# *Helicobacter*-human dual-species transcriptomics *in* *vitro* and *in* *vivo*

Kelly M. Robinson1, Lindsay Morningstar-Wright2, Matthew Chung1, Nikhil Kumar1, Javier Torres3, Thomas G. Blanchard2, Julie C. Dunning Hotopp1,4\*

1Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA

2Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD, USA

3Unidad de Investigación en Enfermedades Infecciosas, UMAE Pediatría, IMSS, México City, DF, México

4Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA

\*To whom correspondence should be addressed: jdhotopp@som.umaryland.edu

KMR: krobinson@umaryland.edu

LMW: lmorn001@umaryalnd.edu

MC: [mattchung@umaryland.edu](mailto:mattchung@umaryland.edu) (ORCID: 0000-0002-9545-523X)

NK: nkumar@som.umaryland.edu

JT: jtorresl57@yahoo.com.mx

TJB: tblanchard@peds.umaryland.edu

JCDH: [jdhotopp@som.umaryland.edu](mailto:jdhotopp@som.umaryland.edu) (ORCID: 0000-0003-3862-986X)

## Abstract

**Background**: Gastric cancer is responsible for the second highest incidence of cancer-related deaths and is the fourth most common type of cancer. One of the main etiological factors contributing to gastric cancer is the presence of *Helicobacter pylori* in the stomach*.* *H. pylori* is the only bacteria considered to be a carcinogen, due to its ability to promote carcinogenesis in gastric tissue. We sought to further investigate differentially expressed genes by co-culturing *H. pylori* with gastric epithelial cells as well as assessing the transcriptome of gastric samples that are colonized with *H. pylori*.

**Methods**: N87 gastric epithelial cells were cultured with WT *H. pylori* 26695 or *cag-* *H. pylori* 26695. RNA was extracted and sequenced from (a) the co-cultures, (b) N87 cells alone, and (c) both of the *H. pylori* strains in tissue culture media at 2, 4, and 24 h. Further RNA samples were extracted from a tumor-adjacent pair, an additional tumor sample, and one intestinal metaplasia sample.

**Results**: Very few transcriptomic differences were observed between the WT and *cag*- *H. pylori* strains across the time course. Similarly, few differences were observed in the transcriptome response of the N87 gastric epithelial cells infected with the WT and *cag- H. pylori* strains. At 24 h, we find that the N87 cells upregulate pro-inflammatory pathways in response of *H. pylori* infection, which was also observed in one of the gastric tumor samples analyzed.

**Conclusions**: This study is the first in-depth transcriptomics analysis of *H. pylori* and human cells in co-culture. *H. pylori* strains in co-culture with and without the pathogenicity island elicited similar responses from the gastric cells and had similar transcriptomes despite strain differences. Overall, this work highlights the benefit of using a dual-species transcriptomics approach on samples in co-culture and how this experimental approach translates to gastric tissue samples.

## Keywords

*Helicobacter pylori*, gastric cancer, *cagA*, *cagE,* transcriptomics

## Introduction

*Helicobacter pylori* is a Gram-negative bacterium found in the stomach of ~50% of humans across the world. *H. pylori* has been thought to colonize humans for an immense period of time that precedes human migration out of Africa (1), with *H. pylori* infections being more common in developing countries due to low socioeconomic status, ethnicity, and household crowding (2). Infection with *H. pylori* can cause many health risks and diseases ranging from peptic ulcers (3), gastritis (4), gastric mucosa associated lymphoid tissue (MALT) lymphoma (5) and gastric adenocarcinoma (6, 7). *H. pylori* is responsible for 36.3% of infection-related cancers worldwide including 50.3% of infection-related carcinomas in developed countries and 32.7% in less developed countries (8). *H. pylori* contributes to gastric carcinoma and gastric MALT lymphoma by causing persistent gastric inflammation (9, 10) through altering host signaling pathways (9) and methylation of host genes (11).

Similar to infection with most microbes, the presence of *H. pylori* invokes an inflammatory response in host cells. However, compared to other tissues, there are relatively few immune cells in the stomach (12), allowing *H. pylori* to use its host’s Th1 immune response to obtain nutrients from damaged tissue (13).

Some strains of *H. pylori* contain the cytotoxin-associated gene(*cag*) pathogenicity island (PAI), a 40 kb region that contains approximately 26 genes encoding for a bacterial type IV secretion system and the oncogene, cytotoxin-associated gene A, *cagA* (14, 15). Upon adherence of *H. pylori* to gastric epithelial cells, the CagA protein is translocated into the gastric epithelial cells via the PAI-encoded type IV secretion system. Once inside the gastric cells, CagA can alter numerous signaling pathways. The mitogenic signaling pathway is upregulated by CagA through activation of the MAP kinases MEK1/2 and ERK1/2 (32). Activation of MEK1/2 and ERK1/2 causes increased expression of *c-fos* and *c-jun*, leading to the activation of the AP-1 transcription factor (9, 32, 33). CagA also affects host signaling through its tyrosine-phosphorylation site, which is phosphorylated by host Src kinases (34). Phosphorylated CagA can bind to SHP-2 and induce a “hummingbird” phenotype, characterized by gastric epithelial cell elongation, motility, migration, and adhesion and thought to be associated with malignant transformation (34). The “hummingbird” phenotype can also be triggered by interactions between CagA and Grb2, which activate the Ras/MEK/ERK pathway (35).

Numerous studies have reported that the presence *H. pylori cag*+infections have exacerbated the risk of gastritis, atrophy, dysplasia, and gastric adenocarcinoma relative to their *cag*- counterparts {Noto, 2012 #1457}{Torres, 1998 #1458}{Blaser, 1995 #1459}{Kuipers, 1995 #1460}.

Despite *cagA* typically being associated with higher risk of progression to gastric cancer, other genes encoded on the *cag* PAI are essential for CagA secretion and can alter the host response on their own. For example, the *cag* PAI-encoded type IV secretion system has been shown to induce pro-inflammatory response on its own via IL-8 induction (36). The type IV secretion system can also deliver peptidoglycan, causing the bacterium to be recognized by the host through the intracellular pathogen recognition receptor Nod1, leading to pro-inflammatory response (37).

We investigated the transcriptional effect of the *cag* PAI in inducing its host’s proinflammatory response, using a dual-species transcriptomics approach. Gastric epithelial cells were co-cultured with *cag*+ and *cag*- *H. pylori* 26695 strains and the human and *H. pylori* transcriptomes were evaluated over the course of 24 h in co-culture. Additionally, we also assess the human and *H. pylori* transcriptomes *in vivo*,using tumor, adjacent, and intestinal metaplasia samples from patients.

## Results

### Validation of *cag* mutant

A *cag-*knockout strain of *H. pylori* 26695 was used where all 26 genes of the *cag* pathogenicity island are replaced with a chloramphenicol resistance gene cassette (39). To validate the strain, weused whole genome sequencing to assess the sequencing depth of two replicates each of the *cag*+ and *cag*- strains over the *cag* PAI. Aligning the genomic reads from the *cag*- strain to the *H. pylori* 26995 reference genome shows that 37,386 bp (547,165 bp-584,551 bp) have 0 sequencing depth (**Supplemental Figure 1**). This region corresponds to all 26 protein-coding genes in the *cag* PAI*.* Comparatively, the two *cag*+ replicates have averages of 135.1x and 200.5x sequencing depth over the same region, confirming the knockout of the *cag* PAI in the *cag*- strain (**Supplemental Figure 1**).

Additionally, we also used RNA-Seq to measure the transcript per million (TPM) values of the 26 *cag* PAIgenes in the *cag*+ and *cag*- *H. pylori* 26695 strains in both tissue culture media and in co-culture with N87 gastric epithelial cells. In the *cag*- strain, TPM values indicated that there was a low level of *cag* transcription ranging from 0-26.5 TPM (average: 1.6 TPM)while in the *cag+* strain, transcription of the PAI ranged from 27.7-9,707 TPM (average: 926.8 TPM). The low TPM values in the *cag-* strain are likely a result of read pairs aligning to the portion of the *cag* PAI surrounding where the chloramphenicol resistance gene cassette was inserted.

### Differentially expressed *H. pylori* genes *in vitro*

The *H. pylori* strains differ in the presence/absence of *cag* PAI. To assess whether the presence of the *cag* PAI has affects *H. pylori* transcription in co-culture*,* we used RNA-Seq to conduct a differential expression analysis between the *cag*+ and *cag*- *H. pylori* 26995. Both strains of *H. pylori* were cultured in broth and transferred to tissue culture media with or without N87 cells. Samples were done in triplicate and sampled for sequencing at 2, 4, and 24 hr post co-culture, with the exception of the 24 h *cag*- *H. pylori* co-culture samples having only two replicates. To minimize erroneous read mappings in multi-species RNA-Seq experiments firstHISAT2 {Kim, 2015 #1445} a combined reference consisting of the human (Ensembl: GRCh38) and *H. pylori* 26695 genomes (RefSeq: NC\_000915.1), allowing for spliced mappingsAll reads mapping to the *H. pylori* genome were then taken and mapped to only the *H. pylori* genome, disallowing spliced mappings. from the resulting alignment file were calculated using the prokaryotic RNA-Seq quantification tool FADU (ref). A rarefaction analysis shows that 5 samples were not sequenced to saturation and thus excluded from all downstream analyses (**Supplemental Figure 2**).

A differential expression analysis across all samples reveals that all 1,445 protein-coding genes in the *H. pylori* 26995 genome are differentially expressed. A principal component analysis (PCA) splits the transcriptional profiles of the samples groups into 3 distinct clusters consisting of (a) both the *cag*+ and *cag*- strains taken before co-culture; (b) both the *cag*+ and *cag*- strains taken 24 h post-co-culture; and (c) all other samples (**Figure 1a**). Similarly, hierarchical clustering shows that both the *cag*+ and *cag*- strains cluster together both before co-culture. Additionally, the *cag+* and *cag*- strains cluster together 24 h post co-culture (**Figure 2b**). Hierarchical clustering also shows that both the *cag*+ and *cag*- co-culture samples taken at 2 and 4 h all cluster together, with the *cag*+ and *cag-* broth samples at 2 and 4 h form another cluster. Transcriptional differences were noted in the transcription of *H. pylori* strains in co-culture, *H. pylori* strains in broth, and those in tissue culture media, but *H. pylori* transcription was not observed to differ based on the presence of the *cag* PAI.

WGCNA was used to perform unsupervised clustering of the *H. pylori* genes into modules of genes based on expression profile similarity across all samples (**Figure 2**). Each individual module was further split into two clusters depending on whether the expression pattern of a gene has a higher Pearson correlation to the eigengene or the inverse eigengene. In total, WGCNA returned 24 expression modules, with 19 of the modules having ≥10 genes (**Figure 2**). For each expression module, significantly over-represented functional terms were identified using Fisher’s exact test (FDR < 0.05).

Only one cluster of genes illustrated any difference between the *cag*+ and *cag*- strains (**Figure 2, ⬛**). This module of genes contains 31 genes upregulated in only the *cag*+ samples and as expected, all 26 *cag* PAI genes are identified in this cluster. The *cag* PAI genes and HP0120, a protein of unknown function that has been predicted to be an outer membrane protein (43), are grouped together in this block of genes that is consistently induced in *cag*+ strains and repressed in *cag*- strains. The similarity in expression profiles between HP0120 and the *cag* PAI genes indicates that expression of this predicted outer membrane protein may be affected by alterations to the *cag* PAI or may be important to pathogenesis in *cag*+ strains.

Despite the paucity of expression modules describing transcriptional differences between the *cag*+ and *cag*- strains, we observed several expression modules describing transcriptional differences based on the time course or the culture method. The largest recovered expression module consists of 507 genes that are upregulated specifically at 24 h in both the *cag*+ and *cag*- *H. pylori* strains in tissue culture media alone (**Figure 2, ⬛**). While there were no over-represented functional terms, this module contains several predicted outer membrane proteins (HP0726, HP1055, HP1327, HP1467) and a virulence factor (HP1407). In the inverse module upregulated only at 24 h in the co-culture samples, we find the outer membrane protein and virulence factor *oipA*. The *oipA* gene encodes an outer membrane protein and is a virulence factor associated with enhanced inflammation and IL-8 production (39).

A expression module containing 40 genes upregulated in both *cag*+ and *cag*- *H. pylori* 2 and 4 h in cell culture media, regardless of the presence of N87 cells, was significantly over-represented in genes encoding for structural constituents for ribosomes (**Figure 2, ⬛**). Similarly, a set of 28 *H. pylori* genes upregulated only in the co-culture samples and a set of 15 genes upregulated at 2 and 4 h in co-culture alone were both over-represented in ribosomal proteins (**Figure 2, ⬛**).

We identified a module of 24 genes over-expressed in only the 24 h samples that is over-represented in genes with hydrolase activity that specifically acts on carbon-nitrogen bonds, with the exception of peptide bonds (**Figure 2, ⬛**). Additionally, we identify another module containing 20 genes that are downregulated in broth and at 24 h in tissue culture media alone, but not in co-culture (**Figure 2, ⬛**). This module is over-represented in genes encoding for SabA, a N-terminal extracellular adhesion domain. We also observe a module of 15 genes that appears to be upregulated only in the *cag*- strain at 24 h in tissue culture media only and 2, 4, and 24 h in co-culture that is over-represented in genes that encode for proteins with structural molecule activity (**Figure 2, ⬛**). However, the incongruence between the *cag*- strain in tissue culture media alone and co-culture makes it unclear whether this set of genes is truly specific to the *cag*- strain or an artifact of the WGCNA clustering method.

We also find the *ureE* gene to be relatively upregulated only in co-culture at 2, 4, and 24 h (**Figure 2, ⬛**). The *ureE* gene encodes a urease accessory protein that contributes to nickel sequestration and is required for urease production (44). Urease is essential for *H. pylori* colonization as it converts urea to ammonia and neutralizes the stomach acidity (45). While *ureE* was detected in this cluster, other urease genes were not. This is particularly noteworthy as all urease accessory proteins (UreIEFGH) are encoded in an operon, transcribed, and then cleaved into separate mRNA (46). While it is unclear why *ureIEFGH* would not be present in the same cluster despite their presence in an operon, as both urease operons are regulated by mRNA decay in response to pH (46). However, transcripts of *ureIE* and *ureFGH* can be detected in normal broth, *ureIE* and *ureEFGH* can be detected at pH 6, and *ureIE* is more stable at pH 7 (46). The co-culture specific upregulation of *oipA* and *ureE* may be representative of a set of genes that are that are induced in response to gastric epithelial cells, but not when cultured in broth alone.

### Differentially expressed human genes *in vitro*

To determine the difference in host response when infected with *H. pylori cag*+ and *cag*- strains, we examined human gene expression in the uninfected N87 gastric epithelial cells and N87 cells exposed to *cag*+ and *cag*- strains of *H. pylori* for 2, 4, and 24 h. Samples were taken in triplicate, sequenced, and aligned using the alignment-free RNA-Seq quantification tool kallisto {Bray, 2016 #1447} to a combined human and *H. pylori* reference genome. A rarefaction curve shows that all samples taken reach saturation and of the 188,753 human transcripts analyzed, 919 genes were determined to be differently expressed (FDR < 0.05). A PCA of the differentially expressed genes from the 7 different samples reveals 3 discrete clusters of samples consisting of (1) the N87 cells alone; (2) the 24 h N87 + HP WT and *cag-* samples; and (3) the 2 and 4 h N87 + HP WT and *cag-* samples (**Figure 3a**). A similar clustering pattern was recovered using a hierarchical clustering analysis (**Figure 3b**), indicating that the differentially expressed genes recovered across the time course primarily reflects the transcriptional differences to *H. pylori*, regardless of the presence of the *cag* PAI. When edgeR was used to identify differentially expressed genes from pairwise comparisons of N87 + *H. pylori* WT and *cag*- for 2, 4, and 24 h, only 1, 0, and 2 differentially expressed transcripts were recovered, respectively (FDR <0.05). This suggests that when looking across the time course, the presence or absence of the *cag* PAI did not greatly influence the host response.

Using WGCNA, the 919 differentially expressed human genes were divided into 17 expression modules, with only 4 modules containing ≥10 transcripts (**Figure 4**). For each of the 4 modules, we used a functional term enrichment analysis and the Ingenuity Pathway Analysis (IPA) (53) canonical pathway analysis pipeline to identify over-represented functions in each of the modules. The largest module contains 339 and 215 differentially expressed transcripts up- and down-regulated in the 24 h co-culture samples, respectively. The 339 up-regulated genes were identified to be significantly over-represented in cholesterol biosynthesis, sialidase activity, neutrophil degranulation, and translation. Similarly, the IPA canonical pathway analysis identified this subset of transcripts to be significantly enriched in EIF2 and EIF4 signaling; cholesterol biosynthesis, protein synthesis, and mTOR signaling. The 215 genes downregulated at the 24 h timepoint are over-represented in intermediate filaments, oxidoreductase activity, and protein folding and binding. The IPA canonical pathway analysis indicates that this subset of genes is enriched in genes with roles in the sirtuin signaling pathway, mitochondrial dysfunction, and death receptor signaling.

The second largest module has an expression pattern similar to the largest module, in that it contains genes differentially expressed at specifically the 24 h timepoint. The major module contains 182 transcripts downregulated at the 24 h timepoint that are over-represented in ubiquitin-dependent protease activity, mitotic cell cycle regulation, MAPK signaling, and NF-κB signaling. Similarly, IPA identifies the major module to be over-represented in encoded genes with roles in protein ubiquitination, IL-17A in psoriasis, BAG2 signaling, and PI3K/AKT signaling. The minor module contains 76 genes upregulated in the 24 h samples that is significantly enriched in nonsense-mediated mRNA decay and translation, with IPA also identifying an enrichment in EIF2 signaling, integrin signaling, and Cdc42 signaling.

The next largest module contains 28 genes that are gradually upregulated from 2 to 4 h in the co-culture samples. Both the functional term and IPA canonical pathway analyses identify a significant over-representation in amino acid metabolism, including tRNA charging, amino acid transport, and amino acid biosynthesis and degradation. The last module contains 20 genes that are upregulated at 4h co-culture. These genes are over-represented in PFKFB4 signaling, AMPK signaling, and glycolysis and gluconeogenesis.

Because many of the over-represented functional terms in the largest expression module contained proinflammatory roles, we investigated potential upstream regulators responsible of this observed enrichment. Using the IPA upstream regulator analysis, we identified many proinflammatory cytokines, including IL-1α, IL-1β, IL-6, TNFα ,TNFSF10, and TNFSF11, as being induced in the gastric epithelial cells in co-culture with *H. pylori* 26695 strains at 24 h.

TNFR1 is a key receptor in TNFα signaling, which contributes to the risk of gastric cancer (55). The inflammasome promotes maturation of and Il-1β (56), while TREM1 stimulates innate immunity by stimulating the release of and TNFα (57). The Th1 pathway, invoked by Th1 cells, is a key part of the immune response to intracellular bacteria (13). In addition to these pathways, numerous cytokine genes were induced in response to *H. pylori* 26695 *cagE*+, consistent with further stimulation of the host immune response. Overall, these analyses highlight the increased immune response at 24 h in gastric cells infected with *H. pylori* 26695, even in the absence of the *cag* PAI.

### Differentially expressed human genes *in vivo*

While the co-culture samples provide a glimpse into the relationships between gastric epithelial cells and strains of *H. pylori,* we sought to interrogate the transcriptional response in gastric tissues and cancers with *H. pylori*. The same IPA analysis was conducted on four gastric tissues samples: two tumor samples (3CG\_046T\_R and 3CG\_051T\_R), one adjacent sample (3CG\_051T\_A\_R), and one intestinal metaplasia sample (4GB011\_R). Illumina paired-end transcriptome reads were quantified using kallisto {Bray, 2016 #1447} with a combined human and *H. pylori* reference genome. Since edgeR could not be used due to a lack of replicates for each sample, the TPM values derived from kallisto were compared to each other for all four samples and log10 ratios were calculated. Because the IPA core analysis only allows for 3000 genes, genes were only kept for subsequent analyses if their log2 ratios for the 4 samples were not all between -2 and 2 and if their expression met a minimum CPM threshold of xx, equivalent to 2,000 reads in the lowest sequenced sample, in at least one sample. From this, the log2 ratio values of the remaining xx genes were used an input to IPA for core and comparative analyses.

We evaluated differentially enriched canonical pathways were evaluated across the four samples (**Figure 5a**). While the tumor sample 3CG\_051T\_R was more similar to the adjacent and metaplasia sample, there were several pathways upregulated in only the other tumor sample, 3CG\_046T\_R. Compared to the other samples, the 3CG\_046T\_R sample was upregulated in numerous pro-inflammatory canonical pathways including the neuroinflammation signaling pathway; leukocyte extravasation signaling; and IL-8, NFAT, and mTOR signaling. Similarly, when we assess associated disease functions for each of the 4 samples, we see an enrichment of genes with innate immunity roles in 3CG\_046T\_R, including the activation and migration of leukocytes and phagocytes (**Figure 5b**). Using the IPA upstream analysis pipeline, 3CG\_046T\_R upregulates genes that are downstream of pro-inflammatory cytokines such as NF-kB, IFNα, IFNg, IL-1, and IL-6 (**Figure 5c**).

## Discussion

### Increased inflammatory response to *H. pylori* 26695

In concordance with the literature (10, 16, 29, 34), multiple inflammatory pathways and genes, were found to be induced in the gastric epithelial cells exposed to *H. pylori* 26695, including MAPK signaling, TNFα, and the other markers of the innate immune response. Despite the induction of an inflammatory response in the gastric epithelial cells, we observed this response to be independent of the presence of the *cag* PAI in *H. pylori.*

### Lack of transcriptional differences between *H. pylori* strains

Few transcriptional differences between the *cag*+ and *cag*- strains of *H. pylori* 26695 were noted, with almost all of the observed differentially expressed genes being between co-culture and broth comparisons. few genes outside of the *cag* PAI were identified in cluster 6 as having different expression patterns between the *cagE*+ and *cagE*- strains of *H. pylori* 26695. Transcriptional differences in the *oipA* and *ureE* genes were identified as the bacteria were transferred from broth to co-culture. Differences were also observed over the time course as virulence genes *vacA* and *flaB* were induced at 2 h and 4 h in both strains of *H. pylori* 26695. Overall, transcription is altered upon the transition from broth to co-culture and these differences can increase over time in response to gastric epithelial cells as the bacteria up-regulate virulence genes.

### Changes in human gene expression over the co-culture time course

One evident trend across this analysis was the increase in differentially expressed genes in response to the *cag*+ and *cag*- strains as the time course progressed. A hierarchical clustering and PCA revealed the 24 h human samples to be most different from the other samples, and that regardless of bacterial strain all samples within a time point were most similar to each other. This suggests that a 24 h time course may reflect early gastric cancer stages (60), and it was the most reflective of the samples in this study. Extension of a time course analysis beyond 24 h may be informative in future studies.

## Conclusion

This co-transcriptomics approach has been a valuable technique to study both the gastric epithelial response to *H. pylori* and the response of *H. pylori* to tissue culture media and gastric epithelial cells. While the transcriptional difference observed was not as profound as expected, we observed the the delayed upregulation of *vacA* in co-culture, the increased expression of *oipA* and *ureE* upon exposure to gastric epithelial cells, and the similarity in inflammatory response between the 24 h co-culture samples and two gastric tumors.

## Methods

### Bacterial strains

Both WT and *cag- H. pylori* strains 26695 were grown on Columbia agar plates (Difco, Detroit, MI, USA) supplemented with 7% horse blood with 20 µg/mL trimethoprim, 6µg/mL vancomycin, 16µg/mL cefsulodin and 2.5µg/mL amphotericin B (all antibiotics from Sigma-Aldrich, St. Louis, MO, USA). Briefly, *cag- H. pylori* 26695 was created and described previously by Yamaoka et al. by amplifying the *cag* PAIand inserting it into the *Eco*RV restriction enzyme site of pBluescriptSK+ (39) followed by insertion of a chloramphenicol resistance gene into an *Eco*47III site of the *cag* PAI fragment (39). Strains were grown at 37 °C in a humidified incubator with 10% CO2. Both WT and *cag-*H*. pylori* 26695 bacteria were harvested from Columbia blood agar plates in 1 mL of Brucella Broth (Difco, Detroit, MI, USA) and transferred to 10 mL *Brucella* broth supplemented with 10% fetal bovine serum (FBS) (Invitrogen, Waltham, MA, USA) and 6 µg/mL vancomycin (Sigma-Aldrich, St. Louis, MO, USA) in T-25 flasks (Corning, Corning, NY, USA). Liquid cultures were maintained at 37°C with 10% CO2. The bacterial density of liquid cultures for each strain was determined by measuring optical density (OD) at 450 nm with comparison to a previously established growth curve. Bacteria were resuspended in gastric cell culture media supplemented with 1% v/v cholesterol at the time of co-culture.

### Culture of human epithelial cell lines

NCI-N87 gastric epithelial cell line was obtained from the ATCC and cultured in 1:1 Dulbecco's Modified Eagle's Medium (DMEM) and Ham's F-12 Nutrient Mixture with L- glutamine (Corning #10090, Corning, NY, USA) supplemented with 10% FBS (Invitrogen, Waltham, MA, USA) and 1% v/v Anti-/Anti- (Gibco, Carlsbad, CA, USA). Cell cultures were grown and expanded at 37 °C with 5% CO2 prior to experimentation.

### Co-culture of WT *H. pylori* 26695 and *cag- H. pylori* 26695 with human NCI-N87 cells

N87 cells were seeded in 24-well plates at a cell density of 1x106 cells/well 24 h prior to co-culture with WT *H. pylori* 26695 and *cag- H. pylori* 26695. Cells were co-cultured at 50:1 bacteria to cell ratio for 2 h, 4 h, and 24 h in triplicate in cell media supplemented with 1% v/v cholesterol.

### DNA and RNA extraction

DNA and RNA were extracted at each time point from cells using the AllPrep DNA/RNA Mini Kit (Qiagen, Cat #80204, Valencia, CA, USA). Total RNA was quantified using the NanoDrop 2000 spectrophotometer, treated in solution with DNase I from the Qiagen RNeasy kit (Qiagen, Valencia, CA, USA), quantified, and stored at -80 °C for transcriptome analysis. DNA and RNA from the gastric samples were extracted using the Qiagen DNA Mini Kit (Qiagen, Valencia, CA, USA) and Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA, USA), respectively.

RNA was extracted from pure cultures of both bacterial strains cultured in tissue culture media using the Qiagen RNeasy Kit (Qiagen, Valencia, CA, USA). The samples were then treated in solution with DNase I from the Qiagen RNeasy kit (Qiagen, Valencia, CA, USA) and underwent a Ribo-Zero reduction (Illumina, San Diego, CA, USA).

### RNA sequencing

*H. pylori* RNA from co-culture was captured using a custom designed Agilent SureSelectXT Library Prep Kit (Agilent Technologies, Santa Clara, CA) designed with probes for 1,441 genes in the *H. pylori* 26695 genome. The only genes that did not have a probe were the *rpmJ* gene and seven hypothetical proteins (HP0024, HP0131, HP0341, HP0511, HP0756, HP0881, HP1097, HP1297). *H. pylori* RNA from pure cultures in tissue culture media was sequenced from libraries constructed from the NEBNext Ultra RNA Library Prep Kit for Ilumina (New England Biolabs, Ipswich, MA, USA). RNA was fragmented with the Covaris E210. Cell line poly-A selected libraries were constructed using the NEBNext Ultra RNA Library Prep Kit for Ilumina (New England Biolabs, Ipswich, MA) according to the manufacturer’s protocol. RNA was fragmented with the Covaris E210. The RNA was purified between enzymatic reactions and the size selection of the library was performed with SPRIselect beads (Beckman Coulter Genomics, Danvers, MA). Indexed libraries were pooled and sequenced on a 2x150 bp run using the Illumina HiSeq 4000 sequencer.

Transcriptomic RNA libraries for gastric tissues samples were polyA-selected and constructed for sequencing with the Illumina TruSeq RNA Library Prep Kit (Illumina, San Diego, CA). RNA was fragmented with the Covaris E210. The RNA was purified between enzymatic reactions and the size selection of the library was performed with SPRIselect beads (Beckman Coulter Genomics, Danvers, MA). All libraries were indexed, pooled, and sequenced on a 2x101 bp run on a Hiseq 2500 sequencer (Illumina, San Diego, CA).

Pass-filtered Illumina reads may undergo one or both of adaptor and quality trimming upon completion of FastQC assessment. Sequence files that showed adaptor contamination of 1% or higher underwent adaptor trimming using Trimmomatic (67) enforcing a minimum Q20 trailing quality for every read. Quality trimming uses fastx\_trimmer (68) to truncate reads where the quality drops below Phred scale quality of 20 (Q20). Reads that fail to meet a minimum read length of 30 bp (or 5bp for miRNA libraries) after adaptor and quality trimming, as well as their mate pairs, are excluded from the trimmed files.

### Genome sequencing

Genomic libraries were constructed for sequencing using the Kapa Hyper Prep with PCR (Kapa Biosystems, Wilmington, MA) according to manufacturer protocol. DNA was fragmented with the Covaris E210. The DNA was purified between enzymatic reactions and the size selection of the library was performed with SPRI-select beads (Beckman Coulter Genomics, Danvers, MA). Genomic libraries from *H. pylori* 26995 *cag*+ and *cag*- strains taken 2 hr post co-culture with N87 cells were sequenced using an Illumina HiSeq 4000 to yield 151 bp reads.

**Validation of *cag* knockout using whole genome sequencing**

Paired-end reads from *H. pylori* 26995 *cag*+ and *cag*- strains were mapped to the *H. pylori* 26995 reference genome (RefSeq: NC\_000915.1) using Bowtie2 {Langmead, 2012 #1443}. The sort and depth functions of SAMtools v1.9 {Li, 2009 #1442}were used to sort the resultant alignment file and assess sequencing depth over the *cag* PAI, respectively.

### RNA-Seq analysis

For analysis of the human data, kallisto {Bray, 2016 #1461} was used to quantify human transcripts with the GRCh38 genome assembly as the reference. The bacterial data was first aligned with HISAT2 (41) to a combined reference consisting of the GRCh38 human genome and the *H. pylori* 26695reference genome (RefSeq: NC\_000915.1), allowing for spliced mappings. Reads mapping to *H. pylori*, were taken and remapped to the *H. pylori* 26695reference genome with HISAT2, disallowing spliced mappings. *H. pylori* genes were then quantified using the prokaryotic RNA-Seq quantification tool FADU (ref). Following these alignments, edgeR v3.24.3{Robinson, 2010 #1452} was used to determine differential expression genes for both the human and *H. pylori* datasets across the study time course (FDR <0.05).

### WGCNA, IPA and functional term enrichment analysis

For both the *in vitro* human and *H. pylori* transcriptomics data sets, WGCNA v1.68 {Pei, 2017 #1463} was used to bin the differentially expressed genes into discrete expression modules. For the human and *H. pylori* expression modules, all modules with ≥10 genes were subject to a functional term enrichment analysis to detect significantly over-represtend GO terms or InterPro descriptions. The human modules were also used as an input for Ingenuity Pathway Analysis (IPA), version 01-07 (Ingenuity Systems) (53) to determine altered expression of canonical pathways, disease functions, and upstream regulators based on FDR values. Pairwise differential expression analyses was also conducted using edgeR (54) for each co-culture time point (WT and *cag*-).

For the *in vivo* human data set, the log2 ratio

In this analysis, z-scores of log2(fold change) were filtered with an absolute value ≥ 2 and p-values <0.05. A comparison analysis was also conducted with the *cagE*+ vs. *cagE*- at 2, 4, and 24 h. These comparison analyses were filtered for p-values <0.05 and z-scores with an absolute value ≥ 1.5. Text files containing z-scores and p-values were exported from IPA and used to make heat maps with heatmap.3 in R version 3.2.0 (69).

## Ethics approval and consent to participate

The University of Maryland, Baltimore Institutional Review Board deemed the study of de-identified nucleic acids from gastric tissue as non-human subjects. Participants were consented by the group of Dr. Javier Torres under Institutional Review Board-CNIC (Mexican National Commission of Scientific Investigation): 2008-785-001 titled, “Infection, inflammation and cancer; Study for the development of diagnostic tests and identification of risk biomarkers, prognosis and prevention.”

## Consent for publication

NA

## Availability of data and materials

The sequencing data for all of the cell line samples, *H. pylori* 26695 wild-type and *cagE* knockout sequencing data has been deposited to the SRA (BioProject #: PRJNA378649). The sequencing data for the gastric tissues samples is in the process of being deposited to dbGaP.

## Competing Interests

The authors declare they have no competing interests.

## Funding

This work was funded by the National Institutes of Health through an NIH Director’s Transformative Research Award (1-R01-CA206188) to JCDH. JT was supported by El Consejo Nacional de Ciencia y Tecnología-Fronteras de la Ciencia (clave 773) and Instituto Mexicano del Seguro Social (FIS/IMSS/PROT/PRIO/13/027).

## Author’s Contributions

KMR, MC, and JCDH wrote the manuscript. KMR, MC, and JCDH edited the manuscript. LMW cultured the N87 cell lines and both strains of *H. pylori* for co-culture experiments and preformed all co-culture experiments. LMW and KMR extracted all DNA and RNA from co-culture experiments. NK preformed all DNase and Ribo-Zero treatments. JT graciously donated DNA and RNA for the gastric tissue samples. MC performed bioinformatics analyses. TJB, LMW, and JCDH contributed to study design and interpretation of results. All authors read and approved the final manuscript.

## References

1. Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, et al. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature. 2007;445(7130):915-8.

2. Everhart JE. Recent developments in the epidemiology of *Helicobacter pylori*. Gastroenterology clinics of North America. 2000;29(3):559-78.

3. Peterson WL. *Helicobacter pylori* and peptic ulcer disease. The New England journal of medicine. 1991;324(15):1043-8.

4. Karnes WE, Jr., Samloff IM, Siurala M, Kekki M, Sipponen P, Kim SWR, et al. Positive serum antibody and negative tissue staining for *Helicobacter pylori* in subjects with atrophic body gastritis. Gastroenterology.101(1):167-74.

5. Parsonnet J, Hansen S, Rodriguez L, Gelb AB, Warnke RA, Jellum E, et al. *Helicobacter pylori* Infection and Gastric Lymphoma. New England Journal of Medicine. 1994;330(18):1267-71.

6. Nomura A, Stemmermann GN, Chyou P-H, Kato I, Perez-Perez GI, Blaser MJ. *Helicobacter pylori* Infection and Gastric Carcinoma among Japanese Americans in Hawaii. New England Journal of Medicine. 1991;325(16):1132-6.

7. Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, Orentreich N, et al. *Helicobacter pylori* infection and the risk of gastric carcinoma. The New England journal of medicine. 1991;325(16):1127-31.

8. Plummer M, Franceschi S, Vignat J, Forman D, de Martel C. Global burden of gastric cancer attributable to *Helicobacter pylori*. International journal of cancer Journal international du cancer. 2015;136(2):487-90.

9. Kim SS, Ruiz VE, Carroll JD, Moss SF. *Helicobacter pylori* in the pathogenesis of gastric cancer and gastric lymphoma. Cancer letters. 2011;305(2):228-38.

10. Peek RM, Jr., Blaser MJ. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. Nature reviews Cancer. 2002;2(1):28-37.

11. Chan AO, Peng JZ, Lam SK, Lai KC, Yuen MF, Cheung HK, et al. Eradication of *Helicobacter pylori* infection reverses *E-cadherin* promoter hypermethylation. Gut. 2006;55(4):463-8.

12. Wang J BT, Ernst PB. Host inflammatory response to infection. In: Mobley HLT MG, Hazell SL, editor. Helicobacter pylori: Physiology and genetics. Washington (DC): ASM Press; 2001.

13. Hazell SL, Lee A, Brady L, Hennessy W. Campylobacter pyloridis and gastritis: association with intercellular spaces and adaptation to an environment of mucus as important factors in colonization of the gastric epithelium. The Journal of infectious diseases. 1986;153(4):658-63.

14. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, et al. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature. 1997;388(6642):539-47.

15. Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, et al. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(25):14648-53.

16. Segal ED, Lange C, Covacci A, Tompkins LS, Falkow S. Induction of host signal transduction pathways by *Helicobacter pylori*. Proceedings of the National Academy of Sciences of the United States of America. 1997;94(14):7595-9.

17. Williams CS, Smalley W, DuBois RN. Aspirin use and potential mechanisms for colorectal cancer prevention. Journal of Clinical Investigation. 1997;100(6):1325-9.

18. Ben Ayed-Guerfali D, Charfi S, Khabir A, Sellami-Boudawara T, Gargouri A, Mokdad-Gargouri R. Clinical and prognosis relevance of COX-2 expression in Tunisian patients with primary gastric adenocarcinoma. Cancer biomarkers : section A of Disease markers. 2016;17(1):67-73.

19. Levi S, Beardshall K, Haddad G, Playford R, Ghosh P, Calam J. Campylobacter pylori and duodenal ulcers: the gastrin link. Lancet. 1989;1(8648):1167-8.

20. Hocker M, Henihan RJ, Rosewicz S, Riecken EO, Zhang Z, Koh TJ, et al. Gastrin and phorbol 12-myristate 13-acetate regulate the human histidine decarboxylase promoter through Raf-dependent activation of extracellular signal-regulated kinase-related signaling pathways in gastric cancer cells. The Journal of biological chemistry. 1997;272(43):27015-24.

21. Berthenet E, Sheppard S, Vale FF. Recent “omics” advances in *Helicobacter pylori*. Helicobacter. 2016;21:14-8.

22. Eaton KA, Morgan DR, Krakowka S. *Campylobacter pylori* virulence factors in gnotobiotic piglets. Infection and immunity. 1989;57(4):1119-25.

23. Maixner F, Krause-Kyora B, Turaev D, Herbig A, Hoopmann MR, Hallows JL, et al. The 5300-year-old *Helicobacter pylori* genome of the Iceman. Science. 2016;351(6269):162-5.

24. Montano V, Didelot X, Foll M, Linz B, Reinhardt R, Suerbaum S, et al. Worldwide population structure, long-term demography, and local adaptation of *Helicobacter pylori*. Genetics. 2015;200(3):947-63.

25. Alm RA, Ling L-SL, Moir DT, King BL, Brown ED, Doig PC, et al. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. Nature. 1999;397(6715):176-80.

26. Vinella D, Fischer F, Vorontsov E, Gallaud J, Malosse C, Michel V, et al. Evolution of *Helicobacter:* acquisition by gastric species of two histidine-rich proteins essential for colonization. PLoS pathogens. 2015;11(12):e1005312.

27. Scott DR, Marcus EA, Wen Y, Oh J, Sachs G. Gene expression *in vivo* shows that *Helicobacter pylori* colonizes an acidic niche on the gastric surface. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(17):7235-40.

28. Amilon KR, Letley DP, Winter JA, Robinson K, Atherton JC. Expression of the *Helicobacter pylori* virulence factor *vacuolating cytotoxin A* (*vacA*) is influenced by a potential stem-loop structure in the 5′ untranslated region of the transcript. Molecular Microbiology. 2015;98(5):831-46.

29. Naumann M, Sokolova O, Tegtmeyer N, Backert S. *Helicobacter pylori*: A paradigm pathogen for subverting host cell signal transmission. Trends in microbiology.

30. Cover TL, Blanke SR. *Helicobacter pylori* VacA, a paradigm for toxin multifunctionality. Nature reviews Microbiology. 2005;3(4):320-32.

31. Palframan SL, Kwok T, Gabriel K. Vacuolating cytotoxin A (VacA), a key toxin for *Helicobacter pylori* pathogenesis. Frontiers in cellular and infection microbiology. 2012;2:92.

32. Meyer-ter-Vehn T, Covacci A, Kist M, Pahl HL. *Helicobacter pylori a*ctivates mitogen-activated protein kinase cascades and induces expression of the proto-oncogenes *c-fos* and *c-jun*. Journal of Biological Chemistry. 2000;275(21):16064-72.

33. Hatakeyama M. Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. Nature reviews Cancer. 2004;4(9):688-94.

34. Blaser MJ, Atherton JC. *Helicobacter pylori* persistence: biology and disease. The Journal of clinical investigation. 2004;113(3):321-33.

35. Mimuro H, Suzuki T, Tanaka J, Asahi M, Haas R, Sasakawa C. Grb2 is a key mediator of *Helicobacter pylori* CagA protein activities. Molecular cell. 2002;10(4):745-55.

36. Gorrell RJ, Guan J, Xin Y, Tafreshi MA, Hutton ML, McGuckin MA, et al. A novel NOD1- and CagA-independent pathway of interleukin-8 induction mediated by the Helicobacter pylori type IV secretion system. Cellular microbiology. 2013;15(4):554-70.

37. Viala J, Chaput C, Boneca IG, Cardona A, Girardin SE, Moran AP, et al. Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. Nature immunology. 2004;5(11):1166-74.

38. Shariq M, Kumar N, Kumari R, Kumar A, Subbarao N, Mukhopadhyay G. Biochemical analysis of CagE: A VirB4 homologue of *Helicobacter pylori* Cag-T4SS. PLOS ONE. 2015;10(11):e0142606.

39. Yamaoka Y, Kwon DH, Graham DY. A M(r) 34,000 proinflammatory outer membrane protein (oipA) of *Helicobacter pylori*. Proceedings of the National Academy of Sciences of the United States of America. 2000;97(13):7533-8.

40. Frick-Cheng AE, Pyburn TM, Voss BJ, McDonald WH, Ohi MD, Cover TL. Molecular and Structural Analysis of the Helicobacter pylori cag Type IV Secretion System Core Complex. mBio. 2016;7(1).

41. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology. 2009;10(3):R25.

42. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics (Oxford, England). 2015;31(2):166-9.

43. Voss BJ, Gaddy JA, McDonald WH, Cover TL. Analysis of Surface-Exposed Outer Membrane Proteins in Helicobacter pylori. Journal of bacteriology. 2014;196(13):2455-71.

44. Benoit S, Maier RJ. Dependence of Helicobacter pylori Urease Activity on the Nickel-Sequestering Ability of the UreE Accessory Protein. Journal of bacteriology. 2003;185(16):4787-95.

45. Evans DJ, Jr., Evans DG, Kirkpatrick SS, Graham DY. Characterization of the Helicobacter pylori urease and purification of its subunits. Microbial pathogenesis. 1991;10(1):15-26.

46. Akada JK, Shirai M, Takeuchi H, Tsuda M, Nakazawa T. Identification of the urease operon in Helicobacter pylori and its control by mRNA decay in response to pH. Mol Microbiol. 2000;36(5):1071-84.

47. Bellucci M, Zambelli B, Musiani F, Turano P, Ciurli S. Helicobacter pylori UreE, a urease accessory protein: specific Ni(2+)- and Zn(2+)-binding properties and interaction with its cognate UreG. The Biochemical journal. 2009;422(1):91-100.

48. Yang X, Li H, Lai TP, Sun H. UreE-UreG complex facilitates nickel transfer and preactivates GTPase of UreG in Helicobacter pylori. The Journal of biological chemistry. 2015;290(20):12474-85.

49. Eaton KA, Suerbaum S, Josenhans C, Krakowka S. Colonization of gnotobiotic piglets by *Helicobacter pylori* deficient in two flagellin genes. Infection and immunity. 1996;64(7):2445-8.

50. Tucker TP, Gray BM, Eaton KA, Merchant JL. Helicobacter pylori Induction of the Gastrin Promoter Through GC-Rich DNA Elements. Helicobacter. 2010;15(5):438-48.

51. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology. 2013;14(4):R36.

52. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9.

53. Krämer A, Green J, Pollard JJ, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics (Oxford, England). 2014;30(4):523-30.

54. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics (Oxford, England). 2010;26(1):139-40.

55. Oshima H, Ishikawa T, Yoshida GJ, Naoi K, Maeda Y, Naka K, et al. TNF-alpha/TNFR1 signaling promotes gastric tumorigenesis through induction of Noxo1 and Gna14 in tumor cells. Oncogene. 2014;33(29):3820-9.

56. Martinon F, Burns K, Tschopp J. The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta. Molecular cell. 2002;10(2):417-26.

57. Schmausser B, Endrich S, Beier D, Moran AP, Burek CJ, Rosenwald A, et al. Triggering receptor expressed on myeloid cells-1 (TREM-1) expression on gastric epithelium: implication for a role of TREM-1 in Helicobacter pylori infection. Clinical and experimental immunology. 2008;152(1):88-94.

58. Dixon BREA, Radin JN, Piazuelo MB, Contreras DC, Algood HMS. IL-17a and IL-22 Induce Expression of Antimicrobials in Gastrointestinal Epithelial Cells and May Contribute to Epithelial Cell Defense against Helicobacter pylori. PLoS ONE. 2016;11(2):e0148514.

59. Couper KN, Blount DG, Riley EM. IL-10: The Master Regulator of Immunity to Infection. The Journal of Immunology. 2008;180(9):5771-7.

60. Kim YH, Liang H, Liu X, Lee J-S, Cho JY, Cheong J-H, et al. AMPKα Modulation in Cancer Progression: Multilayer Integrative Analysis of the Whole Transcriptome in Asian Gastric Cancer. Cancer research. 2012;72(10):2512-21.

61. Bronte-Tinkew DM, Terebiznik M, Franco A, Ang M, Ahn D, Mimuro H, et al. Helicobacter pylori CagA activates the Signal Transducer and Activator of Transcription 3 (STAT3) pathway in vitro and in vivo. Cancer research. 2009;69(2):632-9.

62. Byun E, Park B, Lim JW, Kim H. Activation of NF-kappaB and AP-1 Mediates Hyperproliferation by Inducing beta-Catenin and c-Myc in Helicobacter pylori-Infected Gastric Epithelial Cells. Yonsei medical journal. 2016;57(3):647-51.

63. Basque JR, Chenard M, Chailler P, Menard D. Gastric cancer cell lines as models to study human digestive functions. Journal of cellular biochemistry. 2001;81(2):241-51.

64. Conlin VS, Curtis SB, Zhao Y, Moore ED, Smith VC, Meloche RM, et al. Helicobacter pylori infection targets adherens junction regulatory proteins and results in increased rates of migration in human gastric epithelial cells. Infection and immunity. 2004;72(9):5181-92.

65. Roy RK, Hoppe MM, Srivastava S, Samanta A, Sharma N, Tan KT, et al. CEACAM6 is upregulated by Helicobacter pylori CagA and is a biomarker for early gastric cancer. Oncotarget. 2016;7(34):55290-301.

66. Zang M, Zhang B, Zhang Y, Li J, Su L, Zhu Z, et al. CEACAM6 Promotes Gastric Cancer Invasion and Metastasis by Inducing Epithelial-Mesenchymal Transition via PI3K/AKT Signaling Pathway. PLoS ONE. 2014;9(11):e112908.

67. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics (Oxford, England). 2014;30(15):2114-20.

68. Hannon. FASTX-Toolkit. Available from: <http://hannonlab.cshl.edu/fastx_toolkit/index.html>.

69. R Core Team. R: A language and environment for statistical computing Vienna, Austria: R Foundation for Statistical Computing; 2015. Available from: <http://www.R-project.org/>.

70. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Current Protocols in Bioinformatics: John Wiley & Sons, Inc.; 2002.

## Figure Legends

Figure 1. PCA and hierarchical cluster of the differentially expressed *H. pylori* genes

The z-scores of the log2 TPM values for the 919 differentially expressed *H. pylori* genes across all co-culture samples and the *H. pylori* controls were clustered in a (a) principal component analysis and (b) hierarchical clustering analysis. For the principal component analysis, variation observed is given in parentheses next to the axes labels. Bootstrap probabilities for the dendrogram of all samples are indicated next to their corresponding nodes. For both analyses, points are sized relative to their respective transcriptome library sizes.

Figure 2. WGCNA heatmap of differentially expressed *H. pylori* genes

The z-scores of the log2 TPM values for the 919 differentially expressed *H. pylori* genes are displayed on the heatmap. The horizontal color bar indicates the sample for each column of the heatmap. The outermost vertical color bar indicates the WGNCA module assignment for each gene, while the innermost vertical color bar indicates the major (grey) and minor modules (black) within each WGCNA module.

Figure 3. PCA and hierarchical cluster of the differentially expressed human genes

The z-scores of the log2 TPM values for the 1,445 differentially expressed *in vitro* human genes across all co-culture samples and the N87 control were clustered in a (a) principal component analysis and (b) hierarchical clustering analysis. For the principal component analysis, variation observed is given in parentheses next to the axes labels. Bootstrap probabilities for the dendrogram of all samples are indicated next to their corresponding nodes. For both analyses, points are sized relative to their respective transcriptome library sizes.

Figure 4. WGCNA heatmap of differentially expressed humangenes

The z-scores of the log2 TPM values for the 1,445 differentially expressed human genes are displayed on the heatmap. The horizontal color bar indicates the sample for each column of the heatmap. The outermost vertical color bar indicates the WGNCA module assignment for each gene, while the innermost vertical color bar indicates the major (grey) and minor modules (black) within each WGCNA module.

Figure 5. Upstream regulators of gastric response to *H. pylori cagE*+ vs. *cagE*-

The xx *in vivo* human genes chosen for the IPA comparative analysis were used to determine differentially expressed (a) canonical pathways; (b) diseases and functions; and (c) upstream regulators between the two tumor samples, one metaplasia sample, and one adjacent sample. For each of the three analyses, a heatmap was generated to visualize the activation z-score of the four samples for the 30 most significant functional terms in each category.

## Supplemental Figure Legends

Supplemental Figure 1. Sequencing depth across *cag* PAI island in *cag+* and *cag*- *H. pylori* 26995 strains

Two replicates each of the *cag*+ (SRR410345, SRR410345) and *cag*- (SRR7191641, SRR7191642) *H.pylori* 26995 strains were taken after 2 hr of co-culture with N87 cells and used for whole genome sequencing to examine sequencing depth over the *cag* PAI. Paired-end reads from all four samples were mapped to the *H. pylori* 26995 reference and in both *cag*- strains, there is 0 sequencing depth from the 547,165-584,551 positions, corresponding to the *cag* PAI. Comparatively, the *cag*+ strains average 135.1x and 200.5x over the same region.