Development of *Helicobacter pylori* Whole-Proteome Arrays and Identification of Serologic Biomarkers for Noncardia Gastric Cancer in the MCC-Spain Study

Rima Jeske1,2, Dennis Reininger1, Busra Turgu1, Amber Brauer1, Christoph Harmel1, Nerea Fernández de Larrea-Baz3,4, Vicente Martín4,5, Victor Moreno4,6, Manolis Kogevinas4,7, Marina Pollán3,4, Jörg D. Hoheisel8, Tim Waterboer1, Julia Butt1, Nuria Aragonés4,9, and Katrin Hufnagel1

**ABSTRACT**

**Background:** *Helicobacter pylori* (*H. pylori*) is a bacterial carcinogen and the leading risk factor for noncardia gastric cancer (NGCC). Detecting antibodies against specific *H. pylori* proteins in peripheral blood can be applied to characterize infection and determine disease associations. Most studies analyzing the association between *H. pylori* infection and gastric cancer have focused on previously identified antigens, predominantly the virulence factor cytotoxin-associated gene A (CagA). Selecting antigens in an unbiased approach may, however, allow the identification of novel biomarkers.

**Methods:** Using a combination of multiple spotting technique and cell-free, on-chip protein expression, we displayed the *H. pylori* genome (strain 26695) on high-density microarrays. Immunogenic proteins were identified by serum pool incubations and henceforth analyzed in individual samples. To test its applicability, we used sera from a multicase–control (MCC)-Spain study. Serologic responses between NGCG cases and controls were assessed by conditional logistic regression estimating ORs and 95% confidence intervals.

**Results:** We successfully expressed 93% of the 1,440 *H. pylori* open reading frames in situ. Of these, 231 (17%) were found to be immunogenic. By comparing 58 NGCG cases with 58 matched controls, we confirmed a higher seroprevalence of CagA among cases (66%) than controls (31%). We further identified a potential novel marker, the *Helicobacter* outer membrane protein A (HopA).

**Conclusions:** In this study, we provide evidence that our *H. pylori* whole-proteome microarray offers a platform for unbiased de novo identification of serologic biomarkers.

**Impact:** Given its versatile workflow, antibody responses against other *H. pylori* strains and possible associations with diverse *H. pylori*-related outcomes can be systematically analyzed.

**Introduction**

*Helicobacter pylori* (*H. pylori*) is a Gram-negative bacterium that colonizes the human gastric mucosa. About 89% of noncardia gastric cancer (NGCG) cases, which constitute 78% of all gastric cancer cases, are considered attributable to an infection with this pathogen (1). In 2018, gastric cancer was the fifth most frequently diagnosed cancer and the third leading cause of cancer-related deaths. Randomized clinical trials supported the preventive effect of eradication in an infected asymptomatic population (2, 3). The prevalence of infection can, however, exceed 70% in some countries making screen-and-treat approaches extremely costly, especially using diagnostic assays such as the fecal antigen test or the urea breath test (4–6). Therefore, risk stratification is paramount to identifying people who need to be treated, reducing the number of eradication therapies, and thereby increasing the success of cancer prevention programs.

A widely investigated approach to identify these patients at risk is assessing the presence of specific antibodies in the blood. Furthermore, serologic assays based on multiple recombinant *H. pylori* proteins offer the potential advantage of measuring antibody patterns, which may be specific for disease (as opposed to detecting mere infection) or suited to characterize disease progression.

Currently, a well-studied marker is the presence of antibodies against the cytotoxin-associated gene A (CagA). This virulence factor is a product of the cag pathogenicity island encoded by a subgroup of *H. pylori* strains. Epidemiologic studies have shown that seropositivity to CagA increases the risk for both the precancerous state atrophic gastritis [OR, 3.48; 95% confidence interval (CI), 1.02–12.18; ref. 7] and NGCG [OR, 2.01; 95% CI, 1.21–3.32; ref. 8].

Especially in countries with low prevalence for *H. pylori* infections, like the United States, serologic testing for CagA seropositivity yields promising results (9). In East Asian countries, however, CagA-positive *H. pylori* strains are omnipresent, and testing for CagA antibodies does not substantially reduce the numbers of patients at risk. Because these are the countries with the most reported gastric cancer–related deaths by far, the need for further risk stratification markers continues to be high.

Antibodies against other *H. pylori* proteins have been tested and significant associations with gastric cancer have been reported (10–13). However, these findings are not consistent and most probably

1Infections and Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. 2Faculty of Biosciences, Heidelberg University, Heidelberg, Germany. 3Environmental and Cancer Epidemiology Area, National Center of Epidemiology, Carlos III Health Institute (ISCIII), Madrid, Spain. 4Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP)-CIBER of Epidemiology and Public Health (CIBERESP), Madrid, Spain. 5The Research Group in Gene-Environment and Health Interactions, University of León, León, Spain. 6Cancer Prevention and Control Program, Catalan Institute of Oncology (ICO), Hospital de Lliria, Barcelona, Spain. 7Global Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain. 8Functional Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany. 9Epidemiology Section, Division of Public Health, Department of Health, Madrid, Spain.

**Note:** Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

**Corresponding Author:** Rima Jeske, German Cancer Research Center, Im Neuenheimer Feld 280, Heidelberg 69120, Germany. Phone: 49-6221-4605; E-mail: r.jeske@dkfz.de

Cancer Epidemiol Biomarkers Prev 2020;29:2235–42

**doi:** 10.1158/1055-9965.EPI-20-0348

©2020 American Association for Cancer Research.
dependent on the population of interest. Therefore, an easily adaptable workflow to identify serologic markers specific for certain populations or outcomes is desirable.

In this study, we successfully adapted our recently published method to generate bacterial whole-proteome microarrays to the gastric bacterium, *H. pylori* (14, 15). We present high-density microarrays displaying the entire proteome [1,440 open reading frames (ORF)] of the strain 26695, and thereby, provide evidence that our method to generate bacterial whole-proteome microarrays can be adapted easily to different microorganisms in various areas of infection research.

As a proof-of-concept study, we used serum samples of NCGC cases and matched controls from the multicase–control (MCC)–Spain study with previously characterized *H. pylori* serostatus (16). Thus, we show the applicability of our *H. pylori* microarray and pave the way for upcoming serologic studies to include a prior antigen identification using the workflow presented here.

### Materials and Methods

#### Generation of *H. pylori* whole-proteome microarrays

*H. pylori* whole-proteome microarrays were generated as described for *Chlamydia trachomatis* (15, 17). Briefly, expression constructs were generated by two successive PCRs based on genomic DNA of the *H. pylori* strain 26695 (RefSeq: NC_000915). Gene-specific primers were used to amplify all 1,440 ORFs and to add common N- and C-terminal adaptor sequences.

In a second PCR, a set of primers was used that was complementary to the terminal adaptor sequences, which were added during the first PCR. They further encoded all sequences necessary for in vitro transcription and translation (15, 17). Successful amplification was verified by agarose gel electrophoresis. A band of the expected size was detectable in 97% of the amplification reactions.

Negative controls to determine background reactivity were generated by performing template-free PCR reactions. As positive control, viral capsid antigen (VCA) p18 from Epstein–Barr virus (EBV) was included. The reported seroprevalence for this antigen in adults is above 97% (18).

Final expression constructs, as well as positive and negative controls, were spotted onto nickel-nitrilotriacetic acid–coated glass slides (15, 17). Subsequently, each spot was overlaid with cell-free expression mixture (S30, T7 High-Yield Protein Expression Kit, Promega). Slides were incubated at 37°C for 1 hour and at 30°C overnight prior to storage at −20°C.

#### Assessment of on-chip protein expression

To determine the success of the on-chip protein expression, N-terminal 6xHis and C-terminal V5 tags were detected by immunostaining, as described previously (14, 15). Anti–6xHis–DyLight650 and anti–V5–Cy3 were used at 1:10,000 dilution in 2% BSA. Signals were detected with a PowerScanner (Tecan) at 532 and 635 nm excitation wavelengths and analyzed using the microarray analysis and acquisition software, GenePix Pro 6.0 (Molecular Devices). Spot morphology was evaluated visually, and spilled or smeared signals were marked as invalid; this affected <0.5% of spots in total (14).

A protein was considered to be expressed in full length if its raw median fluorescence intensity (MFI) for the C-terminal V5 tag exceeded the median plus five median absolute deviations (MAD) of all spotted negative control signals (PCR products lacking DNA template). Partial expression was presumed if only the N-terminal signal exceeded the cutoff.

### Proteome immunoassay

Proteome immunoassays were performed by incubating human sera on *H. pylori* protein microarrays diluted 1:33 in SuperBlock Blocking Buffer (ThermoFisher Scientific) for 1 hour (14, 15). To block unspecific reactions against the expression kit, *Escherichia coli* wild-type lysate (1 μg/ml) was used to preincubate the sera. Bound antibodies were visualized by a secondary AlexaFluor647-conjugated goat anti-human IgA, IgG, and IgM antibody (Jackson ImmunoResearch Laboratories). Signal intensities, given as MFI, were considered to be proportional to the amount of bound primary antibody from human sera to the antigen on the microarray. Signals were defined as seroreactive if the antigen-specific MFI exceeded the serum sample–specific cutoff defined as the median plus five MADs of all valid negative control spots for the respective serum sample.

#### Study population

Sera were obtained from the MCC-Spain project (www.mccspain.org), a study of multiple cancer types and population controls. Case recruitment was carried out between 2008 and 2013 in 12 Spanish provinces focusing on the five most prevalent tumor types in Spain: colorectal, breast, esophagogastric, prostate, and chronic lymphocytic leukemia. The MCC-Spain study was conducted in accordance with the Declaration of Helsinki. The study protocol was approved by the ethics committees of the participating institutions and written informed consent was obtained from each participant. Details regarding recruitment and study design are described elsewhere (16). Serum samples of these patients have been previously characterized by multiplex serology measuring antibodies to 16 specific *H. pylori* antigens (19). *H. pylori* seropositivity was defined as being positive to at least four of these antigens (20). Seroprevalences of 88% (95% CI, 87%–90%) and 95% (95% CI, 91%–98%) were reported among controls and NCGC cases, respectively.

For this study, 65 NCGC cases were randomly selected from the total of 202 cases included in the previous study, together with 65 sex-, age-, and region-matched healthy controls from the general practitioner’s lists at primary health care centers of hospital catchment areas from the provinces Barcelona, Leon, and Madrid. Characteristics of the NCGC cases and matched controls are summarized in Supplementary Table S1. Matching was performed for sex, age, and province. None of the additional covariates we examined (i.e., education, smoking status, family history of gastric cancer, and previous serology results) significantly differed between cases and controls.

### Experimental design for the identification of serologic biomarkers

To effectively target highly reactive antigens, and to reduce the required volume of patient sera, our workflow includes a preselction of immunogenic proteins (Fig. 1).

For this purpose, serum pools were generated according to case–control status, irrespective of previous *H. pylori* serology results (i.e., 26 pools comprising five sera each). *Post hoc* analysis confirmed that the few *H. pylori*–negative samples were distributed evenly across all serum pools. Seroresponses of these pools to each of the 1,440 potential antigens were determined on whole-proteome microarrays. Thereby, 231 immunogenic proteins were identified (see below).

Individual seroresponses to 242 antigens (see below) were subsequently determined on minimized microarrays. Antibodies potentially associated with NCGC were identified by comparing the seroresponses in cancer cases with controls. To apply conditional logistic regression, 14 samples were excluded because of erroneous matching. Both the serum pool incubations and the single-serum sample incubations
Two-stage microarray work are minimized. By preselecting immunogenic proteins, sample volume requirements responses were assessed on minimized microarrays displaying reactive antigens serum pools were tested on whole-proteome microarrays, individual serum statistical analysis each) to minimize batch-to-batch variation. Correlations were assessed by Spearman rank correlation coefficient (rho).

Negative controls (n = 56) were distributed across the slides and positive controls were placed in the corners of the microarray (Fig. 2, left).

Subsequently, cell-free protein expression was performed, and its success rate was determined by detection of terminal tags with fluorescence antibodies. Full-length expression was observed for 90% (n = 1,297) of the proteins. For an additional 3% (n = 43), partial expression was observed (Fig. 2, right).

Preselection of immunogenic proteins

Testing individual sera on whole-proteome microarrays was expected to produce mainly negative signals because most H. pylori proteins do not induce a humoral immune response in their host (21). To effectively target for strongly immunogenic proteins, and to reduce the required volume of patient sera, our work included a preselection of immunogenic proteins on the whole-proteome array (1,440 proteins; Fig. 1). For this purpose, we generated serum pools according to case–control status (i.e., consisting of either five cases or five controls).

Signal intensities of 231 antigens exceeded the cutof in at least one of the serum pools (Fig. 3A and B). These were considered for further testing on minimized, more targeted, microarrays.

Generation of minimized H. pylori microarrays

On the basis of the results of the pooled serum incubations, 231 immunogenic proteins were selected. In addition, further potentially interesting antigens were included, even though they did not emerge from the serum pool experiments. DnaK (HP0109) was chosen due to a recent report of oncogenic properties in another pathogen (22). VacA (HP0087), HpaA (HP0410), CagB (HP0522), Cad (HP1104), CagM (HP0537), HcpC (HP1096), and the hypothetical protein HP0231 had previously been used as serologic markers and were included to enable a comparison with published serology data from the MCC-Spain study (19, 23).

Furthermore, the genes for VacA, CagA, and HyuA were split into N- and C-terminal fragments to enhance protein expression. Altogether, 242 gene constructs were spotted on the minimized microarrays (Supplementary Table S3). Each slide fitted four sets of selected proteins (Fig. 4) and was subsequently used to test four individual sera.

Expression control of the minimized arrays was performed on four slides with four blocks each. For this purpose, the first slide of the spotting batch, the last slide, and two slides in between were used. A gene was considered overall expressed if it was detectable in at least half of the slides used for expression controls (nmax = 16). Full-length expression was observed for 96% of the spotted proteins and partial (N-terminal) expression for an additional 2%. Proteins without a detectable expression were HP1329, HP0410 (HpaA), HP0459, HP0547C (CagA–C), and HP0717. Of these, HP0547C exhibited reactivity in serum incubation experiments, indicating that the

Results

Generation of H. pylori whole-proteome microarrays

Sequences of 1,440 genes were amplified from genomic DNA of H. pylori strain 26695 and spotted individually onto the microarray. Testing individual sera on whole-proteome microarrays was expected to produce mainly negative signals because most H. pylori proteins do not induce a humoral immune response in their host (21). To effectively target for strongly immunogenic proteins, and to reduce the required volume of patient sera, our work included a preselection of immunogenic proteins on the whole-proteome array (1,440 proteins; Fig. 1). For this purpose, we generated serum pools according to case–control status (i.e., consisting of either five cases or five controls).

Signal intensities of 231 antigens exceeded the cutof in at least one of the serum pools (Fig. 3A and B). These were considered for further testing on minimized, more targeted, microarrays.

Generation of minimized H. pylori microarrays

On the basis of the results of the pooled serum incubations, 231 immunogenic proteins were selected. In addition, further potentially interesting antigens were included, even though they did not emerge from the serum pool experiments. DnaK (HP0109) was chosen due to a recent report of oncogenic properties in another pathogen (22). VacA (HP0087), HpaA (HP0410), CagB (HP0522), Cad (HP1104), CagM (HP0537), HcpC (HP1096), and the hypothetical protein HP0231 had previously been used as serologic markers and were included to enable a comparison with published serology data from the MCC-Spain study (19, 23).

Furthermore, the genes for VacA, CagA, and HyuA were split into N- and C-terminal fragments to enhance protein expression. Altogether, 242 gene constructs were spotted on the minimized microarrays (Supplementary Table S3). Each slide fitted four sets of selected proteins (Fig. 4) and was subsequently used to test four individual sera.

Expression control of the minimized arrays was performed on four slides with four blocks each. For this purpose, the first slide of the spotting batch, the last slide, and two slides in between were used. A gene was considered overall expressed if it was detectable in at least half of the slides used for expression controls (nmax = 16). Full-length expression was observed for 96% of the spotted proteins and partial (N-terminal) expression for an additional 2%. Proteins without a detectable expression were HP1329, HP0410 (HpaA), HP0459, HP0547C (CagA–C), and HP0717. Of these, HP0547C exhibited reactivity in serum incubation experiments, indicating that the
Figure 2. *H. pylori* microarray displaying the proteome of strain 26695. Schematic representation of the spotting layout (left). Each spot represents a different protein. Positive controls (EBV VCA p18) are marked in dark blue (four corners) and negative control spots (n = 56) in light blue. Determination of on-chip expression by anti-tag staining (right). Green signals indicate anti-V5 (C-terminal) signal, while red signals indicate anti-6xHis (N-terminal) antibody. Merged signals appear yellow.
terminal tags may potentially not be accessible for the antibodies used for quality control.

**Proof-of-concept evaluation of potential serologic *H. pylori* markers for NCGC**

Sera of the matched cases and cancer-free controls were applied individually to the minimized microarrays. Representative pictures are depicted in Fig. 5A and B. The mean number of reactive antigens did not significantly differ between cases and controls ($n = 17$ vs. $20$; $P = 0.13$), neither did the mean fold change values considering all antigens (0.75 vs. 0.77; $P = 0.18$) or only reactive antigens (2.68 vs. 2.86; $P = 0.28$). Also, reactivities toward positive antigen controls (EBV VCA p18) were similar between cases and controls (7.22 vs. 7.06; $P = 0.78$). Furthermore, no correlation was observed between the number of reactive antigens and the patients’ age ($\rho = 0.01$) or sex ($P = 0.31$).

After application of a cutoff for seropositivity, the seroprevalence of the established NCGC serologic marker, HP0547N (CagA-N), was higher in NCGC cases (66%) than in controls (31%; OR, 3.64; 95% CI, 1.69–7.87). We further identified seropositivity to the protein HP0229 [membrane protein, *Helicobacter* outer membrane protein A (HopA)] to be higher in NCGC cases (22%) than in controls (7%; OR, 4.58; 95% CI, 1.23–17.01). Controls were more likely to have antibodies against HP1325 [fumarate hydratase (FumC)], HP0294 [acylamide amidohydrolase (AmiE)], HP0385 [hypothetical protein], and HP0243 [neutrophil activating protein (NapA)] than cases (Supplementary Table S4).

Antibodies against HP0512 [glutamine synthetase] and HP1071 [phosphatidylserine synthase (PSS)] were exclusively detected in controls with a prevalence of 17% and 12%, respectively.

We also compared the magnitude of the seroresponse with the eight above-described antigens between seropositive cases and controls.
Representative pictures of single-serum testing on minimized microarrays. **A**, Controls (top). **B**, NCGC cases (bottom). Red signals indicate reactivity with the respective antigens; saturated signals appear white. Positive control EBV VCA p18 was placed at the four corners of the microarray. Negative controls (n = 25) were distributed across the slide, as illustrated in Fig. 4, left. **C**, Representative reactivity plot (NCGC case from **B**, middle). MFI signals were normalized by division through the patient-specific cutoff (mean reactivity to negative controls plus five SDs) and displayed as fold change values on the y-axis. The threshold (y = 1) is represented by a red line. Each signal exceeding this threshold is annotated with the respective locus tag (HP number). Signals that were derived from positive controls are not shown in this plot. The two top hits (HP0797 and HP0547N) correspond to the two brightest signals located near the left lower corner of the picture.

Only CagA signals showed a significant increase in cases, with a mean fold change of 3.45 (SD, 4.00) versus 1.29 (SD, 1.93) in cancer-free controls (P < 0.001).

**Discussion**

In this study, we generated bacterial whole-proteome microarrays to develop an unbiased approach for the detection of seromarkers for the gastric bacterium, *H. pylori*. We were able to display 93% of the proteome (1,340 of 1,440 ORFs) of the *H. pylori* strain 26695 on a high-density microarray and selected 242 antigens for a subsequent proof-of-concept analysis. By analyzing 116 NCGC cases and matched controls, we were able to reproduce the well-known NCGC marker, CagA, which was associated with a 3.6-fold increased odds for NCGC in our study (8, 24, 25, 26). Furthermore, we identified a new NCGC-associated seromarker, HopA (HP0229).

The microarray technique was initially developed to study the antibody response to the sexually transmissible pathogen, *Chlamydia trachomatis*, in detail. Thereby, we found novel infection markers (e.g., CT_813) and also risk markers associated with cervical cancer (e.g., CT_117; ref. 15). Identified markers are being currently validated in seroepidemiologic studies using high-throughput multiplex serology. The success of the method led us to its transfer to the gastric bacterium, *H. pylori*, and the generation of *H. pylori* microarrays.

For the initial testing and a proof-of-concept analysis, the MCC-Spain study was chosen due to a high baseline prevalence of *H. pylori* in the study population (95% in NCGC cases and 88% in controls), which was reported in a previous study. Fernández de Larrea-Baz and colleagues measured reactivities to 16 antigens by multiplex serology, of which 13 were included in our antigen selection (19).

The results we generated by measuring a subset of the same serum samples compared very well with the findings of Fernández de Larrea-Baz and colleagues (Supplementary Table S5), who reported antibody responses to CagA and VacA as significantly associated with NCGC. We replicated these results for CagA. Our risk estimates for both VacA-N and VacA-C were also elevated, but not statistically significant. This may be due to small sample size and thus, diminished power to detect significant associations.

Overall, we reproduced the trend that were shown by Fernández de Larrea-Baz and colleagues for the antigens present in both studies. However, the detected prevalence for each antigen was lower using the microarray platform compared with the bead-based multiplex assay, for example, anti-UreA was detected in 48% of cases and 66% of controls by multiplex serology and in 16% and 31%, respectively, by microarray. Patients who showed reactivity to one of the tested proteins on the microarrays generally tested seropositive for the same antigen on the bead-based multiplex assay (Supplementary Table S5). Thus, we were able to confirm the findings from the newly developed microarray-based technique with results from the well-established *H. pylori* multiplex serology assay developed by Michel and colleagues (20). However, comparing these two methods, antibody responses were presumably detected with a higher sensitivity in the bead-based multiplex assay, or alternatively with a higher specificity using the microarray.

Some expected signals did not emerge as seroreactive on the miniimized microarrays, such as HP1564 (Omp) and HP0305 (hypothetical protein), which were previously published as NCGC-associated risk markers by Cai and colleagues and Epplein and colleagues using multiplex serology (10, 12). This may be caused by population differences. While we examined a Western population, the above-mentioned associations were identified in Chinese cohorts.

By comparing cases with controls, we also identified a potential new NCGC serologic marker, HopA. The protein is a monomeric porin enabling passive diffusion of small molecules and nutrients (27). To our knowledge, its potential functional role in, or its association with, cancer development has not been characterized yet. However, other members of the Hop family sharing a common N-terminal motif (27), including the blood group antigen-binding A adhesin (BabA), the sialic acid-binding adhesion (SabA), and the outer inflammatory protein A (OipA), were described as risk factors for gastric cancer and MALT lymphoma (28–30). Gastric cancer is accompanied by a reduced rate of basal gastric acid secretion, leading to a higher pH (31). Allan and colleagues reported the downregulation of HopA protein expression at low pH. HopA could, therefore, be important for the adaptation of the organism to its surrounding conditions (32). Upregulation of HopA may result in higher antibody levels to this protein and reflect the bacteria adapting to pathologic changes. Because our...
Further significant associations showed an inverse relation, that is, higher seroprevalences in controls. Antibodies against the FumC (HP1325), the AmiE (HP0294), the DNA protection during starvation protein, NapA (HP0243), and the hypothetical protein (HP0385) were predominantly found in controls; antibodies to glutamine synthetase (HP0512) and PSS (HP1071) were exclusively found among controls. NapA is already reported as inversely associated with NCGC in the MCC-Spain study using multiplex serology (19). Although we obtained the same results using a subset of these samples, the finding is not consistent throughout the literature and might be dependent on the investigated population (10). It remains to be seen whether antibodies against NapA or one of the further inversely associated proteins play a direct role in the host’s immune defense. Alternatively, the loss of these antibodies can also be an accompanying effect when losing the H. pylori infection due to cancer formation (33). The value of these targets as serologic biomarkers needs to be further evaluated in larger studies.

Apart from seroréponses discriminating cases from controls, we also noticed antigens that were predominantly detected in sera from patients who tested positive for H. pylori by multiplex serology, but rarely in patients who tested negative (Supplementary Table S6; ref. 20). They are potentially interesting for serologic diagnosis of mere H. pylori infection, but this has to be evaluated in future experiments with a suitable study design.

One limitation of whole-proteome microarrays is the fact that the expression and conformation of the expressed proteins cannot be verified for every single antigen, and batch effects cannot be excluded. We assessed biological variation (individual sample background reactivity) by normalizing MFI values to fold changes. To account for technical variation, we controlled intermediate results, for example, PCR product sizes, and performed expression controls.

Generally, we found that long genes more often led to partial expression. Therefore, the genes cagA, vacA, and hpyA were split for expression on the minimized microarrays. Apart from the length of the proteins, we did not identify further factors which might systematically impact efficient protein expression, for example, expression of membrane proteins. Although the hydrophobicity index did not significantly differ between expressed and nonexpressed proteins, it remains unclear to what extent these proteins are folded in a physiologic manner. Overall, we assume that their expression does not pose a major barrier to the validity of the microarray technique because antigens usually contain multiple epitopes, some of them conformational, others linear. In our experiments, the signals derived from membrane proteins were equally present and strong. We, therefore, conclude that they are sufficiently folded and assessable to measure serum antibody responses. Despite these limitations, we believe that the microarray platform offers unique advantages in the parallel representation and testing of numerous antigens. It obviates the need for work- and time-intensive protein production by simultaneous on-chip expression. In addition, it is easily transferable to other pathogens and outcomes. The technique compares well with other immunomasys, which we demonstrated by replicating the association between NCGC and antibodies to the well-known antigen, CagA. Furthermore, it fills a missing link in H. pylori research by identifying novel associations without prior assumptions, that is, in an unbiased fashion.

Given the easily adaptable workflow to generate bacterial microarrays, experiments can be tailored to a specific research question or a specific population. While we used a matched case–control setup to identify associations between specific serum antibodies and NCGC, the platform can also be used to characterize sera from prospective studies and thereby identify potential prognostic markers. It further offers the possibility to assess both population-specific differences and differences in the infecting H. pylori strain by representing proteomes from different strains. In conclusion, our microarray platform serves as methodologic extension for classical high-throughput systems to initially identify novel and strongly reactive targets.

**Disclosure of Potential Conflicts of Interest**

N. Fernández de Larrea-Baz reports grants from Instituto de Salud Carlos III and FEDER funds during the conduct of the study. V. Moreno reports grants from Agency for Management of University and Research Grants (AGAUR) of the Catalan Government and Institut de Salud Carlos III during the conduct of the study. N. Aragónes reports grants from Instituto de Salud Carlos III (FEDER funds, grant no. PI11/01403) during the conduct of the study. No potential conflicts of interest were disclosed by the other authors.

**Disclaimer**

The content and views of this article are those of the authors and do not necessarily reflect the official position of the Instituto de Salud Carlos III.

**Authors’ Contributions**

R. Jeske: Conceptualization, formal analysis, investigation, methodology, writing–original draft. D. Reiminger: Investigation, methodology, writing–original draft. B. Turgu: Investigation, methodology, writing–review and editing. A. Brasera: Formal analysis, visualization, writing–review and editing. C. Harme: Formal analysis, visualization, methodology, writing–review and editing. N. Fernández de Larrea-Baz: Resources, data curation, project administration, writing–review and editing. V. Martin: Resources, data curation, writing–review and editing. V. Moreno: Resources, data curation, writing–review and editing. N. Kogevinas: Resources, data curation, writing–review and editing. M. Pollan: Resources, data curation, writing–review and editing. N. Pollan: Resources, data curation, writing–review and editing. J. D. Hoheisel: Formal analysis, supervision, writing–review and editing. T. Waterboer: Conceptualization, supervision, writing–review and editing. J. Butt: Conceptualization, supervision, writing–review and editing. N. Aragonés: Conceptualization, resources, data curation, supervision, writing–review and editing. K. Hufnagel: Conceptualization, supervision, methodology, writing–original draft.

**Acknowledgments**

The authors thank the cooperating hospitals, as well as the participants, for enabling this study. The MCC-Spain study, represented by N. Fernández de Larrea-Baz, V. Martin, V. Moreno, M. Kogevinas, M. Pollan, and N. Aragónes, received funding from the following sources: “Acció Transversal del Càncer,” approved by the Spanish Ministry Council on October 11, 2007, the Instituto de Salud Carlos III–FEDER funds—a way to build Europe (grant nos. PI08/1770, PI08/0533, PI08/1359, PS09/00775, PS09/01662, PS09/01286, PS09/01903, PS09/02078, PI11/00236, PI11/0483, PI11/03810, PI11/01889, PI11/02213, PI12/00150, PI12/00265, PI12/00488, PI12/00715, PI12/01270, PI14/00613, and PI17/00092); the Fundación Marques de Valdecilla (grant no. API 10/09); Obra Social CAJAUSTUR (grant no. SV-CAJAUSTUR-1); Recercaixa (grant no. 2010ACUP 00310); Spanish Association Against Cancer Scientific Foundation; Agència de Gestió d’Autors Universitaris i de Recerca—Generalitat de Catalunya (Catalonian Government, grant nos. 2009SGR1026, 2009SGR1465, and 2017SGR1723); and Junta de Castilla y León (grant no. LE022A10-2. Sample collection and storage was partially supported by the Instituto de Salud Carlos III–FEDER (grant no. RD09/0076/00364), Xarsa de Bancs de Tumors de Catalunya sponsored by Pla Direcció d’Oncolegia de Catalunya. N. Aragonés received funding for the conduct of this study from the Instituto de Salud Carlos III–FEDER funds (grant no. PI11/01403).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 5, 2020; revised June 25, 2020; accepted September 4, 2020, published first September 30, 2020.
References


Development of *Helicobacter pylori* Whole-Proteome Arrays and Identification of Serologic Biomarkers for Noncardia Gastric Cancer in the MCC-Spain Study

Rima Jeske, Dennis Reininger, Busra Turgu, et al.


Updated version

Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-20-0348

Supplementary Material

Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2020/09/11/1055-9965.EPI-20-0348.DC1

Cited articles

This article cites 33 articles, 10 of which you can access for free at:
http://cebp.aacrjournals.org/content/29/11/2235.full#ref-list-1

E-mail alerts

Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions

To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions

To request permission to re-use all or part of this article, use this link
http://cebp.aacrjournals.org/content/29/11/2235.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.