Open Access Full Text Article

ORIGINAL RESEARCH

Long noncoding RNA THAP9-AS1 is induced by *Helicobacter pylori* and promotes cell growth and migration of gastric cancer

Wenxiao Jia[1,*]
Jiaqin Zhang[2,*]
Fang Ma[1]
Shengjie Hao[1]
Xue Li[1]
Ruiting Guo[1]
Qianqian Gao[1]
Yundong Sun[1]
Jihui Jia[1]
Wenjuan Li[1]

[1]Key Laboratory for Experimental Teratology of Chinese Ministry of Education, The Shandong Provincial Key Laboratory of Infection and Immunology, Department of Microbiology, School of Basic Medical Sciences, Shandong University, Jinan, People's Republic of China; [2]Department of Clinical Laboratory, The First Affiliated Hospital of Xiamen University, Xiamen, People's Republic of China

*These authors contributed equally to this work

Correspondence: Wenjuan Li
Key Laboratory for Experimental Teratology of Chinese Ministry of Education, The Shandong Provincial Key Laboratory of Infection and Immunology, Department of Microbiology, School of Basic Medical Sciences, Shandong University, No. 44, Wenhuaxi Road, Jinan, Shangdong Province, People's Republic of China
Tel +86 5 318 838 2579
Fax +86 5 318 838 2502
Email wenjli@sdu.edu.cn

Background: Long noncoding RNAs (LncRNAs) have been confirmed to play crucial roles in cancer biology. Gastric cancer (GC) is the third leading cause of cancer related death, and *Helicobacter pylori* (*H. pylori*) is the major risk factor for GC. In this study, we focused on the roles of *H. pylori*-related lncRNAs in the progression of GC.

Method: Differentially expressed lncRNAs were identified through RNA-seq analysis of *H. pylori*-infected GC cells.

Results: We found that the expression of the lncRNA THAP9-AS1 was up-regulated after infection of GC cells with *H. pylori* and was higher in GC tissues than in gastritis tissues. Colony formation, CCK8 and transwell assays were executed to show that THAP9-AS1 can promote GC cell proliferation and migration in vitro. Our study identified the pro-oncogenic lncRNA THAP9-AS1, which has a higher expression level in GC tissues than in gastritis tissues and which promoted the proliferation and migration of GC cells in vitro.

Conclusion: These findings may provide a potential therapeutic target for *H. pylori*-associated GC.

Keywords: long noncoding RNA, THAP9-AS1, gastric cancer, *Helicobacter pylori*

Introduction

Gastric cancer (GC) is the fifth most common cancer and the third leading cause of cancer-related death in the world. Approximately 700,000 people succumb to this malignancy each year.[1] Although multifactorial pathways are involved in the development of GC, the role of *Helicobacter pylori* (*H. pylori*) infection in GC cannot be ignored.[2,3] The relationships between *H. pylori* and chronic gastritis and peptic ulcers were found by Warren and Marshall in 1983.[4] *H. pylori* is one of the most common human pathogens worldwide, infecting an estimated 50% of the global population; approximately 90% of new cases of noncardia GC worldwide are attributable to *H. pylori* infection.[1] Consequently, *H. pylori* infection is considered strongly associated with gastritis, peptic ulcer disease, mucosa-associated lymphoid tissue lymphoma (MALT) and GC in most countries but not for some Asian countries such as India.[5–9] However, the key molecular mechanism linking *H. pylori* to GC has not been determined.

It has been reported that nearly 75% of the human genome is transcribed into RNA, but less than 2% is translated into protein, and the remaining RNA is called noncoding RNA (ncRNA).[10] Long noncoding RNAs (lncRNAs) are defined as

RNA transcripts that are longer than 200 nt but have no significant protein-coding capacity, and they have been found to play a vital role as regulators of biological functions.[11,12] Accumulating evidence also suggests that lncRNAs play an important role in cell proliferation, differentiation, migration, apoptosis and the immune response through their effects on DNA, RNA and protein expression and interactions.[13–19] Some recent publications showed that the aberrant expression of some lncRNAs, such as SAMMSON, HOTAIR, and GAPLINC, may be a diagnostic biomarker and therapeutic target for melanoma, pancreatic cancer and GC.[16,20–22] Many products of protein-coding genes, such as CEA, HER2 and EGFR, have been investigated widely and function as diagnostic markers and therapeutic targets in the clinic.[23–25] Up to now, the early diagnosis and treatment of GC have been a global problem, and thus, we need a better understanding of the expression pattern of lncRNAs in GC cells. In view of the important cause of GC, we explored the expression patterns of lncRNAs in *H. pylori*-infected GC cells.

In this study, we performed RNA-seq analysis of *H. pylori*-infected AGS cells, and PBS-treated cells were used as a control group. We identified an expression pattern of lncRNAs and mRNAs in *H. pylori*-infected AGS cells. We determined that the expression of THAP domain containing 9 antisense RNA 1 (THAP9-AS1), a long noncoding RNA that contains six exons and is located on 4q21.22, was induced by *H. pylori* infection, which plays an important role in the proliferation and migration of GC cells. THAP9-AS1 was more highly expressed in GC tissues than in gastritis tissues, demonstrating that THAP9-AS1 may have important functions in the progression of GC.

Materials and methods
Cell culture and siRNA transfection
The human GC cell lines AGS, BGC-823, HGC-27, MGC-803 and SGC-7901 were purchased from the Cell Biology Center at the Shanghai Institute of Biochemistry and Cell Biology at the Chinese Academy of Sciences (Shanghai, PR China). AGS cells were cultured with F12 medium (Gibco, USA). BGC-823, MGC-803, HGC-27 and SGC-7901 cells were cultured with RPMI-1640 medium (Gibco, USA). All media were supplemented with 10% fetal bovine serum (Gibco, USA). The cells were all incubated in humidified air with 5% CO_2 at 37 °C. Two THAP9-AS1-specific siRNAs and control siRNA were synthesized by GenePharma (Shanghai, PR China). The siRNA sequences were as follows:

siRNA-1: GGGACAGUGCUCCAAAGAATT; siRNA-2: CCAGAUCAGCUACAGACAATT; and control siRNA: UUCUCCGAACGUGUCACGUTT. The siRNAs were transfected into cells using Lipofectamine 2000 (Invitrogen, USA) according to the manufacturer's protocol.

H. pylori and the infection model
H. pylori strains (NCTC 26695) were kindly provided by Dr. Jianzhong Zhang (Chinese Disease Control and Prevention Center; Beijing, China). The strains were inoculated in Brucella broth containing 5% FBS under microaerobic conditions (5% O_2, 10% CO_2, and 85% N_2) at 37 °C. The AGS cells were seeded in six-well cell culture plates before infection. The next day, the cell monolayer was washed with PBS and cultured in fresh medium. Then, *H. pylori* was inoculated into AGS cells at a multiplicity of infection (MOI) of 100 for 12 hrs as the experimental group. Cells without *H. pylori* infection were cultured in the same environment for 12 hrs. After 12 hrs, the cells were harvested in TRIzol reagent (Invitrogen, USA) and stored at −80 °C until RNA isolation.

RNA isolation
Total RNA was isolated with TRIzol reagent (Invitrogen, USA) according to the manufacturer's instructions. RNA degradation and contamination were monitored using 1% agarose gels. RNA purity was assessed via a NanoPhotometer spectrophotometer (IMPLEN, CA, USA). RNA concentration was measured using the Qubit RNA Assay Kit and a Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). RNA integrity was assessed by the RNA Nano 6000 Assay Kit with the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

Library preparation for lncRNA sequencing and data analysis
Three micrograms of RNA per sample was used as input material for RNA sample preparation. First, ribosomal RNA was removed using the Epicentre Ribo-zero rRNA Removal Kit (Epicentre, USA), and the rRNA-free residue was removed through ethanol precipitation. Subsequently, sequencing libraries were created using the rRNA-depleted RNA with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's recommendations. Briefly, fragmentation was conducted using divalent cations under elevated temperature in the NEBNext First Strand Synthesis Reaction Buffer (5X). First-

strand cDNA was synthesized using a random hexamer primer and M-MuLV Reverse Transcriptase. The synthesis of second-strand cDNA was subsequently performed using DNA polymerase I and RNase H. In the reaction buffer, dNTPs and dTTP were replaced with dUTP. The remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After the adenylation of 3′ ends of DNA fragments, the NEBNext Adaptor with a hairpin loop structure was ligated to prepare for hybridization. The library fragments were purified with the AMPure XP system (Beckman Coulter, Beverly, USA) in order to preferentially select cDNA fragments of 150–200 bp in length. Then, 3 μl of USER Enzyme (NEB, USA) was incubated with size-selected, adaptor-ligated cDNA at 37 °C for 15 min, followed by 5 min at 95 °C before PCR. Then, PCR was carried out with Phusion High-Fidelity DNA polymerase, Universal PCR primers and the Index (X) Primer. Finally, the products were purified (AMPure XP system), and the library quality was assessed using an Agilent Bioanalyzer 2100 system.

The raw reads in fastq format were first processed through in-house PERL scripts. In this step, clean reads were obtained by removing reads containing the adapter, reads containing poly-N and low quality reads from the pool of raw reads. At the same time, the Q20, Q30 and GC content of the clean reads were calculated. All the downstream analyses were based on clean reads of high quality. Reads were mapped with TopHat2 to the human genome sequence assembly (Gencode v23). The mapped reads of each sample were assembled with Scripture and Cufflinks.[26,27]

Coding potential analysis and quantification of gene expression level

To obtain data of high quality, we chose CNCI, CPC, Pfam-scan and phyloCSF to identify candidate lncRNAs. CNCI (Coding-Non-Coding-Index) was used with default parameters to effectively distinguish protein-coding and non-coding sequences independent of annotations.[28–31]

Cuffdiff (v2.1.1) was used to calculate the FPKM of both lncRNAs and coding genes in each sample for the quantification of gene expression levels.[27] Gene FPKM values were computed by summing the FPKM of the transcripts in each gene group. FPKM (fragments per kilo-base of exon per million fragments mapped) was calculated based on the length of the fragments and the read count mapped to this fragment.

RNA extraction and quantitative real-time PCR (qRT-PCR)

Total RNA was isolated using TRIzol reagent (Invitrogen, USA). After quantification, total cDNA was synthesized using the RevertAid First-Strand DNA Synthesis (RT) Kit (Fermentas, Canada). Then, quantitative real-time PCR was performed with the SYBR Premix Ex Taq system (TaKaRa, Japan) and the Bio-Rad CFX96TM Real-Time PCR System (Bio-Rad, USA). The endogenous control was β-actin, and the primer sequences are shown in Table 1. The $2^{-\Delta\Delta Ct}$ method was used to determine the relative expression level of target genes.

Colony formation assay

MGC-803 and SGC-7901 cells were transfected with THAP9-AS1-specific siRNAs and control siRNA for the corresponding time, inoculated into 6-well plates (300 cells/well) with RPMI-1640 medium supplemented with 10% FBS, and cultured at 37 °C with 5% CO_2 for 14 days. The colonies were fixed with methanol for 20 min and stained with crystal violet. Then, colonies with more than 50 cells were counted.

Table 1 The primers of β-actin and nine randomly selected lncRNAs

LncRNA	Primer
RP11-823E8.3	F:5′-TTCTTTAGGCGGTGTGTGGC-3′ R:5′-TCGGTCACTAAAGCCCAGTC-3′
THAP9-AS1	F:5′-ACTGGCTGCTATGGAAAAAGT-3′ R:5′-TCCCTTCCCCTCCTGTCTGC-3′
MIR22HG	F:5′-AAGAACTGTTGCCCTCTGCC-3′ R:5′-AATCTGGGCAAAGGCTCTCC-3′
SNHG1	F:5′-CTTATTGGGCTCCTGCTCGC-3′ R:5′-ACCAGTAAGCTCTTGTGGGC-3′
LRRC75A-AS1	F:5′-AGAGTGCTGAAGACGGGGTA-3′ R:5′-TCAAAACCTCATGGCAGGCT-3′
GAS5-001	F:5′-TGCAGGCAGACCTGTTATCC-3′ R:5′-CCATGAGACTCCATCAGGCA-3′
GAS5-016	F:5′-GCCGAGTCACCCGAGTAAG-3′ R:5′-CTACCACCGACAGCCTTTCA-3′
XLOC_314741-TCONS_00422456	F:5′-TTCTTCCTGTGTGTTGGGGC-3′ R:5′-GTAGTCGGGCACTGGTTTCA-3′
XLOC_158512-TCONS_00206520	F:5′-TTCATCAATGTGTTCGTGAAGC-3′ R:5′-ACGGTTGTGCTTCTCTGTCC-3′
β-actin	F:5′-AGTTGCGTTACACCCTTTCTTG-3′ R:5′-CACCTTCACCGTTCCAGTTTT-3′

Abbreviations: F, forward primer; R, reverse primer.

CCK-8 assay

CCK-8 assays were performed using the Cell Counting Kit-8 (CCK-8; Dojindo Laboratories, Kumamoto, Japan). MGC-803 and SGC-7901 cells transfected with THAP9-AS1-specific siRNAs and control siRNA were inoculated into 96-well plates (2000 cells/well) with RPMI-1640 medium supplemented with 10% FBS and cultured at 37 °C with 5% CO_2. Ten microliters of CCK-8 solution was added to each well on days 1, 2, 3, 4 and 5. After incubating the cells for 3 hrs at 37 °C, the absorbance at 450 nm was measured using the microplate reader Infinite 200 PRO (TECAN, Austria).

Transwell assay

After 12 hrs of serum-free culture, MGC-803 and SGC-7901 cells transfected with THAP9-AS1 specific siRNAs and control siRNA were inoculated into the upper well (10^5 cells/well) of a transwell chamber (pore size: 8 μm, Costar, USA) without a matrigel-coated membrane. The lower chamber contained RPMI-1640 medium supplemented with 20% FBS. The cells were cultured for 24 hrs at 37 °C. The cells that did not migrate were removed with a cotton swab. The cells that migrated through the pores were fixed with methanol for 20 min and stained with crystal violet. The results are presented as the number of cells per field at 400× magnification.

Patient specimens

A total of 31 human gastritis specimens were collected from patients undergoing gastroscopy, and none of these patients had taken drugs, such as antibiotics or anti-inflammatory drugs, before the examination. Twenty GC tissues were obtained from patients with GC during surgery; none of the patients had received radiotherapy or chemotherapy before surgery. GC and gastritis tissues was used to detect the infection of *H. pylori* in the way of C14 urea breath test and rapid urease test. All the patient tissues were obtained from the Qilu Hospital at Shandong University. The study was authorized by the Ethics Committee of Cheeloo College of Medicine, Shandong University. The written informed consents were provided by all participants included in the study and this process was in accordance with the Declaration of Helsinki.

Statistical analysis

Comparisons between different groups were analyzed by Student's *t*-test. $P < 0.05$ indicated statistical significance.

Data availability

Transcriptome sequencing raw read data were submitted to the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra/) under accession number: SRP136039.

Results

Overview of the AGS transcriptome infected by *H. pylori*

In view of the important role of *H. pylori* in the progression of GC, we performed RNA-seq analysis of *H. pylori*-infected AGS cells. 3752 lncRNAs were obtained according to five criteria (1) transcripts were longer than 200 nt; (2) the number of exons was not less than 2; (3) the coverage of the transcripts was greater than 3; (4) transcripts that were same as the known lncRNAs were directly screened, and the rest were compared with other ncRNAs; similar transcripts were discarded; and (5) compared transcripts with known mRNAs and identified intergenic lncRNAs, intronic lncRNAs and anti-sense lncRNAs. To further confirm these 3752 lncRNAs, we performed coding potential analysis using four software programs, Coding Potential Calculator (CPC), Pfam-scan (PFAM), phylogenetic codon substitution frequency (phyloCSF) and Coding-Non-Coding-Index (CNCI), to remove the potential coding transcripts. Finally, 2505 putative non-coding transcripts were obtained (Figure 1A), including 1738 newly found lncRNAs and 767 known lncRNAs collected in GENCODE_v23 (Figure 1B).

We also performed quantitative analysis of all transcripts in *H. pylori*-infected AGS cells (Figure 1C). Given a false discovery rate (FDR) of 2%, the correlation between samples was 0.891 (*p*-value<0.05). Overall, 3661 transcripts (125 lncRNAs, 3465 mRNAs and 71 other ncRNAs) were significantly induced or suppressed in AGS cells after *H. pylori* infection. Among them, 47.36% (1734 transcripts) were up-regulated, and 52.64% (1927 transcripts) were down-regulated (Figure 1D).

Validation of the expression levels of lncRNAs by qRT-PCR

To confirm the accuracy and repeatability of the RNA-seq data, 9 candidate lncRNAs were randomly selected for validation by qRT-PCR. First, we carried out qRT-PCR to confirm the expression of the candidate lncRNAs in AGS cells. The lncRNA expression pattern detected by
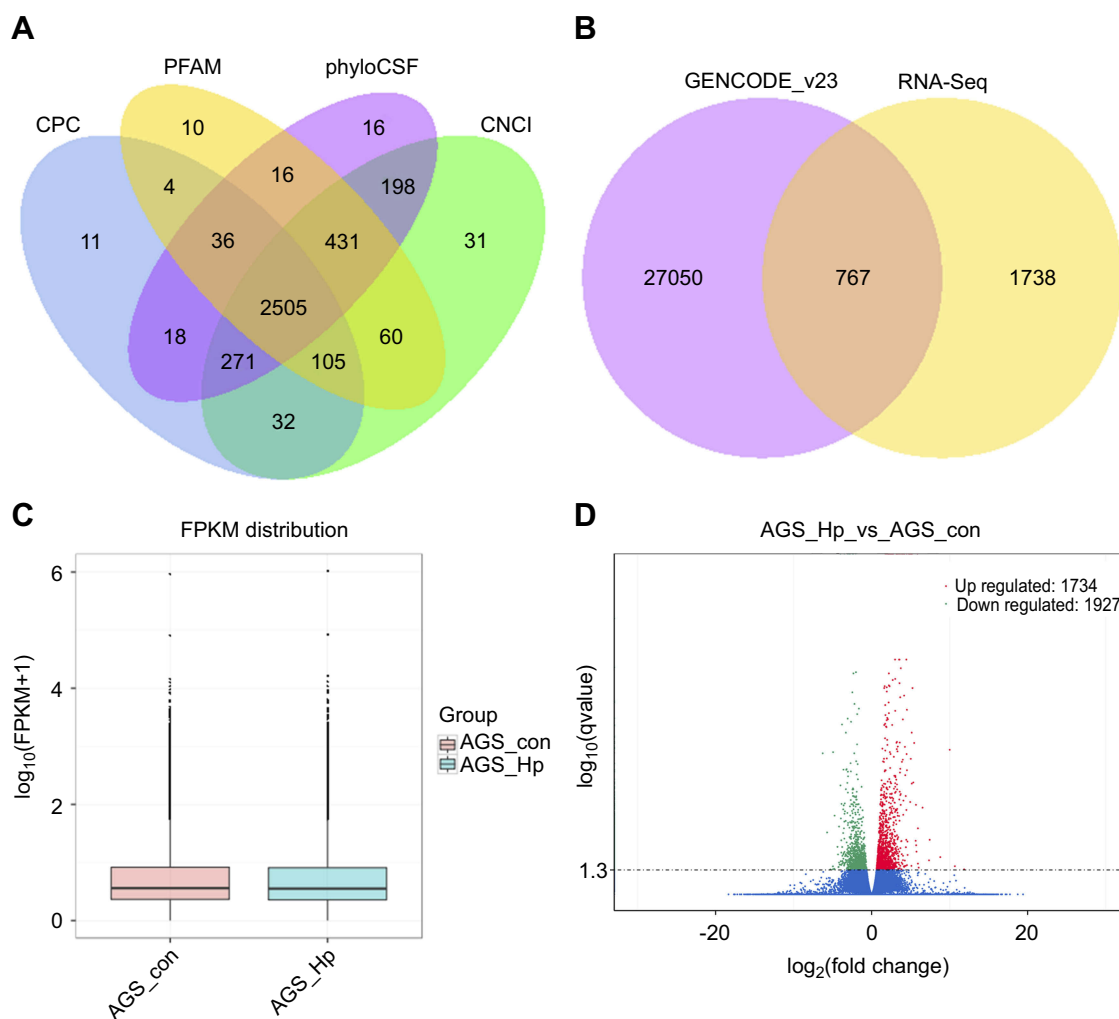
Figure I LncRNA identification and differential expression analysis of the detected transcripts. (**A**) Four tools (CPC, CNCI, PFAM, and PhyloCSF) were employed to analyze the coding potential of candidate lncRNAs. (**B**) Venn diagram showing the distribution of the obtained lncRNAs in GENCODE_v23 and RNA-seq. (**C**) Box plots showing the FPKM distribution for all the detected transcripts. (**D**) Volcano plot of the significantly differentially expressed transcripts between *H. pylori*-infected and non-infected AGS cells (*p*-value<0.05).

Abbreviations: lncRNAs, long noncoding RNAs; CPC, coding potential calculator; CNCI, coding-non-coding-index; PhyloCSF, phylogenetic codon substitution frequency; FPKM, fragments per kilo-base of exon per million fragments mapped.

qRT-PCR was consistent with the RNA-seq data (Figure 2A). Combine with the RNA-seq results, innovativeness and feasibility, lncRNA THAP9-AS1 and XLOC_158512 was chosen because they have not been studied, lncRNA SNHG1 was chosen for its function to promote tumor growth.[32,33] We assessed the expression of THAP9-AS1, SNHG1 and XLOC_158512 in other GC cell lines infected by *H. pylori* and found that THAP9-AS1 and SNHG1 had better conservation and consistency with the results in AGS cells (Figure 2B, D and F). The expression level of THAP9-AS1, SNHG1 and XLOC_158512 in gastritis tissues and GC tissues was also detected. THAP9-AS1 and SNHG1 were upregulated in GC tissues compared to gastritis tissues (Figure 2C, E and G), with a more significant

difference in the expression of THAP9-AS1. Meanwhile, we analyzed *H. pylori*-negative and *H. pylori*-positive GC tissues and found that THAP9-AS1 was more highly expressed in *H. pylori*-positive GC tissues than in *H. pylori*-negative GC tissues. Because of the small quantitiy of specimens, we did not obtain a statistically significant result.

THAP9-AS1 enhanced GC cell proliferation and migration in vitro

THAP9-AS1 was chosen for further functional exploration in the progression of GC based on the above results. First, we quantified the expression level of THAP9-AS1 in four types of GC cells by qRT-PCR
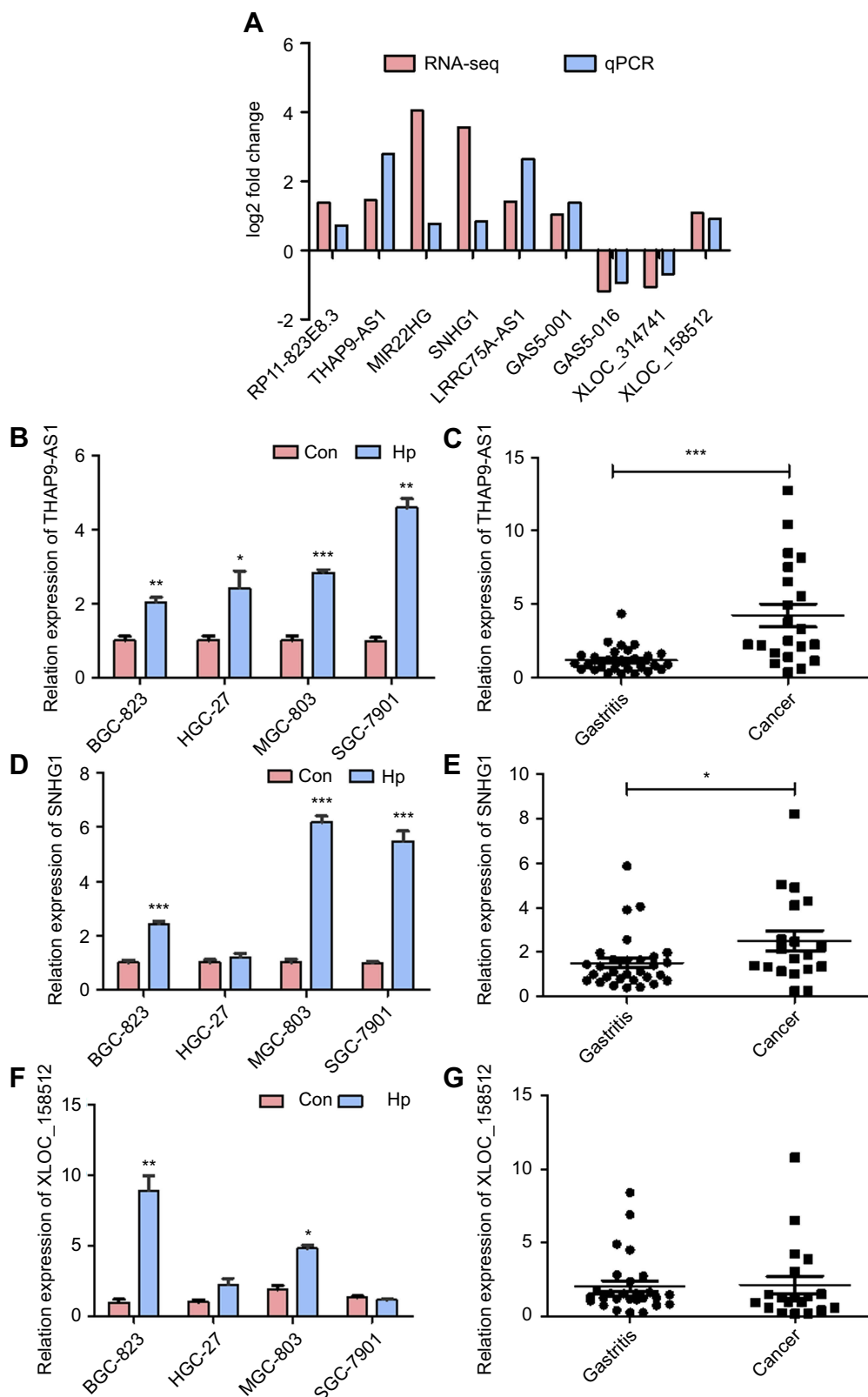
Figure 2 Validation of the RNA-seq results and the expression level in GC cells and gastric tissues. (**A**) The expression level of nine randomly selected lncRNAs in *H. pylori*-infected and non-infected GC AGS cells detected by RNA-seq and qRT-PCR, the result between RNA-seq and qRT-PCR were consistent. (**B**) After the infection of *H. pylori*, THAP9-AS1 expression level was increased in the four GC cell lines. (**C**) THAP9-AS1 had higher expression in GC tissues than gastritis tissues. (**D**) After the infection of *H. pylori*, SNHG1 expression level was increased in the four GC cell lines. (**E**) SNHG1 had higher expression in GC tissues than gastritis tissues. (**F**) After infection of *H. pylori*, XLOC_158512 expression level was increased in BGC-823, HGC-27 and MGC-803, but decreased in SGC-7901. (**G**) XLOC-158512 had higher expression in GC tissues than gastritis tissues. (*$P<0.05$, **$P<0.01$, ***$P<0.001$).
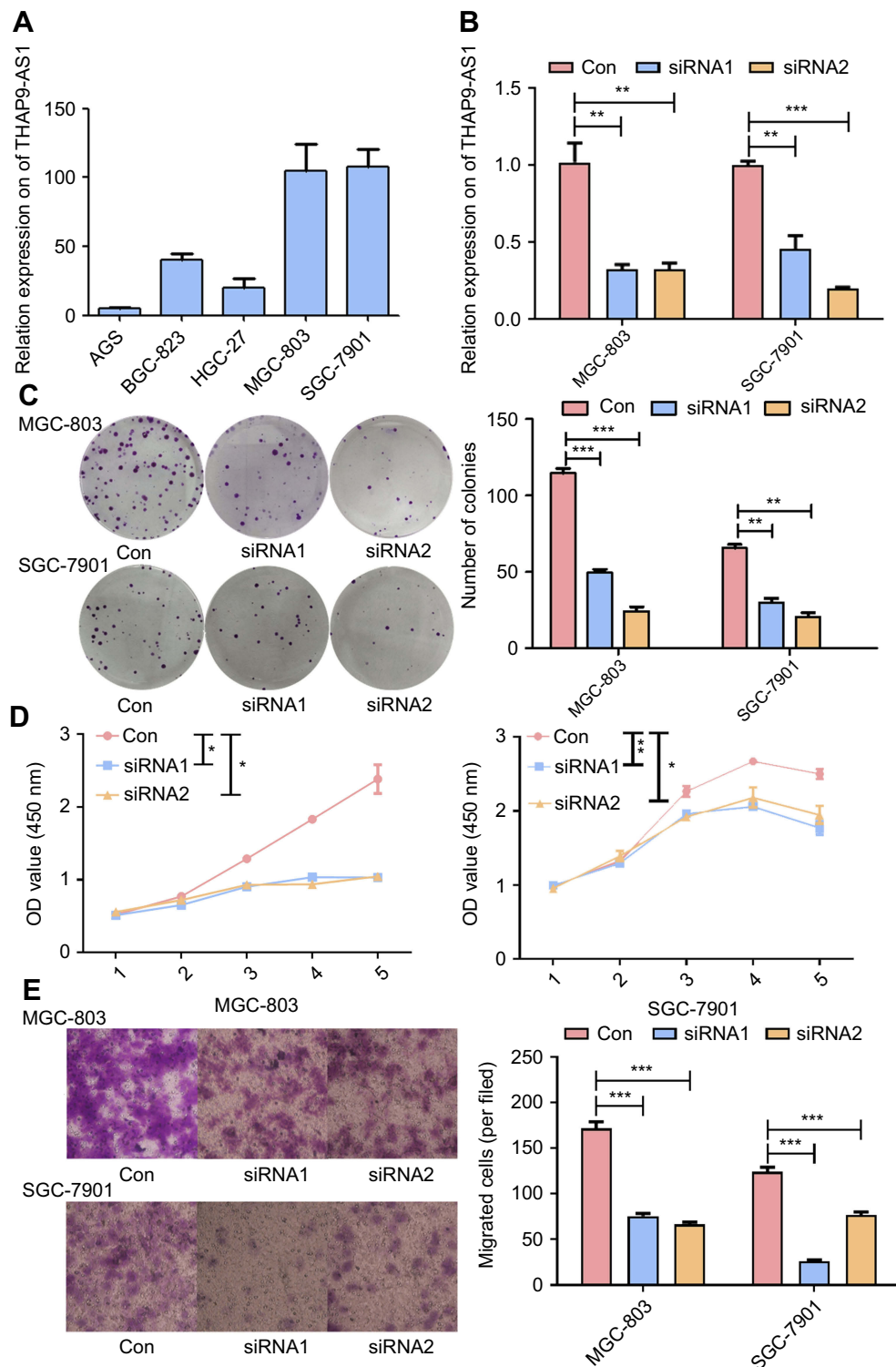
Figure 3 THAP9-AS1 knockdown inhibited GC cell proliferation and migration in vitro. (**A**) The expression level of THAP9-AS1 in five GC cell lines, MGC 803 and SGC-7901 have a rich content of THAP9-AS1. (**B**) After transfecting siRNA-THAP9-AS1, the expression of THAP9-AS1 was decreased significantly in MGC-803 and SGC-7901. (**C, D**) Colony formation assay and CCK8 assay indicated THAP9-AS1 knockdown inhibited the proliferation of MGC-803 and SGC-7901. (**E**) Transwell assay indicated THAP9-AS1 knockdown inhibited migration of MGC-803 and SGC-7901. (*$P<0.05$, **$P<0.01$, ***$P<0.001$).

and found that MGC-803 and SGC-7901 cells exhibited the highest expression level among all the tested cells (Figure 3A). Thus, MGC-803 and SGC-7901 cells

were chosen for functional studies. Subsequently, we inhibited the expression of THAP9-AS1 by transfecting THAP9-AS1-specific small interference RNAs

(siRNAs) into MGC-803 and SGC-7901 cells. Next, qRT-PCR was used to assess the expression of THAP9-AS1, and the siRNA control was used as a negative control (Figure 3B). Colony formation and CCK-8 assays were used to assess cell proliferation. After knocking down the expression of THAP9-AS1 in MGC-803 and SGC-7901, the number of colonies was decreased significantly, and the OD value was also reduced (Figure 3C and D), indicating that inhibiting THAP9-AS1 expression suppressed the proliferation in MGC-803 and SGC-7901 cells. In addition, another phenotype, cell migration, was assessed by a transwell assay. The number of cells passing through the well was decreased obviously after knocking down THAP9-AS1 expression, demonstrating that decreasing THAP9-AS1 expression suppressed the migration of MGC-803 and SGC-7901 cells (Figure 3E). These results indicated that THAP9-AS1 may play important roles in the proliferation and migration of GC cells.

H. pylori enhanced GC cell proliferation and migration by up-regulating the expression of THAP9-AS1 in vitro

Though we have found that THAP9-AS1 can promote the proliferation and migration of GC cell, it is not clear about the exact role for *H. pylori* in this process. In order to explore the relationship between *H. pylori* and THAP9-AS1, we infected GC cell MGC-803 and SGC-7901 with *H. pylori*. After 12 hrs, the *H. pylori* was washed and THAP9-AS1-siRNA1 was transfected to the *H. pylori* infected GC cells. Then we explored the proliferation and migration ability of GC cell MGC-803 and SGC-7901 through CCK-8 and transwell assay. The relative expression of THAP9-AS1 was detected with qRT-PCR. THAP9-AS1 was induced by *H. pylori* infection and inhibited by THAP9-AS1-siRNA1 subsequently (Figure 4A). Through CCK-8 assay, we found that after the infected of *H. pylori*, the proliferation ability of GC cells was promoted
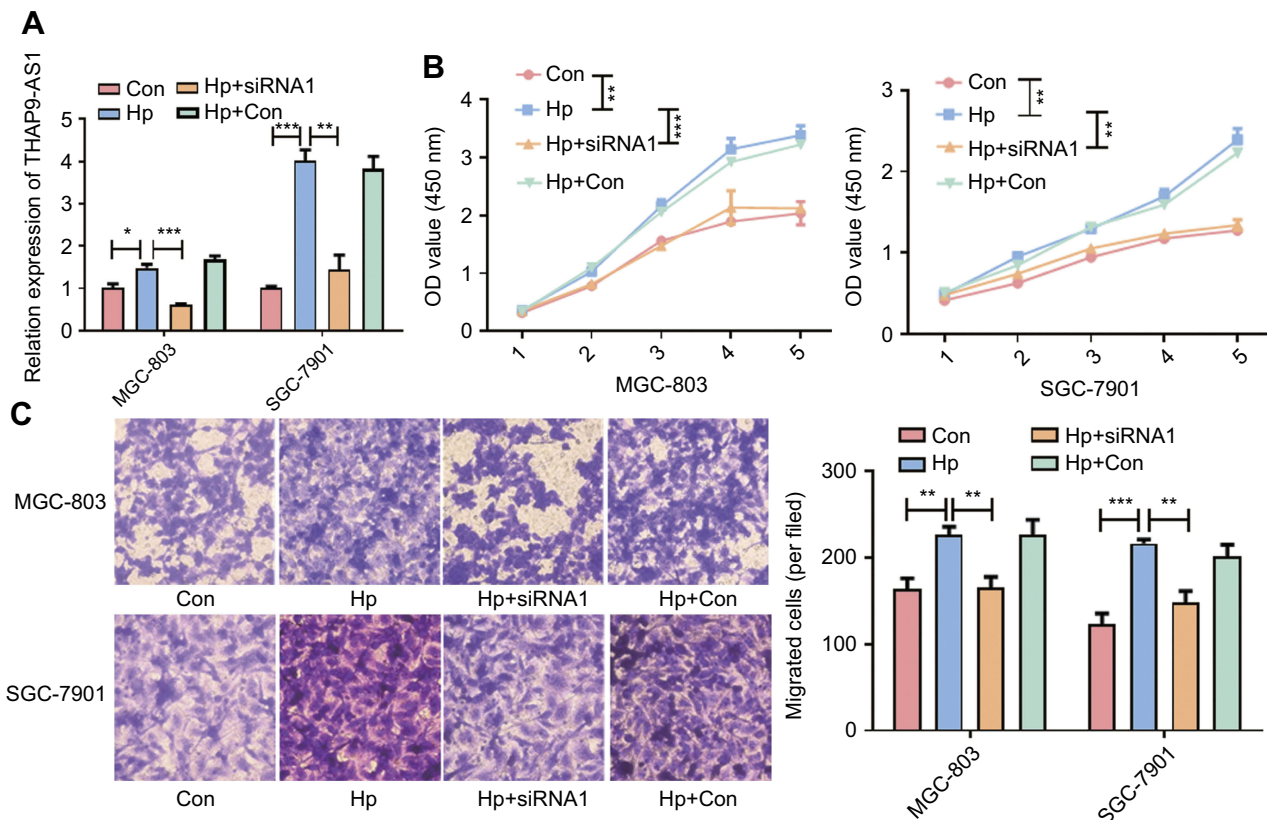

Figure 4 *H. pylori* enhanced GC cell proliferation and migration by up-regulating the expression of THAP9-AS1 in vitro. (**A**) The expression level of THAP9-AS1 after the *H. pylori* infection and transfection of THAP9-AS1-siRNA1 in MGC-803 and SGC-7901 cells. (**B**) CCK-8 assay revealed that *H. pylori* infection promoted the proliferation ability of MGC-803 and SGC-7901, which was abolished by the knockdown of THAP9-AS1. (**C**) Transwell assay indicated that *H. pylori* infection enhanced the migration ability of MGC-803 and SGC-7901 which was inhibited by the down-regulation of THAP9-AS1. (*P<0.05, **P<0.01, ***P<0.001).

significantly, but was inhibited by the knockdown of THAP9-AS1 subsequently (Figure 4B). In addition to the proliferation ability, transwell assay was also performed and an enhanced migration potential was induced by *H. pylori*, which was also abolished by the down-regulation of THAP9-AS1 (Figure 4C). The above results indicated that *H. pylori* promoted GC cell proliferation and migration by up-regulating the expression of THAP9-AS1.

Discussion

As the major risk factor for GC, the molecular mechanisms of *H. pylori* in GC have been widely studied in the past several decades, but the pathogenesis of *H. pylori* is still undetermined.[2,3,5] With the development of high throughput technologies for large-scale expression studies, RNA-seq has accelerated the discovery and identification of lncRNAs.[27,34] An increasing number of lncRNAs have been found to play important roles in the development of cancer through a variety of mechanisms, including chromatin remodeling, RNA processing, localization, translation, mRNA stability and even competing with endogenous RNA to interact with functional proteins.[35–38] In this paper, we explored the role of lncRNAs in *H. pylori*-associated GC.

Though the expression pattern of lncRNAs upon *H. pylori* infection of a gastric epithelial cell line, GES-1, has been studied, the expression pattern of lncRNAs influenced by *H. pylori* in GES-1 cells was explored using a microarray, revealing the relationship between lncRNAs and *H. pylori*-associated diseases.[39,40] We explored the expression pattern of lncRNAs in *H. pylori*-infected AGS cells using RNA-seq which revealed a full-scale lncRNAs map for *H. pylori*-infected AGS cells and identified seven to eight-times more lncRNAs than microarray. Meanwhile, our study also revealed the lncRNAs expression in *H. pylori* infected GC cells, which was not reported yet. Moreover, 1738 new lncRNAs were discovered, and a deeper understanding of the expression pattern of lncRNAs in *H. pylori*-associated GC was revealed. For the differentially expressed lncRNAs, we found that THAP9-AS1 was also induced by *H. pylori* in other GC cells and was richer in GC tissues than in gastritis tissues. We also discovered THAP9-AS1 plays an important role in the proliferation and migration in the progression of GC.

In this study, 2505 lncRNAs were identified. In total, 3661 transcripts were found to be significantly differentially expressed between *H. pylori*-infected AGS cells and control cells. Many lncRNAs were found to play important roles in

the progression of GC, such as GAPLINC, which was found to be up-regulated in human GC tissues compared to paired normal tissues and to regulate CD44 expression by competing for miR-211-3p to enhance GC proliferation and invasion in vitro and in vivo.[22] The function of lncRNAs in the relationship between *H. pylori* and GC has not been thoroughly investigated. In this paper, we used RNA-seq to identify a new *H. pylori*-associated lncRNA, THAP9-AS1, which had not been studied before. Its expression increased significantly after the infection of GC cells with *H. pylori* and was higher in GC tissues than in gastritis tissues. We also found that THAP9-AS1 knockdown inhibited the proliferation and migration of GC cells induced by *H. pylori* in vitro, which revealed that *H. pylori* promoted the proliferation and migration of GC cell in a THAP9-AS1-dependent manner. Lnc-SGK1 was reported to be induced in GC cells by *H. pylori* infection and a high-salt diet, and the activation of SGK1/Jun B signaling promotes Th2 and Th17 cell differentiation.[41] The *H. pylori* infection-related lncRNA AF147447 was also found to function as a tumor suppressor by targeting MUC2 and up-regulating miR-34c.[42] Lnc-SGK1, which mainly functions in T cell differentiation, and AF147447, which acts as a tumor suppressor, inhibited the progression of GC. THAP9-AS1 is a new *H. pylori*-associated lncRNA that has a pro-oncogenic function and enhances the proliferation and migration of GC cells in vitro. Similar to the way in which AF147447 exerts its function through miR-34c and MUC2, we thought some miRNAs and protein-coding genes may be regulated by THAP9-AS1 and exert their pro-oncogenic function. According to the cis target prediction results, SEC31A and THAP9, as predicted target genes, may be regulated by THAP9-AS1 and exert their pro-oncogenic function. As for the specific molecular mechanism, more studies need to be performed in the future. The higher expression of THAP9-AS1 in GC tissues also supports its potential role in the progression of GC; meanwhile, the higher expression level in GC tissues indicates that THAP9-AS1 has the potential to be a diagnostic and prognostic marker of GC.

Conclusion

RNA-seq revealed the expression pattern of *H. pylori*-infected AGS cells, and functional enrichment analysis indicated that many lncRNAs may play important roles in the development of cancer. In addition, we noted that the lncRNA THAP9-AS1 induced by *H. pylori* functions as a pro-oncogenic factor to enhance the proliferation and migration of GC cells in vitro. The higher expression in GC tissues

indicated that THAP9-AS1 might be a potential diagnostic and prognostic marker. However, there are more questions that need to be answered, and the specific molecular mechanism underlying the relationship between lncRNAs and *H. pylori*-associated GC is still elusive. There are more lncRNAs related to *H. pylori*-associated GC that need to be discovered to construct a *H. pylori*, lncRNA and GC network to reveal the full molecular mechanism of *H. pylori* and GC. In summary, THAP9-AS1 is a new *H. pylori*-associated lncRNA and possesses a pro-oncogenic function that enhances the proliferation and migration of GC cells in vitro. Additionally, it may have higher expression in GC tissues than in gastritis tissues, and it has the potential to be a therapeutic target and diagnostic marker.

Author contributions

Wenxiao Jia: performed the experiments and wrote the manuscript; Jiaqin Zhang: performed the experiments and analyzed the data; Fang Ma: performed the experiments and provided technical support; Shengjie Hao: performed the experiments; Xue Li: performed the experiments; Ruiting Guo: analyzed the data; Qianqian Gao: analyzed the data; Yundong Sun: revised the manuscript; Jihui Jia: revised the manuscript; Wenjuan Li: designed the experiments, revised the manuscript and gave final approval of the manuscript. All authors contributed to data analysis, drafting or revising the article, gave final approval of the version to be published, and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin*. 2015;65(2):87–108.
2. Parsonnet J, Friedman GD, Vandersteen DP, et al. *Helicobacter pylori* infection and the risk of gastric carcinoma. *N Engl J Med*. 1991;325 (16):1127–1131. doi:10.1056/NEJM199110173251603
3. Peek RM Jr, Blaser MJ. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer*. 2002;2(1):28–37. doi:10.1038/nrc703
4. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet*. 1984;1 (8390):1311–1315. doi:10.1016/s0140-6736(84)91816-6
5. Wroblewski LE, Peek RM Jr, Wilson KT. *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. *Clin Microbiol Rev*. 2010;23(4):713–739. doi:10.1128/CMR.00011-10
6. Salama NR, Hartung ML, Muller A. Life in the human stomach: persistence strategies of the bacterial pathogen *Helicobacter pylori*. *Nat Rev Microbiol*. 2013;11(6):385–399. doi:10.1038/nrmicro3016
7. Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Int J Cancer*. 2006;118(12):3030–3044. doi:10.1002/ijc.21731
8. Misra V, Pandey R, Misra SP, Dwivedi M. *Helicobacter pylori* and gastric cancer: Indian enigma. *World J Gastroenterol*. 2014;20(6):1503–1509. doi:10.3748/wjg.v20.i6.1503
9. Goh KL. Epidemiology of *Helicobacter pylori* infection in Malaysia–observations in a multiracial Asian population. *Med J Malaysia*. 2009;64(3):187–192.
10. Djebali S, Davis CA, Merkel A, et al. Landscape of transcription in human cells. *Nature*. 2012;489(7414):101–108. doi:10.1038/nature11233
11. Banfai B, Jia H, Khatun J, et al. Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res*. 2012;22(9):1646–1657. doi:10.1101/gr.134767.111
12. Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev*. 2009;23(13):1494–1504. doi:10.1101/gad.1800909
13. Hung T, Wang Y, Lin MF, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet*. 2011;43(7):621–629. doi:10.1038/ng.848
14. Tripathi V, Shen Z, Chakraborty A, et al. Long noncoding RNA MALAT1 controls cell cycle progression by regulating the expression of oncogenic transcription factor B-MYB. *PLoS Genet*. 2013;9(3): e1003368. doi:10.1371/journal.pgen.1003368
15. Guttman M, Donaghey J, Carey BW, et al. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*. 2011;477(7364):295–300. doi:10.1038/nature10398
16. Gupta RA, Shah N, Wang KC, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*. 2010;464(7291):1071–1076. doi:10.1038/nature08975
17. Huarte M, Guttman M, Feldser D, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*. 2010;142(3):409–419. doi:10.1016/j.cell.2010.06.040
18. Gomez JA, Wapinski OL, Yang YW, et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. *Cell*. 2013;152(4):743–754. doi:10.1016/j.cell.2013.01.015
19. Wang GY, Zhu YY, Zhang YQ. The functional role of long noncoding RNA in digestive system carcinomas. *Bull Cancer*. 2014;101 (9):E27–E31. doi:10.1684/bdc.2014.2023
20. Leucci E, Vendramin R, Spinazzi M, et al. Melanoma addiction to the long non-coding RNA SAMMSON. *Nature*. 2016;531(7595):518–522. doi:10.1038/nature17161
21. Xie Z, Chen X, Li J, et al. Salivary HOTAIR and PVT1 as novel biomarkers for early pancreatic cancer. *Oncotarget*. 2016;7(18):25408–19.
22. Hu Y, Wang J, Qian J, et al. Long noncoding RNA GAPLINC regulates CD44-dependent cell invasiveness and associates with poor prognosis of gastric cancer. *Cancer Res*. 2014;74(23):6890–6902. doi:10.1158/0008-5472.CAN-14-0686
23. Feng F, Tian Y, Xu G, et al. Diagnostic and prognostic value of CEA, CA19-9, AFP and CA125 for early gastric cancer. *BMC Cancer*. 2017;17(1):737. doi:10.1186/s12885-017-3738-y
24. Doi T, Shitara K, Naito Y, et al. Safety, pharmacokinetics, and antitumour activity of trastuzumab deruxtecan (DS-8201), a HER2-targeting antibody-drug conjugate, in patients with advanced breast and gastric or gastro-oesophageal tumours: a phase 1 dose-escalation study. *Lancet Oncol*. 2017;18:1512–1522. doi:10.1016/S1470-2045(17)30604-6

25. Wang D, Wang B, Wang R, et al. High expression of EGFR predicts poor survival in patients with resected T3 stage gastric adenocarcinoma and promotes cancer cell survival. *Oncol Lett*. 2017;13 (5):3003–3013. doi:10.3892/ol.2017.5827

26. Guttman M, Garber M, Levin JZ, et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat Biotechnol*. 2010;28 (5):503–510. doi:10.1038/nbt.1633

27. Trapnell C, Williams BA, Pertea G, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28 (5):511–515. doi:10.1038/nbt.1621

28. Sun L, Luo H, Bu D, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res*. 2013;41(17):e166. doi:10.1093/nar/gkt646

29. Kong L, Zhang Y, Ye ZQ, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res*. 2007;35(Web Server issue):W345–W349. doi:10.1093/nar/gkm391

30. Punta M, Coggill PC, Eberhardt RY, et al. The Pfam protein families database. *Nucleic Acids Res*. 2012;40(D1):D290–D301. doi:10.1093/nar/gkr1065

31. Lin MF, Jungreis I, Kellis M. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics*. 2011;27(13):i275–i282. doi:10.1093/bioinformatics/btr209

32. Sun Y, Wei G, Luo H, et al. The long noncoding RNA SNHG1 promotes tumor growth through regulating transcription of both local and distal genes. *Oncogene*. 2017;36(49):6774–6783. doi:10.1038/onc.2017.286

33. Xu M, Chen X, Lin K, et al. The long noncoding RNA SNHG1 regulates colorectal cancer cell growth through interactions with EZH2 and miR-154-5p. *Mol Cancer*. 2018;17(1):141.

34. Pauli A, Valen E, Lin MF, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res*. 2012;22(3):577–591. doi:10.1101/gr.133009.111

35. Cesana M, Cacchiarelli D, Legnini I, et al. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*. 2011;147(2):358–369. doi:10.1016/j.cell.2011.09.028

36. Xing Z, Lin A, Li C, et al. lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. *Cell*. 2014;159 (5):1110–1125. doi:10.1016/j.cell.2014.10.013

37. Heo JB, Sung S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science*. 2011;331(6013):76–79. doi:10.1126/science.1197349

38. Chen LL. Linking long noncoding RNA localization and function. *Trends Biochem Sci*. 2016;41(9):761–772. doi:10.1016/j.tibs.2016.07.003

39. Yang L, Long Y, Li C, et al. Genome-wide analysis of long noncoding RNA profile in human gastric epithelial cell response to *Helicobacter pylori*. *Jpn J Infect Dis*. 2015;68(1):63–66. doi:10.7883/yoken.JJID.2014.149

40. Zhu H, Wang Q, Yao Y, et al. Microarray analysis of Long noncoding RNA expression profiles in human gastric cells and tissues with *Helicobacter pylori* Infection. *BMC Med Genomics*. 2015;8:84. doi:10.1186/s12920-015-0109-x

41. Yao Y, Jiang Q, Jiang L, et al. Lnc-SGK1 induced by *Helicobacter pylori* infection and highsalt diet promote Th2 and Th17 differentiation in human gastric cancer by SGK1/Jun B signaling. *Oncotarget*. 2016;7(15):20549–20560. doi:10.18632/oncotarget.7823

42. Zhou X, Chen H, Zhu L, et al. *Helicobacter pylori* infection related long noncoding RNA (lncRNA) AF147447 inhibits gastric cancer proliferation and invasion by targeting MUC2 and up-regulating miR-34c. *Oncotarget*. 2016;7(50):82770–82782. doi:10.18632/oncotarget.13165