

## ARTICLE

# Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease

Mark S Silverberg<sup>1,11</sup>, Richard H Duerr<sup>2,3,11</sup>, Steven R Brant<sup>4,5</sup>, Gillian Bromfield<sup>6</sup>, Lisa W Datta<sup>4,5</sup>, Niraj Jani<sup>2</sup>, Sunanda V Kane<sup>6</sup>, Jerome I Rotter<sup>7</sup>, L Philip Schumm<sup>8</sup>, A Hillary Steinhart<sup>1</sup>, Kent D Taylor<sup>7</sup>, Huiying Yang<sup>7</sup>, Judy H Cho<sup>6,11</sup>, John D Rioux<sup>9,11</sup> and Mark J Daly<sup>\*,10,11</sup>, of the NIDDK IBD Genetics Consortium

<sup>1</sup>Departments of Medicine and Surgery, Mount Sinai Hospital IBD Centre, University of Toronto, Toronto, Ontario, Canada; <sup>2</sup>Department of Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; <sup>3</sup>Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; <sup>4</sup>Department of Medicine, Meyerhoff Inflammatory Bowel Disease Center, Johns Hopkins University, Johns Hopkins University School of Medicine, Baltimore, MD, USA; <sup>5</sup>Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; <sup>6</sup>University of Chicago, Chicago, IL, USA; <sup>7</sup>Medical Genetics Institute, Cedars-Sinai Medical Center, UCLA School of Medicine, Los Angeles, CA, USA; <sup>8</sup>Department of Health Studies, University of Chicago, Chicago, IL, USA; <sup>9</sup>Department of Medicine, Université de Montréal, Montréal Heart Institute, Montréal, Québec, Canada; <sup>10</sup>Centre for Human Genetic Research, Massachusetts General Hospital, Simches Research Center, Boston, MA, USA

Although the general association of the inflammatory bowel disease (IBD) 5 region on chromosome 5q31 to Crohn's disease (CD) has been replicated repeatedly, the identity of the precise causal variant within the region remains unknown. A recent report proposed polymorphisms in solute carrier family 22, member 4 (*SLC22A4*) organic cation transporter 1 (*OCTN1*) and solute carrier family 22, member 5 (*SLC22A5*) (*OCTN2*) as responsible for the IBD5 association, but definitive, large-sample comparison of those polymorphisms with others known to be in strong linkage disequilibrium was not performed. We evaluated 1879 affected offspring and parents ascertained by a North American IBD Genetics Consortium for six IBD5 tag single nucleotide polymorphisms (SNPs) to evaluate association localization and ethnic and subphenotypic specificity. We confirm association to the IBD5 region (best SNP IGR2096a\_1/rs12521868,  $P < 0.0005$ ) and show this association to be exclusive to the non-Jewish (NJ) population ( $P = 0.00005$ ) (risk allele undertransmitted in Ashkenazi Jews). Using Phase II HapMap data, we demonstrate that there are a set of polymorphisms, spanning genes from prolyl 4-hydroxylase (*P4HA2*) through interferon regulatory factor 1 (*IRF1*) with equivalent statistical evidence of association to the reported *SLC22A4* variant and that each, by itself, could entirely explain the IBD5 association to CD. Additionally, the previously reported *SLC22A5* SNP is rejected as the potential causal variant. No specificity of association was seen with respect to disease type and location, and a modest association to ulcerative colitis is also observed. We confirm the importance of IBD5 to CD susceptibility, demonstrate that the locus may play a role in NJ individuals only,

\*Correspondence: Dr MJ Daly, Massachusetts General Hospital, Simches Research Center, CPZN-6818, 185 Cambridge St, Boston, MA 02114, USA.  
Tel: +1 617 643 3290; Fax: +1 617 643 3293;  
E-mail: mjdaly@chgr.mgh.harvard.edu

<sup>11</sup>These authors contributed equally to this work.

Received 16 May 2006; revised 7 September 2006; accepted 8 September 2006; published online 10 January 2007

and establish that *IRF1*, *PDLIM*, and *P4HA2* may be equally as likely to contain the IBD5 causal variant as the *OCTN* genes.

*European Journal of Human Genetics* (2007) 15, 328–335. doi:10.1038/sj.ejhg.5201756; published online 10 January 2007

**Keywords:** Crohn's disease; association; IBD5; ethnicity

## Introduction

The inflammatory bowel diseases (IBD), Crohn's disease (CD, MIM# 266600) and ulcerative colitis (UC, MIM# 191390), are chronic inflammatory diseases of the gastrointestinal tract with a peak age of onset in the second and third decades of life and result in symptoms of diarrhea, bleeding, and abdominal pain.<sup>1</sup> CD is a transmural inflammatory process that can occur in any part of the gastrointestinal tract with areas of normal mucosa interspersed among areas of affected bowel. UC manifests as a superficial mucosal inflammation, which begins in the anus and rectum and progresses in a contiguous fashion to affect the entire colon in more than one-third of affected individuals. As CD and UC have many overlapping clinical features and there are no specific diagnostic tests, it can occasionally be difficult to distinguish the two, and in those instances a diagnosis of indeterminate colitis (IC) is applied.<sup>2</sup>

The pathogenesis of IBD is believed to involve exposure to an as yet unidentified environmental trigger in a genetically susceptible individual.<sup>3</sup> The importance of genetic factors in the pathophysiology of IBD has been repeatedly established by the substantially increased relative risk to siblings of affected individuals and higher monozygotic *versus* dizygotic twin concordance rates.<sup>4–7</sup> Although the heritability of CD seems higher than that for UC, the available epidemiological and genetic data suggest that there are common susceptibility factors for both diseases and it is likely that gene–gene and gene–environmental interactions determine the specific phenotype.<sup>8</sup> Although significant progress in identifying genetic factors underlying IBD through replicated linkage has been made (IBD1–IBD6),<sup>9–14</sup> only the susceptibility gene within the IBD1 locus has been conclusively identified. Three variants within the caspase recruitment domain-containing protein 15 (*CARD15*, MIM# 605956) gene<sup>15–17</sup> have been consistently associated with the development of small bowel CD.<sup>17</sup> There is an important ethnic and geographic heterogeneity for *CARD15* (*NOD2*) with slightly different allele frequencies and odds ratios (OR) for the major variants among AJ compared to non-Jewish (NJ) populations and in Asian individuals affected by CD, *CARD15* (*NOD2*) variants are extremely rare and not associated with CD.<sup>17–20</sup>

The IBD5 locus (MIM# 606348) on chromosome 5q31 is the second locus clearly demonstrated to confer increased risk for CD through replicated association. Following an

initial finding of significant linkage in a Canadian population,<sup>14</sup> subsequent studies identified a 250 kb risk haplotype within the IBD5 locus that is significantly associated with CD.<sup>21</sup> As part of this effort, the region was sequenced in eight individuals and was intensively evaluated with dense genotyping in more than 250 trios, which led to the discovery of long segments of complete linkage disequilibrium (LD), described as 'haplotype blocks', which were separated by apparent 'hotspots' of recombination.<sup>22</sup> This original association study described 11 haplotype tagging SNPs in this region which were equally associated with CD. Important candidate genes located within and nearby to this region include interferon regulatory factor 1 (*IRF1*), solute carrier family 22, member 4 (*SLC22A4*) also known as organic cation transporter 1 (*OCTN1*); solute carrier family 22, member 5 (*SLC22A5*), also known as organic cation transporter 2 (*OCTN2*); and prolyl 4-hydroxylase (*P4HA2*) among others. Since the initial identification of this risk haplotype, numerous studies have replicated the association with CD and a few have demonstrated an association between IBD5 and UC as well,<sup>23–28</sup> establishing this as one of a still limited number of *bona fide* associations in human complex disease genetics and in IBD specifically.

In 2004 Peltekova *et al*<sup>29</sup> postulated that functional polymorphisms, L503F (rs1050152) and G-207C (rs2631367) in the *SLC22A4* (*OCTN1*) and *SLC22A5* (*OCTN2*) genes, respectively, comprised a two-locus risk haplotype that accounted for the association findings at IBD5. That functional polymorphisms in significant LD with each other in adjacent, highly homologous genes may act as a functional cassette to increase disease risk would represent a novel, important paradigm in complex disorders. In particular, *SLC22A5* (*OCTN2*) transports carnitine, which plays a key role in mitochondrial transport of long chain fatty acids; furthermore, *SLC22A5* (*OCTN2*)-deficient mice develop intestinal lymphocytic infiltration.

Statistical support for a relatively unique contribution of the two-locus *OCTN* haplotype was initially reported in a German cohort.<sup>28</sup> Subsequent to this study, a number of independent populations have been evaluated with somewhat variable results.<sup>30–32</sup> In particular, a study from Edinburgh demonstrated that although the *OCTN1* and *OCTN2* variants are associated with CD, this association is not independent of the background risk haplotype in the region.<sup>30</sup> This study assessed not only the *OCTN* variants but also a number of surrounding SNPs from the list of

SNPs comprising the IBD5 risk haplotype (IGR2096, IGR2198 and IGR2230). Utilizing these SNPs, they could not demonstrate independent association of the OCTN variants and suggest that the IGR2078 SNP used by <sup>24,25,31–33</sup> is one haplotype block removed from the OCTN variants and therefore was not truly indicative of an independent association (Figure 1). There is also significant inconsistency in the results of phenotypic analyses of the IBD5 locus with associations described for ileal and perianal disease locations as well as for UC.<sup>24,25,31–33</sup> To better evaluate localization of the IBD5 association, ethnic specificity of association, and specific association to subphenotypic groups, we performed genotyping of haplotype-tagging SNPs in the IBD5 region in a much larger North American cohort comprised of 1879 IBD-affected offspring and their parents.

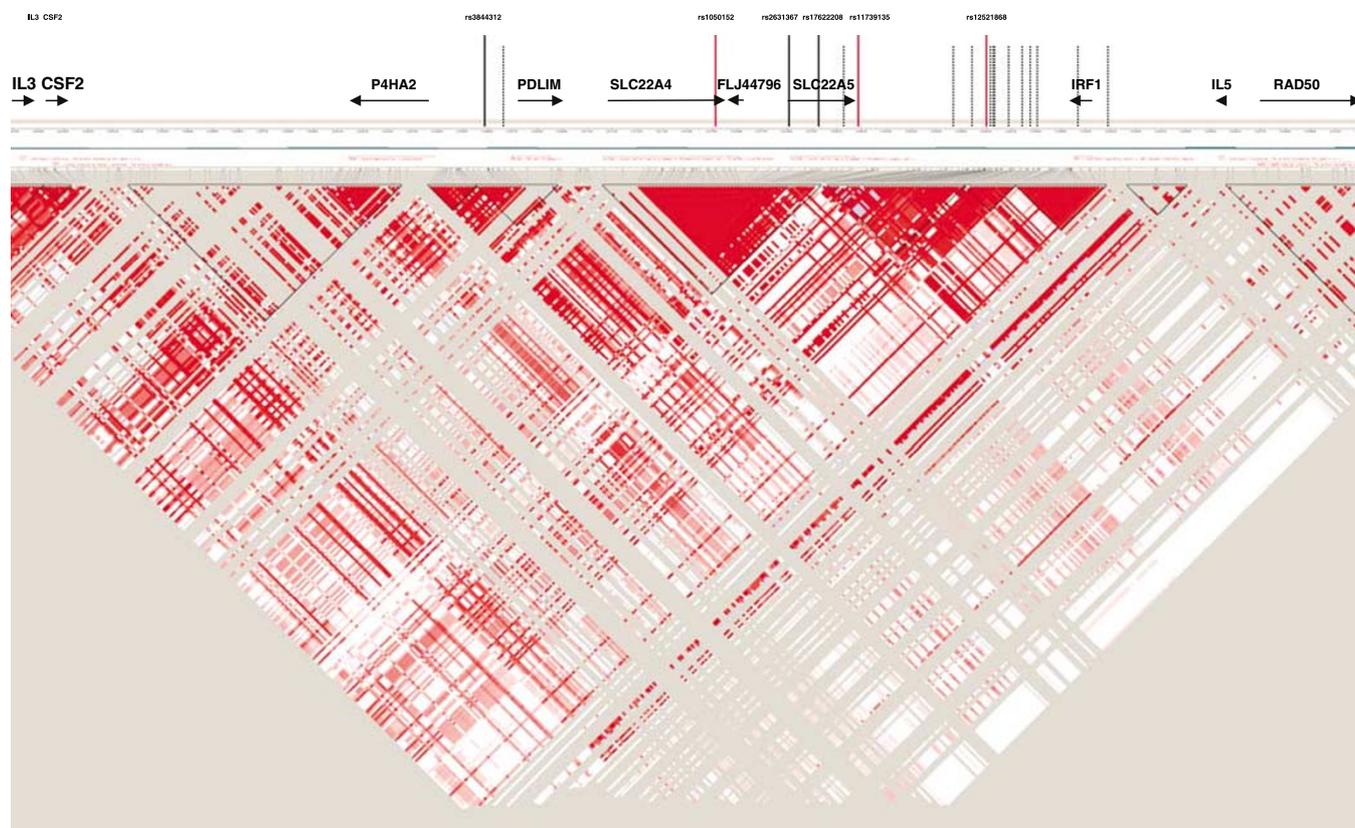
## Materials and methods

### Population

This study was conducted by the NIDDK Inflammatory Bowel Disease Genetics Consortium (NIDDK IBGCG), which is comprised of six Genetic Research Centers (GRCs) located at the Broad Institute/Massachusetts Institute of

Technology (J Rioux), Cedars Sinai Medical Center in Los Angeles (H Yang), University of Chicago (J Cho), Johns Hopkins University (S Brant), University of Pittsburgh (R Duerr) and the University of Toronto (M Silverberg), and a Data Coordinating Center (DCC) located at the University of Chicago. DNA samples and clinical data from patients with inflammatory bowel disease recruited for prior studies in IBD genetics by each GRC were utilized. The investigators listed for each site were responsible for ensuring documentation of the diagnosis of IBD as well as the additional clinical information that was collected. Institutional review board approvals were obtained from each of the sites involved in this study.

Patients considered eligible for this study were those with available DNA and clinical data from complete trio (a proband with two parents available for genotyping) and tetrad (affected siblings with two parents available for genotyping) families. A description of the study sample is shown in Table 1. The following clinical parameters were ascertained: diagnosis, age of diagnosis, gender, race, Hispanic origin, Jewish ethnicity, disease location. The definitions of these categories was deemed to be similar among all sites and therefore included in the analysis. Other disease parameters such as disease behavior, surgical



**Figure 1** IBD5 map demonstrating the location of the SNPs genotyped and regional candidate genes.

**Table 1** (a) Population description<sup>a</sup> and (b) clinical characteristics

	Number of affected offspring	Number of nuclear families
<b>(a)</b>		
CD	1283	1012
UC	565	487
IBD	1879	1442
<b>(b)</b>		
<i>All IBD affected</i>		
Mean age of Dx (years)	19.8 <sup>b</sup>	
Jewish ethnicity (%)	26.4	
Caucasian (%)	98.7	
Gender (% female)	53	
<i>CD disease location (%)</i>		
Ileal only	38.6	
Ileocolonic	45.8	
Colonic only	15.6	
Perianal	28.7	
<i>UC disease location (%)</i>		
Extensive	64.9	
Left sided	25.7	
Proctitis	9.4	

<sup>a</sup>From 1426 independent families.

<sup>b</sup>44% of affected offspring diagnosis age <18.

history, and smoking were deemed to be insufficiently documented with consistency at all centers and were therefore not included for analysis. Definitions of race and Hispanic or Jewish origin were self-reported. There were not a substantial number of Hispanic or non-Caucasian samples and thus the evaluation of these subgroups individually are not addressed in this study. All clinical variables were documented by review of medical records at each GRC.

### Genotyping

DNA from all individuals was genotyped at the Broad Institute of Harvard and MIT using the Sequenom MassArray 7K system and MALDI-TOF MS analysis. For this report, the following seven IBD5 SNPs were tested: IGR2096a\_1 (rs12521868); IGR2198a\_1 (rs11739135); IGR2230a\_1 (rs17622208); SLC22A5/OCTN2 G-207C (rs2631367); SLC22A4/OCTN1 L503F (rs1050152); IGR3081a\_1; and IGR3178a\_1 (rs3844312). The genotypes for IGR3081a\_1 deviated substantially from Hardy-Weinberg equilibrium, so IGR3081a\_1 was dropped from subsequent analysis. Although potentially a sign of non-log-additive association, IGR3081a\_1 was less associated than four other markers and no previous analysis of this association had suggested substantially non-log-additive association so it appeared this was simply an unreliable

assay. The selected IBD5 markers were chosen as tags for the risk haplotype in the consecutive blocks in the maximally associated region as previously described<sup>21,22</sup> with the specific addition of the two OCTN variants<sup>29</sup> for comparison. The Sequenom platform used at the Broad Institute had been used to run more than 50 000 previous assays and was established to be more than 99.5% accurate in HapMap QC exercises. Genotyping was performed blinded to affection and family information.

### Statistical analysis

Analysis was performed using the standard transmission disequilibrium test (TDT) (as implemented in Haploview v3.3). Results are reported with multiple affected siblings counted independently as (a) the present study was not conditioned on nor pursuing linkage in these samples; (b) the modest genotype relative risk of the IBD5 risk haplotype estimated at 1.3<sup>34</sup> suggest a minimal effect on allele sharing owing to this locus ( $\lambda_s < 1.02$ ; expected mean allele sharing <50.5%); and (c) given previous replications we fully expected to observe the effect and the current study is motivated by evaluating subgroups and resolving location rather than establishing whether the factor is truly associated or not. All results were confirmed by running FBAT with the conservative empirical variance adjustment for the presence of affected siblings and no meaningful differences were seen beyond a slight reduction in statistical significance owing to the conservative nature of the correction. The incorporation of a discordant allele test for affected-unaffected parent pairs described by Purcell *et al*<sup>35</sup> was performed using Haploview v3.3. Conditional logistic regression was performed using WHAP (<http://pngu.mgh.harvard.edu/~purcell/whap>), which enables flexible stepwise testing with nested models compared using a likelihood-ratio test. HapMap public release 19 (Phase II data with more than 3.9 million total SNPs passing QC in the CEU sample) was used to identify untested SNPs highly correlated to the most associated SNPs in this study. Testing of subgroup specificity (eg, Jewish *versus* NJ) was performed by permuting subgroup labels at random 10 000 times – the proportion of random subgroups that showed a difference as great or greater than the observed groups is reported as the significance of the observed subgroup difference.

### Results

#### Association of IBD5 with CD

A description of the study sample is shown in Table 1. Utilizing the TDT, significant association between the six IBD5 markers tested and CD was found (Table 2). Importantly, whereas significant results were found for SLC22A4/OCTN1 L503F (rs1050152) and SLC22A5/OCTN2 G-207C (rs2631367), slightly stronger association was seen for marker IGR2096a\_1 (rs12521868), which is located

30 kb telomeric from the end of *IRF1* and 60 kb centromeric of *SLC22A5* (*OCTN2*) (Table 2 and Figure 1). As a considerable number of affected parents were present in this data set, we incorporated a discordant allele test among the affected–unaffected parent pairs, which increased the significance of the replication to a maximum *P*-value of 0.00006 at IGR2096a\_1 (Table 2).

A small number of Toronto samples ( $n=60$ ) were included in the original sample that identified the IBD5 association<sup>21</sup> – removal of these to define formal replication has very little impact (elevating the 0.00006–0.000075). For the remaining subgroup analyses these samples are included as the analyses are evaluating previously unexplored hypotheses and thus inclusion of these samples is biased neither for nor against any specific outcome. Furthermore, a Breslow-Day test of heterogeneity showed no significant site-specific heterogeneity of association across the six sites. The estimated effect size observed in this data set is OR = 1.21 [(1.09, 1.36)] (IGR2096a\_1).

Evaluation of the association results stratified by Jewish and NJ ethnicity demonstrated that the IBD5 association with CD comes exclusively from the NJ CD population

(Table 3). IGR2096a\_1 (rs12521868) and *SLC22A4/OCTN1* L503F (rs1050152) are uniformly the most associated variants in this sample. In fact, within the Jewish CD population, there was a small under-transmission of IBD5 risk alleles. This observed difference in transmission distortion is statistically significant ( $P=0.01$ ) and not simply a factor of the smaller sample of Jewish families. (ie, when ethnicity was permuted, only 1% of subgroups of equivalent size to the Jewish sample here would show the observed lack of transmission.) As a result of this finding, subsequent subgroup and resolution analyzes described here were performed exclusively in the NJ sample. The effect size among NJ only is estimated to be OR = 1.30 (1.14, 1.49).

When evaluated by disease location, we did not observe significant differences between the level of association to IBD5 in patients grouped by disease site (Table 3). When evaluating CD patients by age of diagnosis, both early onset and late onset cases showed significant association.

#### Localization of IBD5 association

We utilized conditional logistic regression approaches to evaluate which, if any, of the significantly associated IBD5

**Table 2** TDT results for CD

Marker	Original name	Transmitted	Untransmitted	$\chi^2$	<i>P</i> -value	$\chi^2$ <sup>a</sup>	<i>P</i> -value <sup>a</sup>
rs3844312	IGR3178a_1	665	573	6.8	0.0045	9.8	0.00088
<b>rs1050152</b>	<b>OCTN1/L503F</b>	695	575	11.3	0.00038	12.7	0.00018
rs2631367	OCTN2/G-207C	646	557	6.6	0.0051	8.4	0.0019
rs17622208	IGR2230a_1	697	591	8.7	0.0016	12.1	0.00026
<b>rs11739135</b>	<b>IGR2198a_1</b>	670	561	9.7	0.00095	10.7	0.00054
<b>rs12521868</b>	<b>IGR2096a_1</b>	673	554	11.5	0.00034	14.8	0.000061

The three SNPs in bold are determined (by conditional logistic regression) to be adequate to explain the entire association signal at IBD5.

<sup>a</sup>With discordant allele test (DAT) with affected parents included.

**Table 3** TDT results stratified by ethnicity, disease location and age of diagnosis

	<i>OCTN1 (SLC22A4)/L503F (rs1050152)</i>				<i>IGR2096a_1 (rs12521868)</i>			
	T	U	$\chi^2$	<i>P</i>	T	U	$\chi^2$	<i>P</i>
All CD	695	575	11.3	0.00038	673	554	11.5	0.00034
All UC	289	249	3.0	0.042	282	242	3.1	0.04
<i>Ethnicity</i>								
CD-NJ	520	402	15.1	0.000051	502	386	15.2	0.00005
CD-J	159	165	—	—	155	161	—	—
<i>Disease location (CD-NJ only)</i>								
Ileocolonic	203	148	8.6	0.0017	197	137	10.8	0.00051
Ileal	148	105	7.3	0.0034	144	100	7.9	0.0024
Colonic	56	38	3.4	0.032	48	39	0.9	NS
<i>Age of diagnosis (CD-NJ only)</i>								
AOD ≤ 18	243	193	5.7	0.0083	236	188	5.4	0.0098
AOD > 18	277	209	9.5	0.001	266	198	10.0	0.0008

AOD, age of diagnosis; AJ, Ashkenazi Jewish; NJ, non-Jewish; NS, nonsignificant; T, transmitted; U, untransmitted.

SNPs could fully explain the association to CD and therefore be potentially implicated as the causal polymorphism(s). We began by adopting the most associated SNP, IGR2096a\_1 (rs12521868) as causal, and asked whether significant association was seen with any other SNP when controlling for the effect of this genetic variant. No other SNPs in this region were associated when IGR2096a\_1 (rs12521868) was considered causal, indicating that this SNP by itself could explain all reports of association to the region and suggesting that the IBD5 association is due to a single factor (rather than a more complex combination of alleles or haplotypes). Similarly, *SLC22A4/OCTN1* L503F (rs1050152) and IGR2198a\_1 (rs11739135) were also found to adequately explain the association to all SNPs (Table 1, Figure 1).

The other three SNPs tested, including *SLC22A5/OCTN2* G-207C (rs2631367) reported to influence, or to be in LD with an SNP that influences *SLC22A5* (*OCTN2*) expression, are rejected as causal under this approach. For example, if one conditions on the association to *SLC22A5/OCTN2* G-207C (rs2631367), very significant association is still seen at IGR2096a\_1 (rs12521868) ( $P < 0.0002$ ). We note that this does not at all compromise the observation that this SNP may alter the expression of *SLC22A5* (*OCTN2*); the data here simply assert that this effect is not the causal factor involved in CD susceptibility. Peltekova *et al*<sup>29</sup> raise the possibility that it is the combination of the *SLC22A4/OCTN1* and *SLC22A5/OCTN2* alleles that is the causal factor. However, these two SNPs are in very strong LD and although there are a substantial number of chromosomes (6.5% in our study) that carry the reported *SLC22A5/OCTN2* risk allele but not *SLC22A4/OCTN1* 503F, fewer than 1% of chromosomes carry 503F without the *SLC22A5/OCTN2* risk allele. Thus, the set of chromosomes that carry the *SLC22A4/OCTN1* 503F risk allele are essentially equal to the set carrying both putative risk alleles and therefore any postulated effects of the 503F allele or the combination of that allele with the *SLC22A5/OCTN2* risk allele cannot be distinguished.

Examining LD among the three potentially causal SNPs and the three rejected SNPs shows, not surprisingly, that the three potentially causal variants are highly correlated with each other (all three pairwise  $r^2$  values  $> 0.8$ ) but less so with the SNPs excluded as being causal (all nine pairwise  $r^2$  values below 0.8). The recently released Phase II HapMap data provides an average of more than one SNP per kb throughout the IBD5 region and using this data, we identify 13 additional SNPs, spanning 243 kb that are highly correlated ( $r^2 > 0.8$ ) with the maximally associated SNP IGR2096a\_1 (rs12521868). This list is far from comprehensive, as the Phase II HapMap has directly examined fewer than half of the common variants in the genome, but suggests that variants near *P4HA2* and *IRF1*, in addition to the two *OCTN* genes are, from a statistical standpoint, equally likely to be the causal factor.

## IBD5 and UC

A modest but significant association is seen between UC and IBD5 variation, supporting the possible role of the IBD5 locus in UC (Table 3). Although subset analysis identified apparently stronger association between IBD5 and extensive UC (as compared to more distal disease) and early onset UC (data not shown), the modest relative risk of this association and the limited size of the UC sample preclude any statistically meaningful subgroup differences to be evaluated until substantially larger study samples are examined.

## Discussion

Although common alleles of modest effect play a very important role in the architecture of complex diseases and can be readily detected in some cases with modest sample sizes, it is now widely recognized that extremely large sample sizes are required in order to conclusively confirm these factors, resolve their location through fine-mapping and identify specific subphenotypes that are either preferentially associated or not-associated with disease. With the large sample size aggregated here, the current study has the opportunity to explore age at diagnosis, disease location and ethnicity in more detail than the previously published replication efforts and to evaluate the positional resolution of the IBD5 association.

In complex disorders, the IBD5 association in CD represents one of the best replicated genomic regions, and therefore warrants comprehensive studies in large, well-phenotyped cohorts to define those population subsets and functional polymorphisms that most likely contribute to disease risk.<sup>23,24,26,28</sup> The report by Peltekova *et al*<sup>29</sup> suggested that an independently associated two-allele risk haplotype of the *SLC22A4* (*OCTN1*) and *SLC22A5* (*OCTN2*) genes were responsible for the association findings of the IBD5 risk haplotype. The data evaluated in this large sample do not support the conclusion that these two SNPs can be assigned responsibility for the genetic susceptibility at IBD5.

In our study, the association of IBD5 to CD was unequivocally confirmed, and as the AJ population has a substantially higher incidence of CD and has been reported to have a potentially different allelic architecture at the *CARD15* (*NOD2*) risk locus, we sought to examine whether the evidence for association was similar in Jews and non-Jews. Strikingly, the 340 Jewish trios show no evidence for association to IBD5, and in fact, demonstrated a slight undertransmission of the risk alleles whereas the 986 NJ trios show strong association (OR  $> 1.3$ ), consistent with a recent meta-analysis of this risk haplotype.<sup>30</sup> We have therefore found statistically meaningful (after permutation testing) evidence that the IBD5 risk factor, whereas present at comparable frequencies in AJ and non-Jews alike, may

constitute a CD risk factor in the NJ population only. The AJ population likely reflects a unique risk group and whereas overall genetic similarity would suggest most risk factors should be shared with NJ individuals, it is conceivable that there will also be risk alleles specific to each. Although variable LD with a true causal factor can lead to apparent differences in association, this seems unlikely given the general similarity in LD patterns among common variants between AJ and NJ populations.<sup>19</sup> Alternately, an interaction which requires the presence of a remote allele or mutation which itself is at different frequencies among populations can give rise to differences in association even when the causal variant is typed and equifrequent among the populations. Studies in other population groups having distinct patterns of LD, such as in African cohorts, may provide additional insight and aid in further resolution of the effect; however, given the relatively modest significance of the observation and the overall genetic similarity between AJ and NJ Caucasians, the effect observed here should be replicated first.

In the context of this study, we also evaluated the relationship of the IBD5 locus to disease location. There has been considerable discrepancy in correlation between IBD5 and phenotype with ileal, ileocolonic, and perianal disease sites all described.<sup>23,24,26,28,31,33</sup> Association to IBD5 is quite strong in ileal disease with and without colonic involvement and is generally comparable in all phenotypic categories in this data set.

Having explored ethnic and phenotypic specificity of IBD5, we then sought to utilize the large sample in conjunction with the identification of this higher penetrance subgroup, to evaluate whether the reported *OCTN1* or *OCTN2* variants were actually the most significantly associated variants in this region and more generally, whether resolution in this extended region of LD was possible. Using stepwise regression, we identified three SNPs (IGR2096a\_1, IGR2198a\_1, and SLC22A4) that were each adequate individually to explain the association to CD in full; that is, if any of these are considered the only causal SNP, no residual association is observed at any other SNP. Therefore, these three SNPs, and by extension SNPs very highly correlated to these SNPs observed on the HapMap, can be considered statistically equivalent in terms of their association to CD. As the list of maximally associated variants does include the *OCTN1* L503F variant, functional evaluation of this variant is certainly warranted. However, the extensive LD in this region suggests that caution must be exerted in asserting which of the SNPs in this region is potentially causal. Several conclusions can be drawn from these data. Given the extensive LD, it is likely that many other SNPs in this region might also demonstrate similar results and therefore be considered potentially causal. Interestingly, whereas the *SLC22A4* variant (but not the *SLC22A5* promoter variant) are included in this group, there is no genetic evidence that *SLC22A4* is

any more associated than the other IBD5 tagging SNPs and in fact the distal IGR2096a\_1 SNP tested is the most associated SNP and appears uniformly on all overtransmitted haplotypes. These data confirm and extend upon, in a much larger sample size, prior findings described in the IBD population from Edinburgh.<sup>30</sup>

Utilizing this large data set, we also examined the reproducibility of previously reported findings regarding UC and IBD5.<sup>24,25,33,36</sup> There was modestly significant association between IBD5 and UC in this population. Despite prior findings that the *CARD15* (*NOD2*) and IBD5 loci interact to enhance this association,<sup>25</sup> we found no evidence to support an interaction between the two (data not shown). Although these results require further investigation in larger sample sizes, it is not unexpected that there may be some effect of IBD5 on the development of UC. There is an increased risk of UC developing in the relatives of those with CD and the reverse is also true suggesting there are genes which confer overlapping susceptibility.<sup>8</sup> In addition, the clinical heterogeneity among the two disorders could result in misspecification of disease type in those with colonic involvement. Moreover, it has been previously noted that there are subsets of CD which have a more 'UC-like' appearance raising further suspicion for overlapping genetic susceptibility factors.<sup>37</sup> One might also speculate on the possibility of a unique 'IBD5 positive', genetically and clinically unique subset of IBD patients, that would require further testing to be better characterized.

To further advance the understanding of the IBD5 locus and its role in IBD susceptibility will require even larger population studies that will have adequate power to dissect the high degree of LD found in this region in conjunction with functional studies that will demonstrate mechanisms. These studies should be focused on the NJ populations given our findings described here. In addition, the findings in UC illustrate the importance of better powered UC genetics studies to confirm and extend the results reported here. Owing to the modest but important genetic relative risk of this locus, large consortia with prospectively collected, well-phenotyped cohorts will be required in order to refine the information regarding this region.

#### Acknowledgements

The NIDDK IBD Genetics Consortium is funded by the following Grants: DK62431 (Brant), DK62422 (Cho), DK62420 (Duerr), DK62432 (Rioux), DK62423 (Silverberg) and DK62413 (Yang). The Data Coordinating Centre funding is from NIDDK Grant DK62429.

#### References

- 1 Podolsky DK: Inflammatory bowel disease. *N Engl J Med* 2002; **347**: 417–429.
- 2 Guindi M, Riddell RH: Indeterminate colitis. *J Clin Pathol* 2004; **57**: 1233–1244.

- 3 Rogler G: Update in inflammatory bowel disease pathogenesis. *Curr Opin Gastroenterol* 2004; **20**: 311–317.
- 4 Tysk C, Lindberg E, Jarnerot G, Floderus-Myrhed B: Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins: a study of heritability and the influence of smoking. *Gut* 1988; **29**: 990–996.
- 5 Thompson NP, Driscoll R, Pounder RE, Wakefield AJ: Genetics versus environment in inflammatory bowel disease: results of a British twin study. *BMJ* 1996; **312**: 95–96.
- 6 Subhani JMS, Ounder RE, Wakefield AJ: Concordance rates of twins and siblings in inflammatory bowel disease. *Gut* 1998; **42**: A40.
- 7 Halfvarson J, Bodin L, Tysk C, Lindberg E, Jarnerot G: Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterology* 2003; **124**: 1767–1773.
- 8 Orholm M, Munkholm P, Langholz E, Nielsen OH, Sorensen IA, Binder V: Familial occurrence of inflammatory bowel disease. *N Engl J Med* 1991; **324**: 84–88.
- 9 Hugot JP, Laurent-Puig P, Gower-Rousseau C *et al*: Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996; **379**: 821–823.
- 10 Satsangi J, Parkes M, Louis E *et al*: Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* 1996; **14**: 199–202.
- 11 Hampe J, Schreiber S, Shaw SH *et al*: A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am J Hum Genet* 1999; **64**: 808–816.
- 12 Ma Y, Ohmen JD, Li Z *et al*: A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm Bowel Dis* 1999; **5**: 271–278.
- 13 Duerr RH, Barmada MM, Zhang L, Pfutzer R, Weeks DE: High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11–12. *Am J Hum Genet* 2000; **66**: 1857–1862.
- 14 Rioux JD, Silverberg MS, Daly MJ *et al*: Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* 2000; **66**: 1863–1870.
- 15 Hugot JP, Chamaillard M, Zouali H *et al*: Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411**: 599–603.
- 16 Ogura Y, Bonen DK, Inohara N *et al*: A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001; **411**: 603–606.
- 17 Economou M, Trikalinos TA, Loizou KT, Tsianos EV, Ioannidis JP: Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a meta-analysis. *Am J Gastroenterol* 2004; **99**: 2393–2404.
- 18 Inoue N, Tamura K, Kinouchi Y *et al*: Lack of common NOD2 variants in Japanese patients with Crohn's disease. *Gastroenterology* 2002; **123**: 86–91.
- 19 Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A: Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet* 2003; **12**: 771–776.
- 20 Sugimura K, Taylor KD, Lin YC *et al*: A novel NOD2/CARD15 haplotype conferring risk for Crohn disease in Ashkenazi Jews. *Am J Hum Genet* 2003; **72**: 509–518.
- 21 Rioux JD, Daly MJ, Silverberg MS *et al*: Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn's disease. *Nat Genet* 2001; **29**: 223–228.
- 22 Daly MJ, Rioux JD, Schaffner SE, Hudson TJ, Lander ES: High-resolution haplotype structure in the human genome. *Nat Genet* 2001; **29**: 229–232.
- 23 Armuzzi A, Ahmad T, Ling KL *et al*: Genotype–phenotype analysis of the Crohn's disease susceptibility haplotype on chromosome 5q31. *Gut* 2003; **52**: 1133–1139.
- 24 Giallourakis C, Stoll M, Miller K *et al*: IBD5 is a general risk factor for inflammatory bowel disease: replication of association with Crohn disease and identification of a novel association with ulcerative colitis. *Am J Hum Genet* 2003; **73**: 205–211.
- 25 McGovern DPB, van Heel DA, Negoro K, Ahmad T, Jewell DP: Further evidence of IBD5/CARD15 (NOD2) epistasis in the susceptibility to ulcerative colitis. *Am J Hum Genet* 2003; **73**: 1465–1466.
- 26 Mirza MM, Fisher SA, King K *et al*: Genetic evidence for interaction of the 5q31 cytokine locus and the CARD15 gene in Crohn disease. *Am J Hum Genet* 2003; **72**: 1018–1022.
- 27 Negoro K, McGovern DP, Kinouchi Y *et al*: Analysis of the IBD5 locus and potential gene–gene interactions in Crohn's disease. *Gut* 2003; **52**: 541–546.
- 28 Torok HP, Glas J, Tonenchi L *et al*: Polymorphisms in the DLG5 and OCTN cation transporter genes in Crohn's disease. *Gut* 2005; **54**: 1421–1427.
- 29 Peltekova VD, Wintle RE, Rubin LA *et al*: Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* 2004; **36**: 471–475.
- 30 Noble CL, Nimmo ER, Drummond H *et al*: The contribution of OCTN1/2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease. *Gastroenterology* 2005; **129**: 1854–1864.
- 31 Newman B, Gu X, Wintle R *et al*: A risk haplotype in the solute carrier family 22A4/22A5 gene cluster influences phenotypic expression of Crohn's disease. *Gastroenterology* 2005; **128**: 260–269.
- 32 Vermeire S, Pierik M, Hlavaty T *et al*: Association of organic cation transporter risk haplotype with perianal penetrating Crohn's disease but not with susceptibility to IBD. *Gastroenterology* 2005; **129**: 1845–1853.
- 33 Waller S, Tremelling M, Bredin F, Godfrey L, Howson J, Parkes M: Evidence for association of OCTN genes and IBD5 with ulcerative colitis. *Gut* 2005; **55**: 809–814.
- 34 Daly MJ, Rioux JD: New approaches to gene hunting in IBD. *Inflamm Bowel Dis* 2004; **10**: 312–317.
- 35 Purcell S, Sham P, Daly MJ: Parental phenotypes in family-based association analysis. *Am J Hum Genet* 2005; **76**: 249–259.
- 36 Palmieri O, Latiano A, Valvano R *et al*: Variants of OCTN1-2 cation transporter genes are associated with both Crohn's disease and ulcerative colitis. *Aliment Pharmacol Ther* 2006; **23**: 497–506.
- 37 Vasiliauskas EA, Plevy SE, Landers CJ *et al*: Perinuclear anti-neutrophil cytoplasmic antibodies in patients with Crohn's disease define a clinical subgroup. *Gastroenterology* 1996; **110**: 1810–1819.