

ARTICLE

Association of NOD2 with Crohn's Disease in a homogenous Irish population

Emer Bairead¹, Dawn L Harmon², Anne M Curtis¹, Yvette Kelly¹, Clare O'Leary¹, Michelle Gardner², Dermot T Leahy², Pat Vaughan⁶, Denise Keegan⁴, Colm O'Morain³, Diarmuid O'Donoghue⁴, Fergus Shanahan¹, Nollaig A Parfrey^{2,5} and Kathleen A Quane¹

¹Department of Medicine, University College Cork, Ireland; ²Department of Pathology, University College Dublin, Ireland; ³Department of Gastroenterology, Adelaide and Meath Hospitals, Tallaght, Dublin, Ireland; ⁴Department of Gastroenterology, St Vincents University Hospital, Dublin, Ireland; ⁵Department of Pathology, University College Cork, Ireland; ⁶HiberGen Ltd, IDA Business Park, Bray, Co. Wicklow, Ireland

Linkage of IBD to the pericentromeric region of chromosome 16 has been widely confirmed by analyses of multiple populations. The NOD2 gene is located in the peak region of linkage on chromosome 16 and thought to be involved in the activation of nuclear factor (NF) κ B in response to bacterial components. Mutations in the NOD2 gene are found to be strongly associated with susceptibility to Crohn's disease (CD). A total of 65 Irish CD families were genotyped to determine if NOD2 mutations conferred susceptibility to CD and the prevalence of these mutations in sporadic and familial forms of the disease. The Irish population is relatively homogenous and thus may provide advantages in genetic studies of complex diseases. We confirmed the IBD1 locus as a susceptibility locus for IBD within the Irish population by linkage analysis followed by linkage disequilibrium studies. No significant evidence of linkage was observed to the previously identified regions on chromosomes 1, 12 and 14. In all, 131 CD affected families were then genotyped for seven of the previously published NOD2 single-nucleotide polymorphisms (SNPs). Allelic transmission distortion was investigated using the pedigree disequilibrium test (PDT). SNP13 (3020insC) was found to be associated with CD ($P=0.0186$). Patients who possessed a rare allele of SNP8, 12 or 13 presented earlier when compared to patients without rare variants (mean age, 20.1 vs 24 years, $P=0.011$) and the rare allele of SNP13 was observed to be predominantly linked to ileal disease ($P=0.02$). This report confirms the importance of NOD2 as a susceptibility gene for CD within the Irish population.

European Journal of Human Genetics (2003) 11, 237–244. doi:10.1038/sj.ejhg.5200954

Keywords: inflammatory bowel disease; Crohn's disease; ulcerative colitis; linkage analysis; NOD2; linkage disequilibrium

Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are the major forms of chronic idiopathic inflammatory bowel

disease (IBD). They occur in young adults with a prevalence of at least one per thousand.¹ Familial aggregation of IBD and greater concordance in monozygotic than dizygotic twins,² indicate a genetic contribution to IBD that is stronger for CD than UC. The relative risk to siblings of affected relatives (λ_s) has been estimated at 36.5 for CD, 16.6 for UC and 24.7 for IBD (CD, UC and indeterminate colitis).³

Genome-wide screens in affected relative and sibling pairs provided significant evidence of linkage on chromo-

*Correspondence: Dr KA Quane, Department of Medicine, University College Cork, Clinical Sciences Building, Cork University Hospital, Cork, Ireland. Tel: +353 21 490 1278; Fax: +353 21 434 5300; E-mail: k.quane@ucc.ie

Received 13 June 2002; revised 17 December 2002; accepted 18 December 2002

some 16 for CD.⁴ This region has been confirmed as a CD susceptibility locus in numerous replication studies in different populations, including an international IBD consortium (as reviewed in Cavanaugh⁵). To date, IBD1 appears to be the most consistently replicated region of linkage in inflammatory bowel disease.

NOD2 is located in the peak region of linkage on chromosome 16 and was found to induce NF- κ B activation in monocytes.⁶ Following a frameshift mutation (caused by a cytosine insertion and known as 3020insC or single-nucleotide polymorphism (SNP)13) leading to the truncation of the leucine-rich repeat region of NOD2, responsiveness to bacterial lipopolysaccharide (LPS) was observed to be greatly diminished.⁶ Associations between this frameshift mutation and CD have been reported in US, UK and European populations.⁶⁻⁸

Genomewide screens have also provided significant evidence of linkage on chromosomes 1,^{7,8} 3,⁷ 6,⁹ 10,⁹ 12,³ 19¹⁰ and X¹¹ to IBD while there have been two independent reports of linkage to CD to chromosome 14q11-12.^{12,13}

In this study, we tested the presence of the chromosome 16, 14, 12 and 1 susceptibility loci in 65 Irish families. We then analysed the contribution of the NOD2 SNPs to CD within the Irish population.

Materials and methods

Family ascertainment and recruitment

Following Ethics Committee approval, recruitment of the IBD families was carried out on a nation-wide basis. Peripheral blood samples were collected from affected individuals and unaffected family members. Informed written consent was obtained from all individuals and affected members were asked to complete a detailed clinical questionnaire. Individuals not of Irish origin were excluded from this study. All affected individuals were Caucasian and had parents that were born in Ireland. DNA was extracted and stored at -20°C prior to genotyping.

A total of 65 multiplex families were recruited (see Table 1 for an overview of family structures). Sib-pair analysis used 77 ASP from these recruited multiplex

Table 1 Families and genotype-affected sibling pairs

	No. of families	Sib-pairs	% of families with both parents	% of families with one parent
CD families	21	29	48	24
UC families	20	20	60	10
Mixed families	24	28	58	8
Total combined disease	65	77		

Table 2 Microsatellite markers and intermarker distances according to Genethon map

Chromosome	Marker	Intermarker distance (cM)
1	D1S228	
	D1S552 ^a	5
	D1S2644	9
	D1S2885	13
12	D12S368	
	D12S83	11
	IFN γ	6
	D12S1708	13
16	D16S401	
	D16S409	10
	D16S3120	8
	D16S419	4
14	D14S261	
	D14S72	2
	D14S990	6
	D14S264	8

^aD1S552 is not a Genethon marker (see text).

families. The 65 families contained 143 affected individuals and 200 unaffected relatives. In addition, 109 CD simplex families consisting of an affected individual and both parents were recruited.

Linkage analysis

A total of 12 microsatellite markers were chosen to span previously identified regions of interest on chromosomes 1, 12, 14 and 16 (Table 2). The polymorphic microsatellites and initial corresponding intermarker distances were from the Genethon map,¹⁴ and database (<ftp://ftp.genethon.fr/pub/Gmap/Nature-1995/data/>) with the exception of D1S552, a CHLC marker. The approximate position of D1S552 relative to Genethon markers was obtained using integrated genome maps available at the National Centre for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>). Fluorescent dye-labelled forward primers and unlabelled reverse primers were custom-synthesised (Applied Biosystems and MWG Biotech). Following multiplex PCRs, amplified products were analysed using an ABI PRISMTM 310 Genetic Analyser (Applied Biosystems). Allele sizes were determined by use of the ABI GENESCAN software (version 2.0) (Applied Biosystems). PedCheck was used to eliminate genotype errors and to check for misinheritance.¹⁵ Allele frequencies for each microsatellite locus were determined using 72 unrelated, nonaffected founder members from the multiplex families.

Pedigrees were primarily analysed by disease subgroup (CD, UC or mixed), then combined to generate an overall result for IBD, with the exception of chromosome 14. The chromosome 14 locus was originally identified as

responsible for susceptibility to CD^{12,13} in two independent studies in the US. However, suggestive evidence of linkage to chromosome 14 was not observed in CD families from another North American panel, but in mixed families.⁷ Therefore, markers on chromosome 14 were genotyped and analysed in the CD and mixed families only. Multipoint nonparametric linkage (NPL) analysis was performed by use of the 'S_{all}' scoring function in the GENEHUNTER-PLUS¹⁶ modification of the GENEHUNTER package.¹⁷ This modification also permits the calculation of a multipoint LOD (MLOD) score based on a linear allele-sharing model.¹⁶ Exclusion maps were constructed for the loci on chromosomes 1, 12 and 14 using GENEHUNTER¹⁷ and λ_s values of 1.2, 1.5, 2.0, 2.5, 3.0 and 5.0. Linkage was considered excluded given a LOD score of less than -2.0 for a specified value of λ_s .

Microsatellite saturation

A further 10 microsatellite markers were chosen to saturate the identified region of interest on chromosome 16. The polymorphic microsatellites were from the Genethon map,¹⁴ and database (<ftp://ftp.genethon.fr/pub/Gmap/Nature-1995/data/>) and the MarshMed database (www.marshmed.org/genetics/). Positioning of markers was carried out using integrated genome maps available at the National Centre for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>). Fluorescent dye-labelled forward primers and unlabelled reverse primers were custom synthesised (Sigma Genosys) (Table 4). Multipoint analysis of the 21 CD families consisting of 29 ASP was carried out using the 'S_{all}' scoring function by the program GENEHUNTER-PLUS¹⁸ as detailed above.

Linkage disequilibrium analysis

Linkage disequilibrium analysis was subsequently carried out in CD families within the region of most positive linkage. In all, 13 microsatellite markers were analysed in 421 individuals consisting of 121 simplex families and 10 CD multiplex families using the pedigree disequilibrium test (PDT).^{18,19}

Mutation detection and analysis

Analysis of SNPs on NOD2 was carried out using a combination of SNaPIT mutation detection method^{20,21} and restriction endonuclease digestions. Information on the relevant SNPs was obtained from the SNP database <http://www.ncbi.nlm.nih.gov/SNP>.

In brief, primers were designed to amplify seven of the published SNPs identified on NOD2 using the computer program OLIGO™ Version 4.0²² (Table 3). Following amplification, six SNPs were genotyped using SNaPIT mutation detection method, which is a specific, versatile method of mutation detection and has previously been used both in large- and small-scale mutation detection.²³ Primers were designed so that during amplification the position of the first uracil incorporated into the extended primers differed depending on whether a polymorphism was present or absent. Subsequent glycosylase excision of the uracil residues followed by cleavage of the apyrimidinic sites allowed detection of the polymorphism in the amplified fragment as a fragment length polymorphism.²⁰ During optimisation of this technique a random number of samples were genotyped using both the SNaPIT method and *MspI* and *HpaII* restriction endonucleases that showed 100% correlation in allele detection. SNP12 (using the nomenclature from Hugot *et al*²⁴) was genotyped utilising the restriction endonuclease *HinPII* (New England Biolabs) as it was not suitable for use with the SNaPIT method of mutation detection.

Following digestion as outlined,²⁰ SNaPIT- and restriction endonuclease-treated samples were analysed on an ABI PRISM™ 310 Genetic Analyser (Applied Biosystems). Heterozygous and homozygous individuals were selected for use as controls to ensure consistent identification of the relevant SNPs between runs.

SNP data were analysed for linkage disequilibrium using the PDT 3.12 program,^{18,19} which analyses all potentially informative data from extended as well as individual pedigrees. A healthy control population was genotyped for the SNPs 8, 12 and 13 polymorphisms and allele frequencies for these SNPs were determined. Fisher's exact test and Student's *t*-test were used to compare qualitative and quantitative variables between groups, respectively.

Table 3 SNP primer sequence and fragment size generated following SNaPIT method or restriction endonuclease digestion (SNP12)

SNP name	Forward primer	Reverse primer	Fragment size in presence of SNP	Fragment size in absence of SNP
SNP2	gaaacctgaggctctacaagactt	cacgtcaaactgctctttgag	22	25
SNP5	gggcagatgtgggcatgg	ccagcacagtgtccgcatcgt	23	45
SNP6	aggcatcagctgtacctga	gtgcagggtgaggtctcttg	28	30
SNP7	tgacatttctcttgcttct	gcgcaagaagcactggaaa	27	36
SNP8	ccagacatctgagaaggcctgct	tggcgggatggagtggaaagtg	23	34
SNP12	actgacactgtctgtactct	acctgatctcccaagaaaactg	42	166
SNP13	gtatctctttccaggtgtcca	ccaggatggtgtcattcctt	29	28

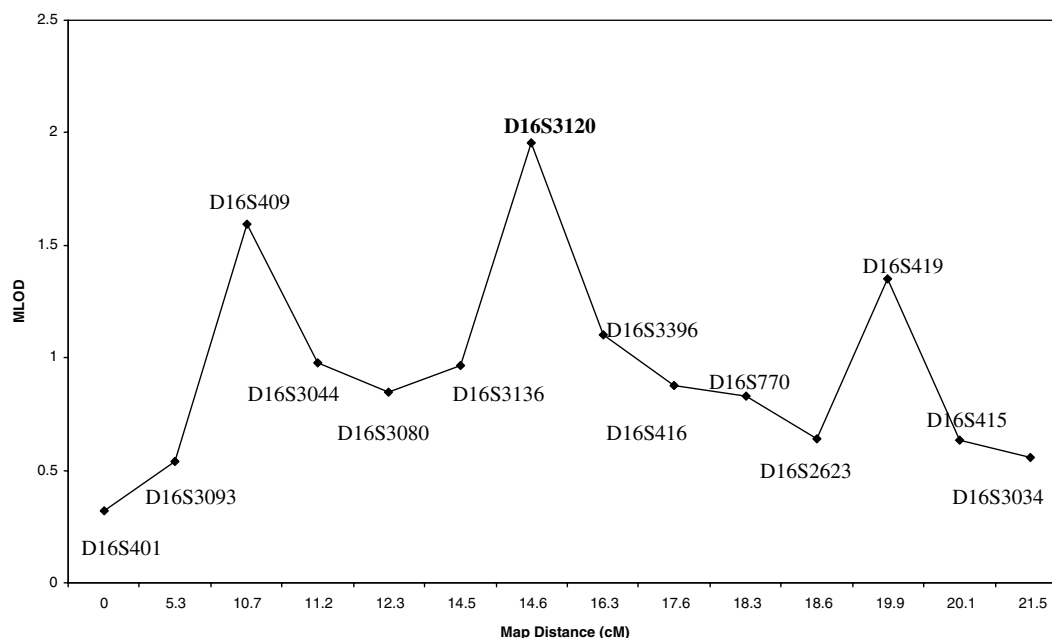


Figure 1 Fine mapping of region of interest on chromosome 16 using CD ASP families. Linkage analysis was performed by using GENEHUNTER-PLUS.

Results

Nonparametric linkage analysis

Multipoint analysis of affected sib-pairs ($n = 77$ pairs) was performed using the 'Sall' algorithm of GENEHUNTER-PLUS.¹⁶ Linkage of CD to chromosome 16 was confirmed²⁵ with a maximum MLOD score of 2.241 at D16S3120 with the UC sib-pairs displaying MLOD scores of zero at this region. There was no significant evidence of linkage to IBD at chromosome 1, 12 or 14 regions (MLOD = 0.786 at D1S552, MLOD = 0.605 at D12S1708, MLOD = 0.107 at D14S264). Using GENEHUNTER,¹⁷ we were able to exclude chromosome 12 as a locus with $\lambda_s = 3.0$ for susceptibility to IBD, although higher λ_s values were needed to exclude CD, UC and mixed subgroups ($\lambda_s > 5$). However, a λ_s value of 2.0 was excluded with respect to the marker D12S83, which was originally reported to confer a sibling relative risk of 2.0 for IBD.³ Chromosome 14 was excluded in mixed disease with a $\lambda_s = 3.0$. However, higher λ_s values were required to exclude chromosome 14 to CD and to exclude chromosome 1 as a susceptibility locus for all subgroups ($\lambda_s > 5$ for all).

Microsatellite saturation experiment

A further 10 microsatellite markers, spanning the interval from D16S3093 to D16S3034 were genotyped in 21 CD multiplex families consisting of 29 ASP with markers at approximate 1.6 cM intervals. Results of GENEHUNTER-PLUS analysis using the 'Sall' algorithm are depicted in graph format (Figure 1). Overall, a triple-peaked configura-

tion of the CD phenotype was observed with the area of maximal MLOD concentrating on the D16S3120 region. Additional peaks were observed near D16S409 and D16S419 markers.

Linkage disequilibrium analysis

A total of 421 individuals from 121 simplex CD families and 10 multiplex CD families were genotyped for 13 microsatellite markers. These multiplex families were selected as both parents and unaffected siblings were available for genotyping, which not only increased the genetic information available for analysis but also acted as further internal controls.²⁶ Analysis of linkage disequilibrium within this data set was carried out using the PDT.^{18,19} PDT analysis of this data set resulted in a significant P -value of 0.0054 at the D16S3080 marker when analysed giving weight to larger families within the data set. This increased to a P -value of 0.017 when weight was averaged out over all families within the data set (Table 4). A weak association was observed for alleles of markers D16S3044 ($P = 0.02$), D16S3080 (Allele 2: $P = 0.009$, Allele 3: $P = 0.008$, Allele 4: $P = 0.04$) and D16S419 ($P = 0.05$). However, after applying the Bonferroni correction for the number of tested alleles, only a weak association between allele 3 (279 bp) of the D16S3080 marker and the CD phenotype remained ($P = 0.05$). Stratification of data on the basis of age of onset < 21 years and disease severity did not cause any significant difference to the observed results (data not shown).

Table 4 Results of PDT analysis using microsatellite markers in the chromosome 16 linkage region

Marker	SUM PDT <i>P</i> -value	AVE PDT <i>P</i> -value	CM
D16S3093	0.9502	0.6983	5.4
<i>D16S409</i>	0.2681	0.3057	0.5
D16S3044	0.2729	0.2722	1.1
D16S3080	0.0054	0.0179	2.2
D16S3136	0.3450	0.7580	0.0
<i>D16S3120</i>	0.4607	0.3701	1.7
D16S3396	0.2305	0.1966	1.3
D16S416	0.5134	0.3456	0.7
D16S770	0.4626	0.4712	0.3
D16S2623	0.6501	0.8493	1.3
<i>D16S419</i>	0.6366	0.1901	0.2
D16S415	0.4823	0.8111	1.4
D16S3034	0.2274	0.2722	

Markers that have been used in the linkage analysis are marked in italics. For each marker, two values obtained from PDT analysis have been given. SUM PDT gives more weight to larger families within the data set. AVE PDT gives equal weight to all families within the data set.

NOD2 SNP analysis

Seven SNPs (SNP2, 5–8, 12 and 13) were genotyped in 121 CD simplex families and 10 CD multiplex families. Following PDT analysis, the SNP13 polymorphism was found to be positively associated with CD with a *P*-value of 0.018 obtained when analysed giving more weight to the larger families within the data set. This value was observed to increase to a *P*-value of 0.028 when weight was averaged out among all of the families within the data set. A *P*-value of 0.0305 was noted for SNP7 when analysed giving weight to the larger families. However, this increased to a nonsignificant *P*-value of 0.0955 when weight was averaged out among all of the families (Table 5). The SNP13 mutation was found to be quite rare with less than 2% of CD patients within the population studied being homozygous and an observed rare allele frequency of 0.04 (Table 5). The frequency of this mutation was observed to be increased in families that were multiply affected with CD than in sporadic cases of CD (22 vs 10%, *P*=0.088); however, this was not observed to be significant.

The rare alleles of SNP8 and 12 did not appear to contribute to the CD risk in this cohort with nonsignificant *P*-values observed. Of the CD patients in this cohort, 22.6% carried a single copy of a NOD2 mutation and 4.8% carried 2 mutations, giving a total of 27.4% of patients with CD carrying at least one copy of a NOD2 mutation. Haplotypes carrying more than one NOD2 mutation (SNP8, 12 and 13) (defined as a 'NOD2 haplotype'²⁸) were not observed in multiplex families but had an estimated frequency of 1.8% in the sporadic cases. Haplotypes carrying all three rare alleles of SNP8, 12 and 13 were not observed in this data set. Allelic frequencies in our control population for SNP8, 12 and 13 (0.04, 0.01, 0.01,

Table 5 Results of PDT analysis of NOD2 SNPs. NOD2 SNP rare allele frequencies as determined from the nonaffected founder members of our cohort are also displayed

NOD2 SNP	SUM PDT <i>P</i> -value	AVE PDT <i>P</i> -value	Rare allele frequency in CD family founders
SNP2	0.4652	0.6729	0.27
SNP5	0.7833	0.7141	0.35
SNP6	0.4568	0.6008	0.50
SNP7	0.0305	0.0955	0.30
SNP8	0.6394	0.8373	0.07
SNP12	0.5485	0.2020	0.03
SNP13	0.0186	0.0280	0.04

For each polymorphism, two values obtained from PDT analysis have been given. SUM PDT gives more weight to larger families within the data set. AVE PDT gives equal weight to all families within the data set.

respectively) were observed to be comparable to other European control populations.^{24,28}

Stratification on NOD2 genotypes

In order to identify the contribution of the NOD2 gene to the original linkage observed on chromosome 16 in the ASP families, we stratified the cohort by removing all of the genotypes of all individuals carrying mutations at SNP7 or 13. For this analysis, all individuals carrying the NOD2 mutations were coded as 'unknown' in the linkage files. MLOD scores at the D16S3120 were observed to drop in significance (from MLOD 2.2 to MLOD 1.9). However, the MLOD scores at the D16S409 and D16S419 locus did not show as significant a drop as displayed at the D16S3120 locus.

NOD2 SNPs and predominant disease distribution

NOD2 mutations have been recently found to be associated with ileal disease.^{28,29} Patients in this cohort who possessed a rare allele of SNP8, 12 or 13 presented earlier when compared to patients without rare variants (mean age, 20.1 vs 24 years, *P*=0.011). An analysis of the available patient information in this cohort indicates that possession of a copy of SNP13 coincides with predominantly ileal disease in 92% of individuals (*P*=0.02). SNP12 and 8 did not appear to be predominantly associated with ileal disease (nonsignificant *P*-values observed). However, as the numbers involved are relatively small (*n*=16, 6, 18 for SNP8, 12 and 13, respectively), a detailed comparative study of the contribution of NOD2 gene mutations to primary disease location in additional Irish CD patients is currently in progress.

Discussion

Genetic linkage of CD to the IBD1 locus on chromosome 16 has been confirmed in many independent groups and

by a large collaborative study.⁵ Altogether, the results of all the independent studies, including this one, are highly consistent, with the recent identification of NOD2 as an IBD susceptibility gene confirming the importance of linkage results on chromosome 16. Nonparametric linkage analysis of the CD-affected sib-pairs, confirmed that a susceptibility locus for CD within the Irish population is located in the pericentromeric region of chromosome 16. Significant MLOD values were obtained at D16S3120 and approaching significance at D16S409.

On chromosome 1, 12 and 14, combined analysis generated positive but nonsignificant LOD scores (MLOD = 0.786 at D1S552, MLOD = 0.605 at D12S1708, MLOD = 0.107 at D14S264). In our study, none of the disease subgroups displayed significant evidence of linkage to these regions, although λ_s values required to exclude loci ranging from 3.0 to >5.0. To date, there has been no suggestion in any previous study that any IBD locus would give such a sibling risk. It has been observed that it is often difficult to replicate linkages within the same disease, with the findings in studies of the same disease often being inconsistent.³⁰ It appears that considerably more families are required to detect susceptibility loci during replication analyses, than for the initial observations in genome-wide screening particularly in the case of loci of modest effect.³¹ Therefore, the lack of linkage to these regions in our replication analysis could be because of the relatively small sample size analysed, and weak susceptibility genes may yet exist in these regions. Alternatively, lack of significant evidence of linkage to these regions may possibly be due to the fact that these loci might not play as significant a role in susceptibility to IBD in the cohort analysed, possibly due to disease or locus heterogeneity. However, this may only be confirmed following the analysis of larger numbers of families for these regions.

Microsatellite saturation of the region of most positive linkage on chromosome 16 in ASP families observed a triple-peaked result with the area of maximal MLOD concentrating on the D16S3120 region, with additional peaks observed near the D16S409 and D16S419 microsatellite markers. In the initial report, the IBD1 region spanned more than 40 cM and two general peaks were observed.⁴ In subsequent studies, peaks in CD families have been reported for many different markers on chromosome 16 (as reviewed in Cavanaugh⁵). Therefore, our result is consistent with other studies in this region and may provide further evidence that more than one susceptibility gene for IBD resides on chromosome 16.

In order to further refine the IBD1 localisation in this region, we looked for an association using the genotyped microsatellite markers. A weak association was observed for CD with the 279 bp allele of D16S3080 ($P=0.05$). This marker is located within 3 cM from the region with the strongest evidence for linkage in our CD ASP multipoint analysis and falls within our detected strongest support

interval for linkage (D16S409–D16S3120). The small difference in positioning from linkage analysis vs PDT analysis could possibly be due to the overlapping but not identical sets of families used in the analyses and the limitations that ASP/haplotype sharing methods are known to have in the fine mapping of complex traits.³² Thus, altogether, linkage studies and linkage disequilibrium analysis centred the same pericentromeric region on chromosome 16.

Evidence was found in this study of an association of a NOD2 frameshift mutation (SNP13) with CD following family-based association analysis with a higher frequency of SNP13 observed in families that were multiply affected with CD than in sporadic cases of CD. However, SNP7 was only associated with CD when more weight was given to the multiplex families within the data set. This result could be due to statistical fluctuation or to underlying genetic differences between multiply affected CD families and sporadic cases. Ogura *et al*⁶ noted that the applicability of NOD2 associations to the more common, sporadic cases would warrant further study. Case-control studies have shown that NOD2 also significantly contributes to the sporadic group.^{33,24,28} Nevertheless, the frequency of the SNP13 polymorphism was previously observed to be significantly higher in individuals from pure CD families than affecteds from mixed families or sporadic CD cases.²⁸ Following genotype-phenotype analysis the SNP13 polymorphism appears to be associated with both a younger age of onset ($P=0.011$) and, similar to other studies,^{28,29} a predominantly ileal disease distribution ($P=0.02$). Analysis of NOD2 polymorphisms in a further cohort of CD patients is currently in progress.

The nonsignificant excess of transmission of the SNP8 and 12 mutations could be because of their very low frequency within the population, as only a small number of transmissions will be informative in even a large cohort.²⁸ Therefore, lack of evidence of association of these SNPs could be because of the relatively small sample size analysed.

On stratification of NOD2 genotype, the central linkage peak drops in significance from a MLOD of 2.2 to MLOD 1.9, indicating that NOD2 played a role in allele sharing in this region. The relatively unchanged flanking peaks may indicate that further susceptibility genes apart from NOD2 exist on chromosome 16. Evidence for an NOD2-independent susceptibility locus close to the D16S3068 microsatellite marker has been observed by Hampe *et al*.³⁴ in the German population, which may lead to the discovery of additional susceptibility genes in this region of chromosome 16.

In summary, we present evidence in this report of an association between NOD2 and CD within the Irish population. No significant evidence of linkage was observed between the chromosome 1, 12 and 14 loci and IBD in this data set. Relative to Britain and much of mainland

Europe, the ethnic and genetic mix of Ireland is long established and relatively homogenous.^{35–37} Thus, analysis of this population may provide advantages in genetic studies of complex diseases.

In terms of position and function, NOD2 is an excellent candidate gene for IBD. Other potentially disease-causing mutations on NOD2 have also been recently identified by Lesage *et al*²⁷ and the role of both these, and the previously identified polymorphisms in the pathophysiological mechanisms of CD and the treatment of this disease are currently under intensive investigation. Hopefully, the precise role of NOD2 in innate immune responses in the gastrointestinal tract may be clarified, thus contributing to the identification of associated environmental factors and guiding the search for specific therapies.

Acknowledgements

We thank the patients and their families who provided blood samples for this study and the 'Irish Society of Colitis and Crohn's Disease' and 'Irish Ileostomy and Internal Pouch Association' for explaining the purposes of the study to members. We thank the Irish Society of Gastroenterology. We acknowledge the helpful contribution of Dr Daniel E Weeks, University of Pittsburgh, USA, with linkage analysis and Dr Meredyth P Bass, Duke University Medical Centre, USA, with linkage disequilibrium analysis. This work was supported by the Wellcome Trust (Reference no. 054234/Z/98/Z) and the Irish Health Research Board.

References

- Podolsky DK: Inflammatory bowel disease (1). *N Engl J Med* 1991; **325**: 928–937.
- Tysk C, Lindberg E, Järnerot 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.
- 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.
- 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.
- Cavanaugh J: International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn's disease and chromosome 16. *Am J Hum Genet* 2001; **68**: 1165–1171.
- 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.
- Cho JH, Nicolae DL, Gold LH *et al*: Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. *Proc Natl Acad Sci USA* 1998; **95**: 7502–7507.
- Cho JH, Nicolae DL, Ramos R *et al*: Linkage and linkage disequilibrium in chromosome band 1p36 in American Chaldeans with inflammatory bowel disease. *Hum Mol Genet* 2000; **9**: 1425–1432.
- 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.
- 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.
- Vermeire S, Satsangi J, Peeters M *et al*: Evidence for inflammatory bowel disease of a susceptibility locus on the X chromosome. *Gastroenterology* 2001; **120**: 834–840.
- 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.
- Duerr RH, Barmada MM, Zhang L, Pfutzer R, Weeks DE: High-density genome scan in Crohn's disease shows confirmed linkage to chromosome 14q11–12. *Am J Hum Genet* 2000; **66**: 1857–1862.
- Dib C, Faure S, Fizames C *et al*: A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 1996; **380**: 152–154.
- O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.
- Kong A, Cox NJ: Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997; **61**: 1179–1188.
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996; **58**: 1347–1363.
- Martin ER, Monks SA, Warren LL, Kaplan NL: A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet* 2000; **67**: 146–154.
- Martin ER, Bass MP, Kaplan NL: Correcting for a potential bias in the pedigree disequilibrium test. *Am J Hum Genet* 2001; **68**: 1065–1067.
- Vaughan P, McCarthy TV: A novel process for mutation detection using uracil DNA-glycosylase. *Nucleic Acids Res* 1998; **26**: 810–815.
- Wynne F, Drummond F, O'Sullivan K *et al*: Investigation of the Genetic Influence of the OPG, VDR (Fok1), and COL1A1 Sp1 Polymorphisms on BMD in the Irish Population. *Calcif Tissue Int* 2002; **20**: 20.
- Rychlik W, Rhoads RE: A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and *in vitro* amplification of DNA. *Nucleic Acids Res* 1989; **17**: 8543–8551.
- Vaughan P, McCarthy TV: Glycosylase mediated polymorphism detection (GMPD)—a novel process for genetic analysis. *Genet Anal* 1999; **14**: 169–175.
- 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.
- Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; **11**: 241–247.
- Speer M: Sample size and power. In: Haines JL, Pericak-vance MA (eds): *Approaches to gene mapping in complex human diseases*. New York: Wiley-Liss, Inc., 1998; pp. 161–200.
- Lesage S, Zouali H, Cezard JP *et al*: CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002; **70**: 845–857.
- 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.
- Ahmad T, Armuzzi A, Bunce M *et al*: The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology* 2002; **122**: 854–866.
- Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M: Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 2001; **69**: 936–950.
- Suarez BK, Hampe CL, Van Eerdewegh P: Problems of replicating linkage claims in psychiatry. In: Cloninger CR (ed): *Genetic approaches to mental disorders*. Washington DC: American Psychiatric Press, 1994, pp. 23–46.
- Ott J: Linkage disequilibrium: in: *Analysis of human genetic linkage*. New York: The John Hopkins University Press, 1999, pp. 280–291.

- 33 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.
- 34 Hampe J, Frenzel H, Mirza MM *et al*: Evidence for a NOD2-independent susceptibility locus for inflammatory bowel disease on chromosome 16p. *Proc Natl Acad Sci USA* 2002; **99**: 321–326.
- 35 Hill EW, Jobling MA, Bradley DG: Y-chromosome variation and Irish origins. *Nature* 2000; **404**: 351–352.
- 36 Relethford JH: Genetic structure and population history of Ireland: a comparison of blood group and anthropometric analyses. *Ann Hum Biol* 1983; **10**: 321–333.
- 37 Relethford JH, Crawford MH: Anthropometric variation and the population history of Ireland. *Am J Phys Anthropol* 1995; **96**: 25–38.