npg

ARTICLE

Association of *DLG5* R30Q variant with inflammatory bowel disease

Mark J Daly[1], Alexandra V Pearce[2], Lisa Farwell[1], Sheila A Fisher[2], Anna Latiano[3], Natalie J Prescott[2], Alastair Forbes[4], John Mansfield[5], Jeremy Sanderson[6], Diane Langelier[7], Albert Cohen[8], Alain Bitton[9], Gary Wild[9], Cathryn M Lewis[2], Vito Annese[3], Christopher G Mathew*[2] and John D Rioux*[1]

[1]*The Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA;* [2]*Department of Medical & Molecular Genetics, Guy's, King's and St Thomas' School of Medicine, London, UK;* [3]*CSS – IRCCS Hospital, San Giovanni Rotondo, Italy;* [4]*St Mark's Hospital, Harrow, UK;* [5]*Department of Gastroenterology & Hepatology, University of Newcastle upon, Tyne, UK;* [6]*Department of Gastroenterology, Guy's and St Thomas' Hospitals, London, UK;* [7]*Montreal Jewish General Hospital, Montreal, Quebec, Canada;* [8]*Centre Hospitalier de Sherbrooke, Sherbrooke, Quebec, Canada;* [9]*Department of Gastroenterology, McGill University Health Centre, McGill University, Montreal, Quebec, Canada*

Crohn's disease (CD) and ulcerative colitis (UC) are chronic inflammatory diseases of the gastrointestinal system known as the inflammatory bowel diseases (IBD). Recently, Stoll and colleagues reported a novel finding of genetic variation in the *DLG5* gene that is associated with IBD (CD and UC combined). We present here a study of the genetic variation described in that report in two well-powered, independent case–control cohorts and one family-based collection, and confirm the proposed association between IBD and the R30Q variant of *DLG5* in two of the three studies. We are, however, unable to replicate the other proposed association to the common haplotype described in Stoll *et al* and suggest that this other finding could conceivably have been partially a statistical fluctuation and partially a result of LD with the replicated R30Q association. This study provides support for the hypothesis that *DLG5* constitutes a true IBD risk factor of modest effect.

Introduction

Numerous genome-wide linkage scans of families with multiple individuals affected with inflammatory bowel disease (IBD) have been reported since 1996. These genome-wide searches have identified many putative loci believed to contain disease susceptibility genes. Two such regions of linkage have yielded to association studies,[1–5] resulting in the identification of genetic variants that are significantly associated with Crohn's disease (CD) and that have subsequently been confirmed by many follow-up studies: *CARD15* (MIM*605956) and *IBD5* (MIM*606348).[6] In 1999, a genome-wide linkage study of 353 affected sibling pairs from the UK and Germany reported modest linkage (LOD = 2.3) over a broad region (approximately 40 cM) of chromosome 10.[7] Recently, Stoll and colleagues performed a comprehensive association mapping study of this linked region in a German–British cohort and identified variation at *DLG5* that was associated with IBD.[1] These authors demonstrated that the *DLG5* gene is

Correspondence and shared senior authorship: Dr JD Rioux, Broad Institute, One Kendall Sq., Bldg 300, Cambridge, MA 02139, USA.
Tel: +1 617 252 1917; Fax: +1 617 252 1902
or Prof. CG Mathew, Guy's King's and St Thomas' School of Medicine, 8th floor, Guy's Tower, Guy's Hospital, London SE1 9RT, UK
Received 19 October 2004; revised 17 January 2005; accepted 21 January 2005

located within a region of strong linkage disequilibrium, having four common haplotypes, which is independent from the surrounding genes because of substantial recombination. Importantly, the most significant association signals were located within the haplotype block containing the *DLG5* gene. It was reported that alleles that specifically tagged the most common (frequency = 34%) haplotype (labeled as the 'A' haplotype) were significantly undertransmitted in IBD trios and therefore potentially represented a protective haplotype. In addition, it was reported that two missense mutations, R30Q and P1371Q (also known as 113A and 4136C, respectively), conferred significant risk to developing IBD. R30Q uniquely tagged the less common (10%) haplotype (labeled 'D' in Stoll *et al*), whereas P1371Q was a rare alteration (2–3%). In their study, significant association was observed in a large set of trios (>500), and then confirmed in an independent cohort of over 500 cases and controls. It was unclear, however, to what extent linkage disequilibrium between the proposed causal variants might be contributing to the appearance of multiple association signals (since the presence of an overtransmitted allele or haplotype requires that the alternate allele or haplotype(s) have a net undertransmission). Given the potential importance of this finding for our understanding of the genetic mechanisms of susceptibility to IBD, it is essential that significant evidence of independent replication be reported before embarking on costly functional analyses of this gene.

Materials and methods

Since the original report by Stoll *et al* described association to the broad definition of IBD (both CD and ulcerative colitis (UC) combined), we therefore decided to study the putative associated variation in *DLG5* in a well-powered replication cohort of samples from three European-derived populations and consider both CD and UC samples together. Specifically, we tested an htSNP for the 'A' haplotype as well as the R30Q and P1371Q missense variants (rs2289311, rs1248696 and rs2289310, respectively) in samples collected from hospitals in Canada (a CD population collected from multiple sites in the province of Quebec), Italy (a mixed CD and UC population enrolled at S Giovanni Rotondo 'CSS' Hospital) and the United Kingdom (a mixed CD and UC British population described previously[8]) (Table 1). In all populations considered, the diagnosis of IBD and classification as CD or UC was confirmed by established criteria of clinical, radiological and endoscopic analysis, and from histology reports.[9,10] The Italian and Quebec samples (Table 1; Experiment 1) were typed in tandem in the same laboratory using the Sequenom MassArray system and, in advance of the replication analysis, these samples were typed on an additional 10 *DLG5* SNPs, which were evaluated and

Table 1 Description of association experiments.

Experiment	Study design	Sample description[a]	Country	Typing[b]
1	Case/control	249 CD patients 90 UC patients 207 controls	Canada, Italy	A
2	Case/control	353 CD patients 336 UC patients 493 controls	UK	B
3	Trios	306 affected offspring (88% CD)	Canada, UK	A, B

[a]All samples are independent from patients/families described in Stoll *et al*. The samples in Experiment #3 are independent from Experiments 1 and 2. In Experiment #1, 166 CD patients and 104 controls were from Canada and the remaining were from Italy. In Experiment 2, an additional set of 630 IBD cases and 152 controls from the UK were typed at site B for the Haplotype A tag SNP and included in the analysis in Table 2 to confirm the lack of association to this previously reported finding. Expanded details on Experiment 3 are available in Table 3.
[b]The typing of the variants was performed at two different sites: A, The Broad Institute using the Sequenom platform; B, Guy's, King's and St Thomas' School of Medicine using the TaqMan platform for haplotype A tag SNP and automated sequencing for R30Q.

confirm the precise haplotype structure and frequencies described in the Stoll *et al* paper (Supplementary Information). Owing to their individually modest sizes and joint genotyping, these two samples (Italian and Quebec) were combined for analysis (the control frequencies and ratio of cases to controls were identical in these populations) and constitute replication sample 1. The large UK sample (replication sample 2) was typed using the ABI TaqMan system (for the haplotype 'A' tag SNP) and by direct resequencing (R30Q). In addition to the case/control samples, the UK and Quebec patient collections included parent–parent-affected offspring trios – these family-based samples did not overlap with the case/control samples nor with the previously published samples in Stoll *et al*, and were combined into a third sample to examine the putatively associated alleles in a robust family-based design (replication sample 3). Case–control associations were evaluated with 2×2 χ^2. Homogeneity across studies was evaluated using the Breslow–Day test. The family-based samples were evaluated with TRANSMIT.[11]

Results

In sample 1 (Quebec/Italy), we found significant replication of association to the minor allele of R30Q ($\chi^2 = 7.8$; $P = 0.003$), but not to the 'A' haplotype ($\chi^2 = 1.4$; $P = 0.12$)

(Table 2). These results strongly confirm the association to the common R30Q variant described in Stoll *et al*, but do not replicate the association to the more common 'A' haplotype. While a trend toward replication of haplotype 'A' is present ($f_{cases} = 36.5\%$, $f_{controls} = 40.1\%$), more than half of that difference would be eliminated if we accept and correct for the effect at R30Q ($f_{cases} = 11.0\%$, $f_{controls} = 5.9\%$, odds ratios (OR) = (1.2–3.1)). The P1371Q site appeared rare in this sample, but was not robustly assayed – for both reasons we could not evaluate its association in this population.

A further replication study was simultaneously attempted on a larger mixed IBD cohort collected in the UK. These samples (Table 1; Experiment 2), completely independent from the original cohort described in Stoll *et al* that was used to identify this association, were typed for the same three variants. As can be seen from the results summarized in Table 2, neither the R30Q variant ($f_{cases} = 9.3\%$, $f_{controls} = 9.7\%$) nor the haplotype 'A' effect ($f_{cases} = 34.0\%$, $f_{controls} = 35.7\%$) was replicated in IBD cases. P1371Q was not significantly associated in this population ($f_{cases} = 3.6\%$, $f_{controls} = 3.1\%$). These results, specifically for R30Q, are in apparent contrast with the results of the first replication attempt (Figure 1).

We then sought to examine the discrepancies between the two case/control studies further on both numerical and phenotypic grounds. From a numeric standpoint, the results appear inconsistent: power analysis indicates that the second study had >90% power to detect the effect originally described by Stoll *et al* (OR of 1.62), while a replication of the significance observed in the pooled Canadian and Italian cohorts ($P < 0.003$) is clearly unlikely to have arisen by chance in two attempts. Indeed, a formal test of heterogeneity excludes the hypothesis that these two studies are statistically consistent ($P < 0.01$, Breslow–Day test for homogeneity), but it is worth noting that the 95% confidence intervals of the replication attempts overlap in a small region corresponding to an OR between 1.2 and 1.25.

Since these studies were statistically inconsistent, we attempted to see if there were any obvious differences between the studies. Extensive quality examination of genotype data was undertaken to see whether one of the studies could have mistyped R30Q sufficiently to induce a false-positive or false-negative association. While different technologies were utilized, both data sets generated data for this SNP that appeared to be of high quality, with very

Table 2 Summary of association results from case–control experiments

Variant	Expt.	Allele	Cases	Controls	P-value
Haplotype A tag[a]	1	C	415	243	
		T	239 (0.365)	165 (0.401)	NS
	2	C	1742	829	
		T	896 (0.340)	461 (0.357)	NS
R30Q[b]	1	A	73 (0.110)	24 (0.059)	0.003
		G	591	380	
	2	A	128 (0.093)	96 (0.097)	NS
		G	1250	890	

NS = not significant.
[a]T allele of this variant, also known as rs2289311, tags the 'A haplotype' described by Stoll and colleagues.[1] As noted in Table 1, additional samples were added to the 'A haplotype' analysis in Experiment 2.
[b]A allele of this variant, also known as rs1248696 (or as 113A in Stoll *et al*), codes for the minor Q allele.
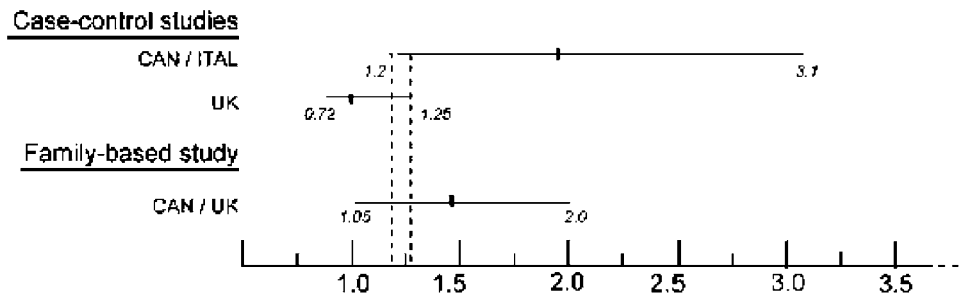
Figure 1 Estimated OR for the R30Q variant in our two case–control and one family-based (father–mother–patient trios) association studies (Experiments 1–3). Each horizontal bar indicates the 95% confidence interval for each OR. The dashed vertical lines indicate the region of overlap between these three independent studies and correspond to an OR between 1.2 and 1.25.

Table 3 Summary of association results from family-based experiment (Experiment #3)

			Number of transmissions[a]		
SNP	Population	No. of affected offspring	Observed	Expected	P-value[b]
Haplo A tag	Canada	183	131	126.2	NS
	UK	289	173	180.8	NS
	Pooled	472	304	307.8	NS
R30Q	Canada	182	43	38.7	NS
	UK	124	33	26.4	0.017
	Pooled	306	76	65.1	0.018

NS = not significant.
[a]Calculated using TRANSMIT.[11]
[b]Obtained from 5000 bootstrap simulations, one-tailed.

few missing data points, expected allele frequencies and conformance with Hardy–Weinberg equilibrium. The Italian and Quebec samples, typed on the Sequenom system, had typing that created the precise haplotype structure and frequencies as observed in the previously published German population.[1] In addition, no Mendelian inheritance errors were discovered in more than 100 Quebec families from which the TDT sample (Table 3) was drawn and manual review of the mass spectrometry traces was performed as a final check of allele calling. In the UK samples, manual review of all sequence-based calls were made, and an independent set of samples typed both in the UK and in Germany (described previously[1]) were completely concordant. Having established high quality for both data sets, we then asked whether phenotypic differences between the two replication studies could shed light. In fact, significant differences in patient phenotypes were observed, but none of these appeared to correspond to the replication differences observed here (for example, study 1 (73% CD) showed replication of R30Q, while study 2 (51% CD) did not; however, study 1 showed significant association in both the CD and UC samples ($P < 0.01$ in both) and division of study 2 along CD/UC bounds did not reveal a consistent CD association). In fact, the Breslow–Day test indicates the same significant lack of homogeneity between the two studies when considering the CD category from each sample alone. Since differences in the patient populations do exist, however, we cannot rule out that an unexamined phenotypic difference in these samples would, in fact, explain the incompatibility.

As noted, in the population samples from Quebec and the UK, smaller and completely independent family-based collections were also available. For a final evaluation of R30Q, we combined these samples in a TDT analysis of parent–parent-affected offspring trios. In both samples, modest excess transmission of allele Q was seen (combined sample producing a $P = 0.018$), further indicating that R30Q appears to constitute a true IBD risk factor. In the parents of affected offspring, the frequency of the risk allele was concordant in both geographical groups (10.5%

Quebec parents, 11.0% UK parents). In both populations, these frequencies are modestly elevated above the population mean as would be expected in the case of a true risk factor, since we have conditioned on these individuals having an offspring affected with a rare disease. Similar family-based evaluation of the proposed haplotype 'A' association again observed no association (Table 3).

Discussion

Thus, in total, it appears that the most parsimonious explanation of the set of observations is that R30Q does influence IBD risk modestly (perhaps with an OR in the vicinity of 1.25 – compatible with the overlap between the two case–control studies and the estimate from the family-based sample) and that heterogeneous replication results are likely due to phenotypic differences among collections. However, it is important to emphasize that, although phenotypic differences may exist, very similar criteria were used by physicians in all centers to establish the diagnosis of CD and UC. Even in the presence of identical diagnostic criteria though, ascertainment and environmental exposures can differ and lead to patient population differences as well. As the results of these replication studies were not statistically compatible, this gene should be examined in a larger consortium effort with more complete power and with large samples of specific subphenotypes of disease to evaluate the potential phenotypic specificity of this association. If the effect is as modest as suggested here (weaker than originally reported by Stoll *et al*), the power of each individual study presented here is below 50% – several thousand cases and controls would be required to have >90% power to replicate this effect with confidence ($P < 0.01$). Thus, the observation that one of the three attempts replicated R30Q below 0.01, with another below 0.05, while clearly unlikely under the null hypothesis of no effect, actually constitute a rather routine and representative set of results expected of a true but modest effect. Rather than the universal validation of the *CARD15* and

IBD5 results, these data resemble the similar effects of variants in *PPARG* and *SUR1/KIR6.2* in type 2 diabetes, where many individual samples failed to see significant effects, but for which there is now highly significant pooled evidence across many studies.[12,13]

References
1 Stoll M, Corneliussen B, Costello CM *et al*: Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nat Genet* 2004; **36**: 476–480.
2 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.
3 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.
4 Hampe J, Cuthbert A, Croucher PJ *et al*: Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001; **357**: 1925–1928.
5 Rioux JD, Daly MJ, Silverberg MS *et al*: Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001; **29**: 223–228.
6 Wild GE, Rioux JD: Genome scan analyses and positional cloning strategy in IBD: successes and limitations. *Best Pract Res Clin Gastroenterol* 2004; **18**: 541–553.
7 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.
8 Cuthbert AP, Fisher SA, Mirza MM *et al*: The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. *Gastroenterology* 2002; **122**: 867–874.
9 Lennard-Jones J: Classification of inflammatory bowel disease. *Scand J Gastroenterol* 1989; **24**: 2–6.
10 Podolsky DK: Inflammatory bowel disease. *N Engl J Med* 1991; **325**: 928–937.
11 Clayton D: A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am J Hum Genet* 1999; **65**: 1170–1177.
12 Altshuler D, Hirschhorn J, Klannemark M *et al*: The common PPARg Pro12Ala polymorphisms is associated with decreased risk of type 2 diabetes. *Nat Genet* 2000; **26**: 76–80.
13 Gloyn AL, Weedon MN, Owen KR *et al*: Large-scale association studies of variants in genes encoding the pancreatic B-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* 2003; **52**: 568–572.
14 Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.

Supplementary Information accompanies the paper on the European Journal of Human Genetics website (http://www.nature.com/ejhg)