Human Molecular Genetics, 2004, Vol. 13, No. 7 ***763–770***
DOI: 10.1093/hmg/ddh090
Advance Access published on February 19, 2004

Inflammatory bowel disease susceptibility loci defined by genome scan meta-analysis of 1952 affected relative pairs

David A. van Heel[1,*], Sheila A. Fisher[2], Andrew Kirby[3], Mark J. Daly[3], John D. Rioux[3], Cathryn M. Lewis[2] and the Genome Scan Meta-Analysis Group of the IBD International Genetics Consortium[†]

[1]Department of Gastroenterology, Imperial College London W12 0NN, UK, [2]Division of Medical and Molecular Genetics, Guy's King's and St Thomas' School of Medicine, London SE1 9RT, UK and [3]Whitehead Institute/MIT Center for Genome Research, Cambridge, MA 02139-1561, USA

Received December 17, 2003; Revised and Accepted February 5, 2004

Crohn's disease and ulcerative colitis (the inflammatory bowel diseases) have a strong genetic component. Although over 20 putative susceptibility loci have been identified by individual genome scans, the majority of these loci have not been replicated. Many individual studies are at the lower limit of acceptable power for complex disease linkage analysis. Genome scan meta-analysis (GSMA), by use of sample sizes an order of magnitude greater than individual linkage studies, has increased power to detect novel loci, may confirm or refute regions detected in smaller individual studies, and enables regions to be prioritized for further gene identification efforts. Genome scan data (markers, significance scores) were obtained from 10 separate studies and meta-analysis was performed using the GSMA method. These studies comprised 1952 inflammatory bowel disease, 1068 Crohn's disease and 457 ulcerative colitis affected relative pairs. Study results were divided into 34 cM chromosomal bins, ranked, weighted by study size, summed across studies and bin-by-bin significance obtained by simulation. A region on chromosome 6p (containing the HLA) met genome wide significance for inflammatory bowel disease. Loci meeting suggestive significance for inflammatory bowel disease were 2q, 3q, 5q, 7q and 16 (NOD2/CARD15 region); Crohn's disease, 2q, 3q, 6p, 16 (NOD2/CARD15 region), 17q, 19p; and ulcerative colitis, 2q. Clustering of adjacent bins was observed for chromosomes 6p, 16, 19p. The meta-analysis has identified novel loci and prioritized genomic regions for further gene identification studies.

INTRODUCTION

Crohn's disease (CD) and ulcerative colitis (UC)—the inflammatory bowel diseases (IBD)—are chronic debilitating conditions of the gastrointestinal tract, with combined prevalence of ~0.2% in Caucasians. The two conditions share some clinical and pathological features, including some treatment options, although disease type can usually be distinguished by features including anatomical site and histology. Epidemiological studies have revealed a significant genetic contribution to their pathogenesis, with sibling relative risks of 30–40 for CD and 10–20 for UC (1). Greater than expected occurrence of both CD and UC within the same family suggests that some inherited variants might predispose to both types of intestinal inflammation.

There have been 10 published inflammatory bowel disease genome scans, in addition to several as yet unpublished scans and locus-specific replication studies (recently reviewed in 2). More than 20 susceptibility loci have been identified from these individual studies, and the majority of these have not been reproduced across studies. This finding is unsurprising as statistical considerations suggest that much larger cohorts are needed for reliable detection of linkage (3) and even larger studies required for replication. One notable success has been the confirmation of the chromosome 16 (*IBD1*) locus by the pooling and analysis of locus specific raw genotype data from

*To whom correspondence should be addressed at: Department of Gastroenterology, Imperial College, Hammersmith Campus, Du Cane Road, London W12 0NN, UK. Tel: +44 2083838067; Fax: +44 2087493436; Email: d.vanheel@imperial.ac.uk
[†]The above named authors take primary responsibility for data analysis and interpretation, a full author list is found in the Appendix.

12 centres (4), and Crohn's disease-associated variants in a gene from this locus (*NOD2*/*CARD15*) have subsequently been identified (5,6).

A pooled analysis of IBD studies at a genome wide level offers the potential to identify novel susceptibility loci of relatively weak effect, confirm regions suspected from smaller individual studies and prioritize regions for further gene identification. A genome-wide pooled analysis might be performed using original genotypes, or by meta-analysis of genome scan results. Use of original genotypes followed by new linkage analysis is complicated by lack of original data availability, industry relationships, difficulties of constructing a combined map and use of different marker sets. We used the Genome Scan Meta-Analysis (GSMA) approach to analyse IBD data (7). This approach has the advantage of straightforward data submission, no assumptions of inheritance models and is robust to different analysis methods. On the other hand, this method reduces the precision of linkage signal (34 cM bins were used in this study). Simulation methods, recently applied to bipolar disorder and schizophrenia genome scan data, were used to assess statistical significance of results (8–10).

We contacted all investigators known to be undertaking linkage analysis in IBD, in part through an IBD International Genetics Consortium, and performed GSMA on data from 10 independent studies. This dataset comprises much the largest study in IBD to date, and one of the largest analysed in any complex disease. The meta-analysis has detected novel regions, confirmed some previously identified loci, and prioritized regions for further gene identification studies.

RESULTS

Weighted genome scan meta-analyses for the three phenotypes

Figure 1A shows bin-by-bin *Psumrnk* for the IBD phenotype. Genome-wide significance was observed for bin 6.2 (containing the HLA region) and suggestive significance for the adjacent bin 6.1. Three other bins (2.6, 3.7, 16.2) met suggestive significance. Four out of five of the bins meeting suggestive significance also met the empirical criteria suggesting these bins were likely to contain linked loci. An additional two bins (5.3, 7.3) also met the empirical criteria but did not quite reach suggestive significance. Table 1 shows the 20 highest bins for each of the three phenotypes. Full data for all bins and phenotypes are available online in Supplementary Information 1.

Figure 1B shows bin-by-bin *Psumrnk* for the CD phenotype. No bins met genome-wide significance and five bins met suggestive significance. Interestingly these bins only represented three distinct loci on chromosomes 16 (bin 16.2), 19 (bins 19.1, 19.2) and 6 (bins 6.1, 6.2) with in each case an adjacent bin also meeting one or more significance criteria. A further four bins (2.6, 3.7, 12.3, 17.3) met the empirical criteria suggesting these bins were likely to contain linked loci.

Figure 1C shows bin-by-bin *Psumrnk* for the UC phenotype. No genome-wide significant results were observed for the UC phenotype. Suggestive significance was seen for bin 2.6 although the empirical criteria were not also met, and one bin out of 102 having a *P*-value of less than 0.01 can be expected by chance.

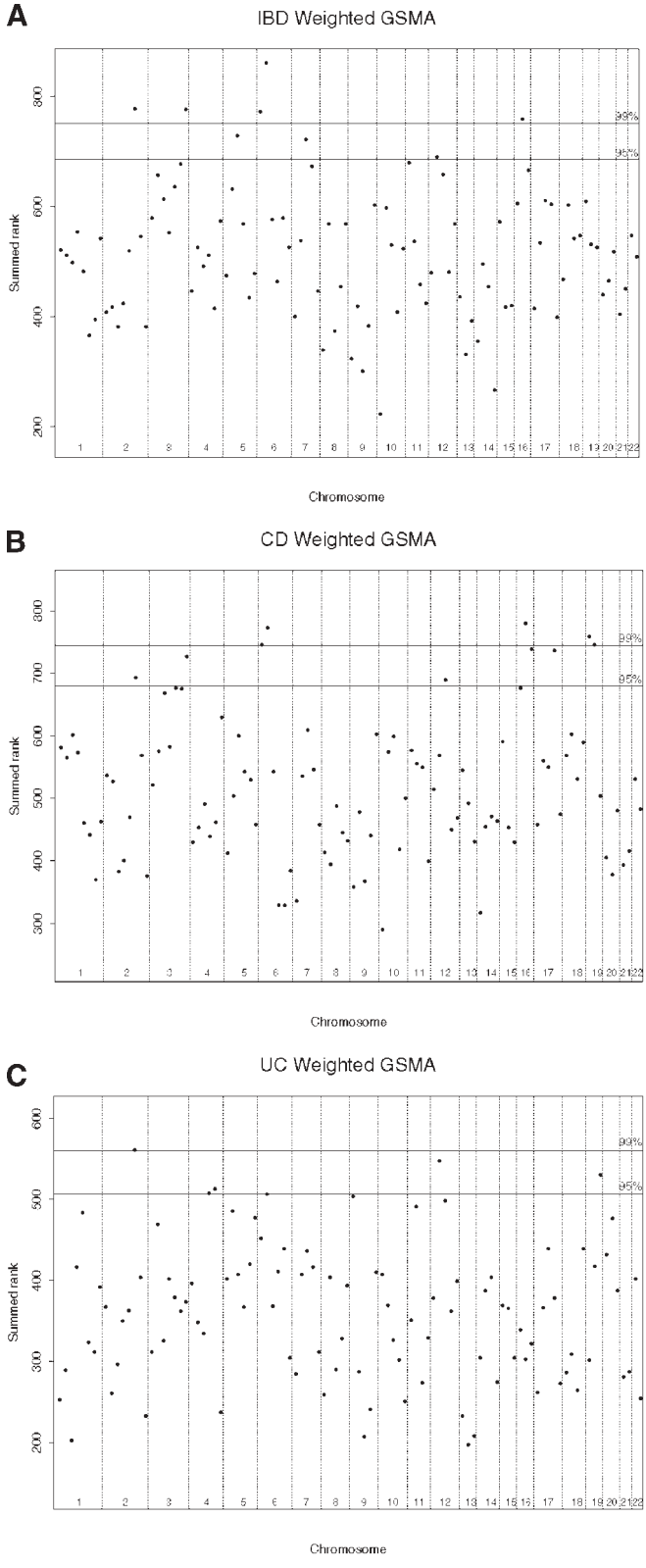
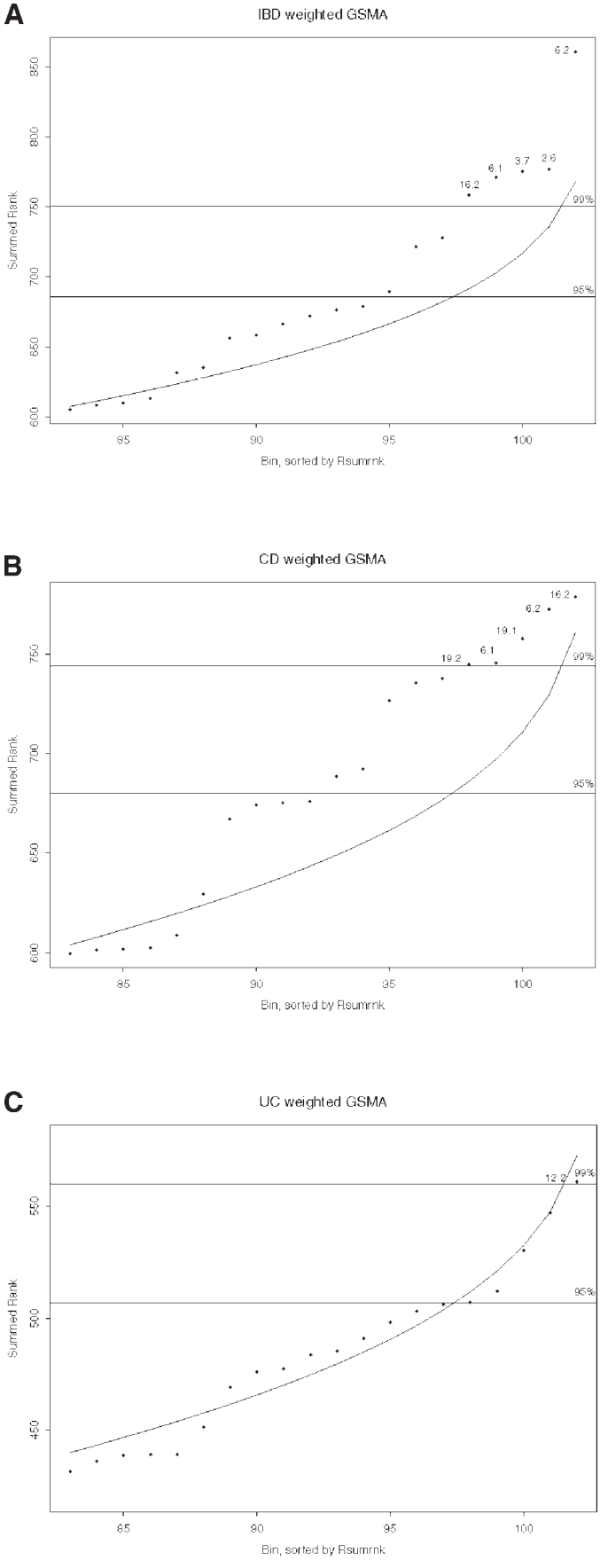


Figure 1. Genome scan meta-analyses results for all chromosomes. (**A**) Inflammatory bowel disease phenotype. (**B**) Crohn's disease phenotype. (**C**) Ulcerative colitis phenotype. Individual chromosomes are sub-divided into ~34 cM bins (represented by a dot), and bins ranked by significance after summing weighted data across studies. Significance levels corresponding to 99% and 95% shown (*Psumrnk* < 0.01 and 0.05, respectively).

Table 1. *Psumrnk* and *Pord* for IBD, CD and UC bins with the twenty highest summed ranks

Phenotype	Bin	Bin location (cM)[a]		Cytogenetic location	Weighted analyses			
		Begin	End		Rank	*Psumrnk*	*Pord*	Both *Psumrnk* and *Pord* < 0.05
IBD	6.2	34	69	6p23–6p21.1	860.9	0.00012[b]	0.01194	+
IBD	2.6	171	206	2q24–2q34	777.5	0.00434[c]	0.06924	
IBD	3.7	201	228	3q27–3qter	775.8	0.00460[c]	0.01012	+
IBD	6.1	0	34	6pter–6p23	771.7	0.00525[c]	0.00126	+
IBD	16.2	34	69	16p13.1–16q12.2	758.7	0.00791[c]	0.00054	+
IBD	5.3	68	103	5q13–5q15	728.3	0.01855	0.00682	+
IBD	7.3	70	104	7q11.1–7q21	722.0	0.02183	0.00308	+
IBD	12.2	35	70	12p12–12q13	690.1	0.04571	0.06824	
IBD	11.1	0	39	11pter–11p11.2	679.4	0.05705	0.09422	
IBD	3.6	167	201	3q23–3q27	676.7	0.06026	0.05580	
IBD	7.4	104	138	7q21–7q31	672.7	0.06520	0.03878	
IBD	16.3	69	101	16q12.2–16q22	666.7	0.07309	0.03714	
IBD	12.3	70	105	12q13–12q21	658.8	0.08467	0.05016	
IBD	3.2	32	65	3p25–3p21	656.7	0.08793	0.02792	
IBD	3.5	134	167	3q13.3–3q23	635.8	0.12525	0.24926	
IBD	5.2	33	68	5p13–5q13	632.0	0.13300	0.22792	
IBD	3.3	65	99	3p21–3p11	613.6	0.17488	0.68336	
IBD	17.2	32	67	17p11.2–17q21	610.6	0.18258	0.65296	
IBD	19.1	0	34	19qter–19p13.2	608.9	0.18687	0.57642	
IBD	16.1	0	34	16pter–16p13.1	605.6	0.19533	0.55772	
CD	16.2	34	69	16p13.1–16q12.2	778.9	0.00323[c]	0.28504	
CD	6.2	34	69	6p23–6p21.1	772.7	0.00400[c]	0.05952	
CD	19.1	0	34	19pter–19p13.2	757.7	0.00655[c]	0.02594	+
CD	6.1	0	34	6pter–6p23	745.8	0.00951[c]	0.01246	+
CD	19.2	34	68	19p13.2–19q13.2	745.0	0.00972[c]	0.00190	+
CD	16.3	69	101	16q12.2–16q22	738.0	0.01193	0.00054	+
CD	17.3	67	100	17q21–17q24	735.7	0.01272	0.00008	+
CD	3.7	201	228	3q27–3qter	726.8	0.01631	0.00010	+
CD	2.6	171	206	2q24–2q34	692.4	0.03824	0.00742	+
CD	12.3	70	105	12q13–12q21	688.7	0.04161	0.00352	+
CD	16.1	0	34	16pter–16p13.1	676.0	0.05459	0.00926	
CD	3.5	134	167	3q13.3–3q23	675.3	0.05541	0.00288	
CD	3.6	167	201	3q23–3q27	674.5	0.05633	0.00090	
CD	3.3	65	99	3p21–3p11	667.3	0.06525	0.00122	
CD	4.6	176	212	4q32–4qter	629.7	0.12895	0.29826	
CD	7.3	70	104	7q11.1–7q21	608.9	0.17786	0.82942	
CD	9.5	136	169	9q33–9qter	602.5	0.19471	0.88260	
CD	18.2	31	64	18p11.1–18q12	602.2	0.19570	0.80812	
CD	1.3	73	108	1p32–1p22	601.6	0.19717	0.71350	
CD	5.3	68	103	5q13–5q15	599.7	0.20248	0.65622	
UC	2.6	171	206	2q24–2q34	561.4	0.00950[c]	0.63912	
UC	12.2	35	70	12p12–12q13	547.6	0.01555	0.48160	
UC	19.3	68	105	19q13.2–19qter	530.5	0.02658	0.52588	
UC	4.5	141	176	4q28–4q32	512.5	0.04357	0.68670	
UC	4.4	107	141	4q23–4q28	507.5	0.04942	0.59536	
UC	6.2	34	69	6p23–6p21.1	506.5	0.05065	0.41334	
UC	9.1	0	34	9pter–9p13	503.4	0.05460	0.30658	
UC	12.3	70	105	12q13–12q21	498.7	0.06110	0.26106	
UC	11.2	39	74	11p11.2–11q13	491.2	0.07260	0.29882	
UC	5.2	33	68	5p13–5q13	485.7	0.08185	0.29916	
UC	1.5	142	176	1p13–1q21	484.1	0.08473	0.20610	
UC	5.6	172	198	5q33–5qter	477.8	0.09669	0.24454	
UC	20.2	34	67	20p11.2–20q13.1	476.3	0.09959	0.16540	
UC	3.2	32	65	3p25–3p21	469.4	0.11428	0.22728	
UC	6.1	0	34	6pter–6p23	451.6	0.15780	0.72092	
UC	17.2	32	67	17p11.2–17q21	439.4	0.19302	0.93390	
UC	6.5	138	170	6q22–6q25	439.3	0.19343	0.87752	
UC	18.4	99	126	18q21–18qter	439.1	0.19398	0.79376	
UC	7.3	70	104	7q11.1–7q21	436.5	0.20192	0.77246	
UC	20.1	0	34	20pter–20p11.2	431.9	0.21650	0.81800	

[a]cM position from Marshfield sex-averaged genetic map.
[b]Bins meeting *Psumrnk* genome wide significance.
[c]Bins meeting *Psumrnk* suggestive significance.



The observed versus expected distributions of *Rsumrnk* for the highest scoring bins for the three phenotypes are shown in Figure (2A–C). The *Pord* statistic reflects, for a given bin position, the significance associated with the difference between observed and expected *Rsumrnk.* The observed distributions for IBD and CD clearly show deviation from the expectation under the null hypothesis, whereas the UC phenotype does not differ.

Clustering of adjacent bins was observed for bins on chromosomes 3, 6 and 16 for the IBD phenotype and chromosomes 3, 6, 16 and 19 for the CD phenotype (albeit for both phenotypes the adjacent chromosome 3 bins do not quite reach significance).

Weighted versus unweighted, adjacent bin pooling and bin omission analyses

Table 2 shows a comparison of weighted (by study size) versus unweighted results for bins that meet one or more significance criteria in the weighted analyses. Combining the data from all phenotypes, slightly more bins showed higher significance in the weighted analyses than unweighted analyses, two bins significant in the weighted analyses were not significant in the unweighted data, and no additional significant bins were observed in the unweighted analyses. Full data for all bins and phenotypes are available online in Supplementary Information 1.

In the analysis of adjacent pairs of bins, performed in case linkage evidence was split across bin boundaries, no new regions were identified. All significant paired bins contained a region suggested by single-bin analysis (data not shown).

The weighted GSMA for the IBD and CD phenotypes was repeated with the omission of bins 6.1, 6.2 (corresponding to *HLA*/*IBD3*) and 16.2, 16.3 (corresponding to *CARD15*/*IBD1*). All bins previously showing suggestive significance ($Psumrnk < 0.01$) were again observed—in each case with marginal increases in significance (data not shown). For the CD phenotype, two bins which had previously only met empirical criteria (bins 17.3, 3.7) now additionally reached suggestive significance ($Psumrnk = 0.0076$ and 0.0099, respectively).

DISCUSSION

Genome-wide evidence for linkage was obtained in this meta-analysis of 10 inflammatory bowel disease genome scans. Bin 6.2 (chromosome 6p23–6p21.1, containing the HLA region) was associated with a genome-wide significant *P*-value for the inflammatory bowel disease phenotype after correction for multiple testing. Suggestive linkage (expected to occur once

Figure 2. Observed versus expected distribution of weighted *Rsumrnk* for the top 20 bins. (**A**) Inflammatory bowel disease phenotype. (**B**) Crohn's disease phenotype. (**C**) Ulcerative colitis phenotype. Individual chromosomes are sub-divided into ∼34 cM bins, and bins ranked by significance after summing weighted data across studies (top 20 bins, ordered by significance, represented by a dot). Bin locations are labelled where empirical criteria, or genome wide/suggestive criteria for linkage are met. Significance levels corresponding to 99 and 95% are shown ($Psumrnk < 0.01$ and 0.05, respectively). The solid line represents the expected distribution of the top 20 bins under the null hypothesis of no linked loci.

per meta-analysis by chance) was observed for four other bins for the IBD phenotype, five bins for the CD disease phenotype and once for the UC phenotype. An additional two (IBD) and five (CD) bins met empirical criteria for linkage (both *Psumrnk* and *Pord* < 0.05) established in studies of simulated complex disease datasets of similar size to the current study. There is no straightforward way to determine which of these latter bins represents a true positive result. A noteworthy observation from the studies of simulated complex disease datasets was that clusters of adjacent bins meeting significance criteria are unlikely to be false positives. Clustering of bins was observed on chromosomes 6 (6pter–6p21.1) for both IBD and CD, 16 (16p13.1–16q22) and 19 (19qter–19q13.2) for CD.

Individual genetic variants showing association with disease have now been reported for chromosome 6 (the HLA region) and for chromosome 16 (the CARD15/NOD2 gene). Convincing evidence for linkage of CD to chromosome 16 was reported in a multicentre analysis of genotypes from six microsatellite markers spanning the pericentromeric region. More recently, disease-associated variants within CARD15 have now been widely replicated for CD. This locus therefore provides a positive control for the GSMA, and some indication of the significance that might be observed elsewhere in the genome for a locus with similar allele frequencies and effect size. It is noteworthy that the bins for this region did not quite reach genome wide significance, and it is possible that the Bonferroni correction applied is over-conservative. Multiple genetic variants within the HLA region have been reported to be associated with IBD, CD and UC over the last two decades. This region contains over 100 genes of known immunological and other functions, and genetic association analyses are complicated by long-range linkage disequilibrium. Both extended haplotypes and individual variants have been reported to be associated with disease in family based studies, case–control studies and meta-analyses. Thus, although the precise disease-causing variant/haplotype is not yet clear, there is little doubt that genetic variation within this region predisposes to IBD and sub-phenotypes, providing a further positive control for the GSMA. It is important that linkage scores in complex traits should not be penalized by the fact that other better linkage scores exist somewhere else in the genome, and we attempted to control for this possibility in the GSMA by dropping the four bins corresponding to the HLA region and CARD15 gene in further analyses. It might also be argued that the null hypothesis for CD was not one of 'no linked loci', and these analyses address in part this issue. Marginally increased significance levels were observed for all bins previously meeting significance criteria, and an additional two bins previously only meeting empirical criteria for the CD phenotype now met suggestive significance.

Previous simulation studies suggested that weighting the meta-analysis by study size led to a small increase in power. Results were similar for the four bins corresponding to the HLA region and CARD15 gene in both weighted and unweighted analyses. Overall, however, slightly more significant bins were observed in the weighted analysis and significance levels were slightly higher in the weighted analyses. We therefore present the weighted data in this paper, and both sets of analyses are available for comparison as Supplementary Information 1. We next tested whether the choice of bin size and bin position influenced the meta-analysis results, as a broad linkage peak might be split between two bins. Results from all combinations of paired adjacent bins were similar to the original weighted analysis.

We report suggestive evidence for linkage to regions on chromosomes 2q, 3q, 5q and 7q for the IBD phenotype, 2q, 3q, 12q, 17q and 19p for the CD phenotype and 2q for the UC phenotype in addition to the previously discussed chromosome 6 and 16 bins. The chromosome 5, 6 and 16 loci have also been confirmed in an earlier smaller meta-analysis of five IBD studies (11), although the power of this earlier study may have been reduced by the lack of availability of significance values and chromosomal locations for individual markers and failure to use of simulation methods for significance testing. The significance criteria used for reporting of linkage findings in individual IBD genome-wide studies vary considerably; at nominal values ($P < 0.05$ uncorrected) linkage has been reported to all but one chromosome. Findings at the suggestive significance level (i.e. Lander and Kruglyak criteria of one false positive per genome scan) which overlap with regions from the current meta-analysis for the IBD phenotype include the 3q region [LOD 2.3 (12)] and 7q region [single-point LOD 3.1 (13)]. Interestingly a recent paper has described association of variants in the MDR1 gene (located within the linked GSMA 7q bin at 7q21) with IBD (14). The region on chromosome 5q reported here for the IBD phenotype has been identified in two studies comprising mostly CD pairs, with LOD scores of 2.2 (15) and 2.4 (16). Suggestive linkage to chromosome 2q has been reported for the UC phenotype [LOD 2.2 (17)]. Chromosome 12q has been identified previously for CD (LOD score not reported) in an early single-point study (13) and chromosome 19p has been identified for CD [LOD 4.6 (16)]. Thus the loci for 2q (for IBD and CD phenotypes) and 17q (CD phenotype) are novel, and reported for the first time at the suggestive level from this meta-analysis. Greater confidence in linkage can now be assumed for the other loci observed in this meta-analysis, which have previously been reported in only one or two genome scans at the suggestive level.

The IBD phenotype comprises both the CD and UC phenotypes, but also over 400 mixed CD–UC pairs not included in the CD and UC analyses. We analysed the combined IBD phenotype, given the biological plausibility of a locus contributing generally to intestinal inflammation, the greater than expected occurrence of mixed CD–UC pairs in families, and the clinical observation that disease cannot be sub-classified (into CD or UC) in 10% of cases. All disease types contribute to the linkage signal observed on chromosome 2q and 6, whereas the much larger signal observed on 3q for IBD than CD might suggest a contribution from the mixed pairs. In contrast, the greater signal on chromosome 16 observed for CD than IBD suggests the majority of linkage signal is contributed by the CD phenotype. Interestingly, linkage to chromosome 5q is reported only for the IBD phenotype. A CD associated haplotype has been identified from the 5q region (18), although recent studies suggest this confers a weaker effect than initially reported (19–21). It is therefore possible that different 5q variants might additionally influence the IBD phenotype, and the 5q13–15 locus for IBD may be distinct. However simulation studies have shown that significant results from the GSMA are often obtained in bins that flank the bin

containing the susceptibility locus. An analysis of pooled genotypes would help clarify whether the GSMA 5q result is due to the IBD5 haplotype or a novel locus.

The current study cannot determine how much of the genetic load of IBD can be attributed to the identified loci, as the GSMA method does not measure effect size (in contrast to traditional meta-analysis). This calculation will also depend on assumptions including the nature of locus–locus interactions, which are difficult to test until the disease causing variants are known. It is likely, however, that the current study has identified the majority of IBD loci detectable through linkage analysis of realistically acheivable cohorts.

The two most promising loci for future gene identification efforts would appear to be chromosome 6 for IBD and CD, and chromosome 19 for CD. No one population/cohort would appear more amenable than others to gene identification efforts at these loci. The cohorts included in the current study were all of predominantly European Caucasian descent. Published data is available for three common Crohn's disease associated NOD2/CARD15 variants from most of these cohorts, and to a more limited extent for the IBD5 risk haplotype. The observation that NOD2/IBD5 allele frequencies are similar across the cohorts used in the current study suggests the populations might be broadly similar at other IBD loci. The only population amenable to specific gene identification studies in ethnic groups is the Ashkenazi Jewish population present in some North American samples. This group was too small to usefully include in the meta-analysis. We did not find evidence for heterogeneity in the linkage data between the cohorts from the current study (data not shown).

The meta-analysis has therefore provided strong evidence for both novel and previously suspected loci involved in IBD susceptibility, using the largest cohort in any complex disease linkage study to date. The success of gene identification strategies on chromosome 16 (NOD2/CARD15) in inflammatory bowel disease will encourage efforts to identify disease-causing mutations from the loci identified and prioritised by the meta-analysis.

MATERIALS AND METHODS

Selection of genome scans

All worldwide centres known to be involved in linkage studies of inflammatory bowel disease were invited to participate, and genome scans further identified through publications and oral presentations. All scans were performed using an even density of microsatellite markers across the genome. Partial scans and candidate region studies were excluded. All groups with published full-genome scans agreed to participate, and the meta-analysis therefore provides a comprehensive overview. Centres provided a computer file containing data on marker names and, for each phenotype, linkage statistics (as either LOD or NPL score) or *P*-values. All scans used multipoint non-parametric analysis methods. Family size, number of affected relative pairs (ARP) and methodology for individual studies are shown in Table 3. All centres ascertained families according to standard clinical, laboratory, endoscopic, histological and radiological criteria for the diagnosis of inflammatory bowel disease.

Table 2. Comparison of weighted and unweighted analyses

Phenotype	Bin	Weighted analyses		Unweighted analyses	
		Rsumrnk	*Psumrnk*	*Rsumrnk*	*Psumrnk*
IBD	6.2	860.9	0.00012	808.5	0.00047
IBD	2.6	777.5	0.00434	749.0	0.00504
IBD	3.7	775.8	0.00460	720.0	0.01265
IBD	6.1	771.7	0.00525	730.5	0.00918
IBD	16.2	758.7	0.00791	747.5	0.00531
IBD	5.3	728.3	0.01855	726.0	0.01057
IBD	7.3	722.0	0.02183	671.5	0.04547
CD	16.2	778.9	0.00323	789.5	0.00111
CD	6.2	772.7	0.00400	760.0	0.00345
CD	19.1	757.7	0.00655	717.0	0.01382
CD	6.1	745.8	0.00951	704.0	0.01997
CD	19.2	745.0	0.00972	649.0	0.07440
CD	16.3	738.0	0.01193	723.5	0.01140
CD	17.3	735.7	0.01272	738.5	0.00717
CD	3.7	726.8	0.01631	700.5	0.02196
CD	2.6	692.4	0.03824	625.5	0.11742
CD	12.3	688.7	0.04161	679.0	0.03805
UC	2.6	561.4	0.00950	558.5	0.00434

Genome scan meta-analysis

GSMA was performed as described (7). Briefly the autosomes (X and Y chromosomes were not studied) were divided into 102 bins of mean 34.2 cM size (range 23.2–47.8 cM) bins. Marker positions were defined using an integrated marker map developed by the Whitehead Institute (based on Marshfield 1999 map markers) (22). For each study, each bin was assigned a within-study rank (*Rstudy*) based on the maximum linkage score within the bin. Bins were ranked in descending order (102 the most significant result). Analyses were performed for the IBD, CD and UC phenotypes. We analysed the IBD phenotype, comprising both CD, UC and CD–UC affected pairs, as it is biologically plausible that a locus might predispose to both types of intestinal inflammation. Previous meta-analyses on simulated complex disease datasets had shown a small increase in power by weighting for individual study size (8). In the weighted analyses, for each phenotype, bins were weighted by multiplying each *Rstudy* value by the number of study ARP divided by the mean number of ARP in all studies. *Rsumrnk* was obtained by summing individual (weighted) *Rstudy* ranks across all studies. Weights used are shown in Table 4.

Two pointwise *P*-values were determined, *Psumrnk* and *Pord*, as described and determined by 50 000 permutations of the weighted dataset (8). *Psumrnk* is the probability of observing a bin's summed rank by chance; and *Pord* the probability of observing the *j*th place bin's summed rank in *j*th place bins in randomly permuted data. As an example of these statistics, if the fourth highest result bin had summed rank 772, the *Psumrnk* gives the probability that an arbitrary bin obtains this summed rank or higher, and *Pord* gives the probability that the fourth most significant result in a GSMA has summed rank 772 or higher. *Psumrnk* gives an analysis by bin, and we therefore expect 5% of bins to achieve a summed rank with *P*-value below 0.05. *Pord* analyses all bins concurrently, but *P*-values are not independent across bins: a significant *Pord* for the *n*th ranked bin increases the probability that the $n-1$th ranked bin has a significant *Pord*.

Table 3. Characteristics of inflammatory bowel disease genome scans

Study site	Reference	Program	Output used	Number of markers	IBD (CD + UC + M)		CD		UC	
Canada/Whitehead	(16)	MM/Sibs	LOD	304	158	190	106	121	23	25
Chicago/Baltimore	(12)	GH+	LOD	371	174	297	99	175	13	26
Finland	(26)	GH	NPL	382	92	152	24	35	56	98
France/Europe	(23)	T test	P value	270	25	52	25	52	—	—
London/Germany	(25)	MM/Sibs	LOD	343	268	353	129	162	90	114
Los Angeles	(15)	MM/Sibs	LOD	350	46	65	46	65	—	—
NovaScotia/Whitehead	(11)	GH	NPL	304	61	168	40	86	12	27
Oxford	(27)	MERLIN	LOD	387	228	288	112	137	77	95
Pittsburgh	(24)	GH+	LOD	588	62	127	62	127	—	—
Pittsburgh/Baltimore/Chicago	(17)	Allegro	NPL	358	139	260	68	108	43	72
Total					1253	1952	711	1068	314	457

IBD, inflammatory bowel disease; UC, ulcerative colitis; M, mixed CD/UC affected relative pairs; ARP, affected relative pairs; GH+/GH, Genehunter Plus/Genehunter; MM/Sibs, MapMaker Sibs; MERLIN, Multipoint Engine for Rapid Linkage Inference; LOD, logarithm of odds; NPL, non-parametric linkage.

Table 4. Weighting function applied to individual studies

Study site	IBD-weight	CD-weight	UC-weight
Canada/Whitehead	1.0	1.1	0.4
Chicago/Baltimore	1.5	1.7	0.4
Finland	0.8	0.3	1.5
France/Europe	0.3	0.5	—
London/Germany	1.8	1.5	1.7
Los Angeles	0.3	0.6	—
NovaScotia/Whitehead	0.9	0.8	0.4
Oxford	1.5	1.3	1.5
Pittsburgh	0.7	1.2	—
Pittsburgh/Baltimore/Chicago	1.3	1.0	1.1

In each meta-analysis we applied both theoretical and empirical criteria for assessing significance at a genome wide level. *Psumrnk* <0.0005 and *Psumrnk* <0.01 correspond to the widely used Lander and Kruglyak criteria for genome wide (expected by chance once per 20 meta-analyses) and suggestive significance (expected by chance once per meta-analysis), respectively, corrected by the Bonferroni method for 102 bins. No further correction was made for the testing of three phenotypes, as these phenotypes were not independent and the genome wide correction was already highly conservative (due to non-independence of linkage statistics in adjacent bins).

Previous genome scan meta-analyses in schizophrenia and bipolar disorder had established further empirical criteria for bins most likely to contain linked loci (those with both *Psumrnk* <0.05 and *Pord* <0.05) (8–10). The number of studies and study size for the three phenotypes in the current analysis falls between the larger schizophrenia dataset and the smaller bipolar disorder-narrow dataset. The conclusions from these studies are therefore likely to be generally valid for the current analyses, and were included in the assessment of significance.

Broad peaks are observed in non parametric linkage analyses in complex disease, which might reduce the power of GSMA by splitting linkage evidence between two bins where the locus lies close to a bin boundary. Therefore an additional analysis was performed where for each study, the most significant linkage statistic/*P*-value within each possible pair of adjacent bins within a chromosome were re-ranked (ranks 1–80). *Rsumrnk*, weighted by study size, was obtained as previously.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

This study was supported by grants from the Wellcome Trust (to C.M.L.; Clinician Scientist Fellowship to D.A.vH.), the Charitable Foundation of Guy's and St Thomas' Hospitals (to C.M.L.) and the NIH (DK64869 to J.D.R.). The Wellcome Trust Centre for Human Genetics, University of Oxford, kindly provided access to computing facilities. Meetings of the IBD International Genetics Consortium were kindly supported by the Crohn's and Colitis Foundation of America.

REFERENCES

1. van Heel, D.A., Satsangi, J., Carey, A.H. and Jewell, D.P. (2000) Inflammatory bowel disease: Progress toward a gene. *Can. J. Gastroenterol.*, **14**, 207–218.
2. Mathew, C.G. and Lewis, C.M. (2004) Genetics of inflammatory bowel disease: progress and prospects. *Hum. Mol. Genet.*, **13**, R161–R168.
3. Risch, N. and Merikangas, K. (1996) The future of genetic studies of human complex diseases. *Science*, **273**, 1516–1517.
4. The International IBD Genetics Consortium (2001) 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.*, **68**, 1165–1171.
5. Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cezard, J.P., Belaiche, J., Almer, S., Tysk, C., O'Morain, C.A., Gassull, M. *et al.* (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature*, **411**, 599–603.
6. Ogura, Y., Bonen, D.K., Inohara, N., Nicolae, D.L., Chen, F.F., Ramos, R., Britton, H., Moran, T., Karaliuskas, R., Duerr, R.H. *et al.* (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature*, **411**, 603–606.

7. Wise, L., Lanchbury, J. and Lewis, C. (1999) Meta-analysis of genome scans. *Ann. Hum. Genet.*, **63**, 263–272.
8. Levinson, D.F., Levinson, M.D., Segurado, R. and Lewis, C.M. (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder. Part I: methods and power analysis. *Am. J. Hum. Genet.*, **73**, 17–33.
9. Lewis, C.M., Levinson, D.F., Wise, L.H., DeLisi, L.E., Straub, R.E., Hovatta, I., Williams, N.M., Schwab, S.G., Pulver, A.E., Faraone, S.V. *et al.* (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder. Part II: schizophrenia. *Am. J. Hum. Genet.*, **73**, 34–48.
10. Segurado, R., Detera-Wadleigh, S.D., Levinson, D.F., Lewis, C.M., Gill, M., Nurnberger, J.I., Jr., Craddock, N., DePaulo, J.R., Baron, M., Gershon, E.S. *et al.* (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder. Part III: bipolar disorder. *Am. J. Hum. Genet.*, **73**, 49–62.
11. Williams, C.N., Kocher, K., Lander, E.S., Daly, M.J. and Rioux, J.D. (2002) Using a genome-wide scan and meta-analysis to identify a novel IBD locus and confirm previously identified IBD loci. *Inflamm. Bowel Dis.*, **8**, 375–381.
12. Cho, J.H., Nicolae, D.L., Gold, L.H., Fields, C.T., Labuda, M.C., Rohal, P.M., Pickles, M.R., Hanauer, S.B., Bayless, T.M. and Brandt, S.R. (1998) 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*, 95, 7502–7507.
13. Satsangi, J., Parkes, M., Louis, E., Hashimoto, L., Kato, N., Welsh, K.I., Bell, J.I. and Jewell, D.P. (1996) Two-stage genome-wide search in inflammatory bowel disease: evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat. Genet.*, **14**, 199–202.
14. Brant, S.R., Panhuysen, C.I., Nicolae, D., Reddy, D.M., Bonen, D.K., Karaliukas, R., Zhang, L., Swanson, E., Datta, L.W., Moran, T. *et al.* (2003) MDR1 Ala893 polymorphism is associated with inflammatory bowel disease. *Am. J. Hum. Genet.*, **73**, 1282–1292.
15. Ma, Y., Ohmen, J.D., Li, Z., Bentley, L.G., McElree, C., Pressman, S., Targan, S.R., Fischel-Ghodsian, N., Rotter, J.I. and Yang, H. (1999) A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm. Bowel Dis.*, **5**, 271–278.
16. Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhart, A.H., McLeod, R.S., Griffiths, A.M., Green, T., Brettin, T.S., Stone, V., Bull, S.B. *et al.* (2000) Genomewide search in canadian families with inflammatory bowel disease reveals two novel susceptibility Loci. *Am. J. Hum. Genet.*, **66**, 1863–1870.
17. Barmada, M.M., Brant, S.R., Nicolae, D.L., Achkar, J.-P., Panhuysen, C.I., Bayless, T.M., Cho, J.H. and Duerr, R.H. (2004) A genome scan in 260 inflammatory bowel disease-affected relative pairs. *Inflamm. Bowel Dis.*, **10**, 15–22.
18. Rioux, J.D., Daly, M.J., Silverberg, M.S., Lindblad, K., Steinhart, H., Cohen, Z., Delmonte, T., Kocher, K., Miller, K., Guschwan, S. *et al.* (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat. Genet.*, **29**, 223–228.
19. Negoro, K., McGovern, D.P., Kinouchi, Y., Takahashi, S., Lench, N., Shimosegawa, T., Carey, A., Cardon, L., Jewell, D. and van Heel, D.A. (2003) Analysis of the IBD5 locus and potential gene-gene interactions in Crohn's disease. *Gut*, **52**, 541–546.
20. Giallourakis, C., Stoll, M., Miller, K., Hampe, J., Lander, E.S., Daly, M.J., Schreiber, S. and Rioux, J.D. (2003) IBD5 is a general risk factor for inflammatory bowel disease: replication of association with Crohn disease and identification of a novel association with ulcerative colitis. *Am. J. Hum. Genet.*, **73**, 205–211.
21. Armuzzi, A., Ahmad, T., Ling, K.L., de Silva, A., Cullen, S., van Heel, D., Orchard, T.R., Welsh, K.I., Marshall, S.E. and Jewell, D.P. (2003) Genotype–phenotype analysis of the Crohn's disease susceptibility haplotype on chromosome 5q31. *Gut*, **52**, 1133–1139.
22. Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L. and Weber, J.L. (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.*, **63**, 861–869.
23. Paavola-Sakki, P., Ollikainen, V., Helio, T., Halme, L., Turunen, U., Lahermo, P., Lappalainen, M., Farkkila, M. and Kontula, K. (2003) Genome-wide search in Finnish families with inflammatory bowel disease provides evidence for novel susceptibility loci. *Eur. J. Hum. Genet.*, **11**, 112–120.
24. Hugot, J.P., Laurent-Puig, P., Gower-Rousseau, C., Olson, J.M., Lee, J.C., Beaugerie, L., Naom, I., Dupas, J.L., Van Gossum, A., Orholm, M. *et al.* (1996) Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature*, **379**, 821–823.
25. Hampe, J., Schreiber, S., Shaw, S.H., Lau, K.F., Bridger, S., Macpherson, A.J., Cardon, L.R., Sakul, H., Harris, T.J., Buckler, A. *et al.* (1999) A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am. J. Hum. Genet.*, **64**, 808–816.
26. van Heel, D., Dechairo, B., Dawson, G., McGovern, D., Negoro, K., Carey, A., Card1on, L., Mackay, I., Jewell, D. and Lench, N. (2003) The *IBD6* Crohn's disease locus demonstrates complex interactions with *CARD15* and *IBD5* disease associated variants. *Hum. Mol. Genet.*, **12**, 2569–2575.
27. Duerr, R.H., Barmada, M.M., Zhang, L., Pfutzer, R. and Weeks, D.E. (2000) High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11–12. *Am. J. Hum. Genet.*, **66**, 1857–1862.

APPENDIX

The Genome Scan Meta-Analysis Group of the International Inflammatory Bowel Disease Genetics Consortium

Canadian/Whitehead Scan (15): J.D. Rioux, M.S. Silverberg, M.J. Daly, A.H. Steinhart, R.S. McLeod, A.M. Griffiths, T. Green, T.S. Brettin, V. Stone, S.B. Bull, A. Bitton, C.N. Williams, G.R. Greenberg, Z. Cohen, E.S. Lander, T.J. Hudson, K.A. Siminovitch.

Chicago/Baltimore Scan (11): J.H. Cho, D.L. Nicolae, R. Karaliukas, E. Swanson, T.M. Bayless, S.R. Brant.

Finland Scan (22): P. Paavola-Sakki, V. Ollikainen, T. Helio, L. Halme, U. Turunen, P. Lahermo, M. Lappalainen, M. Farkkila, K. Kontula.

France/Europe Scan (23): J.P. Hugot, P. Laurent Puig, C. Gower Rousseau, J.M. Olson, J.C. Lee, L. Beaugerie, I. Naom, J.L. Dupas, A. van Gossum, GETAID, M. Orholm, C. Bonaiti-Pellie, J. Weissenbach, C.G. Mathew, J.E. Lennard-Jones, A. Cortot, J.F. Colombel, G. Thomas.

London/Germany Scan (23): J. Hampe, C.G. Mathew, S. Schreiber, M.E. Curran.

Los Angeles Scan (14): Y. Ma, J.D. Ohmen, Z. Li, L.G. Bentley, C. McElree, S. Pressman, S.R. Targan, N. Fischel-Ghodsian, J.I. Rotter, H. Yang.

NovaScotia/Whitehead Scan (10): C.N. Williams, K. Kocher, K. Miller, L. Farwell, E.S. Lander, M.J. Daly, J.D. Rioux.

Oxford Scan (25): D.A. van Heel, B.M. Dechairo, G. Dawson, D.P.B. McGovern, K. Negoro, A.H. Carey, L.R. Cardon, I. MacKay, D.P. Jewell, N.J. Lench.

Pittsburgh Scan (26): R.H. Duerr, M.M. Barmada, L. Zhang, R. Pfutzer, D.E. Weeks.

Pittsburgh/Baltimore/Chicago Scan (16): M.M. Barmada, S.R. Brant, D.L. Nicolae, J.P. Achkar, C. Panhuysen, T.M. Bayless, J.H. Cho, R.H. Duerr