Signature biomarkers in Crohn's disease: toward a molecular classification

R Arsenescu[1], MEC Bruno[2], EW Rogier[2], AT Stefka[2], AE McMahan[2], TB Wright[2], MS Nasser[1], WJS de Villiers[1] and CS Kaetzel[2]

In an effort to develop a molecular classification scheme for Crohn's disease (CD), mucosal biopsies from 69 CD patients and 28 normal controls were analyzed for expression of the RelA subunit of nuclear factor (NF)-κB, A20 (a negative regulator of NF-κB), polymeric immunoglobulin receptor (pIgR), tumor necrosis factor (TNF), and interleukin (IL)-8. Principal component analysis was used to classify individuals into three subsets based on patterns of biomarker expression. Set 1 included normal subjects and CD patients with mild disease and good responses to therapy, thus defining "normal" biomarker expression. CD patients in set 2, characterized by low expression of all five biomarkers, had moderate to severe disease and poor responses to immunosuppressive and anti-TNF therapy. Patients in set 3, characterized by low expression of RelA, A20, and pIgR, normal TNF and elevated IL-8, had acute inflammation that responded well to therapy. Classification of CD patients by these biomarkers may predict disease behavior and responses to therapy.

INTRODUCTION

Crohn's disease (CD) is a chronic disorder characterized by patchy transmural inflammation of the small and/or large bowel, thought to result from an inappropriate inflammatory response to normal components of the intestinal microbiota in genetically predisposed individuals. [1–3] Progress has been made in characterizing the effector cells and molecules that comprise the inflammatory response, and current biological therapies are targeted toward neutralizing these effectors.[4–7] However, recent evidence suggests that the primary defect in CD may actually be a defective innate immune response within the intestinal mucosa.[8,9] There is general agreement that molecular biomarkers for CD are needed to improve diagnosis and guide therapies.[10] The aim of this study was to identify a set of biomarkers representative of the innate immune response in the gut mucosa, which could be used to classify CD patients into molecular phenotypic subsets predictive of disease behavior and responses to therapy.

The association of nuclear factor (NF)-κB signaling with induction of pro-inflammatory cytokines has led to the hypothesis that excessive activation of NF-κB is central to the pathogenesis of CD.[11,12] Recent findings, however, have suggested that activation of NF-κB may be crucial for regulation of intestinal inflammation and maintenance of epithelial barrier function.[13–15] Thus "physiological" inflammation induced by RelA signaling may be important for maintaining epithelial homeostasis in the gut, and reduced expression of RelA could contribute to the "innate immunodeficiency" seen in CD.

NF-κB induces expression of A20, a ubiquitin-modifying enzyme that negatively regulates the NF-κB activation pathway.[16,17] Mice genetically deficient in A20 are hypersensitive to tumor necrosis factor (TNF)- and Toll-like receptor (TLR)-induced NF-κB activation, and develop severe intestinal inflammation.[18] We hypothesized that reduced expression of A20 in the intestinal mucosa could therefore be a risk factor for CD.

Secretory antibodies of the IgA class (SIgA) form the first line of antigen-specific immune protection at mucosal surfaces and promote homeostasis between the intestinal epithelium and the commensal microbiota.[19,20] Transport of IgA across intestinal epithelial cells into the gut lumen is mediated by the polymeric immunoglobulin receptor (pIgR).[21,22] Following transport, the extracellular domain of pIgR is cleaved to form secretory component (SC), which remains associated with SIgA and confers additional innate immune functions.[23,24] Targeted deletion

[1]Division of Digestive Diseases & Nutrition, Department of Internal Medicine, University of Kentucky College of Medicine, Lexington, Kentucky, USA. [2]Department of Microbiology, Immunology & Molecular Genetics, University of Kentucky College of Medicine, Lexington, Kentucky, USA. Correspondence: CS Kaetzel (cskaet@uky.edu) or R Arsenescu (razvan-arsenescu@uky.edu)

of the *Pigr* gene in mice causes elevated serum IgA, increased mucosal permeability, and increased susceptibility to experimental colitis.[25–27] Expression of pIgR in intestinal epithelial cell is upregulated by TNF and TLR signaling via activation of NF-κB.[21,22] We hypothesized that diminished expression of pIgR in the intestinal mucosa could lead to increased inflammatory responses to the commensal microbiota.

The pro-inflammatory cytokine TNF is associated with mucosal inflammation in CD, and therapies designed to neutralize TNF activity have shown promise in a subset of CD patients who are refractory to conventional therapies.[6,28] TNF also has protective roles in innate immunity, including induction of NOD2 and pIgR expression, epithelial restitution, and control of potentially pathogenic luminal bacteria.[29–32] We hypothesized that although optimal expression of TNF in the intestinal mucosa may promote physiological inflammation, reduced TNF could lead to defective innate immunity.

Interleukin (IL)-8 is a potent chemoattractant for neutrophils, the major component of the cellular infiltrate in acute inflammation.[33] Although neutrophils contribute to mucosal inflammation through the release of soluble mediators, they also play an important role in removal of bacteria and foreign debris that could otherwise promote a granulomatous response by mucosal macrophages.[8,9] Accordingly, we investigated IL-8 as a potential biomarker for acute inflammation in CD patients.

Given the heterogeneous clinical presentation in CD, we anticipated that no single biomarker would accurately predict disease behavior. We utilized a multifactorial approach to develop a molecular classification scheme, based on expression levels of RelA, A20, pIgR, TNF, and IL-8 in the intestinal mucosa of CD patients and normal controls.

RESULTS
Biomarker expression in CD patients and normal controls
Mucosal biopsies from the colon and/or terminal ileum were obtained at colonoscopy from 69 CD patients and 28 normal controls (**Figure 1a**). On the basis of visual inspection and histology, tissues were classified as inflamed or non-inflamed. To test the hypothesis that the innate immune response is inherently dysregulated in CD patients, we compared mRNA levels in the colon and ileum for five potential biomarkers in non-inflamed tissue from CD patients and healthy controls (**Figure 1b** and **c**). Levels of RelA, A20, and TNF mRNA were significantly lower in non-inflamed colon mucosa of CD patients than in normal controls. Expression of pIgR in colon mucosa was highly heterogeneous in normal controls and was not significantly different from CD patients. However, the observation that many CD patients had low pIgR expression suggested that this biomarker might be used to identify a subset of CD patients with deficient pIgR-mediated IgA transport. IL-8 expression was extremely low in the majority of normal controls and CD patients. However, elevated IL-8 expression in some CD patients suggested that this biomarker could be useful in a multifactorial analysis of gene expression. There were no significant differences in expression in the ileum mucosa between normal controls and CD patients for any of the five biomarkers, suggesting that

dysregulated expression of RelA, A20, and TNF may be a unique feature of the large bowel in CD. To determine whether expression of these biomarkers varies between the colon and ileum, paired biopsies (within the same individual) were analyzed from 19 normal controls (**Figure 2a**) and 22 CD patients (**Figure 2b**). Expression of pIgR was higher in the colon than in the ileum of both normal controls and CD patients, likely due to enhanced TLR signaling from greater numbers of commensal bacteria.[34] We have previously reported that pIgR expression is significantly higher in the colon than in the ileum in mice and is reduced in mice with deficient TLR signaling.[35] Expression of RelA was slightly but significantly higher in the colon than in the ileum of CD patients, but not of normal controls. No differences in expression of A20, TNF, or IL-8 were observed between the colon and ileum of either normal controls or CD patients. We conclude that regional differences in gene expression between the colon and ileum do not account for the observation that dysregulated expression of RelA, A20, and TNF in CD patients was restricted to the large bowel.

To test the hypothesis that local inflammation influenced biomarker expression, we analyzed paired biopsies of non-inflamed and inflamed colon mucosa from 17 CD patients (**Figure 3a**). Remarkably, there were no significant differences in mRNA levels for any of the five biomarkers, suggesting that the observed dysregulation of gene expression in CD patients was an inherent condition of the colon mucosa, and not secondary to local inflammation. Because access to the ileum during colonoscopy was limited to the terminal 10–20 cm, we were only able to obtain a single biopsy of the ileum mucosa per individual. Accordingly, each CD patient was classified as having either non-inflamed or inflamed ileum, even though regional differences in inflammation could have existed in other parts of the small bowel. Interestingly, no differences in the expression of RelA, A20, pIgR, or TNF were observed among different CD patients with or without ileal inflammation (**Figure 3b**). However, expression of IL-8 was significantly elevated in mucosa of CD patients with visible inflammation in the terminal ileum, suggesting that this gene may be a biomarker for acute ileal disease.

Multifactorial analysis of five biomarkers defines molecular phenotypic subsets in CD
The heterogeneous patterns of gene expression in both normal controls and CD patients suggested that a multifactorial scheme might be more robust than individual biomarkers for classifying molecular phenotypes. Principal component analysis (PCA) is a statistical approach that is used to reduce the number of variables in a complex data set and to classify relationships between variables.[36] A recent study utilized PCA to identify a subset of IBD patients with significant abnormalities in the composition of the colonic microbiota.[37] We used PCA to classify our study cohort (including both normal controls and CD patients) into molecular phenotypic subsets based on expression of RelA, A20, pIgR, TNF, and IL-8 mRNA in non-inflamed intestinal mucosa. The first step in this analysis was to examine patterns of correlated expression among the five potential biomarkers
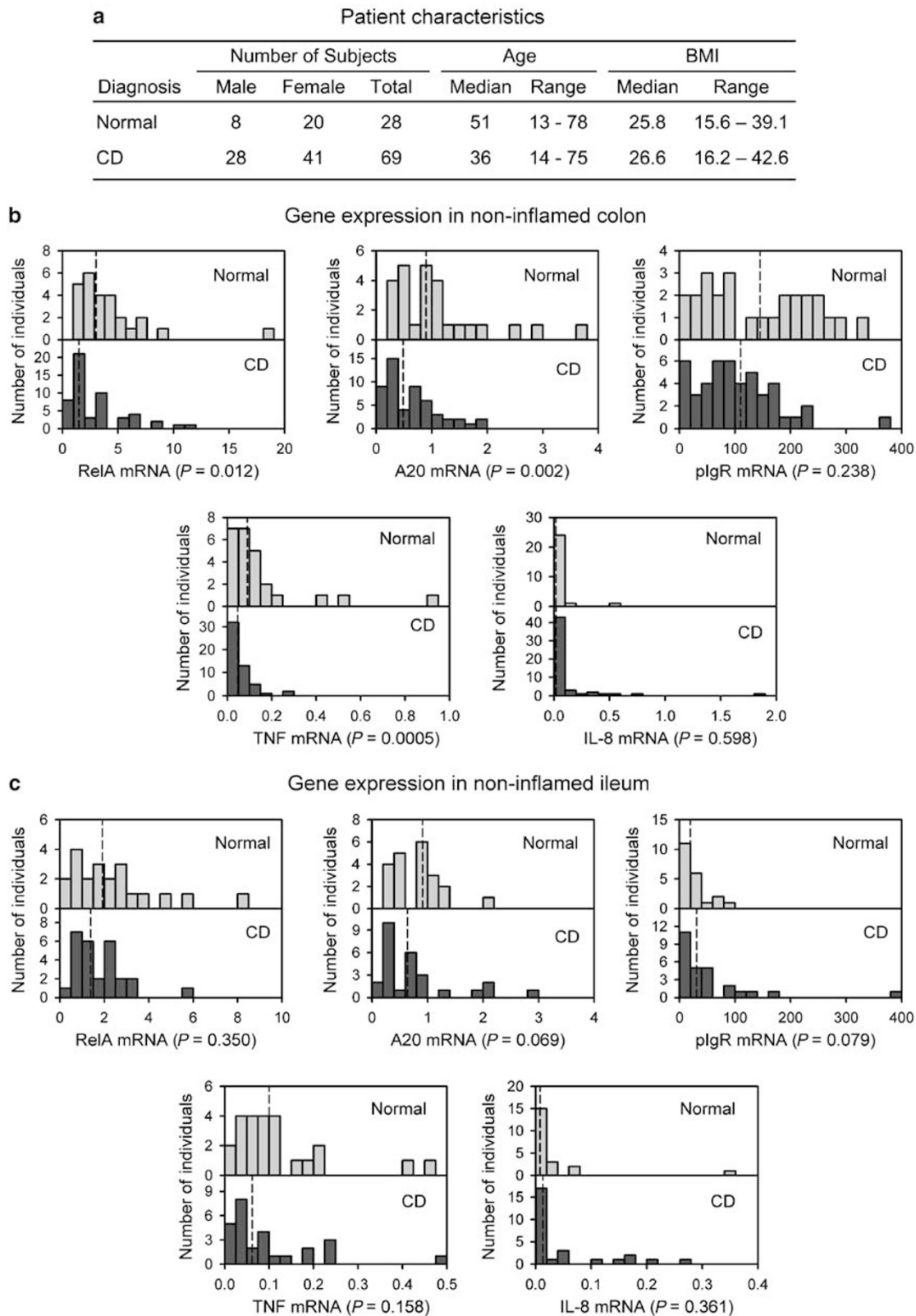
a Patient characteristics

	Number of Subjects			Age		BMI	
Diagnosis	Male	Female	Total	Median	Range	Median	Range
Normal	8	20	28	51	13 - 78	25.8	15.6 – 39.1
CD	28	41	69	36	14 - 75	26.6	16.2 – 42.6

b Gene expression in non-inflamed colon

c Gene expression in non-inflamed ileum

Figure 1 Biomarker expression in CD patients and normal controls. Patient characteristics (**a**) and gene expression in non-inflamed colon (**b**) and ileum (**c**) mucosa were compared between CD patients and normal controls. mRNA levels were measured by RT-qPCR and normalized to β2-microglobulin mRNA. Data are displayed as histograms to illustrate the heterogeneity in expression of individual biomarkers. Dashed lines indicate median expression levels for each group. Significant differences among groups were tested by Mann–Whitney non-parametric analysis; *P*-values are listed for each comparison of normal controls vs. CD patients. CD, Crohn's disease; RT-qPCR, real-time quantitative PCR.
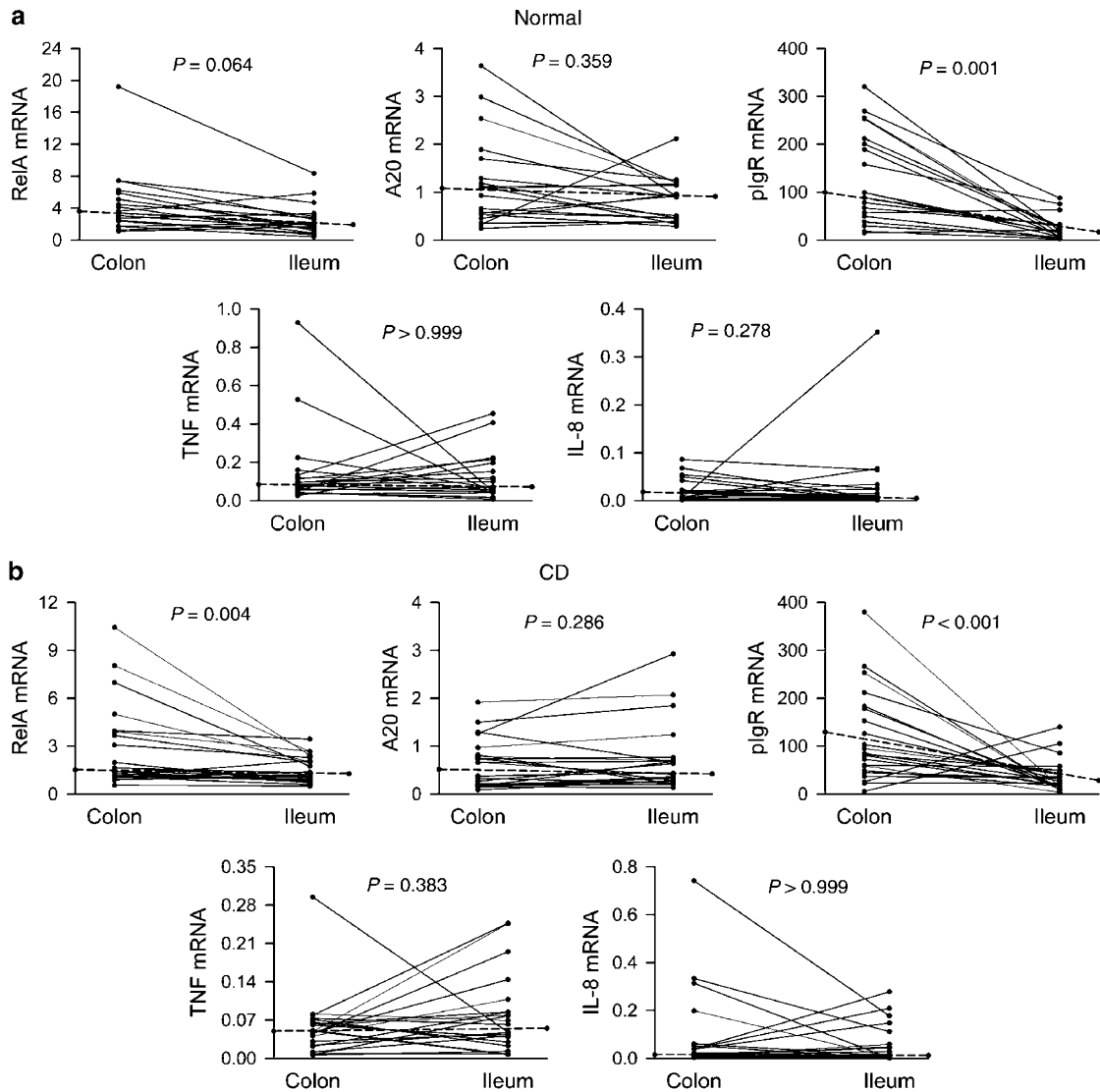
Figure 2 Comparison of gene expression in colon and ileum. Paired biopsies were obtained from non-inflamed mucosa of the colon and ileum for 19 normal controls (**a**) and 22 CD patients (**b**). mRNA levels were measured by RT-qPCR and normalized to β2-microglobulin mRNA. Dashed lines indicate the difference in median expression levels between colon and ileum for each gene. Significant differences among groups were tested by paired sign non-parametric analysis; *P*-values are listed for each comparison of colon vs. ileum. CD, Crohn's disease; RT-qPCR, real-time quantitative PCR.

(**Figure 4a**). In the colon, highly significant correlations were observed among the expression levels of RelA, A20, pIgR, and TNF, consistent with the function of NF-κB in inducing transcription of the A20, pIgR, and TNF genes. Expression of TNF but not RelA was positively correlated with IL-8, suggesting that TNF may promote IL-8 gene transcription through a RelA-independent mechanism. In the ileum, A20 expression was positively correlated with RelA and TNF, but RelA and TNF were not correlated with each other. In contrast to the colon, IL-8 expression was positively correlated with pIgR but not TNF.

Principal component analysis was used to reduce the five variables to two factors or principal components (PCs) (**Figure 4b**). The χ^2 analysis demonstrated that most of the variability in the data set could be described by PC1 and PC2 ($P < 0.0001$ for the colon and $P = 0.0255$ for the ileum). On the basis of colon gene expression, PC1 was strongly weighted toward RelA, A20,

and pIgR, with an intermediate weight for TNF and a negative weight for IL-8 (see Factor tables in **Figure 4b**). PC2 was strongly weighted toward TNF and IL-8, with a low weight for RelA and negative weights for A20 and pIgR. For ileum gene expression, PC1 was strongly weighted toward RelA, A20, pIgR, and TNF and weakly weighted toward IL-8. PC2 was strongly weighted toward IL-8, weakly weighted toward pIgR and TNF, and negatively weighted toward RelA and A20. Each normal control and CD patient was assigned a score for PC1 and PC2 based on the sum of weighted expression levels for all five biomarkers (see Methods). Scatter plots of the scores for PC1 vs. PC2 demonstrated that individuals could be classified into three subsets based on colon or ileum gene expression (**Figure 4c**). Set 1 is defined as individuals with a high score for PC1 and a low score for PC2, set 2 is defined as individuals with low scores for both PC1 and PC2, and set 3 is defined as
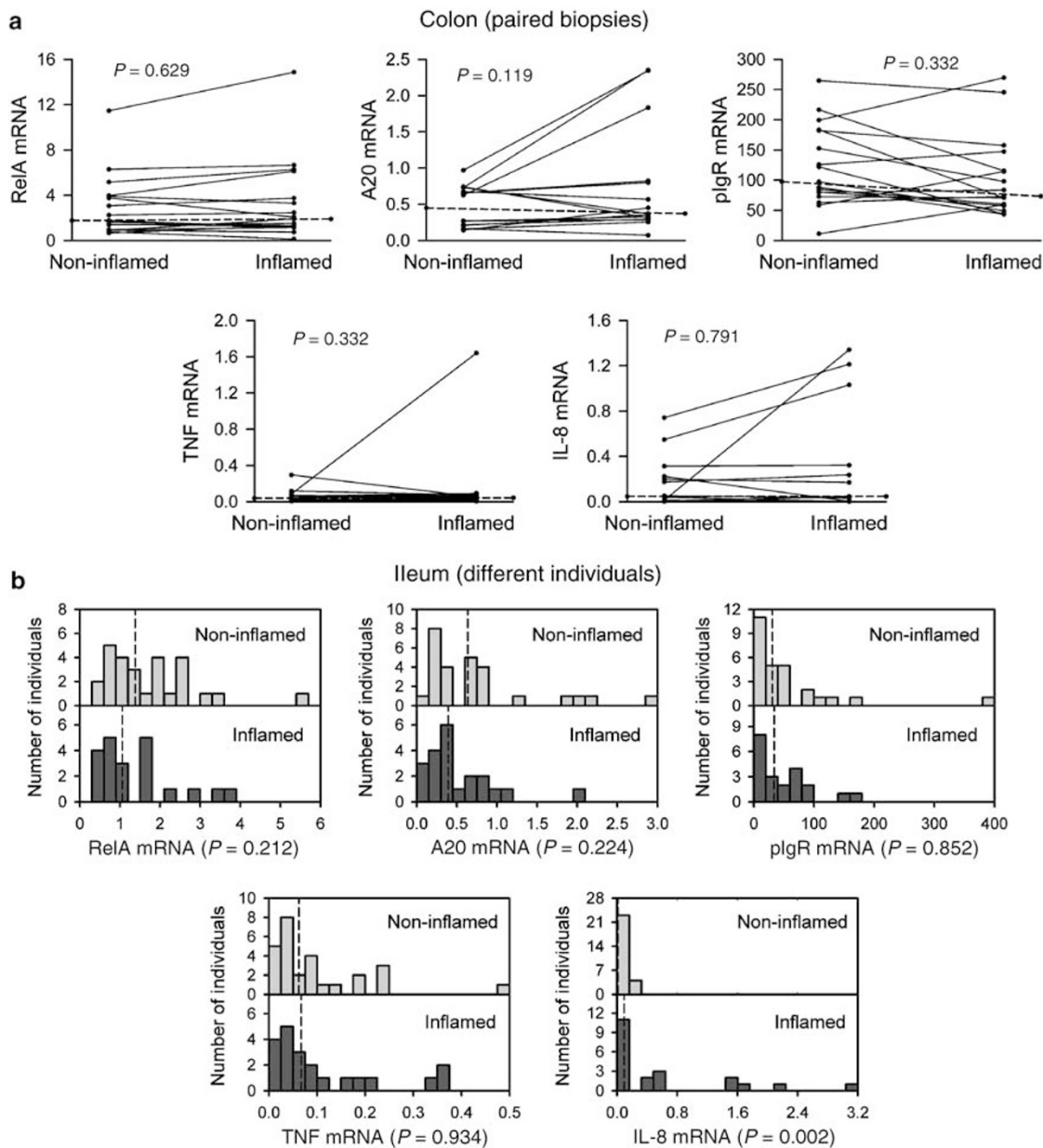
Figure 3 Effects of local inflammation on gene expression. mRNA levels were measured by RT-qPCR and normalized to β2-microglobulin mRNA. (**a**) Paired biopsies were obtained from non-inflamed and inflamed colon mucosa from 17 CD patients. Dashed lines indicate the difference in median expression levels between non-inflamed and inflamed mucosa for each gene. Significant differences among groups were tested by paired sign non-parametric analysis; *P*-values are listed for paired comparisons between inflamed vs. non-inflamed colon mucosae. (**b**) Histograms of gene expression in inflamed or non-inflamed ileum mucosae from different individuals. Because of limited access to the ileum during colonoscopy, a single biopsy from the terminal ileum for each CD patient was classified as either inflamed or non-inflamed, and paired biopsies could not be collected. Significant differences among groups were tested by Mann–Whitney non-parametric analysis; *P*-values are listed for each comparison of individuals with inflamed vs. non-inflamed ileum mucosa. CD, Crohn's disease; RT-qPCR, real-time quantitative PCR.

individuals with a high score for PC2. On the basis of colon gene expression, 81% of normal controls were classified into set 1, with 8% in set 2 and 11% in set 3 (**Figure 4d**). On the basis of ileum gene expression, the percentages of normal controls in sets 1, 2, and 3 were 57, 38, and 5%, respectively. These findings suggested that high scores for PC1 and low scores for PC2, particularly for the colon, define a "normal" molecular phenotype. In contrast, 42% of CD patients were classified in set 1 based on colon gene expression, with 45% in set 2 and 13% in set 3. On the

basis of ileum gene expression, the percentages of CD patients in sets 1, 2, and 3 were 33, 45, and 22%, respectively.

Patterns of colon gene expression differed significantly for CD patients classified into the three molecular phenotypic subsets (**Figure 5**). Individuals in set 2 had significantly reduced expression of RelA, A20, pIgR, and TNF compared to individuals in set 1 (the "normal" phenotype). Individuals in set 3 had significantly reduced expression of RelA, A20, and pIgR, normal expression of TNF, and elevated expression of IL-8. On the basis
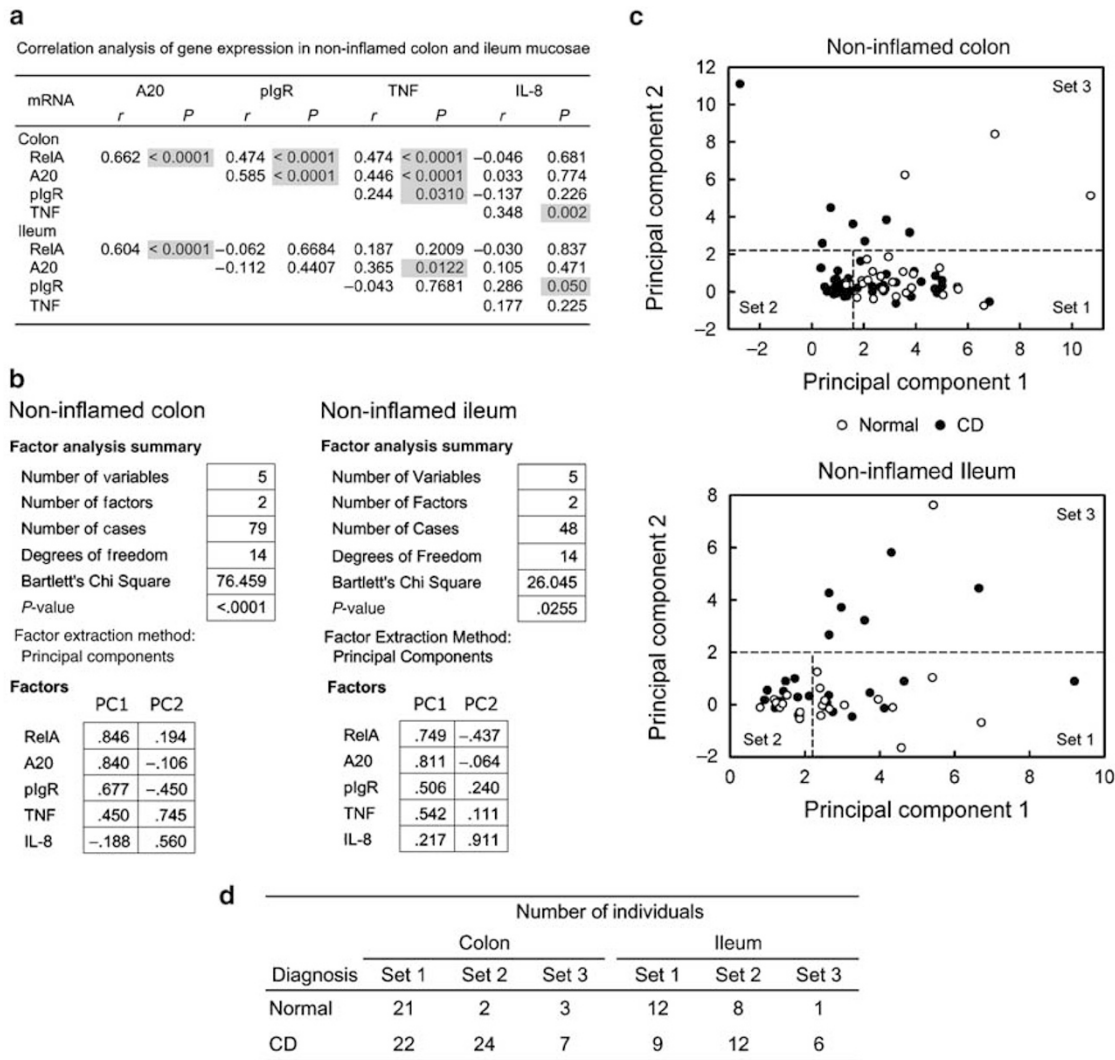
Figure 4 Multifactorial analysis of gene expression patterns. (**a**) Non-parametric Spearman correlation analysis of gene expression data from **Figure 1**. Correlation coefficients (r) and P-values are listed for each comparison. Statistically significant correlations ($P<0.05$) are shaded. (**b**) Factor analysis of NF-κB, RelA, A20, pIgR, TNF, and IL-8 mRNA levels in non-inflamed colon and ileum mucosa, including both normal controls and CD patients. Weighted factors are listed for principal components (PCs) 1 and 2 for each gene. (**c**) Classification of individuals into molecular phenotypic subsets. Scores for PC1 and PC2 were calculated for each individual based on the sum of weighted expression levels for all five biomarkers (see Methods). Set 1 comprises individuals with a high PC1 score and low PC2 score. Set 2 comprises individuals with low scores for both PC1 and PC2. Set 3 comprises individuals with high scores for PC2. The table lists the distribution of normal controls and CD patients among colon and ileum subsets. The scatter plots represent the scores for PC1 and PC2 for each individual; open circle=normal control; filled circle=CD patient. (**d**) Distribution of individuals into molecular phenotypic subsets based on scores for PC1 and PC2. CD, Crohn's disease; IL, interleukin; NF-κB, nuclear factor-κB; PC, principal component; pIgR, polymeric immunoglobulin receptor; TNF, tumor necrosis factor.

of these findings, we propose that CD patients in set 2 have an inherent deficiency in innate immunity in the colon mucosa. CD patients in set 3 had selected features of innate immunodeficiency (reduced RelA, A20, and pIgR), but elevated IL-8 expression in this group was suggestive of acute inflammation. Classification of individuals based on ileum gene expression was less informative than for the colon. CD patients in ileum set 2 had some evidence of innate immunodeficiency (reduced levels of RelA and A20), but expression of pIgR and TNF was normal. CD patients in ileum set 3 had elevated IL-8 expression, but no evidence of innate immunodeficiency. We conclude that RelA, A20, pIgR, TNF, and IL-8 are potentially useful biomarkers for classification of CD patients into molecular phenotypic sub-

sets based on gene expression in non-inflamed colon mucosa. Accordingly, we compared clinical findings in CD patients distributed into the three colon subsets.

Aberrant localization of pIgR and SIgA in a subset of CD patients

Deficient pIgR expression in mice has been associated with reduced transport of SIgA into the colon lumen and elevation of serum IgA.[25,26] We observed that serum IgA levels tended to be lower in CD patients than in normal controls (**Figure 6a**), but given the overall heterogeneity, this difference did not achieve statistical significance. However, serum IgA levels were significantly elevated in CD patients in set 2 (**Figure 6b**), who had
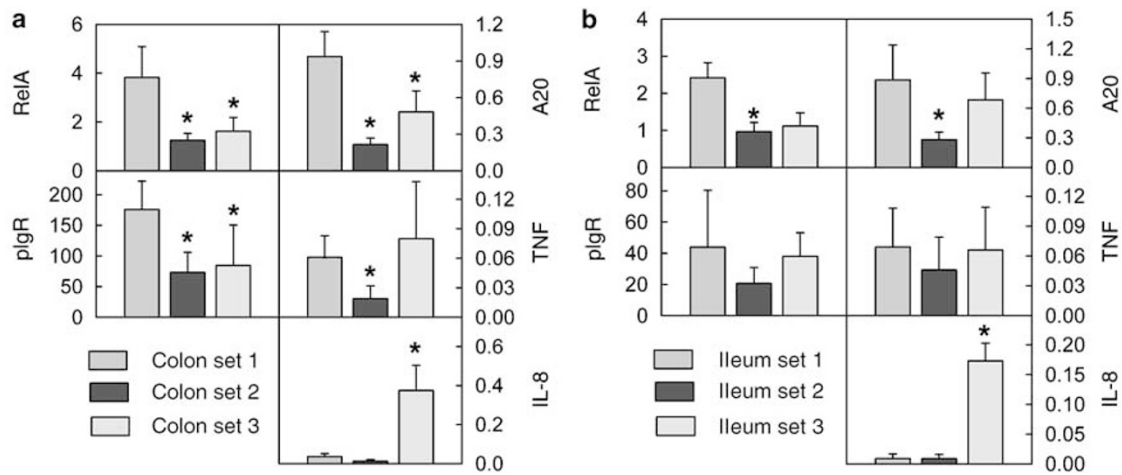
Figure 5 Comparison of individual biomarker expression in CD patients grouped by molecular phenotype. CD patients were classified in sets 1–3 by principal component analysis, as described in **Figure 4**. mRNA levels in colon (**a**) and ileum (**b**) mucosa were measured by RT-qPCR and normalized to β2-microglobulin mRNA. Gene expression levels for sets 1, 2, and 3 were compared by non-parametric Mann–Whitney analysis and are expressed as median+median absolute deviation. Asterisks indicate that gene expression in set 2 or 3 is different from that in set 1 ($P<0.05$). CD, Crohn's disease; RT-qPCR, real-time quantitative PCR.

decreased expression of pIgR in the colonic mucosa (**Figure 5a**). Serum IgA levels also appeared to be elevated in CD patients in set 3, although due to the relatively small number of individuals in set 3 the difference from set 1 was not statistically significant. To examine whether reduced pIgR-mediated transport of IgA contributed to the elevation in serum IgA, we compared the localization of pIgR and IgA in colonic biopsies from a normal individual in set 1 and from a CD patient in set 2 (**Figure 6c**). In the normal colon mucosa, pIgR and IgA were colocalized at the basolateral surface, within the cytoplasm, and at the apical surface of epithelial cells, consistent with normal transport of IgA. In the lamina propria, IgA was observed in association with plasma cells, but no pIgR was detected. In contrast, staining for pIgR was markedly reduced in epithelial cells of non-inflamed mucosa from the set 2 CD patient, and there was reduced evidence of transepithelial transport of IgA. In addition, more IgA was observed in the lamina propria, compared with the normal control. A punctate border of colocalized pIgR/SC and IgA at the apical surface of the epithelium most likely represented SIgA bound to the mucus layer. In the inflamed mucosa from the same CD patient, dense colocalization of IgA and pIgR/SC in the lamina propria was suggestive of deposition of SIgA along the basement membrane. Leakage of excess IgA (either free or bound to SC) from the lamina propria into the systemic circulation could account in part for the elevated levels of serum IgA in set 2 CD patients.

Correlations between molecular phenotypes and clinical outcomes of CD patients

A review of the clinical histories of CD patients in this cohort indicated that, in general, patients in set 1 tended to have mild disease and good responses to therapy; patients in set 2 had moderate to severe disease and poor responses to immunosuppressive and anti-TNF therapy; and patients in set 3 had acute disease that responded promptly and durably to therapy.

Although young age at diagnosis of CD has been reported to be associated with more severe disease,[38] we observed no differences among the three subsets in age at diagnosis (Mann–Whitney test, $P>0.05$). We did find that set 3 patients tended to be more recently diagnosed, independent of the age at diagnosis (**Figure 7**). Set 3 patients also had elevated serum C-reactive protein, a marker of acute inflammation in CD[10] (**Figure 7**). These data suggest that some CD patients may present initially with acute inflammation (set 3), but the inherent innate immunodeficiency characterized by reduced expression of RelA, A20, and pIgR (and likely other genes) results in a failure to resolve the underlying trigger of inflammation. Over time, these patients may progress from acute to chronic inflammation, characteristic of set 2.

Recent evidence suggests that inflammatory mediators produced by visceral adipose tissue may promote intestinal inflammation in CD.[39] However, we observed no differences in body mass index among molecular phenotypic subsets, indicating that excess body fat did not influence expression of the biomarkers we analyzed. There were also no significant differences among subsets in Harvey–Bradshaw index of disease activity[40] at the time of colonoscopy, suggesting that subjective evaluation of disease activity at one point in time may not be informative with respect to the underlying molecular pathogenesis of the disease. Other clinical characteristics that did not affect distribution of CD patients into molecular phenotypic subsets included gender, smoking behavior, and disease location (colon vs. ileum, or both) (**Table 1**).

Crohn's disease patients with an inflammatory or stricturing phenotype (Montreal classification[41]) were evenly distributed among the three molecular phenotypic subsets (**Table 1**). In contrast, patients with penetrating disease were significantly biased toward set 3 and away from set 2 ($P=0.019$). Given that only three patients in our cohort had penetrating disease, these data must be interpreted with caution. However, the possibility that the combination of innate immunodeficiency and acute
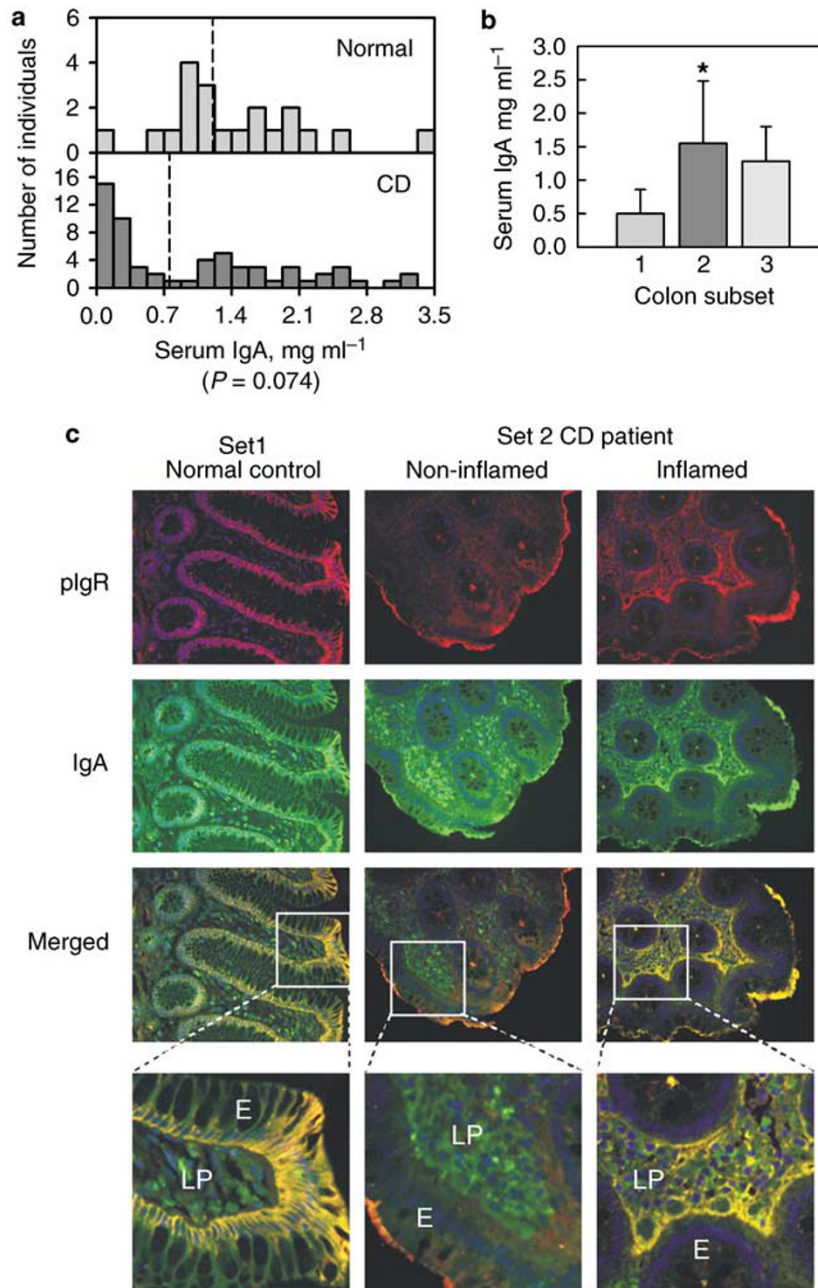
Figure 6 Serum IgA levels and localization of pIgR and IgA in colonic mucosa of normal controls and CD patients. (**a**) Histograms of serum IgA levels in normal controls and CD patients. Dashed lines indicate median levels for each group. Differences in serum IgA levels for normal controls vs. CD patients were tested by Mann–Whitney non-parametric analysis. (**b**) Comparison of serum IgA in CD patients classified in sets 1–3 by principal component analysis, as described in **Figure 4**. Serum IgA levels in sets 1, 2, and 3 were compared by non-parametric Mann–Whitney analysis and are expressed as median + median absolute deviation. An asterisk indicates that gene expression in set 2 or 3 is different from that in set 1 ($p < 0.05$). (**c**) Localization of pIgR/SC and IgA in colonic mucosa by immunofluorescence. Representative images are shown from a normal subject (representing set 1) or matched biopsies from visibly inflamed and non-inflamed regions of the colon from a CD patient with low pIgR mRNA expression (representing set 2). Red staining indicates binding of antibody to human SC and green staining indicates binding of antibody to human IgA. All samples were counterstained with DAPI to visualize nuclei. Stained tissue sections were imaged with a ×20 objective; the bottom panels are enlargements of the designated regions from merged images. CD, Crohn's disease; DAPI, 4′,6-diamidino-2-phenylindole dihydrochloride; E, epithelium; LP, lamina propria; pIgR, polymeric immunoglobulin receptor; SC, secretory component.

inflammation that characterizes set 3 may predispose patients to penetrating disease deserves further investigation.

Only one patient in our cohort (classified in set 1) was homozygous for the L1007fsinsC polymorphism in the *NOD2* gene, which confers increased risk for CD[42] (**Table 1**). Several additional patients were heterozygous for other *NOD2* polymorphisms, but these did not affect their distribution among the molecular phenotypic subsets. These findings suggest that the
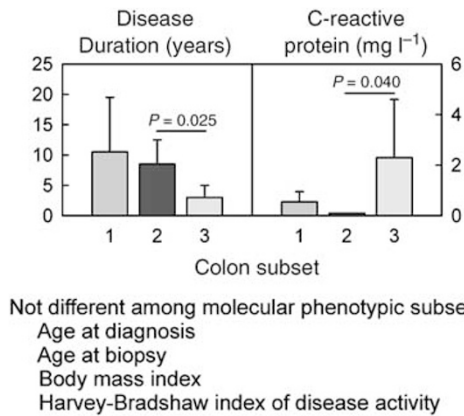
Figure 7 Clinical characteristics of CD patients classified in sets 1–3 based on colon gene expression. Disease duration is defined as the time from initial diagnosis of CD to the time when the mucosal biopsy was collected for each individual. Values for sets 1, 2, and 3 were compared by non-parametric Mann–Whitney analysis and are expressed as median + median absolute deviation. P-values indicate significant differences between CD patients in set 2 and set 3. No significant differences were observed between set 1 and sets 2 or 3. CD, Crohn's disease.

abnormalities in innate immunity identified by our classification scheme are unrelated to *NOD2* polymorphisms.

We next investigated whether the medications that CD patients were receiving at the time of biopsy affected their molecular phenotypes. The χ^2 analysis revealed no significant effect of any medication on distribution of CD patients among the three subsets (**Table 2**). We conclude that the observed abnormalities in mucosal gene expression in sets 2 and 3 reflected an underlying pathology characteristic of each individual, and did not result from acute effects of medications they were receiving at the time of biopsy. This finding does not rule out the possibility that medications received earlier in the course of the disease may have had long-term effects on mucosal gene expression.

DISCUSSION
Signature biomarkers for CD
Here we describe a molecular classification scheme, based on expression of RelA, A20, pIgR, TNF, and IL-8 in the colon mucosa, which identifies unique phenotypic subsets of CD patients. Several findings support the concept that these are

Table 1 Comparison of clinical characteristics and molecular phenotypes of CD patients

| | Number of subjects | | | | | | | |
| | Set 1 | | Set 2 | | Set 3 | | | |
Factor	Observed	Expected	Observed	Expected	Observed	Expected	χ^2	*P*
Gender								
Male	7	8.7	9	9.5	5	7.0	2.153	0.341
Female	15	13.3	15	14.5	2	4.2	1.413	0.493
Smoking behavior								
Smoker	7	6.6	8	7.2	1	2.1	0.684	0.710
Non-smoker	15	13.7	13	14.9	5	4.4	0.471	0.790
Disease location								
Ileum only	7	5.0	4	5.4	1	1.6	1.413	0.493
Colon only	4	4.6	5	5.0	2	1.5	0.276	0.871
Ileum+colon	11	12.0	14	13.1	4	3.8	0.154	0.926
Disease behavior								
Inflammation only	10	8.3	6	9.1	4	2.6	2.078	0.354
Stricturing	11	11.6	14	12.7	3	3.7	0.303	0.860
Penetrating	1	1.2	0	1.4	2	0.4	7.898	0.019
NOD2 *polymorphisms*								
R702W$^{+/-}$	4	2.1	0	2.3	1	0.7	4.223	0.121
G908R$^{+/-}$	0	0.4	1	0.5	0	0.1	1.208	0.547
L1007fsinsC$^{+/+}$	1	0.4	0	0.5	0	0.1	1.409	0.494

Abbreviation: CD, Crohn's disease.
The χ^2 analysis tested the hypothesis that CD patients displaying each of the listed clinical characteristics would be distributed among the molecular phenotypic subsets in the same proportions in which the entire cohort of CD patients was distributed, i.e., set 1:set 2:set 3; 22:24:7 (see Figure 4d).

Table 2 Effects of medications at the time of biopsy on molecular phenotypes of CD patients

Medication	Number of subjects						χ^2	P
	Set 1		Set 2		Set 3			
	Observed	Expected	Observed	Expected	Observed	Expected		
5-ASA								
No	10	13.3	17	14.5	5	4.2	1.388	0.500
Yes	12	8.7	7	9.5	2	2.8	2.114	0.347
Immunosuppressants								
No	14	12.0	14	13.1	1	3.8	2.469	0.291
Yes	8	10.0	10	10.9	6	3.2	2.983	0.225
Steroids								
No	13	14.9	18	16.3	5	4.8	0.442	0.802
Yes	9	7.1	6	7.7	2	2.2	0.937	0.626
Anti-TNF biologics								
No	16	17.8	21	19.5	6	5.7	0.330	0.848
Yes	6	4.2	3	4.5	1	1.3	1.417	0.492

Abbreviation: CD, Crohn's disease.
The χ^2 analysis tested the hypothesis that CD patients receiving each of the listed medications would be distributed among the molecular phenotypic subsets in the same proportions in which the entire cohort of CD patients was distributed, i.e., set 1:set 2:set 3; 22:24:7 (see Figure 4d). Medications: 5-ASA: mesalamine; immunosuppressants: azathioprine, 6-mercaptopurine, methotrexate; steroids: budesonide, prednisone; anti-TNF: Remicade (infliximab), Humira (adalimumab). Some patients were receiving more than one medication at the time of biopsy.

"signature biomarkers" for CD. First, biomarker expression appeared to be an inherent characteristic of each CD patient, and was not affected by local variations in tissue inflammation or current medications. Second, these biomarkers identified a potential state of innate immunodeficiency in the majority of CD patients in our cohort, accompanied by chronic (set 2) or acute (set 3) mucosal inflammation. Third, classification of CD patients into these molecular phenotypic subsets was predictive of disease behavior and responses to therapy. Our findings provide strong support for prospective studies of this molecular classification scheme in a larger population of CD patients.

A model for dysregulated innate immunity in CD

The pathology of CD arises from an imbalanced immune response to exogenous and endogenous molecular cues. Traditionally, CD has been viewed as an exaggerated inflammatory response to microbial antigens caused by a dysregulated adaptive immune response. However, recent evidence suggests that the primary defect in CD may actually be a deficient innate immune response in the intestinal mucosa, resulting in inadequate clearance of bacteria and other foreign material from the lamina propria and generation of a chronic inflammatory state.[8,9] The intestinal epithelium is a crucial player in maintaining this balance, going beyond the simple role of a physical barrier to orchestrate the physiological inflammatory response and the innate immune response to dietary and microbial antigens.

We propose that expression of RelA, A20, pIgR, TNF, and IL-8 may be useful for assessing innate immune function in the colon mucosa. A model describing the activities of these molecules in normal and diseased mucosa is presented in **Figure 8**. In healthy mucosa, cross talk between epithelial cells and constituents of the commensal microbiota is essential for maintaining normal expression of the RelA subunit of NF-κB. Constitutive low-level activation and nuclear translocation of NF-κB result in physiological inflammation, characterized by high expression of pIgR and other proteins involved in epithelial polarity and function, as well as moderate expression of TNF, IL-8, and other effectors of innate immunity. Coordinate expression of negative regulatory molecules such as A20 protects epithelial cells from excessive activation and nuclear translocation of NF-κB. IgA produced by plasma cells in the lamina propria is transported across the epithelial layer by pIgR, thus promoting intracellular and extracellular neutralization of microbes and their products and regulating immune effector molecules such as IL-8.[23,24,43] Reduced expression of RelA in the colon mucosa of some CD patients may compromise the ability of epithelial cells to mount a physiological inflammatory response to the commensal microbiota. Reduced expression of pIgR could inhibit normal transport of IgA, resulting in diminished regulation of microbial products and IL-8, compromised barrier function, and deposition of SIgA complexes in the lamina propria. Low mucosal expression of RelA may also predispose CD patients to periodic flares of pathological inflammation, due to reduced expression of negative regulators such as A20, diminished
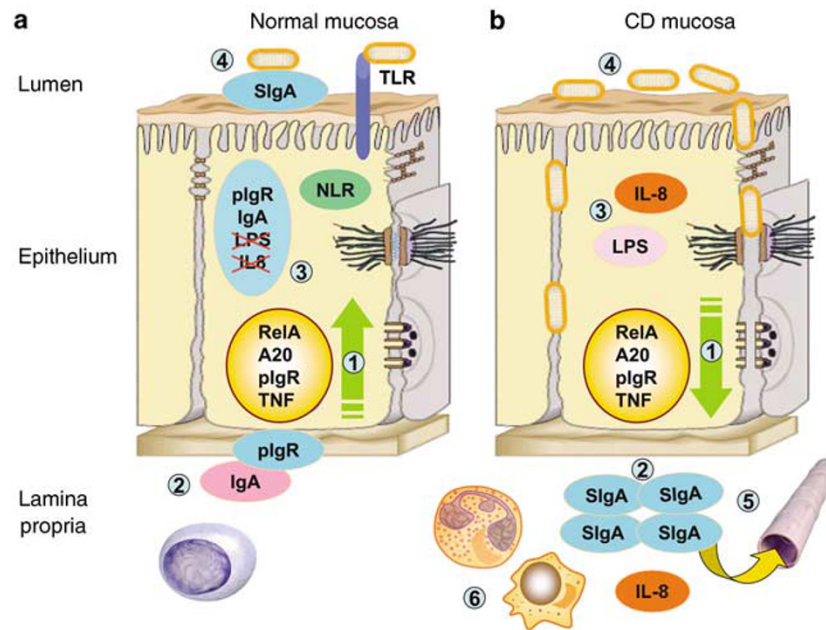
Figure 8 Model for inflammation due to dysregulated mucosal gene expression in Crohn's disease. **(a)** (1) In normal mucosa, bacterial–epithelial cross talk upregulates expression and activation of NF-κB (RelA) and expression of downstream targets (A20, pIgR, and TNF). Modulation of these responses by A20 and other negative regulators limits the "physiologic inflammation." (2) IgA secreted by lamina propria plasma cells binds to pIgR on the basolateral surface of epithelial cells and is transported to the apical surface. (3) During transport, SIgA facilitates clearance of bacterial antigens and inflammatory chemokines. (4) SIgA promotes immune barrier function at the luminal surface. **(b)** (1) Defective expression of NF-κB and downstream genes by epithelial cells in CD. (2) Abnormal deposits of SIgA in the lamina propria. (3) Failure of IgA-mediated clearance of antigens and IL-8 due to defective epithelial transport. (4) Reduced SIgA at the luminal surface results in diminished barrier function. (5) Excess SIgA in the lamina propria enters the systemic circulation. (6) Chronic inflammation with increased IL-8 and influx of neutrophils and macrophages. CD, Crohn's disease; IL, interleukin; NF-κB, nuclear factor-κB; pIgR, polymeric immunoglobulin receptor; SIgA, secretory antibodies of the IgA class; TNF, tumor necrosis factor.

anti-inflammatory activity of SIgA, increased access of microbial antigens to the lamina propria, activation of macrophages and generation of an adaptive immune response.

Protective roles for NF-κB/RelA in the intestinal epithelium
Association of NF-κB signaling with induction of pro-inflammatory cytokines in T cells and inflammatory cells such as macrophages and neutrophils has supported the use of therapies designed to inhibit activation of NF-κB.[1] When considering these therapeutic approaches, it should be kept in mind that physiological NF-κB signaling is important for maintenance of intestinal homeostasis. It has recently been reported that selective inhibition of NF-κB activation in intestinal epithelial cells, by targeted deletion of genes encoding subunits of the IκB-kinase complex, actually increases the susceptibility of mice to experimental colitis.[13,14] A critical role of the RelA subunit of NF-κB in intestinal homeostasis was highlighted by the recent report that mice with a targeted deletion of *Rela* in intestinal epithelial cells exhibited reduced expression of pro-restitution genes, elevated epithelial apoptosis and proliferation, and increased susceptibility to chemically induced colitis.[15] Our data suggest that reduced expression of the RelA subunit of NF-κB in the gut mucosa, with concomitant downregulation of key target genes, may be a contributing factor for the development of CD in humans.

NF-κB signaling in the intestine is kept under tight control by a complex network of negative regulatory molecules, many of which are target genes of NF-κB.[17] To represent this group of negative regulators in our panel of biomarkers, we analyzed expression of A20, a ubiquitin-editing enzyme that downregulates NF-κB signaling initiated by pro-inflammatory cytokines and TLRs.[44] A20 null mice have been shown to develop severe intestinal inflammation,[18] suggesting that A20 and other negative regulators may fulfill the role of a "brake" and set the threshold for NF-κB activation. We consistently observed downregulation of mucosal A20 expression in CD patients, which may hamper their ability to regulate pathological NF-κB activation induced by acute inflammatory responses. Significantly, we found that A20 levels did not increase in inflamed regions of the colon in CD patients (**Figure 3**). Paradoxically, CD may be characterized both by a defect in physiological NF-κB signaling and by a reduced capacity to regulate pathological NF-κB signaling.

Compartmentalization of IgA in the intestinal epithelium is crucial for barrier function
Compartmentalization of IgA antibodies in the gut enhances barrier function by allowing a local antigen-specific response to commensal microorganisms without stimulating potentially inflammatory systemic immune responses.[20] We observed that reduction of pIgR expression in some CD patients (set 2) led to reduced transport of IgA and accumulation of dense membranous deposits of SIgA (**Figure 6**). Elevated serum IgA levels in

this group of CD patients could have resulted from leakage of IgA from the lamina propria into the systemic circulation, which has been reported in pIgR-deficient mice.[25,26] Elevated serum IgA could also result from deficient epithelial barrier function and access of the commensal microbiota to the lamina propria, resulting in a loss of tolerance and a systemic immune response to microbial antigens.

Potential applications of a molecular classification scheme for diagnosis and management of CD

The results of this study support the concept of developing a molecular classification scheme for CD based on patterns of gene expression in the colon mucosa. Using this scheme, we classified a cohort of CD patients into three molecular phenotypic subsets that were generally predictive of clinical findings and responses to therapy. It will be important to validate this model in a larger cohort and to evaluate the utility of additional biomarkers. Molecular classification of CD patients at the time of diagnosis may be helpful in targeting specific therapies. For example, patients classified in set 3, with evidence of acute inflammation, may respond better to TNF blockers and other biological therapies than would patients classified in set 2, who exhibited subnormal expression of TNF in the colon mucosa. This classification system may also be useful for monitoring the course of the disease and modifying therapy over time. We found that recently diagnosed patients were more likely to be classified in set 3, whereas patients with longstanding disease were more likely to be classified in set 1 or set 2 (**Figure 7**), suggesting that patients may undergo transition over time into different subsets. Our finding of reduced mucosal expression of RelA in both set 2 and set 3, as well as the need for physiological NF-κB activation to maintain epithelial barrier function, suggests that the application of therapies designed to inhibit NF-κB activation should be used with caution.

METHODS

Study subjects. Peripheral blood and mucosal biopsies were obtained from individuals undergoing colonoscopy at the University of Kentucky Medical Center, after institutional review board approval and written informed consent. For CD patients, the indication for colonoscopy was either to evaluate disease exacerbation or to screen for dysplasia and colorectal cancer. Diagnosis of CD was based on clinical, radiological, and endoscopic criteria according to the Montreal classification,[41] supported by histopathological findings. Active disease was defined by a Harvey–Bradshaw index[40] >4 or endoscopic evidence of active inflammation. Control subjects aged 50 years or older underwent screening colonoscopies for colon cancer in accordance with current guidelines. Control subjects under 50 years of age underwent colonoscopy for evaluation of constipation or chronic abdominal pain. Individuals were classified as "normal" when endoscopic, radiologic, and pathologic evaluation of randomly obtained biopsies revealed no disease of the small or large bowel.

Analysis of mRNA levels in colonic and ileal mucosa. Biopsied tissue was immediately immersed in an RNA-stabilizing solution (RNAlater; Qiagen, Valencia, CA) and stored at −80 °C. Total RNA was purified using the Qiagen RNeasy Protect mini kit (Qiagen) according to the manufacturer's directions. RNA was reverse-transcribed to generate cDNA templates using the TaqMan Gold RT-PCR kit (Applied Biosystems, Foster City, CA). Specific mRNA levels were quantified by real-time

reverse transcription PCR, using the ABI Prism 7700 Sequence Detection System (Applied Biosystems) as previously reported.[32] The level of β2-microglobulin mRNA, which did not vary between normal subjects and CD patients, was used to normalize mRNA levels for test genes according to the following formula: $(2^{-(C_T\text{test}-C_T\text{β2-microglobulin})}) \times 100\%$. Preliminary studies indicated that gene expression did not vary significantly in different regions of the large bowel (cecum, transverse colon, and rectum) within each individual. Accordingly, biopsies of non-inflamed mucosa from different regions of the colon were pooled into a single sample for mRNA analyses.

Measurement of serum IgA and immunolocalization of IgA and pIgR. Serum IgA levels were measured using a commercial ELISA kit for human IgA (Bethyl Laboratories, Montgomery, TX). Mucosal biopsies were immediately fixed in formalin and then embedded in paraffin. The sections (5–7 μm) were mounted on glass slides, deparaffinized, rehydrated, quenched with 3% H_2O_2, and treated for 10 min with a citric acid-based antigen-unmasking solution (Vector Laboratories, Burlingame, CA). Sections were blocked with normal goat serum (1% in phosphate-buffered saline) and incubated overnight at 4 °C with primary antibodies: monoclonal mouse anti-human IgA, (Santa Cruz Biotechnology, Santa Cruz, CA; catalog no. sc66185) (diluted 1:50) and polyclonal rabbit anti-human SC (the extracellular domain of pIgR)[45] (diluted 1:200). Sections were washed and incubated for 1 h with a mixture of Cy2-conjugated goat anti-mouse IgG and Cy3-conjugated goat anti-rabbit IgG (Jackson ImmunoResearch Laboratories, West Grove, PA; catalog nos. 115-225-146 and 111-165-144). Sections were counterstained with 4′,6-di-amidino-2-phenylindole dihydrochloride (Molecular Probes Invitrogen, Eugene, OR) to visualize nuclei and mounted with VECTASHIELD medium (Vector Laboratories). Stained sections were imaged with an Olympus BX51 microscope, using a ×20 objective. ImagePro software was used to generate the composite images.

Analysis of *NOD2* mutations. Genomic DNA was extracted from whole blood using the AquaPure Genomic DNA Isolation kit (BioRad, Hercules, CA). The R702W, G908R, and 1007fsinsC polymorphisms in the *NOD2* gene were detected by restriction fragment length polymorphism analysis of PCR-amplified DNA, as described.[42]

Statistical analyses. All statistical analyses were performed using StatView software (SAS Institute, Cary, NC). Because biomarker expression did not follow normal Poisson distributions, non-parametric statistical analyses were used for all comparisons. Mann–Whitney analysis was used to test for significant differences among groups of individuals. Paired sign analysis was used to compare differences in gene expression in paired biopsies within individuals. Spearman correlation analysis was used to test for correlations in mRNA levels among different biomarkers. PCA[36] was used to reduce the five variables (biomarkers) into two PCs. Expression levels for biomarkers from non-inflamed colon or ileum mucosa from all subjects in the cohort (normal controls and CD patients) were included in the PCA. The goodness-of-fit of the two PCs for the colon and ileum data sets was tested by χ^2 analysis. To assign individual scores for PC1 and PC2, the expression level for each of the five biomarkers for each individual was first normalized to the mean for that biomarker for all subjects in the cohort. Individual scores for PC1 and PC2 were then calculated as the sum of the normalized value for each biomarker times the weight for that biomarker. Individuals were classified into subsets according to their scores for PC1 and PC2. The χ^2 analysis was used to test the effects of clinical parameters and medications at time of biopsy on distribution of CD patients into phenotypic subsets.

REFERENCES

1. Kucharzik, T. *et al.* Recent understanding of IBD pathogenesis: implications for future therapies. *Inflamm. Bowel. Dis.* **12,** 1068–1083 (2006).
2. Strober, W., Fuss, I. & Mannon, P. The fundamental basis of inflammatory bowel disease. *J. Clin. Invest.* **117,** 514–521 (2007).
3. Xavier, R.J. & Podolsky, D.K. Unravelling the pathogenesis of inflammatory bowel disease. *Nature* **448,** 427–434 (2007).
4. Pallone, F. & Monteleone, G. Mechanisms of tissue damage in inflammatory bowel disease. *Curr. Opin. Gastroenterol.* **17,** 307–312 (2001).
5. Baumgart, D.C. & Sandborn, W.J. Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* **369,** 1641–1657 (2007).
6. Brown, S.J. & Mayer, L. The immune response in inflammatory bowel disease. *Am. J. Gastroenterol.* **102,** 2058–2069 (2007).
7. Atreya, R. & Neurath, M.F. New therapeutic strategies for treatment of inflammatory bowel disease. *Mucosal. Immunol.* **1,** 175–182 (2008).
8. Korzenik, J.R. Is Crohn's disease due to defective immunity? *Gut* **56,** 2–5 (2007).
9. Marks, D.J. & Segal, A.W. Innate immunity in inflammatory bowel disease: a disease hypothesis. *J. Pathol.* **214,** 260–266 (2008).
10. Nikolaus, S. & Schreiber, S. Diagnostics of inflammatory bowel disease. *Gastroenterology* **133,** 1670–1689 (2007).
11. Kumar, A., Takada, Y., Boriek, A.M. & Aggarwal, B.B. Nuclear factor-κB: its role in health and disease. *J. Mol. Med.* **82,** 434–448 (2004).
12. Strnad, J. & Burke, J.R. IκB kinase inhibitors for treating autoimmune and inflammatory disorders: potential and challenges. *Trends Pharmacol. Sci.* **28,** 142–148 (2007).
13. Nenci, A. *et al.* Epithelial NEMO links innate immunity to chronic intestinal inflammation. *Nature* **446,** 557–561 (2007).
14. Zaph, C. *et al.* Epithelial-cell-intrinsic IKK-β expression regulates intestinal immune homeostasis. *Nature* **446,** 552–556 (2007).
15. Steinbrecher, K.A., Harmel-Laws, E., Sitcheran, R. & Baldwin, A.S. Loss of epithelial RelA results in deregulated intestinal proliferative/apoptotic homeostasis and susceptibility to inflammation. *J. Immunol.* **180,** 2588–2599 (2008).
16. Heyninck, K. & Beyaert, R. A20 inhibits NF-κB activation by dual ubiquitin-editing functions. *Trends Biochem. Sci.* **30,** 1–4 (2005).
17. Shibolet, O. & Podolsky, D.K. TLRs in the Gut. IV. Negative regulation of Toll-like receptors and intestinal homeostasis: addition by subtraction. *Am. J. Physiol. Gastrointest. Liver Physiol.* **292,** G1469–G1473 (2007).
18. Lee, E.G. *et al.* Failure to regulate TNF-induced NF-κB and cell death responses in A20-deficient mice. *Science* **289,** 2350–2354 (2000).
19. Kaetzel, C.S. *Mucosal Immune Defense: Immunoglobulin A* (Springer, New York, 2007).
20. Macpherson, A.J., McCoy, K.D., Johansen, F.E. & Brandtzaeg, P. The immune geography of IgA induction and function. *Mucosal. Immunol.* **1,** 11–22 (2008).
21. Kaetzel, C.S. The polymeric immunoglobulin receptor: bridging innate and adaptive immune responses at mucosal surfaces. *Immunol. Rev.* **206,** 83–99 (2005).
22. Kaetzel, C.S. & Bruno, M.E.C. Epithelial transport of IgA by the polymeric immunoglobulin receptor. In *Mucosal Immune Defense: Immunoglobulin A* (Kaetzel C.S., ed) 43–89 (Springer, New York, 2007).
23. Phalipon, A. & Corthésy, B. Novel functions of the polymeric Ig receptor: well beyond transport of immunoglobulins. *Trends Immunol.* **24,** 55–58 (2003).
24. Phalipon, A. & Corthésy, B. Novel functions for mucosal SigA. In *Mucosal Immune Defense: Immunoglobulin A* (Kaetzel C.S., ed) 183–202 (Springer, New York, 2007).
25. Johansen, F.E. *et al.* Absence of epithelial immunoglobulin A transport, with increased mucosal leakiness, in polymeric immunoglobulin receptor/secretory component-deficient mice. *J. Exp. Med.* **190,** 915–922 (1999).
26. Shimada, S. *et al.* Generation of polymeric immunoglobulin receptor-deficient mouse with marked reduction of secretory IgA. *J. Immunol.* **163,** 5367–5373 (1999).
27. Murthy, A.K., Dubose, C.N., Banas, J.A., Coalson, J.J. & Arulanandam, B.P. Contribution of polymeric immunoglobulin receptor to regulation of intestinal inflammation in dextran sulfate sodium-induced colitis. *J. Gastroenterol. Hepatol.* **21,** 1372–1380 (2006).
28. Brown, S.J. & Abreu, M.T. Antibodies to tumor necrosis factor-α in the treatment of Crohn's disease. *Curr. Opin. Drug Discov. Devel.* **8,** 160–168 (2005).
29. Goncalves, N.S. *et al.* Critical role for tumor necrosis factor alpha in controlling the number of lumenal pathogenic bacteria and immunopathology in infectious colitis. *Infect. Immun.* **69,** 6651–6659 (2001).
30. Corredor, J. *et al.* Tumor necrosis factor regulates intestinal epithelial cell migration by receptor-dependent mechanisms. *Am. J. Physiol. Cell Physiol.* **284,** C953–C961 (2003).
31. Rosenstiel, P. *et al.* TNF-α and IFN-γ regulate the expression of the NOD2 (CARD15) gene in human intestinal epithelial cells. *Gastroenterology* **124,** 1001–1009 (2003).
32. Bruno, M.E.C. & Kaetzel, C.S. Long-term exposure of the HT-29 human intestinal epithelial cell-line to TNF causes sustained up-regulation of the polymeric Ig receptor and pro-inflammatory genes through transcriptional and post-transcriptional mechanisms. *J. Immunol.* **174,** 7278–7284 (2005).
33. Kobayashi, Y. The role of chemokines in neutrophil biology. *Front. Biosci.* **13,** 2400–2407 (2008).
34. Schneeman, T.A. *et al.* Regulation of the polymeric Ig receptor by signaling through Toll-like receptors 3 and 4: linking innate and adaptive immune responses. *J. Immunol.* **175,** 376–384 (2005).
35. Bruno, M.E.C., Stefka, A.T., McMahan, A., Thompson, S.N. & Kaetzel, C.S. Colonic bacteria regulate expression of the polymeric immunoglobulin receptor through MyD88-dependent signaling. In *Proceedings of the 13th International Congress of Immunology, Rio de Janeiro (Brazil), August 21–25, 2007* (Kalil J., Cunha-Neto E., Rizzo L.V. eds) 157–160 (Medimond International Proceedings, Bologna, Italy, 2007).
36. Ringnér, M. What is principal component analysis? *Nat. Biotechnol.* **26,** 303–304 (2008).
37. Frank, D.N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. USA* **104,** 13780–13785 (2007).
38. Polito, J.M. *et al.* Crohn's disease: influence of age at diagnosis on site and clinical type of disease. *Gastroenterology* **111,** 580–586 (1996).
39. Schaffler, A., Scholmerich, J. & Buchler, C. Mechanisms of disease: adipocytokines and visceral adipose tissue—emerging role in intestinal and mesenteric diseases. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **2,** 103–111 (2005).
40. Harvey, R.F. & Bradshaw, J.M. A simple index of Crohn's-disease activity. *Lancet* **1,** 514 (1980).
41. Silverberg, M.S. *et al.* Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of a working party of the 2005 montreal world congress of gastroenterology. *Can. J. Gastroenterol.* **19** (Suppl A), 5–36 (2005).
42. Esters, N. *et al.* Transmission of CARD15 (NOD2) variants within families of patients with inflammatory bowel disease. *Am. J. Gastroenterol.* **99,** 299–305 (2004).
43. Marshall, L.J., Perks, B., Ferkol, T. & Shute, J.K. IL-8 released constitutively by primary bronchial epithelial cells in culture forms an inactive complex with secretory component. *J. Immunol.* **167,** 2816–2823 (2001).
44. Wertz, I.E. *et al.* De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-κB signalling. *Nature* **430,** 694–699 (2004).
45. Chintalacharuvu, K.R., Piskurich, J.F., Lamm, M.E. & Kaetzel, C.S. Cell polarity regulates the release of secretory component, the epithelial receptor for polymeric immunoglobulins, from the surface of HT-29 colon carcinoma cells. *J. Cell. Physiol.* **148,** 35–47 (1991).