nature communications

Article

https://doi.org/10.1038/s41467-024-53493-x

Peptidylprolyl isomerase A guides *SENP5/GAU1* DNA-lncRNA triplex generation for driving tumorigenesis

Received: 11 April 2024

Accepted: 14 October 2024

Published online: 21 October 2024

Check for updates

Xiaoyu Zhang[1,2,3,4], Tianyi Ding[1,2,3,4], Fan Yang[1,2,3,4], Jixing Zhang[1,2,3,4], Haowen Xu[1,2,3], Yiran Bai[1,2,3], Yibing Shi[1,2,3], Jiaqi Yang[1,2,3], Chaoqun Chen[1,2,3], Chengbo Zhu[1,2,3] & He Zhang[1,2,3]

The three-stranded DNA-RNA triplex hybridization is involved in various biological processes, including gene expression regulation, DNA repair, and chromosomal stability. However, the DNA-RNA triplex mediating mechanisms underlying tumorigenesis remain to be fully elucidated. Here, we show that peptidylprolyl isomerase A (PPIA) serves as anchor to recruit *GAU1* lncRNA by interacting with exon 4 of *GAU1* and enhances the formation of *SENP5/GAU1* DNA-lncRNA triplex. Intriguingly, TFR4 region of *GAU1* exon 3 and TTS4 region of *SENP5* promoter DNA constitute fragments forming the *SENP5/GAU1* triplex. The *SENP5/GAU1* triplex subsequently triggers the recruitment of the methyltransferase SET1A to exon 1 of *GAU1*, leading to the enrichment of H3K4 trimethylation and the activation of *SENP5* transcription for driving the tumorigenesis of gastric cancer in vitro and in vivo. Our study reveals a mechanism of PPIA-guided *SENP5/GAU1* DNA-lncRNA triplex formation in tumorigenesis and providing a concept in the dynamics of isomerase assisted DNA-RNA hybridization.

DNA-RNA hybridization is a process in which single-stranded or double-stranded DNA pairs with single-stranded RNA through base complementarity pairing forming a heteroduplex or a triplex structure[1,2]. Depending on the mode of pairing, RNA can create R-loops with single-stranded DNA or form DNA-RNA triplexes with double-stranded DNA[1,3]. Specifically, DNA-RNA triplexes are formed through complementary base pairing between RNA strands and purine-rich DNAs, relying on Hoogsteen hydrogen bonds for stability[3]. Compared to traditional Watson-Crick base pairing, Hoogsteen hydrogen bonds are more flexible, that RNA can bind to the genome more widely and flexibly, thereby providing increased plasticity for dynamic interactions between molecules[3–5]. Computational analysis has unveiled numerous potential triplexes in the mammalian genome, with the majority of annotated genes, promoters, and intergenic regions harboring at least one potential sequence for triplex formation[6–10]. Nonetheless, the precise functional roles of these triplexes remain to be further explored. Therefore, the discovery of functional DNA-RNA triplexes is of importance for a comprehensive understanding of biological processes.

Although the DNA-RNA triplex was identified from a structural standpoint in 1957, the physiological functions of triplexes have only begun to be uncovered in recent years[11]. Studies have demonstrated that DNA-RNA triplexes play a diverse array of roles in various cellular processes, including transcriptional regulation, DNA replication and repair, telomere maintenance, genome stability, and RNA processing[2,12,13]. The non-coding RNA (ncRNA) derived from the

[1]State Key Laboratory of Cardiology and Medical Innovation Center, Institute for Regenerative Medicine, Shanghai East Hospital, Frontier Science Research Center for Stem Cells, School of Life Science and Technology, Tongji University, Shanghai, P. R. China. [2]Jiangxi Province Key Laboratory of Organ Development and Epigenetics, Clinical Medical Research Center, Affiliated Hospital of Jinggangshan University, Medical Department of Jinggangshan University, Ji'an, Jiangxi Province, P. R. China. [3]School of Life Science, Jinggangshan University, Ji'an, Jiangxi Province, P. R. China. [4]These authors contributed equally: Xiaoyu Zhang, Tianyi Ding, Fan Yang, Jixing Zhang. e-mail: zhanghe@tongji.edu.cn

secondary promoter of *DHFR* can form a triplex structure with the G-track sequence in the primary promoter, preventing the formation of a transcription initiation complex and inhibiting the transcription of *DHFR* mRNA[14]. Promoter-associated RNA (*pRNA*) can form a triplex structure with the *rRNA* promoter sequence. This *pRNA-rRNA* triplex recognizes the DNA methyltransferase DNMT3b and promotes heterochromatin formation at the *rDNA* promoter, ultimately leading to the silencing of *rRNA* transcription[13]. Studies have confirmed that ncRNA can participate in triplex formation, as a subtype of ncRNA, long non-coding RNA (lncRNA) has also attracted significant research interest in the field of triplex formation[5]. Recent studies have indicated that lncRNA-mediated triplexes play a wide-ranging role in stem cell differentiation, angiogenesis, cell proliferation, and epigenetic regulation. LncRNA *HOTAIR* can form DNA-RNA triplexes with both the *PCDH7* promoter and the *HOXB2* promoter, influencing the differentiation of mesenchymal stem cells[15]. The lncRNA *HIF-α-AS1* forms an *EPHA2-HIF-α-AS1* triplex with intron 1 of *EPHA2*, ultimately repressing angiogenesis by reducing *EPHA2* gene expression[16]. *KCNQ1OT1* lncRNA forms a DNA-RNA triplex structure by pairing with Alu and L1 repeat element-rich regions, binding to heterochromatin protein HP1α, and guiding epigenetic silencing of specific repeat DNA elements[17]. However, the currently functionally characterized DNA-lncRNA triplexes represent only a small fraction of the overall landscape. Therefore, the exploration of potential functions of DNA-lncRNA triplexes in pathological processes is potentially interesting.

Studies have shown that DNA-lncRNA triplexes exert various roles in tumorigenesis by recruiting and excluding regulatory factors at specific genomic loci[14]. LncRNA *MAT1* blocks the interaction between the MLL1 complex and *PCDH20* promoter through the formation of DNA-RNA triplex structure, inactivating the transcription of the tumor suppressor *PCDH20* and accelerating tumorigenesis[18]. The lncRNA *Khps1* forms a DNA-RNA triplex with the *SPHK1* promoter, recruiting the histone acetyltransferase p300/CBP to the *SPHK1* promoter, leading to localized changes in chromatin structure, ultimately resulting in the activation of the oncogene *SPHK1* transcription[19]. However, the DNA-lncRNA triplex mediating mechanisms underlying tumorigenesis remain to be fully understood.

In this work, we aim to identify the mechanism of DNA-lncRNA triplex that regulates gastric tumorigenesis on the example of lncRNA *GAU1* and elucidate the pivotal factors responsible for orchestrating triplex formation within tumors. Through an integrated analysis of Triplex-seq and ChIP-seq, we identify a triplex driver PPIA that recruits lncRNA *GAU1* to the *SENP5* promoter, regulating the formation of *SENP5/GAU1* triplex to accelerate tumor progression by activating oncogene *SENP5*, thereby providing a concept for understanding the dynamics of DNA-lncRNA triplex in the context of tumorigenesis.

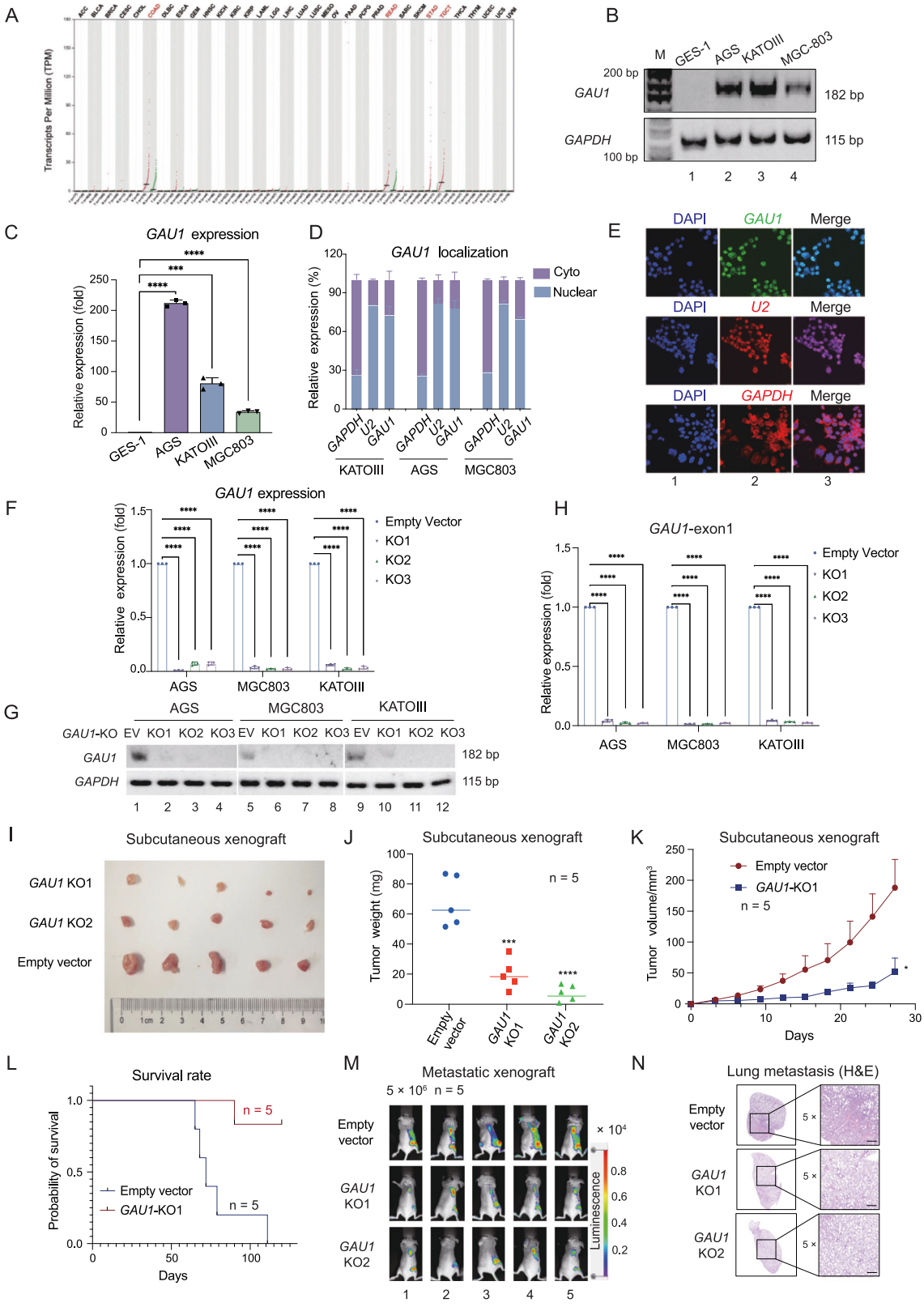
Results

GAU1 served as an oncogenic long non-coding RNA in gastric cancer

To investigate the DNA-lncRNA triplex mediating mechanisms underlying tumorigenesis, suitable functional lncRNA candidates involved in regulating tumorigenesis were identified as the first step. We focused on the *GAU1* lncRNA that we discovered and named[20]. Subsequently, employing bioinformatics prediction, we ascertained the potential of *GAU1* to form DNA-lncRNA triplexes across the entire genome. We found that there was a total of 48,086 genomic sites with the potential to engage in triplex formation with *GAU1* (Supplemental Data 1). Next, we investigated the expression of *GAU1* in various types of tumors. The Cancer Genome Atlas (TCGA) database analysis revealed that *GAU1* exhibited higher expression in gastric cancer, colorectal cancer, and tenosynovial giant cell tumor, whereas its expression was lower or even absent in normal cells (Fig. 1A). Considering that our previous investigations have already substantiated the overexpression of *GAU1* in colorectal carcinomas, we have opted to concentrate our subsequent analytical efforts on gastric cancer, a ubiquitously encountered malignancy, serving as a paradigmatic model for our research endeavors. As expected, we initially observed elevated levels of *GAU1* expression in three human gastric tumor cells, including AGS, MGC803, and KATOIII (Fig. 1B, lanes 2–4 and 1C, square, triangle and inverted triangle) in contrast to the normal human gastric mucosal epithelial cell GES-1 (Fig. 1B, lane 1 and 1C). Next, we conducted a rapid amplification of cDNA ends (RACE) assay to detect the full-length transcript of *GAU1*. RACE assay revealed a 1249 bp full-length *GAU1* transcript with four exons (Supplementary Fig. 1A, B). We further examined the cellular location of the mature *GAU1* transcript. By isolating nuclear and cytoplasmic RNA, we demonstrated that *GAU1* was predominantly localized in the nucleus of AGS, MGC803, and KATOIII cells. We found that 72.6% of *GAU1* was localized within the nucleus of AGS cells, 77.9% within the nucleus of MGC803 cells, and 69.6% within the nucleus of KATOIII cells (Fig. 1D). We designed specific probes for *GAU1*, a nuclear positive control *U2* probe, and a cytoplasmic positive control *GAPDH* probe to investigate the subcellular localization of *GAU1*. Through RNA fluorescence in situ hybridization (RNA-FISH) experiments, we found that the majority of the *GAU1* signal highly overlapped with the nucleus, further substantiating that *GAU1* primarily localizes within the nucleus (Fig. 1E). Subsequently, we employed the CRISPR/Cas9 system to simultaneously knockout exons 2–3 and exon 4 of the *GAU1* (Supplementary Fig. 2A) and observed that the expression of *GAU1* in tumor cells was reduced to 1.3–6.2% of the original level (Fig. 1F and G). Additionally, we assessed the expression levels of *GAU1* exon 1 following the knockout of exons 2-3 and exon 4. We indicated a significant decline in the expression of *GAU1* exon 1 concurrent with the removal of exons 2–4 (Fig. 1H). Next, we conducted in vitro colony formation assays to evaluate the potential of *GAU1* in promoting tumor formation. After *GAU1* knocking out, we observed that the number of viable single-cell colonies significantly decreased, dropping to approximately 25–35% (Supplementary Fig. 2B, lanes 2–3 and Supplementary Fig. 2C, square and triangle) as compared to the empty vector group (Supplementary Fig. 2B, lane 1 and Supplementary Fig. 2C, circle). To evaluate the impact of *GAU1* on tumor metastasis in vitro, we carried out the transwell assay. The results indicated a significant reduction in the migratory capability of the three tumor cells following *GAU1* knockout (Supplementary Fig. 2D, lanes 2–3) as compared to the empty vector group (Supplementary Fig. 2D, lane 1). Quantitative analysis using Microplate Spectrophotometer following elution with glacial acetic acid revealed that the migration rate was reduced by 56–73% following *GAU1* knockout (Supplementary Fig. 2E, square and triangle) as compared to the empty vector group (Supplementary Fig. 2E, circle) in AGS, MGC803, and KATOIII cells. Therefore, the lncRNA *GAU1* plays an oncogene role in gastric tumor cells in vitro.

To further investigate the oncogenic function of *GAU1* lncRNA in vivo, we conducted animal experiments using a subcutaneous xenograft model in nude mice. The mice were injected with *GAU1* empty vector and two *GAU1* KO cells (KO1 and KO2) in the left underarm, with 2 million tumor cells per mouse (Supplementary Fig. 3A). After tumor formation, tumor volume was measured every three days. On day 30, euthanasia was performed on the mice, followed by tumor removal, weight measurement, and photographic documentation. We demonstrated a significant reduction in tumor volume and weight (Fig. 1I, $n = 5$) in the *GAU1* KO groups as compared to the empty vector group, with tumor weight decreased by 66.7–85% (Fig. 1J, square and triangle) and tumor volume ($^{**}P < 0.01$) decreased by 73.7% (Fig. 1K, square). Furthermore, knockout of *GAU1* significantly extended the lifespan of mice ($n = 5$). At the conclusion of the 120-day experiment, only one mouse had perished, whereas all mice in the empty vector group had succumbed (Fig. 1L). In addition, to assess the metastasis potential of *GAU1*, we established a metastatic mice model through intravenous tail injection of gastric tumor cells in nude mice. Systemic metastasis was

https://doi.org/10.1038/s41467-024-53493-x



evaluated using bioluminescence and fluorescence imaging. We revealed that knockout of *GAU1* impaired systemic and pulmonary metastasis of gastric tumor cells and significantly inhibited tumor cell growth. The luciferase signaling exhibited a significant reduction in both proximal (lung) and distal (abdominal cavity) metastasis, as evidenced by a notable decrease in the total luciferase signal values per animal in the *GAU1*-KO groups (Fig. 1M). After 5 weeks, the mice were sacrificed by cervical dislocation, and the lungs were collected for further analysis. Histological examination confirmed smaller metastatic nodules in the *GAU1*-KO groups as compared to the empty vector group, indicating knockout of *GAU1* impaired lung seeding and growth (Fig. 1N 5 ×, Supplementary Fig. 3B, 200 × and 400 ×). Taken together, these results demonstrate that *GAU1* plays a crucial role in promoting tumor growth and metastasis in gastric tumors in vivo.

Fig. 1 | *GAU1* served as an oncogenic long non-coding RNA in gastric cancer. **A** The expression profile of *GAU1* in various tumor and normal cell types, highlighted in red to denote high expression in tumors. **B**–**C** PCR and Real-time PCR results demonstrated the expression levels of *GAU1* in tumor and normal cells (Gastric cancer cells: AGS, $P < 0.0001$; MGC803, $P = 0.0001$; KATOIII, $P < 0.0001$. The normal cell was GES-1). Data are presented as mean ± SD from three independent experiments an unpaired two-tailed t-test. ****$P < 0.0001$ and ***$P < 0.001$as compared to the normal cell line GES-1. **D** Localization of mature *GAU1* transcripts was determined. Representative blots from three independent experiments. Source data are provided as a Source Data file. **E** Representative RNA-FISH images showed that the *GAU1* signal overlapped with DAPI staining. The scale bars represent 10 μm. Representative images from three independent experiments. Source data are provided as a Source Data file. **F** Real-time PCR (above) and **G** PCR (below) analyses of *GAU1* CRISPR/Cas9 transcripts in AGS ($P < 0.0001$), MGC803 ($P < 0.0001$), and KATOIII ($P < 0.0001$) cells. Data are presented as mean ± SD from three independent experiments an unpaired two-tailed t-test. ****$P < 0.0001$ compared with the empty vector (Empty vector: (EV) cells were transfected with the empty CRISPR/Cas9 vector. **H** The expression of exon1 after *GAU1* exon2–3 and exon4 knockout in AGS ($P < 0.0001$), MGC803 ($P < 0.0001$) and KATOIII ($P < 0.0001$) cells was analyzed by Real-time PCR. Data are presented as mean ± SD from three independent experiments an unpaired double-tailed T-test, ****$P < 0.0001$. **I** Photographs of mouse xenograft tumors ($n = 5$ mice). Representative images from five independent samples. **J** Bar graphs showed the weight of allograft tumors in mice treated with Empty vector (Circle), *GAU1*-KO1(Square, $P = 0.0006$), and *GAU1*-KO2 (Triangle, $P < 0.0001$), respectively. Tumor weight (mg) was measured and presented as the mean ± SD ($n = 5$ mice) using an unpaired two-tailed t test. ****$P < 0.0001$ and ***$P < 0.001$. **K** Tumor sizes (mm^3) calculated as length × width × width/2. Average volume of subcutaneous tumors after treatment with Empty vector (Circle), *GAU1*-KO1 (Square, $P = 0.0323$). Data are presented as mean ± SD ($n = 5$ mice) using an unpaired two-tailed t-test. *$P < 0.05$. **L** Survival analysis of mice following *GAU1*-KO (Red) or in the empty vector (Black) group ($n = 5$ mice). Source data are provided as a Source Data file. **M** *GAU1* KO inhibited the establishment of GC metastasis. In vivo total-body bioluminescence images of nude mice ($n = 5$ mice). Source data are provided as a Source Data file. **N** Representative images of the histological analysis of lung seeding in mice (original magnification, 5×, scale bar: 500 μm). Representative images from five independent samples. Source data are provided as a Source Data file.

p300 promotes H3K27 acetylation at the *GAU1* promoter

To elucidate the transcriptional activation mechanism of *GAU1* in gastric tumor cells, we investigated changes in histone acetylation and methylation at the *GAU1* promoter. Specific detection sites were designed on the promoter (site b) and negative control site (site a) of *GAU1* (Fig. 2A). Given the high expression of *GAU1* in gastric cancer, we chose the active histone modifications H3K4 trimethylation (H3K4me3) and H3K27 acetylation (H3K27ac) as the detected candidates. Using chromatin immunoprecipitation (ChIP) assays followed by quantitative PCR (qPCR), we observed increased enrichment of H3K27ac in the promoter of *GAU1* (site b) in AGS, MGC803, and KATOIII gastric tumor cells (Fig. 2B, circle and Supplementary Fig. 4A, left lanes 2–4) as compared to normal control cell GES-1 (Fig. 2B and Supplementary Fig. 4A, left lane 1). However, H3K4me3 at the *GAU1* promoter in gastric tumor cells showed no significant changes as compared to normal GES-1 cells (Fig. 2B, diamond and Supplementary Fig. 4A left lanes 2–4). Notably, no significant changes were observed in H3K4me3 and H3K27ac at the negative control site (site a) of *GAU1* in gastric tumor cells (Supplementary Fig. 4B). Next, we focused on the acetylation modification of acetyltransferase p300/CBP complex at the *GAU1* promoter. Through ChIP-qPCR analysis, we showed stronger binding affinity at *GAU1* promoter (site b) for p300 in gastric tumor cells (Fig. 2C, circle and Supplementary Fig. 4A, right lanes 2–4) compared with GES-1 cells (Supplementary Fig. 4A, right lane 1), while no significant changes were observed in CBP binding at *GAU1* promoter in gastric tumor cells (Fig. 2C, diamond). Notably, there were no significant changes observed in the binding of p300 and CBP at the negative control site (site a) both in gastric tumor cells and normal cells (Supplementary Fig. 4C). Subsequently, we selectively treated gastric tumor cells with the p300-specific inhibitor C646 at concentrations of 5 μM and 10 μM for 48 h. Through ChIP-qPCR analysis, we observed a significant reduction in the binding affinity of p300 to the promoter of *GAU1* (site b) with both C646 treatment at 5 μM and 10 μM (Supplementary Fig. 4D), while the negative control site (site a) showed no discernible change (Supplementary Fig. 4E). Through ChIP-qPCR analysis of the modification status of H3K27ac on the *GAU1* promoter (site b), we observed a significant decrease in H3K27ac levels after p300 silencing using C646 treatment at 5 μM and 10 μM (Fig. 2D), while the negative control site (site a) showed no apparent change (Supplementary Fig. 4F). Next, we used real-time PCR to assess the impact of the C646 inhibitor on *GAU1* expression. After a 48-hour C646 treatment with 5 μM and 10 μM in gastric tumor cells, we observed that the transcription levels of *GAU1* decreased to 18.7–52.9% of the original levels (Fig. 2E). These findings suggest that p300 catalyzes the acetylation of H3K27 at the *GAU1* promoter and activates *GAU1* transcription in gastric tumor cells.

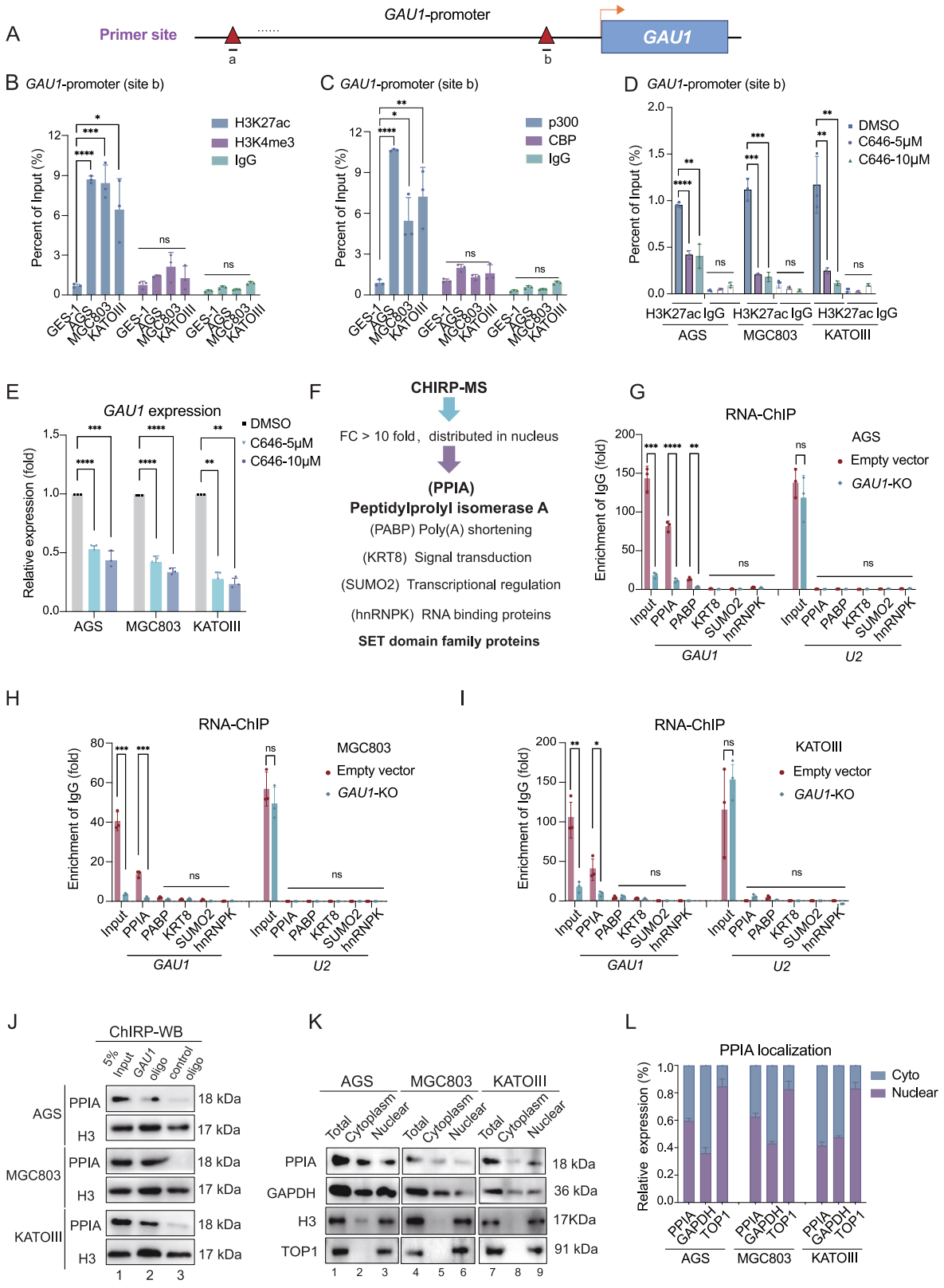
GAU1 interacts with PPIA in gastric tumor cells

To elucidate the cofactors specifically interacting with *GAU1* lncRNA, we employed the biotin-labeled oligos Chromatin Isolation by RNA Purification (ChIRP) method to capture proteins that interacted with *GAU1* in gastric tumor cells (Supplementary Fig. 5A). The captured proteins were labeled and analyzed using mass spectrometry to determine their types and enrichment levels Supplementary Fig. 5B–E). Through screening and analyzing peptide signals, we identified five enriched proteins (PPIA, PABP, KRT8, SUMO2, and hnRNPK) and the SET protein family in the ChIRP lysates (Fig. 2F and Supplementary Table 3). Since there were many members of the SET family of proteins, we firstly used RNA immunoprecipitation (RIP) assay to explore the interactions between the potential five candidates (PPIA, PABP, KRT8, SUMO2, and hnRNPK) and *GAU1*. We showed a stronger interaction between *GAU1* and PPIA proteins in three tumor cells transfected with the empty vector (circle) as compared to *GAU1* KO tumor cells (diamond) (Fig. 2G–I and Supplementary Fig. 6). However, there were not significant enrichments at the non-specific *U2* control both in *GAU1* empty vector and *GAU1* knockout cells. To further clarify this interaction, we performed a ChIRP-western blot assay and confirmed that PPIA could interact with *GAU1* (Fig. 2J). Next, we performed nuclear-cytoplasmic fractionation to examine the subcellular localization of PPIA in gastric tumor cells. Topoisomerase I and H3 were used as nuclear localization controls, and GAPDH protein served as a cytoplasmic localization control (Fig. 2K). As expected, we found that PPIA was widely distributed in both the nucleus and cytoplasm, with approximately 38.4% to 64.5% localized in the nucleus of cells (Fig. 2L). These results suggest that PPIA, but not hnRNPK, PABP, SUMO2 and KRT8 proteins, interacts with *GAU1* lncRNA in gastric tumor cells.

SENP5/GAU1 is a DNA-lncRNA triplex

After confirming the binding of PPIA with *GAU1* through ChIRP and RNA-ChIP assays, we intended to utilize whole-genome co-localization to identify the DNA sites that formed the DNA-lncRNA triplexes with *GAU1* lncRNA. In the experimental workflow of our Triplex-seq analysis, Proteinase K was employed to dismantle protein-mediated DNA-RNA conjugations, while RNase H was utilized to specifically target and eliminate the DNA-RNA duplexes of R-loops. This approach refined the enrichment process, ensuring the isolation of authentic DNA-RNA triplexes. Triplex-seq assay showed 914 potential binding sites where *GAU1* interacted directly with the genome via complementary base

https://doi.org/10.1038/s41467-024-53493-x



pairing (Fig. 3A). We then performed PPIA ChIP-seq to identify PPIA binding sites across the genome, which showed 798 binding sites (Fig. 3A). By aligning the results from above assays, we identified a total of 637 co-localized binding sites of *GAU1* triplex and PPIA across the entire genome. Next, we preferred gene promoters with potential transcriptional regulatory roles across 637 co-localized sites, and 26 co-localized promoter sites distributed in the promoter regions of 15 genes were identified (Fig. 3A and Supplementary Fig. 7A, B). We utilized TIMER2.0 to analyze 15 target genes and found that only *Sentrin-specific protease 5* (*SENP5*) exhibited significant differences between gastric cancer tumor tissues and normal tissues (Supplementary Fig. 7C), while the other 14 targets were excluded due to their lower expression levels or lack of potential tumor-related risks. To further validate this result, we conducted an analysis of *GAU1* triplex-seq (blue)

Fig. 2 | *GAU1* interacts with PPIA in gastric tumor cells. A The schematic diagram of the *GAU1* promoter region and primer locations (Supplementary Table 1) with ChIP detection sites (site a and b). **B** ChIP analysis examined histone modifications (H3K4me3, H3K27ac) at the *GAU1* promoter ($n = 3$). Rabbit normal IgG served as the negative control. ChIP enrichment presented as percentage of bound/input signal. Concentration of H3K4me3 on *GAU1* promoter in tumor cells (AGS, $P < 0.0001$; MGC803, $P = 0.0006$; KATOIII, $P < 0.0132$) compared to GES-1. Data are presented as mean ± SD from three independent experiments an unpaired two-tailed t-test. $^{****}P < 0.0001$, $^{***}P < 0.001$ and $^{*}P < 0.05$; ns no significance. **C** ChIP analysis of CBP/p300 at the *GAU1* promoter ($n = 3$). Concentration of p300 on *GAU1* promoter in tumor cells (AGS, $P < 0.0001$; MGC803, $P = 0.0103$; KATOIII, $P = 0.0071$) compared to GES-1. Data are presented as mean ± SD from three independent experiments an unpaired two-tailed t-test. $^{****}P < 0.0001$, $^{**}P < 0.01$ and $^{*}P < 0.05$; ns: no significance. **D** ChIP analysis on *GAU1* promoter (site b) ($n = 3$) after p300 inhibition with C646 (5 μM and 10 μM). The concentration of P300 on *GAU1* promoter in C646 treated cells (AGS, C646 = 5 μM, $P < 0.0001$, C646 = 10 μM, $P = 0.0002$; MGC803, C646 = 5 μM, $P = 0.0003$, C646 = 10 μM, $P = 0.0019$; KATOIII, C646 = 5 μM, $P = 0.0065$, C646 = 10 μM, $P = 0.0039$) compared with DMSO treated cells. Data are presented as mean ± SD from three independent experiments an unpaired two-tailed t-test. $^{***}P < 0.001$, $^{**}P < 0.01$; ns: no significance. **E** Real-time PCR assessed *GAU1* expression in tumor and normal cells after p300 inhibition with C646 (5 μM and 10 μM). Gastric cancer cells: AGS (C646 = 5 μM, $P < 0.0001$, C646 = 10 μM, $P = 0.0003$), MGC803(C646 = 5 μM, $P < 0.0001$, C646 = 10 μM, $P < 0.0001$), KATOIII (C646 = 5 μM, $P < 0.0001$, C646 = 10 μM, $P < 0.0001$); Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. $^{****}P < 0.0001$ and $^{***}P < 0.001$. **F** ChIRP-MS detected proteins that bound to *GAU1*. **G–I** Real-time PCR assessed *GAU1* binding to PPIA (AGS, $P < 0.0001$; MGC803, $P = 0.0003$; KATOIII, $P = 0.0188$), PABP, KRT8, SUMO2, and hnRNPK in RNA-ChIP assay. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. $^{****}P < 0.0001$, $^{***}P < 0.001$, $^{**}P < 0.01$ and $^{*}P < 0.05$. **J** Western blotting verified ChIRP-MS results. *GAU1* and lacZ oligo pulldown. PPIA antibody detected interaction, H3 antibody as positive control. Representative blots from three independent experiments. Source data are provided as a Source Data file. **K** Western blot analysis was conducted to determine the subcellular localization of PPIA in gastric tumor cells AGS, MGC803, and KATOIII. Representative blots from three independent experiments. Source data are provided as a Source Data file. **L** Western blot data were analyzed using ImageJ to automatically determine the subcellular localization of PPIA protein. Data are presented as mean ± SD from three independent experiments.

and PPIA ChIP-seq (red), revealing a significant enrichment of *GAU1* triplex and PPIA signals at the *SENP5* promoter in gastric cancer (Fig. 3B). Subsequently, we further validated the potential role of *SENP5* in gastric cancer through qPCR and western blot assays. We found that the mRNA expression of *SENP5* in tumor cells were higher than the normal GES-1 cells (Fig. 3C, square, triangle and inverted triangle). Similarly, the protein expression of SENP5 in tumor cells (Fig. 3D, lanes 2–4) were significantly elevated as compared to normal cells (Fig. 3D, lane 1). Next, we used qPCR and western blot to assess *SENP5* expression in *GAU1* empty vector and *GAU1* knockout tumor cells. We found that *SENP5* was significantly decreased after knocking out of *GAU1* in tumor cells (Fig. 3E, square and triangle). We also confirmed that protein level of SENP5 significantly reduced after knocking out of *GAU1* in tumor cells (Fig. 3F, lanes 2–3, 5–6 and 8–9). We further analyzed the Cancer Genome Atlas (TCGA) open data set (Versteeg database; http://www.cbioportal.org) and found that the expression of *SENP5* in gastric cancer tissues was also positively proportional to PPIA (Fig. 3G). Taken together, these results indicate that *GAU1* and PPIA are co-anchored at the *SENP5* promoter.

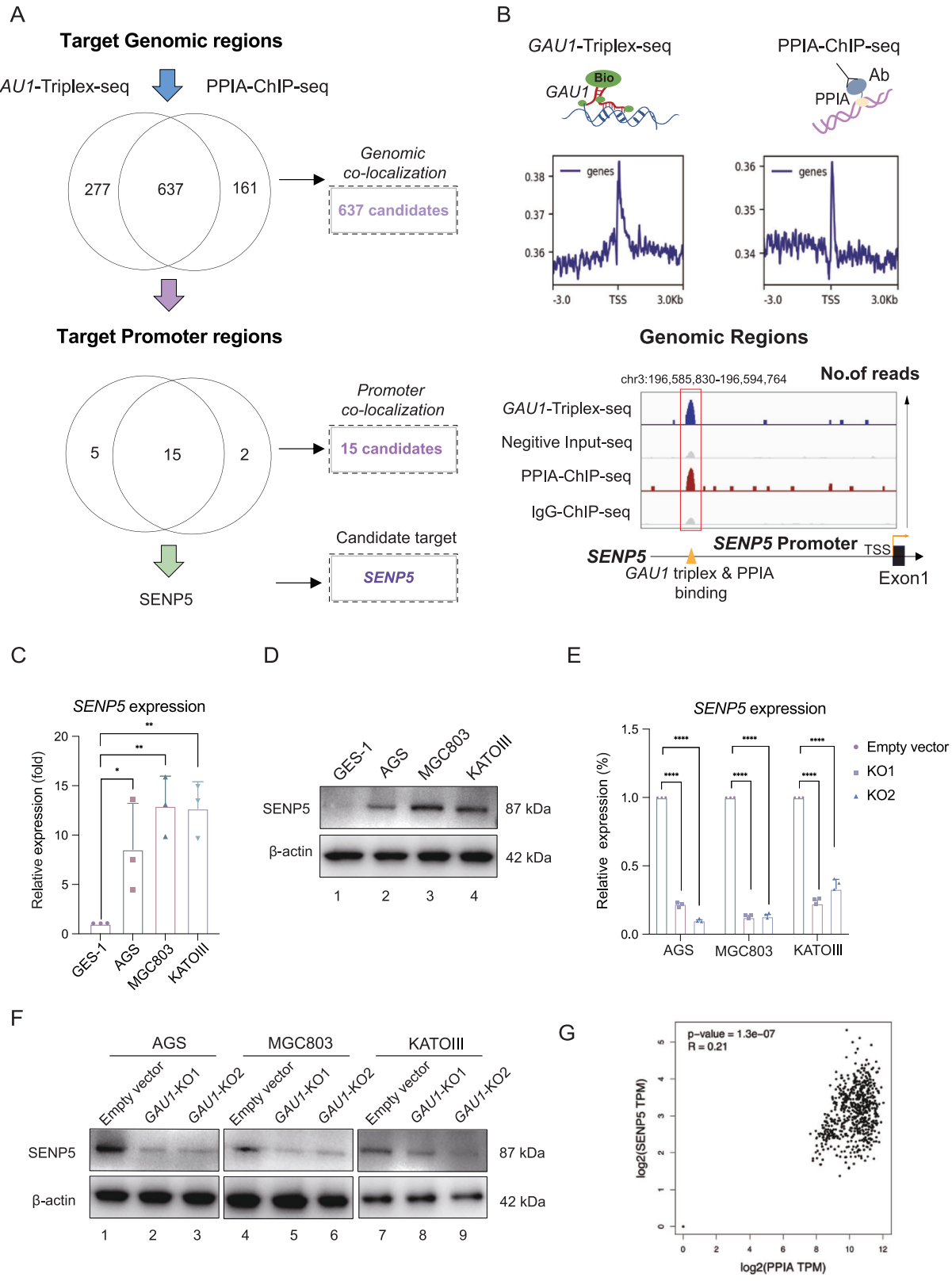
SENP5 is an oncogene in gastric cancer

To investigate the oncogenetic role of *SENP5* in gastric cancer, we conducted in vitro and in vivo experiments. We used short hairpin RNA (shRNA) to knockdown (KD) the expression of *SENP5* and found that the mRNA level of *SENP5* was dropped to 17.6–43.4% of the original level by qPCR in tumor cells (Supplementary Fig. 8A, inverted triangle, circle and diamond). These findings were further confirmed by western blot analysis, which showed that protein level of SENP5 were significantly reduced after transfecting with *SENP5* shRNAs in tumor cells (Fig. 4A, lanes 2–4, 6–8 and 10–12). We selected *shSENP5*–1 and *shSENP5*–2 with the efficient efficacy of *SENP5* knockdown for further studies. Next, we tested the ability of tumor cells to form colonies using a classical cell colony formation assay in vitro to investigate the impact of *SENP5* knockdown on cell proliferation of gastric cancer. We demonstrated a significant reduction in the growth rate of tumor cells following *SENP5* knockdown in tumor cells (Fig. 4B, lanes 2–3) as compared to the empty vector group (Fig. 4B, lane 1). The number of efficient colonies in the *SENP5* silencing tumor cells were only 20–30% (Fig. 4C, square and triangle) of that observed in the empty vector group (Fig. 4C, circle). To further evaluate the potential role of *SENP5* in promoting tumor metastasis in gastric cancer, we employed the transwell assay and revealed that *SENP5* knockdown led to a significant reduction in the metastatic ability of gastric tumor cells (Fig. 4D, lanes 2–3) as compared to the empty vector group (Fig. 4D, lane 1). Quantitative analysis using Microplate Spectrophotometer following elution with glacial acetic acid showed a 62–75% decrease in the metastasis rate in *SENP5* silencing tumor cells (Fig. 4E, square and triangle) as compared to the empty vector group (Fig. 4E, circle). Thus, these results suggest that *SENP5* plays an oncogene role in gastric cancer in vitro.

To further investigate the ability of *SENP5* to promote tumor growth and metastasis in vivo, we established subcutaneous xenograft model in nude mice. *SENP5* empty vector and two *SENP5*-shRNA knockdown (KD) cells were injected into the left underarm of 5-week female nude mice, with 2 million tumor cells per mouse. The tumor volume was measured every 3 days, and the mice were euthanized after 30 days, and tumor weight and photos were collected (Fig. 4F, $n = 5$). Our results showed a 60–64% reduction in tumor volume (Fig. 4G, square and triangle) and a 50–61.8% reduction in tumor weight (Fig. 4H, square and triangle) in the *SENP5* KD-groups as compared to the empty vector group (Fig. 4G, H, circle). Furthermore, we observed a significant extension in the survival time of the *SENP5* KD-groups. At the end of the 120-day experiment, only one mouse out of the two *SENP5* KD-groups had succumbed (Fig. 4I, square and triangle), while all mice in the empty vector group had died (Fig. 4I, circle). To explore the metastasis-promoting potential of *SENP5*, we constructed systemic metastasis mice models by tail vein injection of *SENP5*-empty vector and *SENP5*-KD tumor cells. Thirty days after constructing a metastatic tumor mouse model, we showed that *SENP5* knockdown significantly inhibited systemic metastasis and lung metastasis of gastric tumor cells by bioluminescence and fluorescence imaging analysis (Fig. 4J). The luciferase signal of each animal in the *SENP5* KD-group was significantly decreased as compared to the empty vector group. Additionally, lung histology analysis revealed smaller metastatic nodules in the *SENP5* KD-group (Fig. 4K and Supplementary Fig. 8B, lanes 3–6). Moreover, analysis of the (Versteeg database; http://www.cbioportal.org) showed that *SENP5* expression was significantly increased in gastric cancer tissues ($n = 408$) as compared to normal tissues ($n = 211$) (Fig. 4L). To explore the prognostic role of *SENP5* in tumors, we investigated the overall survival probability. As expected, the R2 Dorsman database showed that higher level of *SENP5* (High expression of *SENP5*, $n = 271$) correlated with a worse prognosis in gastric cancer (Fig. 4M). Taken together, these findings suggest that *SENP5* functions as an oncogene in gastric cancer.

GAU1 recruits SET1A to increase the H3K4 trimethylation of *SENP5* promoter

To further elucidate the activated mechanism of *SENP5* expression mediated after the formation of *SENP5/GAU1* triplex in gastric tumor



cells, we screened for histone modification changes, specifically focusing on H3K4me3 and H3K27ac. We have designed two detected sites to cover the *SENP5* promoter (sites d and e) and an upstream negative control site (site c) in DNA-ChIP assay. Site d was corresponding to the PPIA binding site at *SENP5* promoter identified from the PPIA-ChIP-seq and situated 1780-nucleotide (nt) away from the transcription start site (TSS) of *SENP5*. Site e was located at a distance of 226-nt away from the *SENP5* TSS (Fig. 5A). Through ChIP-qPCR, we found a notable increase in H3K4me3 at the *SENP5* promoter (site e) of gastric tumor cells compared with normal GES-1 cells (Fig. 5B, circle).

Furthermore, at another detection site (site d) on the *SENP5* promoter (Supplementary Fig. 9A), we only observed a slight increase in H3K4me3 in AGS cells, but the change was far less pronounced as compared to the substantial alteration observed at site e (Fig. 5B and

Fig. 3 | *SENP5/GAU1* is a DNA-lncRNA triplex. A Overlapping maps of PPIA ChIP-seq and *GAU1* Triplex-seq were presented. Above: On the left were genomic loci bound only by *GAU1* triplex, on the right were genomic loci bound only by PPIA, and in the middle were genomic loci bound by both *GAU1* and PPIA. Below: On the left were genomic promoter sites that were bound by *GAU1* triplex only, on the right were genomic promoter sites that were bound by PPIA only, and in the middle were genomic promoter sites that were bound by both *GAU1* and PPIA. **B** Peaks showed *GAU1* triplex and PPIA binding at *SENP5* promoter. IgG and non-biotin-labeled samples were set as negative controls. Left: *GAU1* triplex-binding site in tumor cells. Right: PPIA-binding site in tumor cells. **C** Real-time PCR results showed *SENP5* mRNA expression in AGS ($P = 0.0490$), MGC803($P = 0.0024$), KATOIII ($P = 0.0018$) cells, and normal cell GES-1. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. $^{**}P < 0.01$, $^{*}P < 0.05$ compared with normal cell GES-1. **D** Western blot results showed SENP5 protein expression in AGS (lane 2), MGC803(lane 3), KATOIII (lane 4) cells and normal cell GES-1(lane 1). Representative blots from three independent experiments. Source data are provided as a Source Data file. **E** Real-time PCR results showed *SENP5* mRNA expression in *GAU1* knockout gastric tumor cells AGS (KO1 and KO2, $P < 0.0001$), MGC803(KO1 and KO2, $P < 0.0001$), KATOIII (KO1 and KO2, $P < 0.0001$) cells. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. $^{****}P < 0.0001$ compared with empty vector. **F** Western blot results showed significantly decreased SENP5 protein in *GAU1* knockout cells AGS (lanes 2–3), MGC803 (lanes 4–6), and KATOIII (lanes 8–9) as compared to empty vector cells. Representative blots from three independent experiments. Source data are provided as a Source Data file. **G** In the clinical tumor samples of TCGA database, the expression of PPIA was positively correlated with the expression of SENP5; STAD Tumor ($n = 408$ samples) and STAD Normal ($n = 211$ samples).

Supplementary Fig. 9B, circle). When compared to normal GES-1 cells, there were no significant differences in the H3K4me3 levels at site d in MGC803 and KATOIII (Supplementary Fig. 9B, circle). Simultaneously, we did not detect significant changes in H3K27ac at *SENP5* promoter (both site d and e) in gastric tumor cells as compared to normal GES-1 cells (Fig. 5B, square and Supplementary Fig. 9B, square). Meanwhile, there were no significant differences at the negative control (site c) in H3K4me3 and H3K27ac between tumor cells and normal cells (Supplementary Fig. 9C, circle and square). Therefore, we consider choosing H3K4me3 at site e as the key modification for *SENP5* transcriptional activation.

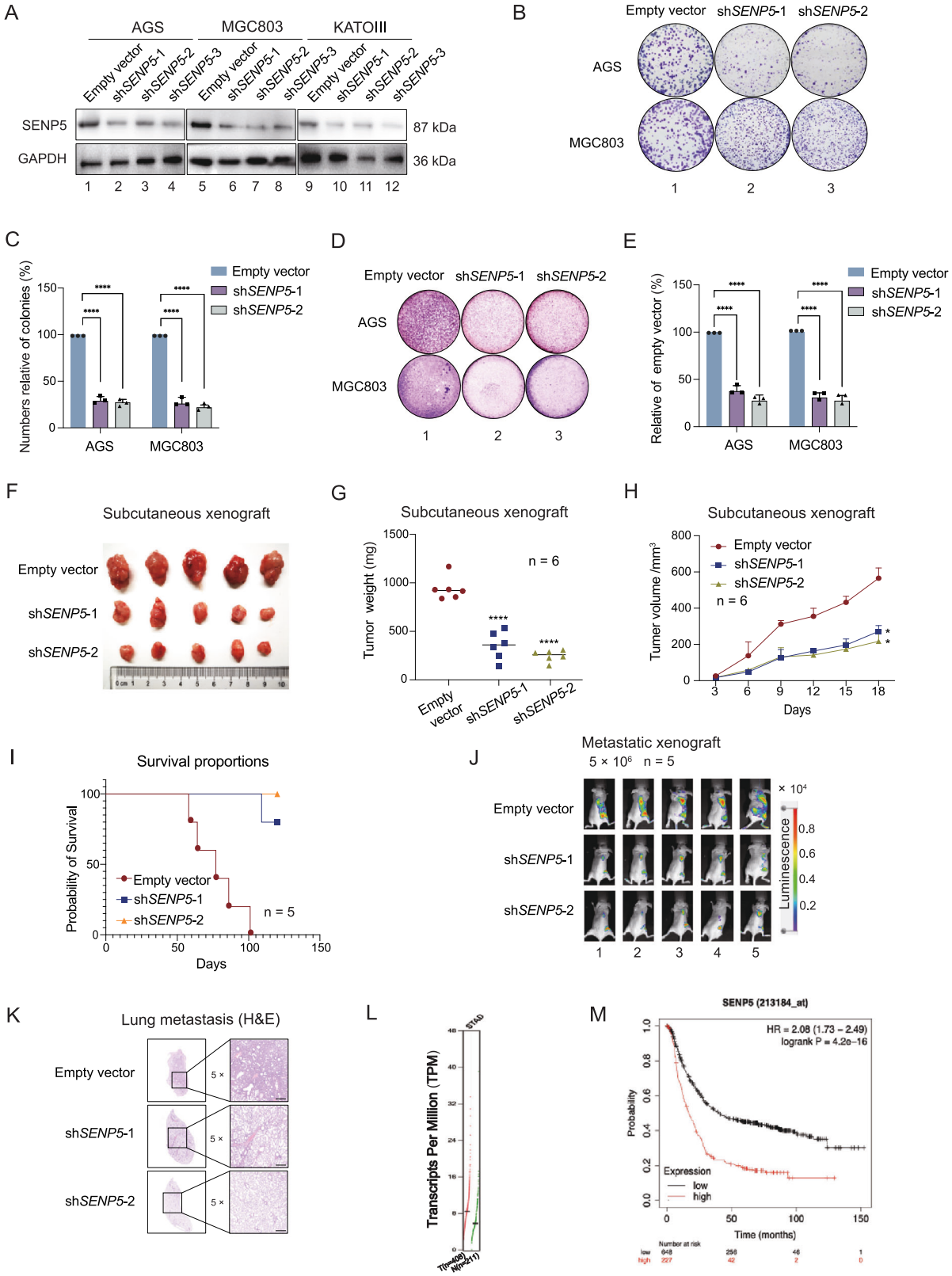
Since above analysis of the enriched proteins from *GAU1*-ChIRP-MS had revealed a significant enrichment of peptide segments from the SET protein family with *GAU1* (see Fig. 2F). We then evaluated the core components of SET1/MLL protein complexes, MLL2 and SET1A, in response to H3K4me3. As expected, we demonstrated an increased binding affinity of SET1A (circle) at the *SENP5* promoter (site e) of gastric tumor cells in contrast to normal GES-1 cells (Fig. 5C, circle). No significant changes in binding affinity were observed for SET1A at the promoter of site d (Supplementary Fig. 9D, circle). In contrast, no significant differences were observed in the binding of MLL2 between gastric cancer and normal cells at the two examined binding sites (Fig. 5C, square and Supplementary Fig. 9D, square). In addition, no significant differences were shown in the binding of SET1A and MLL2 with the negative control site between tumor and normal cells (Supplementary Fig. 9E). Therefore, we are inclined to consider SET1A as the methyltransferase regulating H3K4me3 at the *SENP5* promoter site e.

Subsequently, through ChIRP-western blot analysis, we further validated the interaction between *GAU1* and SET1A (Fig. 5D, lane 2). We then assessed the binding of *GAU1* at the *SENP5* promoter by ChIRP-qPCR and observed that the binding of *GAU1* at both sites of the *SENP5* promoter (site d and site e) were significantly reduced in gastric tumor cells following *GAU1* knockout (Fig. 5E and F, left). In contrast, the binding of positive control (*U2* oligo) and negative control (*lacZ* oligo) at the *SENP5* promoter remained unchanged after *GAU1* knocking out (Fig. 5E, F, middle and right). Simultaneously, we did not observe significant changes in *GAU1* binding at the negative control site (site c) (Supplementary Fig. 9F). Furthermore, we used RIP-qPCR to demonstrate that the interaction of *GAU1* and SET1A were significantly decreased after *GAU1* knocking out (Fig. 5G, circle and triangle). Additionally, the interaction of SET1A and *SENP5* promoter (site e) was significantly reduced after *GAU1* knocking out (Fig. 5H, circle and triangle) in three gastric tumor cells using a ChIP-qPCR assay, while no significant change was observed in the binding of SET1A to another *SENP5* promoter site d and the negative control site (site c) (Supplementary Fig. 9G and Supplementary Fig. 9H). Subsequently, we utilized ChIP-qPCR to examine the enrichment of H3K4me3 at the *SENP5* promoter. We indicated a reduction in H3K4me3 at site e following *GAU1* knocking out (Fig. 5I, circle and triangle), while almost no change in H3K4me3 was observed at site d (Supplementary Fig. 9I). Regardless of *GAU1* silencing, we did not detect a significant difference in the level of H3K4me3 at the negative control site (site c) (Supplementary Fig. 9J). Overall, our results suggest that *GAU1* recruits SET1A to promote the enrichment of H3K4me3 at the *SENP5* promoter in gastric tumor cells.

PPIA determines the formation of *SENP5/GAU1* triplex and SET1A enrichment at *SENP5* promoter

Next, we examined the role of PPIA in the formation of the *SENP5/GAU1* triplex and SET1A enrichment at the *SENP5* promoter. We unexpectedly found that the binding of PPIA to the *SENP5* promoter (Fig. 6A, site d) remained relatively unchanged (Fig. 6A, site d) after *GAU1* knocking out using ChIP-qPCR (Fig. 6B, circle and triangle). Additionally, our analysis revealed specific binding of PPIA to site d, with almost negligible binding to site e, and no significant changes in PPIA binding were observed after *GAU1* knocking out (Supplementary Fig. 10A and Supplementary Fig. 10B). Furthermore, we observed no significant alterations detected at the negative control site (site c) (Supplementary Fig. 10C). Next, we aimed to ascertain whether PPIA alteration would impact the triplex formation between *GAU1* and the *SENP5* promoter. To validate this hypothesis, we employed CRISPR/Cas9 to knockout (KO) PPIA by targeting two sgRNAs to the promoter region of *PPIA* to excise its promoter sequence. We then confirmed the efficacy of knockout by amplifying the gDNA sequence of the *PPIA* promoter through PCR (Fig. 6C). Subsequently, we observed a significant decrease of *PPIA* expression by qPCR assay (Fig. 6D) and western blot assay (Fig. 6E). Next, we performed ChIP-qPCR and demonstrated that the PPIA binding at the *SENP5* promoter (site d) was significantly reduced after *PPIA* knocking out (Fig. 6F, circle and triangle), while no significant changes were observed at another *SENP5* promoter site (site e) and the negative control (site c) (Supplementary Fig. 10D and Supplementary Fig. 10E). As expected, the ChIRP-qPCR analysis showed a substantial decrease in the binding of *GAU1* to both site d and site e on the *SENP5* promoter following *PPIA* knocking out, while no significant change was observed at the negative control (Fig. 6G, H, circle and triangle and Supplementary Fig. 10F). Next, ChIP-qPCR revealed a significant reduction in SET1A binding to the *SENP5* promoter at site e following *PPIA* knocking out (Fig. 6I, circle and triangle). Moreover, no significant changes were observed in SET1A binding at the site d and negative control site c (Supplementary Fig. 10G and Supplementary Fig. 10H). Finally, we validated the role of PPIA in the *SENP5* expression through qPCR and western blot assays. We observed a significant reduction in *SENP5* expression in gastric tumor cells following *PPIA* knocking out (Fig. 6J, K). Taken together, these findings indicate that knockout of *PPIA* diminishes the formation of the *SENP5/GAU1* triplex and SET1A enrichment at *SENP5* promoter, thereby reducing the *SENP5* transcription.

https://doi.org/10.1038/s41467-024-53493-x



Variant exons of *GAU1* interact with PPIA and SET1A respectively

Next, we aimed to precisely identify the specific PPIA and SET1A binding region on *GAU1* lncRNA, we constructed *GAU1* plasmids containing different *GAU1* exons and full length, respectively (Fig. 7A). Subsequently, we stably transfected plasmids containing distinct segments of *GAU1* into *GAU1*-KO gastric tumor cells. Through ChIRP-qPCR analysis, we observed that exon 3 of *GAU1* (star), exon 4 of *GAU1* (inverted triangle), and the full-length *GAU1* (triangle) were able to interact with the *SENP5* promoter. No significant changes were observed at the negative control site (site c) (Fig. 7B and Supplementary Fig. 11). Interestingly, the binding capacity of full-length *GAU1* to the *SENP5* promoter was significantly higher than that of exon 3 or exon 4 alone. Next, we synthesized different exon and full-length RNA of *GAU1* using the T7 RNA kit and incubated with purified PPIA protein

Fig. 4 | *SENP5* is an oncogene in gastric cancer. A Western blot results showed a significant decrease in SENP5 protein in *SENP5* knockdown cells (AGS, MGC803 and KATOIII) as compared to empty vector cells. Representative blots from three independent experiments. Source data are provided as a Source Data file. **B** Cloning formation assay assessed the colony formation ability of empty vector, *shSENP5*-1, and *shSENP5*-2 in gastric tumor cells AGS and MGC803; Representative images from three independent experiments. **C** The colony count in the empty vector group was set as 100%. AGS ($P < 0.0001$), MGC803($P < 0.0001$). Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. ****$P < 0.0001$. **D** Images of migrated tumor cells in empty vector, *shSENP5*-1, and *shSENP5*-2 cells (AGS and MGC803). Representative images from three independent experiments. **E** Quantification of migrated tumor cells in empty vector, *shSENP5*-1, and *shSENP5*-2 cells AGS and MGC803. The empty vector group was set as 100%. AGS ($P < 0.0001$), MGC803($P < 0.0001$). Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. ****$P < 0.0001$. **F** Mice xenograft tumors were photographed. Representative images from five independent samples. **G** Bar graphs showed the weight of allograft tumors in mice treated with empty vector (Circle), *shSENP5*-1 (Square, $P < 0.0001$), and *shSENP5*-1 (Triangle, $P < 0.0001$), respectively. Data are presented as mean ± SD ($n = 6$ mice) using an unpaired two-tailed t-test. ****$P < 0.0001$. **H** Average volume of subcutaneous tumors after treatment with empty vector (Circle), *shSENP5*-1 (Square, $P = 0.0104$), and *shSENP5*-2 (Triangle, $P = 0.0151$) was assessed. Tumor sizes (mm^3) were calculated as the length × width × width/2. Values represent the means ± SD ($n = 6$ mice) using an unpaired two-tailed t-test. *$P < 0.05$. **I** Survival analysis of mice following *shSENP5*-1 (Square), *shSENP5*-2 (Triangle), or empty vector (Circle); $n = 5$ mice. **J** Establishment of gastric cancer metastatic tumor model after *SENP5* silencing. Representative images from five independent samples. **K** Representative images of the histological analysis of lung seeding in mice (original magnification, 5 ×, scale bar: 500 μm). Representative images from five independent samples. Source data are provided as a Source Data file. **L** SENP5 was highly expressed in gastric cancer tissues compared with normal tissues. **M** Correlation between *SENP5* expression and survival rate of gastric cancer patients in TCGA data set survival curves were used to analyze the overall survival rate of patients with high and low *SENP5* expression.

and SET1A protein respectively in vitro. Followed by the pull-down of *GAU1* RNA fragments using in vitro RNA immunoprecipitation, our results revealed that PPIA interacted directly with *GAU1* full-length (circle) and *GAU1* exon 4 (diamond) (Fig. 7C). Moreover, the SET1A protein interacts directly with the *GAU1* full length (circle) and *GAU1* exon 1(square) (Fig. 7D). Employing an in vitro co-IP assay, we observed no direct binding between PPIA and SET1A without *GAU1* (Fig. 7E, lane 2), yet their association was evident upon adding *GAU1* and RNase H (Fig. 7E, lane 3). Conversely, the addition of *GAU1* and RNase A led to the disruption of this binding (Fig. 7E, lane 4). These results suggest that PPIA predominantly interacts with the exon 4 of *GAU1* for recruiting *GAU1* to the *SENP5* promoter, while SET1A is recruited to the *SENP5* promoter through direct interaction with exon 1 of *GAU1*. Significantly, PPIA and SET1A do not display direct binding but rather interact in a *GAU1*-dependent manner, with *GAU1* serving as a molecular intermediary.

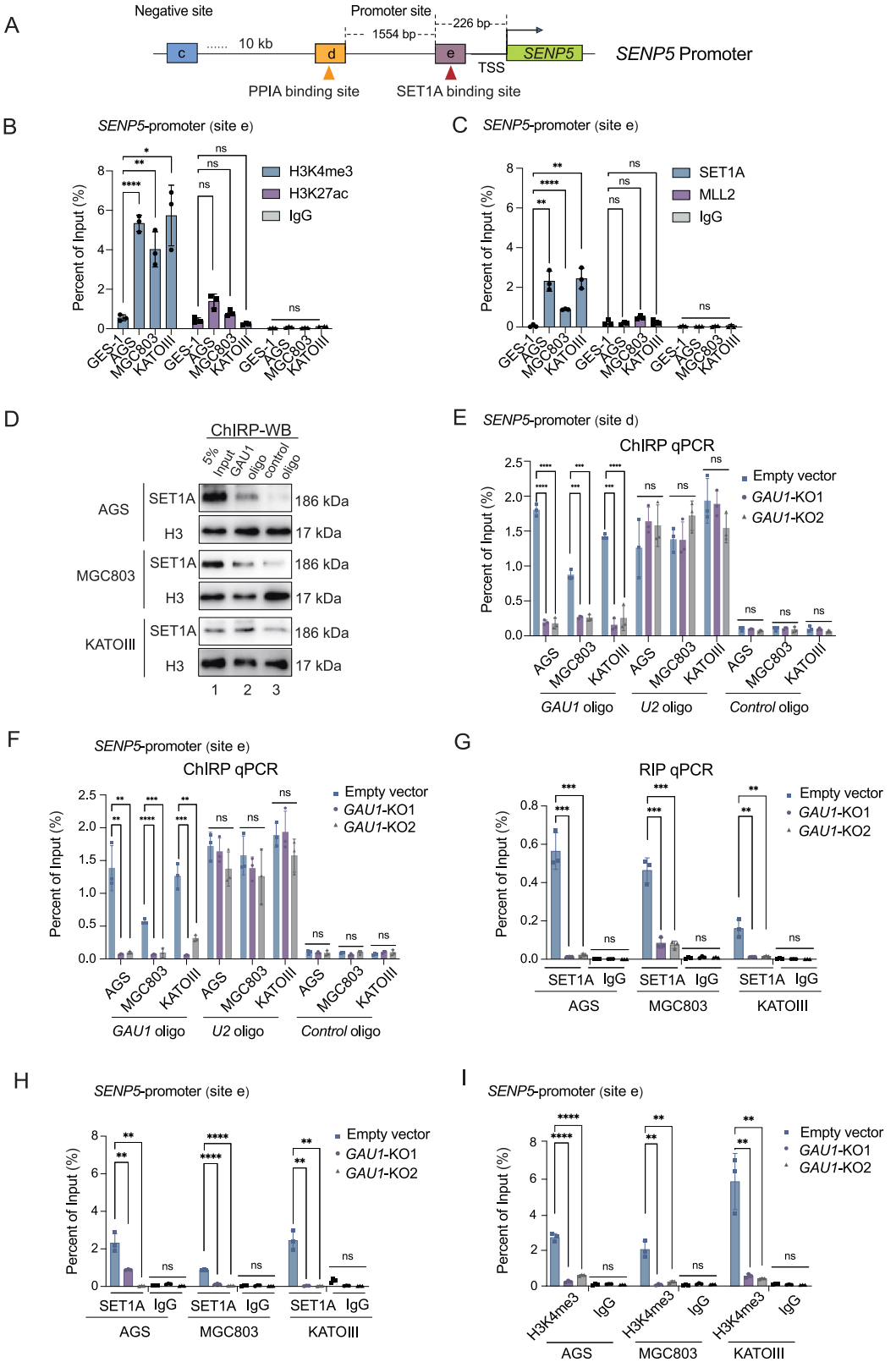
The TFR4-TTS4 is instrumental in the formation of the *SENP5/GAU1* triplex

To precisely identify the interacting regions of the *SENP5/GAU1* triplex, we initially identified the precise sites of *GAU1* triplex formation on the *SENP5* promoter. Employing bioinformatic prediction tools (http://www.gaemons.net/LongTarget), we identified potential Triplex-Forming Regions 1–5 (TFRs 1–5) on *GAU1* that formed a triplex with the *SENP5* promoter (Fig. 7F, Supplementary Fig. 12A and Supplementary Data 2). Subsequently, we proceeded to corroborate the predictive outcomes through repeated validation utilizing the *LongTarget* computational tool (Supplementary Data 3). Subsequently, to validate the actual binding sites between *GAU1* RNA and *SENP5* DNA, we synthesized biotin labeled RNAs containing each of the five predicted TFRs and co-incubated them with a 2 kb DNA fragment of the amplified *SENP5* promoter in vitro, followed by DNA electrophoretic mobility shift assay (EMSA) experiments. Before conducting the assay, we assessed the optimal amount of biotin-labeled *GAU1*-TFRs RNA used in this detection. As depicted in Fig. 7G, we observed that 10 pM of biotin-labeled *GAU1*-TFR4 RNA was sufficient to form the DNA-lncRNA triplex with 1 μg of purified 2 kb-*SENP5* promoter DNA (Fig. 7G, lanes 4–5). However, none of the other TFRs could establish a *SENP5/GAU1* triplex (Supplementary Fig. 12B). These results suggest that *GAU1* TFR4 is the core RNA fragment for the formation of *SENP5/GAU1* DNA-lncRNA triplex.

Next, we further identified the specific Triplex Targeting Sequence (TTS) on the *SENP5* promoter using an in vitro triplex pull-down assay (Supplementary Fig. 12C). Initially, we employed bioinformatics (http://www.gaemons.net/LongTarget) to predict the potential TTS sites on the *SENP5* promoter with the capability to form a DNA-lncRNA triplex (Fig. 7I and Supplementary Data 2 and 3). These sites were found to be non-overlapping with the confirmed PPIA binding site (yellow triangle) and were located between the PPIA binding site and the transcription start site of *SENP5* (Fig. 7H). Subsequently, we co-incubated a biotin labeled TFR4 RNA with a 2 kb *SENP5* promoter DNA fragment. After enzymatic digestion and sonication, we isolated the DNA fragment capable of forming a triplex with TFR4 RNA. We then showed that only TTS4 had the capacity to form a *SENP5/GAU1* triplex with TRF4, while the other TTSs did not exhibit this capability (Fig. 7I, diamond and Supplementary Fig. 12D, lane 4). Additionally, we found that the binding site of PPIA on the *SENP5* promoter was not the site where the *SENP5/GAU1* triplex formed (Fig. 7I, circle). We then obtained the sequence of the TTS4 fragment through deep sequencing (Fig. 7J). After verification, we identified that TTS4 was situated at 1250-nt upstream of the transcription start site of *SENP5*. To further substantiate the ability of *GAU1*-TFR4 and *SENP5*-TTS4 to form DNA-RNA triplexes, we employed EMSAs utilizing RNase H, RNase A, TFR4 mutants, and DNA degradation. After treatment with RNase H, we revealed that the hybridization bands of TFR4-TTS4 triplex formation remained observable (Fig. 7K, lane 3). Conversely, upon treatment with RNase A, we found that these hybridization bands denoting TFR4-TTS4 triplex formation were disappeared (Fig. 7K, lane 2). Moreover, neither the introduction of mutations into TFR4 nor the degradation of DNA allowed for the formation of the TFR4-*SENP5*-promoter triplex structure (Fig. 7K, lane 5 and 6). Next, we performed thermal melting assays and showed that melting temperatures of DNA-RNA triplex Tm (red) was 80.78 ± 0.22 °C. Moreover, RNA-DNA heteroduplex and DNA-DNA duplex Tm were 82.89 ± 0.22 °C (gray), and 84.76 ± 0.32 °C(black), respectively (Fig. 7L). We also used CD-spectroscopy to confirm triplex formation of *GAU1* TFR4 and *SENP5* TTS4. The CD spectrum indicated typical features for triplex formation including a positive small peak at 220 nm, two negative peaks at 210 nm and 240 nm, and a blue shift of the peak at 270 nm (Fig. 7M, red). These typical peaks of triplex were distinct from the *SENP5* DNA duplex (Fig. 7M, black) or the heteroduplex spectra (Fig. 7M, grey). These results indicate that TTS4 of *SENP5* promoter and TFR4 of *GAU1* can form a *SENP5/GAU1* DNA-lncRNA triplex.

To further verify whether fragments that form the *SENP5/GAU1* triplex in vitro could also form triplex under physiological conditions in vivo, we performed an in vivo triplex capture assay (Fig. S12E). In this assay, we transferred five biotin-labeled *GAU1* TFR oligos into *GAU1*-silenced gastric tumor cells. Subsequently, we enriched and purified DNA fragments capable of forming *SENP5/GAU1* triplex using standard DNA-RNA triplex capture assay and analyzed them via qPCR. As anticipated, our in vivo experiments confirmed that only *GAU1* TFR4 could form a triplex with the TTS4 site of the *SENP5* promoter in AGS cell (Fig. 7N, diamond and Supplementary Fig. 12F, lane 4). Similarly, we observed the same phenomenon in MGC803 and KATOIII cells

https://doi.org/10.1038/s41467-024-53493-x



(Fig. 7O, P, diamond, and Supplementary Fig. 12F, lane 4). These results suggest that TFR4 fragment at exon 3 of *GAU1* can form a *SENP5/GAU1* DNA-lncRNA triplex with the TTS4 fragment.

Discussion

DNA-RNA triplexes participate in the regulation of various biological functions by either facilitating or inhibiting gene transcription[4,21]. LncRNA-mediated DNA-RNA triplexes in tumors are beginning to be revealed[22]. However, the mechanism and functional characterization of DNA-lncRNA triplexes in tumorigenesis have not been thoroughly explored. In this study, we revealed a PPIA-guided *SENP5/GAU1* DNA-lncRNA triplex driving gastric cancer tumorigenesis. PPIA served as an anchor to recruit *GAU1* lncRNA and enhance the stability of the *SENP5/GAU1* triplex, triggering the recruitment of the methyltransferase

Fig. 5 | *GAU1* recruits SET1A to increase the H3K4 trimethylation of *SENP5* promoter. A The schematic diagram of the *SENP5* promoter region and primer locations (Supplementary Table 1). ChIP detection sites include negative site c, site d (identified in PPIA-ChIP-seq), and site e (SET1A binding site). **B** ChIP analysis was conducted to examine histone modifications of H3K4me3 and H3K27ac at the *SENP5* promoter (site e) in AGS ($P < 0.0001$), MGC803($P = 0.0027$), KATOIII ($P = 0.0122$) and GES-1. ChIP enrichment was presented as the percentage of input signal. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. ****$P < 0.0001$, **$P < 0.01$, and *$P < 0.05$. **C** ChIP analysis of SET1/MLL protein at the *SENP5* promoter (site e) was conducted in AGS ($P = 0.0014$), MGC803 ($P < 0.0001$), KATOIII ($P = 0.0013$) and GES-1. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. ****$P < 0.0001$ and **$P < 0.01$. **(D)** ChIRP-Western blot was used to verify the interaction between *GAU1* and SET1A in AGS, MGC803, KATOIII. *GAU1* oligo and control (lacZ) oligo pulldown were utilized for all proteins interacting with *GAU1* and lacZ, respectively. Representative blots from three independent experiments. Source data are provided as a Source Data file. **E** ChIRP analysis of *GAU1* binding at the *SENP5* promoter (site d) was conducted in AGS (KO1, $P < 0.0001$; KO2, $P < 0.0001$), MGC803(KO1, $P = 0.0001$; KO2, $P = 0.0002$), KATOIII (KO1, $P = 0.0001$; KO2, $P = 0.0004$) after *GAU1* KO. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. ****$P < 0.0001$and ***$P < 0.001$. **F** ChIRP analysis of *GAU1* binding at the *SENP5* promoter (site e) was conducted in AGS (KO1, $P = 0.0027$; KO2, $P = 0.0029$), MGC803(KO1, $P < 0.0001$; KO2, $P = 0.0005$), KATOIII (KO1, $P = 0.0005$; KO2, $P = 0.0013$) after *GAU1* KO. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. ****$P < 0.0001$, ***$P < 0.001$, and **$P < 0.01$. **G** Real-time PCR analysis was performed to assess the binding of *GAU1* to SET1A using samples(AGS: KO1, $P = 0.0006$; KO2, $P = 0.0006$; MGC803: KO1, $P < 0.0007$; KO2, $P = 0.0006$; KATOIII: KO1, $P = 0.0049$; KO2, $P = 0.0054$) from the RNA-ChIP assay after *GAU1* KO. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. ***$P < 0.001$ and **$P < 0.01$. **H** ChIP analysis of SET1A protein at the *SENP5* promoter (site e) was conducted in AGS (KO1, $P = 0.0081$; KO2, $P = 0.0014$), MGC803(KO1, $P < 0.0001$; KO2, $P < 0.0001$), KATOIII (KO1, $P = 0.0013$; KO2, $P = 0.0013$) after *GAU1* KO. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. ****$P < 0.0001$ and **$P < 0.01$. **I** ChIP analysis of histone modifications of H3K4me3 at the *SENP5* promoter (site e) was conducted in AGS (KO1, $P < 0.0001$; KO2, $P < 0.0001$), MGC803(KO1, $P = 0.0017$; KO2, $P = 0.0024$), KATOIII (KO1, $P = 0.0043$; KO2, $P = 0.0039$) after *GAU1* KO. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. ****$P < 0.0001$ and **$P < 0.01$.

SET1A to the *SENP5* promoter, thereby leading to an enrichment of H3K4me3 and the activation of *SENP5* transcription for tumorigenesis (Fig. 7Q). Our study reveals a mechanism of PPIA-guided *SENP5/GAU1* triplex formation in the tumorigenesis of gastric tumors and illustrates a precise model in the dynamics of DNA-lncRNA triplex.

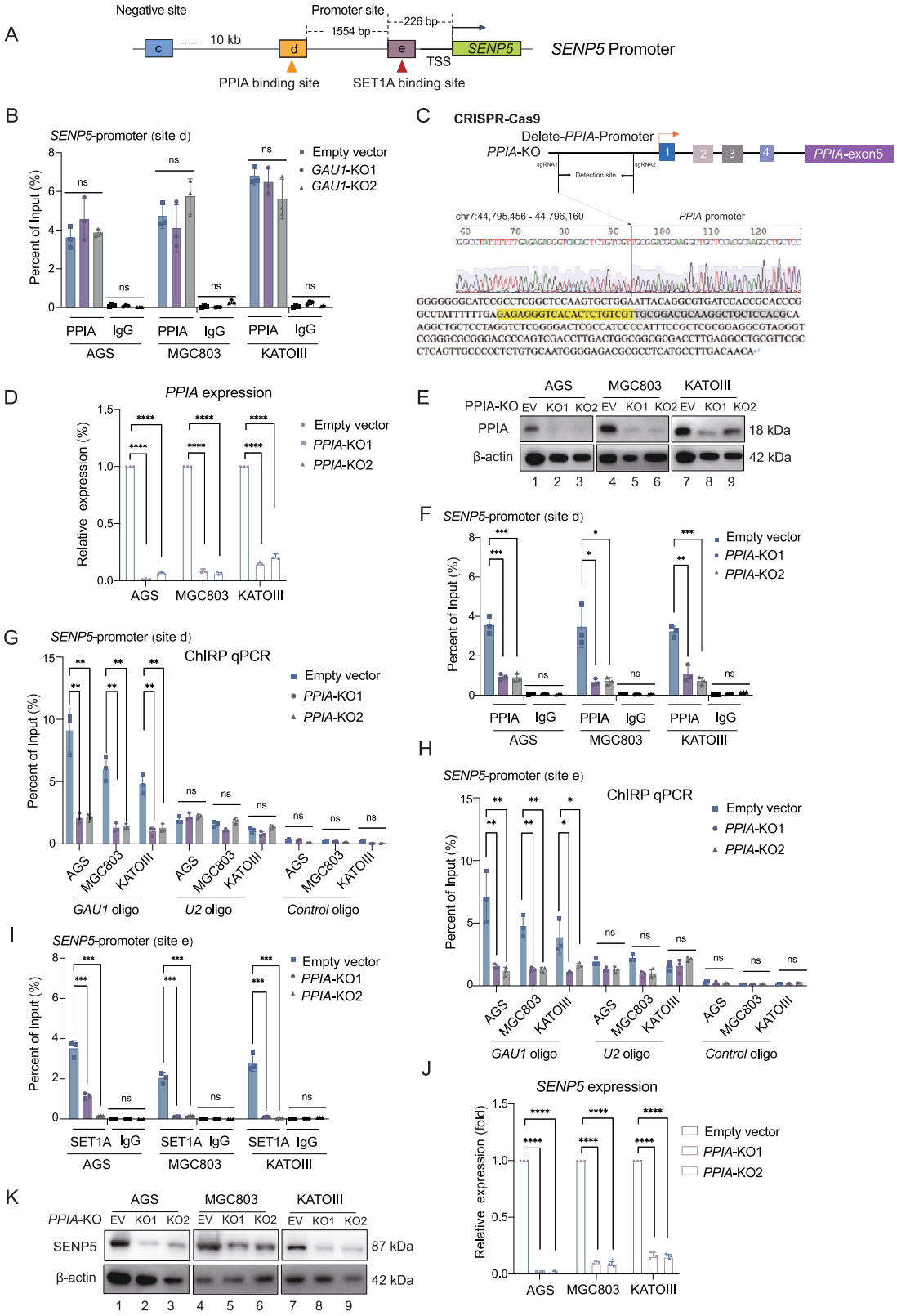
It is imperative to note that the formation of DNA-RNA triplexes hinges on Hoogsteen hydrogen bonding, rendering them significantly less stable than double-stranded DNA, owing to their non-canonical pairing[22]. Consequently, beyond the base sequence, the regulatory factors governing the formation of DNA-RNA triplex remain largely elusive. Previous studies have primarily focused on the influence of pH fluctuations on DNA-RNA triplex formation[23]. pH variations could impact the protonation or deprotonation of specific functional groups within nucleotides, potentially altering hydrogen bonding patterns and overall triplex stability[24,25]. Additionally, histones not only affect hydrogen bonding between DNA and RNA but also interact with the bases within the DNA-RNA triplex, influencing overall structural stability[26]. Chemical modifications, such as methylation on DNA and RNA molecules can alter their complementarity and structure, thereby influencing the formation of triplexes[27,28]. However, to the best of our knowledge, there has been limited in-depth research investigating the potential role of proteins in the formation and stabilization of DNA-RNA triplexes. In this study, we identified PPIA as a key peptidylprolyl isomerase protein regulating the formation of the *SENP5/GAU1* triplex. It should also be further emphasized that PPIA, also known as Cyclophilin A (CyPA), belongs to the protein family with peptidyl-prolyl cis-trans isomerase (PPIase) activity, catalyzing the conversion of prolyl peptide bonds between cis and trans conformations which was first discovered in 1984[29]. Recent research has revealed that PPIA, in addition to its cis-trans isomerase function, also serves an additional role as a nuclear molecular chaperone[30]. Although our report is only one case of a PPIA protein involved in the stabilization of a DNA-lncRNA triplex, it cannot be theoretically ruled out that other proteins are involved in the formation and stabilization of DNA-lncRNA triplexes in either a positive or negative manner. Certainly, we cannot exclude the possibility that PPIA may function as a broad regulatory factor involved in the formation of DNA-RNA triplexes. Therefore, it will be interesting to focus on gaining a deeper understanding of whether PPIA functions as a broad regulator of triplex formation and to identify other proteins to better understand the role of DNA-lncRNA triplex-protein interactions.

It should be noted that while the TFR and TTS sequences capable of forming DNA-RNA triplexes are generally based on complementary base pairing, they do not strictly adhere to the principles of complementary base pairing[31,32]. Therefore, the TFR and TTS sequences that can form DNA-RNA triplexes possess a certain degree of uniqueness and versatility[33]. Recent DNA-RNA triplex capture experiments, EMSA, melting assays, CD spectroscopy have greatly facilitated the elucidation of DNA-RNA triplexes[16,31,34–36]. Through bioinformatics analysis and DNA-RNA triplex capture experiments, our study demonstrated that *GAU1*-TFR4 could form DNA-RNA triplexes with *SENP5*-TTS4 and served as critical element for driving tumorigenesis. To the best of our knowledge, the *GAU1*-TFR4/*SENP5*-TTS4 triplex has not been previously reported. Furthermore, we have showed that the TFRs of *GAU1* has the potential to form DNA-RNA triplexes with other genomic loci by using bioinformatic analysis. Thus, it would be of great interest to investigate whether these predicted triplexes could regulate tumorigenesis through a similar or unknown mechanism.

The classic regulatory mechanisms employed by lncRNA entail their function as molecular scaffolds, orchestrating the recruitment or exclusion of proteins to coordinate the precise localization of transcription factors onto target chromatin[37]. LncRNA *MREF* interacts with Smarca5 to promote chromatin accessibility upon activation and differentiation of muscle satellite cells, thereby promoting the genomic binding of p300/CBP/H3K27ac and the expression of myogenic regulatory factors such as MyoD and cellular differentiation[38]. LncRNA *NORAD* upholds genomic stability by sequestering PUMILIO proteins, thereby facilitating mitosis and DNA replication[39]. However, the extent to which proteins may influence the genomic localization of lncRNA remains an area yet to be exhaustively elucidated. In this study, we showed that PPIA regulated the enrichment of *GAU1* lncRNA to the *SENP5* promoter. This discovery diverges from the conventional concept that lncRNA serves as scaffolds to recruit functional proteins. Instead, it introduces a paradigm for understanding the genomic localization of lncRNA, underscoring the pivotal role of proteins as intermediary factors in recruiting functional lncRNA.

It should be noted that *GAU1* is a lncRNA that we identified and characterized in our previous study[20]. However, the intricate mechanisms underlying the transcriptional activation of *GAU1* remain elusive. In this study, we identified the p300 mediating H3K27 acetylation modification on the *GAU1* promoter as a key regulatory determinant of *GAU1* activation. It should also be emphasized that we cannot theoretically exclude that other chromatin modifications are involved in *GAU1* activation. Therefore, it would be great interesting to focus on identifying other causes to better understand *GAU1* activation. Furthermore, we did not proceed with the construction of *GAU1* knockout (KO) mice, as despite the high homology rate of 96%

https://doi.org/10.1038/s41467-024-53493-x



between the mouse and human *GAU1* lncRNA, there is still a portion of the sequence that is not conserved. Consequently, there remains some potential risk in constructing KO mice for this study. Nonetheless, the generation of *GAU1* knockout models would be of great interest to further investigate the biological role of *GAU1* in tumorigenesis for our future research.

SUMOization is widely present in a variety of crucial physiological and pathological processes[40,41]. This process is reversible and mediated by the SENP (Sentrin/SUMO-specific protease) family, which can remove SUMO from conjugated substrates[42]. Among the key members of the SUMO system, SENP5 plays a vital role in cell division[43,44]. It has been reported that SENP5 is implicated in the regulation of various

Fig. 6 | PPIA determines the formation of *SENP5/GAU1* triplex and SET1A enrichment at *SENP5* promoter. A The schematic diagram of the *SENP5* promoter region and primer locations (Supplementary Table 1). Negative site c and *SENP5* promoter site d (PPIA binding site) were designated as ChIP detection sites. **B** ChIP analysis of PPIA protein at the *SENP5* promoter (site d) in AGS, MGC803, KATOIII after *GAU1* KO. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test, ns: no significance. **C** The schematic diagram illustrates the *PPIA* knockout site. Two sgRNAs were used to knock out the *PPIA* promoter. Sequencing results and peak maps of gDNA fragments after knockout can also be seen in the illustration. The sgRNA sequence information is available in Supplementary Table 1. **D** Real-time PCR showed the expression of *PPIA* mRNA after CRISPR/Cas9 knockout of the *PPIA* gene in AGS (KO1, $P < 0.0001$; KO2, $P < 0.0001$), MGC803(KO1, $P < 0.0001$; KO2, $P < 0.0001$), and KATOIII (KO1, $P < 0.0001$; KO2, $P < 0.0001$) gastric tumor cells. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. $^{****}P < 0.0001$ compared with empty vector. **E** Western blot results showed that SENP5 protein decreased significantly in *SENP5* knockdown cells AGS, MGC803 and KATOIII as compared to empty vector cells. Representative blots from three independent experiments. Source data are provided as a Source Data file. **F** ChIP analysis of PPIA protein at the *SENP5* promoter (site d) was conducted in AGS (KO1, $P = 0.0006$; KO2, $P = 0.0007$), MGC803(KO1, $P = 0.0105$; KO2, $P = 0.0112$), KATOIII (KO1, $P = 0.0020$; KO2, $P = 0.0002$) after PPIA KO. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. $^{***}P < 0.001$, $^{**}P < 0.01$ and $^{*}P < 0.05$. **G** ChIRP analysis the binding of *GAU1* in the *SENP5* promoter (site d) with *GAU1* oligo pulldown was assessed in AGS (KO1, $P = 0.0024$; KO2, $P = 0.0023$), MGC803 (KO1, $P = 0.0013$; KO2, $P = 0.0012$), KATOIII (KO1, $P = 0.0014$; KO2, $P = 0.0019$) after *PPIA* KO. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. $^{**}P < 0.01$. **H** ChIRP analysis the binding of *GAU1* in the *SENP5* promoter (site e) with *GAU1* oligo pulldown was assessed in AGS (KO1, $P = 0.0097$; KO2, $P = 0.0081$), MGC803(KO1, $P = 0.0027$; KO2, $P = 0.0029$), KATOIII (KO1, $P = 0.0230$; KO2, $P = 0.0476$) after *PPIA* KO. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test, $^{**}P < 0.01$ and $^{*}P < 0.05$. **I** ChIP analysis of PPIA protein at the *SENP5* promoter (site e) was performed in AGS (KO1, $P = 0.0006$; KO2, $P = 0.0001$), MGC803(KO1, $P = 0.0002$; KO2, $P = 0.0002$), KATOIII (KO1, $P = 0.0004$; KO2, $P = 0.0003$) after *PPIA* KO. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. $^{***}P < 0.001$. **J** Real-time PCR showed the expression of *SENP5* mRNA after PPIA CRISPR/Cas9 knockout in gastric tumor cells AGS (KO1, $P < 0.0001$; KO2, $P < 0.0001$), MGC803(KO1, $P < 0.0001$; KO2, $P < 0.0001$), and KATOIII (KO1, $P < 0.0001$; KO2, $P < 0.0001$). Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test. $^{****}P < 0.0001$ compared with empty vector. **K** The Western blot results showed that as compared to cells with empty vector, the expression level of SENP5 protein decreased in gastric tumor cells AGS, MGC803, and KATOIII after *PPIA* knockout. Representative blots from three independent experiments. Source data are provided as a Source Data file.

cancers, including breast cancer, osteosarcoma, oral squamous cell carcinoma, and hepatocellular carcinoma[45,46]. However, in this study, we have elucidated the oncogenic role of *SENP5* in gastric cancer. Furthermore, we have unveiled that SET1A mediating H3K4 trimethylation is required for *SENP5* transcriptional activation in gastric cancer. Further studies should focus on identifying other causes involving in *SENP5* activation.

It is well known that lncRNAs play a wide range of epigenetic regulatory roles by binding to the genome mainly indirectly through interactions with RNAs and proteins or directly by forming hybrids with DNA[47,48]. The indirect binding mode of lncRNAs is well known, but the role of DNA-RNA hybrids, especially DNA-RNA triplexes, has only been revealed in recent years[1,3,49]. A comprehensive genome-wide analysis of triplex-forming sites revealed that many potential triplex-forming sites in *GAU1* are distributed throughout the genome. Through bioinformatic predictions, we discovered that across the entire genome, there were 48,086 sites harboring the potential to form DNA-RNA triplexes with *GAU1*, of which 39,684 sites specifically exhibit the capacity to engage in triplex formation with TFR4. However, this finding was not unexpected as the mystery of DNA-RNA triplex has been gradually revealed with the development of DNA-RNA triplex genome-wide detection technology[21,31]. In addition to the lncRNA *GAU1*, lncRNA such as *HIF1α-AS1, HOTAIR, FENDRR*, and *PARTICLE* exhibit numerous DNA-RNA triplex formation sites across the entire genome, suggesting that the triplex may be a generalized mode of gene expression, chromatin remodeling, and regulation of disease development[16,50,51]. These studies prompted us to reevaluate the mode of genomic localization of lncRNA and propose a model for their involvement in gene expression across the entire genome. Here, we proposed a double localization model in which lncRNA co-localizes in the genomic region by both lncRNA-protein interaction and lncRNA-DNA triplex formation. This model significantly enhances the stability of *GAU1* binding at the *SENP5* promoter. To our knowledge, this represents the case of a lncRNA being positioned at the genomic locus participating in gene regulation with simultaneous help of lncRNA interacting protein and lncRNA forming triplex. It is also emphasized that lncRNA mediating modes of gene regulation are diverse. Therefore, focusing on identifying additional lncRNA-genomic DNA interacting patterns will be an intriguing avenue for gaining deeper insights into lncRNA mediating mechanisms of gene regulation.

Methods

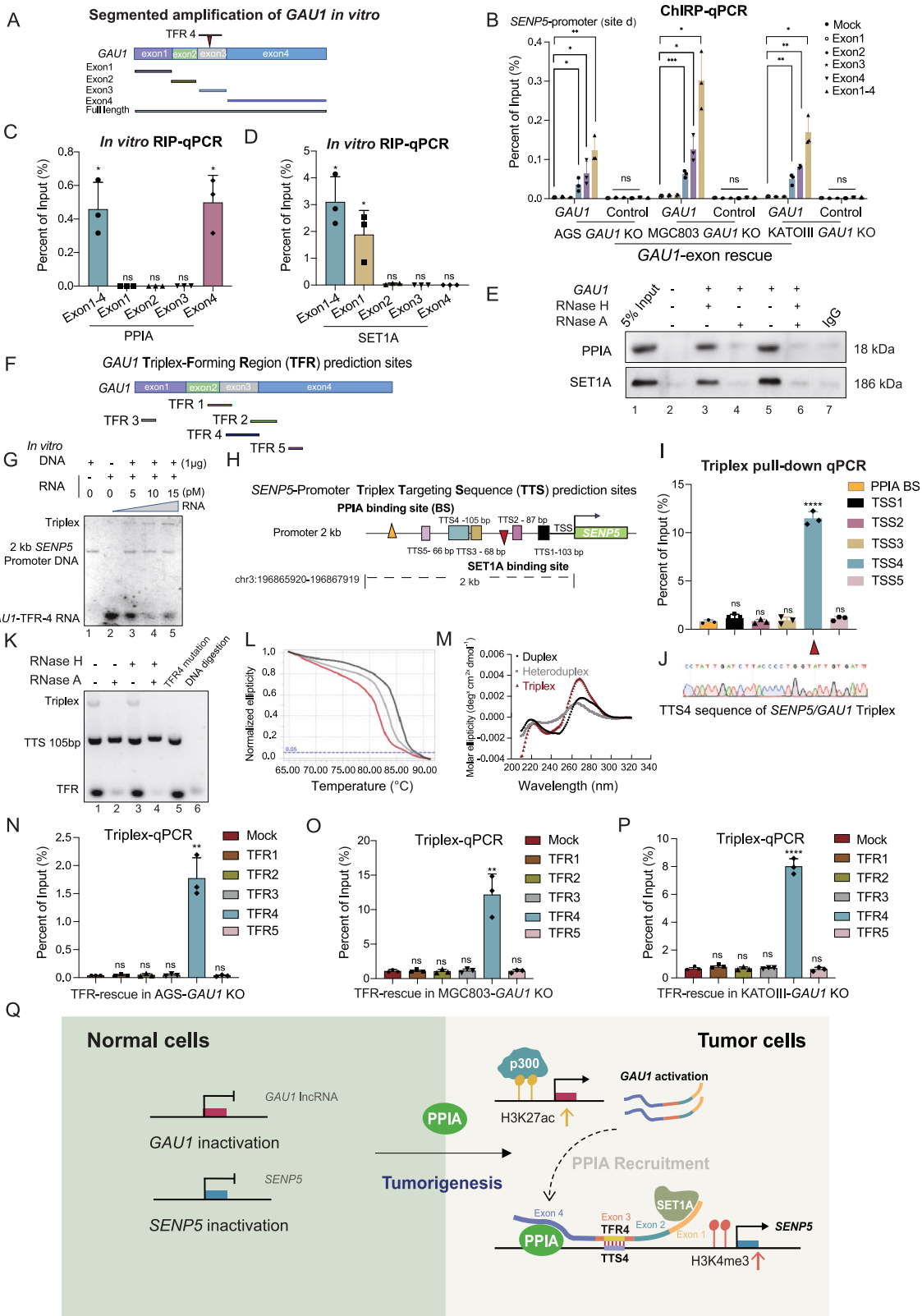
Ethical compliance statement

This research strictly adheres to all relevant ethical regulations and guidelines. The study protocol was reviewed and approved by the Institutional Animal Care and Treatment Committee of Tongji University, ensuring that all procedures involving animals were conducted in accordance with the highest ethical standards. The study protocol was approved by the College of Life Science and Technology of Tongji University. All animal studies were conducted in accordance with the guidelines provided by the Institutional Animal Care and Treatment Committee of Tongji University. The ethical approval and experimental protocol number is TJAB04823101. The animals were treated following relevant institutional and national guidelines and regulations. The maximum allowable tumor size/burden (with a diameter of less than 1 cm) was not exceeded. Mice were housed at an ambient temperature of 24 ± 2 °C, with circulating air, constant humidity of 50 ± 10%, and a 12-hour light/dark cycle. Female BALB/c nude mice (GemPharmatech), specifically five weeks of age, were selected for this study. The minimum number of animals in a single experiment is 5, and the number of animals used is detailed in the main article and figure legend. The decision to use female mice was made after careful consideration and analysis, which indicated that the gender of the mice was not a confounding factor in the study's findings. Extensive preliminary data supported the conclusion that gender-related biological variables did not significantly alter the experimental results or the interpretation of the data. Therefore, the use of female mice was justified to ensure the consistency and reproducibility of the study's observations and conclusions. Mice were euthanized by cervical dislocation at the end of the experiment, after which tumors and major organs were collected using surgical scissors.

Cell culture

The human AGS (Caucasian, female, 54 years adult, ATCC, CRL-1739), MGC803(Asian, male, 58 years adult, Shanghai cell Bank, CBP60485), GES-1(Asian, female, 22 years adult, Shanghai cell Bank, CBP60512) and 293 T(ATCC, CRL-3216) cell lines were cultured in DMEM (Sigma–Aldrich, Wisconsin, GER) supplemented with 10% certified heat-inactivated fetal bovine serum (Yeasen, Shanghai, China), penicillin (100 U/mL), and streptomycin (100 mg/mL) at 37 °C in a humidified 5% CO_2 atmosphere, while KATOIII(Asian, Male, 55 years adult, gastric cancer with pleural effusion, ATCC, HTB-103) was cultured in 1640.

https://doi.org/10.1038/s41467-024-53493-x



RNA-FISH

The cells were fixed with 4% formaldehyde and stored overnight in 70% ethanol. The fluorescence-labeled single-strand probes were synthesized and then hybridized. *GAU1*, *U2* and *GAPDH* oligos were purchased from BGI. To increase the stability of RNA foci, RNA signals were detected with a tyramide-Alexa Fluor 546 signal amplification kit (Beyotime, Shanghai, China). After labeling, fluorescence signals were detected using a confocal microscope (FV1000; Olympus).

CRISPR/Cas9

Four single guide RNAs (sgRNAs) were separately cloned into lenti-CRISPRv2 (Addgene #52961, https://www.addgene.org/52961/) plasmids. To delete the *GAU1* exon2-exon3 and *GAU1* exon4 region from

Fig. 7 | The TFR4-TTS4 is instrumental in the formation of the *SENP5/GAU1* triplex. A Schematic illustration of *GAU1* fragment for in vitro amplification. TRF4 was contained in exon 3 of *GAU1*. **B** Different *GAU1* exons were rescued in *GAU1*-knockout gastric tumor cells AGS (Exon3, $P = 0.0318$; Exon4, $P = 0.0403$; Exon1–4, $P = 0.0061$), MGC803(Exon3, $P = 0.0004$; Exon4, $P = 0.0231$; Exon1–4, $P = 0.0198$), and KATOIII(Exon3, $P = 0.0053$; Exon4, $P < 0.0012$; Exon1– 4, $P = 0.0173$), and CHIRP-qPCR was used to detect the binding of different fragments of *GAU1* in the *SENP5* promoter. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test, $^{****}P < 0.0001$, $^{***}P < 0.001$, $^{**}P < 0.01$, and $^{*}P < 0.05$. **C** Real-time PCR was used to detect the binding of purified PPIA protein to *GAU1* RNA (Exon1–4, $P = 0.0390$; Exon4, $P = 0.0338$) of different exons and full-length *GAU1* RNA transcribed in vitro. PPIA antibodies were used for RIP. Data are presented as mean ± SD from three independent experiments. One sample t and Wilcoxon test of Row Stats, $^{*}P < 0.05$. **D** Real-time PCR was used to detect the binding of purified SET1A protein to *GAU1* RNA (Exon1–4, $P = 0.0296$; Exon4, $P = 0.0478$) of different exons and full-length *GAU1* RNA transcribed in vitro. SET1A antibodies were used for RIP. Data are presented as mean ± SD from three independent experiments. One sample t and Wilcoxon test of Row Stats, $^{*}P < 0.05$. **E** In vitro co-IP assay shows the interaction between PPIA and the SET1A. Groups were delineated as: Control (no *GAU1*), *GAU1* + RNase H, *GAU1* + RNase A, *GAU1* alone, and *GAU1* + RNase A + RNase H. Representative blots from three independent experiments. Source data are provided as a Source Data file. **F** Prediction sites for *GAU1* triplex forming regions (TFR) were provided. **G** EMSA was conducted to detect the interaction between the *SENP5* promoter and *GAU1* TFR. Representative images from three independent experiments. Source data are provided as a Source Data file. **H** Prediction sites for *SENP5* promoter triplex targeting sequences (TTS) were outlined. **I** Triplex pull-down qPCR was used to detect *SENP5* promoter DNA fragments that could form *SENP5/GAU1* triplexes with *GAU1* TRF4(TSS4, $P < 0.0001$) in vitro. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test, $^{****}P < 0.0001$. **J** It was shown that the TTS4 sequence can form a triplex. **K** EMSA was utilized to probe the interaction between the *SENP5*-TTS4 and *GAU1*-TFR4. Representative images from three independent experiments. Source data are provided as a Source Data file. **L** Thermal melting assay of the *SENP5* DNA duplex (black), the heteroduplex (gray) and *SENP5*-TTS4:*GAU1*-TFR4 triplex (red). **M** Circular dichroism spectra of the *SENP5* DNA duplex (black), the heteroduplex (gray) and *SENP5*-TTS4:*GAU1*-TFR4 triplex (red) measured at 25 °C. **N**–**P** Real time PCR was used to detect the formation of biotin labeled TRFs in AGS (TFR4, $P = 0.0012$), MGC803(TFR4, $P = 0.0032$) and KATOIII (TFR4, $P < 0.0001$) gastric tumor cells with *GAU1* KO. Data are presented as mean ± SD from three independent experiments using an unpaired two-tailed t-test, $^{****}P < 0.0001$, and $^{**}P < 0.01$. **Q** Schematic diagram of the research model. PPIA served as an anchor to recruit *GAU1* lncRNA and enhance the stability of the *SENP5/GAU1* triplex, triggering the recruitment of the methyltransferase SET1A to the *SENP5* promoter, thereby leading to an enrichment of H3K4me3 and the activation of *SENP5* transcription for tumorigenesis.

the genome, AGS, MGC803, and KATOIII cells were transfected with plasmids containing single guide RNAs (and Cas9) targeting the left or right side of the region to be deleted. Colonies were derived from single cells and tested for the loss of the deletion region. The control group was transfected with the sgRNA-empty vector using Lipofectamine 3000 (Invitrogen, Carlsbad, USA) according to the manufacturer's instructions.

Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) assay

Total RNA in AGS, MGC803, and KATOIII were extracted by Trizol (Invitrogen, Carlsbad, USA), RNAs were reversed-transcribed using qPCR SYBR Green Master Mix (Yeasen, Shanghai, China) and analyzed with a LightCycle96 quantitative PCR system (Roche, Basel, CH). The relative mRNA expression level of genes was calculated using the $2^{-\Delta\Delta Ct}$ method after normalization to *GAPDH*, which served as an internal loading control. To detect the expression levels of *GAU1, SENP5*, we designed qPCR-primers, which were synthesized by Sangon (Sangon, Shanghai, China), according to the manufacturer protocol. The primer sequences are in the Supplemental Table 1: Primers, oligos, sgRNAs used in the experiment.

Colony formation assay

The cells were counted and adjusted to 1000 cells per well in 6-well plates and then placed into a humidified incubator for 1 weeks. Subsequently, the cells were washed twice in phosphate-buffered saline (PBS), fixed with 70% methanol for 30 min and stained for 30 min with crystal violet. The colonies were counted in assays. Finally, GraphPad prism8.0 was used to analyze the number of cell colonies in each well.

Western blot assay

Cells were collected at the specified time and rinsed twice with PBS. The cell extracts were prepared with lysis buffer and centrifuged at 13,000 g for 30 min at 4°C and protein concentration was quantified with a BCA kit (Yeasen, Shanghai, China). The supernatant was collected and 5×SDS protein loading buffer was added in proportion and incubated at 100°C for 10 min. Proteins were separated on a 10–12% SDS-PAGE and electrophoretically transferred to polyvinylidene difluoride (PVDF) membranes (Millipore, Massachusetts, USA). Membranes were then blocked in 5% non-fat milk with Tris-buffered saline and 0.1% Tween 20 at room temperature for 2 h and the antibody was diluted with 1×TBST according to the instructions and incubated overnight at 4°C. Membranes were subsequently incubated with secondary antibodies for another 2 h at room temperature. Immunoblots were visualized by an enhanced chemiluminescence detection kit (ECL kit; Yeasen, Shanghai, China) under a chemiluminescence imaging analysis system (Tanon, Shanghai, China).

Nuclear fractionation

Cells were collected at the specified time and washed twice with PBS. Subsequently, 10^7 cells were harvested; resuspended in 1 mL of ice-cold DEPC-PBS, 1 mL of buffer C1 (1.28 M sucrose, 40 mM Tris-HCl [pH 7.5], 20 mM $MgCl_2$, 4% Triton X-100), and 3 mL of RNase-free water, and incubated for 15 min on ice. Then, cells were centrifuged for 15 min at 3000 rpm; The resulting supernatant containing the cytoplasmic component and the pellet containing the nuclear fraction were both retained for RNA extraction.

Lentivirus packaging and generation of stable cell lines

The Lipofectamine 3000 reagent (Invitrogen, Carlsbad, USA) was incubated with Opti-MEM I Reduced Serum Medium (GIBCO, the USA) and used to transfect 239 T cells with 3 mg of the target plasmids (pLentiCRISPR V2 plasmid) and 3.0 mg of the PMD plasmid and 6.0 mg of the PsPAX2 plasmid. Six hours after transfection, the medium was replaced with 10 mL of fresh medium. The supernatant containing the viruses was collected at 48 and 72 h. The virus-containing solution was filtered and concentrated. Add 50 µL of the virus carrying Cas9 and sgRNA to 1 mL of serum-free, antibiotic-free medium. Twenty-four hours prior to transfection, tumor cells were seeded at 2.0×10^5 cells per well in a six-well plate. The medium was replaced with virus-containing supernatant supplemented with 0.5 ng/mL polybrene (Sigma-Aldrich, Wisconsin, GER). After 48 h, the medium was replaced with fresh medium. Cells were selected by incubation with 4 mg/mL puromycin (InvivoGen, French) for 2 weeks. Colonies were selected and expanded for further analysis.

Tumor xenograft and metastasis model in nude mice

The cells were harvested by trypsinization and washed twice with PBS (Gibco, USA). BALB/c nude mice (female, 5 weeks old) were used for the study. The minimum number of animals in a single experiment is 5. The ethical approval and experimental protocol number is TJAB04823101. The xenograft model of MGC803-KO or MGC803 empty vector cell tumor was established by injecting 2×10^7 cells into the right armpit of female nude mice. All mice were sacrificed by

cervical dislocation 30 days after implantation for the tumor formation analysis or 120 days after injection for the survival analysis. Tumor size was measured every 3 days with a vernier caliper. For the metastasis model, 5-week-old nude female mice were deeply anesthetized. Cells were pre-transfected with a lentivirus encoded by the plvx-luciferase -Blasticidin vector (MGC803 *GAU1*-empty vector; MGC803 *GAU1*-KO1; MGC803 *GAU1*-KO2; MGC803 *shSENP5*-empty vector; MGC803 *shSENP5*−1and MGC803 *shSENP5*−2 sgRNA2). Next, 5×10^6 cells in a 0.1 mL volume of sterile saline solution were injected through the caudal vein. Bioluminescence was detected after 40 days by in vivo small animal imaging systems, the numbers of nude mice presented with stronger luminescence signal in empty vector groups have been calculated as compared with *GAU1*-KO or *shSENP5* group. Animal experiments were carried out in accordance with the guidelines of SPF Animal Center of Tongji University. We confirm that the maximal tumor size/burden did not exceed the limits permitted (tumor volume: 1 cm^3) by our ethics committee or institutional review board.

In vivo triplex capture assay with biotinylated RNA

Biotinylated RNA was segmented synthesis in vitro, and the cells were grown on six-well plates at 50–60% cell confluency. Lipofectamine 3000 reagent (Invitrogen, Carlsbad, USA) assisted transfection of biotin-labeled target 10 pmol RNA or negative control. Then the medium was changed after 8 h, and the cells continued to grow for 24 h before harvest. The cells were washed 3 times with cold PBS, and the cells were scraped off with a disposable scraper. After centrifugation, cells are lysed on ice with cell lysis buffer (40 mM Tris-acetate, pH 7.5, 20 mM KCl, 50 mM magnesium−acetate, 50% glycerol) containing RNase inhibitor for 15 min. After centrifugation, the supernatant is removed. Then incubate on ice with nuclei lysis buffer (10 mM KCl, 10 mM Hepes, pH 8.0, 5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, ribonuclease inhibitors. Ice-cold) containing RNase inhibitor for 10 min. After the pellet was suspended by Triplex buffer (10 mM Tris-HCl, pH 7.5, 20 mM KCl, 10 mM $MgCl_2$, ribonuclease inhibitors), Proteinase K was added, and the pretreated protein was incubated at 37 °C for 30 min. Then the DNA-RNA mixture was broken with sonicate, and the fragment was about 500 bp. RNase H was added to treat the samples, and the R-loop interference was removed after incubation at 37 °C for 30 min. Add Streptavidin C1 Dynabeads and incubate samples at 37 °C for 2 h. Subsequently, the beads were washed five times with 1 mL of precooled wash buffer (150 mM KCl, 10 mM Tris−HCl, pH 7.5, 5 mM $MgCl_2$, 0.5% NP-40, ribonuclease inhibitors) for 5 min per wash. Then, 100 µL of elution buffer (10 mM Tris-HCl, pH 7.5, 20 mM KCl, 10 mM $MgCl_2$, RNase A) was added, and the protein was eluted at 37 °C for 1 h. The supernatant was transferred to a new low-binding tube. And the beads were eluted again with 100 µL of elution buffer. The two supernatants were combined. After sodium acetate precipitated samples, used 20 µL TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 7.5) resuspended pellet, Furthermore, DNA was subjected to sequencing analysis.

In vitro triplex capture assay with biotinylated RNA

The 5′-untranslated region transcription start site (TTS) fragment of the genome SENP5 was designed, and the PCR products were digested by exonuclease. Next, 1 µg PCR-fragments were incubated with 5–15 pmol biotin-labeled *GAU1* triplex-forming oligonucleotide (TFO) probe in triplex hybridization solution for 30 min at 37 °C. PCR-fragments without *GAU1* TFO sequence served as an empty vector, and that without RNA served as a blank control. RNA-DNA complexes were bound to streptavidin-coated beads at 37 °C for 40 min, washed 3 times with buffer containing 150 mmol/L KCl, 10 mmol/L Tris-HCl (pH 7.5), 5 mmol/L $MgCl_2$, 0.5% NP-40 and RNasin, and washed once with buffer containing 15 mmol/L KCl, 10 mmol/L Tris-HCl (pH 7.5) and 5 mmol/L $MgCl_2$. RNA-bound DNA was eluted with a mixture of 1% sodium dodecyl sulfate, 50 mmol/L Tris-HCl (pH 8.0), and 10 mmol/L ethylene diamine tetra acetate (EDTA) at 65 °C for 5 min. They were digested with RNase A (50 ng/mL, at 37 °C for 30 min) and protease K (200 ng/mL, at 15 °C for 15 min). The recovered DNA was analyzed by qPCR and normalized to input DNA. The *SENP5* TTS sequences are in Supplemental Table 2: TFOs and TTSs.

Electrophoretic mobility shift assay (EMSA)

The DNA and the labeled synthesized *GAU1* TFO were reacted in 10 µL binding reaction system (the buffer contained 1 mmol/L $MgCl_2$. 0.5 mmol/L EDTA, 0.5 mmol/L dithiothreitol, 50 mmol/L NaCl, 10 mmol/L Tris-HCl (pH 7.5), and 0.05 mg/mL poly-(dI-dC)) at room temperature for 20 min. Subsequently, the samples were loaded onto a 4% denaturing polyacrylamide gel for electrophoresis. The corresponding bands indicative of triplex formation could be visualized through staining. The strength of gel bands was quantitatively analyzed by imaging analysis system (Tanon, Shanghai, China). The *GAU1* TFO sequences are in Supplemental Table 2: TFOs and TTSs.

Circular dichroism (CD) and melting curve analysis

Circular dichroism spectra were obtained using an Applied Photophysics (Leatherhead, Surrey, UK) Chirascan polarimeter. The measurement range was from 210 to 320 nm, and the spectra were recorded at 25 °C using a 1 cm path length quartz cuvette. Each DNA duplex, DNA-RNA heteroduplex, and DNA-DNA-RNA triplex sample at 8 µM was measured in 10 mM Tris-HCl (pH 7.4), 50 mM KCl, and 5 mM $MgCl_2$ (pH 7.4). Spectra were acquired through eight scans and data were smoothed using a Savitzky-Golay filter. Observed ellipticity values were converted to molar ellipticity [θ] = deg x cm^2 x $dmol^{-1}$. Melting curves were obtained at a constant wavelength within the temperature range of 65 °C to 95 °C, with a heating rate of 1 °C/min. All melting temperature data were converted to normalized ellipticity and evaluated using the following equation.

Chromatin isolation by RNA purification (ChIRP)

Biotin labeled probes against *GAU1* were designed according to the online tool (www.singlemoleculefish.com). Probes against *LacZ* RNA were used as the negative control. Probes against *U2* used as positive control. A total of 2×10^7 AGS, MGC803, KATOIII cells were resuspended in precooled PBS buffer and crosslinked with 3% formaldehyde and the reaction was quenched by glycine. Cells were then pelleted at 1000 g for 10 min and resuspended in Swelling Buffer containing 0.1 M Tris pH7.0, 10 mM KOAc, 15 mM MgOAc. 1% NP-40, 1 mM DTT, 1 mM PMSF, complete protease inhibitor, and 0.1U/µL Superase-in for 10 min on ice. Nuclei was further resolved in 50 mM Tris 7.0, 10 mM EDTA, 1% SDS, add DTT, PMSF and Superase-in. The chromatin was then sonicated to 100–500 bp size and diluted in 500 mM NaCl, 1% SDS, 100 mM Tris 7.0, 10 mM EDTA, 15% formamide. Pre-binding probes oligos were added to streptavidin beads. The beads were mixed with the cell lysate and hybridized at 37 °C overnight on an end-to-end shaker. Subsequently, the beads were washed five times with 1 mL of prewarmed wash buffer for 5 min per wash. Then, 100 µL of elution buffer containing (50 mM $NaHCO_3$, 1%SDS, 200 mM NaCl, 20U RNase A and 20U RNase H) was added, and the mixtures were eluted at 37 °C for 1 h. Subsequently, the supernatant was collected, 10 µL protease K (Thermo Fisher, Massachusetts, USA) was added and incubated at 55 °C for 2 h. DNA was then extracted with phenol: chloroform: isoamyl and precipitated with ethanol at −80 °C. Furthermore, purified DNA was subjected to qPCR analysis. The primers sequences are in Supplemental Table 1: Primers, oligos, sgRNAs used in the experiment

Chromatin isolation by RNA purification mass spectrometry (ChIRP-MS)

In ChIRP-MS, we conducted a mass spectrometry assay with one biological replicate using AGS-NC, AGS-KO, positive control *U2* oligo, and negative control *lacZ* oligo. For the identification of proteins from gel

strips, samples are first separated through gel electrophoresis, and strips are excised at distinct positions on the gel. Following enzymatic digestion, peptides are extracted and subjected to desalting before being lyophilized. These peptides are then reconstituted in mobile phase A (2% acetonitrile, 0.1% formic acid), centrifuged at 20,000 g for 10 min, and the supernatant is injected into a Thermo UltiMate 3000 UHPLC system. Separation occurs on a self-packed C18 column using a gradient program at a flow rate of 300 nL/min, which includes increasing mobile phase B (98% acetonitrile, 0.1% formic acid) from 5% to 25% over 40 min, further to 35% in 5 min, up to 80% over another 2 min, maintaining this for 2 min, and returning to initial conditions over the last 6 min. The eluted peptides are ionized by a nanoESI source and analyzed by a tandem mass spectrometer like the Q-Exactive HF X, operating in data-dependent acquisition (DDA) mode with key parameters set to optimize detection, including ion source voltage at 1.9 kV, MS1 scan range from 350 to 1,500 m/z, resolution at 60,000 for MS1 scans, and 15,000 for MS2 scans. Fragmentation targets the top 30 most intense ions with charges 2^+ to 6^+ and exceeds a threshold of 10,000 counts, using HCD mode with fragment ions detected in the Orbitrap, and dynamic exclusion set to 30 sec. Automatic gain control (AGC) targets are set to 3E6 for MS1 and 1E5 for MS2 scans. Finally, protein identification software is utilized to analyze the acquired mass spectra and identify the proteins present in the samples.

RNA-chromatin immunoprecipitation (RIP)

RNA immunoprecipitation was performed following Part of the Methods in Molecular Biology book series (MIMB, volume 1480) Protocol. AGS, MGC803 and KATOIII cells at 80–90% confluency in 15 cm dish were scraped off, then lysed in complete RIP lysis buffer, after which 100 µL of whole-cell extract was incubated with RIP buffer containing magnetic beads conjugated with human anti-SET1A (D3V9S) Rabbit mAb #61702 (1:50, CST, Mass, USA); anti-PPIA CyPA Antikörper (6-YD13): sc-134310(1:50, Santa Cruz, USA); anti-IgG Rabbit (DA1E) mAb IgG XP® Isotype Control #3900(1:50, CST, Mass, USA); anti-Histone H3 (D1H2) XP® Rabbit mAb #4499. Samples were incubated with Proteinase K with shaking to digest the protein and then immunoprecipitated RNA was isolated. Furthermore, purified RNA was subjected to qRT-PCR analysis to demonstrate the presence of the binding targets using respective primers. The primers sequences are in Supplemental Table 1: Primers, oligos, sgRNAs used in the experiment.

Chromatin immunoprecipitation (ChIP)

ChIP was conducted using an EZ-Magna ChIP A/G kit (Millipore, Massachusetts, USA) according to the manufacturer instructions. The anti-H3K27ac, Acetyl-Histone H3 (Lys27) (D5E4) XP® Rabbit mAb #8173 (1:50, CST, Mass, USA); anti-H3K4me, Mono-Methyl-Histone H3 (Lys4) (D1A9) XP® Rabbit mAb #5326 (1:50, CST, Mass, USA); anti-SET1A (D3V9S) Rabbit mAb #61702 (1:50, CST, Mass, USA); anti-PPIA CyPA Antikörper (6-YD13): sc-134310(1:50, Santa Cruz, USA) antibody used for RNA-ChIP was also applied for ChIP. Anti IgG was used as a negative control. The primers sequences are Supplemental Table 1: Primers, oligos, sgRNAs used in the experiment.

TCGA dataset

To validate the potential role of *SENP5* in GC, we queried the TCGA (http://www.cbioportal.org), TIMER2.0 (http://timer.cistrome.org/) and GEPIA (gepia.cancer-pku.cn), R2 provided the transcriptional landscape and follow-up information on 1222 gastric cancer samples.

Statistical analysis

All experiments were performed in triplicate, and the data were expressed as mean ± SD. The comparative CT method was applied in the quantitative real-time RT-PCR assay according to the delta-delta CT method. The data were analyzed with t-test or by one-way analysis of variance, and results were considered statistically significant at $P \le 0.05$. If $P < 0.05$, it is considered to have statistical difference, it is represented by '*'; If $P < 0.01$, it is indicated by '**'; If $P < 0.001$, it is indicated by '***'; If $P < 0.0001$, it is indicated by '****'.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The ChIP-seq and Triplex-seq data generated in this study are available at the Gene Expression Omnibus GSE247292 and the Sequence Read Archive (SRA) BioProject PRJNA1037069. The ChIRP-MS data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD056041. Source data are provided with this paper.

References

1. Petermann, E., Lan, L. & Zou, L. Sources, resolution and physiological relevance of R-loops and RNA-DNA hybrids. *Nat. Rev. Mol. Cell Biol.* **23**, 521–540 (2022).
2. Leisegang, M. S., Warwick, T., Stotzel, J. & Brandes, R. P. RNA-DNA triplexes: molecular mechanisms and functional relevance. *Trends Biochem Sci.* **49**, 532–544 (2024).
3. Warwick, T. et al. A universal model of RNA.DNA:DNA triplex formation accurately predicts genome-wide RNA-DNA interactions. *Brief Bioinform* **23** https://doi.org/10.1093/bib/bbac445 (2022).
4. Xu, Y., McSally, J., Andricioaei, I. & Al-Hashimi, H. M. Modulation of Hoogsteen dynamics on DNA recognition. *Nat. Commun.* **9**, 1473 (2018).
5. Li, Y., Syed, J. & Sugiyama, H. RNA-DNA triplex formation by long noncoding RNAs. *Cell Chem. Biol.* **23**, 1325–1333 (2016).
6. Buske, F. A., Bauer, D. C., Mattick, J. S. & Bailey, T. L. Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res* **22**, 1372–1381 (2012).
7. Goni, J. R., de la Cruz, X. & Orozco, M. Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res* **32**, 354–360 (2004).
8. Jalali, S., Singh, A., Scaria, V. & Maiti, S. Genome-wide computational analysis and validation of potential long noncoding RNA-Mediated DNA-DNA-RNA triplexes in the human genome. *Methods Mol. Biol.* **2254**, 61–71 (2021).
9. Wu, Q. et al. High-affinity triplex-forming oligonucleotide target sequences in mammalian genomes. *Mol. Carcinog.* **46**, 15–23 (2007).
10. Warwick, T., Brandes, R. P. & Leisegang, M. S. Computational Methods to Study DNA:DNA:RNA Triplex Formation by lncRNAs. *Noncoding RNA,* **9** https://doi.org/10.3390/ncrna9010010 (2023).
11. Xu, H. et al. Single-cell profiling of long noncoding RNAs and their cell lineage commitment roles via RNA-DNA-DNA triplex formation in mammary epithelium. *Stem Cells* https://doi.org/10.1002/stem.3274 (2020).
12. Liang, Y. et al. Identification of Long Noncoding RNAs That Exert Transcriptional Regulation by Forming RNA-DNA Triplexes in Prostate Cancer. *Int. J. Mol. Sci.* **24**, https://doi.org/10.3390/ijms24032035 (2023).
13. Schmitz, K. M., Mayer, C., Postepska, A. & Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* **24**, 2264–2269 (2010).
14. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445**, 666–670 (2007).
15. Kalwa, M. et al. The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. *Nucleic Acids Res* **44**, 10631–10643 (2016).

16. Leisegang, M. S. et al. HIF1alpha-AS1 is a DNA:DNA:RNA triplex-forming lncRNA interacting with the HUSH complex. *Nat. Commun.* **13**, 6563 (2022).
17. Zhang, X. L. et al. KCNQ1OT1 promotes genome-wide transposon repression by guiding RNA-DNA triplexes and HP1 binding. *Nat. Cell Biol.* **24**, 1617–1629 (2022).
18. Pan, H. et al. Chromosomal instability-associated MAT1 lncRNA insulates MLL1-guided histone methylation and accelerates tumorigenesis. *Cell Rep.* **41**, 111829 (2022).
19. Postepska-Igielska, A. et al. LncRNA Khps1 regulates expression of the proto-oncogene SPHK1 via triplex-mediated changes in chromatin structure. *Mol. Cell* **60**, 626–636 (2015).
20. Chai, P. et al. Dynamic chromosomal tuning of a novel GAU1 lncing driver at chr12p13.32 accelerates tumorigenesis. *Nucleic Acids Res* **46**, 6041–6056 (2018).
21. Cecconello, A., Magro, M., Vianello, F. & Simmel, F. C. Rational design of hybrid DNA-RNA triplex structures as modulators of transcriptional activity in vitro. *Nucleic Acids Res.* **50**, 13172–13182 (2022).
22. Han, J. J. LncRNAs: the missing link to senescence nuclear architecture. *Trends Biochem Sci.* **48**, 618–628 (2023).
23. Li, T., Ackermann, D., Hall, A. M. & Famulok, M. Input-dependent induction of oligonucleotide structural motifs for performing molecular logic. *J. Am. Chem. Soc.* **134**, 3508–3516 (2012).
24. Gehring, K., Leroy, J. L. & Gueron, M. A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature* **363**, 561–565 (1993).
25. Choi, J. & Majima, T. Conformational changes of non-B DNA. *Chem. Soc. Rev.* **40**, 5893–5909 (2011).
26. Maldonado, R., Schwartz, U., Silberhorn, E. & Langst, G. Nucleosomes stabilize ssRNA-dsDNA triple helices in human cells. *Mol. Cell* **73**, 1243–1254.e1246 (2019).
27. Tabe-Bordbar, S. & Sinha, S. Integrative modeling of lncRNA-chromatin interaction maps reveals diverse mechanisms of nuclear retention. *BMC Genomics* **24**, 395 (2023).
28. Ogunleye, A. J., Romanova, E. & Medvedeva, Y. A. Genome-wide regulation of CpG methylation by ecCEBPalpha in acute myeloid leukemia. *F1000Res* **10**, 204 (2021).
29. Hadpech, S. & Thongboonkerd, V. Current update on theranostic roles of cyclophilin A in kidney diseases. *Theranostics* **12**, 4067–4080 (2022).
30. Wang, P. & Heitman, J. The cyclophilins. *Genome Biol.* **6**, 226 (2005).
31. Senturk Cetin, N. et al. Isolation and genome-wide characterization of cellular DNA:RNA triplex structures. *Nucleic Acids Res* **47**, 2306–2321 (2019).
32. Kandimalla, E. R. & Agrawal, S. Single strand targeted triplex-formation. Destabilization of guanine quadruplex structures by foldback triplex-forming oligonucleotides. *Nucleic Acids Res* **23**, 1068–1074 (1995).
33. Renneberg, D. & Leumann, C. J. Exploring hoogsteen and reversed-hoogsteen duplex and triplex formation with tricyclo-DNA purine sequences. *Chembiochem* **5**, 1114–1118 (2004).
34. Liquier, J. et al. Spectroscopic studies of chimeric DNA-RNA and RNA 29-base intramolecular triple helices. *Nucleic Acids Res* **23**, 1722–1728 (1995).
35. Maldonado, R. & Langst, G. Analyzing RNA-DNA triplex formation in chromatin. *Methods Mol. Biol.* **2161**, 247–254 (2020).
36. Postepska-Igielska, A., Blank-Giwojna, A. & Grummt, I. Analysis of RNA-DNA Triplex Structures In Vitro and In Vivo. *Methods Mol. Biol.* **2161**, 229–246 (2020).
37. Kopp, F. & Mendell, J. T. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**, 393–407 (2018).
38. Lv, W. et al. Long noncoding RNA lncMREF promotes myogenic differentiation and muscle regeneration by interacting with the Smarca5/p300 complex. *Nucleic Acids Res* **50**, 10733–10755 (2022).
39. Lee, S. et al. Noncoding RNA NORAD regulates genomic stability by sequestering pumilio proteins. *Cell* **164**, 69–80 (2016).
40. Mevissen, T. E. T. & Komander, D. Mechanisms of deubiquitinase specificity and regulation. *Annu Rev. Biochem* **86**, 159–192 (2017).
41. Hendriks, I. A. et al. Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation. *Nat. Struct. Mol. Biol.* **24**, 325–336 (2017).
42. Chang, H. M. & Yeh, E. T. H. SUMO: from bench to bedside. *Physiol. Rev.* **100**, 1599–1619 (2020).
43. Di Bacco, A. et al. The SUMO-specific protease SENP5 is required for cell division. *Mol. Cell. Biol.* **26**, 4489–4498 (2006).
44. Mukhopadhyay, D. & Dasso, M. Modification in reverse: the SUMO proteases. *Trends Biochemical Sci.* **32**, 286–295 (2007).
45. Jin, Z. L., Pei, H., Xu, Y. H., Yu, J. & Deng, T. The SUMO-specific protease SENP5 controls DNA damage response and promotes tumorigenesis in hepatocellular carcinoma. *Eur. Rev. Med Pharm.* **20**, 3566–3573 (2016).
46. Ding, X. J. et al. Overexpression of SENP5 in oral squamous cell carcinoma and its association with differentiation. *Oncol. Rep.* **20**, 1041–1045 (2008).
47. Yao, Z. T. et al. New insights into the interplay between long non-coding RNAs and RNA-binding proteins in cancer. *Cancer Commun.* **42**, 117–140 (2022).
48. Qian, X. Y., Zhao, J. Y., Yeung, P. Y., Zhang, Q. C. & Kwok, C. K. Revealing lncRNA structures and interactions by sequencing-based approaches. *Trends Biochemical Sci.* **44**, 33–52 (2019).
49. Park, M. K. et al. NEAT1 is essential for metabolic changes that promote breast cancer growth and metastasis. *Cell Metab.* **33**, 2380–2397.e9. (2021).
50. O'Leary, V. B. et al. PARTICLE triplexes cluster in the tumor suppressor WWOX and may extend throughout the human genome. *Sci. Rep.* **7**, 7163 (2017).
51. Kuo, C. C. et al. Detection of RNA-DNA binding sites in long non-coding RNAs. *Nucleic Acids Res* **47**, e32 (2019).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 82372705 and 31870748) to H.Z., the Shanghai Oriental Elite Project (Grant No. 2000152009) to H.Z., the National Key Research and Development Program of China (Grant No. 2017YFE0196300) to H.Z., the Shanghai Natural Science Foundation (Grant No. 22ZR1466100) to H.Z., the Fundamental Research Funds for the Central Universities (Grant No. 22120230292) to H.Z., the Key Laboratory of Organ Development and Epigenetics of Jiangxi Province (Grant No. 2024SSY07141) to H.Z., the Shuguang Project of Shanghai Municipal Education Commission and Shanghai Education Development Foundation (Grant No. 17SG19) to H.Z., the Outstanding Young Medical Scholar of Shanghai Municipal Commission of Health and Family Planning (Grant No. 2017YQ067) to H.Z., the Outstanding Yong Scholar Grant of Tongji University (Grant No. PA2019000239) to H.Z., and the Startup Funding of Frontier Science Research Center for Stem Cells & Shanghai East Hospital of Tongji University (Grant No. DFRC2019003) to H.Z.

Author contributions

X.Z.: Conceptualization, Investigation, Methodology, Experiment, Data curation, Writing – original draft, Writing – review & editing. T.D.: Conceptualization, Investigation, Methodology, Experiment, Writing – original draft, Writing – review & editing. F.Y.: Validation. J.Z.: Methodology. H.X.: Resources. Y.S.: Investigation. Y.B.: Investigation. J.Y.: Methodology. C.C.: Data curation. C.Z.: Resources. H.Z.: Project administration, Funding acquisition, Conceptualization, Supervision, Writing – original draft. All authors have read and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53493-x.

Correspondence and requests for materials should be addressed to He Zhang.

Peer review information *Nature Communications* thanks Ji-Fan Hu, Matthias Leisegang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.