Functional annotation of a novel **NFKB1**

**promoter polymorphism that increases risk for ulcerative colitis**

Amir S. Karban\(^1,4\), Toshihiko Okazaki\(^1,4\), Carolien I.M. Panhuysen\(^2\), Thomas Gallegos\(^1\), James J. Potter\(^1\), Joan E. Bailey-Wilson\(^3\), Mark S. Silverberg\(^4\), Richard H. Duerr\(^5\), Judy H. Cho\(^6\), Peter K. Gregersen\(^7\), Yuqiong Wu\(^1\), Jean-Paul Achkar\(^8\), Themistocles Dassopoulos\(^1\), Esteban Mezey\(^1\), Theodore M. Bayless\(^1\), Franklin J. Nouvet\(^1\) and Steven R. Brant\(^1,9,\)

\(^1\)The Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center, Gastroenterology Division, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA, \(^2\)Boston University School of Medicine, Boston, MA, USA, \(^3\)National Human Genome Research Institute, National Institutes of Health, Baltimore, MD, USA, \(^4\)Departments of Medicine and Surgery, University of Toronto and Mount Sinai Hospital, Toronto, Canada, \(^5\)Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA, \(^6\)The Martin Boyer Laboratories, Section of Gastroenterology, Department of Medicine, The University of Chicago Hospitals, Chicago, IL, USA, \(^7\)Center for Genomics and Human Genetics, North Shore Long Island Jewish Research Institute, Manhasset, NY, USA, \(^8\)Department of Gastroenterology, The Cleveland Clinic Foundation, Cleveland, OH, USA and \(^9\)Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

Received July 18, 2003; Revised October 20, 2003; Accepted October 29, 2003

Nuclear Factor-κB (NF-κB) is a major transcription regulator of immune response, apoptosis and cell-growth control genes, and is upregulated in inflammatory bowel disease (IBD), both ulcerative colitis (UC) and Crohn’s disease. The **NFKB1** gene encodes the NF-κB p105/p50 isoforms. Genome-wide screens in IBD families show evidence for linkage on chromosome 4q where **NFKB1** maps. We sequenced the **NFKB1** promoter, exon 1 and all coding exons in 10 IBD probands and two controls, and identified six nucleotide variants, including a common insertion/deletion promoter polymorphism (−94ins/delATTG). Using pedigree-based transmission disequilibrium tests, we observed modest evidence for linkage disequilibrium (LD), independent of linkage, between the −94delATTG allele and UC in 131 out of 235 IBD pedigrees with UC offspring (\(P=0.047–0.052\)). This allele was also more frequent in the 156 non-Jewish UC probands from the 235 IBD pedigrees than in 149 non-Jewish controls (\(P=0.015\)). The −94delATTG association with UC was replicated in a second set of 258 unrelated, non-Jewish UC cases and 653 new, non-Jewish controls (\(P=0.021\)). Nuclear proteins from normal human colon tissue and colonic cell lines, but not ileal tissue, showed significant binding to −94insATTG but not to −94delATTG containing oligonucleotides. **NFKB1** promoter/exon 1 luciferase reporter plasmid constructs containing the −94delATTG allele and transfected into either HeLa or HT-29 cell lines showed less promoter activity than comparable constructs containing the −94insATTG allele. Therefore, we have identified the first potentially functional polymorphism of **NFKB1** and demonstrated its genetic association with a common human disease, ulcerative colitis.

\(^4\)The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Human Molecular Genetics, Vol. 13, No. 1 © Oxford University Press 2004; all rights reserved
INTRODUCTION

Ulcerative colitis (UC) and Crohn’s disease (CD) are idiopathic, chronic, frequently disabling, inflammatory bowel diseases (IBD) (12). UC is characterized by mucosal inflammation limited to the colon, always involving the rectum and a variable extent of the more proximal colon in a continuous manner. CD inflammation is transmural, most often discontinuous and may involve any portion of the gastrointestinal tract but most commonly involves the distal ileum. The prevalence of IBD in the United States is 200–300/100,000 with a similar prevalence for UC and CD (3). IBD is considered a complex genetic disorder predicted to involve multiple genes of relatively low penetrance, since the familial patterns of inheritance do not conform to simple Mendelian models (4). Overall, 10–20% of individuals with IBD report one or more additional relatives with IBD (5). Relatives of CD patients have an 10-fold risk of developing CD. Relatives of UC patients have an 8-fold risk of developing UC. However, these diseases appear to be genetically related, as relatives of UC patients have a 2-fold risk of developing CD (6). Relatives of CD patients have a 4-fold risk of developing UC and relatives of CD patients have a 2-fold risk of developing UC (6).

An important candidate gene for IBD is the NFKB1 gene located at chromosome 4q24 (7,8). Nuclear Factor-κB (NF-κB) proteins are a family of transcription factors that regulate various biological defense processes, most notably innate and adaptive immune responses, acute phase reaction and apoptosis (9). There are five members of the NF-κB family in mammals: p50/p105, p65/RelA, c-Rel, RelB and p52/p100. Although many dimeric forms of NF-κB have been detected, the major form of NF-κB is a heterodimer of the p50 and p65/RelA subunits, encoded by the genes NFKB1 and NFKB2, respectively (10). Human NFKB1 encodes two proteins, a 105 kDa, non DNA-binding, cytoplasmic molecule (p105) and a 50 kDa DNA-binding protein (p50) that corresponds to the N-terminus of p105. The NFKB1 gene spans 156 kb and has 24 exons with introns varying between 40,000 and 323 bp in length (Fig. 1) (11).

In most cells before stimulation, NF-κB primarily resides in the cytoplasm in inactive complexes through association with a sequestering inhibitory protein, termed IκB (12). A wide range of stimuli, including bacterial and viral products, cytokines and oxidant-free radicals, activate NF-κB (9). These stimuli promote NF-κB nuclear translocation by a mechanism that involves IκB phosphorylation and the ubiquitin-proteosome pathway. This phosphorylation appears to target IκB for degradation and leads to its dissociation from the NF-κB complex and subsequent translocation of NF-κB to the nucleus (13). There, active NF-κB binds to genomic DNA at promoter regions and thereby regulates gene transcription.

Inappropriate activation of NF-κB has been implicated in inflammation associated with a variety of human diseases and pathologic conditions, among them asthma, inflammatory arthritis, septic shock, lung fibrosis, diabetes, cancer, AIDS, atherosclerosis, stroke and IBD (9,10). Furthermore, several anti-inflammatory and anti-cancer drugs work in part through inhibition of NF-κB activation (9,14). For example, aspirin and glucocorticoids inhibit NF-κB (15,16). Consistent with NF-κB regulation of genes involved in the immune and inflammatory responses, mice null for several of the NF-κB subunits show defects in clearing bacterial infection along with defects in B-cell and T-cell functions (17).

NF-κB has a central pathogenic role in chronic intestinal inflammation (18). Using immunohistochemistry methods, in the inflamed intestinal mucosa of CD and UC patients, activated NF-κB was increased and found localized to the macrophages and epithelial cells (19). Schreiber et al. (20) similarly found CD and UC patients had increased NF-κB activity in intestinal lamina propria cells. Additionally, the therapeutic properties of mesalazine and sulfasalazine (the most common specific medical therapies for mild to moderate UC), rely in part on inhibition of NF-κB activation (21,22). Three CD associated mutations in the NOD2/CARD15 gene on chromosome 16 all have a defect in their ability to activate NF-κB (23). Recent evidence suggests that this may result in a defect in the innate immune system's ability to protect the gut against invasive bacteria (24).

Interestingly, the major locus, cdc51, for the severe colitis phenotype of C3H/HeBir-IL10 knockout mice is located where the mouse (nfkbl) homolog to human NFKB1 maps (25), and thus nfkb1 has been proposed as a candidate gene for this mouse model and NFKB1 as a candidate gene for human colitis (26). In our 1998 North American genome-wide screen in multiplex IBD pedigrees (27), there was evidence for linkage present on chromosome 4q24 where NFKB1 maps (multipoint non-parametric logarithm of the odds, MLOd = 1.71, P = 2.5 × 10⁻³). Evidence for linkage in this region was greater for the ‘mixed’ families (containing at least one UC and one CD patient); the uncorrected MLOd was 2.76 (P = 1.9 × 10⁻⁴). A British/German (28) and a Canadian (29) IBD genome-wide screens both found evidence to support linkage in the same overall region, in UC sibling pairs and ‘all IBD’ pedigrees, respectively.

Because of the physiological relevance of NFKB1 to IBD, the linkage evidence between IBD or UC and the NFKB1 locus region in genome-wide screens, and localization of a mouse model of colitis to the mouse nfkb1 locus, we examined NFKB1 as a candidate gene for IBD, and in particular UC.

RESULTS

Polymorphism detection in the NFKB1 gene

We sequenced the NFKB1 promoter, exon 1 and all 23 coding exons and their flanking introns (Fig. 1, top) using DNA from 12 unrelated subjects: two Centre d’Etude du Polymorphisme Humain (CEPH) controls and 10 probands from pedigrees with the greatest evidence for linkage, as noted by maximal family non-parametric linkage (NPL) scores (30), to the NFKB1 region in our 1998 IBD genome screen (27). Six nucleotide variations were detected (Table 1). Five were novel: an insertion/deletion polymorphism of four bases in the 5′ promoter region (−94ins/delATTG) (Fig. 1, bottom), an exon 1 polymorphism located within the 5′ untranslated region of NFKB1 message and three intronic variants. A previously described exon 12+77 C > T silent polymorphism was also observed.

Genetic association of the NFKB1 promoter −94delATTG allele with UC

Of the six variations detected, only the −94ins/delATTG allele appeared to have a potential functional role. It involved the
deletion of multiple nucleotides and is located between two putative key promoter regulatory elements (Fig. 1), the most proximal is a functional kB binding site located 19 base pairs 3′ (31,32). Therefore, we analyzed the ∆94ins/delATTG polymorphism in 235 singleton and multiplex IBD pedigrees for association with UC, CD or IBD phenotype.

TDT analysis using the program Genehunter 2.1 (33,34) showed 100 transmissions to 71 non-transmissions of the ∆94delATTG (D) allele to UC offspring (P = 0.027, Table 2). There was also increased transmission of the D allele in all IBD pedigrees (206 to 170) although this trend did not reach the 0.05 level of significance. There was no evidence for association with the CD phenotype. The results of non-parametric linkage analysis using Genehunter 2.1 on the 126 families (96 informative) that contained either one or more siblings or other non-parent child IBD affected relative pairs showed slight evidence of linkage for the IBD phenotype (NPL 1.7; P = 0.04).

Two additional TDT programs were used, Family Based Association Test (FBAT) (35) and the Pedigree Disequilibrium Test (PDT) (36). Both packages provide valid tests of LD independent of linkage using different analytic schemes. Using the different analytic outcomes provided, both tests showed borderline significant LD evidence for the association of the D allele with the UC phenotype, independent of linkage (FBAT, P = 0.052; PDT, global score sum P = 0.047) (Table 2). The PDT also showed significant evidence for IBD (P = 0.035, Table 2).

Case–control analysis and replication of the ∆94delATTG association with UC

Based on these TDT results, we extended the study to compare the frequency of the D allele and DD genotype in cases and controls. Such case–control studies are frequently more powerful measures of allelic association (depending on allele frequencies) than comparable TDT analyses (37). These analyses were also performed to determine the specific genotypes [homozygote insertion or wildtype (WW), heterozygote (WD) or homozygote deletion (DD)] that resulted in increased risk of the D allele in our UC and IBD pedigrees. The D allele was more frequent in non-Jewish, unrelated UC or IBD probands (from the same IBD pedigrees examined in the TDT analyses) than among ethnically matched controls (P = 0.015 and 0.014, respectively; Table 3).
The −94ins/delATTG polymorphism influences nuclear protein binding to the NFKB1 promoter

Electrophoretic Mobility Shift Assays (EMSA) were performed to assess if the −94ins/delATTG polymorphism is within a binding domain for nuclear proteins. Oligonucleotides that contained the wildtype sequence (‘W’) showed strong binding to nuclear protein extracted from two human colonic epithelial cell lines, CaCo2 and HT-29 (Fig. 2A). In contrast, the deletion oligonucleotide (‘D’) showed no binding. The ‘DL’ oligonucleotide (containing only a single ATTG deletion allele but with four additional NFKB1 nucleotides added 5’ to 3’ to make a deletion oligonucleotide with the same length as ‘W’) appeared to allow minimal binding of proteins of similar mobility. This binding, however, was markedly less than that of the ‘W’ oligonucleotide.

To assess the specificity of the observed DNA–protein interaction, mutations were made of the tandem ATTG residues at the polymorphic site. Mutating the most 5’ ATTG to CAGT (Fig. 2B, lane 4) resulted in a near complete loss of binding to HeLa-cell derived nuclear protein (as compared to the ‘W’ oligonucleotide, lane 1), whereas mutating the second (i.e. 3’) ATTG to CAGT resulted in no detectable binding (lane 5). Furthermore, mutating the first ‘T’ of the second ATTG at this position (lane 6) decreased the binding of this oligonucleotide. The point mutation L118P, which has been reported to affect DNA binding, was also analyzed. It appeared that mutation at this position decreased the binding of the deletion oligonucleotide (Fig. 2B, lane 7) compared to its wildtype counterpart (lane 10).

The NFKB1 promoter-luciferase reporter constructs show decreased promoter activity for the deletion polymorphism in transient transfection experiments

HeLa and HT-29 cells were transiently transfected with either pGL3-W or pGL3-D reporter constructs (Fig. 3A). These constructs contained 736 bp of the 3’ region of the NFKB1 promoter with the W allele (ATTG3) or 732 bp of the same region with the D allele (ATTG1). Each construct also included the most 5’ 245 bp of exon 1 (i.e. the same sequences as shown in Fig. 1). The regions cloned into both constructs include the
Activator Protein-1 (AP-1) and κB nuclear protein binding consensus sequences in the promoter, and the putative HIP-1, Housekeeping Initiator Protein I, motif in exon 1. These transcriptional regulatory elements of \(NFKB1\) have been previously identified and shown to be important for \(NFKB1\) gene promoter activity (31,32). The constructs did not include the exon 1+252C>G polymorphism sequence. The thymidine kinase (TK) promoter-Renilla luciferase plasmid (phRL-TK) was co-transfected to control for differences in transfection efficiency.

\(pGL3-D\) transfected HeLa cells showed significantly reduced relative luciferase activity at baseline (\(pGL3-W, 3.28 \pm 0.08\) versus \(pGL3-D, 2.52 \pm 0.26\), \(P = 0.005\); Fig. 3B). Incubation for 6 h with 1 μg/ml lipopolysaccharide extract (LPS), a potent activator of both NF-κB and \(NFKB1\) transcription (38–40), markedly increased relative luciferase activity by more than 3-fold from baseline for both \(pGL3-W\) and \(pGL3-D\) transfected HeLa cells. Yet LPS stimulated \(pGL3-D\) activity remained significantly lower than stimulated \(pGL3-W\) activity (Fig. 3B). At 24 h of LPS exposure, \(pGL3-W\) but not \(pGL3-D\) transfected HeLa cells showed a further increase in relative luciferase activity from that observed at 6 h. In fact, \(pGL3-W\) induced relative luciferase activity was 82% greater than \(pGL3-D\) relative luciferase activity at 24 h of LPS exposure (\(pGL3-W, 17.46 \pm 0.34\) versus \(pGL3-D, 9.57 \pm 0.16\), \(P < 0.0001\)).

For HT-29 colonic epithelial cells, baseline relative luciferase activity for both constructs was very low (<5% of HeLa cells). There was a slight, but non-significant decrease in the \(pGL3-D\) transfected versus \(pGL3-W\) transfected cells. Similar to that observed for the HeLa cells, transfected HT-29 \(pGL3-W\) activity was significantly higher than \(pGL3-D\) activity following 6 h of LPS stimulation. Higher \(pGL3-W\) relative luciferase activity was most pronounced following 24 h of LPS stimulation (\(pGL3-W, 4.42 \pm 0.21\) versus \(pGL3-D, 2.85 \pm 0.15\), \(P = 0.0001\)). Transfected \(pGL3\)-basic vector plasmid alone showed <0.01% of the relative luciferase activity as compared to the \(pGL3-W\) or \(pGL3-D\) constructs at baseline and after LPS stimulation for both cell types (data not shown).

### DISCUSSION

\(NFKB1\) encodes the genes for the p50 and p105 NF-κB isoforms, ubiquitous transcription regulators important for multiple diseases and pathological states associated with inflammation and immunity, including IBD. \(NFKB1\) is therefore a candidate gene for IBD, and particularly UC, given the increased linkage evidence observed for the region of chromosome 4q24 containing \(NFKB1\) in UC or mixed pedigrees and given that an important mouse colitis model links to the region of mouse \(nfkbl\). Of six nucleotide variations detected from probands with increased linkage evidence to the region, we chose to further analyze a 4 bp promoter polymorphism, \(−94\)ins/delATTG, because it produced a relatively large sequence change and due to its location proximal to binding sites important to promoter regulation. Ultimately, we observed that promoter-exon 1 constructs that contained the ATTG deletion (\(D\)) allele showed significantly reduced promoter activity in vitro. This was particularly pronounced following 24 h of exposure to LPS, a potent activator of NF-κB. We further observed that nuclear protein extracts from HT-29 human colonic epithelial cells and from HeLa cell lines, and extracts from mucosal biopsies from normal human colon tissues bound avidly and specifically to ATTG insertion (\(W\)) containing oligonucleotides. Conversely, nuclear proteins bound only weakly—or not at all—to ATTG deletion containing oligonucleotides (\(D\)). These results suggest that the \(−94\)ins/delATTG polymorphism may: (i) affect promoter activity of the \(NFKB1\) gene, particularly following stimulation of the innate immune system by bacterial cell wall components (e.g. LPS); and (ii) contain nucleotides that, depending on the specific allele, differentially bind to an unidentified nuclear protein. Whether or not potential up-regulation of \(NFKB1\) promoter activity by nuclear protein binding to the \(W\) and not the \(D\) allele accounts for the observed differences in \(NFKB1\) in vitro promoter activity or whether the differences in activity is independent of this binding will require further experimentation. This will likely involve the identification of the nuclear protein that binds well to the \(W\) and not the \(D\) oligonucleotides. For our purposes in this genetic study, the major importance of these

### Table 3. Case–control analyses

<table>
<thead>
<tr>
<th>Set</th>
<th>Phenotype</th>
<th>No.</th>
<th>WW</th>
<th>WD</th>
<th>DD</th>
<th>D allele frequency</th>
<th>P-value, a D allele frequency</th>
<th>P-value, a DD versus DW or WW</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Controls-NJ</td>
<td>149</td>
<td>0.389</td>
<td>0.470</td>
<td>0.141</td>
<td>0.376</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UC-NJ</td>
<td>92</td>
<td>0.283</td>
<td>0.478</td>
<td>0.239</td>
<td>0.478</td>
<td><strong>0.015</strong></td>
<td><strong>0.040</strong></td>
</tr>
<tr>
<td></td>
<td>CD-NJ</td>
<td>74</td>
<td>0.311</td>
<td>0.527</td>
<td>0.162</td>
<td>0.426</td>
<td>0.170</td>
<td>0.410</td>
</tr>
<tr>
<td></td>
<td>IBD-NJ</td>
<td>156</td>
<td>0.295</td>
<td>0.481</td>
<td>0.224</td>
<td>0.465</td>
<td><strong>0.014</strong></td>
<td><strong>0.041</strong></td>
</tr>
<tr>
<td></td>
<td>Controls-J</td>
<td>142</td>
<td>0.430</td>
<td>0.458</td>
<td>0.113</td>
<td>0.342</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UC-J</td>
<td>39</td>
<td>0.308</td>
<td>0.513</td>
<td>0.179</td>
<td>0.436</td>
<td>0.088</td>
<td>0.182</td>
</tr>
<tr>
<td></td>
<td>CD-J</td>
<td>48</td>
<td>0.417</td>
<td>0.438</td>
<td>0.146</td>
<td>0.365</td>
<td>0.385</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td>IBD-J</td>
<td>79</td>
<td>0.354</td>
<td>0.506</td>
<td>0.139</td>
<td>0.392</td>
<td>0.167</td>
<td>0.353</td>
</tr>
<tr>
<td>B</td>
<td>Controls-NJ</td>
<td>653</td>
<td>0.369</td>
<td>0.481</td>
<td>0.150</td>
<td>0.391</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UC-NJ</td>
<td>258</td>
<td>0.318</td>
<td>0.477</td>
<td>0.205</td>
<td>0.444</td>
<td><strong>0.0212</strong></td>
<td><strong>0.0285</strong></td>
</tr>
<tr>
<td>A + B</td>
<td>Controls-NJ</td>
<td>802</td>
<td>0.372</td>
<td>0.479</td>
<td>0.148</td>
<td>0.388</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>UC-NJ</td>
<td>350</td>
<td>0.309</td>
<td>0.477</td>
<td>0.214</td>
<td>0.453</td>
<td><strong>0.0020</strong></td>
<td><strong>0.0043</strong></td>
</tr>
</tbody>
</table>

NJ, non-Jewish; J, Ashkenazi Jewish.

a P-values are for comparisons between UC, CD or IBD cases and ethnically matched controls for each set.

Numbers in bold denote P-values ≤ 0.05.
cellular findings is that the −94ins/delATTG polymorphism has evidence from two independent functional assays, in vitro promoter activity and differential nuclear protein binding, that the specific allele inherited likely has functional consequences. The −94ins/delATTG polymorphism thus represents the first potential functional NFKB1 polymorphism. Its association with diseases (like UC, that appear to be mediated by NF-κB) is therefore of greater interest than the potential association of non-functional NFKB1 polymorphisms.

We tested the −94ins/delATTG polymorphic alleles for association with UC, CD and IBD, initially in 235 pedigrees containing one or more affected offspring. Using several different analytic schemes, the D allele was observed to be in LD with the UC phenotype. However, the association was of borderline significance, perhaps because of the limited sample size given the modest transmission to non-transmission ratio. To strengthen our findings, we compared the allele and genotype frequencies in probands with those of controls. We found that there was stronger evidence of D allele association with UC using this method. The fact that our TDT results are consistent with the case–control results suggests that the observed case–control association is unlikely to be secondary to population stratification between cases and controls, as the TDT use of within family controls precludes this potential problem. The case–control association required separation of non-Jewish and Jewish Caucasian cases and controls because we observed that allele frequencies were different for controls based on ethnicity. The non-Jewish results were significant, yet the Jewish results were not, perhaps secondary to small sample size and/or a weaker genetic effect. Nonetheless, the trend of greater D alleles in UC cases as compared to controls was similar for both the Jewish and non-Jewish populations studied. It is more expected for a potential functional polymorphic allele that associations will be present for diseases independent of ethnicity, although this finding is not always observed, even for established associations. For example, the functionally
demonstrated 702Trp NOD2 allele (also less common in Jewish than non-Jewish Caucasian patients) has been observed to be even less common in Jewish CD patients than controls (23). However, the importance of the NFKB1 promoter polymorphism in the Jewish population will remain uncertain, until it is studied in a larger set of subjects.

We replicated the −94delATTG–UC association using an independent, second set of non-Jewish UC cases and healthy controls. The overall odds ratio (calculated from both sets of samples) of the DD homozygote genotype was modest (odds ratio 1.59). The modest genotypic risk observed fits with models of inheritance proposed for complex genetic disorders; multiple low penetrant risk alleles of different genes have been hypothesized to account for overall genetic risk (41). The weak linkage evidence found in the family samples is not surprising (and may be even greater than expected) given the low odds ratio of the risk genotype. There can be other polymorphisms on other genes in the region and even within non-coding regions of NFKB1 that may be functional and contribute even greater risk to developing UC, yet this would not invalidate our observations that −94delATTG is associated and functional. The heterozygote (WD) genotype was not associated with IBD risk. This suggests that a single W allele may abrogate risk from (and be dominant over) the UC associated D allele.

The *in vitro* promoter expression studies suggest that the D allele may result in relatively decreased NFKB1 message and hence decreased p50/p105 NF-κB protein production. This is in contrast to our initial expectations, since UC has been associated with increased levels of NF-κB. It is noted that parallel findings have been observed for the CD associated NOD2 mutations: *in vitro* studies in NOD2 transfected cell lines show that NOD2 mutations result in a decrease rather than an expected increase in NF-κB activity (23). Recently, mutant NOD2 has been shown to be defective in clearing invasive bacteria in comparison to wildtype NOD2 (24). Thus, it has been hypothesized that poor activation of NF-κB may weaken the normal cellular defenses against intestinal bacteria by the innate immune system. This defect may allow bacteria that cross the intestinal lumen to not be properly cleared by the immune system, and hence contribute to ongoing intestinal inflammation characteristic of CD. Thus a potential explanation of decreased NFKB1 D allele gene expression may be that a resulting decrease in NF-κB p50/p65 heterodimers, major mediators of inflammation, could decrease the ability of the colon to be protected from colonic bacteria.

Support for this concept is that p50 deficient mice have been found to have greater susceptibility to infection from some (*Listeria monocytogenes* and *Streptococcus pneumoniae*) but not all (*Haemophilus influenzae* and *Escherichia coli*) types of bacteria (42).

A second, and perhaps more provocative explanation of how reduced NFKB1 gene expression may result in increased risk of UC, is that p50—which, unlike p65, does not contain a transactivation domain—can in some cases inhibit inflammation: p50 homodimers (also products of NFKB1) may be involved in blocking p65 dimers from binding to promoters and activating genes involved in inflammatory cascades (43). In mouse macrophage cell lines, p50 over-expression was shown to inhibit tumor necrosis factor-α (*TNF-α*) gene expression, and the mechanism of p50 inhibition appears to depend on NF-κB binding sites, found within the TNF-α promoter, that have preferential affinity for p50 homodimers (39). Overexpression of p50 homodimers has also been suggested to be the mechanism of LPS refractoriness following repetitive stimulation of mononuclear phagocytes (44). Nonetheless, p50 may have dual roles as p50-deficient mice are refractory to the induction of arthritis models (45), and p50 alone can stimulate C-reactive protein expression, although this induction is considerably less than p50/p65 heterodimer stimulation (46). Our functional studies with the promoter polymorphism use an *in vitro* model system, and the actual effect of the polymorphism on NFKB1 message and protein expression will ultimately require studies using human tissues.

It is unlikely that additional polymorphisms or mutations in the promoter, exon 1 and exon 2 regions evaluated will be found to account for the observed D association with UC. To examine this, we sequenced these same regions using DNA from an additional 12 unrelated patients all with DD genotypes. However, no additional polymorphisms were observed. Hence, including the seven D chromosomes reported in Table 1, we find only the −94ins ATTG and exon 1+252C>G polymorphisms in 31 D containing chromosomes. Additionally, these two polymorphisms are found in complete LD. The two most common haplotypes (−94insATTG–exon 1+252C and −94delATTG–exon 1+252G) were observed in 72 out of 74 total chromosomes genotyped. Therefore, either polymorphism will yield essentially equivalent information for testing potential NFKB1 promoter/exon-1 associations, and genetically, it is unlikely that the D association with UC can be easily separated from the expected corresponding exon 1+252G association with UC. In future studies it will be useful to extend our promoter/exon-1 luciferase construct studies to include the exon 1+252 polymorphism region and test constructs that contain the D and W alleles with either the C or G exon 1+252 alleles, to determine if the exon 1+252C>G polymorphism also has an effect on gene expression.

It is also unlikely that there exist common, functionally relevant, NFKB1 coding polymorphisms in exons 3 through 24. Although we only screened a modest number of chromosomes, Wintemeyer et al. (47) screened 96 Parkinson’s disease patients and Mitorski et al. (48) screened a large number (exact figures not reported) of multiple sclerosis patients and healthy controls (apparently 100 controls given the Leu614Phe frequency noted below) for NFKB1 coding polymorphisms/mutations. These studies both used very high sensitivity methods of single-strand conformation polymorphism analysis (SSCP) and reported success in completely screening all exons, except for exons 1 and 2. The two studies observed the relatively common exon 12+77C>T silent polymorphism, a rare exon 8 silent polymorphism in one Parkinson’s disease patient and a Leu614Phe exon 17 mutation in 0.5% of controls in the multiple sclerosis study.

The genetic contribution to the pathogenesis of UC remains largely unclear. While genome-wide searches have identified several loci in linkage with the disease, case-control studies have only shown a reproducible association between UC and HLA class II genes, especially DRB1*0103 and DRB1*15 (49). Although most studies have focused on HLA class II genes, there is an increasing interest in the role of cytokines in UC pathogenesis and on the polymorphic genes that may...
influence cytokine secretion (50). NFKB1 may be the first of perhaps several modest UC risk genes that are involved with these pathways. Other cytokine regulators of the pathway, and ultimately NF-κB protein activation, include interleukin 1 receptor antagonist (IL1RN) and IκB-like gene (NFKBIL1). There has been evidence, albeit inconsistent, for an association of allele 2 of IL1RN, the gene that encodes the interleukin 1 receptor antagonist (51,52) and preliminary evidence of an association of NFKBIL1 with UC (53). It will be interesting to determine if these associations may be clarified by examining for evidence of epistasis with the NFKB1 -94ins/delATTG polymorphism.

NFKB1 is a candidate gene for numerous other inflammatory diseases and risk for immune-mediated conditions. An association was reported between an NFKB1 microsatellite and type 1 diabetes (54) but could not be replicated (55). No associations were found with NFKB1 and the exon 12+77C>T polymorphism for multiple sclerosis or Parkinson’s disease (47,48). LD is likely incomplete between the exon 12 SNP and the -94delATTG polymorphism. Therefore, the -94ins/delATTG polymorphism should be tested in these and other NF-κB mediated complex genetic disorders, particularly since we have provided initial evidence that this polymorphism may have functional attributes and appears to be an important risk factor for one immune mediated, complex genetic disorder, ulcerative colitis.

MATERIALS AND METHODS

Subjects for the TDT and case–control studies

From all study subjects, informed consent for participation in molecular genetic studies was obtained and ethical approval was given from each center’s institutional review boards. In addition, DNA samples from two CEPH controls (133101, 133102) were obtained from Coriell Institute for Medical Research (NIGMS Human Genetic Mutant Cell Repository Camden, New Jersey).

For TDT studies, we used DNA samples from all available parent/child pedigrees with a UC offspring and a similar number of pedigrees with a CD offspring. These pedigrees were from an extended set of an IBD family collection, ascertained by the IBD Genetic Studies of Johns Hopkins University, University of Chicago and University of Pittsburgh, that has been described previously (56). Briefly, DNA was purified from blood samples obtained from North American, non-Hispanic Caucasian families with one or more cases of IBD, diagnosed as UC, CD or indeterminate colitis. The case notes of all patients were reviewed and diagnoses were confirmed by standard endoscopic, histopathological, and radiological criteria (1,2). Subjects were classified as Ashkenazi Jewish, as previously described (27). UC, CD or IBD probands from these pedigrees were also compared with controls ascertained by Johns Hopkins University and University of Chicago, as described (56).

For the case–control replication, we genotyped DNA samples from a separate set of non-Jewish, non-Hispanic, Caucasian UC patients, that were not members of families genotyped for the TDT studies and for whom DNA samples on parents were unavailable. Additional DNA samples were genotyped from non-Jewish, non-Hispanic, Caucasian UC patients recruited from the University of Toronto IBD center. The ‘set B’ non-Jewish, non-Hispanic control DNA samples genotyped were from healthy individuals, randomly ascertained from a population based cohort study, the NYCP for longitudinal follow up for future development of cancer. The NYCP has enrolled approximately 20000 normal subjects from the New York Metropolitan area between the ages of 35 and 60 since 1999. In addition to blood samples, data on ethnicity of the subject, their parents and grandparents, as well as a general medical history and a family history of cancer is obtained during a face to face interview of each subject.

Sequencing NFKB1 for polymorphisms detection

To detect NFKB1 sequence variations, we initially sequenced DNA samples from 12 subjects (all Caucasian, three Jewish) to give 95% power to detect polymorphisms with a frequency of >5% (57). Using NFKB1 specific primers (Supplementary Material Table 4), designed on the basis of the published NFKB1 genomic DNA sequences (accession AF213884, gi 7012904), we amplified by PCR overlapping fragments of the promoter and exon 1 (from position -889 5' of a NFKB1 major transcription initiation site) (31) and all 23 coding exons as well as >25 bp of each coding exon’s flanking intron sequence. PCR was performed, in a 50 μl reaction mixture containing 15 ng of genomic DNA, under the following conditions: denaturation at 95°C for 30 s, annealing at 56°C and extension at 72°C for 1 min, amplification for 35 cycles. The annealing temperature for amplifying GC rich and promoter regions was 60°C. Amplified DNA fragments were purified by spin column centrifugation through a selective adsorption silica-gel matrix (QIAquick PCR Purification Kit Qiagen Cat. No. 28104) and then sequenced on an ABI 3700 fluorescent capillary sequencer. Sequences of amplified fragments were compared with each other and with the published NFKB1 genomic DNA sequence to identify variants. Additional UC DD homozygotes were sequenced to identify more rare variants for the promoter, exon 1 and exon 2 region that may be in LD with the D allele.

Genotyping the -94delATTG promoter polymorphism

A restriction enzyme digestion assay was used to genotype the -94ins/delATTG polymorphism for Johns Hopkins, University of Chicago and New York Cancer Project samples. A 289 bp PCR fragment was amplified from genomic DNA using the ‘promoter e’ forward and ‘promoter f’ reverse primers (Supplementary Material Table 4). Products were digested by the enzyme PflM1, which cleaves the -94insATTG containing product twice and the -94delATTG containing product once (Fig. 1), and analyzed on a 2.5% agarose gel.

University of Pittsburgh samples were genotyped for the polymorphism using the same primers but with the forward primer end labeled with fluorescent dye, and the presence or absence of the 4 bp deletion was determined by the size of the labeled PCR product on an ABI 3700 sequencer.

The University of Toronto samples were genotyped using the ABI Prism SNapShot kit. A 190 bp fragment of DNA, encompassing the site of the -94delATTG polymorphism,
was first amplified by PCR. The PCR product was purified of unincorporated dNTPs as well as single stranded DNA/primers using shrimp alkaline phosphatase and exonuclease I, respectively. The purified fragment was then used as the template for the SNAPShot reaction. Primers that were complimentary to the wildtype −94 ins ATTG and deletion −94 del ATTG sequence were designed in the 5′ forward and 3′ reverse directions and differentially labeled by a fluorophore. Each primer ended at the nucleotide immediately preceding the 5′ most ‘A-nucleotide’ of the PflMI restriction enzyme cleavage site (Fig. 1). A single base pair extension revealed each allele discriminated by size and differential fluorophore emissions detected by 3100 and 3700 ABI sequencers and data was analyzed with GeneScan and/or GeneMapper software.

Twelve DNA samples, four for each of the three possible genotypes (homozygote wildtype, heterozygote and homozygote deletion) whose sequences were determined by direct sequencing, were used as blinded controls for all three genotyping methods.

Statistical analysis

Transmission disequilibrium tests. We tested for the presence of LD between the NFKB1 promoter polymorphism and UC, CD and IBD using the family-based association tests in Genehunter 2.1, FBAT (Family Based Association Test) and the PDT (Pedigree Disequilibrium Test).

The TDT analysis implemented in Genehunter 2.1 (33) performs the traditional TDT (58) using all genotyped parent–child trios in the families. In this analysis, transmissions from homozygous parents are not counted (they provide a transmitted and an untransmitted copy of the same allele) and cases where one parent is missing are used only when the genotyped parent and the proband are both distinct heterozygotes (59). The cases where both parents and the proband have the same heterozygous genotypes are counted (as a transmission and non-transmission of each allele).

To test for LD independent of linkage, we used the programs FBAT (35) and PDT (36). Both FBAT and PDT allow for inclusion of triads, discordant sibships as well as extended families and will incorporate data from multiple affected sibships in the analysis while adjusting for their non-independence. The FBAT (35) also uses data from nuclear families, sibships or a combination of the two, to test for association between traits and genotypes. If data are available on pedigrees, the program decomposes each pedigree into individual nuclear families or sibships. The program constructs, by default, a test of the null hypothesis: no linkage and no association; testing for both, linkage in the presence of LD. Using option ‘-e’, it computes the test statistic using the empirical variance, as described by Lake et al. (60). This option should be used when testing for association in an area of known linkage and data from multiple sibs in a family are used. Distortion of transmission from parents to offspring is assessed by an observed/expected chi-square test.

The PDT (36) summarizes the results in two global scores the ‘sumPDT’, summarizing the level of significance from all families, and the ‘avePDT’, weighting the contribution of larger families to ensure that their contribution to the end result does not exceed that of the smaller families.

Case–control analyses. Comparison of allele frequencies and genotypes between cases and controls was done using Fisher’s exact test of proportions (61). For individuals from multiply affected IBD pedigrees, only one individual from each pedigree, specifically, the first individual with UC, CD or IBD enrolled from their family into the study, was used for the respective analyses.

Electrophoretic mobility-shift assays (EMSAs)

Nuclear protein extracts were made from 90% confluent human tissue-culture cells grown at 37°C with 5% CO2 in DMEM supplemented with 10% fetal bovine serum (FBS) and streptomycin/ampicillin, or were extracted from colonic and ileal biopsies of normal mucosa from two individuals without IBD nor other inflammatory disorders or diarrheal diseases, that had undergone colonoscopy screening for colonic polyps. Biopsies were obtained following informed consent. Each of the nuclear protein extracts was made using the NE-PER kit from Pierce (Milwaukee, MI, Cat. #78833) as per manufacturer’s instructions. Complimentary single-stranded oligonucleotide probes were synthesized based on the NFKB1 promoter (‘W’, ‘D’ and ‘DL’; Fig. 2B) or the canonical NF-kB p50/p65 protein binding consensus sequence (62) (5′-AGTTGAGGGACTTCCAGGC-3′) was used as a control for equal protein loading. An additional 4-base overhang (gatc) was added at the 5′ ends of each oligonucleotide to optimize end-labeling with 32P. Complimentary oligomers were allowed to anneal, and then radioactively labeled with dATP [α-32P] and dCTP [α-32P] according to the method of Feinberg and Vogelstein (63). Following purification by Qiagen the labeled, double-stranded DNA oligomers were then incubated for 30 min with individual nuclear extract samples at room temperature. Electrophoretic Mobility Shift Assays (EMSAs) were performed as previously described (64).

Plasmid construction of luciferase reporter genes

The promoter-exon 1 region of the NFKB1 containing genomic sequence from nucleotides −736 to +245 (Fig. 1) was prepared by PCR amplification of either −94 ins ATTG homozygote or −94 del ATTG homozygote human genomic DNA using primers Pro c-F and Pro h-R (Supplementary Material Table 4). The PCR products were purified by agarose gel electrophoresis, extracted from gel slices (QiAprep miniprep kit; Qiagen Inc., Chatsworth, CA), and cloned into the pCR II - TOP vector (Invitrogen, San Diego, CA). After restriction digestion with KpnI and XhoI, the NF-kB promoter fragment was cloned directionally into the pGL3-Basic firefly luciferase expression vector (Promega, Madison, WI) between unique KpnI and XhoI sites. Restriction analysis and complete DNA sequencing confirmed the orientation and integrity of each construct’s inserts.

Transient transfection/reporter assay

HeLa human cervical adenocarcinoma and HT-29 human epithelial colon cancer cell lines were obtained from American Type Culture Collection (Rockville, MD). Cells were grown in Dulbecco’s modified Eagle’s (DME)/high glucose medium.
supplemented with 10% fetal bovine serum, 1 mM sodium pyruvate, 100 U/ml of penicillin G and 100 μg/ml streptomycin (Life Technologies, Inc.) at 37°C in 5% CO₂. Subconfluent cells cultured in 24-well dishes were transiently co-transfected with 0.4 μg of either pGL3-W or pGL3-D reporter vector (Fig. 3) and 5 ng of the thymidine kinase promoter–Renilla luciferase control vector (phRL-TK, Promega) using 0.6 μl of FuGENE6 as per manufacturer’s specifications (Roche Molecular Biochemicals, Indianapolis, IN). The phRL-TK vector contains the herpes simplex virus thymidine kinase promoter and was co-transfected as an internal control for transfection efficiency (65). The concentrations of each PGL3-W and PGL3-D vectors were determined, following Qiagen purification procedure in parallel, by an average of 10 spectrophotometric readings. Transfections using pGL3-Basic vector without an insert were used as a negative control. Twenty-four hours after transfection, the cells were cultured in 10% serum medium or with exposure to 1 μg/ml of E. coli derived lipopolysaccharides (LPS: Serotype O55:5B) for 6–24 h. Cells were then lysed, and firefly and Renilla luciferase activities were measured simultaneously in each sample using the Dual-Luciferase Reporter Assay System according to the manufacturer’s instructions (Promega). Firefly luciferase activities were normalized to Renilla luciferase activity as ‘relative luciferase activity’. The data presented are means of six independent experiments. The results are expressed as the mean plus standard error of the mean. Statistical analyses were performed using Stat View software for Macintosh version 5.0 (SAS Institute Inc.). Unpaired Student’s t-tests were used for comparisons. A P-value of 0.05 was considered to be statistically significant.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG Online.

**ACKNOWLEDGEMENTS**

We are grateful to the IBD patients, their families and healthy control individuals for their participation in this study. We thank Dr Ethylin Wang Jabs for her critical review of the manuscript, and Drs Mary Harris, Carmelo Cufarri and Michael F. Picco for assistance in patient recruitment. Typing assistance was provided by Marcia Tepper, and Tara Hardy assisted in subject recruitment and data entry. This study was funded in part by National Institutes of Health Grants R01DK58189 and RR00052 (S.R.B.), and RR00055 (J.H.C.); Crohn’s and Colitis Foundation of America (S.R.B., J.H.C., R.H.D.); The Harvey M. and Lyn P. Meyerhoff IBD Center (A.S.K., T.O., T.D., T.M.B., S.R.B.); The Israeli Society of Gastroenterology (A.S.K.); ToyoBio Biotechnology Foundation (T.O.); Johns Hopkins University Minority Student Internship Summer Program (T.G.); Reva and David Logan Foundation (J.H.C.); The Seafar Family Foundation (R.H.D.); Crohn’s and Colitis Foundation of Canada, Canadian Association of Gastroenterology, and the Canadian Institutes of Health Research (M.S.S.). New York Control Samples were funded by Academic Medicine Development Corporation (P.K.G.).

**REFERENCES**


