

Title: Crohn's patients and healthy infants share immunodominant B cell response to commensal flagellin peptide epitopes

Authors:

Qing Zhao^{1,*}, Lennard Wayne Duck¹, John T. Killian, Jr.², Alexander F. Rosenberg^{3,4}, Peter J. Mannon⁵, R. Glenn King³, Lee A. Denson⁶, Subra Kugathasan⁷, Edward N. Janoff⁸, Maria C. Jenmalm⁹, and Charles O. Elson^{1,*}

Affiliations:

¹Department of Medicine, University of Alabama at Birmingham; Birmingham, USA.

²Department of Surgery, University of Alabama at Birmingham; Birmingham, USA.

³Department of Microbiology, University of Alabama at Birmingham; Birmingham, USA.

⁴Informatics Institute, University of Alabama at Birmingham; Birmingham, USA.

⁵Department of Internal Medicine, University of Nebraska Medical Center; Omaha, USA.

⁶Schubert-Martin Inflammatory Bowel Disease Center, Department of Pediatrics, University of Cincinnati; Cincinnati, USA.

⁷Division of Pediatric Gastroenterology, Emory University School of Medicine, Children's Healthcare of Atlanta; Atlanta, USA.

⁸Department of Medicine, University of Colorado Denver, Denver Veterans Affairs Medical Center; Aurora, USA.

⁹Department of Biomedical and Clinical Sciences, Linköping University; Linköping, Sweden.

*Corresponding author.

Qing Zhao: qingzhao@uabmc.edu

Charles O. Elson: coelson@uabmc.edu

Abstract:

About half of patients with Crohn's disease (CD) develop selective serum IgG response to flagellin proteins of the *Lachnospiraceae* family. Here, we identified a dominant B cell peptide epitope in CD, locating in the highly conserved "hinge region" between the D0 and D1 domains at the amino-terminus of *Lachnospiraceae* flagellins. Serum IgG reactive to this epitope is present at an elevated level in adult CD patients and in pediatric CD patients at diagnosis. Most importantly, high levels of serum IgG to the hinge epitope were found in most infants from 3 different geographic regions (Uganda, Sweden, and the USA) at one year of age. This vigorous homeostatic response decrements with age as it is not present in healthy adults. These data identify a distinct subset of CD patients, united by a shared reactivity to this dominant flagellin epitope that may represent failure of a homeostatic response beginning in infancy.

INTRODUCTION

Inflammatory bowel disease (IBD) is characterized as chronic inflammation in the gut mucosa, with Crohn's disease (CD) and ulcerative colitis (UC) as the two main forms. It is generally accepted that aberrant immune responses to environmental factors, especially to the gut microbiota, play a critical role in the pathogenesis of IBD(1-3). Although CD and UC share various genetic polymorphisms(4), dominant antigens identified in CD are considerably different from that in UC(5-7). Flagellins expressed by commensal *Lachnospiraceae* bacteria and CD4⁺ T cells specific to these flagellins are found in blood and intestinal lesions of Crohn's patients(8, 9), suggesting that *Lachnospiraceae* flagellin is a keystone antigen driving gut inflammation in at least a subset of Crohn's patients. Serum IgG reactivity to CBir1, a *Lachnospiraceae* flagellin isolated from mice, is elevated and associated with disease complications in patients with CD, but not in patients with UC or healthy controls (HC)(6, 8). In addition, anti-flagellin IgG reactivity was shown to predict the development of CD years prior to disease onset(10-13). Therefore, sero-reactivity to *Lachnospiraceae* flagellins has been used as a biomarker for CD diagnosis. However, although *Lachnospiraceae* flagellin has been identified as an immunodominant antigen in CD at the protein level, the dominant B cell peptide epitopes and where they are located remain unclear.

The motility of various commensal and pathogenic bacteria, especially those that belong to the Proteobacteria and Firmicutes (reclassified as Bacillota now) phyla, is driven by their surface expression of flagella, which can consist of as many as 30,000 flagellin monomers(14). Although flagellin proteins have a vast span of diversity among different organisms, they share a characteristic molecular structure consisting of amino- and carboxyl- domains (D0 and D1) that are conserved within a given bacterial family(15, 16). The amino- and carboxyl- domains are separated by a hypervariable region (D2 and D3 domains). In addition to conferring motility, flagellin is also one of the most potent immune activators known in the microbiota(17-21). As a result, flagellin proteins expressed by intestinal pathogens such as *Salmonella* spp. and *Escherichia coli* (*E. coli*) serve as major constituents mediating intestinal infection and inflammation(22, 23). However, immune reactivity targeting flagellins expressed by commensal bacteria, especially those belong to the *Lachnospiraceae* family, is only elevated in patients with CD, but not in patients with UC or adult HC, despite the presence of abundant intestinal *Lachnospiraceae* bacteria in the latter(24).

Previously, using a microbiota protein antigen microarray, we have shown that ~30% of CD patients in our regional IBD cohort recruited at the University of Alabama at Birmingham (UAB) had elevated serum IgG reactivity to more than 10 different flagellins expressed by *Lachnospiraceae* bacteria (from both human and mouse origin), and reactivity to individual flagellins was highly correlated to each other(24), suggesting that serum IgG response to *Lachnospiraceae* flagellins was driven by reactivity to shared epitopes, which remained unidentified. Consistently, these multi-flagellin reactive patients had significantly increased flagellin-specific effector/memory CD4⁺ T cells in the circulation(24), which could travel back to the gut and drive colitis (25-27). We proposed that patients with multi-flagellin reactivity are more likely to develop disease complications in CD, and early identification of these patients through reactivity to the shared epitopes would greatly help with their disease treatment.

Here, using a novel flagellin peptide microarray and a flagellin peptide cytometric bead array assay, we report the identification of a key IgG B cell epitope in *Lachnospiraceae* flagellins in CD patients, which locates at the “hinge region” between the D0 and D1 domain in the highly conserved amino-terminus (N-term) of *Lachnospiraceae* flagellins. Sero-reactivity to the dominant epitope was identified in our regional adult CD patients, and in a multi-site-recruited cohort of pediatric CD patients. In addition, the magnitude of this epitope-specific reactivity was positively correlated with multi-flagellin reactivity in CD patients, and predicted the development of a complicated disease course in pediatric CD patients 2-3 years in advance. Interestingly, we observed a robust level of serum IgG response to this dominant epitope in over 90% of infants of diverse geographic origins, which peaked at one year of age, indicating that the elevated IgG reactivity to *Lachnospiraceae* flagellins in CD patients may result from an aberrant immune recall response of sensitization that occurred in early life.

RESULTS

Cohort description

Three cohorts were used in this study, including a discovery cohort of adult IBD patients and HC recruited at UAB(24), a validation cohort of treatment-naïve pediatric CD patients and non-IBD controls collected at 28 sites in the USA and Canada (the RISK Stratification Study)(28), and a healthy and non-IBD control cohort of infants and their mothers recruited in Uganda, the USA(29), and Sweden(30).

Lachnospiraceae flagellin peptide microarray identifies key IgG binding epitopes in CD patients

With our previous microbiota protein microarray, Pearson correlation analysis showed that serum IgG response to most *Lachnospiraceae* flagellins strongly correlated with each other in CD patients(24). This observation suggested that serum IgG reactivity in CD patients was targeting shared epitopes among different *Lachnospiraceae* flagellins. To test this hypothesis, we generated a novel flagellin peptide microarray consisting of 1512 peptides derived from 19 different *Lachnospiraceae* flagellins of both human and mouse origin. These were sequential 15mer peptides overlapping by 5 amino acids, which covered the entire sequences of 19 flagellin proteins. Sera from CD patients, HC, and UC patients were probed against the array, and peptide-specific IgG or IgA reactivities were detected with fluorescently labeled secondary antibodies (**Figure 1A**). In the heatmap where serum IgG reactivity to peptides in the D0 and D1 domains of the N-term of *Lachnospiraceae* flagellins is shown (**Figure 1B**), those in CD patients clustered apart from those in UC patients and HC, whereas the response to *Lachnospiraceae* flagellin peptides in UC and HC subjects was not different. Notably, reactivities from CD sera fell in two major clusters, with increased IgG binding identified at p26-41 and p41-61 between the D0 and D1 domains in the amino chain of *Lachnospiraceae* flagellins (**Figure 1B and 1C**), which was not seen in UC or HC subjects. There was no significant difference seen in serum IgG reactivity among CD, UC, and HC subjects against other peptides included in the microarray (**Supplementary Figure 1**). Differences in the IgA response to *Lachnospiraceae* flagellin peptides among groups were not observed. Sequence analysis and motif discovery analysis revealed that, regardless of human or mouse origin, the majority of the 19 *Lachnospiraceae* flagellins included in the peptide microarray have

significant amino acid homology in both the N-term and C-term of their D0 and D1 domains. Specifically, such homology peaks at the “hinge region” between the D0 and D1 domains at the N-term (**Figure 1D and 1E**), which is significantly different from flagellins expressed by Proteobacteria *Salmonella dublin* (*S. dublin*) and *E. coli* (**Supplementary Table 1**), although the latter two are homologous to each other. Multiple likely B cell epitopes were predicted in this hinge region when *Bepipred Linear Epitope Prediction 2.0* was performed through Immune Epitope Database(31), consistent with this region being highly immunogenic (**Supplementary Table 2**). The dominant B cell epitope in *Lachnospiraceae* flagellins identified in CD patients with the flagellin peptide microarray locates right at this “hinge region”. It includes sub-epitopes p25-44, which locates at the D0 domain, and p41-59 which locates at the D1 domain. Hereafter, we will refer to the dominant *Lachnospiraceae* flagellin B cell epitope in CD, D0-1N p25-59, as the “hinge peptide”.

Serum IgG multi-flagellin reactivity in CD patients is driven by reactivity to the dominant *Lachnospiraceae* flagellin peptide epitope

In order to confirm the findings of the flagellin peptide microarray, we developed a flagellin peptide cytometric bead array, where serum IgG, IgA, and IgM reactivity to different bead-bound antigens can be simultaneously detected in a multiplexed fashion (**Figure 2A**). Eight biotinylated conserved peptides (D0N p1-19, D0N p25-44, D1N p41-59, D0-1N p25-59, D1N p74-102, D1C p391-407, D0-1C p410-435, D0C p437-460) covering the N-term and C-term of D0 and D1 domains of *Lachnospiraceae* flagellins were synthesized and individually conjugated to 4-micron fluorescent beads (**Figure 2B and 2C**). Beads conjugated with different peptides can be differentiated by their fluorescent intensity in the APC/Cy7 channel with flow cytometry (**Figure 2C**). Serum samples (UAB cohort) from CD patients with multi-flagellin reactivity on the microbiota protein microarray (≥ 10 different flagellins, CD_{high}), CD patients who reacted to less than 10 or no *Lachnospiraceae* flagellins (CD_{low}), UC patients and HC were assayed and IgG, IgA, and IgM reactivity to individual peptides were detected with fluorescently labeled secondary antibodies. Standard curves using human IgA, IgM, and IgG capture beads and known concentrations of polyclonal human IgA, IgM, and IgG were obtained and applied to quantify peptide-specific antibodies, allowing cross-comparison of samples run at different times. Data from this cytometric bead array assay strongly validated the results we obtained from the flagellin

peptide microarray (**Figure 2D**). IgG reactivity to the “hinge peptide” (D0-1N p25-59) is highly represented in CD_{high} patients compared to CD_{low}, UC, and HC (**Figure 2D-F**). In addition, CD_{high} patients have significantly elevated serum IgG specific to the sub-epitopes (D0N p25-44 and D1N p41-59) compared to CD_{low} patients, UC patients, and HC (**Figure 2G and 2H**), but not to other conserved peptides of the *Lachnospiraceae* flagellins including D0N p1-19, D1N p74-102, D1C p391-407, and D0-1C p410-435 (**Figure 2I-K, and 2M**). These data indicate that multi-flagellin IgG reactivity in Crohn’s patients is driven by reactivity to the hinge region of the N-terminus. Of note, CD_{high} patients also have significantly elevated serum IgG reactivity to D0C p437-460 (**Figure 2L**), which is the terminal conserved region of *Lachnospiraceae* flagellin C-terminus, although the magnitude of reactivity is much lower than that to the N-term hinge peptide. Serum IgA and IgM responses to individual conserved peptides were also analyzed, but consistent with the microbiota protein microarray and flagellin peptide microarray data, no differential reactivity among groups was observed.

***Lachnospiraceae* hinge peptide is the dominant B cell epitope in pediatric CD patients and predicts the development of disease complications**

Although we were able to confidently identify the dominant B cell peptide epitope in CD patients using our IBD cohort recruited at UAB, there were several limitations involved. First, it is a regional adult cohort of IBD, which might not represent a broader IBD population. Also, the refractory nature of this cohort (rate of biologic therapy in CD was 78% and rate of surgery in CD was 66%)(24) limited the power to analyze the association between the epitope-specific IgG sero-response and disease complications in CD. Therefore, we decided to validate this finding using a multi-site-recruited pediatric cohort consisting of patients who were newly diagnosed with CD as well as non-IBD controls (the RISK Stratification Study)(28). CD patients in this cohort represented the variety of disease phenotypes regarding disease location, severity, and behavior, and they were recruited at diagnosis prior to any treatment. Furthermore, patients were followed up to 36 months after the initial sample collection, and their disease course was well-documented, thus the prospective nature of this cohort enabled us to test whether serum IgG to the dominant epitope in CD patients was able to predict the development of disease complications later. In specific, we referred to the Montreal classification and defined developing stricturing (B2) or penetrating (B3) disease behavior as a complicated disease course.

To confirm that pediatric CD patients in the RISK cohort are reactive to *Lachnospiraceae* flagellins, we first performed the microbiota protein microarray as previously described(24). Sera at recruitment of CD patients who remained inflammatory (B1 behavior) at 3-year follow-up, CD patients who developed strictures or penetrating disease (B2/3 behavior) at 3-year follow-up, and non-IBD controls were probed with 19 *Lachnospiraceae* flagellins of human and mouse origin, in addition to *S. dublin* FliC and *E. coli* FliC. Previous study using this cohort has reported an augmented serum IgG specific to *L. CBir1* Fla in 35% of CD patients who remained B1 at follow-up, and in 62% of CD patients who developed B2/3 at follow-up(28). Consistent with this data, our results showed that CD patients that remained B1 exhibited a trend of increased, although not significant, serum IgG reactivity to *Lachnospiraceae* flagellins when compared to control subjects (**Figure 3A**). In contrast, CD patients that developed B2/3 at follow-up showed significantly elevated serum IgG reactivity to the majority of *Lachnospiraceae* flagellins compared to controls and to patients who remained B1 (**Figure 3A and Supplementary Figure 2**). Consistent with what we observed in our adult CD cohort recruited at UAB, a subset of pediatric CD patients also exhibited serum IgG reactivity to the majority of *Lachnospiraceae* flagellins tested (**Figure 3B**), and the proportion of patients that were multi-flagellin reactive was significantly higher in those who developed B2/3 at follow-up (**Figure 3C and 3D**).

Next, we tested whether the *Lachnospiraceae* flagellin hinge peptide also served as the dominant B cell epitope in CD patients in the RISK cohort by performing the flagellin peptide microarray with sera of the multi-flagellin reactive CD patients. Similar to what we observed in our adult IBD cohort recruited at UAB, serum IgG response in multi-flagellin reactive CD patients in the RISK cohort mainly targeted p26-41 and p41-61 between the D0 and D1 domains in the N-term of *Lachnospiraceae* flagellins (**Supplementary Figure 3**). Finally, patient and control sera from the whole cohort were assayed with the flagellin peptide cytometric bead array. Similar to adult CD patients, the dominant serum IgG binding sites in pediatric CD patients located at the conserved *Lachnospiraceae* flagellin hinge region between D0 and D1 domains in the N-term (D0-1N p25-59, D0N p25-44, and D1N p41-59, **Figure 3E-G**). IgG sero-reactivity to other peptides was either not significantly different among groups (D1N p74-102 and D1C p391-407, **Supplementary Figure 4**) or at a much lesser magnitude compared to the hinge peptide (D0-1C p410-435, D0C p437-460, and D0N p1-19, **Supplementary Figure 4**). Moreover, response to the N-term hinge peptide in patients who developed B2/3 at follow-up was significantly higher than

that of patients who remained B1 at follow-up, indicating the prognostic value of hinge peptide specific serum IgG in CD patients early on, even at initial disease diagnosis. In conclusion, the dominant B cell peptide epitope in *Lachnospiraceae* flagellins in CD patients were validated with both the flagellin peptide microarray and flagellin peptide cytometric bead array, using the RISK cohort which represents a broad variety of disease phenotypes and behaviors. More importantly, results from the flagellin peptide cytometric bead array demonstrated that elevated serum IgG to *Lachnospiraceae* flagellin hinge peptide was associated with stricturing or penetrating disease at 3-year follow-up, thus can be used to predict the development of disease complications in CD.

Serum IgG to *Lachnospiraceae* hinge peptide strongly correlates with multi-flagellin reactivity in CD

Next, we sought to determine whether serum IgG reactivity to the conserved peptides included in the flagellin peptide cytometric bead array correlates with reactivities to *Lachnospiraceae* flagellins included in the microbiota protein microarray. First, we performed Pearson correlation analysis between serum IgG reactivity against individual flagellins and serum IgG reactivity specific to individual peptide epitopes for CD, HC, and UC subjects from the UAB cohort and CD patients from the RISK cohort (**Figure 4A**). The most significant correlations with serum reactivities specific to *Lachnospiraceae* flagellin proteins were observed in serum reactivities against D0N p25-44, D1N p41-59, and D0-1N p25-59, i.e. the hinge peptide, especially in CD patients from the UAB cohort and the RISK cohort. Sero-reactivities to the hinge peptide did not universally correlate with all flagellins tested, in that the correlation with that against Proteobacteria flagellins such as *E. coli* FliC and *S. dublin* FliC was significantly reduced. Moreover, sero-reactivity to the C-term conserved epitopes D0-1C p410-435 and D0C p437-460 showed slight significance in correlating with that to *Lachnospiraceae* flagellin proteins in RISK cohort CD patients. To better view the correlation in each subject, we plotted individual sero-reactivity against the most significant peptide, D0-1N p25-59, and one of the least significant peptides, D1C p391-407, with that of *R. intestinalis* Fla1, and *E. coli* FliC, respectively (**Figure 4B**). Reactivity to D0-1N p25-59, especially in CD patients from both UAB and the RISK cohort correlated significantly with that of *R. intestinalis* Fla1, which was expressed by bacteria in the *Lachnospiraceae* family, but poorly with *E. coli* FliC. In contrast, reactivity to D1C p391-407 did not correlate with that of *R. intestinalis* Fla1 in any subject group, demonstrating that reactivity to

the hinge peptide, but not to other conserved flagellin peptides, underlies sero-reactivity to *Lachnospiraceae* flagellins. Furthermore, a good concordance was found between sero-reactivity to the hinge peptide obtained from the flagellin peptide cytometric bead array and flagellin seropositivity determined by the microbiota protein microarray in both the UAB cohort and the RISK cohort (**Figure 4C and 4D**). This further demonstrates that multi-flagellin reactivity in CD patients is largely due to IgG binding to the dominant hinge peptide epitope.

Serum IgG reactivity to *Lachnospiraceae* flagellin hinge peptide is a homeostatic response in most infants

Our previous study showed a surprisingly similar level of serum IgG reactivity to Firmicutes flagellins in healthy and allergic Swedish infants at 2 years of age compared to that in adult CD patients(30). We wondered whether this high level of flagellin reactivity existed universally in infants raised in distinct environments. To test this hypothesis, sera were obtained from HIV-exposed but uninfected healthy infants from Uganda, healthy and allergic infants from Sweden, and healthy infants from the USA at 6 and 12 months of age, as well as their cord blood representing antibody reactivity of the mothers. Data from the Swedish and the American infants were grouped together, representing reactivity from developed countries, whereas data from the Ugandan infants represented reactivity in developing countries.

Serum samples were probed with the flagellin peptide cytometric bead array and the level of IgG reactivity to the hinge peptide in each individual at different time points was linked and compared to that of CD patients in both the UAB cohort and RISK cohort. Hinge peptide specific IgG response in the cord blood of both Uganda and US/Sweden cohorts was low and comparable to that identified in adult HCs from the UAB cohort (**Figure 5A**). Consistent with our previous microbiota protein microarray data(30), serum IgG response to the hinge peptide was significantly elevated at 12 months of age in both cohorts (**Figure 5A**), indicating that IgG sero-reactivity to *Lachnospiraceae* flagellins is a homeostatic response present in infants regardless of environmental exposures, and this heightened response is driven by reactivity to the hinge epitope. Accompanied with this, there was an increase in sero-reactivity to other microbiota antigens at 12 months of age(30), reflecting the development of homeostatic IgG response to endogenous gut microbiota in early life. Interestingly, increase in hinge-specific sero-reactivity already happened at 6 months of age in Ugandan infants, but remained low in infants in the US/Sweden cohort;

whereas reactivity was much greater in US/Sweden infants than Ugandan infants at 12 months of age (**Figure 5A**). Furthermore, sero-reactivity to the hinge peptide was attributed to both sub-epitopes (D0N p25-44 and D1N p41-59) in the US/Sweden cohort (**Figure 5B and 5C**), but there was a lack of potent sero-reactivity to sub-epitope D0N p25-44 in Uganda infants at 12 months of age (**Figure 5B**). Distinct from response to the hinge peptide, serum IgG reactivity to the other peptide epitopes tested in the cytometric bead array in Ugandan infants exhibited a significant increase at 6 months of age and started to decline at 12 months of age, which was not observed in the US/Sweden cohort (**Figure 5A-C and Supplementary Figure 5**). Collectively, these data indicate that there is an environmental contribution to the kinetics of the development of homeostatic IgG response to the gut microbiota.

DISCUSSION

The dynamic interactions of the host with its microbiota begins at birth and continues throughout the life of the host. The composition of human gut microbiota changes over a lifetime, but most critically early in life where the microbiota is unstable until 3 years of age(32-34). A collection of genes is important in establishing and maintaining homeostasis with the microbiota, and when such genes are deleted from mice, colitis ensues(35-37). A similar process occurs in humans resulting in IBD, which eventuates in immune reactivity to the microbiota. Indeed, serologic responses to gut microbial antigens have been identified, such as perinuclear Anti-Neutrophil Cytoplasmic Antibodies (pANCA) in UC(5, 38), and anti-*Saccharomyces cerevisiae* antibodies (ASCA) in CD(39, 40). In addition, sero-response to flagellins, which is preferentially directed at flagellins of the *Lachnospiraceae* family of bacteria, was shown elevated specifically in patients with CD, but not in UC patients or HC(6, 8, 24). Data in this paper provides further insight into such response, identifying a shared dominant B cell peptide epitope in CD patients for the first time. Importantly, heightened sero-reactivity to this dominant epitope at CD diagnosis is able to predict the future development of disease complications up to 3 years, making the *Lachnospiraceae* flagellin hinge peptide a promising biomarker candidate for both CD diagnosis and prognosis.

IgG⁺ plasma cells only constitute a small proportion of plasma cells in the intestinal lamina propria at homeostasis, with most of the plasma cells being IgA⁺⁽⁴¹⁻⁴³⁾. However, anti-commensal IgG responses are significantly induced during intestinal inflammation, both in mouse models of colitis(25, 44), and in patients with IBD(45-48). It has been shown that IgG antibodies specific to flagellins derived from *Salmonella* spp. or *E. coli* aggravated colitis progression in mouse models(48, 49). However, these microorganisms are neither abundant mucus colonizers in healthy human intestines(33, 43), nor primary targets of serum IgG from CD patients(24). *Lachnospiraceae* of the Firmicutes phylum are one of the most abundant mucus-associated bacterial families in both human and mouse intestines(33, 50). They are very diverse: analysis of the genome of 273 isolates from human donors found only 60% of genes shared among multiple species(51). At homeostasis, they are beneficial commensals due to their ability to convert dietary fibers into short-chain fatty acids and thus promote the differentiation and maintenance of epithelial cells and regulatory CD4⁺ T cells (Treg) in the gut(52, 53). They are obligate anaerobes which are resistant to host antimicrobial peptides(54), and they produce lantibiotics that contribute

to resistance to pathogen colonization, protecting the host(51). Many, but not all, are flagellated, a property facilitating their colonization of the mucosa. Therefore, flagellin, the monomeric subunit of the flagellum, is one of the most expressed bacterial proteins in the intestinal lumen and the mucosal layer. Crystallization studies of flagellin have revealed four domains, D0 and D1 that are localized internally in the flagellum, and D2 and D3 that are on the exposed surface of the flagellum(16). A recent study by Bourgonje *et al.* showed that serum IgG response to conserved motifs (D0 and D1 domains) in both the N-terminus and C-terminus of *Lachnospiraceae* flagellins was overrepresented in CD patients. Such response was identified using a massive antigen library including microbial peptides, immune epitope and food allergens using a high-throughput phage-display immunoprecipitation sequencing technique(55). Our study confirmed this finding, and beyond that, we defined the dominant B cell peptide epitope (D0-1N p25-59) of *Lachnospiraceae* flagellins in CD patients for the first time. Another interesting finding was that little or no reactivity to the Toll-like receptor 5 (TLR5) binding site of the flagellin (D1N p74-102) was detected in any of the subjects tested. As one of the receptors for flagellins, TLR5 is expressed by epithelial cells, including Paneth cells at the base of crypts, innate immune cells and some Treg cells(56). The TLR5 binding site on the flagellin locates in the D1 domain(18, 57), where only the monomeric form of flagellin is able to bind and activate host TLR5; those in the flagellum with D1 domain internalized thus do not bind TLR5. Antibodies targeting the conserved D0/D1 region were shown to neutralize flagellin TLR5 activation(58), but it is unclear which epitope contributed to the neutralization. It is possible that antibody directly blocking the flagellin TLR5 binding site inhibits its TLR5 activation. However, antibodies targeting the flagellin TLR5 allosteric binding site (located in the D0 domain)(59-61) may also alter its TLR5 activating function. Considering that the dominant B cell epitope identified in this study locates at the hinge region of D0-D1 domain of *Lachnospiraceae* flagellins, one can ask what role these antibodies may play in flagellin TLR5 activation in sero-positive subjects. In addition, it has been shown that the level of *Lachnospiraceae* bacteria was significantly reduced preceding the onset of CD and in newly diagnosed, treatment-naïve pediatric CD patients(62, 63), raising the question whether flagellin-specific antibodies in patients with CD are involved in the depletion of gut *Lachnospiraceae* bacteria. Do antibodies specific to D0N p25-44 and D1N p41-59 sub-epitopes play different roles? These are important questions that need to be further investigated in future studies.

Another important finding of this study was that most human infants at one year of age have a strong IgG response to *Lachnospiraceae* flagellin, targeting the exact dominant peptide epitope identified in CD patients. CD4⁺ T cell help is required for B cell responses to peptides, and T cells are restricted by major histocompatibility complex (MHC) class II molecules. The almost universal IgG response to the hinge peptide despite human immune diversity indicates that the T cell response is not restricted to any MHC class II allele. Indeed, T cell epitopes that are promiscuous, i.e., can bind to multiple MHC class II alleles, are enriched in *Lachnospiraceae* flagellins. Starting from 6 months, infants are often first introduced to solid food and thus experiencing a massive shift in the composition of the gut microbiota(32, 33). This is also a critical time window for the colonization and diversification of Firmicutes in the human gut. The increased serum IgG response to *Lachnospiraceae* flagellins reflects an augmented homeostatic T and B cell response to the colonizing bacteria, in that this process does not involve intestinal inflammation. Interestingly, the magnitude of hinge-specific serum IgG is higher in Sweden/US infants than that of Ugandan infants. This might be due to how soon infants raised in different lifestyles are introduced to solid food, and how long breast feeding is continued. However, whether the generation of this adaptive IgG response is simply a result of antigen encounter, or the flagellin-specific antibodies are actively involved in *Lachnospiraceae* colonization needs further investigation. Interestingly, this heightened sero-reactivity to *Lachnospiraceae* flagellins declines with age, in that the flagellin epitope response is hardly detected in healthy adults, consistent with an active regulation of this B cell reactivity and the associated T cell reactivity. It is unclear which immune pathways participate in the downregulation of such immune responses, and the robust IgG sero-reactivity to *Lachnospiraceae* flagellins in CD patients indicates a loss of immunological control. Because flagellins are potent activators of several innate immune receptors, it is possible that defective innate immune pathways are involved. Interestingly, sero-reactivity to the *Lachnospiraceae* flagellin hinge peptide was also found elevated in a subset of patients with severe chronic fatigue syndrome(64). These data are consistent with a homeostatic immune response to flagellin in infancy being converted to a pathologic response later in life.

Data in this study raise many interesting and important questions. The strong serum IgG response observed in both healthy infants and CD patients makes us ask what role the hinge epitope specific B cells play in homeostasis and in the early phases of CD. Does the flagellin specific B cell (and CD4⁺ T cell) response in Crohn's patients decline with age as in most healthy children,

and then increase at the onset of the pathology, or does it remain elevated from infancy? Why do flagellin specific IgG antibodies predominantly target the highly conserved hinge peptide and are these antibodies germline or somatically hypermutated? Although we mainly focused on characterizing the flagellin specific sero-reactivity in this study, we are also interested to see whether flagellin specific, especially hinge epitope specific IgG and secretory IgA exist in the intestinal mucosa, and if so, do they block the uptake of flagellins into lymphoid tissues and thus prevent systemic activation? Furthermore, we previously showed that the Treg to T effector ratio of flagellin reactive CD4⁺ cells were significantly reduced in CD patients compared to HCs. This finding leads us to ask whether *Lachnospiraceae* bacteria express flagellins and produce short chain fatty acids such as butyrate synergistically to stimulate predominately Treg cells in most humans, but not in patients with CD. The answers to these questions will await further studies which will be facilitated by the development of the flagellin peptide cytometric bead array that can readily identify patients early in their disease course. Considering that an augmented flagellin-specific CD4⁺ T cell response accompanies multi-flagellin IgG reactivity, identification of CD patients with heightened hinge reactivity might also benefit patient treatment by offering a novel flagellin peptide-guided immunotherapy, which ablates circulating flagellin-specific memory CD4⁺ T cells by concomitant cell activation and metabolism inhibition(27).

MATERIALS AND METHODS

Study Design. The objective of this research was to identify immunodominant B cell peptide epitopes in the *Lachnospiraceae* bacterial family in patients with Crohn's disease and in healthy/non-IBD control infants, and correlate this immune reactivity with clinical disease course in the former. This was accomplished by performing a novel flagellin peptide array and flagellin peptide cytometric bead array with an adult discovery cohort containing patients with Crohn's disease, ulcerative colitis, and healthy controls, and a validation cohort of pediatric treatment-naïve Crohn's patients and non-IBD controls. Furthermore, a cohort of healthy and non-IBD control infants and their mothers from three distinct geographic locations was utilized for the discovery of the homeostatic antibody response to the same dominant B cell epitope in *Lachnospiraceae* flagellins. Sample sizes were selected on the basis of expected variance and effect size in well-characterized experimental systems and were large enough to achieve a $\geq 90\%$ power to detect differences $> 20\%$ among groups. Replicated samples were randomly omitted. Experiments of the validation cohort were conducted blinded until data analyses.

Patient sample collection. Peripheral blood samples along with medical metadata of IBD patients and healthy volunteers in the UAB cohort were collected at the Kirklin Clinic of UAB Hospital and in research labs in Birmingham, Alabama upon informed consent. All steps followed UAB's Institutional Review Board guidelines (IRB X300001155, X140515004, and F081003001) and the Declaration of Helsinki Principles. Serum samples of subjects in the RISK cohort were obtained from Dr. Subra Kugathasan at Emory University, USA. Serum samples of subjects in the Uganda and US cohort were obtained from Dr. Edward N. Janoff at the University of Colorado Denver, USA. Serum samples of healthy and allergic subjects in the Sweden cohort were obtained from Dr. Maria C. Jenmalm at Linköping University, Sweden. The Regional Ethics Committee for Human Research at Linköping University approved the study (Dnr 99323). Informed consent was obtained from both parents before inclusion. All serum samples of the RISK, Uganda, and US/Sweden cohorts were obtained through Material Transfer Agreement.

Motif discovery. Protein sequences (D0-1N p0-75) of *Lachnospiraceae* flagellins included in the flagellin peptide microarray plus *S. dublin* FliC and *E. coli* FliC were analyzed with MEME Suite 5.5.0 for the discovery of recurring protein motifs with fixed length (6-35 width), with 0-order model of sequences used as the background model. Motifs discovered were ordered by E-value and the most significant motif was shown in **Figure 1**.

Microbiota protein microarray. Microbiota protein microarray was performed as previously described(24). In brief, recombinant bacterial flagellin proteins were expressed and diluted in 10mM Tris (PH7.4) with 20% glycerol and 0.1% sodium dodecyl sulfate at 0.2mg/ml. Antigens were printed in quadruplicate on slides coated with nitrocellulose pads (Maine Manufacturing) using a Spotbot Personal microarrayer (ArrayIt). Antigen printed pads were blocked with SuperBlock (ThermoFisher) for 1 hour before probed with serum samples (diluted at 1:100 in SuperBlock). After 1 hour, pads were washed 3 times with PBS plus 0.05% Tween 20, and then incubated with Dylight 550 Goat anti-human IgA (ImmunoReagents, Cat # GtxHu-001-E550NHSX) and Dylight 650 Donkey anti-human IgG (Invitrogen, Cat # SA5-10129) at 1:1000 dilution for 1 hour. Finally, pads were washed 3 times with PBS plus Tween, air-dried, and scanned using GenePix 4000B imager (Axion).

Flagellin peptide microarray. 1512 overlapping 15mer peptides derived from 19 *Lachnospiraceae* flagellins of human and mouse origin were synthesized and printed on individual glass slide as previously described(65). Slides were blocked with SuperBlock for 1 hour before they were probed with human sera (dilution factor indicated in corresponding Figure Legends). Slides were washed 3 times with PBS plus 0.05% Tween 20 after 1 hour incubation with serum samples, and then incubated with Dylight 550 Goat anti-human IgA and Dylight 650 Donkey anti-human IgG at 1:1000 dilution for 1 hour. Finally, slides were washed 3 times with PBS plus Tween, air-dried, and scanned using GenePix 4000B imager.

Generation of flagellin peptide microarray heatmaps. UAB cohort: 1477 of the original 1512 peptides are depicted (columns in the heatmap). A peptide was excluded if more than 190 subjects had signal greater than 2.16 (i.e., $\log_{10}(\text{signal} + 1) = .5$). 265 of the original 288 subjects are depicted (rows in the heatmap). A subject was excluded if more than 500 peptides had signal greater than 2.16. These thresholds were determined by identifying cutoffs in long tails of distributions subjects and peptides. RISK cohort: 474 peptides and 120 subjects are depicted.

Flagellin peptide cytometric bead array. 8 conserved peptides from *Lachnospiraceae* flagellins were synthesized, with biotin and a GSGSG linker attached to their N-terminus. D1N p41-59 and D0-1N p25-59 include a 50/50 mixture of 2 consensus peptides with one amino acid different, respectively. Sequence of each peptide is listed as follows:

D0N p1-19: [Biotin]GSGSGMVVQHNLRAMNSNRMLGIT[COOH]

D0N p25-44: [Biotin]GSGSGKSTEKLSSGYKINRAADDAA[COOH]

D1N p41-59: [Biotin]GSGSGDDAAGLTISEKMR(S/K)QIRGL[COOH]

D0-1N p25-59:

[Biotin]GSGSGKSTEKLSSGYKINRAADDAAGLTISEKMR(S/K)QIRGL[COOH]

D1N p74-102: [Biotin]GSGSGQTAEGALTEVHDMLQRMNELAVQAANGTN[COOH]

D1C p391-407: [Biotin]GSGSGLGAVQNRLEHTINNLDN[COOH]

D0-1C p410-435: [Biotin]GSGSGENTTAAESQIRDMDMATEMVKYSNNN[COOH]

D0C p437-460: [Biotin]GSGSGLAQAGQSMLAQSNQANQGVLSLLG[COOH]

Note that () indicates that it's a 50/50 mix of peptides at this position.

Flagellin peptide cytometric bead array was performed as previously described(66). In brief, biotinylated flagellin peptides were conjugated on to Streptavidin coated 4um polystyrene beads (Spherotech, Inc. Cat # PAK-4067-8K). Beads conjugated with each individual peptide have different fluorescent intensity in the APC/Cy7 channel so that they can be differentiated from each other by flow cytometry. After incubation of 40ul diluted serum samples (serum samples were diluted at 1:200 and 1:600 in SuperBlock to ensure sufficient detection of epitope-specific IgG response; IgM and IgA responses were detected at 1:200 and 1:600 with IgG, respectively) from individuals with CD, UC, and HCs with the array of peptide-bound beads, antibody response to individual peptides was detected with different fluorescently labeled secondary antibodies (Goat anti-human IgM-AF488, SouthernBiotech, Cat # 2020-30; Goat anti-human IgG-PE, SouthernBiotech, Cat # 2040-09; Goat anti-human IgA-PE, SouthernBiotech, Cat # 2050-09; and Goat anti-human IgG-AF488, SouthernBiotech, Cat # 2040-30) using CytoFLEX cytometer (The Beckman Coulter Life Sciences).

Calculation of epitope-specific antibody concentrations in flagellin peptide cytometric bead array. Biotinylated Goat F(ab)₂ anti-human isotype antibodies (anti-human IgM, SouthernBiotech, Cat # 2022-01; anti-human IgA, SouthernBiotech, Cat # 2052-01, and anti-human IgG, SouthernBiotech, Cat # 2042-01) were conjugated on to selected peaks of Streptavidin coated 4um polystyrene capture beads, respectively, and incubated with 0.75x serial dilutions of purified polyclonal human antibodies (IgG, SouthernBiotech, Cat # 0150-01; IgM, SouthernBiotech, Cat # 0158L-01; IgA, SouthernBiotech, Cat # 0155K-01) at known concentrations. Standard curves were obtained based on MFI detected by secondary antibodies bound using CytoFLEX cytometer. Flow cytometry standard (FCS) files of both samples and the standard curve were applied to a set of MATLAB programs (<https://github.com/UAB->

[Immunology-Institute/cba-toolbox](#)) generated by Dr. Rosenberg at UAB for the quantification of peptide-specific antibodies in the corresponding isotype(66).

Statistical analysis. Statistical analyses were performed with GraphPad PRISM, version 9.5, with nonparametric Kruskal-Wallis test or two-way ANOVA as indicated. Dunn's or Tukey's multiple comparisons test was used when comparing multiple groups. All statistical tests were 2-sided. Specific tests used and values of n are stated in corresponding figure legends. Data were presented as means \pm SEM. The differences were considered statistically significant at $P < 0.05$ (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$) and ns indicates not significant.

Acknowledgments

We thank all the patients and volunteers that participated in this study and the hospital staff for obtaining blood samples from participants. We thank all site investigators from 28 sites and the Crohn's and Colitis Foundation of America for supporting the RISK cohort study. The authors would also like to thank BioRender for providing the platform for figure generation.

Authorship Contributions

COE initiated, designed, and supervised the study. QZ designed and conducted the study, analyzed the data, and wrote the manuscript. LWD participated in study design, conducted the microbiota protein microarray and flagellin peptide microarray, and analyzed the data. JTK and RGK participated in the setup and data analysis of the flagellin peptide cytometric bead array. AFR analyzed and visualized the data, and designed the program for the quantification of epitope-specific antibodies in the cytometric bead array. PJM, LAD, ENJ, MCJ and COE acquired blood samples and clinical diagnoses from study participants.

Conflict of Interest

Conflicts of interest are disclosed by the following authors: COE and UAB hold a patent on Lachnospiraceae A4 Fla2, which has been licensed for clinical use by Prometheus Laboratories. COE, QZ, LWD, RGK, and UAB have filed a provisional patent on the flagellin peptide cytometric bead array. COE is founder and Chief Scientific Officer of ImmPrev Bio, which is developing an antigen-directed immunotherapy for Crohn's disease. The remaining authors disclose no conflicts.

Funding

This work was supported by a Synergy Award from the Rainin Foundation, a grant from the Department of Veterans Affairs (CX0001530), NIAID (U19 AI142737), NIH (R01 HD059527), NIDDK (T32 DK007545), and NIAID (T32 AI007051).

Figure Legends

Figure 1. *Lachnospiraceae* flagellin peptide microarray identifies key IgG binding epitopes in CD patients. (A) Schematic strategy of the flagellin peptide microarray. (B) Heatmap showing serum IgG reactivity for sequential, overlapping fifteen-residue flagellin peptides (108 peptides from each of the 19 different flagellins) in CD (n=135), HC (n=42), and UC (n=88) subjects. Sera were diluted at 1:100. Rows (subjects) were clustered using Euclidean distance and complete linkage applied to log-transformed data. Columns (peptides) are sorted by peptide position (and include peptides from up to 19 different flagellins per position). Cohort groups are indicated by colored dots on the right margin. Peptide starting position (relative to the N-terminus) is indicated on the bottom. For each 15-mer segment, per-residue information content is computed at each position across all flagellins (except *L. CBir11* Fla and *L. CBir66* Fla) and summated to indicate sequence conservation, shown as color. Although all peptides were used for the clustering, only peptides up to and including 121-136 are depicted in this heatmap. (See **Supplementary Figure 1** for version showing all peptides). (C) 3D structure of *L. A4* Fla2 as a representative flagellin showing the location of the hinge region. (D) Sequence alignment of flagellins included in the peptide microarray plus *S. dublin* FliC and *E. coli* FliC. Sequences were aligned in BioEdit using the ClustalW alignment feature. Flagellin D0 and D1 domains in the N-term are indicated by the color bars on top, whereas the hinge region is marked by the red box. (E) Motif discovery of the flagellin hinge region.

Figure 2. Serum IgG multi-flagellin reactivity in CD patients is driven by reactivity to the dominant *Lachnospiraceae* flagellin peptide epitope. (A) Schematic strategy of the flagellin peptide cytometric bead array. (B and C) Gating strategies of the flagellin peptide cytometric bead array. (D) Sera from CDhigh (n=39), CDlow (n=126), UC (n=181), and HC (n=95) subjects in the UAB cohort were probed against the flagellin peptide cytometric bead array. Representative flow plots of serum IgG reactivity to flagellin peptide epitopes in each group are shown. The red box highlights the hinge region peptides. (E) Histogram overlay of serum IgG reactivity to *Lachnospiraceae* D0-1N p25-59 in 3 representative CDhigh, CDlow, UC, and HC, respectively. (F-M) Epitope-specific IgG concentration in the serum of CDhigh, CDlow, UC, and HC subjects to indicated flagellin peptide epitopes. Data are presented as means \pm SEM and analyzed with nonparametric Kruskal-Wallis test. *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

Figure 3. *Lachnospiraceae* hinge peptide is the dominant B cell epitope in pediatric CD patients and predicts the development of disease complications. (A) Sera of CD patients (grouped by disease behavior at 3-year follow-up using the Montreal classification, n=250 for B1 (inflammatory), n=142 for B2 (stricturing) or B3 (penetrating)) and non-IBD controls (n=72) from the RISK cohort were probed against the microbiota protein microarray. Mean fluorescent intensity (MFI) of serum IgG specific for indicated flagellin antigens is presented as means \pm SEM and analyzed with nonparametric Kruskal-Wallis test. *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001. (B) Heatmap of log₁₀ IgG response (color) on the microbiota protein microarray for 21 flagellins (rows) and 392 individuals with CD in the RISK cohort (columns). Rows and columns were hierarchically clustered using Euclidean distance on the log data, average linkage and optimal leaf ordering maximizing the sum of similarity between every leaf and all other leaves in the adjacent cluster. Montreal disease classification at follow-up (B2 and B3) is indicated below the heatmap, with unmarked individuals being B1. The bar graph at the bottom indicates the number of positive flagellin responses for each individual, where a positive response is two standard deviations above the mean (non-log transformed) of a non-IBD control cohort (n=72, not shown) per flagellin. (C and D) Summary of multi-flagellin reactivity in CD patients with B1 or B2/3 behavior at follow-up. (E-G) Sera of CD patients (grouped by disease behavior at 3-year follow-up, n=249 for B1, n=140 for B2/3) and non-IBD controls (n=72) from the RISK cohort were probed against the flagellin peptide cytometric bead array. Epitope-specific IgG concentration in different groups to indicated flagellin peptide epitopes is shown. Data are presented as means \pm SEM and analyzed with nonparametric Kruskal-Wallis test. *P<0.05, **P<0.01, ****P<0.0001.

Figure 4. Serum IgG to *Lachnospiraceae* hinge peptide strongly correlates with multi-flagellin reactivity in CD. (A) Heatmaps representing correlation coefficients between serum IgG anti-flagellin reactivity (log₁₀ MFI from microbiota protein microarray; rows) and serum IgG anti-peptide reactivity (hyperbolic arcsin MFI from flagellin peptide cytometric bead array; columns). Four separate heatmaps are shown for CD (n=148), HC (n=86), and UC (n=169) participants from the UAB cohort and CD (n=414) participants from the RISK cohort, respectively. Color corresponds to Pearson correlation coefficient indicated in the bar scale. Dots indicate correlations that were below Bonferroni-corrected p-value of 0.01 (based on 21 flagellins, 8 peptides, 4 groups

= 672 tests) testing the hypothesis of no correlation against the alternative that there is a non-zero correlation. Rows (flagellins) are sorted based on the correlation for peptide D0-1N p25-59 in CD patients in the RISK cohort, as indicated by the triangle. **(B)** Hyperbolic arcsin MFI of serum IgG anti-*Lachnospiraceae* D0-1N p25-59 and D1C p391-407 of individual CD, HC, and UC participants from the UAB cohort and CD participants from the RISK cohort obtained from the flagellin peptide cytometric bead array was plotted against reactivity (\log_{10} MFI) to *R. intestinalis* Fla1 and *E. coli* FliC obtained from the microbiota protein microarray, respectively. Least squares regression was performed, with the correlation line in red and R value shown for each plot. **(C and D)** The value of hinge-specific IgG obtained from the flagellin peptide cytometric bead array is plotted on X axis, and the number of flagellins bound by sera in the same individual CD patients on the microbiota protein microarray on the Y axis. Data from the UAB cohort and the RISK cohort are presented as indicated.

Figure 5. Serum IgG reactivity to *Lachnospiraceae* flagellin hinge peptide is a homeostatic response in most infants. **(A)** Sera of healthy and non-IBD control infants at 0 month (cord blood, CB), 6 months, and 12 months of age (with all three timepoints) from the Uganda cohort (n=76) and the US (n=10) /Sweden (n=36) cohort were probed against the flagellin peptide cytometric bead array. Concentration of serum IgG specific to *Lachnospiraceae* flagellin D0-1N p25-59 in different groups is shown. Each dot represents an individual and responses of the same individual at different timepoints are linked with the grey lines. The means of serum IgG response to *Lachnospiraceae* flagellin D0-1N p25-59 of CD patients in the UAB (broken line) and RISK (solid line) cohort are plotted. Data are presented as means \pm SEM and analyzed with repeated measures ANOVA and Friedman test. ****P<0.0001. **(B and C)** Sera of healthy and non-IBD control infants at indicated timepoints (including individuals not represented in all three timepoints) from the Uganda cohort (n=92) and the US (n=16) /Sweden (n=54) cohort were probed against the flagellin peptide cytometric bead array. IgG concentration specific to *Lachnospiraceae* D0N p25-44 and D1N p41-59 sub-epitopes in different groups is shown, respectively. Data are presented as means \pm SEM and analyzed with two-way ANOVA and Tukey's multiple comparisons test. *P<0.05, ***P<0.001, ****P<0.0001.

REFERENCE

1. Khor, B., A. Gardet, and R.J. Xavier, Genetics and pathogenesis of inflammatory bowel disease. *Nature* **474**, 307-17, (2011).
2. Elson, C.O. and Y. Cong, Host-microbiota interactions in inflammatory bowel disease. *Gut Microbes* **3**, 332-44, (2012).
3. de Souza, H.S. and C. Fiocchi, Immunopathogenesis of IBD: current state of the art. *Nat Rev Gastroenterol Hepatol* **13**, 13-27, (2016).
4. Cho, J.H., The genetics and immunopathogenesis of inflammatory bowel disease. *Nat Rev Immunol* **8**, 458-66, (2008).
5. Saxon, A., F. Shanahan, C. Landers, T. Ganz, and S. Targan, A distinct subset of antineutrophil cytoplasmic antibodies is associated with inflammatory bowel disease. *J Allergy Clin Immunol* **86**, 202-10, (1990).
6. Lodes, M.J., Y. Cong, C.O. Elson, R. Mohamath, C.J. Landers, S.R. Targan, M. Fort, and R.M. Hershberg, Bacterial flagellin is a dominant antigen in Crohn disease. *J Clin Invest* **113**, 1296-306, (2004).
7. Landers, C.J., O. Cohavy, R. Misra, H. Yang, Y.C. Lin, J. Braun, and S.R. Targan, Selected loss of tolerance evidenced by Crohn's disease-associated immune responses to auto- and microbial antigens. *Gastroenterology* **123**, 689-99, (2002).
8. Targan, S.R., C.J. Landers, H. Yang, M.J. Lodes, Y. Cong, K.A. Papadakis, E. Vasiliauskas, C.O. Elson, and R.M. Hershberg, Antibodies to CBir1 flagellin define a unique response that is associated independently with complicated Crohn's disease. *Gastroenterology* **128**, 2020-8, (2005).
9. Shen, C., C.J. Landers, C. Derkowski, C.O. Elson, and S.R. Targan, Enhanced CBir1-specific innate and adaptive immune responses in Crohn's disease. *Inflamm Bowel Dis* **14**, 1641-51, (2008).
10. Torres, J., F. Petralia, T. Sato, P. Wang, S.E. Telesco, R.S. Choung, R. Strauss, X.J. Li, R.M. Laird, R.L. Gutierrez, C.K. Porter, S. Plevy, F. Princen, J.A. Murray, M.S. Riddle, and J.F. Colombel, Serum Biomarkers Identify Patients Who Will Develop Inflammatory Bowel Diseases Up to 5 Years Before Diagnosis. *Gastroenterology* **159**, 96-104, (2020).
11. Lee, S.H., W. Turpin, O. Espin-Garcia, J.A. Raygoza Garay, M.I. Smith, H. Leibovitzh, A. Goethel, D. Turner, D. Mack, C. Deslandres, M. Cino, G. Aumais, R. Panaccione, K. Jacobson, A. Bitton, A.H. Steinhart, H.Q. Huynh, F. Princen, P. Moayyedi, A.M. Griffiths, M.S. Silverberg, A.D. Paterson, W. Xu, K. Croitoru, Crohn's, and C. Colitis Canada Genetics Environment Microbial Project Research, Anti-Microbial Antibody Response is Associated With Future Onset of Crohn's Disease Independent of Biomarkers of Altered Gut Barrier Function, Subclinical Inflammation, and Genetic Risk. *Gastroenterology* **161**, 1540-1551, (2021).
12. van Schaik, F.D., B. Oldenburg, A.R. Hart, P.D. Siersema, S. Lindgren, O. Grip, B. Teucher, R. Kaaks, M.M. Bergmann, H. Boeing, F. Carbonnel, P. Jantchou, M.C. Boutron-Ruault, A. Tjonneland, A. Olsen, F.L. Crowe, P.H. Peeters, M.G. van Oijen, and H.B. Bueno-de-Mesquita, Serological markers predict inflammatory bowel disease years before the diagnosis. *Gut* **62**, 683-8, (2013).
13. Choung, R.S., F. Princen, T.P. Stockfish, J. Torres, A.C. Maue, C.K. Porter, F. Leon, B. De Vroey, S. Singh, M.S. Riddle, J.A. Murray, J.F. Colombel, and P.S. Team, Serologic

- microbial associated markers can predict Crohn's disease behaviour years before disease diagnosis. *Aliment Pharmacol Ther* **43**, 1300-10, (2016).
14. Minamino, T. and K. Namba, Self-assembly and type III protein export of the bacterial flagellum. *J Mol Microbiol Biotechnol* **7**, 5-17, (2004).
 15. Maki-Yonekura, S., K. Yonekura, and K. Namba, Conformational change of flagellin for polymorphic supercoiling of the flagellar filament. *Nat Struct Mol Biol* **17**, 417-22, (2010).
 16. Yonekura, K., S. Maki-Yonekura, and K. Namba, Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424**, 643-50, (2003).
 17. Hayashi, F., K.D. Smith, A. Ozinsky, T.R. Hawn, E.C. Yi, D.R. Goodlett, J.K. Eng, S. Akira, D.M. Underhill, and A. Aderem, The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099-103, (2001).
 18. Smith, K.D., E. Andersen-Nissen, F. Hayashi, K. Strobe, M.A. Bergman, S.L. Barrett, B.T. Cookson, and A. Aderem, Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. *Nat Immunol* **4**, 1247-53, (2003).
 19. Franchi, L., A. Amer, M. Body-Malapel, T.D. Kanneganti, N. Ozoren, R. Jagirdar, N. Inohara, P. Vandenabeele, J. Bertin, A. Coyle, E.P. Grant, and G. Nunez, Cytosolic flagellin requires Ipaf for activation of caspase-1 and interleukin 1beta in salmonella-infected macrophages. *Nat Immunol* **7**, 576-82, (2006).
 20. Miao, E.A., C.M. Alpuche-Aranda, M. Dors, A.E. Clark, M.W. Bader, S.I. Miller, and A. Aderem, Cytoplasmic flagellin activates caspase-1 and secretion of interleukin 1beta via Ipaf. *Nat Immunol* **7**, 569-75, (2006).
 21. Zhao, Y., J. Yang, J. Shi, Y.N. Gong, Q. Lu, H. Xu, L. Liu, and F. Shao, The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature* **477**, 596-600, (2011).
 22. Zeng, H., A.Q. Carlson, Y. Guo, Y. Yu, L.S. Collier-Hyams, J.L. Madara, A.T. Gewirtz, and A.S. Neish, Flagellin is the major proinflammatory determinant of enteropathogenic Salmonella. *J Immunol* **171**, 3668-74, (2003).
 23. Giron, J.A., A.G. Torres, E. Freer, and J.B. Kaper, The flagella of enteropathogenic Escherichia coli mediate adherence to epithelial cells. *Mol Microbiol* **44**, 361-79, (2002).
 24. Alexander, K.L., Q. Zhao, M. Reif, A.F. Rosenberg, P.J. Mannon, L.W. Duck, and C.O. Elson, Human Microbiota Flagellins Drive Adaptive Immune Responses in Crohn's Disease. *Gastroenterology*, (2021).
 25. Brandwein, S.L., R.P. McCabe, Y. Cong, K.B. Waites, B.U. Ridwan, P.A. Dean, T. Ohkusa, E.H. Birkenmeier, J.P. Sundberg, and C.O. Elson, Spontaneously colitic C3H/HeJBir mice demonstrate selective antibody reactivity to antigens of the enteric bacterial flora. *J Immunol* **159**, 44-52, (1997).
 26. Feng, T., L. Wang, T.R. Schoeb, C.O. Elson, and Y. Cong, Microbiota innate stimulation is a prerequisite for T cell spontaneous proliferation and induction of experimental colitis. *J Exp Med* **207**, 1321-32, (2010).
 27. Zhao, Q., L.W. Duck, F. Huang, K.L. Alexander, C.L. Maynard, P.J. Mannon, and C.O. Elson, CD4(+) T cell activation and concomitant mTOR metabolic inhibition can ablate microbiota-specific memory cells and prevent colitis. *Sci Immunol* **5**, (2020).
 28. Kugathasan, S., L.A. Denson, T.D. Walters, M.O. Kim, U.M. Marigorta, M. Schirmer, K. Mondal, C. Liu, A. Griffiths, J.D. Noe, W.V. Crandall, S. Snapper, S. Rabizadeh, J.R. Rosh, J.M. Shapiro, S. Guthery, D.R. Mack, R. Kellermayer, M.D. Kappelman, S. Steiner, D.E.

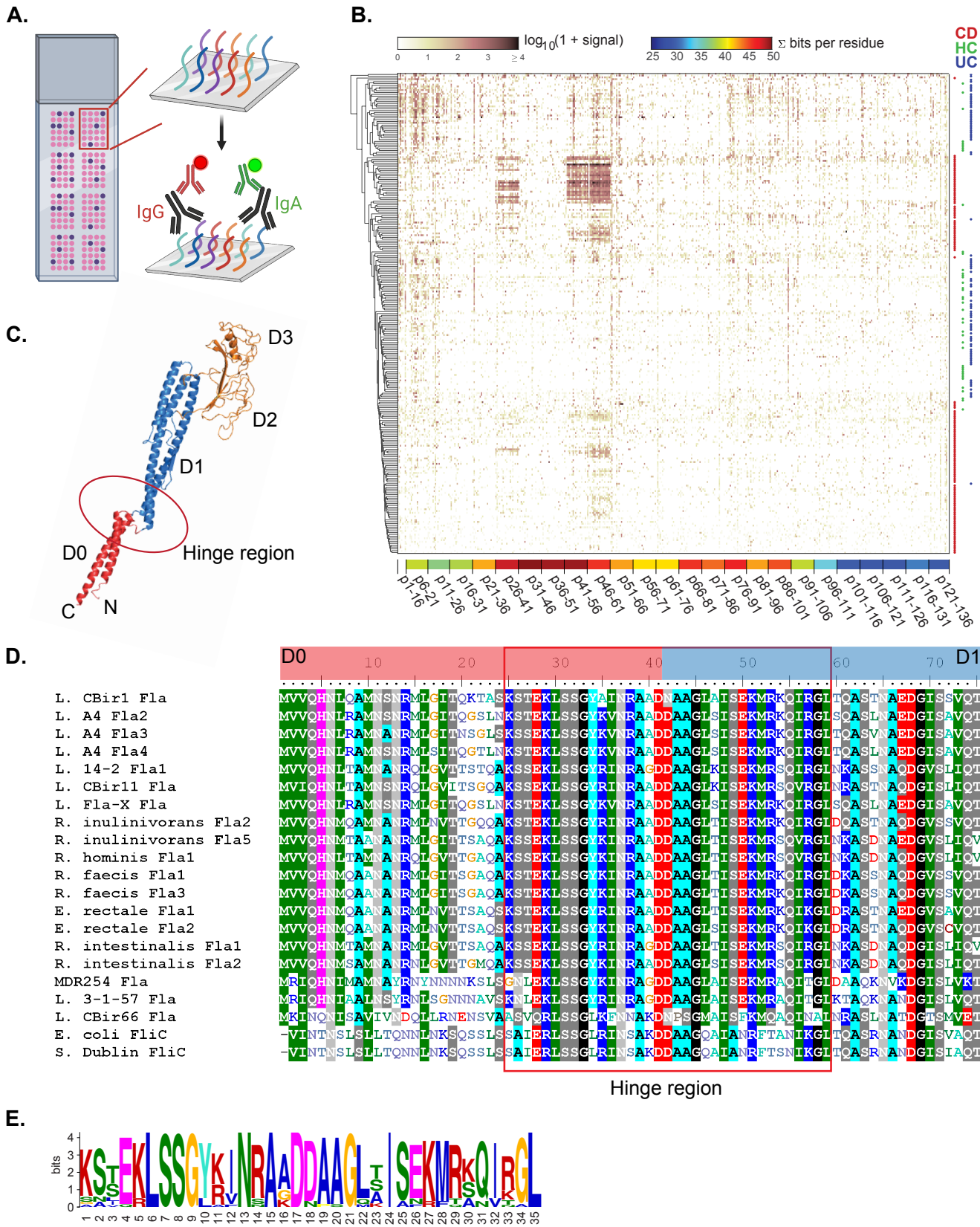
- Moulton, D. Keljo, S. Cohen, M. Oliva-Hemker, M.B. Heyman, A.R. Otley, S.S. Baker, J.S. Evans, B.S. Kirschner, A.S. Patel, D. Ziring, B.C. Trapnell, F.A. Sylvester, M.C. Stephens, R.N. Baldassano, J.F. Markowitz, J. Cho, R.J. Xavier, C. Huttenhower, B.J. Aronow, G. Gibson, J.S. Hyams, and M.C. Dubinsky, Prediction of complicated disease course for children newly diagnosed with Crohn's disease: a multicentre inception cohort study. *Lancet* **389**, 1710-1718, (2017).
29. Gaensbauer, J.T., J.T. Rakhola, C. Onyango-Makumbi, M. Mubiru, J.E. Westcott, N.F. Krebs, E.J. Asturias, M.G. Fowler, E. McFarland, and E.N. Janoff, Impaired haemophilus influenzae type b transplacental antibody transmission and declining antibody avidity through the first year of life represent potential vulnerabilities for HIV-exposed but -uninfected infants. *Clin Vaccine Immunol* **21**, 1661-7, (2014).
 30. Christmann, B.S., T.R. Abrahamsson, C.N. Bernstein, L.W. Duck, P.J. Mannon, G. Berg, B. Bjorksten, M.C. Jenmalm, and C.O. Elson, Human seroreactivity to gut microbiota antigens. *J Allergy Clin Immunol* **136**, 1378-86 e1-5, (2015).
 31. Jespersen, M.C., B. Peters, M. Nielsen, and P. Marcatili, BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* **45**, W24-W29, (2017).
 32. Koenig, J.E., A. Spor, N. Scalfone, A.D. Fricker, J. Stombaugh, R. Knight, L.T. Angenent, and R.E. Ley, Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4578-85, (2011).
 33. Lozupone, C.A., J.I. Stombaugh, J.I. Gordon, J.K. Jansson, and R. Knight, Diversity, stability and resilience of the human gut microbiota. *Nature* **489**, 220-30, (2012).
 34. Yatsunenko, T., F.E. Rey, M.J. Manary, I. Trehan, M.G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R.N. Baldassano, A.P. Anokhin, A.C. Heath, B. Warner, J. Reeder, J. Kuczynski, J.G. Caporaso, C.A. Lozupone, C. Lauber, J.C. Clemente, D. Knights, R. Knight, and J.I. Gordon, Human gut microbiome viewed across age and geography. *Nature* **486**, 222-7, (2012).
 35. Panwala, C.M., J.C. Jones, and J.L. Viney, A novel model of inflammatory bowel disease: mice deficient for the multiple drug resistance gene, *mdr1a*, spontaneously develop colitis. *J Immunol* **161**, 5733-44, (1998).
 36. Kuhn, R., J. Lohler, D. Rennick, K. Rajewsky, and W. Muller, Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75**, 263-74, (1993).
 37. Sadlack, B., H. Merz, H. Schorle, A. Schimpl, A.C. Feller, and I. Horak, Ulcerative colitis-like disease in mice with a disrupted interleukin-2 gene. *Cell* **75**, 253-61, (1993).
 38. Oudkerk Pool, M., P.M. Ellerbroek, B.U. Ridwan, R. Goldschmeding, B.M. von Blomberg, A.S. Pena, K.M. Dolman, H. Bril, W. Dekker, J.J. Nauta, and et al., Serum antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease are mainly associated with ulcerative colitis. A correlation study between perinuclear antineutrophil cytoplasmic autoantibodies and clinical parameters, medical, and surgical treatment. *Gut* **34**, 46-50, (1993).
 39. Main, J., H. McKenzie, G.R. Yeaman, M.A. Kerr, D. Robson, C.R. Pennington, and D. Parratt, Antibody to *Saccharomyces cerevisiae* (bakers' yeast) in Crohn's disease. *BMJ* **297**, 1105-6, (1988).
 40. Quinton, J.F., B. Sendid, D. Reumaux, P. Duthilleul, A. Cortot, B. Grandbastien, G. Charrier, S.R. Targan, J.F. Colombel, and D. Poulain, Anti-*Saccharomyces cerevisiae*

- mannan antibodies combined with antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease: prevalence and diagnostic role. *Gut* **42**, 788-91, (1998).
41. Mowat, A.M. and W.W. Agace, Regional specialization within the intestinal immune system. *Nat Rev Immunol* **14**, 667-85, (2014).
 42. Spencer, J. and L.M. Sollid, The human intestinal B-cell response. *Mucosal Immunol* **9**, 1113-24, (2016).
 43. James, K.R., T. Gomes, R. Elmentaite, N. Kumar, E.L. Gulliver, H.W. King, M.D. Stares, B.R. Bareham, J.R. Ferdinand, V.N. Petrova, K. Polanski, S.C. Forster, L.B. Jarvis, O. Suchanek, S. Howlett, L.K. James, J.L. Jones, K.B. Meyer, M.R. Clatworthy, K. Saeb-Parsy, T.D. Lawley, and S.A. Teichmann, Distinct microbial and immune niches of the human colon. *Nat Immunol* **21**, 343-353, (2020).
 44. Duck, L.W., M.R. Walter, J. Novak, D. Kelly, M. Tomasi, Y. Cong, and C.O. Elson, Isolation of flagellated bacteria implicated in Crohn's disease. *Inflamm Bowel Dis* **13**, 1191-201, (2007).
 45. Macpherson, A., U.Y. Khoo, I. Forgacs, J. Philpott-Howard, and I. Bjarnason, Mucosal antibodies in inflammatory bowel disease are directed against intestinal bacteria. *Gut* **38**, 365-75, (1996).
 46. Ruthlein, J., M. Ibe, W. Burghardt, J. Mossner, and I.O. Auer, Immunoglobulin G (IgG), IgG1, and IgG2 determinations from endoscopic biopsy specimens in control, Crohn's disease, and ulcerative colitis subjects. *Gut* **33**, 507-12, (1992).
 47. Gouni-Berthold, I., B. Baumeister, H.K. Berthold, and C. Schmidt, Immunoglobulins and IgG subclasses in patients with inflammatory bowel disease. *Hepatogastroenterology* **46**, 1720-3, (1999).
 48. Castro-Dopico, T., T.W. Dennison, J.R. Ferdinand, R.J. Mathews, A. Fleming, D. Clift, B.J. Stewart, C. Jing, K. Strongili, L.I. Labzin, E.J.M. Monk, K. Saeb-Parsy, C.E. Bryant, S. Clare, M. Parkes, and M.R. Clatworthy, Anti-commensal IgG Drives Intestinal Inflammation and Type 17 Immunity in Ulcerative Colitis. *Immunity* **50**, 1099-1114 e10, (2019).
 49. Kobayashi, K., S.W. Qiao, M. Yoshida, K. Baker, W.I. Lencer, and R.S. Blumberg, An FcRn-dependent role for anti-flagellin immunoglobulin G in pathogenesis of colitis in mice. *Gastroenterology* **137**, 1746-56 e1, (2009).
 50. Qin, J., R. Li, J. Raes, M. Arumugam, K.S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada, D.R. Mende, J. Li, J. Xu, S. Li, D. Li, J. Cao, B. Wang, H. Liang, H. Zheng, Y. Xie, J. Tap, P. Lepage, M. Bertalan, J.M. Batto, T. Hansen, D. Le Paslier, A. Linneberg, H.B. Nielsen, E. Pelletier, P. Renault, T. Sicheritz-Ponten, K. Turner, H. Zhu, C. Yu, S. Li, M. Jian, Y. Zhou, Y. Li, X. Zhang, S. Li, N. Qin, H. Yang, J. Wang, S. Brunak, J. Dore, F. Guarner, K. Kristiansen, O. Pedersen, J. Parkhill, J. Weissenbach, H.I.T.C. Meta, P. Bork, S.D. Ehrlich, and J. Wang, A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59-65, (2010).
 51. Sorbara, M.T., E.R. Littmann, E. Fontana, T.U. Moody, C.E. Kohout, M. Gjonbalaj, V. Eaton, R. Seok, I.M. Leiner, and E.G. Pamer, Functional and Genomic Variation between Human-Derived Isolates of Lachnospiraceae Reveals Inter- and Intra-Species Diversity. *Cell Host Microbe* **28**, 134-146 e4, (2020).
 52. Louis, P. and H.J. Flint, Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol Lett* **294**, 1-8, (2009).

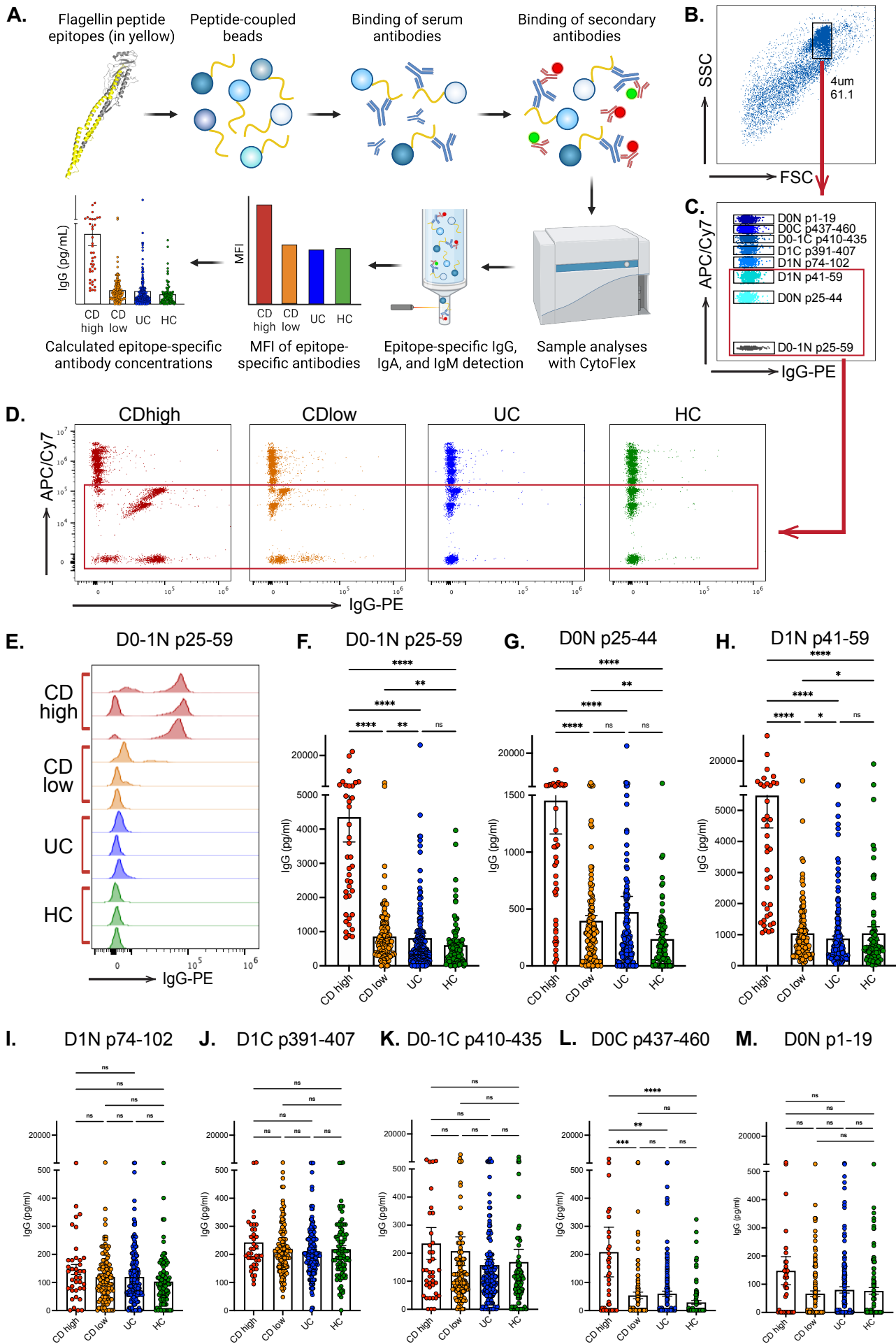
53. Patterson, A.M., I.E. Mulder, A.J. Travis, A. Lan, N. Cerf-Bensussan, V. Gaboriau-Routhiau, K. Garden, E. Logan, M.I. Delday, A.G.P. Coutts, E. Monnais, V.C. Ferraria, R. Inoue, G. Grant, and R.I. Aminov, Human Gut Symbiont *Roseburia hominis* Promotes and Regulates Innate Immunity. *Front Immunol* **8**, 1166, (2017).
54. Sun, D., R. Bai, W. Zhou, Z. Yao, Y. Liu, S. Tang, X. Ge, L. Luo, C. Luo, G.F. Hu, J. Sheng, and Z. Xu, Angiogenin maintains gut microbe homeostasis by balancing alpha-Proteobacteria and Lachnospiraceae. *Gut* **70**, 666-676, (2021).
55. Bourgonje, A.R., S. Andreu-Sanchez, T. Vogl, S. Hu, A. Vich Vila, R. Gacesa, S. Leviatan, A. Kurilshikov, S. Klompus, I.N. Kalka, H.M. van Dullemen, A. Weinberger, M.C. Visschedijk, E.A.M. Festen, K.N. Faber, C. Wijmenga, G. Dijkstra, E. Segal, J. Fu, A. Zhernakova, and R.K. Weersma, Phage-display immunoprecipitation sequencing of the antibody epitope repertoire in inflammatory bowel disease reveals distinct antibody signatures. *Immunity*, (2023).
56. Price, A.E., K. Shamardani, K.A. Lugo, J. Deguine, A.W. Roberts, B.L. Lee, and G.M. Barton, A Map of Toll-like Receptor Expression in the Intestinal Epithelium Reveals Distinct Spatial, Cell Type-Specific, and Temporal Patterns. *Immunity* **49**, 560-575 e6, (2018).
57. Kreuzberger, M.A.B., C. Ewing, F. Poly, F. Wang, and E.H. Egelman, Atomic structure of the *Campylobacter jejuni* flagellar filament reveals how epsilon Proteobacteria escaped Toll-like receptor 5 surveillance. *Proc Natl Acad Sci U S A* **117**, 16985-16991, (2020).
58. Nempont, C., D. Cayet, M. Rumbo, C. Bompard, V. Villeret, and J.C. Sirard, Deletion of flagellin's hypervariable region abrogates antibody-mediated neutralization and systemic activation of TLR5-dependent immunity. *J Immunol* **181**, 2036-43, (2008).
59. Yoon, S.I., O. Kurnasov, V. Natarajan, M. Hong, A.V. Gudkov, A.L. Osterman, and I.A. Wilson, Structural basis of TLR5-flagellin recognition and signaling. *Science* **335**, 859-64, (2012).
60. Forstneric, V., K. Ivicak-Kocjan, T. Plaper, R. Jerala, and M. Bencina, The role of the C-terminal D0 domain of flagellin in activation of Toll like receptor 5. *PLoS Pathog* **13**, e1006574, (2017).
61. Clasen, S.J., M.E.W. Bell, A. Borbon, D.H. Lee, Z.M. Henseler, J. de la Cuesta-Zuluaga, K. Parys, J. Zou, Y. Wang, V. Altmannova, N.D. Youngblut, J.R. Weir, A.T. Gewirtz, Y. Belkhadir, and R.E. Ley, Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. *Sci Immunol* **8**, eabq7001, (2023).
62. Raygoza Garay, J.A., W. Turpin, S.H. Lee, M.I. Smith, A. Goethel, A.M. Griffiths, P. Moayyedi, O. Espin-Garcia, M. Abreu, G.L. Aumais, C.N. Bernstein, I.A. Biron, M. Cino, C. Deslandres, I. Dotan, W. El-Matary, B. Feagan, D.S. Guttman, H. Huynh, L.A. Dieleman, J.S. Hyams, K. Jacobson, D. Mack, J.K. Marshall, A. Otley, R. Panaccione, M. Ropeleski, M.S. Silverberg, A.H. Steinhart, D. Turner, B. Yerushalmi, A.D. Paterson, W. Xu, C.G.P.R. Consortium, and K. Croitoru, Gut Microbiome Composition Is Associated With Future Onset of Crohn's Disease in Healthy First-Degree Relatives. *Gastroenterology*, (2023).
63. Gevers, D., S. Kugathasan, L.A. Denson, Y. Vazquez-Baeza, W. Van Treuren, B. Ren, E. Schwager, D. Knights, S.J. Song, M. Yassour, X.C. Morgan, A.D. Kostic, C. Luo, A. Gonzalez, D. McDonald, Y. Haberman, T. Walters, S. Baker, J. Rosh, M. Stephens, M. Heyman, J. Markowitz, R. Baldassano, A. Griffiths, F. Sylvester, D. Mack, S. Kim, W.

- Crandall, J. Hyams, C. Huttenhower, R. Knight, and R.J. Xavier, The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15**, 382-392, (2014).
64. Vogl, T., I.N. Kalka, S. Klompus, S. Leviatan, A. Weinberger, and E. Segal, Systemic antibody responses against human microbiota flagellins are overrepresented in chronic fatigue syndrome patients. *Sci Adv* **8**, eabq2422, (2022).
65. Legutki, J.B., Z.G. Zhao, M. Greving, N. Woodbury, S.A. Johnston, and P. Stafford, Scalable high-density peptide arrays for comprehensive health monitoring. *Nat Commun* **5**, 4785, (2014).
66. Rosenberg, A.F., J. J.T. Killian, T.J. Green, J. Akther, M.E. Hossain, S. Qiu, T.D. Randall, F.E. Lund, and R.G. King, A high-throughput multiplex array for antigen-specific serology with automated analysis. *bioRxiv*, 2023.03.29.534777, (2023).

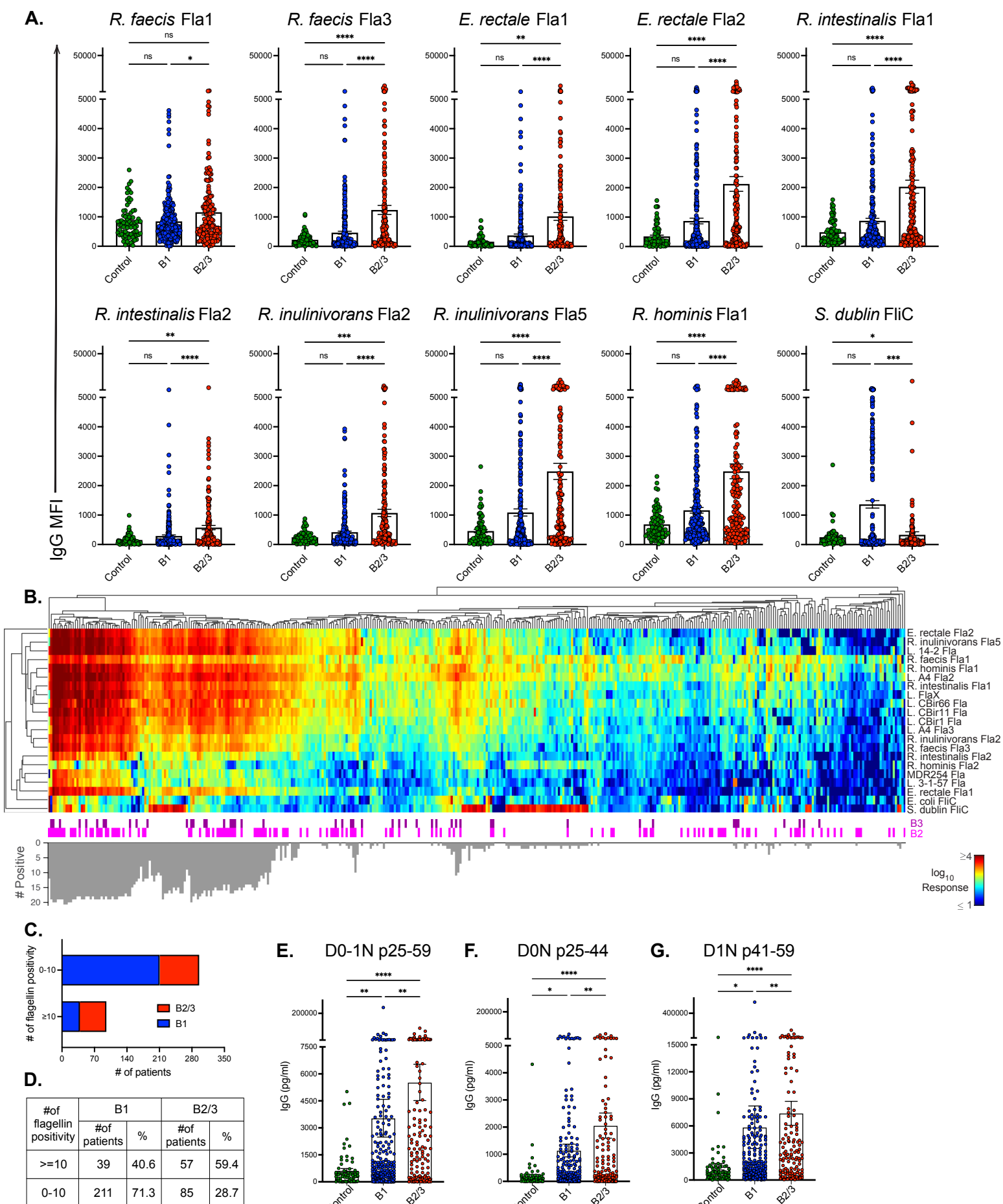
Zhao et al. Figure 1



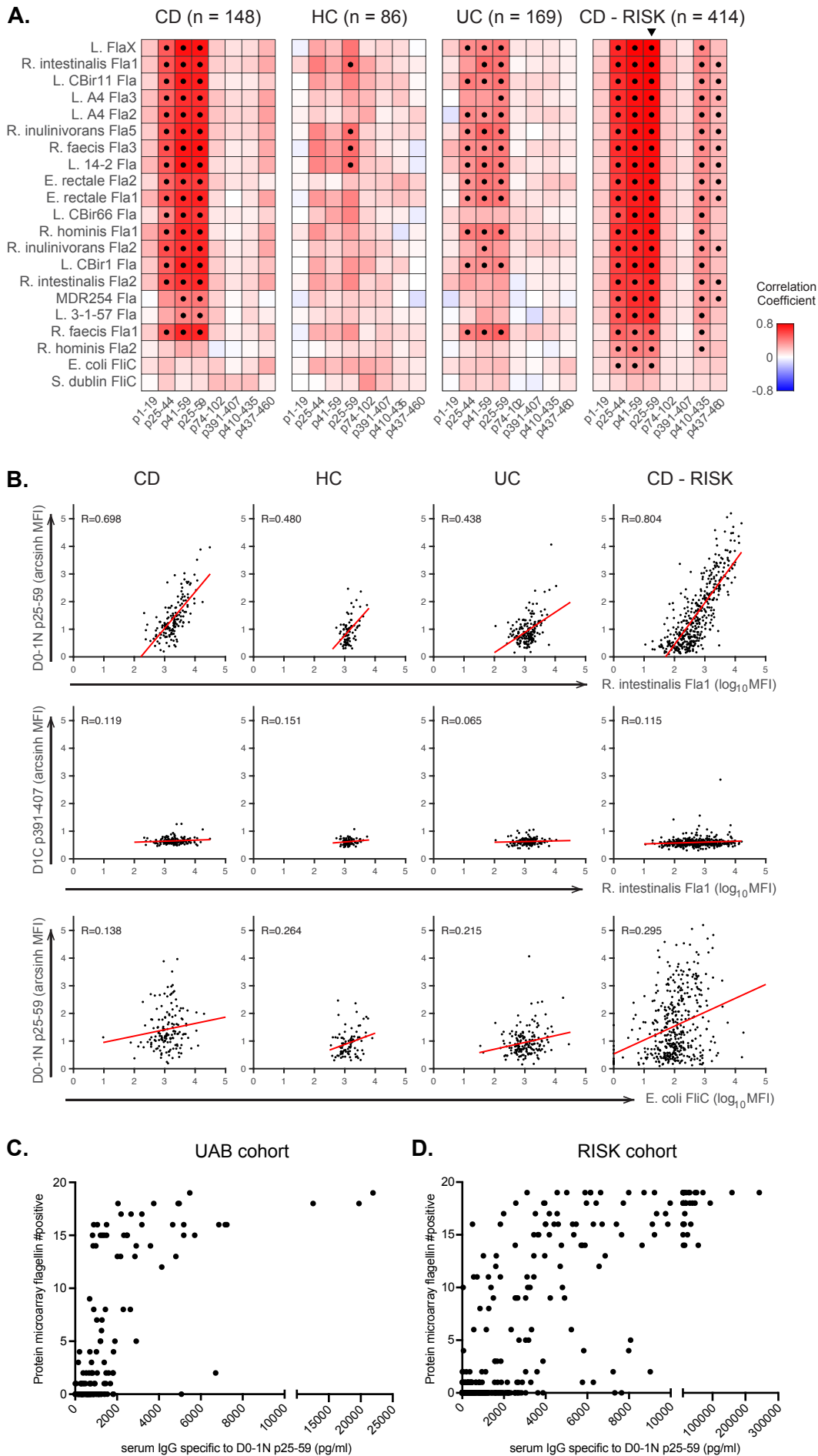
Zhao et al. Figure 2



Zhao et al. Figure 3



Zhao et al. Figure 4



Zhao et al. Figure 5

