Original Article
Comparative transcriptome analysis using RNA-Seq screening reveals critical genes and regulatory mechanisms of gastric carcinogenesis involved in Helicobacter pylori infections and CagA

Yanyan Shi¹, Qiang Li²

¹Peking University Third Hospital, Research Center of Clinical Epidemiology, Beijing 100191, P.R. China; ²The Ministry of Science and Technology (MoST), Exchange, Development & Service Center for Science & Technology Talents, Beijing 100045, P.R. China

Received March 18, 2019; Accepted May 7, 2019; Epub July 15, 2019; Published July 30, 2019

Abstract: Background: Helicobacter pylori (H. pylori) has been hurled into the spotlight as an important pathogenic factor in the development of gastric cancer. Mechanisms of H. pylori, however, need to be clarified. Methods: Gastric cancer AGS cells were infected with CagA+ and CagA− H. pylori strains at multiplicities of 25:1. Comparative transcriptome analysis using RNA-seq screening was used to identify genes regulated by H. pylori and H. pylori CagA. These were validated by real-time PCR. Results: A total of 2,181 genes were identified to be significantly differentially-expressed from the three groups. Genes involved in cellular response to stress and cell death were highly enriched under H. pylori infections. Genes involved in mitotic sister chromatid segregation and mitotic cell cycle were enriched, comparing CagA+ and CagA− H. pylori infections. IL-8, IL-33, TFF1, MUC13, MUC5AC, EGFR, VEGFA, and BCL2L1 genes were verified to be upregulated after H. pylori infections. EGFR and VEGFA gene-expression changes induced by H. pylori did not depend on CagA. Conclusion: Novel and significant differentially-expressed genes were identified in this study. Present data provides a crucial resource, determining specific responses to H. pylori infections or H. pylori CagA in gastric cancer cells.

Keywords: Helicobacter pylori, CagA, RNA-seq, real-time PCR, gastric cancer

Introduction
Gastric cancer is one of the most common cancer, imposing a considerable health burden all over the world [1]. Helicobacter pylori (H. pylori) is a pathogen that colonizes mucosal surfaces of the human stomach. Persistent colonization of H. pylori, combined with the highly inflammatory response of the host, is a critical factor associated with severe manifestations of gastric disease [2]. Evidence has shown that H. pylori infections increase the risk of gastric cancer. The International Agency for Research on Cancer has classified H. pylori as a group I carcinogen, causing gastric cancer in humans [3-5].

H. pylori virulence factors play an important role in determining outcomes of gastric diseases [6]. One well-known H. pylori virulence factor that augments cancer risk is CagA. CagA can be translocated into host cells, subsequently interferes with cellular processes [7]. Infections with cagA-positive H. pylori strains have been associated with an increased risk of developing gastric cancer. CagA has been designated a bacterial oncoprotein [8]. Numerous studies have been conducted concerning the pathogenicity of H. pylori or H. pylori CagA [9-11]. However, the mechanisms of H. pylori or H. pylori CagA on gastric carcinogenesis are complex and remain unclear.

RNA sequencing technology provides researchers with a revolutionary method. It is highly precise, rapid, reproducible, and cost-effective. This technique has been used in comparative transcriptomics, identifying differential-gene expression among various treatment conditions. The present study generated a transcrip-
**Helicobacter pylori** and gastric carcinogenesis

H. pylori infected gastric cell culture models are suitable for in vitro infections of *H. pylori*. Human-derived AGS gastric epithelial cells are used in *H. pylori* infection cell models in vitro. AGS cell line co-culturing with *H. pylori* has been used to investigate the mechanisms of *H. pylori* on gastric carcinogenesis [12, 13]. The present study used the most classic in vitro study model to explore the pathogenic mechanisms of *H. pylori* infections.

**Materials and methods**

**H. pylori** culturing

*H. pylori* strains, ATCC 26695 (CagA+) and Hp8822 (CagA−), were used in this study. *H. pylori* strains were cultured, respectively, on blood agar plates containing 39 g l−1 Columbia solid culture medium (Oxoid), 5% (v/v) sheep’s blood (Curtin Matheson, Jessup, MD, USA), antibiotics amphotericin B 4 µg ml−1 (Life Tech, Carlsbad, CA, USA), trimethoprim 4 µg ml−1, and vancomycin 4 µg ml−1. The plates were incubated at 37 °C for 3-5 days in a microaerobic environment (5% (v/v) O2, 10% (v/v) CO2, and 85% (v/v) N2). Before harvesting, the *H. pylori* cultures were examined using urease tests and Gram staining. Oxidase tests and catalase tests were also used, ensuring that the strains were not contaminated.

**Cell culturing, culture conditions, and co-culture assays**

AGS cells were cultured in RPMI1640 medium supplemented with 10% (v/v) fetal bovine serum (FBS) (Hyclone, Logan, UT, USA). AGS cells were cultured at 37 °C in a humidified incubator in 5% (v/v) CO2. After the cultures had been resuscitated on blood agar plates, *H. pylori* 26695 (CagA+) and Hp8822 (CagA−) were harvested and washed with PBS three times. They were prepared in the cell growth medium and diluted to a final concentration of 1×10⁸ CFU mL⁻¹. AGS cells were plated one day before *H. pylori* treatment. For co-culturing of cells and strains, the cells were rinsed once with PBS before fresh growth medium was added. Bacterial strains were then added to the cell medium at a multiplicity of infection (MOI) of 50:1 for 24 hours. There were three groups, including negative control cells, as AGS1, *H. pylori* 26695 (CagA+) infected cells, as AGS3, and Hp8822 (CagA−) infected cells, as AGS4.

**RNA isolation, cDNA library preparation, and whole transcript analysis via RNA-sequencing**

Total RNA was extracted using TRIzol Reagent (Tiangen, Beijing) and assessed with Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) and QubitFluorometer (Invitrogen). Total RNA samples meeting the following requirements were used in subsequent experiments: 1) RNA integrity number (RIN) > 7.0; 2) 28S:18S ratio > 1.8. RNA-seq libraries were generated and sequenced by CapitalBio Technology (Beijing, China). Triplicate samples of all assays were used to construct an independent library with the following sequencing and analysis. Next Ultra RNA Library Prep Kit for Illumina (NEB) was used to construct libraries for sequencing. Next Poly(A) mRNA Magnetic Isolation Module kit (NEB) was used to enrich the poly(A) tailed mRNA molecules from 1 µg total RNA. The mRNA was fragmented into ~200 base pair pieces. First-strand cDNA was synthesized from the mRNA fragments reverse transcriptase and random hexamer primers. Second-strand cDNA was then synthesized using DNA polymerase I and RNaseH. The end of the cDNA fragment was subjected to an end repair process, including the addition of a single “A” base, followed by ligation of the adapters. Products were purified and enriched by polymerase chain reaction (PCR), amplifying the library DNA. The final libraries were quantified using the KAPA Library Quantification kit (KAPA Biosystems, South Africa) and an Agilent 2100 Bioanalyzer. After quantitative reverse transcription-polymerase chain reaction (RT-qPCR) validation, the libraries were subjected to paired-end sequencing, with pair end 150-base pair reading lengths on an Illumina Hi-Sequencer (Illumina) [14].

**RNA-seq analysis for differentially-expressed genes (DEGs)**

The genome of human genome version of hg19 was used as reference. Sequencing quality was assessed with FastQC (Version 0.11.5). Low-quality data was filtered using NGSQC (v0.4).

---

Clean reads were then aligned to the reference genome using HISAT2 (Johns Hopkins University, USA), with default parameters [15]. Processed reads from each sample were aligned using HISAT2 (Johns Hopkins University, USA) against the reference genome. Gene expression analyses were performed with Cuffquant and Cuffnorm (Cufflinks 2.2.1).

Cuffdiff was used to analyze DEGs between samples. The standardization method of Cuffdiff is geometric, with the per-condition and pooled as the discrete model [16]. Thousands of independent statistical hypothesis testing was conducted on DEGs, separately. P-values were then obtained. They were corrected by the FDR method. Corrected P-values were calculated by the BH method. P-values were used to conduct significance analysis [17]. Parameters for classifying significant DEGs included ≥2-fold differences (|log_2 FC|≥1, FC: the fold change of expressions) in the transcript abundance and P<0.05.

**Gene ontology (GO) and KEGG enrichment analyses of DEGs**

Pearson’s correlation coefficient was computed for fragments per kilobase of exon per million fragments mapped (FPKM) and log_{10} (FPKM) values for genes from each group. DEGs were analyzed using the search tool DAVID (http://david.abcc.ncifcrf.gov/) for gene ontology (GO) annotation and enrichment analysis, including three main modules. These included biological process, cellular component, and molecular function. Web-based Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg), BioCyc (http://biocyc.org/), Reactome (http://www.reactome.org/), and Panther (http://www.pantherdb.org/) were used for pathway analysis.

**Real-time quantitative PCR**

Eight DEGs identified by RNA-seq were validated using real-time PCR. Total RNA was isolated from three groups and was used for real-time quantitative PCR. Real-time quantitative PCR was performed, as described previously [3]. SYBR Green on a Light Cycler 480II Real-time PCR Detection System (Roche) was used. Each real-time PCR procedure was performed in triplicate. Primers included: 18S rRNA: 5’-GTAACC-CGTTAACCACCTT-3’ (forward primer) and 5’-CATCCATCGTGATAGCG-3’ (reverse primer); IL-8: 5’-AAGACATACTCCCAACCTCCACC-3’ (forward primer) and 5’-TTCAAAACCTCTCCACACCCCTCCT-3’ (reverse primer); IL-33: 5’-TGAGC-TGTGGTAGTGAAGATG-3’ (forward primer) and 5’-CCCTGCTGCGACAGTGTGGTT-3’ (reverse primer); TFF1: 5’-AGGCCGACAGAAGCTGACATTACG-3’ (forward primer) and 5’-ACGTGGATGGTATTAGGATAGAAG-3’ (reverse primer); MUC13: 5’-CT-GAAATGCGTGCTGTA-3’ (forward primer) and 5’-AGTCATCCGCAGTCTGGTTTT-3’ (reverse primer); MUC5AC: 5’-TCACCAACACCAGAAGCC-3’ (forward primer) and 5’-TGAGCACCAGAGCCCATCC-3’ (reverse primer); MUC13: 5’-TGAGCACCAGAGCCCATCC-3’ (reverse primer); EGFR: 5’-CTACAACTACCTGAAGCC-3’ (forward primer) and 5’-TGAGCACCAGAGCCCATCC-3’ (reverse primer); VEGFA: 5’-GGCCAGCACATAGGAGAGAT-3’ (forward primer) and 5’-ACGTGGATGGTATTAGGATAGAAG-3’ (reverse primer); BCL2L1: 5’-CCTGCAGACAGACGACGACG-3’ (forward primer) and 5’-TGAGCACCAGAGCCCATCC-3’ (reverse primer).

**Results**

**Overview of RNA-seq results**

Exploring the molecular mechanisms of *H. pylori* infections and *H. pylori* CagA in gastric carcinogenesis, this study treated gastric cancer AGS cells with *H. pylori* 26695 (CagA+) or Hp8822 (CagA−), respectively at a MOI of 25:1 for 24 hours. As described, negative control cells were numbered as AGS1. Cells infected with *H. pylori* 26695 (CagA+) were numbered AGS3 and cells infected with Hp8822 (CagA−) were numbered AGS4. Three sequencing libraries were prepared from AGS1, AGS3, and AGS4 samples. They were sequenced with the Illumina HiSeq platform.

Pearson’s correlation coefficients were calculated for each of the two groups. Pearson’s correlation coefficient for AGS1 and AGS3 was 0.877. For AGS1 and AGS4, Pearson’s correlation coefficient was 0.919. For AGS3 and AGS4, Pearson’s correlation coefficient was 0.962. Pearson’s correlation coefficients of the two groups are shown in a heatmap (Figure 1A). It shows that *H. pylori* 26695 (CagA+) infections...
Helicobacter pylori and gastric carcinogenesis

Figure 1. A. Pearson's correlation coefficients among each two groups are shown in a heatmap. AGS1, negative control AGS cells without H. pylori infection; AGS3, AGS cells with H. pylori 26695 (CagA+) infection; AGS3, AGS cells with CagA- H. pylori infection. B. Differentially-expressed genes were identified in different treatment groups. C. Venn diagrams show overlaps of differentially-expressed genes among the three comparisons. D. Expression levels of dysregulated genes identified by RNA-seq from the three groups (AGS1, AGS3, AGS4) are shown in a heatmap. Red-colored clusters represent the high gene expression quantity, while green-colored clusters represent low gene expression quantity.

(AGS3) produced more changed genes, compared with Hp8822 (CagA-) infections (AGS4).

Gene expression profile and cluster analysis

Based on deep sequencing of the three libraries used in this study, 1,804 genes were identified to be differentially-expressed between AGS1 and AGS3. Moreover, 1,038 genes were identified between AGS1 and AGS4, while 340 genes were identified between AGS3 and AGS4 (Figure 1B). The Venn diagram shows the distribution of expressed genes among the three samples (Figure 1C). A total of 2,181 genes were identified to be significantly differentially-expressed from the three groups, in which 20 proteins were common to all three groups. A heatmap was constructed from the data obtained for differentially-expressed genes. Clustering analysis showed that these genes were differentially-expressed among three groups (Figure 1D).
**Helicobacter pylori** and gastric carcinogenesis

**Functional classification and enrichment of genes regulated by H. pylori or H. pylori CagA**

Annotations of significantly differentially-expressed genes were analyzed. Functional categories of the differentially-expressed genes are shown in Figures 2-4.

For AGS1 and AGS3, the top 20 enriched GO terms within each major functional category are shown in Figure 2, according to P-values of enriched GO terms. The top four most enriched GO terms under “biological process” included “cellular response to stress”, “cell death”, “death”, and “programmed cell death”. The top four most enriched GO terms under “cellular component” included “cytoplasm”, “nuclear lumen”, “nuclear part”, and “nucleoplasm”. In the molecular function category, “protein binding”, “DNA helicase activity”, “binding”, and “anion binding” were the most prominent categories.

For AGS1 and AGS4, the top 20 enriched GO terms within each major functional category are shown in Figure 3. The top four most...
**Figure 3.** Top 20 GO annotation terms of differentially-expressed genes for AGS1 and AGS4, according to $P$ values of enriched GO terms.
Figure 4. Top 20 GO annotation terms of differentially-expressed genes for AGS3 and AGS4, according to P values of enriched GO terms.
Figure 5. Top 20 enriched pathways from KEGG, BioCyc, Reactome, and Panther databases of the differentially-expressed genes for AGS1 and AGS3 (A), AGS1 and AGS4 (B), AGS3 and AGS4 (C), P<0.05.
enriched GO terms under “biological process” included “response to organic substance”, “cellular response to chemical stimulus”, “cell death”, and “death”. The top four most enriched GO terms under “cellular component” included “cytoplasm”, “synapse”, “plasma membrane part” and “post-synapse”. In the molecular function category, “protein binding”, “binding”, “enzyme binding”, and “ubiquitin-like protein ligase binding” were the most prominent categories.

For AGS3 and AGS4, the top 20 enriched GO terms within each major functional category are shown in Figure 4. The top four most enriched GO terms under “biological process” included “mitotic sister chromatid segregation”, “sister chromatid segregation”, “chromosome segregation”, and “mitotic cell cycle”. The top four most enriched GO terms under “cellular component” included “condensed nuclear chromosome, centromeric region”, “condensed nuclear chromosome kinetochore”, “spindle” and “condensed chromosome”. In the molecular function category, “identical protein binding”, “2’-5’-oligoadenylate synthetase activity”, “arachidonate 15-lipoxygenase activity”, and “protein homodimerization activity” were the most prominent categories.

**Pathway analysis of genes regulated by H. pylori or H. pylori CagA**

Differentially-expressed genes were mapped to a pathway database, aiming to identify biological pathways operating during gastric carcinogenesis associated with H. pylori infections or H. pylori CagA. Whole analysis results, using KEGG, BioCyc, Reactome, and Panther databases, are shown in Figure 5. Results of KEGG analysis are shown in Supplementary Figure 1.

For AGS1 and AGS3, the top 20 enriched pathway terms are shown in Figure 5A. Differentially-expressed genes regulated by H. pylori infections were most enriched in “cell cycle”, “DNA strand elongation”, “mitotic G1-G1/S phases”, and “cellular responses to stress”. The top 20 enriched KEGG terms (Supplementary Figure 1A) showed that “cell cycle”, “DNA replication”, “metabolic pathways”, and “base excision repair” were most prominent.

For AGS1 and AGS4, the top 20 enriched pathway terms are shown in Figure 5B. Differentially-expressed genes regulated by CagA H. pylori infections were most enriched in “MAPK signaling pathway”, “Cytokine signaling in immune system”, “cellular responses to stress”, and “development biology”. The top 20 enriched KEGG terms (Supplementary Figure 1B) showed that “MAPK signaling pathway”, “dorso-ventral axis formation”, “bladder cancer”, and “FoxO signaling pathway” were most prominent.

For AGS3 and AGS4, the top 20 enriched pathway terms are shown in Figure 5C. Differentially-expressed genes regulated by H. pylori CagA were most enriched in “cell cycle mitotic”, “cell cycle”, “M phase”, and “mitotic prometaphase”. The top 20 enriched KEGG terms (Supplementary Figure 1C) showed that “alanine, aspartate, and glutamate metabolism”, “cell cycle”, “small cell lung cancer”, and “NF-kappa B signaling pathways” were most prominent.

**Experimental validation via real-time quantitative PCR**

RNA-seq data were validated using real-time-PCR of eight genes. As shown in Figure 6, real-time quantitative PCR expression levels were generally consistent with changes identified by RNA-seq, suggesting that results of RNA-seq were reliable. IL-8, IL-33, EGFR, VEGFA, BCL2L1, TFF1, MUC13, and MUC5AC expression levels were upregulated after H. pylori infections. Regulation of IL-8, IL-33, BCL2L1, TFF1, MUC13, and MUC5AC genes were shown to be CagA-dependent. EGFR and VEGFA expression changes induced by H. pylori did not depend on CagA.

**Discussion**

H. pylori infections are a critical factor in the development of upper gastrointestinal diseases, including peptic ulcers, gastric cancer, and gastric mucosa-associated lymphoid-tissue (MALT) lymphoma. Eradication of H. pylori reduces the progression of atrophic gastritis and prevents metachronous gastric cancer [5, 19]. A recent Kyoto Global Consensus Report on H. pylori gastritis and the Maastricht V Consensus Report recommended H. pylori eradication therapy after infection [20, 21]. H. pylori has been hurled into the spotlight as an important pathogenic factor of gastric diseases, especially gastric carcinogenesis.

During infection, H. pylori CagA is delivered into gastric epithelial cells. This has been consid-
Helicobacter pylori and gastric carcinogenesis

Interleukin 8 (IL-8) was upregulated significantly. This is well-known as a mechanism of H. pylori infected gastric cells [22]. The current study used IL-8 as a positive control marker to certify the infection model were successfully published. In this study, IL-33 was upregulated. IL-33 is a cytokine that acts as a nuclear factor. It has been found to be related with development of many kinds of tumors [23]. IL-33 is expressed in gastric foveolar epithelial cells and is released upon mucosal damage. Downstream signaling of IL-33 induces the alternative activation of macrophages, which then secretes more IL-33 to create a self-sustaining signaling network. It promotes progression of proliferative metaplasia. In the presence of persistent injury and chronic inflammation, risks for gastric cancer development are mediated by IL-33 [24]. Further exploration should be performed concerning the mechanisms of IL-33 in H. pylori-induced gastric cancer.

Epidermal growth factor receptor (EGFR) is a cell surface protein of the EGF family. In tumors of epithelial origin, EGFR overexpression is at a high frequency. This may predict worse survival. EGFR upregulation has been found as the downstream of ROS generation. It is critical for gastric cancer metastasis [25, 26]. Vascular endothelial growth factor (VEGF) signaling synergizing with EGFR can promote epithelial cancer development [27]. In the current study, EGFR and VEGFA were upregulated after H. pylori infections, indicating EGFR and VEGFA might play a critical role in H. pylori related gastric cancer. Moreover, programmed cell death protein ligand 1 (PD-L1) expression has been shown to be regulated by the oncogenic driver EGFR in non-small cell lung cancer [28]. It is an important immune-checkpoint receptor of immune cells and plays a crucial role in tumor-immune-escape and immunotherapy. The term “programmed cell death” was enriched in this study. However, this study did not find that expression of PD-L1 or its receptor PD-1 to be regulated by H. pylori. More studies should be performed, investigating programmed cell death proteins and their ligands in the process of H. pylori-infected gastric carcinogenesis.

BCL-XL and MCL-1 are BCL-2 anti-apoptotic proteins. BCL2L1 (BCL-XL gene) has been reported as an anti-apoptotic target gene [29]. Its overexpression increases cell migration and invasion, facilitating tumor cells to form vasculogenic structures [30]. MCL-1 plays an important role in cervical cancer and has shown potential as a therapeutic target [31, 32]. The current study found significant upregulation of BCL2L1 and MCL-1 in H. pylori-infected AGS cells, indi-
Helicobacter pylori and gastric carcinogenesis

cating that after H. pylori-infection, cells undergo tumorigenesis while DNA damage occurs. It is necessary to investigate whether BCL2L1 or MCL-1 inhibitors have a therapeutic function in H. pylori-infected gastric cancer.

Mucins are major ingredients of the mucus layer attached to gastric mucosa. There is a complex interaction between mucins and H. pylori. Secreting urease and elevating pH, H. pylori can liquefy surrounding mucins, enabling it to move mucins in contact with epithelial cells. Host mucins, in turn, have shown antibiotic effects fighting against H. pylori to control its growth and aggressiveness [33]. Some studies have shown a probable link between H. pylori infections, mucins, and gastric cancer, but mucins expression changes related with H. pylori or gastric cancer have not been clarified [34-36]. In this study, MUC13 and MUC5AC expression levels were upregulated by H. pylori and were CagA-dependent.

Trefoil factor 1 (TFF1) is expressed in a tissue-specific manner in surface mucous cells in gastric mucosa. Some studies have focused on the role of TFF1 in H. pylori-mediated gastric diseases. Serum TFF1 levels have been reported to be higher in populations with H. pylori infections than in those without H. pylori infections [37]. However, some studies have suggested that TFF1 expression is independent of H. pylori infections [38]. Other studies have reported that loss of TFF1 expression is involved in H. pylori-induced gastric carcinogenesis, indicating that TFF1 provides tumor-suppressor functions and that its expression is downregulated when H. pylori infections are present [39]. Recent studies have shown that TFF1 can bind to gastric mucins and help locate H. pylori in a discrete layer of gastric mucus, inhibiting H. pylori contact with epithelial cells [40]. Present results showed the upregulation of mucins and TFF1, hypothesizing that these could help the mucus layer reflect and support host fight against H. pylori infections. Functions of these gene expression changes should be explored more thoroughly.

According to previous studies, cell damage is serious when H. pylori infection MOIs of 50:1 and 100:1 was used. Some drugs could not reverse serious cell damage. The current study used a MOI of 25:1. On one hand, there are fewer experimental reports on low-concentra-

Acknowledgements

This study was funded by the National Natural Science Foundation of China (Grant No. 817-00496) and Interdisciplinary Medicine Seed Fund of Peking University (BMU2017MX015).

Disclosure of conflict of interest

None.

Address correspondence to: Yanyan Shi, Peking University Third Hospital, Research Center of Clinical Epidemiology, Beijing 100191, P.R. China. E-mail: 0729xst@163.com; shiyanyan@bjmu.edu.cn

References

Helicobacter pylori and gastric carcinogenesis


[34] Shi D, Qiu XM and Bao YF. Effects of helicobacter pylori infection on MUC5AC protein expression in gastric cancer. Future Oncol 2013; 9: 115-120.


Supplementary Figure 1. Top 20 enriched KEGG pathways of differentially-expressed genes for AGS1 and AGS3 (A), AGS1 and AGS4 (B), AGS3 and AGS4 (C), P<0.05.