

Gene expression profiling in human gastric mucosa infected with *Helicobacter pylori*

Véronique J Hofman^{1,2,*}, Chimène Moreilhon^{3,4,*}, Patrick D Brest^{1,4}, Sandra Lassalle^{1,2,4}, Kevin Le Brigand^{3,4}, Dominique Sicard⁵, Josette Raymond⁶, Dominique Lamarque⁷, Xavier A Hébuterne^{1,4,8}, Bernard Mari^{3,4}, Pascal JP Barbry^{3,4,*} and Paul M Hofman^{1,2,4,9,*}

¹Faculty of Medicine, INSERM ERI-21, Nice Cedex, France; ²Laboratory of Clinical and Experimental Pathology, Pasteur Hospital, Nice, France; ³CNRS, Institute of Molecular and Cellular Pharmacology, UMR 6097, Sophia Antipolis, France; ⁴University of Nice Sophia Antipolis, Nice, France; ⁵Department of Bacteriology, Archet Hospital, Nice, France; ⁶Department of Bacteriology, Saint-Vincent de Paul Hospital, Paris, France; ⁷Department of Gastroenterology, Hotel Dieu Hospital, Paris, France; ⁸Department of Gastroenterology, Archet Hospital, Nice, France and ⁹Human Tissue Biobank Unit, Pasteur Hospital, Nice, France

Pathogenic mechanisms associated with *Helicobacter pylori* infection enhance susceptibility of the gastric epithelium to carcinogenic conversion. We have characterized the gene expression profiles of gastric biopsies from 69 French Caucasian patients, of which 43 (62%) were infected with *H. pylori*. The bacterium was detected in 27 of the 42 antral biopsies examined and in 16 of the 27 fundic biopsies. Infected biopsies were selected for the presence of chronic active gastritis, in absence of metaplasia and dysplasia of the gastric mucosa. Infected antral and fundic biopsies exhibited distinct transcriptional responses. Altered responses were linked with: (1) the extent of polymorphonuclear leukocyte infiltration, (2) bacterial density, and (3) the presence of the virulence factors *vacA*, *babA2*, and *cagA*. Robust modulation of transcripts associated with Toll-like receptors, signal transduction, the immune response, apoptosis, and the cell cycle was consistent with expected responses to Gram-negative bacterial infection. Altered expression of interferon-regulated genes (*IFITM1*, *IRF4*, *STAT6*), indicative of major histocompatibility complex (MHC) II-mediated and Th1-specific responses, as well as altered expression of *GATA6*, have previously been described in precancerous states. Upregulation of genes abundantly expressed in cancer tissues (*UBD*, *CXCL13*, *LY96*, *MAPK8*, *MMP7*, *RANKL*, *CCL18*) or in stem cells (*IFITM1* and *WFDC2*) may reveal a molecular switch towards a premalignant state in infected tissues. Tissue microarray analysis of a large number of biopsies, which were either positive or negative for the *cagA* virulence factor, when compared to each other and to noninfected controls, confirmed observed gene alterations at the protein level, for eight key transcripts. This study provides 'proof-of-principle' data for identifying molecular mechanisms driving *H. pylori*-associated carcinogenesis before morphological evidence of changes along the neoplastic progression pathway.

Modern Pathology (2007) 20, 974–989; doi:10.1038/modpathol.3800930; published online 20 July 2007

Keywords: *Helicobacter pylori*; DNA microarrays; tissue microarrays; biopsy

Helicobacter pylori is a prevalent Gram-negative bacterium that infects nearly 50% of the human population. Even if the majority of *H. pylori*-colonized individuals remain asymptomatic, infection with *H. pylori* is strongly associated with gastroduodenal diseases such as chronic active gastritis, peptic ulcers,

and gastric malignancies. Propensity to develop diseases depends on a complex set of factors, including (i) bacterial-associated virulence strains, (ii) host genotype, (iii) environmental parameters and (iv) specific associations among these above-mentioned factors.

Three *H. pylori* virulence factors have been associated with more severe diseases.¹ They correspond to the secreted vacuolating cytotoxin (VacA), the blood group antigen-binding adhesion molecule (BabA) and the gene products of the pathogenicity island (*cagA* PAI).

In addition to the bacterial strain, several host determinants have been shown to influence infection outcome. This has been examined most

Correspondence: Dr PM Hofman, MD, PhD, Department of Pathology, INSERM ERI-21, Faculté de Médecine, avenue de Valombrose, 06107 Nice, Cedex 01, France.
E-mail: hofman.p@chu-nice.fr

*These authors contributed equally to this work.

Received 27 February 2007; revised 12 June 2007; accepted 18 June 2007; published online 20 July 2007

thoroughly for IL-1 β , tumor necrosis factor (TNF)- α , IL-10, and MHC II polymorphisms.^{2,3} Moreover, both *in vitro* and *in vivo* studies have demonstrated that *H. pylori* affects intracellular signal transduction in host cells, thereby leading to the activation of different transcription factors and the induction of proinflammatory cytokines.^{4,5} Interest has also been focused on the 'pre-malignant transition' potential of *H. pylori*. Studies in a model of *H. pylori*-induced gastritis in Mongolian gerbils have provided casual evidence for altered gastric epithelial cell apoptosis and proliferation in the development of gastric adenocarcinoma.⁶

Previous published studies have used expression profiling to investigate the changes in host gene expression induced by *H. pylori* infection. These studies have relied mainly on either *in vitro* culture models or *in vivo* animal models.^{7,8} DNA microarray technologies have also been employed to profile human tissue infected by *H. pylori*.⁹ However, this latter study was performed on a limited number of patients and potential protein alterations were not confirmed by immunohistochemical staining of affected tissue.

In the present work, the gene expression profiles of 42 antral biopsies corresponding to 27 patients infected with different *H. pylori* strains and 15 age-matched patients without histological evidence of *H. pylori* infection were examined by differential screening with long-oligonucleotide microarrays. Similarly, the gene expression profiles of 27 fundus biopsies corresponding to 16 patients harboring different *H. pylori* strains and 11 age-matched patients without histological evidence of *H. pylori* infection were examined on the same microarray platform. Alteration of global gene expression was assessed along with the number of infiltrating polymorphonuclear leukocytes (PMNL) at the site of infection, bacterial density, and the status of *cagA*, *vacA*, and *babA2* expression by the different strains isolated after culture.

Materials and methods

Patients

For the microarray study, a total of 69 patients (41 males and 28 females) undergoing upper digestive endoscopy for the evaluation of dyspepsia and upper abdominal pain at Archet II Hospital, (University of Nice, France), and Cochin Hospital (University of Paris, France) were enrolled between June 2001 and September 2004. The subjects ranged in age from 24 to 40 years (mean, 32 \pm 8). The patients were all French Caucasian residents. Patients were classified by endoscopy as having gastritis ($n=27$), duodenal ulcers ($n=8$), or gastric ulcers ($n=3$). Lesions were not detected in the other patients by endoscopy. Patients with malignancy, immunosuppression, gastrointestinal hemorrhage, previous diagnosis of *H. pylori* infection were

excluded from the study, as were patients treated with aspirin, nonsteroidal anti-inflammatory drugs, or antibiotics during the month preceding the endoscopy. Multiple endoscopic biopsies were obtained from the fundus (27 patients), or from the antrum (42 patients). Six biopsies were taken from each individual per site to assess gene expression, bacterial genotype, and histology, in parallel. The Institution Medical Review Board at the Pasteur Hospital (Nice, France) approved the study and informed consent was obtained from all patients.

For tissue microarray (TMA) analysis, 304 subjects were initially included in this study. Patients were hospitalized to undergo an upper digestive tract endoscopy for objective evidence of GI disease (dyspepsia and/or gastric pain) (169 patients) or before gastroplasties (135 asymptomatic patients). Patients were all French Caucasian residents. Clinical information regarding treatment with NSAID, aspirin, or antibiotics during the two months preceding the endoscopy was obtained from the hospital information system.

Histological Assessment

For diagnosis, three gastric biopsies from the antrum and three gastric biopsies from the fundus were obtained from each patient, then fixed in 10% formalin buffer. Hematoxylin and eosin- and Giemsa-stained sections were examined by two pathologists (VH, PH) who were blinded to the other experimental results. Slides were coded and evaluated histologically for the following parameters: (i) activity of gastritis (PMNL infiltration), (ii) chronicity of gastritis (lymphocytic and plasma cell infiltration), (iii) glandular atrophy, and (iv) intestinal metaplasia. Each parameter was graded as none (0), mild (1), moderate (2), or severe (3–4), according to the updated Sydney classification.¹⁰ Density of bacteria, evaluated on Giemsa-stained sections, was graded as none (0), mild (1), moderate (2), intense (3), and very intense (4).

Detection of *H. pylori* Genotypes and Sequencing

The presence of *H. pylori* was assessed by bacterial culture of biopsies, as described previously.¹¹ *H. pylori* strains were isolated from patients, identified, stored, and recovered, as described previously.¹² *H. pylori* *cagA*, *vacA* (*vacAs1/s2* and *vacAm1/m2*), and *babA2* genotypes were determined by polymerase chain reaction (PCR). Genomic DNA was extracted from *H. pylori* using the High Pure PCR Template preparation kit (Roche, Germany). The different primer sequences used in the present work have been described previously.¹³ Sequences of the PCR products were confirmed by automated sequencing (ABI Prism 310 Genetic Analyser; Perkin Elmer, Branchburg, NJ, USA) using the same primer pairs.

Target Preparation and Hybridization

RNA extraction

Total RNA was isolated from 10 to 15 mg of frozen antral and fundic mucosa biopsies (two biopsies from each site) using Trizol extraction (Invitrogen). The quality of total RNA was assessed by the 28S/18S RNA ratios, determined on an Agilent 2100 Bioanalyzer (Agilent Technologies, Massy, France).

Cy3- and Cy5-labelling

Two microgram of total RNA was amplified with the Amino Allyl MessageAmp aRNA kit (AMBION ref. 1752) according to the manufacturer's instructions.¹⁴ cDNA was then transcribed *in vitro* with aaUTP (amplification) and purified using the PCR purification Kit (Qiagen ref. 28106). Cy3- and Cy5-labelled RNA (Cy3: patient and Cy5: universal RNA conditions) were pooled, concentrated with Millipore Microcon columns YM-30 (ref. 42410) and fragmented with the Ambion RNA Fragmentation reagents (ref. 8740), according to the manufacturer's instructions. All group samples (ie, normal or infected biopsies) were competitively hybridized with a common Universal Human Reference RNA (Stratagene Europe, Amsterdam, The Netherlands).

Probe design, microarray printing, and hybridization

Experiments were performed using a custom designed microarray consisting of 1696 distinct oligonucleotide probes. Genes were selected based on their relevance to inflammation, apoptosis, cell cycle, differentiation, and cytoskeleton.¹⁵ The final list of the 1696 probes spotted on the microarray has been stored in GEO (<http://www.ncbi.nlm.nih.gov/geo>) under the reference GPL1460. It is available on <http://www.microarray.fr> (follow the link to 'human local set'). All probes (ranging from 50 to 52 bases) were resuspended in 50% DMSO at a concentration of 50 μ M. Microarrays were printed with a SDDC-2 (Biorad) on commercial UltraGAPS II slides (aminosilane coated slides for printing DNA Microarrays, Corning ref. 40017). Each oligonucleotide was spotted four times on each slide (two distinct pairs of spots), to reduce positional bias of the fluorescence readout. Microarrays were hybridized in a Ventana hybridization station. Arrays were scanned with a ScanArray Express (version 2.0.19 microarray acquisition system, Packard BioScience, Rungis, France).

Microarray Data Analysis

TIFF images containing the data from each fluorescence channel were quantified with the Genepix pro 5.0 program (Axon instruments) using an irregular features quantification method. Cy3 intensities corresponding to biopsies were then used for analysis. The four replicate spots of the same gene were

processed separately. Intensities from Cy3 were log transformed, then mean-centered and reduced using GeneANOVA (Gene Expression Analysis of Variance) software.¹⁶ A *t*-test was performed on corrected intensities (P -value < 0.001), as suggested by Saeed *et al.*¹⁷ For the analysis of variance, the signal was modelled as proposed by Kerr and Churchill,¹⁸ according to:

$$y_{ijk} = \mu + A_i + D_j + T_k + G_g + (TG)_{kg} + (AG)_{ig} + (DG)_{jg} + \varepsilon_{ijk} \quad (1)$$

where y_{ijk} is the fluorescent intensity of array i and replicate j representing treatment k and gene g ; μ , A_i , D_j , T_k , and G_g are factors representing average signals among the whole experiment, one array, one replicate, one treatment, or one gene, respectively. $(TG)_{kg}$, $(AG)_{ig}$, and $(DG)_{jg}$ are two-factor interactions between treatment and gene, array and gene (spot effects), replicate and gene, respectively. ε_{ijk} represents a residual noise component, modelled as a Gaussian distribution with mean 0 and variance σ^2 . $(TG)_{kg}$ denotes differences in expression for particular treatment and gene combinations that are not explained by the average effects on these treatment and genes. They represent the effects of interest, which were modelled using GeneANOVA.¹⁶

For Figure 3, normalization was performed using the limma package available from Bioconductor (<http://www.bioconductor.org>).¹⁹ Background was evaluated according to Kooperberg.²⁰ Interslide normalization was performed to force a uniform distribution of the Cy5 signal (that corresponded to the universal RNA). Data were then analyzed according to the prediction analysis of microarray software PAM (<http://www-stat.stanford.edu/~tibs/PAM/>).²¹

Ontologies attached to each gene were used to classify altered genes according to main biological themes, using the Expression Analysis Systematic Explorer (EASE) program, available at <http://david.niaid.nih.gov/david/ease.htm>,²² and Mediante, an information system containing diverse information about our probe set (<http://www.microarray.fr>).¹⁵ Hierarchical clustering was performed using Bioconductor, as described previously.²³ One meaningful annotation of the probes was provided by the tissue source of all matching expressed sequence tags (EST). This information was more specifically used to identify genes related to stomach, cancer, fibroblasts, immune, or stem cells.

Semiquantitative Real-Time RT-PCR Analysis

Expression levels of *ATP6V1A*, *GATA6*, *IRF4*, *LY96*, *MMP-7*, and *UBD* (already found regulated by the microarray experiments) and of *RANKL*, interferon (*IFN*) γ , *IL-8* (missed by the microarray experiments) were independently assessed by semiquantitative real-time RT-PCR (sqRT-PCR) using cDNAs syn-

thesized from biopsies. *GAPDH* was used as an internal control. Expression levels were measured by sqRT-PCR using a Quantitect™ Probe PCR kit (Qiagen, Valencia, CA, USA) and the ABI Prism 7900 HT sequence detection system (Applied Biosystems, CA, USA). Results are expressed in arbitrary units.

TMA Construction and Immunohistochemical Studies

Among the 135 control volunteers, 85 patients (group 1) did not use NSAID, aspirin, or antibiotics and were negative for *H. pylori* histologically and by culture. Biopsies performed on this latter group of patients did not show mucosal lesions. Among the 169 patients with gastric symptoms, 129 cases did not take NSAID or aspirin, were positive for *H. pylori* infection and showed chronic active gastritis (55 *cagA*⁻/*vacAs2m2*⁺/*babA2*⁻ strains (group 2) and 74 cases *cagA*⁺/*vacAs1m1*⁺/*babA2*⁺ strains (group 3)).

Representative gastric antral biopsies were selected from hematoxylin and eosin-stained sections for building TMAs. TMAs were set up as described previously.²⁴ Briefly, one tissue core (600 μm in diameter) was obtained from the upper mucosa of each specimen where pits and glands were cut longitudinally. Tissue cores were arrayed into new paraffin blocks using a fine steel needle to produce three TMAs as follows: (i) an array of 222 cores from patients in group 1, (ii) an array of 135 cores from patients in group 2, and (iii) an array of 255 cores from patients in group 3. TMAs of gastric specimens from *H. pylori*-infected patients (groups 2 and 3) contained normal gastric antral mucosa (six tissue cores from biopsies performed on volunteers), which served both as control, and to mark the intercore spacing of 1 mm between core centers. A 4-μm hematoxylin and eosin-stained section was reviewed to confirm the presence of morphologically representative areas of the original lesions.

The sources and dilutions of antibodies used for immunohistochemical analyses were as follows: STAT6, Lab Vision Corporation (Francheville, France), monoclonal (1:200); MMP-7, Lab Vision Corporation (Francheville, France), monoclonal (1:500); UBD, Serotech (Oxford, London), monoclonal (1:200); CXCL13, RD Systems (Abingdon, UK), monoclonal (1:200); MAPK8, Abgent (San Diego, CA, USA), polyclonal (1:100); WFDC2, Abcam (Paris, France), polyclonal (1:500); LY96, Abcam (Paris, France), polyclonal (1:500); and GATA6, RD Systems (Abingdon, UK), polyclonal (1:200). Immunostaining was performed on the automated Ventana Benchmark XT immunostainer (Ventana Medical System, Ilckrich, France) using the manufacturer's deparaffinization, antigen retrieval and detection reagents. After staining, slides were examined by two pathologists (VH and PH). Results were evaluated by the quick score method, as performed previously.²⁵

Results

Histological Results

H. pylori was detected (both on HES- and Giemsa-stained sections) in 27 antral biopsies and in 16 fundic biopsies. The bacterium was also detected in cultures from all of these biopsies. Other antral biopsies (*n* = 15) and fundic biopsies (*n* = 11) had no *H. pylori*, contained normal mucosa and served as controls. All *H. pylori*-infected patients displayed evidence of inflammatory infiltrates with PMNL and mononuclear cell infiltration into mucosa. All of the *H. pylori*-positive patients had mononuclear cell infiltration, whereas PMNL infiltration was seen in 87% of antral biopsies and 49% of fundus biopsies.

Characteristics of *H. pylori* Isolates

In the antrum, genotypic identification of isolates from the 27 infected patients indicated that 14 (52%) harbored *cagA*⁺ strains. Moreover, *H. pylori* isolates belonged to the *vacA s1/m1*, *vacA s2/m2*, *vacA s1/m2*, or *vacA s2/m1* genotypes in 8 (31%), 9 (33%), 8 (30%), and 2 (8%) cases, respectively. Twelve (44%) were *babA2*⁺ strains and the simultaneous presence of *cagA*, *vacA s1/m1*, and *babA2* genes (triple positive) was found in six (22%) *H. pylori* isolates. In contrast, nine *H. pylori* isolates belonged to the *cagA*(-), *vacA s2/m2*, and *babA2*(-) haplotype. In the fundus, genotypic profiles of the *H. pylori* isolates from the 16 infected patients were similar to those identified in antrum. Eight patients (50%) harbored *cagA*⁺ strains. *H. pylori* isolates belonged to the *vacA s1/m1* (*n*: 6), *vacA s2/m2* (*n*: 5), *vacA s1/m2* (*n*: 3), and *vacA s2/m1* (*n*: 2) genotypes. Eight were *babA2*⁺ strains and the simultaneous presence of *cagA*, *vacA s1/m1*, and *babA2* genes (triple positive) was found in four *H. pylori* isolates. Finally, for a given patient, no mixed infection by different strains of *H. pylori* was detected.

Correlation between *H. pylori* Genotypes and Histological Results

Although most *H. pylori*-infected patients displayed evidence of chronic gastric inflammation, there was no significant correlation between different *H. pylori* genotypes and the presence of PMNL and mononuclear cell infiltrates in the antrum or fundus of the stomach (*P* > 0.4). However, we noticed that the presence of the *cagA*⁺, *babA2*⁺ genotype was associated with a higher degree of lymphocytic infiltration in the antrum. To confirm that observation, we examined the correlation between the three main virulence factors (*vacA/cagA*⁺/*babA2*⁺ strain) and bacterial density. PMNL scores were assessed as well. In the antrum, the average bacterial density score was equal to 1 for *vacA s2/m2 cagA*⁻/*babA2*⁻

cultures (in nine independent cultures) and to 3.1 for *vacA s1m1/cagA⁺/babA2⁺* cultures (seven independent cultures) ($P < 0.00001$, Student's unpaired *t*-test). The average PMNL score was equal to 1.9 for tissue with positive culture for *vacA s2/m2 cagA⁻/babA2⁻* and to 2.9 for tissue with positive culture for *vacA s1m1/cagA⁺/babA2⁺* ($P < 0.02$, Student's unpaired *t*-test). A slightly distinct picture was obtained in the fundus, as PMNL infiltration was greater in *vacA s1/m1 cagA⁺/babA2⁺* biopsies than in *vacA s2/m2/cagA⁻/babA2⁻* biopsies (average scores: 2.3 and 1, respectively), whereas bacterial densities were similar between the two groups (2 and 1.8 for *vacA s2/m2 cagA⁻/babA2⁻* strains and *vacA s1/m1/cagA⁺/babA2⁺* strains, respectively).

Transcriptional Regulation of Genes in Antrum and Fundus in *H. pylori*-Infected Biopsies Compared to Noninfected Biopsies

Total RNA was isolated from biopsies of *H. pylori*-infected or noninfected patients. The lowest acceptable value for the 28s/18s RNA ratio was defined as 1.3 for this type of sample (average = 1.6; range = 1.3–1.8).

After hybridization and normalization, 165 genes (92 upregulated and 73 downregulated) were defined as differentially expressed between gastric antrum of *H. pylori*-infected and noninfected patients (Figure 1). Similarly, 47 genes (19 upregulated and 28 downregulated) were defined as differentially expressed between fundic mucosa of *H. pylori*-infected and noninfected patients (Figure 2). Of note, 18 transcripts were similarly regulated in antral and fundic samples (8 upregulated and 10 downregulated).

Gastric antral biopsies

Most of the genes regulated in the antrum were associated with protein families such as receptors, co-receptors, or proteases, or they were linked to the MAPK kinase signalling pathway, to immune response or to apoptotic processes (Figure 1; Table 1).

Receptors and co-receptors Several genes encoding proteins with receptor function were regulated in gastric antral mucosa after *H. pylori* infection. For example, the Toll-like receptors *TLR2*- and the *TLR4*-associated surface protein *Ly96* that confers responsiveness to LPS were upregulated, whereas *TLR3* was downregulated. We also observed an upregulation of *ITGB2* (integrin $\beta 2$) and *VCAM1* (a member of the Ig superfamily that mediates leukocyte-endothelial cell adhesion and signal transduction).

Signal transduction *MAPK8* (*JNK*), *RAC2*, *SLA* (*Src*-like-adaptator) were all upregulated, whereas *FARP2* (*FERM*, *RhoGEF* and *pleckstrin domain protein 2*) was downregulated. *STAT6* was downregulated in parallel with *SOCS2*. The *GATA6* transcription factor was downregulated.

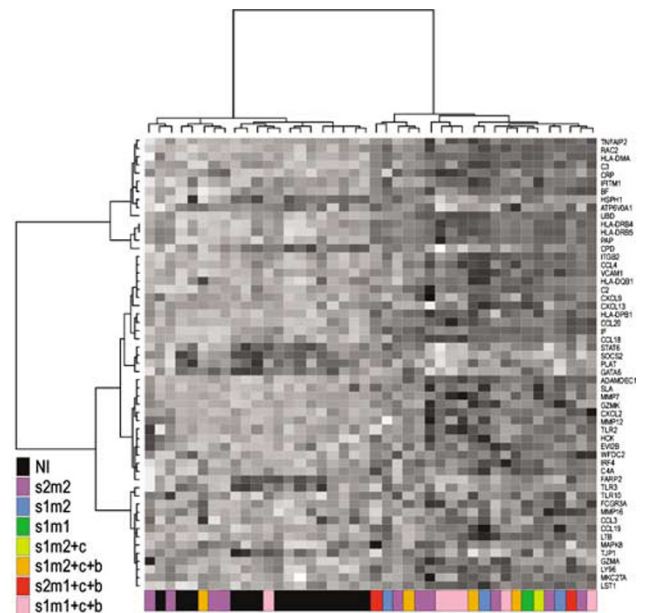


Figure 1 Heatmap comparing the infected *H. pylori* group (comprising the subgroups *s2/m2*, *s1/m2⁺c⁺b*, *s1/m1⁺c⁺b*, *s2m1⁺c⁺b*, *s1/m2*, *s1/m1⁺c⁺b*, etc...)) and the noninfected *H. pylori* group in antral biopsies. The distance corresponds to a Manhattan distance. Clustering was performed using a Ward method of agglomeration. Color varies from white, for the lowest levels of expression, to black for the highest levels of expression.

Immune response and inflammation Transcripts encoding immune response proteins were also associated to *H. pylori* infection. *LTB*, a membrane protein of the TNF family was upregulated. A large proportion of upregulated transcripts encoded chemokines or their receptors, as follows: (i) the chemokine (C-C motif) ligands *CCL-3* (*MIP1 α*), *CCL-4* (*MIP1 β*), *CCL-18* (*MIP4*), *CCL-19* (*MIP3 β*), *CCL-20* (*MIP3 α*); (ii) the chemokine (C-X-C motif) ligands *CXCL-2* (*MIP 2 α* or *Gro β*), *CXCL-9*, *CXCL-13*; and (iii) the chemokine (C-C motif) receptor *CCR-7* (*MIP3 β* receptor). RT-PCR experiments revealed an upregulation of *IL-8* mRNA (data not shown). Moreover, transcripts encoding *FCGR3A* (Fc fragment of IgG, low affinity IIIa, receptor for CD16), *HLA-DMA*, *HLA-DPB1*, *HLA-DQB1*, *HLA-DRB4*, and *HLA-DRB5* were upregulated. These proteins belong to the HLA class II, MHC2TA non-DNA-binding transactivator that functions both in constitutive and inducible MHC Class II expression. Transcripts encoding proteins implicated in the complement cascade, like *BF* (a component of the alternative pathway of complement activation), *C4A* (HLA class III), *C2*, *C3*, and *IF* were also overexpressed. The gene encoding the heat shock protein *HSPH1* was downregulated. *CRP*, an inflammation marker, and *SAA1*, known to be induced by LPS via *TLR4* in hepatocytes and involved in liver inflammation and injury, were upregulated. Upregulation of *RANKL* and *IFN γ* was demonstrated by RT-PCR experiments (data not shown), consistent

Table 1 Differential expression in antral biopsies between *H. pylori* positive (HP+) patients and *H. pylori* negative (HP-) patients revealed by *t*-test analysis on mean-centered reduced normalized intensities

Symbol	Locus link	Chromosome	Ratios	s.d.	NB	PMNL	Virulence
<i>Receptor/co-receptor</i>							
ITGB2	3689	21q22.3	1.962	0.091	+	+	+
LY96 ^a	23 643	8q21.11	1.213	0.013	+	+	+
TLR10	81 793	4p14	1.053	0.168	+	+	+
TLR2	7097	4q31.3	1.437	0.213	+	+	+
TLR3	7098	4q35.1	0.67	0.246	-	-	-
VCAM1	7412	1p21.2	2.711	0.329	+	+	+
<i>Signal transduction</i>							
FARP2	9855	2q37.3	0.659	0.156	-	-	-
HCK	3055	20q11.21	1.428	0.066	-	-	-
MAPK8 ^a	5599	10q11.22	1.152	0.07	+	+	+
RAC2	5880	22q13.1	1.692	0.065	+	+	+
SLA	6503	8q24.22	1.914	0.136	+	+	+
SOCS2	8835	12q22	0.763	0.058	-	-	-
STAT6 ^a	6778	12q13.3	0.387	0.094	-	-	-
<i>Immune response</i>							
BF	629	6p21.32	3.014	0.621	+	+	+
C2	717	6p21.32	1.247	0.011	+	+	+
C3	718	19p13.3	1.274	0.071	+	+	+
C4A	720	6p21.32	1.185	0.038	+	+	+
CCL18 ^a	6362	17q12	446.019	598.108	+	+	+
CCL19	6363	9p13.3	3.023	0.272	+	+	+
CCL20	6364	2q36.3	2.569	0.079	+	+	+
CCL3	6348	17q12	1.326	0.309	+	+	+
CCL4	6351	17q12	1.972	0.154	+	+	+
CRP	1401	1q23.2	2.025	0.089	+	+	+
CXCL13 ^a	10 563	4q21.1	33.328	38.498	+	+	+
CXCL2	2920	4q13.3	3.491	0.612	+	+	+
CXCL9	4283	4q21.1	2.844	0.579	+	+	+
FCGR3A	2214	1q23.3	2.414	0.246	+	+	+
HLA-DMA	3108	6p21.32	4.361	0.416	+	+	+
HLA-DPB1	3115	6p21.32	25.234	13.111	+	+	+
HLA-DQB1	3119	6p21.32	2.295	0.262	+	+	+
HLA-DRB4	3125	6p21.32	2.364	0.605	+	+	+
HLA-DRB5	3125	6p21.32	1.539	0.696	+	+	+
HSPH1	10 808	13q12.3	0.845	0.032	-	-	-
IF	3426	4q25	1.708	0.082	+	+	+
LTB	4050	6p21.33	2.362	0.865	+	+	+
MHC2TA	4261	16p13.13	1.301	0.129	+	+	+
PLAT	5327	8p11.21	0.583	0.105	-	-	-
IFITM1 ^a	8519	11p15.5	1.178	0.034	+	+	+
IRF4 ^a	3662	6p25.3	1.45	0.192	+	+	+
<i>Protein degradation</i>							
ADAMDEC1	27 299	8p21.2	3.166	0.583	+	+	+
GZMA	3001	5q11.2	1.313	0.186	+	+	+
GZMK	3003	5q11.2	4.161	1.006	+	+	+
MMP12	4321	11q22.2	3.631	1.35	+	+	+
MMP16	4325	8q21.3	1.516	0.275	+	+	+
MMP7 ^a	4316	11q22.2	4.942	1.173	+	+	+
<i>Proliferation/apoptosis</i>							
PAP	5068	2p12	2.814	1.537	+	+	+
TNFAIP2	7127	14q32.32	2.646	0.459	+	+	+
WFDC2 ^a	10 406	20q13.12	1.442	0.424	+	+	+
UBD	10 537	6p21.3	195.162	4.782	+	+	+
<i>Other</i>							
CPD	1362	17q11.2	0.34	0.071	-	-	-
TJP1	7082	15q13.1	0.817	0.075	-	-	-
EVI2B	2124	17q11.2	1.308	0.155	+	+	+
LST1	7940	6p21.33	1.336	0.168	+	+	+
ATP6V1A ^a	523	3q13.2	0.741	0.082	-	-	-
GATA6 ^a	2627	18q11.1	0.039	0.011	-	-	-

Rows—genes are clustered in six distinct functional groups—receptor/co-receptor, signal transduction, immune response, protein degradation, proliferation/apoptosis, and miscellaneous. Columns—symbol: gene symbol; locus link: locuslink numbers; chromosome: chromosomal location of gene; ratios: ratios of normalized expression of HP (+)/HP (-) patients; s.d.: standard deviation between spot replicates of normalized expression of HP (+)/HP (-) patients; NB: gene expression positively correlated (+) or inversely correlated (-) to the number of bacteria; PMNL: gene expression positively correlated (+) or inversely correlated (-) to the number of PMNL; virulence: gene expression positively correlated (+) or inversely correlated (-) to the strain virulence.

^aTranscriptional regulation confirmed by RT-PCR.

with the upregulation of the IFN-regulated transcript IFITM1 (IFN-induced transmembrane protein 1) and the IFN-regulatory factor 4 (IRF4) (for which upregulation was also confirmed by RT-PCR experiments).

Proteolysis *GZMA* (granzyme 1, cytotoxic T-lymphocyte-associated serine esterase 3), and *GZMK* (serine protease, granzyme 3, tryptase II), transcripts encoding matrix metalloproteinases: *MMP7* (confirmed by RT-PCR), *MMP12*, and *MMP16*, as well as *ADAMDEC1* and *ADAMTS6*, two members of the disintegrin and metalloproteinase families, were more abundantly expressed in *H. pylori*-infected patients than in noninfected patients. Conversely, we observed a parallel upregulation of the plasmin inhibitor *SERPIN F2* ($\alpha 2$ anti-plasmin) and downregulation of the tissue-type plasmin activator *PLAT*.

Cell proliferation The pancreatitis-associated protein PAP has been observed during pancreatic inflammation and liver carcinogenesis. WFDC2 is a biomarker for ovarian carcinoma and TNFAIP2 may correspond to a retinoic acid target gene in acute promyelocytic leukaemia. Transcripts encoding all three of these proteins were upregulated during infection. Ubiquitin D (*UBD*), also named *FAT10*, was also found to be upregulated in both microarray and RT-PCR experiments.

Correlation of gene expression levels, bacterial density, degree of PMNL infiltration, and bacterial virulence in *H. pylori*-infected antrum For our analyses, the different levels of *H. pylori* infection were classified into three groups: noninfected mucosa ($n=15$); infected mucosa with a bacterial density score of 1 (poorly infected patients) ($n=11$); infected mucosa with a bacterial density score from 2 to 4 (highly infected patients) ($n=16$). The mean normalized intensities of gene expression in each of the three groups were then determined. Genes for which expression levels were regulated in parallel with increasing bacterial density were designated as '+' (upregulated) or '-' (downregulated) (Table 1).

Variation of 88% of the transcripts upregulated in antral biopsies followed the variation of the bacterial density score (ie, the stimulation factor increasing from not infected to score 1, then from score 1 to score 2–4). The same, but opposite, observation was made for 75% of the downregulated transcripts.

A similar classification was performed using PMNL scores: non infected ($n=15$); with a PMNL score equal to 1 ($n=7$); and with a PMNL score from 2 to 4 ($n=20$). Fifty-one percent of the transcripts differentially expressed in antral biopsies from infected and noninfected patients exhibited a regulation that depended on the PMNL score.

The dependence between a selective host transcriptional response and the presence of virulence factors was then analyzed. Most strains were harboring either the 3 *babA2*, *vacA*, and *cagA* gene

factors (22%), or none of them (33%). Moreover, 87% of the strains harboring *cagA* were also harboring *babA2*, and 100% strains that did not harbor *cagA* did not harbor *babA2*. Only 2 *cagA*⁺ strains were found to be negative for *babA2*. Three groups of patients were considered according to the virulence status of the infecting bacteria: group 1 corresponded to non infected patients; group 2 corresponded to patients infected with *vacAs2m2/cagA*⁻/*babA2*⁻ strains (mild virulence); and group 3 corresponded to patients infected with *vacAs1m1/cagA*⁺/*babA2*⁺ strains (strong virulence). Transcripts regulated in parallel with an increased virulence were designated as '+' (gradually upregulated: noninfected <*vacAs2m2/cagA*⁻/*babA2*⁻-infected patients <*vacAs1m1/cagA*⁺/*babA2*⁺-infected patients) or '-' (downregulated: noninfected >*vacAs2m2/cagA*⁻/*babA2*-infected patients >*vacAs1m1/cagA*⁺/*babA2*⁺-infected patients). As indicated in Table 1, 63% of the selected transcripts were regulated in conjunction with bacterial virulence.

Microarray prediction analysis The Prediction Analysis of Microarrays (PAM) algorithm is based on an enhancement of the simple nearest prototype (centroid) classifier.²¹ We have used this approach developed by Tibshirani and co-workers to identify relevant classifier genes, which would be capable of predicting the infection state of patients, according to the different virulence factors of *H. pylori*. We were able to distinguish patients infected with strains harboring *vacAs2m2* and *vacAs1m1*, patients infected with strains *cagA*⁻ and *cagA*⁺ and patients infected with strains *babA2*⁻ and *babA2*⁺ from noninfected patients (Figure 3).

Chromosome localization Upregulated genes coding *BF*, *C2*, *C4a*, *HLA-DMA*, *HLA-DPB1*, *HLA-DQB1*, *HLA-DRB4*, *HLA-DRB5*, *LST1*, *LTB*, and *UBD* were all found in the 6p21.3 chromosomal position. In the same way, the upregulated genes coding *CCL18*, *CCL3*, *CCL4*, *EVI2B* and the downregulated gene coding *CPD* were all found in the 17q 11–13 region (Table 1).

Gastric fundus biopsies

Thirty-eight percent of the genes displaying statistically significant regulation by *H. pylori* in the fundus (Figure 2; Table 2) are differentially regulated in the antral biopsies.

Receptors, co-receptors, and extracellular proteins In contrast to the results obtained in the analysis of antral biopsies, expression of the *TLR* genes was not significantly regulated in fundic mucosa. Most of the genes encoding extracellular proteins that were identified in this analysis were specifically downregulated in the fundic mucosa, including *LAMB1* (laminin $\beta 1$), *DST* (dystonin: a member of the plakin protein family of adhesion junction

Table 2 Differential gene expression in fundic biopsies between *H. pylori* positive (HP+) patients and *H. pylori* negative (HP-) patients revealed by *t*-test analysis on mean-centered reduced normalized intensities

Symbol	Locus link	Chromosome	Ratios	s.d.	NB	PMNL	Virulence
<i>Receptors/co-receptors</i>							
LAMB1	3912	7q31.1	0.892	0.004	—	—	—
DST	667	6p12.1	0.745	0.031	—	—	—
ITGB1	3688	10p11.22	0.761	0.021	—	—	—
<i>Immune response</i>							
CCL20	6364	2q36.3	1.486	0.046	+	+	+
CCL3	6348	17q12	1.134	0.115	+	+	+
CD74	972	5q32	3.688	0.88	+	+	+
CXCL2	2920	4q13.3	1.578	0.102	+	+	+
HLA-DPA1	3113	6p21.32	15.568	5.208	+	+	+
HLA-DPB1	3115	6p21.32	3.779	0.755	+	+	+
HLA-DRA	3122	6p21.32	424	1087	+	+	+
HLA-DRB4	3125	6p21.32	5.897	0.837	+	+	+
IF	3426	4q25	1.188	0.028	+	+	+
HSPH1	10 808	13q12.3	0.85	0.016	+	+	+
<i>Protein degradation and peptidases</i>							
MMP12	4321	11q22.2	1.204	0.078	+	+	+
UBD ^a	10 537	6p21.3	13.591	1.426	+	+	+
CPD	1362	17q11.2	0.483	0.015	—	—	+
<i>Proliferation and signal transduction</i>							
GATA6 ^a	2627	18q11.2	0.348	0.024	—	—	—
RBL2	5934	16q12.2	0.784	0.039	—	—	—
RALGPS1A	9649	9q33.3	0.655	0.052	—	—	—
PSD3	23 362	8p22	0.819	0.068	—	—	—
ARHGEF12	23 365	11q23.3	0.667	0.013	—	—	—
<i>Other</i>							
ATP6V1A ^a	523	3q13.31	0.788	0.043	—	—	—
CAT	847	11p13	0.743	0.06	—	—	—

Rows—genes are clustered in five distinct functional groups: receptor/co-receptor, immune response, protein degradation and peptidases, proliferation and signal transduction, and miscellaneous. Columns—symbol: gene symbol; locus link: locuslink numbers; chromosome: chromosomal location of gene; ratios: ratios of normalized expression of HP (+)/HP (-) patients; s.d.: standard deviation between spot replicates of normalized expression of HP (+)/HP (-) patients; NB: gene expression positively correlated (+) or inversely correlated (-) to the number of bacteria; PMNL: gene expression positively correlated (+) or inversely correlated (-) to the number of PMNL; virulence: gene expression positively correlated (+) or inversely correlated (-) to the strain virulence.

^aTranscriptional regulation confirmed by RT-PCR.

plaque proteins), and *ITGB1* (integrin β 1), receptor for extracellular matrix proteins (fibronectin, collagen, laminin).

Immune response The regulation of transcripts encoding genes related to the immune response represented a large fraction of the transcriptional response to *H. pylori* infection in the fundus. Three chemokine genes were upregulated: *CCL3*, *CCL20*, and *CXCL2*. Several genes found to be regulated in the antrum, including *HLA* genes encoding *HLA-DPB1* and *HLA-DRB4*, were similarly upregulated in the fundus. In contrast, some others like *HLA-DPA1* and *HLA-DRA* were upregulated only in the fundus. The *CD74* antigen (invariant polypeptide of the MFC class II antigen-associated) was specifically upregulated in the fundus. In contrast to the antrum, no transcripts encoding proteins of the complement were regulated in the fundus; however, the *IF* (I factor) was upregulated. Finally, expression of *HSPH1* was lower in *H. pylori*-infected fundic mucosa than in noninfected fundic mucosa.

Proteolysis Similar to data obtained in the antrum, *MMP12* (macrophage elastase) and *CPD* (metallo-carboxypeptidase) expression was decreased in the fundus, whereas *UBD* was upregulated (Figure 2; Table 2). sqRT-PCR experiments revealed a slight upregulation of *MMP-7* in the fundus, in contrast to the marked upregulation found in the antrum (data not shown).

Cell proliferation and signal transduction and other Concerning the regulation of genes related to cell proliferation and signal transduction in the fundus, we observed the downregulation of a similar set of genes that were downregulated in the antrum. This was notably the case for the transcription factors *GATA6* (Figure 2; Table 2) and *RBL2*. Some of these downregulated transcripts encode guanine nucleotide exchange factors such as *RALGPS1A*, *PSD3*, and *ARHGEF12*.

Correlation of gene expression levels, bacterial density, degree of PMN infiltrates, and bacterial

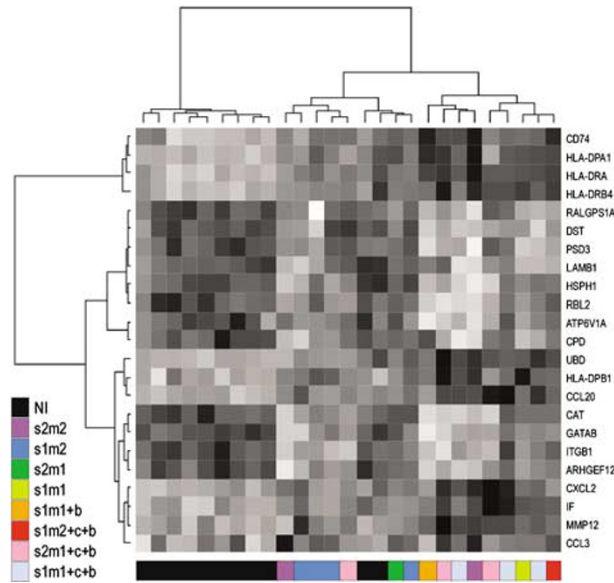


Figure 2 Heatmap comparing the infected *H. pylori* group (comprising the subgroups *s2/m2*, *s1/m2+c+b*, *s1/m1+c+b*, *s2/m1+c+b*, *s1/m2*, *s1/m1+c+b*, etc...) and the noninfected *H. pylori* group in fundic biopsies. The distance corresponds to a Manhattan distance. Clustering was performed using a Ward method of agglomeration. Color varies from white, for the lowest levels of expression, to black for the highest levels of expression.

virulence in H. pylori-infected fundus Assessment of the results with respect to bacterial density, PMNL infiltration and virulence scores was performed using the same parameters as in the analysis of antral biopsies. Results indicated that 55, 53, and 23% of significantly modulated genes were regulated in parallel to the bacterial density, PMNL infiltration, and virulence scores, respectively (Table 2).

Assessment of tissue specificity of gene expression To further define the major traits of the response to *H. pylori* infection, we searched the list of transcripts identified in the analysis of antral biopsies (Table 1) for ESTs that matched each oligonucleotide probe. The number of ESTs matching a specific probe defines the abundance of the corresponding transcript.¹⁵ Fifty-one distinct probes identified in the antrum were associated with at least three ESTs (Figure 4b). These ESTs were then further characterized with respect to cDNA libraries from which they originated and classed into five relevant categories as follows: (1) stomach; (2) immune cells; (3) stem cells; (4) fibroblasts; and (5) cancer cells. The percentage of ESTs associated with each of these categories, and therefore, preferentially expressed in the different cell types, is shown in Figure 4a. Thus, certain transcripts are more specifically expressed in stomach (such as *CPD* or *UBD*), in immune cells (such as *EVI2*, *LTB*, or *TLR10*), in cancer cells (such as *CXCL13*, *LY96*, *MAPK8*), or in stem cells (*IFITM1*, *MMP16*). *GATA6*, *IFITM1*, *MMP16*, *SOCS2*, *TPJ1*,

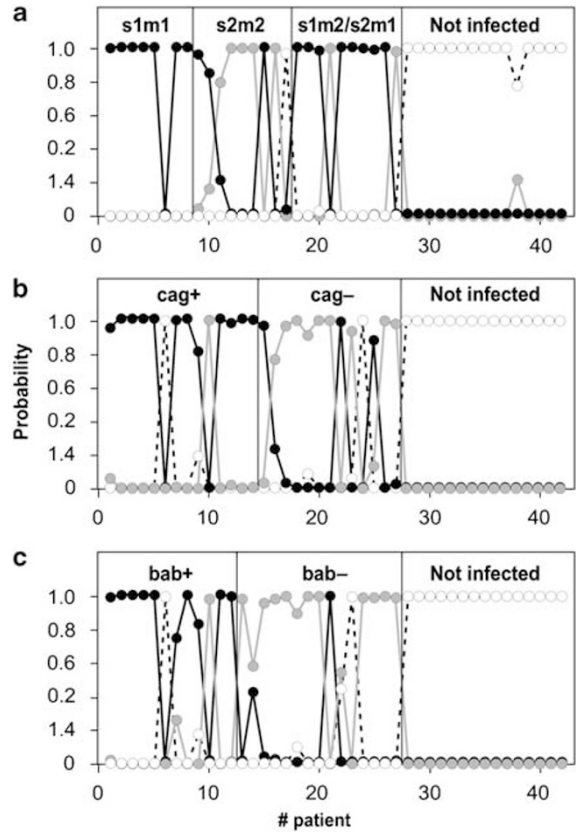


Figure 3 PAM analysis. Estimated probabilities for the Vac (a), cag (b), and babA (c) genotypes in the antrum. Samples are partitioned by their true class and the predicted class corresponds to the points (● for *s1/m1*, *cag+*, or *babA+*; ● for *s2/m2*, *cag-*, or *babA-*; ○ not infected). Samples were classified in different categories according their maximal score. A correct prediction was obtained in the case of *s1/m1*, *cag+*, *babA+*, and not infected. The prediction was less efficient in the case of the other genotypes.

WFDC2 had previously been identified in stem cells or embryonic tissues. *ATP6V1A* falls in the same category, but is also expressed in the stomach. Interestingly, *ATP6V1A*, *GATA6*, *IFITM1*, *SOCS2*, and *TPJ1* have also been found in tissues of tumoral origin. *IFITM1*, upregulated after *H. pylori* infection and present in tumoral tissues, is a good candidate gene associated with pathogenic mechanisms that enhance the susceptibility of gastric epithelial cells to carcinogenic conversion following infection by *H. pylori*.

Immunohistochemistry Results on TMAs

To validate potential new diagnostic markers for tissues infected with *H. pylori*, we selected eight gene products identified in our gene expression studies for immunohistochemical profiling in TMAs. Selection of these genes (*STAT6*, *MMP-7*, *UBD*, *CXCL13*, *LY96*, *MAPK8*, *WFDC2*, and *GATA6*) was based on either our DNA microarray experi-

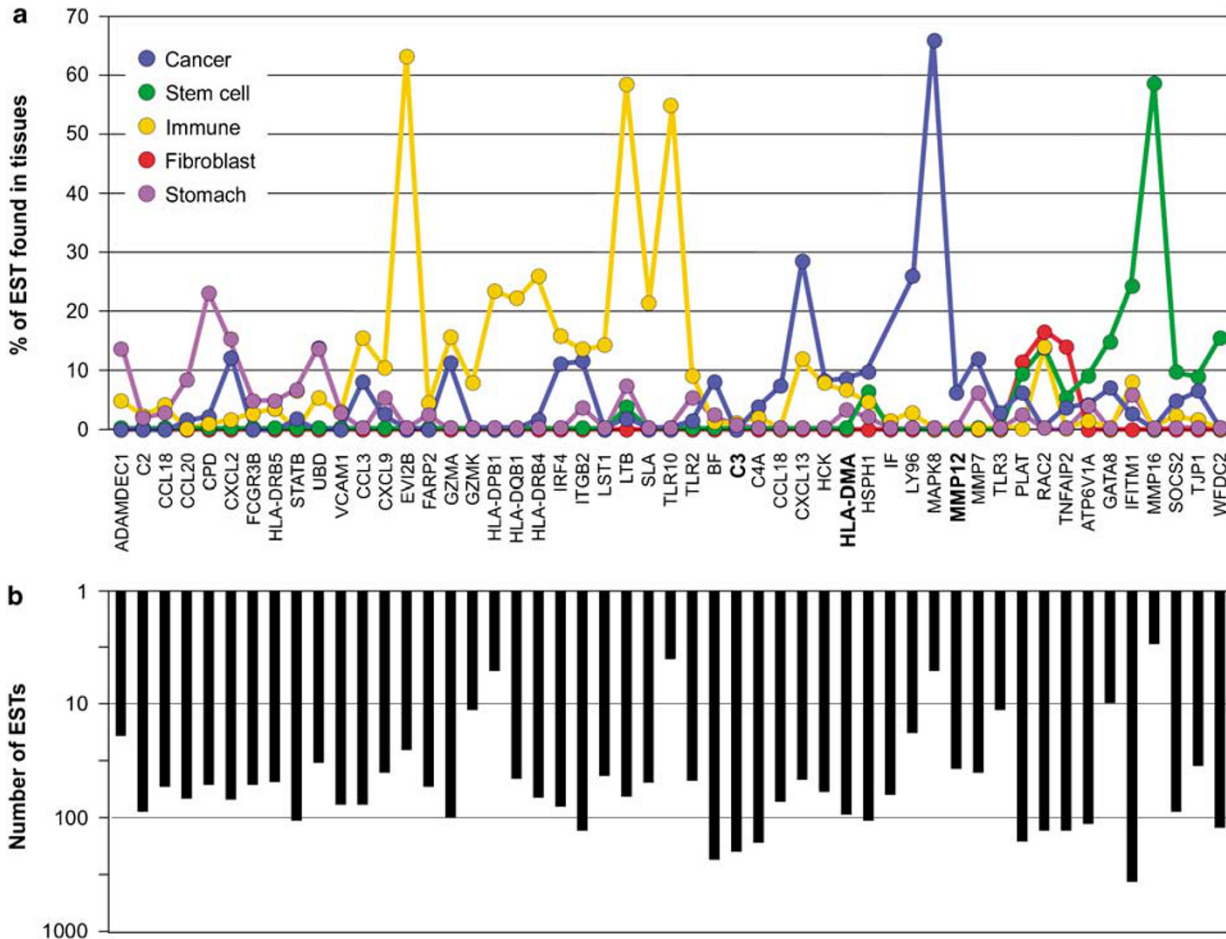


Figure 4 (a) Number of matching ESTs per probe, for each of the 51 probes differentially expressed in infected samples. The percentage of all ESTs identified in libraries constructed from cancer tissues, stem cells, immune cells, fibroblasts, or stomach tissue are indicated. This representation allows an easy identification of cell-specific probes. (b) Total number of ESTs per probe, for each of the 51 probes differentially expressed in infected samples. This value provides an estimation of the abundance of each transcript.

ments or on EST frequency estimates of transcript abundance, presented in Figure 4a. Transcripts encoding these proteins were either upregulated (*MMP-7*, *UBD*, *CXCL13*, *LY96*, *MAPK8*, *WFDC2*) or downregulated (*STAT6*, *GATA6*) in biopsies from patients infected with *cagA*⁺/*vacAs1m1/babA2*⁺ *H. pylori* strains. Samples were classified according to the assignment of patients to group 1 (noninfected), group 2 (infected by *vacAs2m2/cagA*⁻/*babA2*⁻ strains) or 3 (infected by *vacAs1m1/cagA*⁺/*babA2*⁺ strains).

Evaluation of positive cells was performed only in well-representative spots. As shown in Figure 5Aa, *GATA6* was mainly detected in the nuclei of epithelial cells and only weakly present in mesenchymal cells. *GATA6* staining was the strongest in group 1 (Figure 5Aa), less intense in group 2 (Figure 5Ab), and weakest in group 3 (Figure 5Ac). Staining of *STAT6* was robust and predominantly cytoplasmic in epithelial cells from antral biopsies of group 1 patients (Figure 5Aa). Weaker staining was observed in antral biopsies from groups 2 (Figure 5Ab) and 3 (Figure 5Ac). *MMP7* was largely

detected in epithelial cells, with expression being weakest in group 1 patients (Figure 5Aa), intermediate in group 2 (Figure 5Ab), and highest in group 3 (Figure 5Ac). *CXCL13* was mainly present at the periphery of the isolated lymphoid follicles (ILFs). *CXCL13* was also detected in small lymphoid aggregates without ILF organization. *CXCL13* expression was very weak or absent in group 1 patients (Figure 5Aa), more intense in group 2 (Figure 5Ab), and highest in group 3 (Figure 5Ac). *LY96* staining was mainly observed in biopsies from group 3 patients (Figure 5Ba) with some staining in samples from group 2 (Figure 5Bb) and only weakly expressed or absent in group 1 (Figure 5Bc). *LY96* was expressed in two different mucosal cell types. Most of the staining localized to the epithelial cells in the gastric crypts, whereas only a few of the mesenchymal cells within the lamina propria were positive for *LY96*. *WFDC2* expression was strong in the gastric epithelium from group 3 patients (Figure 5Bc). Occasionally, staining was also observed in the epithelium of the gastric biopsies of group 2 patients (Figure 5Bb), but only weak expression was detec-

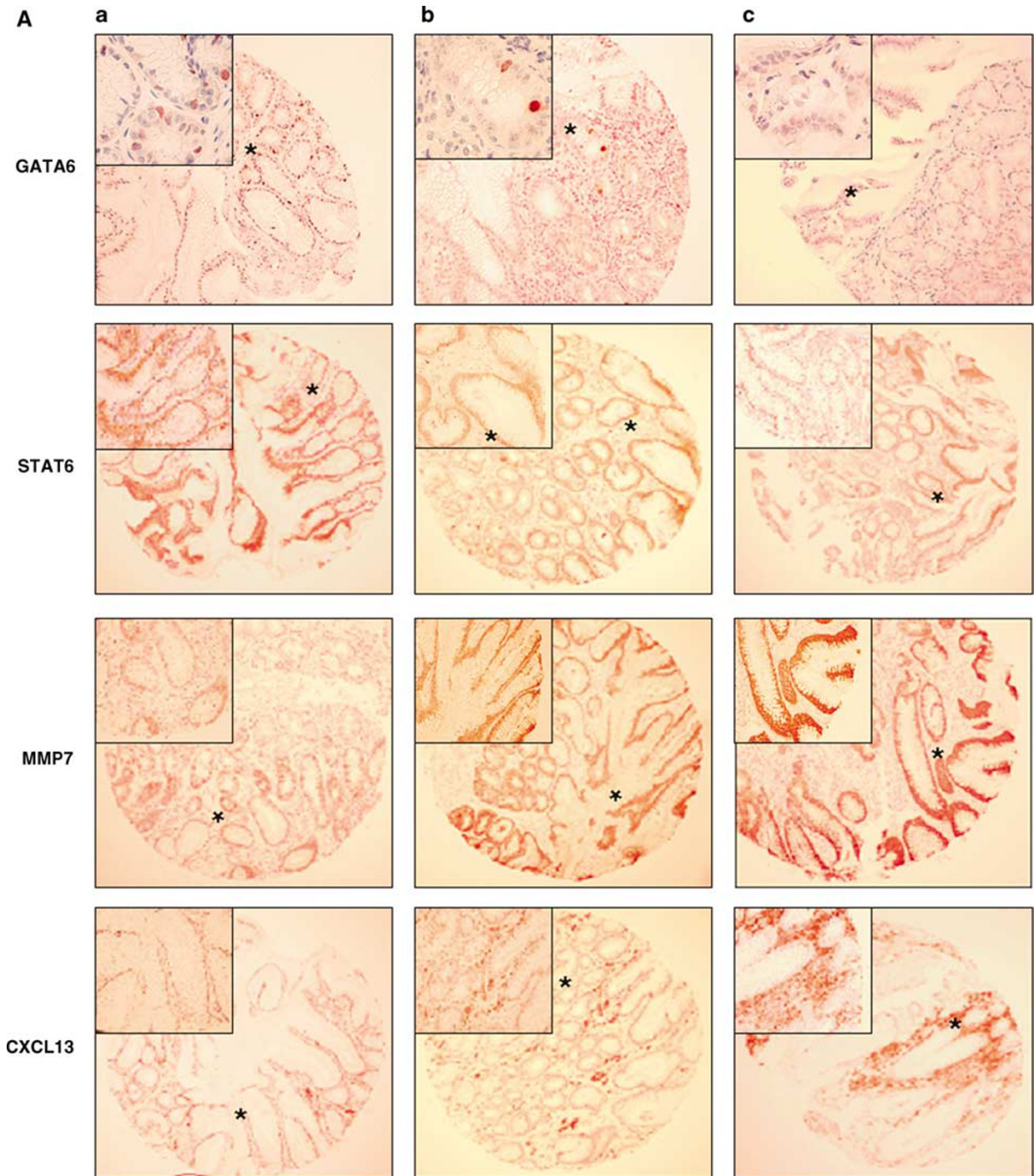


Figure 5 (A) immunohistochemical staining of TMAs from antral biopsies of group 1 (control subjects) (panel a), group 2 (patients infected with *cagA* (-), *vacA* *s2/m2* (+), *babA2* (-) *H. pylori*) (panel b), and group 3 (patients infected with *cagA* (+), *vacAs1/m1* (+), *babA2* (+) *H. pylori*) (panel c). Each spot shows representative immunostaining of GATA6, STAT6, MMP7, and CXCL13 in each tissue, at low original magnification ($\times 100$) and at high original magnification ($\times 400$; insets). In the epithelial component, staining for GATA6 was nuclear, whereas staining of STAT6 and MMP7 was cytoplasmic. CXCL13 expression was exclusively localized in lymphoid cell aggregates. (B) immunohistochemical staining of TMAs from antral biopsies of group 1 (control subjects) (panel a), group 2 (patients infected with *cagA* (-), *vacA* *s2/m2* (+), *babA2* (-) *H. pylori*) (panel b), and group 3 (patients infected with *cagA* (+), *vacAs1/m1* (+), *babA2* (+) *H. pylori*) (panel c). Each spot shows representative immunostaining for LY96, WFDC2, MAPK8, UBD, at low original magnification ($\times 100$) and at high original magnification ($\times 400$; insets).

ted in the gastric epithelium from group 1 patients (Figure 5Ba). MAPK8 was mainly expressed in biopsies of groups 2 (Figure 5Bb) and 3 (Figure

5Bc) patients, with the most intense staining of the kinase observed in the gastric epithelium. The epithelium from group 1 patients stained only

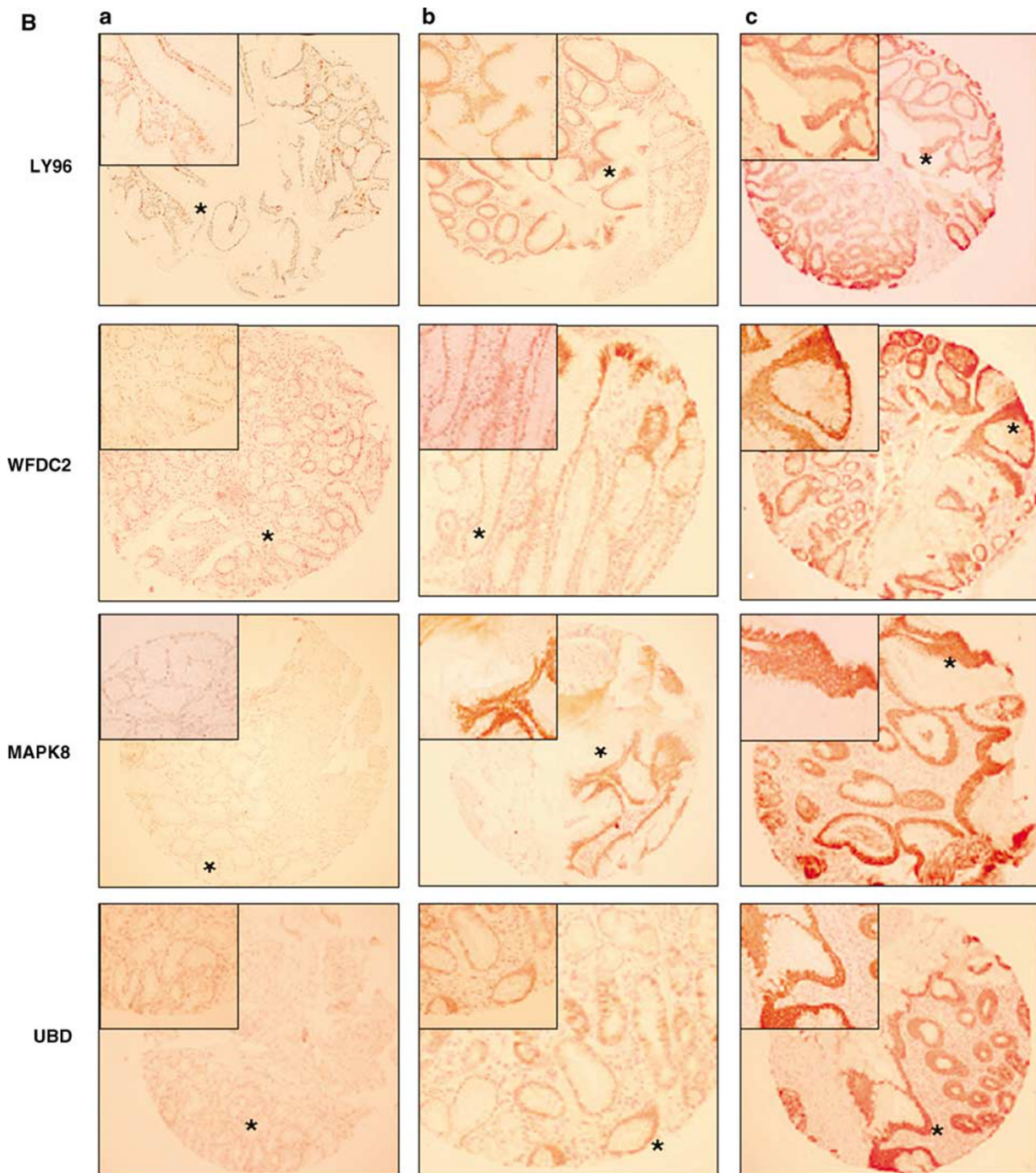


Figure 5 Continued.

weakly (Figure 5Ba). UBD expression exhibited a diffuse cytoplasmic pattern of epithelial staining in group 3 biopsies (Figure 5Bc) and less in those from group 2 (Figure 5Bb). UBD expression was weak or absent in the gastric epithelium of group 1 biopsies (Figure 5Ba).

Evaluation of the density of positive cells was then performed in representative spots from each group. Quantification showed that the density of positive cells stained with GATA6 and STAT6 was

lower in group 3 than in 2 and 1, whereas the density of positive cells stained with MMP7 and CXCL13 was significantly higher in group 3 than in 2 or 1 (Figure 6). The density of positive cells was low in group 1 for LY96, WFDC2, MAPK8, UBD, MMP7, and CXCL13 (Figure 6). The density of positive cells for LY96, WFDC2, MAPK8, and UBD was significantly higher in group 3 than in 2, and largely above the density observed for group 1 (Figure 6).

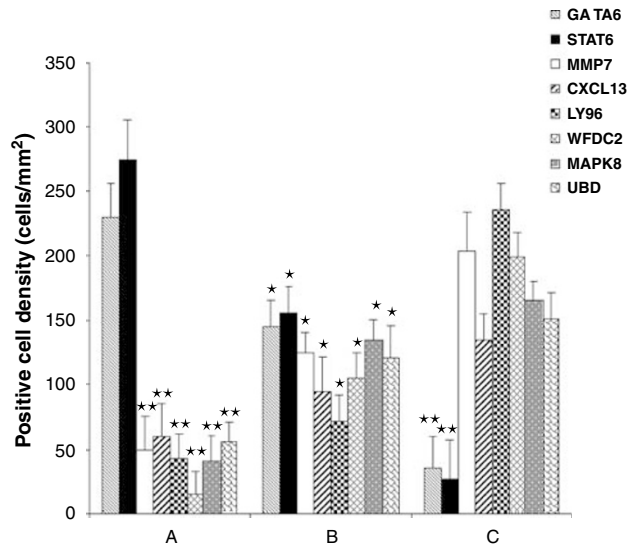


Figure 6 Positive cell density in control subjects (a), in patients infected with *cagA* (-), *vacA s2/m2* (+), *babA2* (-) *H. pylori* (b), and in patients infected with *cagA* (+), *vacAs1/m1* (+), *babA2* (+) *H. pylori* (c) for GATA6, STAT6, MMP7, CXCL13, LY96, WFDC2, MAPK8, and UBD, * $P < 0.05$; ** $P < 0.01$.

Discussion

Most of the previous microarray studies focusing on the host-cell response to *H. pylori* infection have been carried out on gastric epithelial cell lines such as AGS²⁶ and Kato III,²⁷ or in animal models.⁸ Very few analyses have been performed in human tissue infected by *H. pylori*^{9,27} and reported studies have included only few distinct tissue samples. The heterogeneity of tissue origin and nature of the samples (ie, whole gastric mucosa or biopsies), as well as uncertainties about patients' status (chronic active gastritis with or without metaplasia or atrophy), underscore the need for more standardized studies.^{9,27} Our present study was carried out on both antral and fundic biopsies from a large series of patients with 'homogeneous' epidemiological, clinical, and histological data.

In addition to the careful selection of biopsies, we also characterized the potential impact of the three major *H. pylori* virulence factors *cagA*, *babA2*, and *vacA* on both immune cell infiltration and expression profiling. We show here that strains expressing the *cagA*, *babA2*, and *vacA* virulence factors were associated with an increased number of bacteria and increased infiltration of PMNL into mucosa. It is likely that CD4⁺ and CD8⁺ T cells are also present at the site of infection, as described by Bamford *et al.*²⁸ The global data obtained by microarray profiling corresponds to a complex transcriptional response originating from several cell types. Even though the distinct contribution of different cells to the global transcriptional response may not be well established, at least some genes may be assigned to specific cell types. The approach that we developed based on the Mediante annotations (Figure 4)

provides an initial answer to this question. Hence, of the 51 selected transcripts, 17 are more highly expressed in tissues of immune origin, whereas 10 are more highly expressed in the stomach. Results from Figure 4 reveal a major contribution of immune cells to the observed response in addition to an alteration in resident cells. Interestingly, the transcriptional responses of many genes to infection could be correlated to bacterial density, immune cell infiltration, and bacterial virulence (Tables 1 and 2).

Comparison of the transcriptional patterns observed in the antrum or in the fundus revealed two distinct responses to *H. pylori* infection. First, the transcriptional response to infection (as assessed by the number of altered genes) was more pronounced in the antrum than in the fundus. Second, a good correlation between transcriptional changes and *H. pylori* virulence factors, number of bacteria and PMNL density, was clearly found in the antrum, whereas this correlation was less pronounced in fundic mucosa. As the two sites differ in their responses to infection, the two transcriptional responses were analyzed separately. In antral biopsies from infected patients, we found that genes involved in bacterial recognition and signalling events were upregulated in comparison with non-infected biopsies. These genes include *TLR2* and *LY96*. *TLR2* is known to be upregulated in response to peptidoglycan, lipoteichoic acid,²⁹ and bacterial lipoproteins³⁰ or LPS. *TLR4* is activated by the LPS of Gram-negative bacteria³¹ and *LY96* is a protein associated with this receptor. The expression of these two TLR after *H. pylori* infection has been well documented in the literature. Smith *et al.*³² demonstrated that *H. pylori* LPS was an agonist of the *TLR2* and that LPS binding to this receptor, but not to *TLR4*, was responsible for NF- κ B activation. However, the lack of the *TLR4* response was due in part to the lack of the *TLR4* accessory molecule *LY96* in MKN45 cells.³² Moreover, Su *et al.*³³ indicated that *H. pylori* activates *TLR4* expression in epithelial cells. Similarly, Ishihara *et al.*³⁴ showed that *LY96* expression was elevated in gastric epithelial cells during *H. pylori* infection and they proposed that the *TLR4/MD-2* system is an important receptor complex involved in the response to *H. pylori* LPS in the stomach. Interestingly, we found that many chemokines were upregulated in infected patients, mainly in the antrum. Among them, *CXCL2* (*MIP2 α*) was upregulated in infected patients, in concert with increased bacterial virulence. Bach *et al.*²⁶ indicated that an increase in *CXCL2* expression upon infection with *H. pylori* was not influenced by *cagA* expression in AGS cells. We also observed an upregulation in *CCL3* (*MIP-1 α*) in infected patients, as described previously by Yamaoka *et al.*³⁵

Two main processes were specifically initiated in infected antral biopsies. First, infection was associated with an upregulation of genes encoding complement proteins, namely *C2*, *C3*, *C4a*, and *factor B*, suggesting that *H. pylori* engages both the

classical and alternative activation pathways. In this regard, Berstad *et al*³⁶ suggested that complement activation by constituents released from the bacterial cell wall (such as LPS), could explain why *H. pylori* induces gastric injury without invading the mucosa. Second, we observed an upregulation of genes encoding type II MHC in the antrum of *H. pylori*-infected patients. It has previously been shown that MHC II-deficient mice are not protected against *H. pylori* infection.³⁷ In fact, close analysis of the CD4 T-cell response revealed that the specific Th2-polarized response mediates protective immunity, in contrast to the Th1 response induced by *H. pylori*, which is ineffective in combating infection.^{28,38} In addition, it has been suggested that the Th1 response may also mediate gastritis and gastric cancer.³⁹ Several studies have shown that T lymphocytes of Stat6-deficient mice fail to differentiate into Th2 cells in response to IL-4 or IL-13.^{40,41} RANKL, a member of the TNF cytokine family, is also specifically expressed in Th1-type T lymphocytes. Taken together, the downregulation of *Stat6* and the upregulation of *RANKL* observed in our experiments is in agreement with the lack of a Th2 response associated with a prominent Th1 activation. The Th1-mediated response is characterized by high levels of IFN γ and IL-12. IFN γ , one of the major cytokines produced by T cells recruited to the gastric mucosa during infectious and noninfectious gastritis, can increase both class II MHC expression and attachment of the bacteria to gastric epithelial cells, as well as the induction of apoptosis in gastric epithelial cells.⁴² Increased IFN γ expression reported in the present study is supported by an upregulation of molecules such as the *IFITM1*. Indeed, IFN γ upregulation may be related to an increased risk of developing cancers, as suggested by the upregulation of two other genes *FAT10* and *IRF-4*. The expression of *FAT10* (or *UBD*) is specifically increased in infected antral biopsies and has previously been shown to be synergistically induced by IFN γ and TNF α .⁴³ Elevated *FAT10* expression in mature B cells and dendritic cells has also suggested a potential role of *FAT10* in antigen presentation.⁴⁴ *IRF4* is a lymphoid-specific member of the IRF family of transcription factors. It was also shown to be upregulated in antrum infected with *H. pylori*. Genetic studies have demonstrated that *IRF-4* is a critical effector of mature lymphocyte function.⁴⁵ Imaizumi *et al*⁴⁶ have indicated that *IRF-4* is involved in the pathogenesis of T-cell leukemia through its positive effect on the cell cycle, and that *IRF-4* can be used as a molecular marker of clinical subtype in cell leukemia. An alternative explanation of altered expression of IFN γ -regulated genes is provided by the fact that *IFITM1* is notably associated with stem cells. Several observations support the existence of cancer 'stem cells' in several tumor types (as described for at least leukemia and breast tumors), suggesting that only a subset of cells is able to

maintain tumor growth.⁴⁷ This hypothesis is supported by the observation that overexpression of molecules associated with normal self-renewal induces tumor formation by stem cells derived from normal adult tissue.⁴⁸ The reactivation of pluripotency-associated genes has been also observed in various somatic cancer cells.⁴⁹ Finally, manipulation in normal stem cells of some self-renewal pathways, such as Oct4, can also lead to cancer.⁵⁰ However, it is still unclear whether such tumors arise from 'sick' stem cells or from nonstem (progenitor) cells, which acquire a stem cell-like phenotype during carcinogenesis. Although this hypothesis is for the moment purely based on the altered expression following *H. pylori* infection of a few transcripts that have previously been identified in stem cells, the involvement of cancer stem cells in tumor initiation would provide an avenue for early detection of samples prone to cancerization. From that perspective, not only *IFITM1*, but also *GATA6*, *MMP16*, *SOCS2*, *TPJ1*, or *WFDC2* appear to be attractive markers.

Infection in the antrum also elicited many genes of *ADAM* and *MMP* families. Among them, *MMP-7* appeared to be upregulated. Wroblewski *et al*⁵¹ have shown an increased *MMP7* expression in gastric epithelium in response to *H. pylori* through mechanisms that involve activation of Rho and Rac. Moreover, Crawford *et al*⁵² have shown that *H. pylori cagA*⁺ strains enhance levels of *MMP-7* within inflamed mucosa, by an ERK-dependent mechanism.

Other upregulated genes noted in the present work have been involved in signalling pathways, potentially leading to cellular events that regulate proliferation and cancer. Among them, *Rac-2* was found to be upregulated in infected antral biopsies. It has been shown that *Rac-2* is involved in FcR- and CR3-induced activation in PMNL.⁵³ In our study, *MAPK8 (JNK)* was upregulated in infected antral biopsies, and upregulation was increased when patients were infected with *vacAs1m1/cagA*⁺/*babA2*⁺ strains as compared to *vacAs2m2/cagA*⁻/*babA2*⁻ strains. Finally, we found the chromosomal locations where a large proportion of genes were modulated, that is, position 6p21 containing *HLA* genes. Interestingly, this region has been shown to be associated with gastric cancers.⁵⁴

In fundic biopsies, the global transcriptional response to infection appeared weaker than in the antrum, in terms of the number of regulated genes. However, a significant proportion of regulated genes were common between the antrum and the fundus. Thus, we found a downregulation of the *ATP6V1A* (ATPase, H⁺ transporting, lysosomal 70 kDa, V1 subunit A) and a downregulation of the transcription factor *GATA6*, for which a binding site is present in the promoter of several stomach-specific genes including the *H*⁺/*K*⁺-ATPase.⁵⁵ Upregulation of the MHC2 response was also elicited in both parts of the stomach, but regulated genes encoding MHC2

proteins were more numerous in antrum. The observed upregulation of *UBD* and downregulation of catalase have previously been associated with liver cancer.⁵⁶

The distinct gene responses observed between antral and fundic biopsies included an upregulation of genes encoding *TLR* or proteins related to *TLR* in the antrum, but not in the fundus (*t*-test $P=0.001$). This could be relevant to the model proposing that infection with *H. pylori* begins in the antrum and then spreads throughout the fundus. With regard to the complement pathway, *C2*, *C3*, and *C4A* expression levels were not statistically significant, but *IF* was also slightly upregulated.

In conclusion, the results obtained from this work extend previous results drawn largely from cell culture experiments and animal models, or from limited studies on cohorts of patients.^{2,9} Our work identifies for the first time a limited set of molecules (*GATA6*, *STAT6*, *MMP7*, *CXCL13*, *UBD*, *MAPK8*, *LY96*, *WFDC2*) that establish a risk associated with *H. pylori* infection. Our study performed on a large number of *H. pylori*-infected patients has provided new insights into the cellular host responses to *H. pylori* infection and suggests that a premalignant molecular signature can be detected in the most severely infected cases, before any histological evidence of malignancy.

Acknowledgements

We thank Ellen Van Obberghen-Schilling for editing this manuscript. We thank Michelle Batoz for her excellent technical assistance concerning the RT-PCR experiments and Frédéric Berto for his technical assistance concerning the PCR genotyping of *H. pylori*. This work was supported by a PHRC regional grant from CHU of Nice (2002), by the Institut National of Cancer (PL079) and by the Canceropole PACA. VH is a recipient of a fellowship from INSERM (poste d'accueil 2000). We acknowledge the excellent support of the Nice-Sophia Antipolis Transcriptome Platform of the Marseille-Nice Genopole in which the microarray experiments were carried out. Special thanks are due to Virginie Magnone and Géraldine Rios for microarray production, and to Franck Aguila for artwork.

References

- Blaser MJ, Berg DE. *Helicobacter pylori* genetic diversity and risk of human disease. *J Clin Invest* 2001;107:767–773.
- Mattapallil JJ, Dandekar S, Canfield DR, *et al*. A predominant Th1 type of immune response is induced early during acute *Helicobacter pylori* infection in rhesus macaques. *Gastroenterology* 2000;118:307–315.
- Furuta T, Shirai N, Takashima M, *et al*. Effect of genotypic differences in interleukin-1 beta on gastric

- acid secretion in Japanese patients infected with *Helicobacter pylori*. *Am J Med* 2002;112:141–143.
- Keates S, Hitti YS, Upton M, *et al*. *Helicobacter pylori* infection activates NF-kappa B in gastric epithelial cells. *Gastroenterology* 1997;113:1099–1109.
- Yamaoka Y, Kudo T, Lu H, *et al*. Role of interferon-stimulated responsive element-like element in interleukin-8 promoter in *Helicobacter pylori* infection. *Gastroenterology* 2004;126:1030–1043.
- Honda S, Fujioka T, Tokieda M, *et al*. Development of *Helicobacter pylori*-induced gastric carcinoma in Mongolian gerbils. *Cancer Res* 1998;58:4255–4259.
- Backert S, Gressmann H, Kwok T, *et al*. Gene expression and protein profiling of AGS gastric epithelial cells upon infection with *Helicobacter pylori*. *Proteomics* 2005;5:3902–3918.
- Huff JL, Hansen LM, Solnick JV. Gastric transcription profile of *Helicobacter pylori* infection in the rhesus macaque. *Infect Immun* 2004;72:5216–5226.
- Wen S, Felley CP, Bouzourene H, *et al*. Inflammatory gene profiles in gastric mucosa during *Helicobacter pylori* infection in humans. *J Immunol* 2004;172:2595–2606.
- Dixon MF, Genta RM, Yardley JH, *et al*. Classification and grading of gastritis. The updated Sydney System. International workshop on the histopathology of gastritis, Houston 1994. *Am J Surg Pathol* 1996;20:1161–1181.
- Rudi J, Rudy A, Maiwald M, *et al*. Direct determination of *Helicobacter pylori* *vacA* genotypes and *cagA* gene in gastric biopsies and relationship to gastrointestinal diseases. *Am J Gastroenterol* 1999;94:1525–1531.
- Poon SK, Chang CS, Su J, *et al*. Primary resistance to antibiotics and its clinical impact on the efficacy of *Helicobacter pylori* lansoprazole-based triple therapies. *Aliment Pharmacol Ther* 2002;16:291–296.
- Gerhard M, Lehn N, Neumayer N, *et al*. Clinical relevance of the *Helicobacter pylori* gene for blood-group antigen-binding adhesin. *Proc Natl Acad Sci USA* 1999;96:12778–12783.
- Van Gelder RN, von Zastrow ME, Yool A, *et al*. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 1990;87:1663–1667.
- Le Brigand K, Russell R, Moreilhon C, *et al*. An open-access long oligonucleotide microarray resource for analysis of the human and mouse transcriptomes. *Nucleic Acids Res* 2006;34:e87.
- Didier G, Brezellec P, Remy E, *et al*. GeneANOVA—gene expression analysis of variance. *Bioinformatics* 2002;18:490–491.
- Saeed AI, Sharov V, White J, *et al*. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* 2003;34:374–378.
- Kerr MK, Churchill GA. Experimental design for gene expression microarrays. *Biostatistics* 2001;2:183–201.
- Wettenhall JM, Smyth GK. LimmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics* 2004;20:3705–3706.
- Kooperberg C, Fazio TG, Delrow JJ, *et al*. Improved background correction for spotted DNA microarrays. *J Comput Biol* 2002;9:55–66.
- Tibshirani R, Hastie T, Narasimhan B, *et al*. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci USA* 2002;99:6567–6572.

- 22 Hosack DA, Dennis Jr G, Sherman BT, *et al*. Identifying biological themes within lists of genes with EASE. *Genome Biol* 2003;4:R70.
- 23 Pollard KS, van der Laan MJ. Statistical interference for simultaneous clustering of gene expression data. *Math Bio Sci* 2002;176:99–121.
- 24 Giltane JM, Rimm DL. Technology insight: Identification of biomarkers with tissue microarray technology. *Nat Clin Pract Oncol* 2004;1:104–111.
- 25 Ginestier C, Charaffe-Jauffret E, Bertucci F. *et al*. Distinct and complementary information provided by use of tissue and DNA microarrays in the study of breast tumor markers. *Am J Pathol* 2002;161:1223–1233.
- 26 Bach S, Makristathis A, Rotter M, *et al*. Gene expression profiling in AGS cells stimulated with *Helicobacter pylori* isogenic strains (cagA positive or cagA negative). *Infect Immun* 2002;70:988–992.
- 27 Cox JM, Clayton CL, Tomita T, *et al*. cDNA array analysis of cag pathogenicity island-associated *Helicobacter pylori* epithelial cell response genes. *Infect Immun* 2001;69:6970–6980.
- 28 Bamford KB, Fan X, Crowe SE, *et al*. Lymphocytes in the human gastric mucosa during *Helicobacter pylori* have a T helper cell 1 phenotype. *Gastroenterology* 1998;114:482–492.
- 29 Opitz B, Schroder NW, Spreitzer I, *et al*. Toll-like receptor-2 mediates *Treponema glycolipid* and lipoteichoic acid-induced NF-kappaB translocation. *J Biol Chem* 2001;276:22041–22047.
- 30 Lien E, Sellati TJ, Yoshimura A, *et al*. Toll-like receptor 2 functions as a pattern recognition receptor for diverse bacterial products. *J Biol Chem* 1999;274:33419–33425.
- 31 Cario E, Brown D, McKee M, *et al*. Commensal-associated molecular patterns induce selective toll-like receptor-trafficking from apical membrane to cytoplasmic compartments in polarized intestinal epithelium. *Am J Pathol* 2002;160:165–173.
- 32 Smith Jr MF, Mitchell A, Li G, *et al*. Toll-like receptor (TLR) 2 and TLR5, but not TLR4, are required for *Helicobacter pylori*-induced NF-kappa B activation and chemokine expression by epithelial cells. *J Biol Chem* 2003;278:32552–32560.
- 33 Su B, Ceponis PJ, Lebel S, *et al*. *Helicobacter pylori* activates Toll-like receptor 4 expression in gastrointestinal epithelial cells. *Infect Immun* 2003;71:3496–3502.
- 34 Ishihara S, Rumi MA, Kadowaki Y, *et al*. Essential role of MD-2 in TLR4-dependent signaling during *Helicobacter pylori*-associated gastritis. *J Immunol* 2004;173:1406–1416.
- 35 Yamaoka Y, Kita M, Kodama T, *et al*. Chemokines in the gastric mucosa in *Helicobacter pylori* infection. *Gut* 1998;42:609–617.
- 36 Berstad AE, Hogasen K, Bukholm G, *et al*. Complement activation directly induced by *Helicobacter pylori*. *Gastroenterology* 2001;120:1108–1116.
- 37 Pappo J, Torrey D, Castriotta L, *et al*. *Helicobacter pylori* infection in immunized mice lacking major histocompatibility complex class I and class II functions. *Infect Immun* 1999;67:337–341.
- 38 Smythies LE, Waites KB, Lindsey JR, *et al*. *Helicobacter pylori*-induced mucosal inflammation is Th1 mediated and exacerbated in IL-4, but not IFN-gamma, gene-deficient mice. *J Immunol* 2000;165:1022–1029.
- 39 Fox JG, Beck P, Dangler CA, *et al*. Concurrent enteric helminth infection modulates inflammation and gastric immune responses and reduces helicobacter-induced gastric atrophy. *Nat Med* 2000;6:536–542.
- 40 Kaplan MH, Schindler U, Smiley ST, *et al*. Stat6 is required for mediating responses to IL-4 and for development of Th2 cells. *Immunity* 1996;4:313–319.
- 41 Shimoda K, van Deursen J, Sangster MY, *et al*. Lack of IL-4-induced Th2 response and IgE class switching in mice with disrupted Stat6 gene. *Nature* 1996;380:630–633.
- 42 Fan X, Crowe SE, Behar S, *et al*. The effect of class II major histocompatibility complex expression on adherence of *Helicobacter pylori* and induction of apoptosis in gastric epithelial cells: a mechanism for T helper cell type 1-mediated damage. *J Exp Med* 1998;187:1659–1669.
- 43 Raasi S, Schmidtke G, Groettrup M. The ubiquitin-like protein FAT10 forms covalent conjugates and induces apoptosis. *J Biol Chem* 2001;276:35334–35343.
- 44 Bates EE, Ravel O, Dieu MC, *et al*. Identification and analysis of a novel member of the ubiquitin family expressed in dendritic cells and mature B cells. *Eur J Immunol* 1997;27:2471–2477.
- 45 Mittrucker HW, Matsuyama T, Grossman A, *et al*. Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. *Science* 1997;275:540–543.
- 46 Imaizumi Y, Kohno T, Yamada Y, *et al*. Possible involvement of interferon regulatory factor 4 (IRF4) in a clinical subtype of adult T-cell leukemia. *Jpn J Cancer Res* 2001;92:1284–1292.
- 47 Al-Hajj M, Wicha MS, Benito-Hernandez A, *et al*. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci USA* 2003;100:3983–3988.
- 48 Lo Celso C, Prowse DM, Watt FM. Transient activation of beta-catenin signalling in adult mouse epidermis is sufficient to induce new hair follicles but continuous activation is required to maintain hair follicle tumours. *Development* 2004;131:1787–1799.
- 49 Monk M, Holding C. Human embryonic genes re-expressed in cancer cells. *Oncogene* 2001;20:8085–8091.
- 50 Gidekel S, Pizov G, Bergman Y, *et al*. Oct-3/4 is a dose-dependent oncogenic fate determinant. *Cancer Cell* 2003;4:361–370.
- 51 Wroblewski LE, Noble PJ, Pagliocca A, *et al*. Stimulation of MMP-7 (matrilysin) by *Helicobacter pylori* in human gastric epithelial cells: role in epithelial cell migration. *J Cell Sci* 2003;116:3017–3026.
- 52 Crawford HC, Krishna US, Israel DA, *et al*. *Helicobacter pylori* strain-selective induction of matrix metalloproteinase-7 *in vitro* and within gastric mucosa. *Gastroenterology* 2003;125:1125–1136.
- 53 Forsberg M, Druid P, Zheng L, *et al*. Activation of Rac2 and Cdc42 on Fc and complement receptor ligation in human neutrophils. *J Leukoc Biol* 2003;74:611–619.
- 54 Wu CW, Chen GD, Fann CS, *et al*. Clinical implications of chromosomal abnormalities in gastric adenocarcinomas. *Genes Chromosomes Cancer* 2002;35:219–231.
- 55 Yoshida T, Sato R, Mahmood S, *et al*. GATA-6 DNA binding protein expressed in human gastric adenocarcinoma MKN45 cells. *FEBS Lett* 1997;414:333–337.
- 56 Gupta A, Butts B, Kwei KA, *et al*. Attenuation of catalase activity in the malignant phenotype plays a functional role in an *in vitro* model for tumor progression. *Cancer Lett* 2001;173:115–125.