcinogenesis, we found the involvement of SOD2. It has been reported that SOD2 expression is increased in gastric cancers (36). The expression of the *SOD2* gene can be epigenetically silenced in human breast cancer cells. H3K4 methylation and H3K9 acetylation may be the major histone regulatory pattern in *SOD2* gene transcription(39). Consistent with these studies, we show that GCInc1 activates *SOD2* transcription, and SOD2 is functionally responsible for GCInc1-mediated gastric carcinogenesis. Notably, as GCInc1 may be directly and indirectly linked to gene regulatory networks, in addition to SOD2, we do not rule out the involvement of other genes in GCInc1 associated biological function.

Given the clinical, genetic, biochemical, and functional significance of GCInc1 in gastric cancer, we conclude that GCInc1 and its associated pathway is crucial for gastric carcinogenesis, and targeting this pathway may be pivotal in the prevention or treatment of gastric cancer.

Methods

Patient Specimens

Tumor and the adjacent gastric specimens were obtained from GC patients who underwent surgery at Shanghai Renji Hospital from January 2000 to January 2005. The study protocol was approved by the ethics committee of Shanghai Jiao Tong University School of Medicine, Renji Hospital. Written informed consents were obtained from all participants in this study. All the research was carried out in accordance with the provisions of the Helsinki Declaration of 1975. None of these patients had received radiotherapy or chemotherapy prior to surgery. The percentage of tumor cellularity in the tissue section was at least 70% as determined by a pathologist.

Human gastric mucosal tissues (normal tissues, tissues diagnosed with IM or DYS) were collected from patients by gastroscopy inspection in Renji Hospital with written informed consent. None of the patients had taken nonsteroidal anti-inflammatory drugs, H₂ receptor antagonists, proton pump inhibitors, antimicrobials, or bismuth compounds 4 weeks prior to the study. We used 20 samples per group to ensure adequate statistical power to distinguish phenotypes between the groups. The different extent of inflammation in these tissues was examined according to the updated Sydney System (International Workshop on the Histopathology of Gastritis, Houston 1994).

Bioinformatics Analysis

LncRNAs were identified from the gene expression data which can be accessed by GEO Series accession number GSE50710. The RNA sequence and CHIP sequence

data can be accessed by GEO Series accession number GSE63765. The detailed Bioinformatics analyses were described in Supplementary Material and Methods.

Cell Culture and Treatment

The human gastric epithelial cell line, GES-1, and gastric cancer cell lines, MKN45 and AGS were from Dr. Jun Yu, the Chinese University of Hong Kong. BGC823, SGC7901, and MGC803 cells were from Dr. Shuiping Tu, Rutgers, the State University of New Jersey. All cell lines were genotyped for identity by Beijing Microread Genetics Co.,Ltd and tested routinely for *Mycoplasma* contamination. Gastric cells were cultured in RPMI-1640 medium (Gibco, USA) supplemented with 10% fetal bovine serum (FBS) at 37°C in an atmosphere of 5% CO₂. Gastric cells were transfected with different siRNA and GCIn1 expressing plasmids for genetic and functional assays. The detailed cell treatment was described in Supplementary Material and Methods.

Sense-specific RT-PCR

The sense-specific PCR (Supplementary Figure S6B) was developed after pilot and optimizing experiments. The detailed Sense-specific RT-PCR was described in Supplementary Material and Methods.

5' Rapid Amplification of cDNA Ends (RACE)

We used the 5'-RACE analyses to determine the transcriptional initiation site of GCIn1 using a SMARTer™ RACE cDNA Amplification Kit (Clontech, Palo Alto,

CA). The gene-specific primer used for the PCR of the RACE analysis was 5'-TATGTACACAGTGTTGCTCACC-3' (5'RACE).

Cell Proliferation and Transwell Assay

Cell proliferation was assessed by the BrdU incorporation assay (Roche Molecular Biochemicals, Mannheim, Germany). Briefly, control and treated gastric cancer cells were seeded onto the 96-well plates at an initial density of 5×10^3 cells per well. BrdU labeling solution (10 μ l/well) was added to the cells at specified time points. After incubating for 2 hours, the reaction product was quantified according to manufacturer's instructions.

Transwell assays were performed using 24-well transwells (8- μ m pore size; Millipore) precoated with Matrigel (BD Biosciences, USA). Cells at logarithmic phase were transfected with GClnc1 siRNA or control siRNA. Transfected cells were then harvested, and 1×10^5 cells were seeded in serum-free medium into the upper chamber, whereas medium supplemented with 20% FBS was applied to the lower chamber as a chemoattractant. After 48 hours incubation, the migrated cells at the bottom surface of the filter were fixed, stained, and counted.

Hybridization *In Situ* and Immunohistochemical Staining

The *in situ* detection of lncRNA GClnc1 was performed on 6- μ m formalin-fixed, paraffin-embedded (FFPE) sections using DIG-labeled miRCURYTM Detection probe (Exiqon, Denmark). Positive controls (RNU6B, Exiqon) and scrambled control RNAs were included for each hybridization procedure and analyzed using a Nikon 80i microscope with Nikon NIS-Elements F 2.3 software (Nikon, Japan).

The expression of Ki67 was examined with primary antibodies (Ki67:1:100) using the LSAB+ kit (DakoCytomation, Copenhagen, Denmark). The tissue slides were examined independently by two investigators. Protein expression was quantified using a visual grading system based on the extent of staining (the percentage of positive tumor cells on a scale of 0–4: 0, none; 1, 1–25%; 2, 26–50%; 3, 51–75%; 4, >75%) and the intensity of staining (graded on a scale of 0–3: 0, no staining; 1, weak staining; 2, moderate staining; 3, strong staining). The product of the extent and intensity grades was used to define the Ki67 protein expression.

Luciferase Assay

Gastric cancer cells were transfected with the plasmids expressing the designated combinations of pGL3-SOD2PWT and other relevant siRNAs at 1.0 µg and 100 ng of phRL (Renilla luciferase) with Lipofectamine 2000 (Invitrogen) or DharmaFECT 1 siRNA transfection reagents (Dharmacon). Twenty-four hours after transfection, the cells were collected to detect luciferase activity using the Dual-Luciferase reporter assay system (Promega, USA). Luciferase activity was measured by using a BD Monolight 3010 luminometer (BD Biosciences, USA). Transfection efficiency was normalized by dividing the luciferase activity of the construct to the corresponding Renilla luciferase activity.

RNA Analysis, Extraction, and Quantitative Real-time PCR

The RNA expression levels were measured using a real-time quantitative PCR system. Total RNA was extracted by TRIzol reagent (Invitrogen, USA), and 1 µg of total RNA was reverse-transcribed using the PrimeScriptTM RT Reagent Kit (Perfect

Real Time; Takara, Shiga, Japan). The amplified transcript level of each specific gene was normalized to that of *18S*. The primers (**Table S8**) were provided by Sheng gong Company, Shanghai.

Western Blot

Western blot was performed using standard techniques(47). Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) antibody was used as a control for whole cell lysates. Antibodies were purchased from Cell Signaling Technology Inc. (Beverly, MA, USA), Abcam (Cambridge, MA, USA) and Sigma (St. Louis, MO, USA).

Northern Blot

Northern blot was performed with 300–600 ng of purified poly (A) mRNA. RNAs were resolved by denaturing agarose gel electrophoresis (Ambion, USA), and transferred to Hybond-XL membranes (GE Healthcare, USA). LncRNAs were detected using ³²P-labeled DNA probes.

***In Vitro* Transcription and Translation**

LncRNA GCInc1 and lacZ were cloned into pBluescript KSII downstream of the T7 promoter. The plasmid was linearized by single digestion with *KpnI*. 250 ng of linear pCDNA3.1-GCInc1 was transcribed *in vitro* (Epicentre, USA). 700 ng of purified GCInc1 RNA were translated *in vitro* using Biotinylated leucine tRNA (Promega, USA). Biotinylated proteins were detected using a BrightStar BioDetect Kit (Ambion, USA). LacZ mRNA was used as a positive translation control, and mock-translated samples (no RNA template) were used as negative controls.

RNA Pull-Down Assay

Cell nuclear lysates were incubated with biotinylated RNA and streptavidin beads for RNA pull down incubation. Then beads were collected by centrifugation. RNA-associated proteins were eluted and resolved by SDS/PAGE followed by silver staining (Bio-Rad, USA). The detailed RNA Pull-Down Assay is described in Supplementary Material and Methods.

Liquid Chromatography-Mass Spectrometry (LC-MS) Analysis, Database search, and Protein Identification

To identify specific GCLnc1 interactors, GCLnc1 and Antisense GCLnc1 pulled-down eluates were compared. The bands that were predominantly represented in the GCLnc1 pulled-down sample were chosen. The bands were excised to perform LC-MS analysis. The proteins sequenced with at least two peptides were considered a reliable identification. The detailed LC-MS, database and protein identification are described in Supplementary Material and Methods.

RNA Immunoprecipitation

RIP experiments were performed using the Megna RIP RNA-binding Protein Immunoprecipitation Kit (Millipore, USA). The WDR5/KAT2A antibody used for RIP was purchased from Abcam (Cell signaling Technology, USA). The co-precipitated RNAs were detected by reverse transcription PCR. To demonstrate that the detected RNA signals specifically bind to WDR5 or KAT2A, total RNA (input controls) and normal mouse IgG controls were simultaneously assayed. The gene-specific primers used for detecting GCLnc1 are displayed in **Table S8**.

Co-immunoprecipitation (Co-IP)

Co-immunoprecipitation was performed as described previously(48). Both the input and IP samples were analyzed by Western blot using various antibodies at the following dilutions: WDR5 Antibody (1:1000) (Cell Signal Technology, USA), KAT2A Antibody (1:1000) (Cell Signal Technology, USA), and normal rabbit IgG (Upstate, USA).

Chromatin Isolation by RNA Purification (CHIRP)

A total of GClnc1 asDNA(antisense DNA), lacZ asDNA,(antisense DNA), and GClnc1 sDNA(sense DNA) probes were designed using the online probe designer at singlemoleculefish.com. Oligonucleotides were biotinylated at the 3' end with an 18-carbon spacer arm. BGC823 cells were collected and subjected to chromatin isolation by RNA purification (ChIRP) using the method described by Chu et al(49).

***In Vivo* Experiments**

In order to clarify the effect of GClnc1 on tumor growth and invasion *in vivo*, 4-week-old male BALB/c nude mice (Experimental Animal Centre of SIBS) were used. Different numbers of gastric cancer cells were inoculated into the mice. Ten days after subcutaneous inoculation, mice were treated with different virus. Tumor growth and invasion were monitored. All experimental procedures were approved by the Institutional Animal Care and Use Committee of Ren Ji Hospital, School of Medicine, Shanghai Jiao Tong University. The detailed *in vivo* experiments are described in Supplementary Material and Methods.

Adenovirus and Plasmids Construction

The control shRNA, GCInc1 shRNA1/2, SOD2 shRNA, GCInc1 overexpressing adenovirus, and all plasmids were constructed by Shanghai Obio Technology Company, Shanghai, China.

Chromatin Immunoprecipitation (ChIP) and High-Throughput Sequencing

Chromatin was sonicated and immunoprecipitated with different antibodies. Real-time PCR was performed in triplicates. Each PCR assay was carried out in a 20 μ l reaction volume by using the eluted immunoprecipitated DNA. The amount of genomic DNA co-precipitated with the specific antibody was calculated as follows: $CB_{TB} = CB_{TB}(\text{genomic input}) - CB_{TB}(\text{specific antibody})$, where $CB_{TB}(\text{genomic input})$ and $CB_{TB}(\text{specific antibody})$ were the mean threshold cycles of PCR performed in triplicates on DNA samples from the genomic input samples and the specific antibody samples, respectively. The forward and reverse primers were listed in the supplementary table (**Supplementary Table S9**). High-Throughput RNA and ChIP sequencing was performed after knockdown of GCInc1 in BGC823 cells. The detailed ChIP and High-Throughput Sequencing are described in Supplementary Material and Methods.

Drug Sensitivity Assays to Fluorouracil and Cisplatin

Cell proliferation was assessed by the BrdU incorporation assay (Roche Molecular Biochemicals, Mannheim, Germany). Briefly, control and siRNA-transfection gastric cancer cells, which were seeded onto the 96-well plates at an initial density of 5×10^3 cells per well, were treated with different doses of Fluorouracil and Cisplatin, BrdU labeling solution (10 μ l/well) was added to the cells at specified time points. After 2

hours incubation, culture medium was removed and the cells were fixed. Then DNA was denatured by adding FixDenat (200 μ l/well) and anti-BrdU-POD working solution (100 μ l/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: approx. 492 nm).

Tissues and Laser-Capture Microdissection.

Gastric cancer tissues and the adjacent tissues were obtained from 20 patients enrolled in Renji Hospital. Tissues were embedded in TissuesTek OCT medium (Sakura, Tokyo, Japan) and snap frozen on dry ice. The frozen sections were fixed in 70% ethanol for 30 second and stained with H&E, followed by three dehydration steps of 5 seconds each in 70%, 95%, and 99.5% ethanol and a final 5-min dehydration in xylene. Once air-dried, the stained tissues were laser-capture microdissected by a PixCell II LCM system (Arcturus Engineering, Mountain View, CA). Normal gastric epithelial cells, gastric cancer cells, and stroma cells were selectively captured by microscopic visualization.

Generation of *SOD2* Knockout BGC823 (BGC823^{KO}) and (MKN45^{KO}) Cell Lines

LentiCRISPR (Addgene plasmid #49535) is a lentiviral vector expressing cas9 and single guide RNA (sgRNA)(50). Genome engineering experiments using CRISPR-Cas9 systems were performed as described previously(51). Briefly, three sgRNAs targeting *SOD2* were cloned into the lentiCRISPR vector. Lentivirus was

produced in HEK 293T cells. BGC823 and MKN45 cells were transduced with lentivirus and selected with 2 µg/ml puromycin. The *SOD2* knock out cell clones were isolated by limited serial dilution. *SOD2* deletion was confirmed by sequencing and immunoblotting analysis. The sgRNA sequences targeting *SOD2* were as follows: 5'-TTCCAGGGCGCCGTAGTCGT-3' (site 1); 5'-AAGCTGACGGCTGCATCTGT-3' (site 2) and 5'-CTTGGTTTCAATAAGGAACG-3' (site 3) (50).

Statistical Analysis

The detailed statistical analysis is described in Supplementary Material and Methods.

Author Contributions

T.T.S and J.H performed the experiments and analyzed data. Q.L, L.L.R and T.T.Y performed the *in vivo* experiments. T.C.Y, J.Y.T, Y.J.B and Y.H collected clinical specimens. Y.X.C collected the data of the clinical specimens. H.Y.C performed Bioinformatics analyses. J.H, H.Y.C, J.Y.F and W.Z designed the research and wrote the manuscript.

Grant Support

This project was supported (in part) by grants from National Natural Science Foundation of China (81421001, 81320108024, 31371420, 91129724, 31271366, 31371273, 81522008); the Shanghai Natural Science Foundation (Grant No. 13ZR14244000); the Program for Professor of Special Appointment (Eastern Scholar No. 201268 and 2015 Youth Eastern Scholar NO.QD2015003) at Shanghai

Institutions of Higher Learning ; Shanghai Municipal Education
Commission—Gaofeng Clinical Medicine Grant Support (No.20151512) and
Chenxing Project of Shanghai Jiao-Tong University to HYC and JH.

References

1. McLean MH, El-Omar EM. Genetics of gastric cancer. *Nat Rev Gastroenterol Hepatol*. 2014;11:664-74.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65:87-108.
3. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van de Velde CJ, Nicolson M, et al. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med*. 2006;355:11-20.
4. Ueda T, Volinia S, Okumura H, Shimizu M, Taccioli C, Rossi S, et al. Relation between microRNA expression and progression and prognosis of gastric cancer: a microRNA expression analysis. *Lancet Oncol*. 2010;11:136-46.
5. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. 2014;513:202-9.
6. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet*. 2009;10:155-9.
7. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell*. 2009;136:629-41.
8. Kogo R, Shimamura T, Mimori K, Kawahara K, Imoto S, Sudo T, et al. Long noncoding RNA HOTAIR regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res*. 2011;71:6320-6.
9. Zhang X, Gejman R, Mahta A, Zhong Y, Rice KA, Zhou Y, et al. Maternally expressed gene 3, an imprinted noncoding RNA gene, is associated with meningioma pathogenesis and progression. *Cancer Res*. 2010;70:2350-8.
10. Lai MC, Yang Z, Zhou L, Zhu QQ, Xie HY, Zhang F, et al. Long non-coding RNA MALAT-1 overexpression predicts tumor recurrence of hepatocellular carcinoma after liver transplantation. *Med Oncol*. 2012;29:1810-6.
11. Yang F, Bi J, Xue X, Zheng L, Zhi K, Hua J, et al. Up-regulated long non-coding RNA H19 contributes to proliferation of gastric cancer cells. *FEBS J*. 2012;279:3159-65.
12. Hu Y, Wang J, Qian J, Kong X, Tang J, Wang Y, et al. Long noncoding RNA GAPLINC regulates CD44-dependent cell invasiveness and associates with poor prognosis of gastric cancer. *Cancer Res*. 2014;74:6890-902.
13. Endo H, Shiroki T, Nakagawa T, Yokoyama M, Tamai K, Yamanami H, et al. Enhanced expression of long non-coding RNA HOTAIR is associated with the development of gastric cancer. *PLoS One*. 2013;8:e77070.
14. Xu TP, Liu XX, Xia R, Yin L, Kong R, Chen WM, et al. SP1-induced upregulation of the long noncoding RNA TINCR regulates cell proliferation and apoptosis by affecting KLF2 mRNA stability in gastric cancer. *Oncogene*. 2015;34:5648-61.
15. Yang F, Xue X, Bi J, Zheng L, Zhi K, Gu Y, et al. Long noncoding RNA CCAT1, which could be activated by c-Myc, promotes the progression of gastric carcinoma. *J Cancer Res Clin Oncol*. 2013;139:437-45.
16. Xu TP, Liu XX, Xia R, Yin L, Kong R, Chen WM, et al. SP1-induced upregulation of the long noncoding RNA TINCR regulates cell proliferation and

- apoptosis by affecting KLF2 mRNA stability in gastric cancer. *Oncogene*. 2015.
17. Sugano K, Tack J, Kuipers EJ, Graham DY, El-Omar EM, Miura S, et al. Kyoto global consensus report on *Helicobacter pylori* gastritis. *Gut*. 2015;64:1353-67.
 18. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 2013;41:e166.
 19. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*. 2011;27:i275-82.
 20. Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010;464:1071-6.
 21. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science*. 2010;329:689-93.
 22. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature*. 2011;470:284-8.
 23. Sun S, Del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT. Jpx RNA activates Xist by evicting CTCF. *Cell*. 2013;153:1537-51.
 24. Spriggs KA, Bushell M, Willis AE. Translational regulation of gene expression during conditions of cell stress. *Mol Cell*. 2010;40:228-37.
 25. Loughlin FE, Mansfield RE, Vaz PM, McGrath AP, Setiyaputra S, Gamsjaeger R, et al. The zinc fingers of the SR-like protein ZRANB2 are single-stranded RNA-binding domains that recognize 5' splice site-like sequences. *Proc Natl Acad Sci U S A*. 2009;106:5581-6.
 26. Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, et al. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*. 2005;121:859-72.
 27. Narlikar GJ, Fan HY, Kingston RE. Cooperation between complexes that regulate chromatin structure and transcription. *Cell*. 2002;108:475-87.
 28. Brown SE, Szyf M. Epigenetic programming of the rRNA promoter by MBD3. *Mol Cell Biol*. 2007;27:4938-52.
 29. Gao H, Dahlman-Wright K. The gene regulatory networks controlled by estrogens. *Mol Cell Endocrinol*. 2011;334:83-90.
 30. Matera AG, Wu W, Imai H, O'Keefe CL, Chan EK. Molecular cloning of the RNA polymerase I transcription factor hUBF/NOR-90 (UBTF) gene and localization to 17q21.3 by fluorescence in situ hybridization and radiation hybrid mapping. *Genomics*. 1997;41:135-8.
 31. van Riggelen J, Yetil A, Felsher DW. MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat Rev Cancer*. 2010;10:301-9.
 32. Huisinga KL, Pugh BF. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in *Saccharomyces cerevisiae*. *Mol Cell*. 2004;13:573-85.
 33. Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, et al. A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling.

Nature. 2006;442:86-90.

34. Ruthenburg AJ, Wang W, Graybosch DM, Li H, Allis CD, Patel DJ, et al. Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. *Nat Struct Mol Biol.* 2006;13:704-12.

35. Tanner KG, Trievel RC, Kuo MH, Howard RM, Berger SL, Allis CD, et al. Catalytic mechanism and function of invariant glutamic acid 173 from the histone acetyltransferase GCN5 transcriptional coactivator. *J Biol Chem.* 1999;274:18157-60.

36. Janssen AM, Bosman CB, van Duijn W, Oostendorp-van de Ruit MM, Kubben FJ, Griffioen G, et al. Superoxide dismutases in gastric and esophageal cancer and the prognostic impact in gastric cancer. *Clin Cancer Res.* 2000;6:3183-92.

37. Nelson KK, Ranganathan AC, Mansouri J, Rodriguez AM, Providence KM, Rutter JL, et al. Elevated sod2 activity augments matrix metalloproteinase expression: evidence for the involvement of endogenous hydrogen peroxide in regulating metastasis. *Clin Cancer Res.* 2003;9:424-32.

38. Chen PM, Wu TC, Wang YC, Cheng YW, Sheu GT, Chen CY, et al. Activation of NF-kappaB by SOD2 promotes the aggressiveness of lung adenocarcinoma by modulating NKX2-1-mediated IKKbeta expression. *Carcinogenesis.* 2013;34:2655-63.

39. Hitchler MJ, Oberley LW, Domann FE. Epigenetic silencing of SOD2 by histone modifications in human breast cancer cells. *Free Radic Biol Med.* 2008;45:1573-80.

40. Ooi CH, Ivanova T, Wu J, Lee M, Tan IB, Tao J, et al. Oncogenic pathway combinations predict clinical prognosis in gastric cancer. *PLoS Genet.* 2009;5:e1000676.

41. Deng N, Goh LK, Wang H, Das K, Tao J, Tan IB, et al. A comprehensive survey of genomic alterations in gastric cancer reveals systematic patterns of molecular exclusivity and co-occurrence among distinct therapeutic targets. *Gut.* 2012;61:673-84.

42. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A.* 2009;106:11667-72.

43. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, et al. Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell.* 2008;32:232-46.

44. Nagano T, Mitchell JA, Sanz LA, Pauler FM, Ferguson-Smith AC, Feil R, et al. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science.* 2008;322:1717-20.

45. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science.* 2008;322:750-6.

46. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell.* 2007;129:1311-23.

47. Xiong H, Hong J, Du W, Lin YW, Ren LL, Wang YC, et al. Roles of STAT3 and ZEB1 proteins in E-cadherin down-regulation and human colorectal cancer epithelial-mesenchymal transition. *J Biol Chem.* 2012;287:5819-32.

48. Uyama N, Zhao L, Van Rossen E, Hirako Y, Reynaert H, Adams DH, et al. Hepatic stellate cells express synemin, a protein bridging intermediate filaments to focal adhesions. *Gut*. 2006;55:1276-89.
49. Chu C, Quinn J, Chang HY. Chromatin isolation by RNA purification (ChIRP). *J Vis Exp*. 2012;61:3912.
50. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*. 2014;343:84-7.
51. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8:2281-308.

Figure Legends

Fig. 1 LncRNA candidate BC041951 is clinically relevant in gastric cancer.

(A) Survival was analyzed and compared between patients with low and high levels of BC041951 in 165 patients with gastric cancer (cohort 1). Log-rank test.

(B) Statistical analysis of lncRNA BC041951 expression in normal, intestinal metaplasia (IM), dysplasia, and cancerous gastric tissues, non-parametric Mann-Whitney test. All the bars represent SE.

(C) Prediction error curves of the different predictors in cohort 1. Apparent error (AE) and 10-fold cross-validated cumulative prediction error (PE), computed for a five year follow-up using Kaplan-Meier estimation as reference.

(D) Comparing different histological stages, positive or negative vascular invasion, tumor size, and AJCC stage between GClnc1 high expression and GClnc1 low expression tumors of the cohort 1. The heatmap illustrates the association of different clinical characters with GClnc1 high and low expression tumors. Statistical significance was performed by Chi-square test. Pathological differentiation (PD).

(E) Multivariable analysis was performed in cohort 1. All the bars correspond to 95% confidence intervals.

(F) Representative images of GClnc1 expression in gastric cancer (GC) and adjacent gastric tissues using *in situ* hybridization (ISH) analysis in cohort 2. $n = 105$.

(G) Survival was analyzed and compared between patients with high and low levels of GClnc1 expression in tumor in cohort 2. $n = 105$, Log-rank test.

(H) Multivariable analysis was performed in cohort 2. All the bars correspond to 95% confidence intervals.

Fig. 2 The correlation between tumorigenic gene sets and GClnc1 expression

levels is assessed via GSEA.

(A-D) Overview of GSEA used to identify the differential gene profiles between sh-GClncl gastric cancer cells and controls.

(E) GSEA comparison between patients with high GClncl expression (red) and low GClncl expression (blue). Median split, $n = 80$ (GSE 27342). Distinct pathways and biological processes were illustrated between the two patient populations. Cytoscape and Enrichment map were used for visualization of the GSEA results (1% FDR, $P = 0.005$). Nodes represent enriched gene sets, which were grouped and annotated by their similarity according to the related gene sets. Enrichment results were mapped as a network of gene sets (nodes). Node size was proportional to the total number of genes within each gene set. Proportion of shared genes between gene sets was presented as the thickness of the green line between nodes. This simplified network map was manually curated by removing general and uninformative sub-networks, as shown in **Supplementary Fig. S2B**. Enrichment plots were shown for a set of activated genes related to cell proliferation, cell cycle, cancer-related pathways, and metastasis and chemotherapy resistance pathways in 80 gastric cancer patients. The enrichment score (ES, green line) means the degree to which the gene set was over-represented at the top or bottom of the ranked list of genes. Black bars indicated the position of genes belonging to the gene set in the ranked list of genes included in the analysis. A positive value indicated more correlation with “high GClncl expression” and a negative value indicated more correlation with “low GClncl expression”.

Fig. 3 GClncl is an oncogenic lncRNA in gastric cancer.

(A) BrdU assay was performed in BGC823 and MKN45 cells transfected with

GClnc1 siRNA1/2. $n = 3$, non-parametric Mann-Whitney test.

(B) Representative data of tumors in nude mice bearing gastric cancer cells transfected with nothing, control shRNA adenovirus, and GClnc1 shRNA1 adenovirus. $n = 8$.

(C) Tumor volume was measured after GClnc1 shRNA1 adenovirus treatments in the xenograft mouse model. $n = 8$, $*P < 0.01$, non-parametric Mann-Whitney test.

(D) Tumor weight was measured in mice after different treatments. $n = 8$, non-parametric Mann-Whitney test.

(E) Dose-response curve of a representative experiment showing relative fluorouracil sensitivity determined by BrdU incorporation. BGC823 cells were transfected with control or GClnc1 siRNA1/2 and were treated with fluorouracil. $n = 3$, non-parametric Mann-Whitney test, compared with control siRNA group.

(F) Dose-response curve of a representative experiment showing relative Cisplatin sensitivity determined by BrdU incorporation. BGC823 cells were transfected with control or GClnc1 siRNA1/2 and treated with Cisplatin. $n = 3$, non-parametric Mann-Whitney test, compared with control siRNA group.

(G) Transwell Matrigel invasion assay was performed in BGC823 cells transfected with control siRNA, GClnc1 siRNA1/2, and lnc ENST430239 siRNA 1/2, respectively. $n = 3$.

(H) Survival analysis was performed in mice bearing gastric cancer transfected with GClnc1 shRNA1 adenovirus, control shRNA adenovirus, or nothing (PBS injection), respectively. $n = 10$, Log-rank test.

(I) Representative hematoxylin-eosin staining and summarized data on tumor lung foci in nude mice at 13 weeks after injection with GClnc1 shRNA1 adenovirus. $n = 10$.

Fig. 4 GCInc1 interacts with WDR5 and KAT2A epigenetic modification complex.

- (A) Experimental design for pull-down assays and identification of GCInc1-associated cellular proteins. GCInc1 and anti-GCInc1 RNA were biotinylated by *in vitro* transcription, refolded, and incubated with BGC823 total cell lysates.
- (B) Silver staining of biotinylated GCInc1-associated proteins. Two GCInc1-specific bands (arrows) were excised and analyzed by mass spectrometry.
- (C) Western blot of the proteins from antisense GCInc1 and GCInc1 pull-down assays. $n = 3$.
- (D) Western blot of WDR5 and KAT2A in samples pulled down by full-length (FL) or truncated GCInc1 ($\Delta 1$: 1–500, $\Delta 2$: 501–1000, $\Delta 3$: 1001–1500, $\Delta 4$: 1501–2155). $n = 3$.
- (E) RNA immunoprecipitation (RIP) experiments were performed using the WDR5 or KAT2A antibody, and specific primers were used to detect GCInc1 or GAPDH. $n = 3$.
- (F, G) Nuclear lysates of BGC823 were immunoprecipitated with anti-WDR5 antibody (F), anti-KAT2A antibody (G) or control IgG. Aliquots of Nuclear lysates (20 % of input) and the WDR5 or KAT2A immunoprecipitates were separated by SDS-PAGE, and the specific immunoprecipitation of WDR5 and KAT2A was confirmed by Western blot. The complexes were analyzed for the presence of GCInc1 or GAPDH by real-time PCR. Signals were normalized to actin mRNA. Results are mean \pm SE of three independent experiments.
- (H) Co-immunoprecipitation detected the interaction of WDR5 and KAT2A in the BGC823 cells. The 20 % of input (cell lysate) and WDR5 or KAT2A

immunoprecipitates were separated by SDS-PAGE. The specific immunoprecipitation of WDR5 and KAT2A was confirmed by Western blot. $n = 3$.

(I) Immunoprecipitation assay was performed to detect the interaction between WDR5 and KAT2A after transfection of GCInc1 siRNA (upper panel) or RNase treatment (lower panel). $n = 3$.

Fig. 5 GCInc1 coordinates the localization of WDR5 and KAT2A genome-wide.

(A) Heatmap of WDR5 global genomic binding at the target sites in BGC823 cells after transduction of control and GCInc1 shRNA adenovirus. A 2-kb interval centered on the called WDR5 peak is shown.

(B) Heatmap of KAT2A global genomic binding at the target sites in BGC823 cells after transduction of control and GCInc1 shRNA adenovirus. A 2-kb interval centered on the called KAT2A peak is shown. To demonstrate accurately the ChIP-seq data, the same control panels (IgG) were used in panels A and B. The antibodies were from rabbit with the same isotype. The DNA samples for WDR5 and KAT2A ChIP experiments were prepared under identical conditions from the same cells with different experimental treatments.

(C) Venn diagram shows the gene promoters occupied by WDR5 (966 genes), KAT2A (734 genes), or both (147 genes).

(D) Heatmap showing the expression change of WDR5/KAT2A co-occupied genes in BGC823 cells after transfection of GCInc1 shRNA and control shRNA. Gene expression is shown as RPKM after normalization.

(E) GSEA data showing the enrichment of ChIP-seq promoter peaks with significant loss of WDR5/KAT2A binding for the downregulated genes in BGC823 cells after GCInc1 shRNA virus infection, compared with control shRNA virus. Normalized

enrichment score (NES); Family wise error rate (FWER).

Fig. 6 GClnc1 promotes gastric cancer progression via SOD2.

(A) GClnc1 subnetwork in the gastric cancer coexpression network. This subnetwork consists of GClnc1 (center) and its 12 direct neighbors. Genes colored in light blue are protein-coding RNAs with unknown function. Genes colored in light red are protein-coding RNAs involved in tumor growth and metastasis. Node size represents the association *P* value between the neighbor gene and GClnc1.

(B) Correlation between *SOD2* and GClnc1 RNA level in 165 human GC tissues.

(C) Schematic representation of exon composition of the *SOD2* gene and the GClnc1 loci on human chromosome 6q25.3.

(D) Real-time PCR and Western blot assays were performed in BGC823 cells after transfection of GClnc1, *SOD2*, and lnc ENST430239 siRNAs. *n* = 3, non-parametric Mann-Whitney test.

(E) Luciferase reporter vectors were generated by inserting the promoter region (-1000 to 0 bp) of the *SOD2* gene. The reporter vectors were then co-transfected into BGC823 GC cells with either GClnc1 siRNA, lncENST430239 siRNA, or control siRNA. Cells were harvested for luciferase activity assay. Results are shown as the mean \pm SE of triplicate determination from three independent experiments. Non-parametric Mann-Whitney test.

(F) Cell proliferation assay was assessed by BrdU incorporation after transfection of the *SOD2* siRNA, the plasmid encoding GClnc1, and control in BGC823 cells. Results are shown as the mean \pm SE of triplicate determination from three independent experiments. Non-parametric Mann-Whitney test.

(G) Transwell Matrigel invasion assay was performed in BGC823 cells after

transfection of the *SOD2* siRNA, the plasmid encoding *GClnc1*, and control in BGC823 cells. $n = 3$, non-parametric Mann-Whitney test.

Fig. 7 *GClnc1* is a molecular link among *WDR5/KAT2A* and *SOD2*.

(A) Real-time PCR assay was performed to measure *SOD2* mRNA level after transfection of *WDR5/KAT2A* siRNA, *GClnc1* expressing plasmid, and control siRNA plus control plasmid in BGC823 cells. Results are shown as the mean \pm SE of triplicate determination from three independent experiments. Non-parametric Mann-Whitney test.

(B) The *SOD2* DNA was detected in the chromatin sample immunoprecipitated from BGC823 cells using an antibody against *WDR5* or *KAT2A*. $n = 3$.

(C) Real-time PCR of the ChIP samples shows the binding efficiency of *WDR5* or *KAT2A* to the *SOD2* gene promoter after transfection of *WDR5*, *KAT2A*, or *GClnc1* siRNA in BGC823 cells, respectively. Results are shown as the mean \pm SE of triplicate determination from three independent experiments.

(D) The *SOD2* DNA was detected in the chromatin sample immunoprecipitated from BGC823 cells using an antibody against H3K4me3 or H3K9ac. $n = 3$.

(E) Real-time PCR of the ChIP samples shows the H3K4 trimethylation or H3K9 acetylation and H3K27 trimethylation levels of the *SOD2* gene promoter after knockdown of *WDR5*, *KAT2A*, or *GClnc1* in BGC823 cells, respectively. Results are shown as the mean \pm SE of triplicate determination from three independent experiments.

(F) ChIRP analysis of *GClnc1* binding to the *SOD2* promoter is shown. ChIRP was performed as previously described(49). Results are expressed as 1. the fold enrichment of *SOD2* promoter fragments in *GClnc1* ChIRP analysis with *GClnc1*

asDNA probe; 2. the fold enrichment of GAPDH promoter fragment in LacZ ChIRP analysis with LacZ asDNA probe; 3. the fold enrichment of *SOD2* promoter fragments in GCInc1 ChIRP analysis with GCInc1 sDNA probe; 4. the fold enrichment of *SOD2* intron2 fragment in GCInc1 ChIRP analysis with GCInc1 asDNA probe. $n = 3$.

(G) A schematic model of lncRNA GCInc1 functions in gastric carcinogenesis. GCInc1 may function as a scaffold lncRNA to recruit WDR5 and KAT2A and transactivate the transcription of target genes, including *SOD2* gene, and finally promote gastric cancer initiation and progression.

Fig. 1

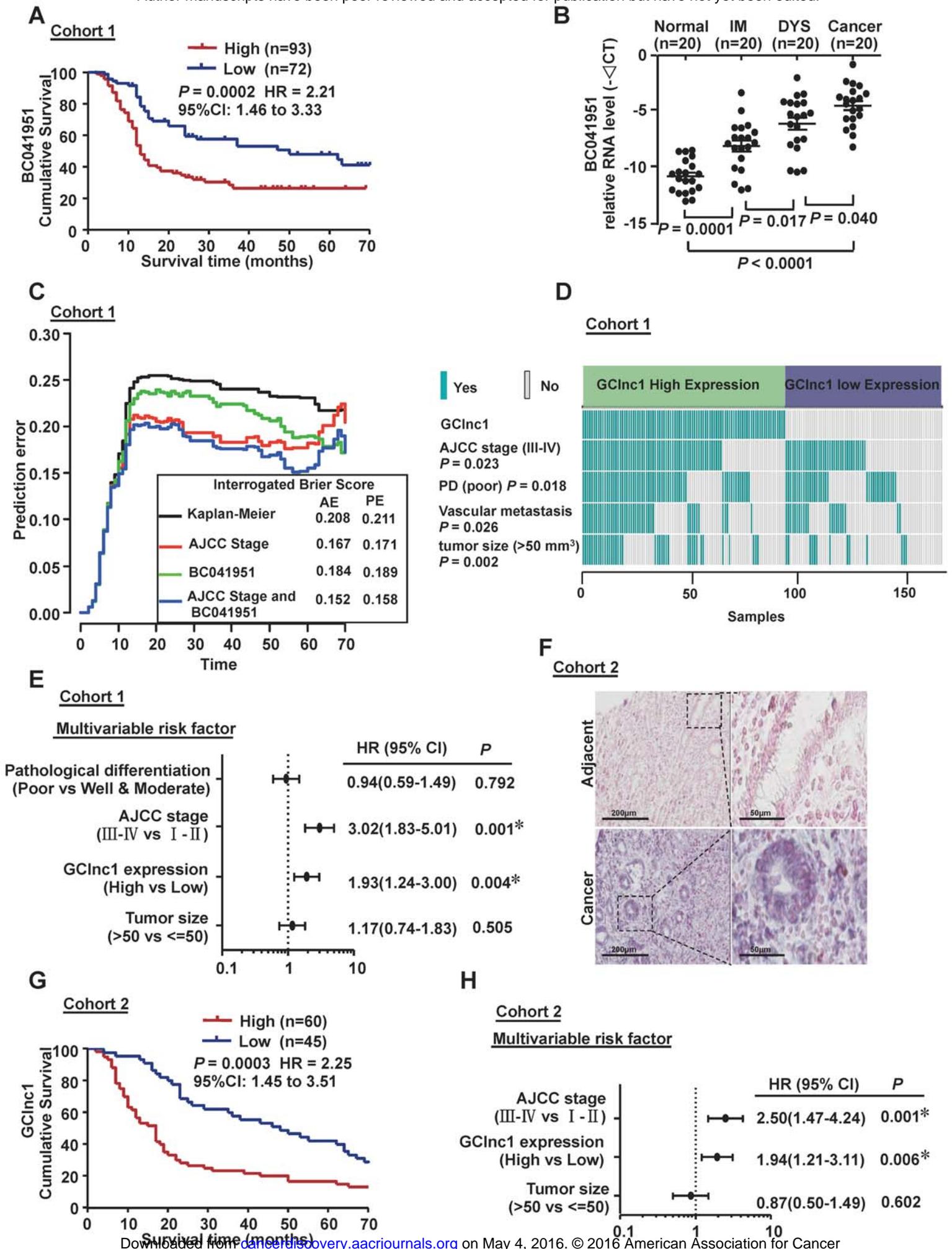


Fig. 2

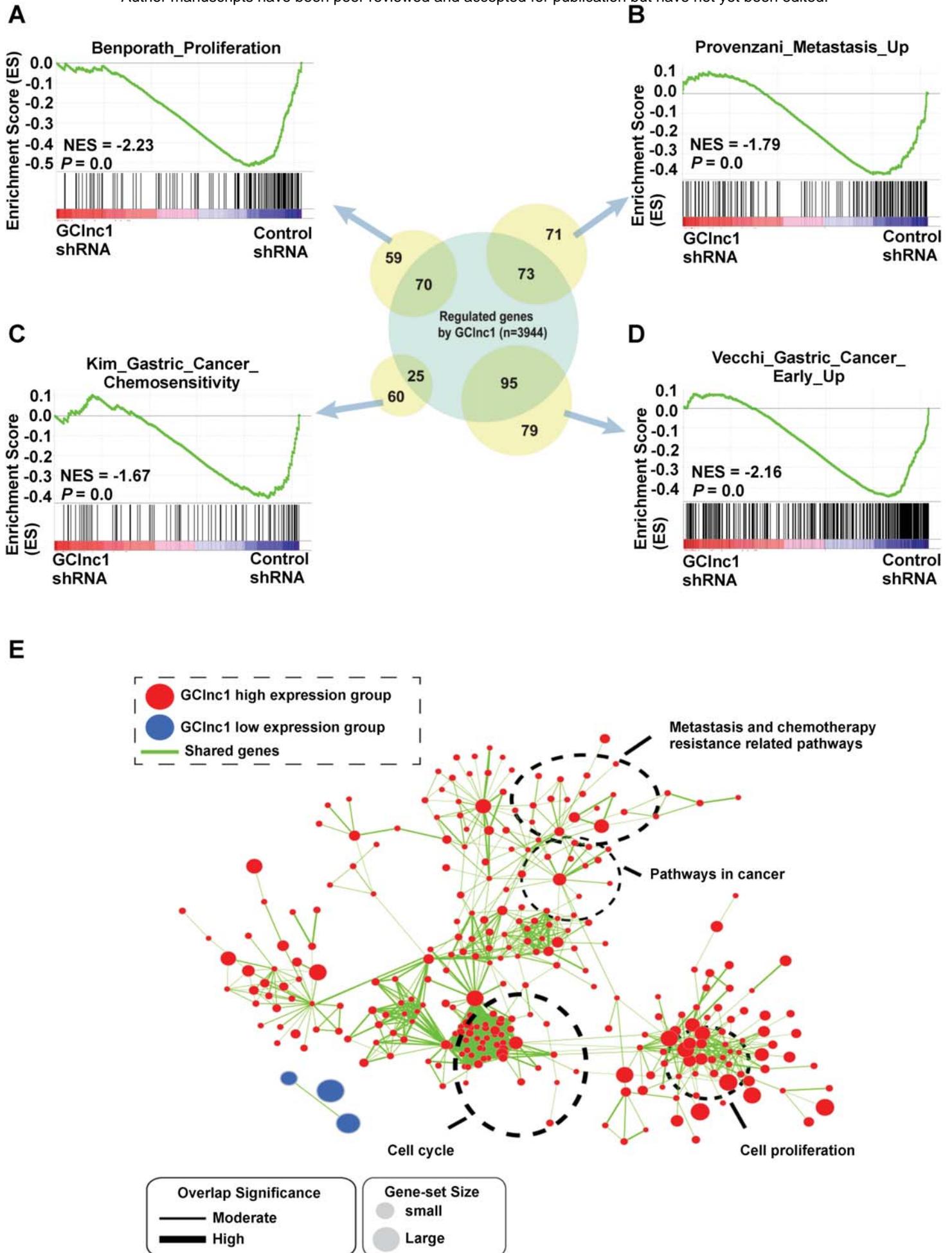


Fig. 3

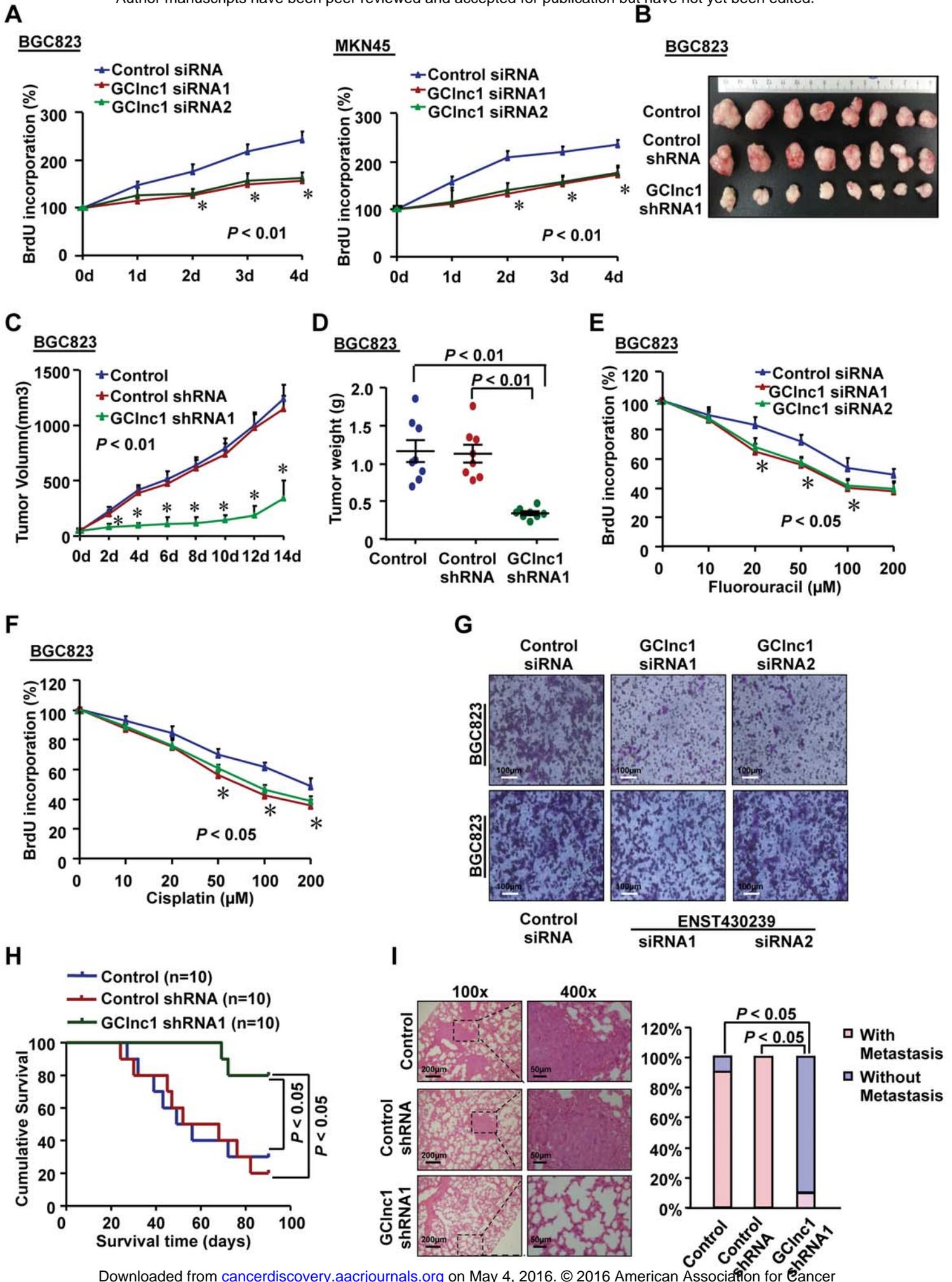


Fig. 4

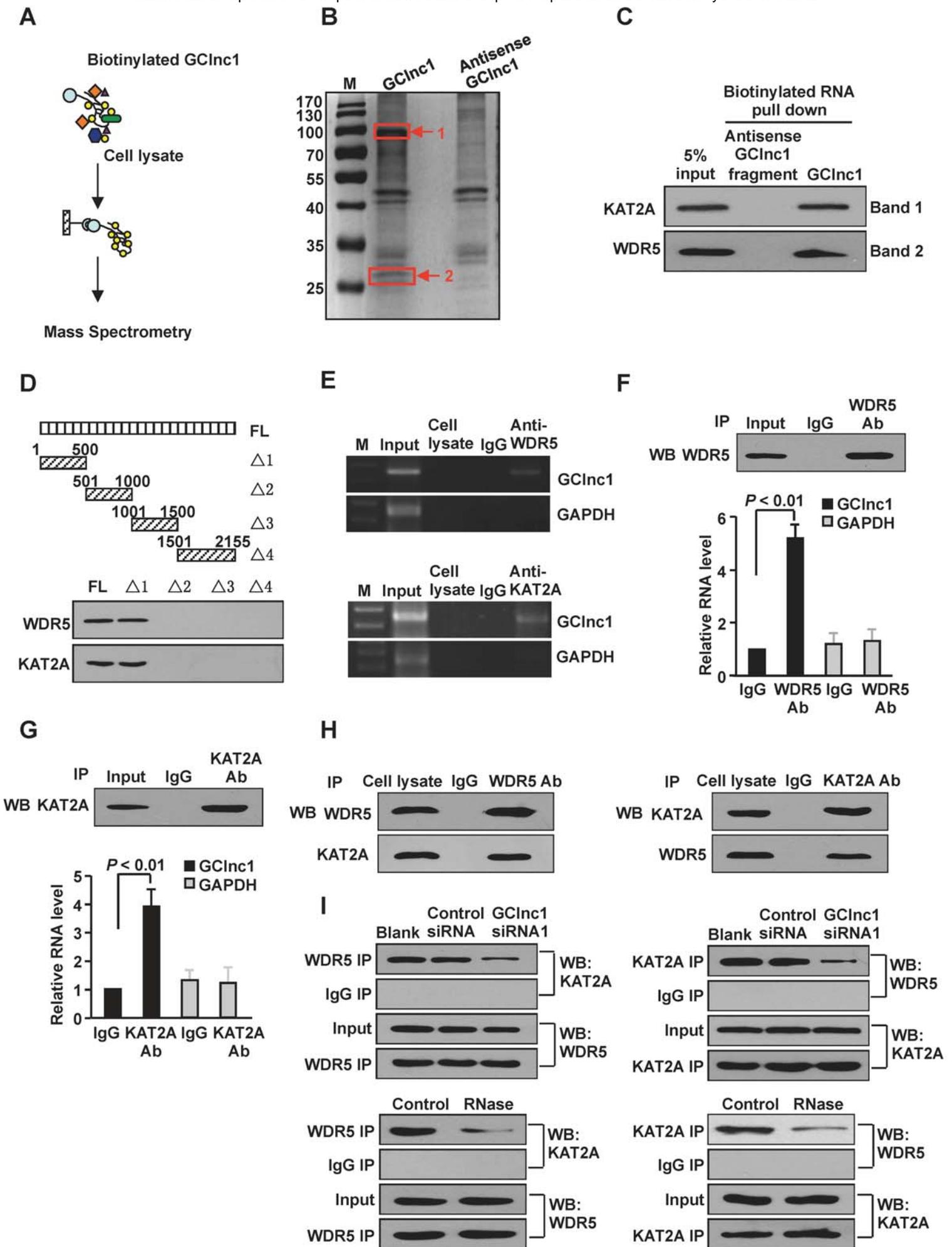


Fig. 5

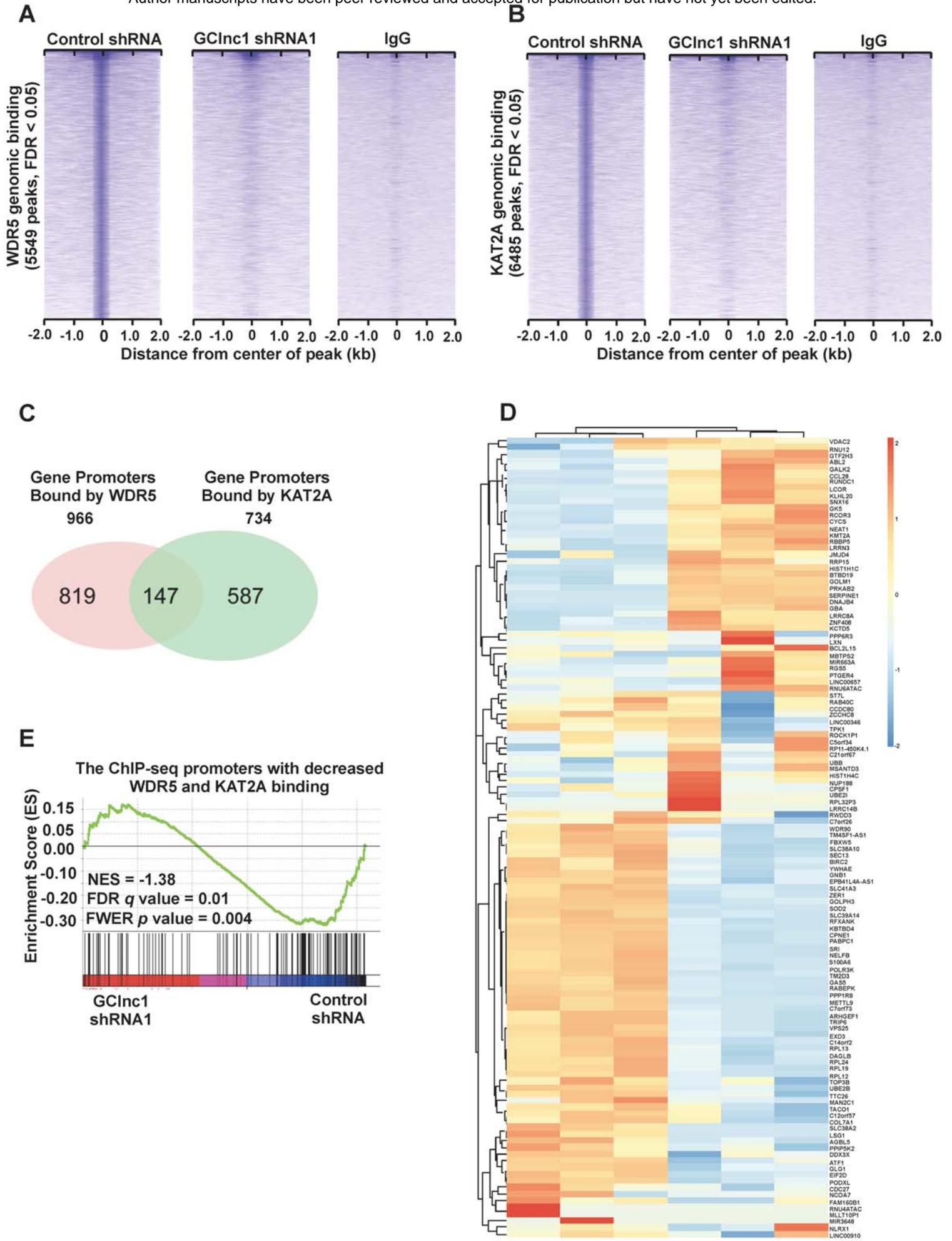
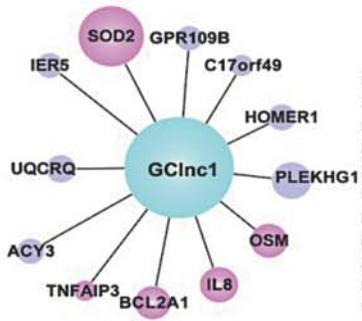
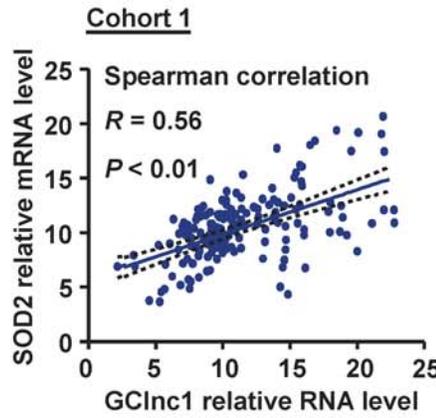


Fig. 6

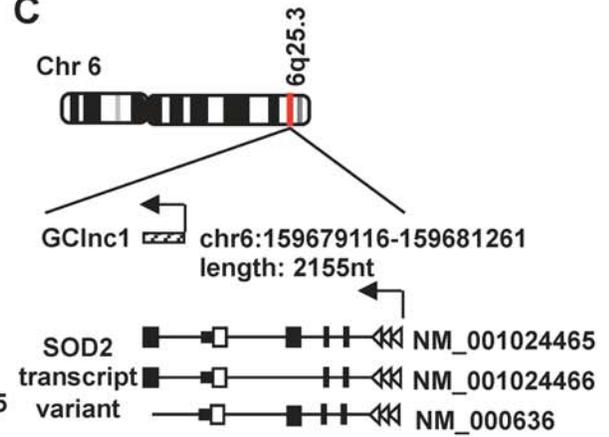
A



B

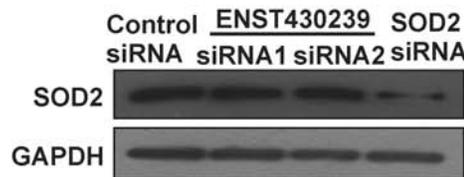
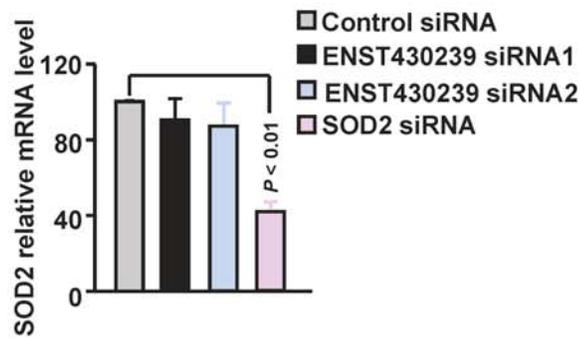
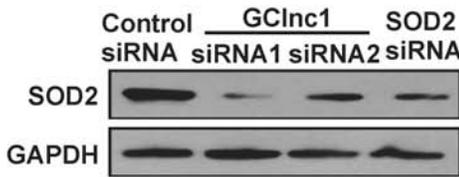
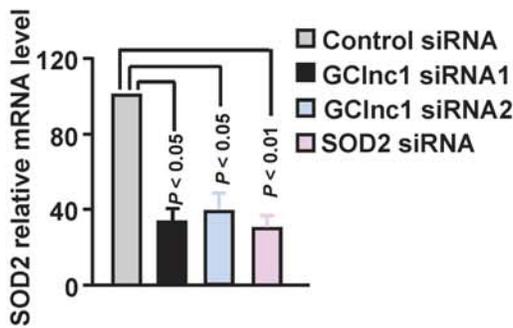


C



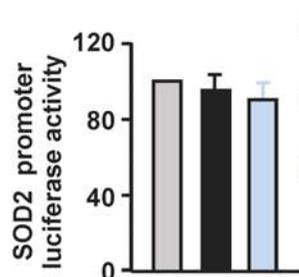
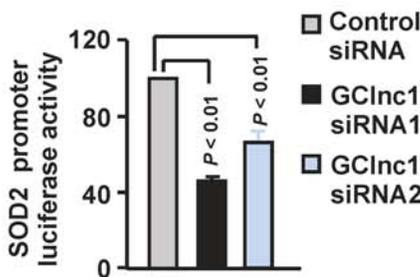
D

BGC823



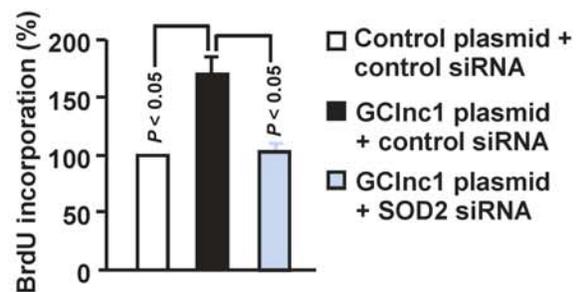
E

BGC823

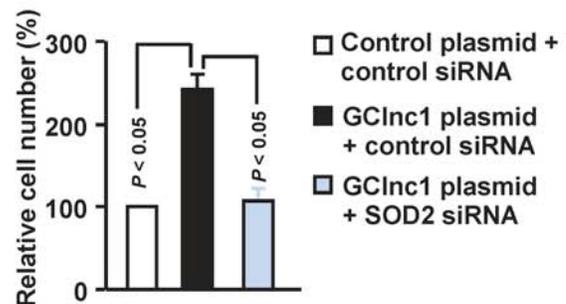
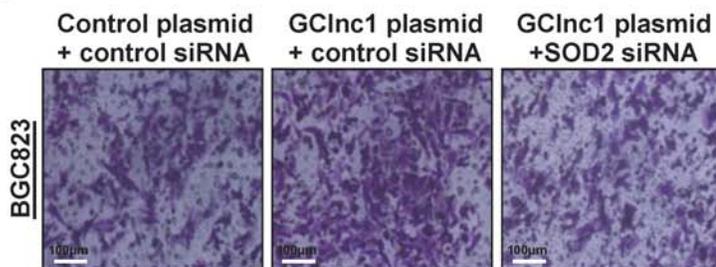


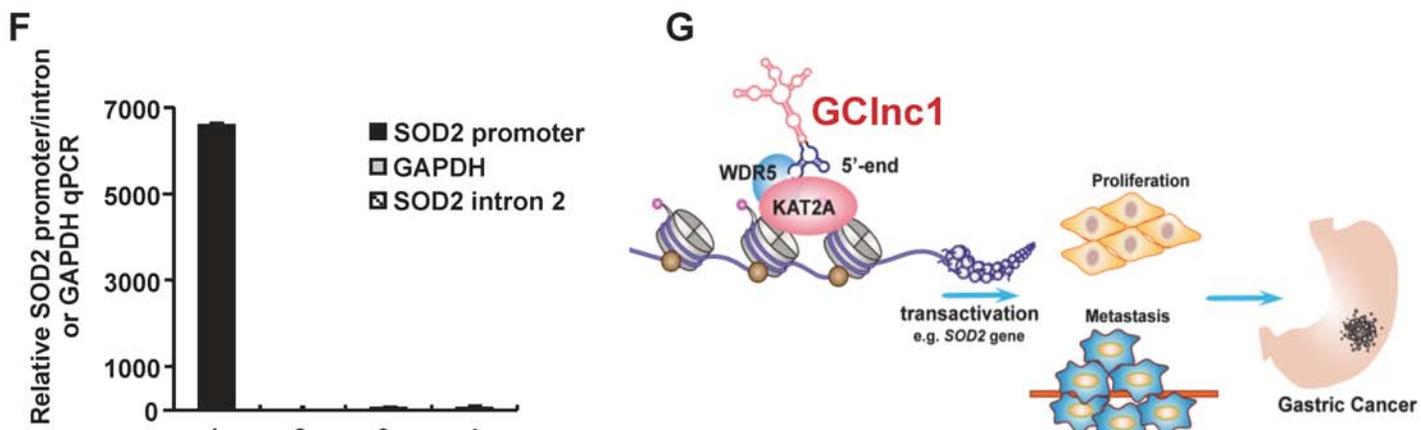
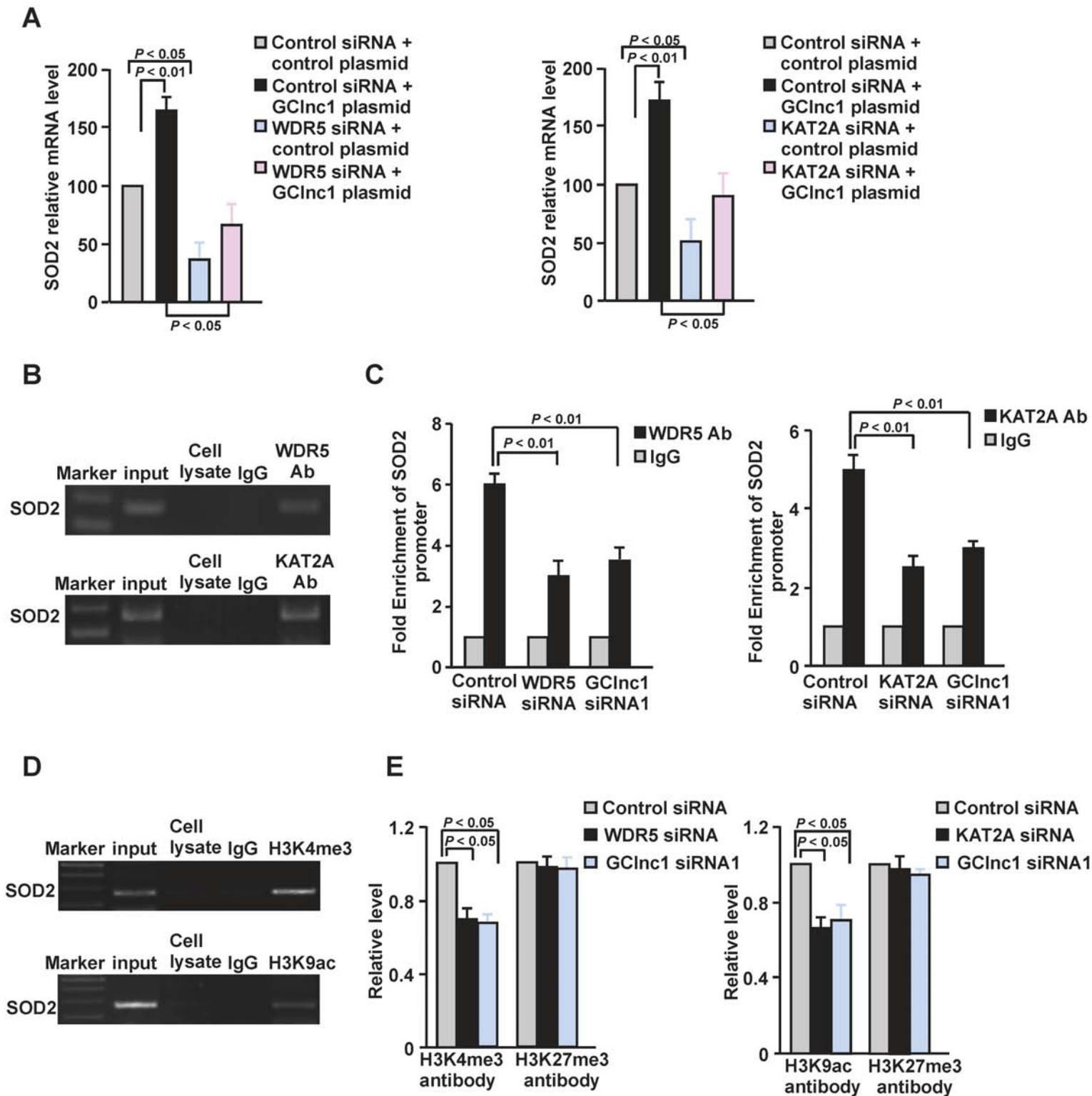
F

BGC823



G





CANCER DISCOVERY

A novel lncRNA GCInc1 promotes gastric carcinogenesis and may act as a modular scaffold of WDR5 and KAT2A complexes to specify the histone modification pattern

Tian-Tian Sun, Jie He, Qian Liang, et al.

Cancer Discov Published OnlineFirst May 4, 2016.

Updated version	Access the most recent version of this article at: doi: 10.1158/2159-8290.CD-15-0921
Supplementary Material	Access the most recent supplemental material at: http://cancerdiscovery.aacrjournals.org/content/suppl/2016/05/04/2159-8290.CD-15-0921.DC1.html
Author Manuscript	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org .