Comparative Medicine
Copyright 2007
by the American Association for Laboratory Animal Science

Vol 57, No 1
February 2007
Pages 57-65

A Reproducible Scoring System for Quantification of Histologic Lesions of Inflammatory Disease in Mouse Gastric Epithelium

Kathryn A Eaton,[†,*] Stephen J Danon, Steven Krakowka, and Steven E Weisbrode

Comparison of experimental groups by microscopic examination is a common and useful method for evaluating animal models of disease. Quantification of lesions is challenging, however, and differences in scoring systems hinder comparison of results from different laboratories. The purpose of this study was to validate a simple and reproducible scoring system for *Helicobacter pylori*-associated gastric disease in mice. The system is based on quantification of the percentage of microscopic fields in which lesions are present, rather than on subjective estimates of lesion severity. Linear regression analyses revealed good agreement between investigators in scoring of all 3 histologic criteria examined. The range of correlation coefficients between individual readers' scores and mean scores for the 3 histologic criteria examined were: neutrophilic inflammation, 0.845 to 0.935; gastritis sufficient to displace glands, 0.919 to 0.943; and epithelial metaplasia, 0.650 to 0.799. Comparison of scores in different experimental groups by analysis of variance and Fisher least significant difference tests revealed significant differences between infected and uninfected groups and between immunodeficient and immunocompetent groups. We propose that this system may be useful in standardizing the morphologic evaluation of rodent models of *H. pylori* and that a similar system could be devised for evaluation of other animal models of enteric disease.

Abbreviations: SCID, severe combined immunodeficient mice (C57BL/6J-*Prkdc*^{scid}); rSCID, recipient SCID (SCID mouse given congenic splenocytes by adoptive transfer); PMN, polymorphonuclear leukocytes (neutrophils); PI, postinfection; PT, post-transfer

Microscopic examination of anatomic lesions in tissues has been used experimentally since reliable methods for fixation and microscopic examination became available. As biomedical research has become more quantitative, investigators have sought reliable methods for quantifying morphologic lesions, so that groups might be compared statistically or at least more objectively. Because by its nature pathologic diagnosis is subjective and dependent on tissue handling, fixation, and other methodologic considerations for its reproducibility, quantification has been challenging.

Histologic lesions in tubular organs such as the gastrointestinal tract are particularly difficult to quantify for a variety of reasons. When gastrointestinal samples are placed into formalin, the muscularis contracts variably, resulting in distortion of the tissue, which complicates embedding, and shrinkage in length, so that total length is not comparable between tissue preparations When gastrointestinal tissue is cut and embedded, the distance between glands and the number of cells per unit area vary depending on how much that particular sample contracted during fixation. In addition, intestine is a complex tissue with many diverse cell populations and microcompartments, which differ not only between histologic sections but also along the linear axis of the tissue. Therefore care must be taken to sample exactly the same region of tissue in each and every animal. Because macroscopic anatomic landmarks may not be available to facilitate the necessary precision, sampling of intestinal tissue is difficult.

Because of the difficulty of quantification, investigators develop semiquantitative scoring systems for gastrointestinal lesions whereby subjective evaluations of, for example, intensity of inflammatory infiltrate are graded (for example, as none, mild, moderate, or severe) blindly by an experienced histopathologist, and the scores are converted into numbers (0, 1, 2, 3) which can be averaged and compared across groups.

Scoring of gastric lesions due to *Helicobacter pylori* is most commonly done using semiquantitative systems.[9,13,14,16,19] *H. pylori* is a human bacterial pathogen that is associated with gastritis, peptic ulcer disease, and gastric cancer. In C57BL/6 mice, *H. pylori* causes slowly progressive chronic gastritis that is detectable by 8 to 12 wk after inoculation and progresses to severe proliferative gastritis with epithelial metaplasia by 6 mo after inoculation.[8] In immunodeficient mice that have been reconstituted by adoptive transfer of congenic splenocytes, this gastritis is severe and rapidly progressive.[8] The severity of lesions varies according to duration of infection, mouse strain, bacterial strain, and bacterial virulence factor expression[5,9,12-14,19] and therefore scoring of histologic lesions in mice is commonly employed. The various semiquantitative scoring methods used are often very accurate when performed by the same observers in the same laboratory, but can be difficult to compare between observers and between laboratories. Scoring criteria may differ between laboratories, and investigators often disagree regarding features that differentiate the severity of lesions. In addition, many laboratories use scoring systems that are modified from systems developed for other

Received: 30 May 2006. Revision requested: 8 Aug 2006. Accepted: 15 Aug 2006.
Ohio State University, Columbus, OH.
[†]Present address: University of Michigan Unit for Laboratory Animal Medicine, Ann Arbor, MI.
[*]Corresponding author. Email: kateaton@umich.edu

models or based on lesions in human tissue but are not validated for relevance to the model in which they are applied.[17,20,22]

For all of these reasons, quantitative scoring of histologic lesions in animal models of *H. pylori* is difficult to perform and interpret. Nevertheless, standardized approaches to determining differences in treatment groups are increasingly vital for evaluation of animal model systems and for identification of changes that merit mechanistic study. Therefore, standardized, biologically relevant, reproducible scoring systems that can readily be used for comparison of results between different laboratories are needed. The goal of this study was to evaluate a scoring system for gastric lesions of *H. pylori* in mice. The system was intended to be 1) easy to perform, 2) reproducible across tissues and observers, 3) easy to learn and transfer between laboratories, and 4) predictive of disease status.

Materials and Methods

Investigators. Four different investigators ('readers'), who were blinded regarding experimental group, evaluated the slides and scored them according to the criteria described later. Three of the readers were board-certified veterinary pathologists, and the remaining one was a microbiologist with extensive experience in evaluating animal models of *H. pylori*. Three readers were actively involved with *H. pylori* research, and 1 (a board-certified pathologist) was not. The 4 readers scored the same set of coded slides and reported the results according to the code number on each slide. When all readers had scored all the slides, the results were tabulated.

Scoring system. The system was based on a protocol developed by one of the readers and modified by agreement of all 4 readers as follows. All 4 readers examined a subset of the available slides after having read a description of the scoring system, and then they all met to review the slides and agree on a modified scoring system. The essential elements were that the system would be objective (not requiring estimates of severity), include elements of both inflammation and epithelial changes, and permit quantitation of large numbers of slides rapidly but reproducibly with minimal observer fatigue.

The examination protocol and scoring system were as follows. Each glass slide contained sections of well-oriented gastric glandular mucosa that extended along the greater curvature of the stomach from cardia to pylorus (Figure 1). The readers examined all sections in their entirety by use of an objective lens with magnification of ×20 (total magnification, ×200). Only fields that contained full-thickness gastric mucosa that was oriented perpendicularly were scored. Because of the method of tissue preparation (described later), most fields could be scored. Only a few fields per slide either lacked the full thickness of the mucosa or were poorly oriented and therefore could not be scored. The number of fields scored per slide (mean ± 1 standard deviation) was 19.2 ± 6.4.

Each ×200 microscopic field was scored separately for the presence or absence of each of the following 3 histologic criteria: 1) neutrophilic infiltration (polymorphonuclear leukocytes [PMN]), 2) gastritis, and 3) epithelial metaplasia. An important aspect of the scoring system was that no attempt was made to grade severity of the change. Each microscopic field was scored only for presence or absence of the lesion according to the agreed-upon definition, and the results were reported as the percentage of fields affected on each slide. The histologic criteria scored were defined as follows:

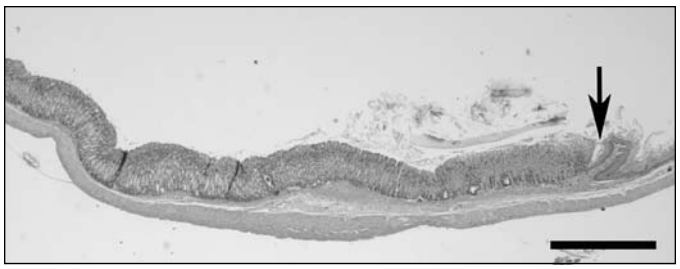

Figure 1. Hematoxylin and eosin-stained section of a well-oriented 5-µm histologic section of gastric mucosa. The majority of the section is perpendicular to the axis of the glands and includes the full thickness of the mucosa. The section extends from the cardiac junction (arrow) through the pylorus (not shown). Bar, 1 mm.

Neutrophilic inflammation was defined as the presence of at least several clusters of neutrophils within the mucosa. A cluster was defined as 3 or more contiguous cells (that is, an imaginary straight line between their nuclei would not cross any other cells). The actual number of clusters was not defined, but this factor did not present a problem because clusters were either numerous or absent. Gland displacement was not a requirement for scoring a field positive for PMN.

Gastritis was defined as inflammatory cell infiltrate (any cell type) that was sufficient to displace gastric glands. Epithelial metaplasia was defined as any evidence of loss of parietal cells with replacement by mucus-type cells. For each slide the number of positive fields (fields in which the lesion was present) was divided by the total number of fields in the section and multiplied by 100% to calculate the percentage of affected fields.

Preparation of tissue samples. To minimize variability, all stomachs were processed in the same way by the same person. Stomachs were removed from the mice immediately after death and bisected along the greater and lesser curvatures. The stomachs were placed serosa-side down on absorbent paper, and 2 adjacent full-thickness longitudinal strips were removed from the greater curvature, placed in a tissue-processing cassette, and immersed in neutral buffered formalin. Each strip included the junction of glandular and nonglandular mucosa on the proximal end and extended into the proximal duodenum at the distal end, although only the gastric mucosa was scored. The strips remained adhered to the paper throughout processing, ensuring that the tissue would not curl during formalin fixation and dehydration. The sections were processed routinely in an automatic processor and embedded in paraffin such that they could be cut to include well-oriented cross-sections of the entire mucosa that spanned from cardia to pylorus in so far as possible; 5-µm sections were stained with hematoxylin and eosin.

Mice. The sections examined were from female specific pathogen-free (including helicobacters) C57BL/6J and B6.CB17-*Prkdc*^{scid}/SzJ (severe combined immunodeficient [SCID]) mice from several different studies.[6-8] In all studies, mice were 6 to 8 wk old when infected, and uninfected controls were age-matched. Age at euthanasia depended on the duration of the individual experiment, as indicated in Results. Mice were housed in standard isolator cages and offered water and lab chow ad libitum. Some groups of SCID mice received congenic splenocytes by adoptive transfer as previously described.[8] For this, splenocytes were isolated by disaggregation, washed, and adjusted to 1×10^7 viable cells/ml, and 0.1 ml per mouse was administered by intraperitoneal injection.

Three categories of mice were included: 1) mice that were not expected to have gastric lesions (uninfected C57BL/6J mice, uninfected SCID mice, and *H. pylori*-infected SCID mice that were not adoptively transferred); 2) mice that, based on previous experience,[5-8] were expected to have moderate gastric lesions (*H. pylori*-infected C57BL/6J mice); and 3) mice expected to have severe gastric lesions (*H. pylori*-infected SCID mice that had received congenic splenocytes by adoptive transfer [recipient SCID mice, rSCID]). Infection by *H. pylori* was established by oral gavage and demonstrated in all cases by culture.[5-8] In experiment 1, all 4 readers scored sections from 146 mice as described. In experiment 2, a single reader scored 314 additional sections by use of the same scoring system. All procedures involving animals had been approved by the Ohio State University Laboratory Animal Care and Use Committee.[5-8]

Statistics. Values were calculated using Statview for Windows, V. 5.0.1 (SAS Institute, Inc.) To compare the level of agreement between investigators in experiment 1, a mean score for each slide was calculated for each histologic lesion. These mean scores were plotted against the individual scores for each investigator, and linear regressions were calculated. Values of R^2 (goodness of fit) and P values for the slope of the regression line are indicated. To determine biologic relevance of the scoring system, group means were compared by analysis of variance with the Fisher Least Significant Difference post-test to compare individual groups. We assigned a threshold of $P < 0.05$ for statistical significance. Finally, mean scores in experiment 1 and individual scores in experiment 2 were evaluated to determine whether there were ranges of scores that included all or most of the animals in specific treatment groups.

Results

Histology. Figure 1 illustrates the overall appearance of the sections. Each slide contained linear full-thickness sections of stomach embedded and cut as closely as possible to the perpendicular axis. Figure 2 demonstrates an example of 2 fields that would be scored positive for PMN. The box in Figure 2 A indicates a large cluster of neutrophils. In Figure 2 B, inflammatory infiltrate containing a mixture of cell types, including neutrophils, displaces much of the base of the gastric epithelium. To be scored as positive for PMN, the field had to have contained at least several clusters as indicated in Figure 2 A. However, the range in intensity of the infiltrate, as illustrated by the 2 examples, was not included in the score; thus, both Figure 2 A and 2 B were scored as positive. Similarly, Figure 3 demonstrates the range of gastric gland displacement by inflammatory infiltrate, scored as gastritis. Displacement that was mild as in Figure 3 A or severe as in Figure 3 B was scored as positive, but fields containing scattered inflammatory cells that did not displace glands (Figure 3 C) were scored as negative for gastritis. Infiltrates included mixtures of lymphocytes, plasma cells, and neutrophils, but individual cell types were not recorded for the gastritis score. Figure 4 illustrates the appearance of metaplastic gastric epithelium. Normal fundic glands contain mostly parietal cells and chief cells with a zone of mucus cells confined to the surface and superficial pits. Fields that were scored as positive for metaplasia contained at least 1 gland in which mucus type cells were present deep in the glands and replaced the normal parietal cells or chief cells or both (Figure 4 A). Often the metaplastic change was severe, resulting in complete replacement of normal gastric glands with metaplastic glands (Figure 4 B). Sections with widespread metaplasia as shown in Figure 4 B also contained intense inflammatory infiltrates.

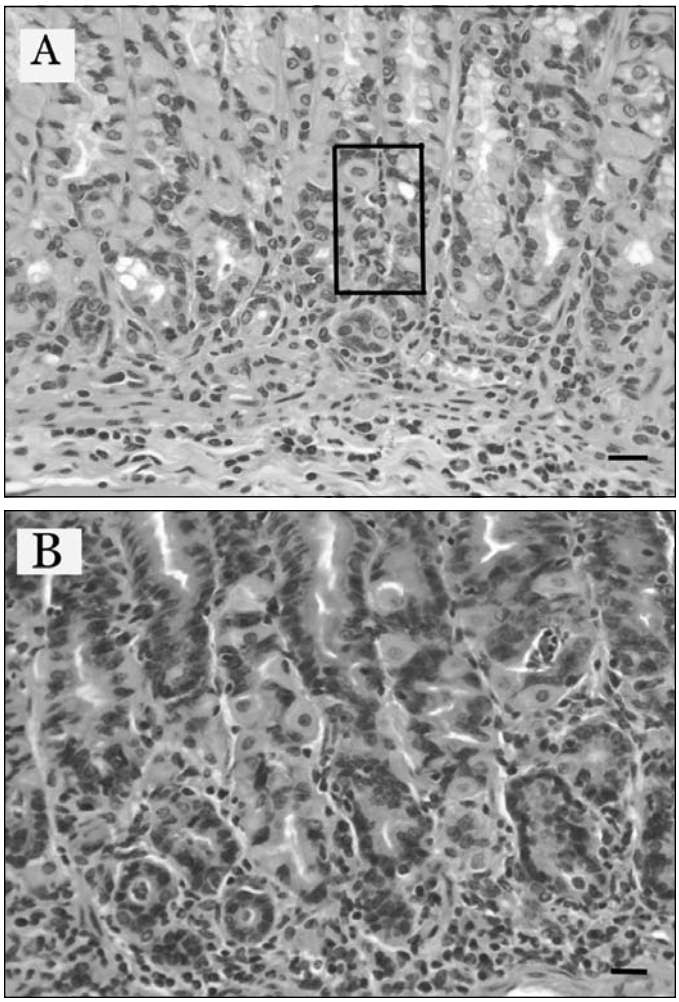

Figure 2. Subsets of histologic fields that were scored positive for PMN. (A) The box indicates a large cluster of neutrophils is indicated; several additional smaller clusters are present in the field. This level of infiltration is mild. (B) Inflammatory infiltrate, including many neutrophils, forms a band across the base of the gastric mucosa. This level of infiltration is moderate to severe. Regardless of the severity of infiltrate, both fields were scored simply as positive, for the purposes of this study. In addition, all sections were scored at ×200 magnification; for illustrative purposes, only partial fields are shown. Bar, 20 µm.

Experiment 1—agreement between investigators. The level of agreement among readers was determined by linear regression comparing the individual and mean scores for each slide for each of the 3 histologic criteria. For this comparison, each reader's scores for each investigator were averaged, and each slide was assigned 3 average scores, 1 for each histologic criterion (PMN, Gastritis, Metaplasia). We then plotted each reader's scores against the average and calculated a linear regression. Figure 5 shows the regression plots of reader A's scores plotted against mean score. Plots for the other 3 readers were similar (not shown). Table 1 shows the correlation coefficients, slopes, and P values of the regressions for each reader for each histologic criterion. As Table 1 shows, this analysis indicated that with 1 exception (discussed later), the readers' scores agreed well.

Agreement between all readers was highest for PMN. Values of R^2 for PMN were all greater than or equal to 0.845, indicating good correlation between each individual reader and the mean

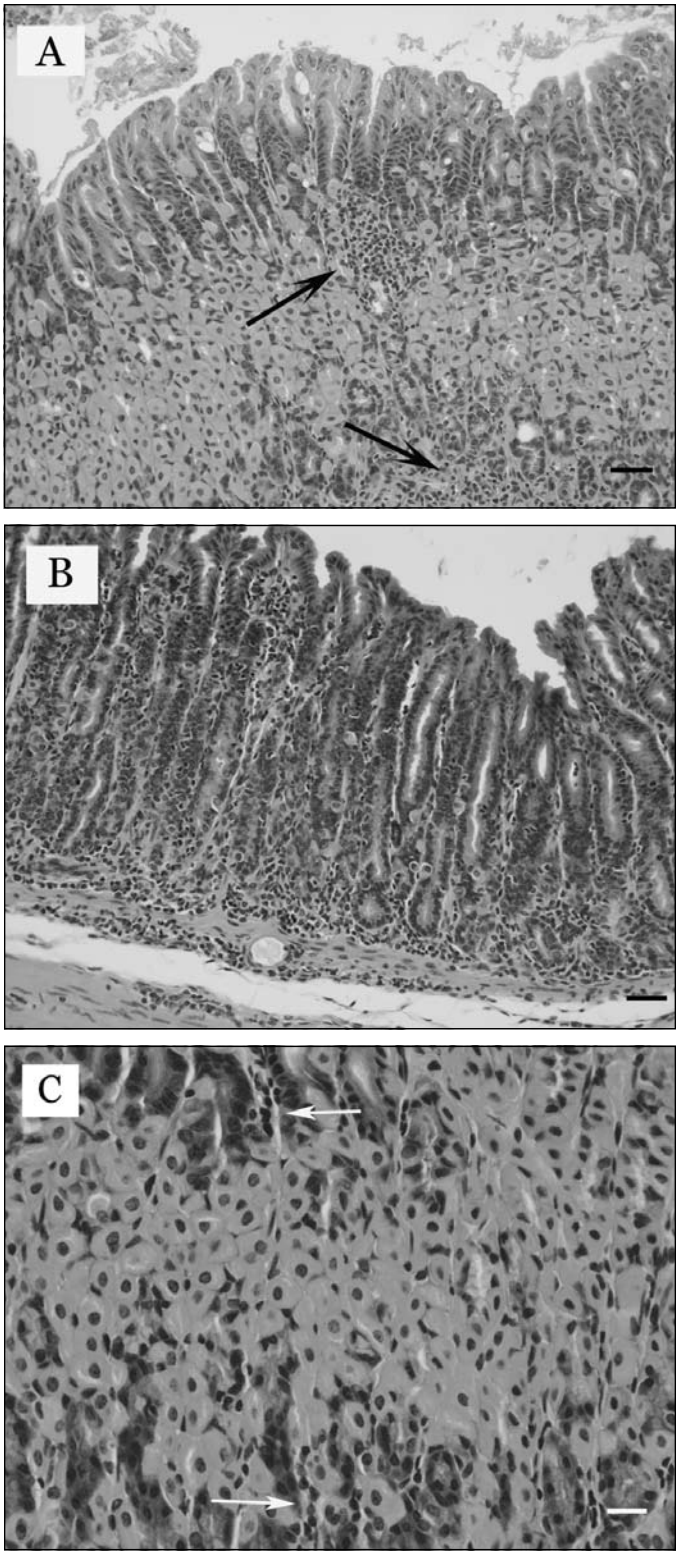

Figure 3. Displacement of gastric glands by inflammatory infiltrate, which was scored as gastritis. (A) Mild displacement by small clusters of inflammatory cells in the superficial and deep mucosa (arrows). Bar, 40 µm. (B) Lower magnification of Figure 2 B shows widespread displacement of the base of glands and extension into the superficial mucosa. Bar, 40 µm. Both A and B were scored as positive for gastritis. (C) Small clusters of lymphocytes that do not displace gastric glands (arrows). This field was scored as negative for gastritis. Bar, 20 µm.

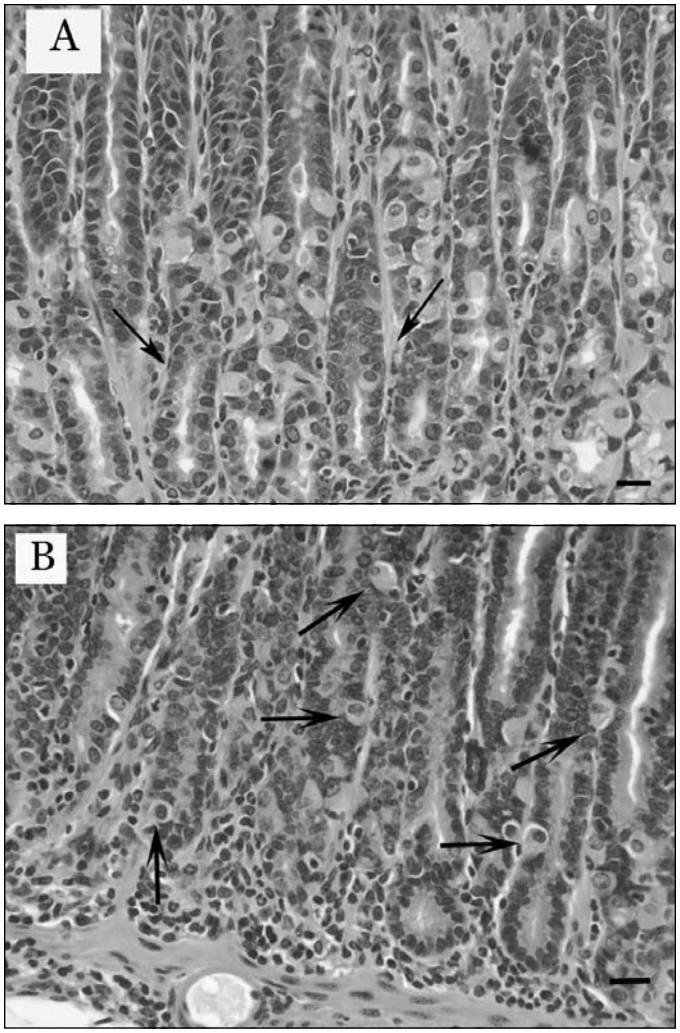

Figure 4. Fields that were scored as positive for metaplasia. (A) Arrows indicate parts of glands in which parietal cells are reduced in number or absent and are replaced by mucus-type cells. (B) More extensive metaplasia in which most of the glands are replaced by mucus-type epithelium. Arrows indicate scattered retained parietal cells. Bar, 20 µm.

scores. Correlation was lower and more variable for metaplasia, for which R^2 values ranged from a high of 0.799 (reader A) to a low of 0.650 (reader C). For gastritis, 3 of the readers agreed well (R^2 ranged from 0.874 to 0.885), but reader B differed markedly, demonstrating minimal correlation with the mean values. Retrospective evaluation revealed that this reader inadvertently had used a different scoring system for gastritis than the one used by the other 3 readers. Recalculation of the mean scores without reader B's scores revealed a better correlation for the scores of readers A, C, and D and the mean ($R^2 = 0.943$, 0.926, and 0.919 for readers A, C, and D, respectively). Reader B's scores correlated poorly ($R^2 = 0.356$).

Biologic relevance of the scores. Statistically significant agreement among scores is only useful if the scores reflect a meaningful aspect of the pathobiology, such as infection or mouse strain. To determine whether scores reflected biologic relevance, we assigned the scores to experimental groups, as shown in Figure 6. The uninfected mice, mice that were infected for a short period of time, and SCID mice that had no adaptive immune response showed few lesions, whereas mice that were infected for longer periods of time had more extensive lesions. Analysis of variance

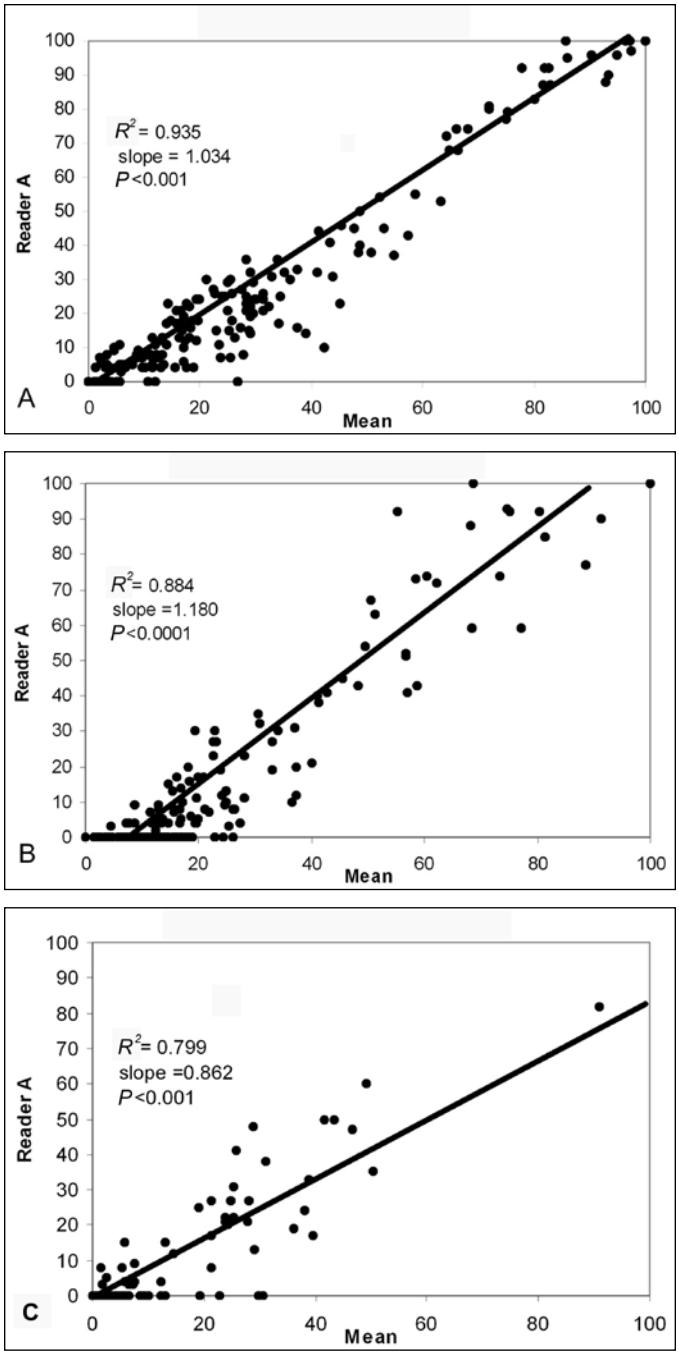

Figure 5. Correlations between mean scores and reader A scores for (A) PMN, (B) gastritis, and (C) metaplasia. Correlation coefficient (R^2), slope of correlation line, and *P* value for the slope are given. Correlation coefficient, slope, and *P* value for all the readers' scores are included in Table 1.

revealed significant differences ($P < 0.05$) among the groups for all 3 histologic criteria, for infected and uninfected mice, and for mice with and without adaptive immune responses. Groups with statistically similar scores were identified by Fisher Least Significant Difference. For this evaluation, the mean scores of 3 groups—uninfected C57BL/6J mice, infected C57BL/6J mice killed 16 wk after inoculation, and infected SCID mice killed 12 wk after adoptive transfer—were chosen (in light of previous studies[7,8]) to represent unaffected, moderately affected, and extensively affected animals, respectively. Groups were categorized as N (no lesions, not statistically different from uninfected C57BL/6 mice), M (moderate lesions, not significantly different from C57BL/6 mice killed 16 wk after inoculation), and E (extensive lesions, not significantly different from recipient SCID mice killed 12 wk after transfer). Groupings are indicated by letter in Figure 6; groups that were statistically identical with more than one of the severity categories (N, M, or E) are not assigned a category in Figure 6. In all histologic categories, uninfected mice and immunodeficient mice (SCID mice with or without *H. pylori*) were in group N (no lesions), and *H. pylori*-infected rSCID mice were in group E (extensive lesions). C57BL/6J mice killed 8 wk after inoculation fell into the N group, consistent with previous findings by our laboratory[7,8] and others[13] that gastritis develops slowly in these mice. Most of the M group animals were infected C57BL/6J mice and uninfected rSCID mice, consistent with previous findings that mild to moderate gastritis develops over several months in infected C57BL/6J mice and rSCID mice develop high background lesions even in the absence of *H. pylori* infection.[6-8]

Table 1. Correlations between mean score and reader score for PMN, gastritis, and metaplasia

Lesion	Reader	R^2	Slope of regression line
PMN			
	A	0.935	1.034
	B	0.845	0.948
	C	0.922	1.018
	D	0.873	0.916
Gastritis			
	A	0.844	1.180
	B	0.542	0.945
	C	0.876	1.026
	D	0.872	1.234
Metaplasia			
	A	0.799	0.862
	B	0.768	1.373
	C	0.650	0.950
	D	0.652	0.776

In all cases, the *P* value for the slope of the regression line was <0.001.

To determine whether individual animals could be identified as uninfected or immunodeficient based on histologic score alone, we used the means and standard deviations for specific groups to set cutoff values for the presence or absence of lesions. Because uninfected C57BL/6J mice are expected to have some background inflammation but should not have lesions attributable to *H. pylori* infection, the mean score of this group was taken as a baseline to indicate no disease. Any value 2 standard deviations higher than the mean of uninfected C57BL/6J mice was taken to indicate a biologically significant lesion. For example, because the PMN score (mean ± 1 standard deviation) for all 22 uninfected C57BL/6J mice was 13.6% ± 14.0%, any value greater than or equal to 42% was considered to be an inflammatory infiltrate of at least moderate extent. Scores less than 42% were defined as background (no significant lesions). To distinguish between moderate and widespread disease groups, we used the mean scores for infected C57BL/6J mice killed 16 wk after inoculation. Because previous work revealed moderate gastritis in these mice,[6-8] scores 2 standard deviations greater than the mean of this group were defined as widespread disease. Table 2 indicates the cutoff values for no

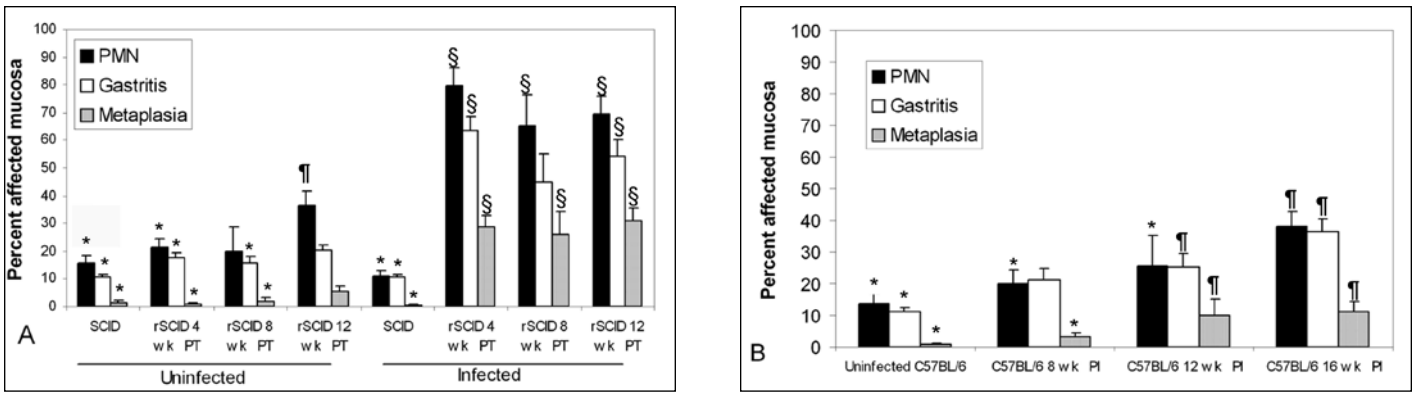

Figure 6. Mean scores for each experimental group, experiment 1. In each group, uninfected mice and SCID mice had lower mean scores for all 3 histologic criteria than did infected and immunocompetent mice. Infected recipient SCID (rSCID) mice had the highest mean scores of all groups. In each histologic category, groups were categorized statistically as having no lesions (N), moderately extensive lesions (M), or extensive lesions (E). * = Group N, ¶ = Group M, § = Group E. In C57BL/6J mice, infection resulted in development of moderate lesions, and adoptive transfer in infected mice led to development of extensive lesions. (A) Infected and uninfected SCID and rSCID mice. (B) Infected and uninfected C57BL/6J mice. PI, postinfection; PT, post-transfer.

Table 2. Cut-off values used to define disease categories for each histologic criterion

	Proportion (%) of affected mucosa used to define disease level		
Lesion	No disease	Disease present	Widespread disease
PMN	<42%	≥43%	>60%
Gastritis	<21%	≥22%	>56%
Metaplasia	<7%	≥8%	>26%

disease, moderate disease, and widespread disease for each of the 3 histologic criteria.

Using these criteria we determined whether presence of moderate or widespread disease correlated with mouse group. For experiment 1, the percentage of each group that fell within each disease category (not present, present, or present and widespread) is indicated in Table 3. Like Figure 6, Table 3 indicates that among uninfected and infected SCID mice generally fewer than 10% of the animals had biologically significant lesions. Some uninfected rSCID mice had lesions in some categories, reaching up to 50% in the case of gastritis in mice killed 12 wk after inoculation; however, none of these mice had widespread lesions. As described earlier, *H. pylori*-infected rSCID mice formed a group in which at least 80% of mice had lesions, and most had widespread lesions. The other groups (uninfected rSCID mice and infected C57BL/6J mice) varied in the percentage of animals with lesions, consistent with previous findings indicating that inflammation is moderate and variable under these conditions.[7,8]

Combining scores from all 3 histologic lesions resulted in a clear division into 3 groups. One group consisted of the uninfected C57BL/6J mice, nonrecipient SCID mice (infected or not), and uninfected rSCID mice killed 4 or 8 wk after inoculation: none of these mice had lesions in any of the 3 categories (Table 3). *H. pylori*-infected rSCID mice, which had the highest percentage of animals with lesions in all 3 categories (more than 80%) formed a second group. Infected C57BL/6J and uninfected rSCID groups killed 12 wk after transfer formed a third group with variable lesions. Between 10% and 50% of the mice in the second 2 groups had lesions in all 3 histologic categories.

To further evaluate the biologic relevance of the scoring system, we applied it to a second study. In experiment 2, a single reader (reader A) scored 314 mice in various infection groups. The 2 studies differed in that 1) experiment 2 was scored by a single investigator, and 2) the experimental mouse groups differed somewhat. Experiment 2 included chronically infected C57BL/6J mice (killed 6 mo and 1 y after inoculation), and some of the groups examined in experiment 1 were not included (Figure 6). Figure 7 shows that the range of scores per group in experiment 2 were similar to those of experiment 1. Uninfected and immunodeficient mice had the lowest scores, and *H. pylori*-infected rSCID mice had the highest scores. In experiment 2, the timing of euthanasia differed slightly from that in experiment 1, so that N, M, and E designations were based on uninfected C57BL/6J mice, C57BL/6J mice killed 12 wk after inoculation, and *H. pylori*-infected rSCID mice killed 8 wk after transfer, respectively. By use of the same cutoff values for presence and extent of disease as in experiment 1, experiment 2 revealed similar results in that disease largely was not present in uninfected and immunodeficient mice, and mice with widespread lesions were mostly *H. pylori*-infected rSCID mice (Table 4). Infected C57BL/6J mice killed 6 mo or 1 y postinoculation also fell into the group with extensive lesions, consistent with previous findings of progressive gastritis in these mice.[7,8,13]

Discussion

The results of the present study demonstrate that the scoring system we described is reproducible between readers and experiments and accurately reflects biologically significant changes in a mouse model of *H. pylori*-associated gastritis. These findings are important for 2 reasons. First, the recent rapid rise of the use of mice and other rodents in the study of disease due to *H. pylori* has been accompanied by a large number of different scoring systems for histologic lesions,[3,9,12,14-16,19] leading to confusion and difficulty when comparing results between laboratories. Second, because some of these systems are based on lesions in human tissues[9,12] or have not been examined for biologic relevance, they are difficult to interpret.

Quantitation of gastric lesions due to naturally occurring *H. pylori*-associated disease in humans is most commonly done subjectively by means of the Updated Sydney System.[4] In this system a specified number of gastric biopsies from specific sites in the stomach are examined and scored based on a published visual

Table 3. Percentage of mice in each group with lesions not present, present, or present and widespread—experiment 1

Group	N	PMN			Gastritis			Metaplasia			All 3 lesions present
		None	Present	Widespread	None	Present	Widespread	None	Present	Widespread	
Uninfected and SCID											
Uninfected C57BL/6J	22	95	5	0	95	5	0	95	5	0	0
Uninfected SCID	22	95	5	0	100	0	0	100	0	0	0
Infected SCID	32	100	0	0	100	0	0	100	0	0	0
Uninfected recipient SCID											
Uninfected recipient SCID 4 wk PT	15	87	13	0	60	40	0	100	0	0	0
Uninfected recipient SCID 8 wk PT	5	80	20	0	100	0	0	100	0	0	0
Uninfected recipient SCID 12 wk PT	6	67	33	0	50	50	0	67	33	0	17
Infected C57BL/6J											
Infected C57BL/6J 8 wk PI	10	90	10	0	60	40	0	90	10	0	10
Infected C57BL/6J 12 wk PI	6	83	17	17	50	50	0	67	33	17	17
Infected C57BL/6J 16 wk PI	6	50	50	50	17	83	0	33	67	67	50
Infected recipient SCID											
Infected recipient SCID 4 wk PT	10	10	90	90	0	100	70	10	90	50	90
Infected recipient SCID 8 wk PT	5	20	80	80	20	80	40	20	80	40	80
Infected recipient SCID 12 wk PT	7	14	86	86	0	100	57	0	100	71	83

PI, postinfection; PT, post-transfer.

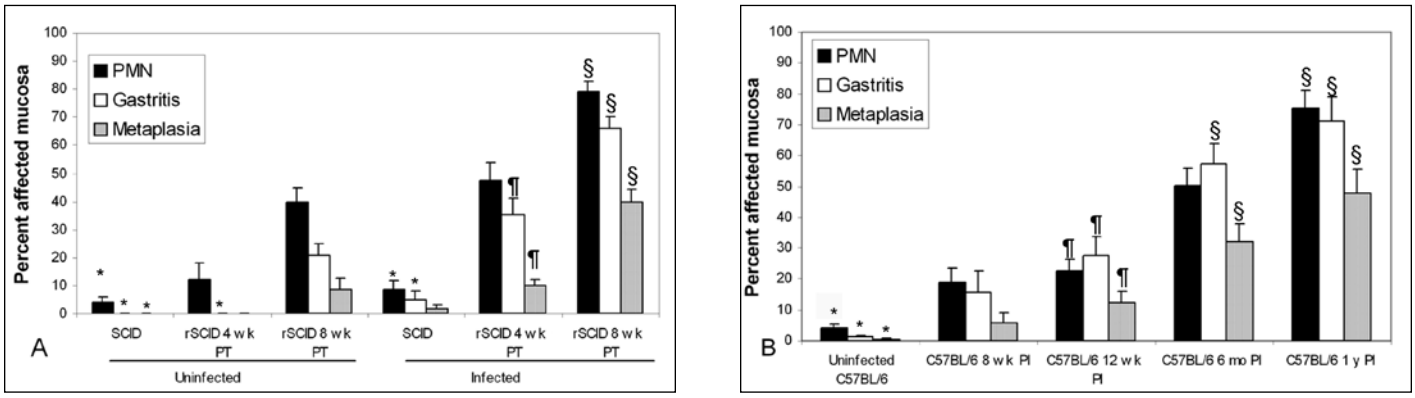

Figure 7. Mean scores for each experimental group, experiment 2. As in experiment 1, uninfected mice and SCID mice had the lowest scores, and H. pylori-infected recipient SCID (rSCID) mice had the highest scores. In each histologic category, groups were categorized statistically as having no lesions (N), moderately extensive lesions (M), or extensive lesions (E). * = Group N, ¶ = Group M, § = Group E (A) Infected and uninfected SCID and recipient SCID mice. (B) Infected and uninfected C57BL/6J mice. PI, postinfection; PT, post-transfer.

analog scale, which is an artist's depiction of inflammatory infiltrate, bacteria, and epithelial changes (metaplasia and atrophy). This system is widely used for both diagnostic and investigational purposes; however, it is influenced by observer subjectivity, and its reproducibility has been questioned.[1,2,10] The Sydney system occasionally is adapted for use with animal models,[17,20-22] but it has not been validated for use in animals and probably is not applicable because 1) it is based on examination of gastric biopsies from specific anatomic sites (typical of clinical material from human patients) rather than large regions of mucosa (typical of animal tissues from experimental studies), and 2) lesions in animals may differ from those in humans. For this reason, other semiquantitative methods have been developed by many laboratories.[3,9,12,14-16,19] All of these are based on quantitation of subjective pathology scores, and they vary markedly, hindering attempts to compare lesions between studies. In addition, with few exceptions,[21] the relevance of the resulting scores to disease status in the animals has not been evaluated. Therefore, a standardized protocol for scoring of gastric lesions in animal model studies of *H. pylori*-associated disease could facilitate comparisons between published studies.

Our study illustrates a general approach to developing and evaluating a system for quantifying histologic lesions in a tissue that does not easily support morphometric evaluation. Traditional morphometric methods involve precise quantification of tissues that have been carefully prepared to minimize variability between specimens and use cell counts or computer-based measurements. Even these approaches include variability due to specimen preparation and require some amount of observer variation, at least in the selection of microscopic fields or lesions to quantify.[11,18] The recent explosion in the use of animal models, particularly mouse models, of enteric disease has led to a need for improved methods of analysis of morphologic findings. The experiments we report here indicate that such scoring systems can be developed. We sug-

Table 4. Percentage of mice in each group with lesions not present, present, or present and widespread—experiment 2

Group	N	PMN			Gastritis			Metaplasia			All 3 lesions present
		None	Present	Widespread	None	Present	Widespread	None	Present	Widespread	
Uninfected and SCID											
Uninfected C57BL/6J	77	99	1	0	97	3	0	99	1	0	0
Uninfected SCID	25	100	0	0	100	0	0	100	0	0	0
Infected SCID	28	92	4	4	93	7	7	93	7	4	4
Uninfected recipient SCID											
Uninfected recipient SCID 4 wk PT	5	100	0	0	100	0	0	100	0	0	0
Uninfected recipient SCID 8 wk PT	40	57	43	30	70	30	10	80	20	8	15
Infected C57BL/6J											
Infected C57BL/6J 8 wk PI	8	100	0	0	75	25	13	75	25	0	0
Infected C57BL/6J 12 wk PI	18	89	11	0	50	50	17	50	50	22	11
Infected C57BL/6J 6 mo PI	10	20	80	30	0	100	40	10	90	60	50
Infected C57BL/6J 1 y PI	13	8	92	77	15	85	85	8	92	92	85
Infected recipient SCID											
Infected recipient SCID 4 wk PT	29	48	52	38	45	55	31	55	45	7	28
Infected recipient SCID 8 wk PT	61	13	87	80	12	88	63	20	80	59	59

PI, postinfection; PT, post-transfer.

gest that similar systems for other enteric disease models could be devised based on the principles described in this study.

Our data indicate that scoring of the percentage of affected fields rather than subjectively estimating severity of infiltrate results in both good agreement among readers and good correlation with treatment group. However, the degree of agreement between readers varied for the different histologic criteria. The highest level of agreement was in evaluation of neutrophilic infiltration, perhaps because this criterion was the best defined and thus easiest to score. The explanation for the marked difference in gastritis scores between reader B and the other 3 readers was that reader B failed to observe the requirement for gland displacement as a prerequisite for a positive gastritis score. Further investigation after the slides were scored revealed that this reader had scored fields with any inflammatory cells at all as positive for gastritis. In retrospect, the definition of gastritis as requiring gland displacement should have been stressed prior to the study. Eliminating the scores from reader B resulted in high agreement among the other three scorers, indicating that scoring of gastritis was accurate when the criteria were made clear to the readers.

Metaplasia scores were the most variable between readers. The reason for this variability is unclear. In the initial discussions to establish scoring criteria, metaplasia was the most difficult criterion to define, perhaps confusing consensus opinion as to what constituted a metaplastic gland. Diagnosis of metaplasia thus may have required some level of subjective judgment, such that variation in the experience or individual biases of the 4 readers led to differences in diagnosis. Alternatively, perhaps metaplasia was difficult to recognize when only a few metaplastic glands were present, and readers simply scored these fields differently. Recognition of metaplastic glands appeared to be more difficult than recognition of inflammatory infiltrate in tissue, and the use of metaplasia as a scoring criterion for gastritis may or may not be useful in comparisons between laboratories.

Comparison of group means in both experiment 1 and experiment 2 revealed significant group differences associated with immune or infection status. Uninfected and immunodeficient groups in all cases had mean scores that were significantly lower than comparable infected or immunocompetent groups. In addition to group mean scores, the individual animal scores in both experiment 1 and experiment 2 reliably identified animals in the different treatment groups. Uninfected animals and animals without an adaptive immune response had scores that fell below a specified value for PMN, gastritis, and metaplasia. Similarly, animals with lesions above a defined value for PMN, gastritis, and metaplasia were with few exceptions in either C56BL/6 mice infected for 6 mo or more or infected rSCID mice. These results are compatible with those of previous studies that demonstrate minimal gastritis in uninfected mice; moderate, slowly progressive gastritis in *H. pylori*-infected C57BL/6J mice; and severe, rapidly progressive gastritis in *H. pylori*-infected rSCID mice.[7,8,13]

In summary, we have described a scoring system for *H. pylori*-associated gastritis in mice that is reproducible between readers and accurately reflects the disease status of the mice. These data demonstrate that 1) this method is biologically and statistically valid for comparison of *H. pylori*-infected mice and 2) histologic lesions can be quantified reproducibly in conventionally prepared sections of mouse stomach. Widespread use of this system to score histologic lesions in rodent stomachs and similar systems to score gastrointestinal lesions in other animal models of enteric disease could improve agreement among laboratories and facilitate comparison of results.

Acknowledgment

This work was supported in part by Public Health Service grant R01AI043643 from the National Institutes of Health.

References

1. **Aydin O, Egilmez R, Karabacak T, Kanik A.** 2003. Interobserver variation in histopathological assessment of *Helicobacter pylori* gastritis. World J Gastroenterol **9:**2232–2235.

2. **Chen XY, Vanderhulst RWM, Bruno MJ, Vanderende A, Xiao SD, Tytgat GNJ, Tenkate FJW.** 1999. Interobserver variation in the histopathological scoring of *Helicobacter pylori* related gastritis. J Clin Pathol **52:**612–615.
3. **Court M, Robinson PA, Dixon MF, Crabtree JE.** 2002. Gastric Helicobacter species infection in murine and gerbil models: comparative analysis of effects of *H. pylor*i and *H. felis* on gastric epithelial cell proliferation. J Infect Dis **186:**1348–1352.
4. **Dixon MF, Genta RM, Yardley JH, Correa P.** 1996. Classification and grading of gastritis. The updated Sydney system. International workshop on the histopathology of gastritis, Houston 1994. Am J Surg Pathol **20:**1161–1181.
5. **Eaton KA, Logan SM, Baker PE, Peterson RA, Monteiro MA, Altman E.** 2004. *Helicobacter pylori* with a truncated lipopolysaccharide O chain fails to induce gastritis in SCID mice injected with splenocytes from wild-type C57BL/6J mice. Infect Immun **72:**3925–3931.
6. **Eaton KA, Mefford M, Thevenot T.** 2001. The role of T cell subsets and cytokines in the pathogenesis of *Helicobacter pylori* gastritis in mice. J Immunol **166:**7456–7461.
7. **Eaton KA, Mefford ME.** 2001. Cure of *Helicobacter pylori* infection and resolution of gastritis by adoptive transfer of splenocytes in mice. Infect Immun **69:**1025–1031.
8. **Eaton KA, Ringler SR, Danon SJ.** 1999. Murine splenocytes induce severe gastritis and delayed-type hypersensitivity and suppress bacterial colonization in *Helicobacter pylori*-infected SCID mice. Infect Immun **67:**4594–4602.
9. **Ghiara P, Marchetti M, Blaser MJ, Tummuru MKR, Cover TL, Segal ED, Tompkins LS, Rappuoli R.** 1995. Role of the *Helicobacter pylori* virulence factors vacuolating cytotoxin, CagA, and urease in a mouse model of disease. Infect Immun **63:**4154–4160.
10. **Guarner J, Herrera-Goepfert R, Mohar A, Sanchez L, Halperin D, Ley C, Parsonnet J.** 1999. Interobserver variability in application of the revised Sydney classification for gastritis. Hum Pathol **30:**1431–1434.
11. **Jarvis LR, Whitehead R.** 1985. Morphometric analysis of gastric dysplasia. J Pathol **147:**133–138.
12. **Lee A, Chen M, Coltro N, O'Rourke J, Hazell S, Hu P, Li Y.** 1993. Long term infection of the gastric mucosa with Helicobacter species does induce atrophic gastritis in an animal model of *Helicobacter pylori* infection. Int J Med Microbiol **280:**38–50.
13. **Lee A, Orourke J, Deungria MC, Robertson B, Daskalopoulos G, Dixon MF.** 1997. A standardized mouse model of *Helicobacter pylori* infection: introducing the Sydney strain. Gastroenterology **112:**1386–1397.
14. **Marchetti M, Arico B, Burroni D, Figura N, Rappuoli R, Ghiara P.** 1995. Development of a mouse model of *Helicobacter pylori* infection that mimics human disease. Science **267:**1655–1658.
15. **Mohammadi M, Czinn S, Redline R, Nedrud J.** 1996. *Helicobacter*-specific cell-mediated immune responses display a predominant Th1 phenotype and promote a delayed-type hypersensitivity response in the stomachs of mice. J Immunol **156:**4729–4738.
16. **Mohammadi M, Redline R, Nedrud J, Czinn S.** 1996. Role of the host in pathogenesis of *Helicobacter*-associated gastritis: *H. felis* infection of inbred and congenic mouse strains. Infect Immun **64:**238–245.
17. **Nagata J, Kijima H, Takagi A, Ito M, Goto K, Yamazaki H, Nakamura M, Mine T, Ueyama Y.** 2004. *Helicobacter pylori* induces chronic active gastritis in p53-knockout mice. Int J Mol Med **13:**773–777.
18. **Ruiz B, Garay J, Correa P, Fontham ET, Bravo JC, Bravo LE, Realpe JL, Mera R.** 2001. Morphometric evaluation of gastric antral atrophy: improvement after cure of *Helicobacter pylori* infection. Am J Gastroenterol **96:**3281–3287.
19. **Sakagami T, Dixon M, O'Rourke J, Howlett R, Alderuccio F, Vella J, Shimoyama T, Lee A.** 1996. Atrophic gastric changes in both *Helicobacter felis* and *Helicobacter pylori* infected mice are host-dependent and separate from antral gastritis. Gut **39:**639–648.
20. **Suzuki H, Mori M, Seto K, Kai AM, Kawaguchi C, Suzuki M, Suematsu M, Yoneta T, Miura S, Ishii H.** 1999. *Helicobacter pylori*-associated gastric pro- and antioxidant formation in Mongolian gerbils. Free Rad Biol Med **26:**679–684.
21. **Walker MM, Worku M, Coggle S, Thursz MR.** 2002. A novel method for assessing gastritis in the murine model demonstrates genetically determined variation in response to *Helicobacter felis* infection. Helicobacter **7:**265–268.
22. **Xiao F, Furuta T, Takashima M, Shirai N, Hanai H.** 2001. Effects of cyclooxygenase-2 inhibitor on gastric acid secretion in *Helicobacter pylori*-infected C57BL/6 mice. Scand J Gastroenterol **36:**577–583.