ARTICLE

Genome-wide search in Finnish families with inflammatory bowel disease provides evidence for novel susceptibility loci

Paulina Paavola-Sakki¹, Vesa Ollikainen^{2,3}, Tiina Heliö¹, Leena Halme⁴, Ulla Turunen⁵, Päivi Lahermo², Maarit Lappalainen¹, Martti Färkkilä⁵ and Kimmo Kontula^{*,1}

¹Department of Medicine, University of Helsinki, Helsinki, Finland; ²Finnish Genome Center, University of Helsinki, Helsinki, Finland; ³HIIT Basic Research Unit, Department of Computer Science, University of Helsinki, Helsinki, Finland; ⁴Department of Surgery, University of Helsinki, Helsinki, Finland; ⁵Department of Gastroenterology, University of Helsinki, Helsinki, Finland

Epidemiological and genetic linkage studies have indicated a strong genetic basis for development of inflammatory bowel disease (IBD) which was recently supported by discovery of the Crohn's disease (CD) susceptibility gene termed *NOD2/CARD15*. We carried out a genome-wide linkage study in Finnish IBD families, providing a particular advantage to map susceptibility genes for ulcerative colitis (UC) within a genetic isolate. Initially, 92 IBD families with 138 affected sib-pairs (ASPs), were genotyped for 429 markers spaced at approximately 10 cM intervals. Next, the loci on chromosomes 2p13-11, 11p12-q13, and 12p13-12 were high-density mapped in the extended family cohort of 130 families with 173 ASPs. In this study, the most significant lod scores were observed for the UC families on chromosome 2p11 (D2S2333), in the vicinity of the *REG* gene cluster which is strikingly overexpressed in the IBD mucosa. The maximum two-point lod score was 3.34 (dominant model), and the corresponding NPL score 2.61. For UC, the second highest two-point NPL score of 2.00 was observed at proximal 12p13, where also some evidence for linkage disequilibrium emerged ($P=0.07$ and $P=0.007$ for the basic and extended IBD cohorts, respectively). The highest two-point NPL score for the CD families was 2.34 at D12S78 (12q23) with significant evidence for linkage disequilibrium ($P=0.004$), and for the mixed (MX) families 2.07 at D4S406 near the linkage peak reported previously. This study confirmed several of the IBD loci that have previously been reported and gives evidence for new IBD loci on chromosomes 2p11, 11p12-q13, 12p13-12, 12q23, and 19q13.

European Journal of Human Genetics (2002) 11, 112–120. doi:10.1038/sj.ejhg.5200936

Keywords: inflammatory bowel disease; genome-wide scan; genetic linkage

Ulcerative colitis (UC) and Crohn's disease (CD) are chronic idiopathic, inflammatory disorders of the gastrointestinal tract, and are thought to arise from the interplay of genetic and environmental factors. Many features suggest a significant familial component to pathogenesis of these chronic

inflammatory disorders. There is an increased risk among the first-degree relatives of IBD patients, with 5–10% of the patients having an affected first degree-relative, and there is a higher concordance rate for monozygotic than dizygotic twins for the same disorder.¹ The concordance rate is two to three times higher in CD than in UC, suggesting that CD would be determined to a higher extent by genetic factors than UC. The assumption is, however, that CD and UC would also share at least some of the susceptibility genes, since many of the genetic loci are clearly contributed by both disorders and there are families segre-

*Correspondence: Dr. Kimmo Kontula, Professor of Medicine, Department of Medicine, University of Helsinki, Haartmaninkatu 4, PL 340, 00290 Helsinki, Finland. Tel: 358-9-4717 2230; Fax: 358-9-4717 4013; E-mail: Kimmo.Kontula@hus.fi

Received 9 August 2002; revised 4 November 2002; accepted 15 November 2002

gating both UC and CD as well as indeterminate colitis. About 10% of IBD patients have indeterminate colitis in which differential diagnosis between CD and UC cannot be established.

Despite seven published genome scans and multiple single locus linkage reports on inflammatory bowel disease (IBD), so far only one gene, termed *NOD2/CARD15*, has been identified to confer susceptibility for CD.^{2,3} In addition, a risk haplotype for CD has been reported within the cytokine gene cluster on chromosome 5q.⁴ The linkage findings for IBD have been the most consistent for the chromosomal loci on 16cen (IBD1, MIM 266600), 12q (IBD2, MIM 601458), 14q11-12 (IBD4), and 6p (IBD3, MIM 604519).⁵⁻¹⁵ In addition, linkage to chromosome 1p has been reported both in the outbred North American population and the isolated Chaldean population.^{8,16} For the loci on chromosomes 5q31-33,¹⁴ 7q22,⁶ and 19p13,¹⁴ the lod scores have exceeded or approached the genome-wide significance level in a single study only. For pure UC, the linkage results have been less significant, and also less consistent between different populations. One problem is genetic heterogeneity, the effect of which is multiplied in attempts to find genes of low or modest effect using relatively small study samples. One of the most significant findings has been that UC seems to make a considerable contribution to the *IBD2* locus.^{17,18}

The Finnish population represents a genetically isolated founder population, which may offer some advantages especially in association mapping in the case of rare susceptibility alleles.^{19,20} Availability of a particularly large sample of UC families comparable in size to UC cohorts of large multinational studies¹⁰ prompted us to carry out a genome-wide scan on the Finnish multiplex IBD families. Here, we report data from our genome-wide scan which confirms several of the loci that have previously been reported and gives evidence for new IBD loci on chromosomes 2p11, 11p12-q13, 12p13-12, 12q23, and 19q13.

Materials and methods

Families

IBD patients having one or more IBD-affected first-degree relatives were recruited by gastroenterologists at all the university centres and regional central hospitals in Finland. The diagnoses of CD, UC and indeterminate colitis were established according to standard endoscopic and histological criteria.²¹ For assessment of the location of IBD, we defined the largest extent of the disease using radiographic and endoscopic data as well as surgical reports. Patients with proctitis only were not accepted as IBD patients. This procedure and the clinical features of familial IBD in Finland have been described in detail previously.^{22,23} We accepted 92 IBD families (designated as basic cohort), containing a total of 138 affected sib-pairs (ASPs), into a genome-wide linkage study (Table 1). In five of the mixed (MX) families, in which sibships with both UC and CD

were present, one of the siblings had indeterminate colitis. For high-density mapping on certain chromosomal regions, a total of 38 new IBD families were recruited to complement the basic cohort, yielding a total collection of 130 families, comprising 173 ASPs (designated as extended cohort, Table 1). Informed written consent was obtained from all study participants, and the study protocol was accepted by the Ethical Review Committee of Helsinki University Central Hospital.

Genotyping

Genomic DNA was extracted from peripheral blood leukocytes using standard procedures. Both parents were available for genotyping in 45 of the 130 families, while 25 families had one parent available. The genome-wide scan, with an average intermarker spacing of 10 cM, was performed by the use of the Applied Biosystems Linkage Mapping Set MD10. Altogether 444 loci were amplified with fluorescently labelled primers in separate polymerase chain reactions (PCRs) which were set up with Tecan Genesis 150 pipetting robot, and electrophoresed on MegaBACE 1000 96-well capillary instrument (Amersham Biosciences). Further details of the genotyping system and the analysis of the results have been described elsewhere.²⁴

A genome-wide scan was initially performed in the basic cohort of 92 families. Subsequently, the regions of interest on chromosomes 2p13-11, 11p12-q13, and 12p13-12, where the multipoint NPL score for IBD (basic cohort) extended over 1.5 on a broader area encompassing at least two adjacent markers, were further investigated with fifteen additional microsatellite markers in the extended family cohort (Table 2). In addition, for chromosomes 3 and 16 we had previous genotype data from a different marker set for the basic family cohort.²² This data was integrated with the genome scan data to achieve denser microsatellite spacing on chromosomes 3 and 16 (Table 2).

Statistical analysis

Two-point and multipoint nonparametric linkage analysis of the data from the genome-wide scan was performed using affected relative pairs as implemented in GENEHUNTER 2.1 r2 beta software package.²⁵ The allele frequencies in the marker loci were estimated from the total study sample. For comparison, we used parametric linkage analyses using an affecteds-only strategy with the dominant and recessive mode of inheritance (data not shown, except for some peak markers). These analyses were performed using the MLINK program of the LINKAGE package,²⁶ FASTLINK version 2.2.²⁷ Affected sib-pair analysis (ASP) was not primarily used as many of our families were extended, but to complement the NPL statistics the ASP analyses were performed using the SIBPAIR program²⁸ of the ANALYZE package.²⁹ Linkage disequilibrium was tested with a haplotype relative risk test as incorporated in the ANALYZE package. This package also contains a transmission disequilibrium test (TDTLIKE)

Table 1 Structure of the basic and extended family cohorts

Sibship size	Basic cohort				Extended cohort			
	UC	CD	MX	Total	UC	CD	MX	Total
2	36 (36)	15 (15)	12 (12)	63 (63)	54 (54)	17 (17)	22 (22)	93 (93)
2+ARP ^a	6 (6)	1 (1)	5 (5)	12 (12)	8 (8)	2 (2)	7 (7)	17 (17)
3	5 (15)	1 (3)	3 (9)	9 (27)	8 (24)	1 (3)	3 (9)	12 (36)
3+ARP	3 (9)	–	–	3 (9)	–	–	–	–
4	1 (6)	–	2 (12)	3 (18)	1 (6)	–	2 (12)	3 (18)
2+2 ^b	–	–	1 (4)	1 (4)	–	–	1 (4)	1 (4)
2+3 ^b	–	–	1 (5)	1 (5)	–	–	1 (5)	1 (5)
ARPs only	–	–	–	–	3	–	–	3
All	51 (72)	17 (19)	24 (47)	92 (138)	74 (92)	20 (22)	36 (59)	130 (173)

Note. The number of families are indicated and the number of sib-pairs are given in parenthesis. Mixed (MX) refers to sibships with both UC and CD. For simplicity, the number of ARPs are not presented in this table since most of the linkage information is obtained from the sib-pair data. ^aARP denotes for affected relative pair. ^bThe number of siblings in successive generations.

Table 2 Phases of the linkage analysis on the genome-wide and chromosome-specific data

Linkage analysis	Family cohort	Chromosomes (region in cM)	Average spacing of microsatellites	Number of markers
Phase 1 (The genome-wide scan)	Basic cohort	All chromosomes	~ 10 cM	(429)
Phase 2 (High-density mapping on the basic cohort)	Basic cohort	2p13-11 (84.2–113.3)	~ 3–4 cM	6 (9)
		3p23-21 (46.1–75.6) ^a		8 (11)
		11p12-q13 (43.0–71.7)		3 (8)
		12p13-12 (0.0–40.8)		6 (11)
		16p12-q12 (32.6–68.6) ^a		6 (10)
Phase 3 (High-density mapping on the extended cohort)	Extended cohort	2p13-11 (84.2–113.3)	~ 3–4 cM	6 (9)
		11p12-q13 (43.0–71.7)		3 (8)
		12p13-12 (0.0–40.8)		6 (11)

Note. The genotype data analysed in phase 2 and 3 originates from the same markers. The numbers describe the additional markers which were genotyped to achieve higher marker density. The total number of markers genotyped for each region are shown in parentheses. ^aPrevious genotype data for the basic cohort integrated with the genome scan data.

which for our data (complex pedigrees) can be used as a test for linkage.

The significance thresholds for the observed NPL statistics were estimated by computer simulations. To accomplish this, we created 1000 artificial data sets with the original pedigree structure and phenotype information. The founder alleles were drawn from the allele frequency distribution estimated from the observed data, after which the inheritance of alleles in the pedigrees was simulated under the null hypothesis of no disease gene (simple allele dropping). Next, missing data were introduced to the same markers and individuals as in real data, after which the artificial data sets were analysed like the real data set.

The data analysis was performed in three phases (Phase 1–3, Table 2). Phase 1 comprises the analyses performed on the whole genome scan data obtained from the basic cohort (92 families). Phase 2 analysis was performed on the basic cohort using genotype data from all available markers, including the markers used in the high-density mapping. Phase 3 analysis was performed on the selected chromosomal regions which were mapped to a higher density in the extended cohort of 130 families.

The data analysis was performed for all families combined (IBD) as well as for different family categories (pure UC, pure CD, and MX families). Families were also stratified according to the age at diagnosis in order to diminish genetic heterogeneity. The cutoff value of 24 years of age or less was selected since it represented the mean for the distribution of the youngest age at diagnosis in different families, and the resulting subgroups were still of reasonable size. In addition, we stratified the CD families according to their *CARD15/NOD2* status. Since in our case-control study (unpublished data) we observed a significant association between the 1007fs variant^{2,3} only and CD, the individuals positive for the 1007fs variant were coded unknown for their phenotypes in the linkage analysis. There were only four *NOD2*-positive CD families ($n=20$) which contained 14 individuals with the 1007fs variant.

Results

Linkage data for IBD

The genome-wide scan performed on the basic cohort (phase 1) revealed four chromosomal regions where the

multipoint NPL score for IBD was >1.5 over a broader area encompassing at least two adjacent markers. These chromosomes were 3p, 11, 12, and 16 (Figure 1). In addition to these loci, multipoint NPL scores over 1.5 were observed for chromosomes 2, 19, and 20 (Figure 1). The highest two-point NPL score observed for the basic IBD cohort was 2.15 at D12S352 (0.0 cM) followed by a value of 2.0 for D11S1314 (71.7 cM). When stratifying the data according to the age at diagnosis ≤ 24 years, the linkage was enhanced on chromosomes 2p11 with the peak multipoint NPL score of 2.34, 11p12-q13 (2.31), 12p13 (2.19), 19p13 (2.07) and 19q13 (2.66). The highest two-point NPL scores for the analyses of phase 2 (basic cohort with denser marker map) and phase 3 (extended cohort with denser marker map) are summarised in Table 3.

Linkage data for the UC families

The most significant genome-wide two-point NPL score for the UC families was 2.61 at D2S2333 (2p11, 100.4 cM, Table 3). In the affecteds-only analysis with a dominant mode of inheritance, the two-point lod score reached 3.34 at recombination fraction (θ) 0.00. For chromosome 2, two other peaks with nominal evidence for linkage were observed (Figure 1). After high-density mapping on the extended family cohort (phase 3) for the locus on 2p11, the linkage of the UC families to this region diminished (Table 3). Two-point NPL score for D2S2333 was still 1.71, but the multipoint NPL declined to 0.50 due to very low NPL-values observed with the adjacent markers which located at 2.0 and 4.0 cM distance from D2S2333. The affecteds-only analysis with a dominant mode of inheritance gave a two-point lod score of 1.20 ($\theta=0.14$) for the extended UC cohort.

In phase 1 analysis for UC, the second highest two-point NPL score of 2.00 (multipoint 1.82) was observed at proximal 12p13 (at D12S352, 0.0 cM) where also the highest two-point NPL score (2.15) for IBD was obtained (Figure 1, Table 3). In phase 2 analysis, the maximum multipoint NPL score for UC was 2.04 at 0.0 cM (Figure 1, Table 3). When selecting families with UC ≤ 24 years ($n=25$, age at diagnosis ≤ 24 years), the NPL scores reached up to 2.72 (two-point) and 3.30 (multipoint, Table 3). The inclusion of 23 new UC families in the extended cohort (phase 3) diminished the NPL scores at D12S352 (Table 3). Two-point NPL score for UC diminished from 2.00 to 0.87 and for UC with earlier onset from 2.72 to 1.64.

Some evidence for linkage disequilibrium was discovered on chromosome 12p13-12 (Table 3). The *P*-value obtained for D12S99 (13.8 cM) using the haplotype relative risk test was 0.07 for the basic cohort, 0.007 for the extended cohort, and 0.06 for all UC families. The test revealed that two adjacent alleles with this marker (2 bp difference in allele size) were preferentially transmitted to the affected offspring. The frequency of the more common allele was 0.32 in the transmitted chromosomes vs 0.22 in the control

chromosomes, and the corresponding frequencies for the +2 bp allele were 0.05 vs 0.01, respectively. A trace of allelic association was also observed with the adjacent marker GATA49D12 at 19.8 cM where the *P*-value was 0.03 for the extended cohort.

In phase 1 analysis, nominal multipoint NPL scores were also observed on chromosomes 13, 15, and 19 (Figure 1). The highest genome-wide multipoint NPL score for the UC families was observed on distal chromosome 19q13 (93.0 cM) where the multipoint NPL score was 2.08 at the most distal marker D19S210 (Figure 1, Table 3). The corresponding two-point NPL score was 1.88. For the families with UC ≤ 24 years, the NPL scores were as high as 2.35 (two-point) and 2.92 (multipoint, Table 3).

Linkage data for the CD families

The highest genome-wide two-point NPL score for the CD subgroup was 2.34 that was observed at D12S78 (12q23) with significant evidence for linkage disequilibrium ($P=0.004$, Figure 1, Table 3). The corresponding multipoint NPL score was 2.36. The transmission disequilibrium test for the CD families gave a *P*-value of 0.00005 providing additional evidence for linkage to this marker. After removing the NOD2/1007fs variant carriers from the analysis, the two-point and multipoint NPL scores were 2.51 and 2.54, respectively (data not shown).

Nominal multipoint NPL scores were observed for chromosomes 7 (peak NPL score; 1.76 at 81.2 cM), 10 (1.52 at 137.1 cM), and 16 (2.11 at 60.6 cM, Figure 1). For chromosome 16, the integration of previous genotype data from a different marker set for the basic family cohort (phase 2) did not significantly enhance the evidence for linkage observed in phase 1. The multipoint NPL score for the CD families peaked at 1.91 for D16S415 (61.6 cM, Figure 1). The highest two-point NPL score of 1.53 was observed on the p-telomeric region at D16S407 (Table 3). When selecting families with CD ≤ 24 years ($n=10$), the multipoint NPL score rose to 2.53 at D16S3068 (41.6 cM). Two other peaks of same magnitude were observed at 7.3 cM (multipoint NPL score of 2.41) and 61.1 cM (multipoint NPL score of 1.95, Table 3). For the *CARD15/NOD2*-stratified set of CD families the maximum multipoint NPL score of 1.57 was observed at 60.6 cM. Another peak with a multipoint NPL score of 1.45 was observed at 32.6 cM.

For chromosome 11, the highest multipoint NPL score obtained in phase 1 analysis was 1.48 at D11S901 (78.6 cM). The haplotype relative risk test gave a *P*-value of 0.01 for D11S901 and D11S4175 (90.5 cM) providing some evidence for linkage disequilibrium. In phase 3 analyses, the NPL scores improved in the CD group in which the peak multipoint NPL score reached 1.78 at D11S4136 (69.0 cM, Table 3). For this marker, the mean identical-by-descent sharing for a sib-pair was 0.72. The *NOD2*-stratification did not significantly alter the results concerning chromosome 11.

Figure 1 Multipoint nonparametric linkage analysis results for the IBD genome scan. The analyses were performed for all families combined (IBD) as well as for different family categories (pure UC, pure CD, and MX families). The vertical axis indicates the NPL score along the length (in cM) of each chromosome. The horizontal bars and arrows indicate the position of the previously reported IBD-susceptibility loci and the *CARD15/NOD2* gene.

Table 3 Summary of two-point linkage analysis in phase 2 and phase 3

Chr	Marker	cM	Phase 2 (basic cohort, n=92 IBD families) Two-point NPL score/Peak multipoint NPL score					Phase 3 (extended cohort, n=130 IBD families) Two-point NPL score/Peak multipoint NPL score					
			IBD	UC	CD	MX		IBD	UC	CD	MX		
							≤24 years All, n=47 CD, n=10 UC, n=25 MX, n=12					≤24 years All, n=65 CD, n=12 UC, n=34 MX, n=19	
2p11	D2S2333	100.4	1.81/ 1.34	2.61/ 1.74	−0.62	0.26	2.15/2.34 (All)	0.98	1.71/ 0.50	−0.89	0.05	1.21/1.19 (All)	
3p21	D3S2432	51.1	1.75/ 1.92	1.47/ 1.42	0.61	0.77	1.90/1.38 (All)						
11p12-q13	D11S987	64.3	1.57/ 2.02	0.34	1.44	1.35	1.43/2.65 (MX) 58 cM	1.32/ 1.31	0.29	1.36	1.10/ 1.05	1.14/2.19 (MX) 46 cM	
	D11S4136	69.0	1.40	0.96	1.11	0.41/ 2.00		0.78	0.15	1.57/ 1.78	0.12		
	D11S1314	71.7	2.00	0.89	1.34	1.46		1.30	0.54	1.44	0.64		
	D11S901	78.6	1.16	0.14	1.27/1.45; *0.01	1.00							
12p13-12	D12S352	0.0	2.15/ 1.56	2.00/ 2.04	0.26	1.07	2.72/3.30 (UC)	1.29/ 1.19	0.87/ 1.40	0.20	1.06	1.64/2.45 (UC)	
	D12S99	13.8	−0.32; *0.07	0.09	−1.48	0.48		0.08; *0.007	0.57	−1.72	0.62		
	GATA49D12	19.8	1.28	1.18/ 1.48	−0.69	1.36	2.20/2.43 (MX) 22 cM	1.02; *0.03	0.79	−0.61	1.26	2.02/2.34 (MX)	
	D12S1697	26.3	1.79/ 1.55	1.41	−0.33	1.74		1.87/ 1.53	1.65/ 1.30	−0.66	1.66	1.81 (MX)	
12q23	D12S78	118.5	1.47	1.18	2.34/2.36; *0.004	−0.82							
12q24	D12S1659	162.8	1.85/ 1.89	1.31/ 1.20	0.14	1.59/ 1.80							
16p13-q12	D16S407	7.3	1.38	0.22	1.53	1.09	1.85/2.41 (CD)						
	D16S3075	14.6	0.28/ 1.84	−0.28	0.25	0.74							
	D16S3068	41.6	0.83	0.43	0.14	0.88	1.40/2.53 (CD)						
19p13	D16S415	61.1	0.55	−0.09	1.26/1.91	0.14							
	D19S216	5.9	0.43	−0.63	0.31	1.49	1.57 (MX)						
	D19S884	11.9	0.75	−0.002	0.07	2.00 1.42/ 2.05	1.92/2.54 (MX)						
19q13	D19S210	93.0	1.73/ 2.02	1.88/ 2.08	0.38	0.33	2.35/2.92 (UC)						

Note. Age-stratified linkage results for specific loci are shown only if significant change in the NPL scores was observed. The family category in which the evidence for linkage was strongest is shown in parentheses. Significant P-values obtained using the haplotype relative risk test are indicated by asterisks. Multipoint peak location is shown if it is deviating from the marker that is presented in the table. As chromosome 19 was not high-density mapped, the NPL scores represent results from phase 1 analysis. Chr indicates the chromosome and cM the marker distance from p-telomere.

Linkage data for the MX families

The highest genome-wide two-point NPL score for the MX families was 2.07 at D4S406. The highest genome-wide multipoint NPL score was 2.15 that was observed on chromosome 22 (Figure 1). The second highest multipoint NPL score of 2.14 was observed on chromosome 2, at 0.0 cM (Figure 1).

Multipoint NPL scores over 2.0 were also observed for chromosomes 10 and 19p13 (Figure 1). On chromosome 19p13, analysis of the families with MX ≤24 years (*n*=12 families) gave a maximum multipoint NPL score of 2.54 at D19S884 (11.9 cM, Table 3). For chromosome 11, the maximum multipoint NPL score was 1.69 at D11S987

(64.3 cM). In the MX families with earlier onset, the maximum multipoint NPL score reached 2.65 (basic cohort) and 2.19 (extended cohort) at 58.0 cM and 46.1 cM, respectively (Table 3).

Simulation experiments

The genome-wide significance of the observed NPL statistics was estimated by simulation. The medians for the highest NPL scores occurring anywhere in the genome by chance only turned out to be 2.25 for the CD families, 2.40 for the UC families, 2.48 for the MX families, and 2.42 for all IBD families. The thresholds for the NPL score occurring with probability of 5% by chance only, were 2.91, 3.11,

3.29, and 3.04 for the CD, UC, MX, and all IBD families, respectively. These values correspond to thresholds of 'suggestive' and 'significant' linkage in a genome-wide scan as defined by Lander and Kruglyak.³⁰ In our study, the NPL scores exceeded the threshold of suggestive linkage on two loci. On chromosome 12q23 at D12S78 the two-point and multipoint NPL scores for the CD families were 2.34 and 2.36, respectively, and on chromosome 2p11 at D2S2333, the two-point NPL score for the UC families was 2.61.

Discussion

There are at least two specific aspects that make the present genome-wide linkage study particularly important. First, we studied the largest UC cohort from one population reported so far, extending the total number of affected UC sib-pairs up to 92. For comparison, in the large international European UC cohort the total number of ASPs was 114.¹⁰ Second, we provide genetic data on the genetically homogeneous Finnish population which may offer some advantages for association mapping of complex disease genes, especially in the case of rare susceptibility alleles.^{19,20}

Our study provides evidence for five potential novel linkages to chromosomes 2, 11p12-q13, 12p13-12, 12q23, and 19q13. On chromosome 2, there were three positive regions showing nominal linkage for IBD. The midmost peak on 2p11 was contributed by the UC families of the basic cohort especially, producing a suggestive two-point NPL score of 2.61 and a significant maximum two-point lod score of 3.34 in the analysis using the dominant mode of inheritance. The linkage, however, diminished upon the high-density mapping on the extended UC cohort. The linked region corresponds almost exactly to the localisation of three genes coding for regenerating islet-derived 1A (*REG1A*), regenerating islet-derived 1B (*REG1B*), and pancreatitis-associated proteins (*PAP*), that have been reported to be highly overexpressed in the diseased mucosa of both the CD and UC patients.^{31,32}

On chromosome 11p12-q13 we observed a broad region of positive NPL scores reaching up to 2.02 that was contributed by all the IBD subgroups. The transcription factor gene encoding the *p65* subunit of nuclear factor kappa-B (*NF-κB p65*) and the TNF receptor-associated factor 6 have previously been assigned to this region. On chromosome 12, for all IBD families both a p-telomeric and a q-telomeric region of nominal linkage were observed. For the CD families, the NPL scores were negative on the most proximal p-telomeric region (12p13-12), where nominal to suggestive NPL scores were obtained for the UC and MX families. On this region (12p13-12), stratification of the UC and MX families by the age at diagnosis (≤ 24 years) strengthened significantly the linkage signal observed. In addition, linkage disequilibrium was observed with two adjacent markers D12S99 and GATA49D12. Allelic association at D12S99 was contributed by two different alleles with a size difference of 2 bp. It is very unlikely that such

association happens just by chance. The infrequent allele may have originated as a result of a historical allele expansion in the more common allele representing the founder chromosome carried by the IBD patients. The NPL scores for the CD group grew positive towards the IBD2 region on 12q. For the CD families the strongest linkage was, however, observed more telomeric to the IBD2 region peaking at D12S78 (12q23) with significant allelic association ($P=0.004$). On chromosome 19q13, nominal NPL scores were observed for the UC subgroup especially, and the linkage was significantly enhanced in the subgroup of UC families with disease onset ≤ 24 years.

This genome-wide linkage study confirmed several of the loci that have previously been reported for IBD. The place and extent of the chromosome 3p21 linkage remained the same as in our previous report.²² Our nominal linkage of IBD and UC to chromosome 3q (multipoint peak at 167.5 cM) was somewhat more centromeric compared to the linkage peak reported by Cho *et al.*⁸ For chromosome 12, our positive NPL scores were localised to 12p13-12 and 12q23-24, and mostly flanked the IBD2 locus, the maximum linkage peaks of which have previously been assigned to 12q11-q21.¹⁵ Only for our CD families, nominal NPL scores were observed on the proximal region of IBD2, at D12S1617 and D12S345.

Our data concerning the IBD1 locus on chromosome 16 were different compared to the earlier reports¹⁵ since our MX families seemed to contribute to the linkage. In the CD families, we observed a broad region of nominal linkage. The maximum two-point NPL score peaked at D16S407 near the p-telomere, corresponding the maximum linkage peak reported by Satsangi *et al.*⁶ The multipoint NPL peaked near the *NOD2/CARD15* gene at D16S415. The linkage signal strengthened in the CD families with disease onset ≤ 24 years with a maximum two-point NPL score of 1.85 observed at D16S407 (7.3 cM) and a maximum multipoint NPL score of 2.53 observed at D16S3068 (41.6 cM). The latter marker is the same for which Hampe *et al.* reported significant allelic association.³³ Collectively, these data imply that there could be an additional CD susceptibility gene on chromosome 16.

Upon subgroup (CD/UC/MX families) analyses, there was confirmatory evidence for four additional IBD loci, comprising chromosomes 4q (MX only), 7q22 (especially CD), 19p13, and 22 (MX only). The strongest linkage to chromosome 4q localised very close to the linkage peak reported by Cho *et al.*⁸ who observed it only in the MX subset of families. The chromosome 7 peak was assigned near to the peak linkage region reported by Satsangi *et al.*⁶ and the linkage of IBD to chromosome 22 has previously been described by Hampe *et al.*¹⁰ The linkage of our IBD families to chromosome 19p13 was the most prominent of these four confirmatory observations, and it was especially significant in the subgroup of the MX families with disease onset ≤ 24 years. The linkage peak localised onto the maximum

region of linkage reported by Rioux *et al.*¹⁴ Even though this result must be considered with caution since multiple comparisons were performed, it has been shown that stratification by age can unmask genetic susceptibility to complex disease. Accordingly, patients carrying two mutant alleles of the *CARD15/NOD2* gene were shown to be characterised by younger age at onset in addition to certain phenotypic features.³⁴ Furthermore, Brant *et al* reported a stronger linkage of early-onset families to the IBD1 region.³⁵

In conclusion, this genome-wide scan on Finnish IBD families revealed several potential novel IBD loci. Although only one of the loci (2p11) showed significant lod score under the dominant inheritance model, there were still some other loci of particular note. Some evidence of linkage disequilibrium was observed on chromosomes 11p12-q13, 12p13-12, and 12q23. In addition, enhanced linkage for the families with earlier onset of the disease was shown for the loci on chromosomes 2p11, 11p13-12, 12p13-12, and 19q13, thus providing additional evidence that these loci could harbor IBD susceptibility genes.

Acknowledgements

We thank members of Finnish IBD families participating in the study. We are obliged to our collaborators Prof Heikki Järvinen (Helsinki University Hospital), Dr Heimo Nurmi (Turku University Hospital), Docent Anna-Liisa Karvonen (Tampere University Hospital), Docent Risto Julkunen and Dr Markku Heikkinen (Kuopio University Hospital), and Docent Seppo Niemelä (Oulu University Hospital). Sirpa Stick and Tiula Soppela-Loponen are acknowledged for excellent laboratory assistance, and Paula Leinonen and Elina Salminen for assisting in collection of patients and hospital records. This study was supported by a grant from the Special State Funds of The Helsinki University Central Hospital.

References

- Binder V: Genetic epidemiology in inflammatory bowel disease. *Dig Dis* 1998; **16**: 351–355.
- 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.
- 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.
- 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.
- 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.
- 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.
- Cavanaugh JA, Callen DF, Wilson SR *et al*: Analysis of Australian Crohn's disease pedigrees refines the localization for susceptibility to inflammatory bowel disease on chromosome 16. *Ann Hum Genet* 1998; **62**: 291–298.
- 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.
- Duerr RH, Barmada MM, Zhang L *et al*: Linkage and association between inflammatory bowel disease and a locus on chromosome 12. *Am J Hum Genet* 1998; **63**: 95–100.
- 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.
- Hampe J, Shaw SH, Saiz R *et al*: Linkage of inflammatory bowel disease to human chromosome 6p. *Am J Hum Genet* 1999; **65**: 1647–1655.
- 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 disease shows confirmed linkage to chromosome 14q11-12. *Am J Hum Genet* 2000; **66**: 1857–1862.
- 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.
- Cavanaugh J: International collaboration provides convincing linkage replication in complex disease through analysis of a large pooled data set: Crohn disease and chromosome 16. *Am J Hum Genet* 2001; **68**: 1165–1171.
- 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.
- Parkes M, Barmada MM, Satsangi J, Weeks DE, Jewell DP, Duerr RH: The IBD2 locus shows linkage heterogeneity between ulcerative colitis and Crohn disease. *Am J Hum Genet* 2000; **67**: 1605–1610.
- Parkes M, Satsangi J, Jewell DP, Weeks DE, Barmada MM, Duerr RH: Ulcerative colitis is more strongly linked to chromosome 12 than Crohn's disease. *Gut* 2001; **49**: 311.
- Eaves IA, Merriman TR, Barber RA *et al*: The genetically isolated populations of Finland and Sardinia may not be a panacea for linkage disequilibrium mapping of common disease genes. *Nat Genet* 2000; **25**: 320–323.
- Kere J: Human population genetics: lessons from Finland. *Annu Rev Genomics Hum Genet* 2001; **2**: 103–128.
- Lennard-Jones JE: Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl* 1989; **170**: 2–6.
- Paavola P, Helio T, Kiuru M *et al*: Genetic analysis in Finnish families with inflammatory bowel disease supports linkage to chromosome 3p21. *Eur J Hum Genet* 2001; **9**: 328–334.
- Halme L, Turunen U, Helio T *et al*: Familial and sporadic inflammatory bowel disease: comparison of clinical features and serological markers in a genetically homogeneous population. *Scand J Gastroenterol* 2002; **37**: 692–698.
- Paunu N, Lahermo P, Onkamo P *et al*: A novel low-penetrance locus for familial glioma at 15q23-q26.3. *Cancer Res* 2002; **62**: 3798–3802.
- 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.
- Lathrop GM, Lalouel JM, Julier C, Ott J: Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 1984; **81**: 3443–3446.
- Cottingham Jr RW, Idury RM, Schaffer AA: Faster sequential genetic linkage computations. *Am J Hum Genet* 1993; **53**: 252–263.
- Kuokkanen S, Sundvall M, Terwilliger JD *et al*: A putative vulnerability locus to multiple sclerosis maps to 5p14-p12 in a region syntenic to the murine locus Eae2. *Nat Genet* 1996; **13**: 477–480.
- Terwilliger JD: A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am J Hum Genet* 1995; **56**: 777–787.

- 30 Lander ES, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; **11**: 241–247.
- 31 Dieckgraefe BK, Stenson WF, Korzenik JR, Swanson PE, Harrington CA: Analysis of mucosal gene expression in inflammatory bowel disease by parallel oligonucleotide arrays. *Physiol Genomics* 2000; **4**: 1–11.
- 32 Lawrance IC, Fiocchi C, Chakravarti S: Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes. *Hum Mol Genet* 2001; **10**: 445–456.
- 33 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.
- 34 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.
- 35 Brant SR, Panhuysen CI, Bailey-Wilson JE *et al*: Linkage heterogeneity for the IBD1 locus in Crohn's disease pedigrees by disease onset and severity. *Gastroenterology* 2000; **119**: 1483–1490.