ABCB1/MDR1 gene determines susceptibility and phenotype in ulcerative colitis: discrimination of critical variants using a gene-wide haplotype tagging approach

G.-T. Ho1,*, N. Soranzo3, E.R. Nimmo1, A. Tenesa2, D.B. Goldstein3 and J. Satsangi1

1Molecular Medicine Unit and Gastrointestinal Unit, University of Edinburgh and 2MRC Human Genetics Unit, Western General Hospital, Edinburgh, UK and 3Goldstein Laboratories, Darwin Building, University College London, London, UK

Received October 1, 2005; Revised and Accepted January 18, 2006

Several lines of evidence suggest a role for the multidrug resistance gene (ABCB1/MDR1) and its product, P-glycoprotein 170, in the pathogenesis of inflammatory bowel disease (IBD). In addition, P-glycoprotein activity determines bioavailability of many drugs used regularly in many medical specialities, and ABCB/MDR1 variation appears to be a critical pharmacogenetic determinant. We have utilized a gene-wide haplotype tagging approach to further define the identity of germ-line variations in the ABCB1/MDR1 gene contributing to IBD susceptibility. Six haplotype tagging single nucleotide polymorphisms (tSNPs) representing the haplotypic variations of the ABCB1/MDR1 gene were identified initially following the characterization of the haplotype structure of this gene in 24 Centre d’Etude du Polymorphisme Humain Caucasian trios. Genotyping was performed in 249 ulcerative colitis (UC) and 179 Crohn’s disease (CD) patients and 260 healthy controls. Using log-likelihood analysis, we observed a highly significant association between the common haplotypes and UC (P = 4.22 × 10⁻⁷) but not CD (P = 0.22). This significant association was critically dependent on one tSNP, intronic variant rs3789243. All haplotypes with this variant retained a highly significant association (P = 3.2 × 10⁻⁷–3.6 × 10⁻¹²), whereas significance was lost when rs3789243 was dropped in systematic haplotypic analysis. The effect of this tSNP was independent of C3435T SNP, previously suggested to be the critical variant in disease susceptibility and drug transport. The association with UC was shown to be strongest with the phenotype of extensive disease (P = 1.7 × 10⁻⁷). This ‘candidate gene’ approach provides compelling evidence to support the contribution of the ABCB1/MDR1 gene in determining risk to UC but not to CD and provides new insights into the localization of the critical susceptibility determinants within the gene. In addition, these findings have potentially important implications in the application of pharmacogenetics across a range of common diseases, including HIV, epilepsy and colorectal cancer.

INTRODUCTION

Crohn’s disease (CD) (MIM 266600) and ulcerative colitis (UC) (MIM 191390) are common chronic inflammatory disorders of the gastrointestinal tract characterized by a dysregulated mucosal immune response (1). Several lines of evidence, including twin concordance, family studies and ethnic aggregation of inflammatory bowel disease (IBD), clearly demonstrated the significant genetic contribution to the pathogenesis of IBD (2). Genetic linkage analyses, through genome-wide screens, have identified a number of susceptibility loci, underlying the complexity of this contribution (3). Complementary techniques of positional cloning and candidate gene approach have led to the established finding of the

*To whom correspondence should be addressed at: Gastrointestinal Unit, Western General Hospital, Edinburgh EH4 2XU, UK. Tel: +44 131 537 1769; Fax: +44 131 537 1007; Email: gwotzerho@aol.com

© The Author 2006. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org
NOD2/CARD15 gene as a susceptibility gene for CD (4,5). Most recently, further progress has been forthcoming as evident in recent studies which suggest a possible role for further genes in determining susceptibility to CD, the DLG5 (discs, large homologue gene), OCTN 1 and 2 (organic cation transporter) and NOD1 (nucleotide oligomerization domain 1) genes in some but not all ethnic groups (6–8).

The role of the multidrug resistance 1 (MDR1) gene and its product, the P-glycoprotein 170, as a potential determinant of susceptibility of IBD has been a subject of considerable recent interest (9). P-glycoprotein 170 functions as an ATP-dependent efflux transporter pump, which is highly expressed in the epithelial surfaces of intestine, biliary ductules, proximal tubules of kidneys and central nervous system where it forms the basis of the blood–brain barrier (10–12). Inter-individual variability in P-glycoprotein expression in the gastrointestinal tract plays a role in determining the pharmacokinetics of a wide-ranging number of substrates. Although the exact physiological role in the gut remains unknown, the high constitutive levels of expression of P-glycoprotein 170 in the gut suggest a role in protection against xenobiotics, including bacterial products. This hypothesis is supported by the phenotype of mdrl−/− mice models, which develop enterocolitis with histopathology resembling human UC, in specific pathogen-free but not germ-free conditions (13). Bone marrow transfer studies involving this model indicate that the deficiency of P-glycoprotein 170 in the epithelial rather than the lymphoid cells may be responsible for the pathology observed. Most recently, compelling data from Langmann et al. (14) have demonstrated that the ABCB1/MDR1 gene expression is significantly reduced in the colonic tissue of patients with UC but not CD.

Specific allelic variations of the MDR1 gene have been shown to be associated with different levels of P-glycoprotein 170 expression. Two exonic variants, namely the C3435T single nucleotide polymorphism (SNP) in exon 26 and a missense G2677T/A (Ala893Ser/Thr) SNP in exon 21, have received most attention. The C3435T SNP was first shown to correlate with expression of P-glycoprotein (15). In this study, the TT genotype was associated with lowered intestinal P-glycoprotein expression (CC > CT > TT, \( P = 0.056 \) and 0.053, respectively). Further pharmacokinetic studies gave support to the hypothesis that the C3435T SNP may be involved in the regulation of the expression of the P-glycoprotein (16–18). The G2677T/A SNP in exon 21 results in two distinct amino acid changes, namely 893Ser (G2677T) and the much rarer 893Thr (G2677T). These two variant alleles have also been suggested to be associated with altered P-glycoprotein 170 expression (17,19).

Although these two SNPs have been shown to be associated with expression of P-glycoprotein 170, the mechanisms whereby these variants may affect gene transcription have yet to be clarified and doubts have arisen as to the functional effects of these variants. The C3435T SNP is a silent ‘wobble’ mutation. Site-specific mutagenesis experiments have demonstrated that the C3435T and G2677T/A substitutions have no effect on P-glycoprotein function in vivo (20). Haplotypes derived from these two SNPs showed a stronger association with P-glycoprotein activity (inferred from digoxin uptake) (18) than either individual variant. Furthermore, the correlation between C3435T SNP and P-glycoprotein activity/expression does not appear consistent across ethnic groups. Studies in Caucasian populations have shown a general trend of association between the MDR1 3435 TT-genotype and lowered P-glycoprotein activity/expression, whereas this association appears generally reversed in the Japanese (21–23). These discrepancies, and the observation that these SNPs display tight linkage disequilibrium (LD) in Europeans (24), suggest to us that the hypothesis that 3435 site may be in tight LD with another causal variant remains the most plausible explanation.

Six studies have examined the effect of these candidate SNPs (C3435T and G2677T) either singly or in combination in determining susceptibility to IBD (25–30). Schwab et al. (25) initially reported that in an age- and sex-matched case–control study that both T-allele and TT-genotype of the 3435 SNP were associated with increased susceptibility to UC but not to CD. In a North American study examining both G2677T and C3435T SNPs, Brant et al. (26) showed an association only between the Ala893 polymorphism (G2677T) and IBD. Significant association was observed from both case–control study and pedigree disequilibrium test with IBD. Two other studies failed to show an association with the C3435T SNP (27,28). Potocnik et al. (29) demonstrated a haplotypic rather than an allele-specific association with UC and corticosteroid refractory CD.

Data from our unit involving a large well-characterized population of IBD, however, replicated the association with UC (30). In addition, several other important findings in our data set suggested that the contribution of the MDR1 gene may be more complex than originally thought. First, on sub-phenotypic analysis, the strongest association was seen with extensive and severe UC. This suggests that phenotypic heterogeneity (i.e. composition of the studied population) may have contributed to the apparently inconsistent results observed in other studies. Secondly, two-locus haplotype analysis (G2677T and C3435T SNP) showed protective and susceptible haplotypes. Thus, it appears that variants of the MDR1 gene can alter the risk of developing UC in a bi-directional fashion. It is possible that specific haplotypes can either increase or decrease susceptibility by altering the level of PgP expression.

To obtain a more robust assessment of the contribution of the ABCB1/MDR1 gene to disease susceptibility, we utilized an approach where highly informative ‘tagging’ SNPs (tSNPs), which represent the haplotypic variations of this gene, are used to test for association. This involved a three-step approach, where first the ABCB1/MDR1 haplotype structure was initially characterized by re-sequencing this gene in 24 Centre d’Etude du Polymorphisme Humain (CEPH) Caucasian trios, previously presented by Soranzo et al. (24). Having established the pattern of LD, we identified a set of SNPs, which represent or ‘tag’ the common variations of the region, in this case, the gene. We then genotyped the tSNPs in the study population and employed likelihood calculations to determine whether there was a difference between case and control populations. In this study, we have identified six tSNPs by using a selection strategy which is ‘block’-free initially described by Weale et al. (31). This is therefore a ‘gene’-wide study and also implicitly a candidate gene rather than a candidate polymorphism study.
RESULTS

The re-sequencing of 12 amplicons distributed along the length of the \textit{ABCB1/MDR1} gene and corresponding to a total of 4.1 kb identified 17 SNP loci in 24 Caucasian CEPH trios as described in an earlier study (24). Three loci had a low minor allele frequency (\(\leq 6\%\)) and were therefore excluded from further analysis as these low-frequency variants were unlikely to be responsible for the previously documented association between UC and the common C3435T SNP (25,30). Using 14 SNPs with high minor allele frequency, a set of six tSNPs were identified, which provide a coefficient of determination of at least 0.80 in predicting all the known SNPs (Table 1) (31,32). We have genotyped these tSNPs in a panel of 249 and 179 individuals with UC and CD, respectively, and 260 healthy controls (HC) in a Scottish-Caucasian population. The genotypic frequencies of the six selected tSNPs are shown in Table 1. The G-allele of rs3789243, tSNP1 \((P = 0.03, \text{OR} 1.31, 95\% \text{CI} 1.03–1.68)\) and GG-genotype \((P = 0.04, \text{OR} 1.76, 95\% \text{CI} 1.06–2.92)\) were significantly higher in patients with UC compared with controls. In addition, there was a trend towards increased frequencies of the T-allele \((P = 0.07, \text{OR} 1.26, 95\% \text{CI} 0.98–1.62)\) and TT-genotype \((P = 0.12, \text{OR} 1.54, 95\% \text{CI} 0.92–2.54)\) of rs1045642, tSNP4 (C3435T) in UC compared with HC. No other allelic associations were observed in UC or CD.

Using the log-likelihood ratio analysis which tests the association between the inferred haplotypes using disease/control partition (using EH+ which makes a likelihood assessment that includes statistical uncertainty in phase estimation), we detected a strong association of haplotypes with UC, with a highly significant \(P\)-value of 4.22 \(\times\) 10\(^{-7}\) (35 degrees of freedom).
freedom). In contrast, no association was seen with CD ($P = 0.22$) (Table 2). The log-likelihood analysis was most significant in the phenotype of extensive UC (colitis extending beyond the splenic flexure) ($P = 1.7 \times 10^{-7}$) and less in recto-sigmoid disease (0.0089). A weaker association was also observed in patients requiring surgery within the UC group ($P = 0.0002$). No association was seen within the subphenotypes of disease location and behaviour in CD (data not shown).

Having found an overall association with UC, we assessed the individual contribution of each of the common haplotypes, according to the strength of association with UC compared with controls. The haplotypes are described in Table 3. In this case–control analysis, we defined haplotypes which were more common in cases than controls as ‘risk’ haplotypes. Haplotypes which were more common in controls were defined as ‘protective’ haplotypes. For UC, two risk haplotypes (111211 and 111212) and one protective haplotypes (211212) were found in the UC group. It is of great interest to note that the risk and protective haplotypes differed only at tSNP1, with the G-allele conferring susceptibility ($P = 0.0054–0.0002$, OR 2.24–3.07, 95% CI 1.26–5.73) and A-allele, protection ($P = 0.0001$, OR 0.36, 95% CI 0.24–0.56) (Table 3).

We further analysed all combinations of haplotypes to (i) determine which tSNP(s) may be responsible for the observed association and (ii) determine whether the effect observed is haplotype-specific. We therefore performed a log-likelihood analysis on all possible sets of two-, three-, four- and five-locus haplotypes (a total of 56 different sets of haplotypes). Only haplotypes containing rs3789243, tSNP1, were found to retain a significant association with UC (Table 4). It is particularly noteworthy that in the 26 haplotypes without rs3789243, tSNP1, none of these approached significance.

Eleven haplotypes (from a total of 30) containing rs3789243, tSNP1, but without rs1045642, tSNP4 (C3435T), demonstrated highly significant association. Conversely, there was no association seen in the total of 15 haplotypes which contained rs1045642, tSNP4 (C3435T), but not rs3789243, tSNP1. This implies that rs1045642, tSNP4 (C3435T) is not responsible for the overall association seen using the six-locus haplotype. We have focused on rs1045642, tSNP4 (C3435T) due to the controversies surrounding the contribution of this SNP to disease susceptibility in previous studies (25–30).

Having found an overall association with UC, we systematically examined the direct haplotypic frequencies for UC and controls were compared using Fisher’s exact test.

For each of the selected haplotypes, the estimated inferred haplotypic frequencies for UC and controls were compared using SNPHAP programme

<table>
<thead>
<tr>
<th>tSNP1</th>
<th>tSNP2</th>
<th>tSNP3</th>
<th>tSNP4</th>
<th>tSNP5</th>
<th>tSNP6</th>
<th>UC (%)</th>
<th>HC (%)</th>
<th>$P$-value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>23.2</td>
<td>29.9</td>
<td>0.80</td>
<td>0.96 (0.72–1.29)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>7.8</td>
<td>2.7</td>
<td>0.0002</td>
<td>3.07 (1.65–5.73)</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>7.5</td>
<td>3.5</td>
<td>0.0054</td>
<td>2.24 (1.26–3.99)</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2.9</td>
<td>2.9</td>
<td>1.00</td>
<td>1.04 (0.50–2.16)</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2.5</td>
<td>3.0</td>
<td>0.57</td>
<td>0.77 (0.47–1.98)</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2.4</td>
<td>3.4</td>
<td>0.36</td>
<td>0.69 (0.33–1.44)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1.6</td>
<td>1.5</td>
<td>1.00</td>
<td>1.04 (0.39–2.81)</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>11.6</td>
<td>14.1</td>
<td>0.22</td>
<td>0.79 (0.55–1.15)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>6.4</td>
<td>6.7</td>
<td>0.89</td>
<td>0.95 (0.58–1.56)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>6.2</td>
<td>15.4</td>
<td>0.0001</td>
<td>0.36 (0.24–0.56)</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2.5</td>
<td>2.3</td>
<td>1.00</td>
<td>1.05 (0.46–2.35)</td>
</tr>
</tbody>
</table>

**DISCUSSION**

This gene-wide approach provides the most robust evidence yet to support the genetic contribution of the ABCB1/MDR1 gene to the susceptibility of UC. We observed a highly significant association (log-likelihood ratio test $P = 4.22 \times 10^{-10}$) with UC and not CD in our population. As this study accounted for the haplotypic variations across the gene, the results are much stronger than the previously published studies utilizing candidate polymorphisms, C3435T and G2677T/A (25–30). In addition, we have again shown that the contribution of germ-line variations of the ABCB1/MDR1 gene is primarily most significant in the phenotype of extensive colitis. There remains a degree of controversy of whether the ABCB1/MDR1 gene contributes only to disease susceptibility in UC and not CD. Previous genetic studies have presented a more inconsistent picture. The negative result in CD, which we have found from our haplotypic approach, sets a statistical limitation to the importance of common variants in this gene and their contribution to CD-susceptibility in our population.

Confirming the initial impression of the complex contribution of the ABCB1/MDR1 gene, our data show that the highly significant association is critically dependent on an upstream intronic variant rs3789243, tSNP1, which is also independent of the effect of the often cited rs1045642, tSNP4 (C3435T). Although we cannot conclusively disprove the case for the rs1045642, tSNP4 (C3435T) as the ‘causal’ variant for previously observed association, this finding...
strongly suggests that this is not likely to be the case. This lends considerable support to the current thinking that C3435T and G2677T/A (in exons 26 and 21, respectively) are not the causal variants but lie in LD with it/them. The replication at the haplotype level, based on the tSNPs identified by our study, in other European Caucasian populations is now crucial and will be stronger than further candidate SNP analysis in this field in IBD, as already more extensive variation has been studied and tagged in this essentially gene-based study (33).

Only haplotypes containing tag 1 retains significance. d.f., degrees of freedom; H1, alternative hypothesis.
For each of the respective two-, three-, four-, five- and six-locus, we have selected the most significant haplotypes (Table 4) for analysis. For each of the selected haplotypes, the inferred haplotypic frequencies for UC and controls were compared using Fisher’s exact test. The most significant specific haplotypes are shown in Table 4.

In our previous study, we had demonstrated an association with rs1045642, tSNP4 (C3435T) comparing 335 UC and 370 HC (3435T allele frequency, 58.2 versus 52.8%; \( P = 0.02, \text{OR} 1.26, 95\% \text{CI} 1.03–1.58 \) (30). In the present study, we did not observe a significant difference in this SNP. However, the size of effect as demonstrated by the OR was similar to our previously published data (\( P = 0.07, \text{OR} 1.26, 95\% \text{CI} 0.98–1.62 \) (30). Assuming an effect based on single locus analysis of C3435T in the magnitude of OR 1.2–1.5 based on published data (25,30), the current cohort is relatively underpowered to detect these differences.

It is of interest to note that modest associations at allelic and genotypic level can translate to very significant observed associations at the haplotype level. A few issues are pertinent. The apparent contrast between the strength of the association with any given SNP, at the single locus level, and the level of significance derived from the log-likelihood analysis reflects the fundamental differences between these statistical techniques. First, the log-likelihood method for haplotype analyses employed in this study provided an indication for the levels of significance of the difference between two hypotheses/models in the distribution of haplotypes but not the measure of the size of effect. Notwithstanding this, when we analysed the estimated frequencies of directly inferred haplotypes between UC and controls, the derived estimated OR in each of the most significant haplotypes using respective two-, three-, four- and five-tSNPs ranged from 2.37 to 3.27 (95% CI 1.59–5.58). These derived ORs are consistent in terms of the magnitude of effect compared with those associated with proven genetic determinants implicated in other complex diseases, including CD (4–8,34–41). This effect size was only modestly greater than that associated with the carriage of the risk G-allele of the critical rs3789243, tSNP1 (G-allele OR 1.31, 95% CI 1.03–1.68). It also remains possible that the haplotypic approach is collectively picking up the effect of rarer (as yet unknown) variant(s) in which the respective individual variants cannot achieve on a single locus level.

Secondly, although we have shown that haplotype association observed was dependent on rs3789243, tSNP1, it is premature to suggest that this SNP itself is the causal variant. On the basis of these data, it is conceivable that the observed contribution of germ-line variations of ABCB1/MDR1 may represent haplotypic association rather than the specific contribution of one or two individual SNPs. Future work should be directed towards deep re-sequencing of the associated interval surrounding rs3789243, tSNP1, in a cohort of patients with extensive UC, which we have demonstrated to be the sub-phenotype most implicated by this gene. Although we have further fine localized the region of interest in this gene, the ease or difficulty in pinpointing the causal variant(s) remains unpredictable. It is possible that the extent of LD could be so high that the causal variant(s) could lie anywhere in the block on the associated chromosome. This difficulty has been highlighted in the recent studies focusing on the OCTN1/2 variants within IBD5 locus, cytokine gene cluster on chromosome 5q31–33, a 250 kb haplotype of high LD which confers susceptibility to CD (7,34,40). It seems increasingly certain that once haplotypic replication has been attained in the case of ABCB1/MDR1 gene, functional studies examining the effect of the list of putative causal variants derived from further fine localization will be the direction to take.

The block-free haplotype tagging approaches utilized by this study, which circumvents the heuristically defined block boundaries, originally described by Weale and Goldstein (31,32,42), have been shown recently following extensive empirical evaluations to be efficient and effective method to represent both known and unknown common variations (43). In the case of this study, we have successfully used this method in an area where controversy still exists, e.g. the genetic contribution of the ABCB1/MDR1 gene to susceptibility of IBD. In our current approach, we have not determined the extent of LD with the ABCB1/MDR1 gene. Therefore, it is conceivable that this set of tSNPs can also pick up long-range LD signals from neighbouring regions in contrast to the more conventional concept of tagging major haplotypes in blocks of high LD (44).

It is pertinent that the implications of our findings extend beyond the field of IBD. As suggested earlier, there has been much interest surrounding the issue of the identification of the causal variant(s) underlying reported genetic associations of the candidate SNP rs1045642, tSNP4 (C3435T), with other clinical conditions such as drug-resistant epilepsy (24,45), immune recovery after initiation of anti-retroviral therapy in HIV (46), increased risk of renal cell carcinoma (47) and other drug responses in the field of pharmacogenetics. It would be of great interest to examine the contribution of these tSNPs in these other conditions. The economy and applicability afforded by this approach is now well demonstrated. If variations of the ABCB1/MDR1 gene are indeed implicated in these conditions, similar degrees of associations can be anticipated.

By using a gene-wide haplotype approach, which is increasingly accepted as the model for future genetic association.
studies (33), we have now provided very strong evidence to implicate the genetic involvement of the ABCB1/MDR1 gene in susceptibility to UC but not CD in our population. These data also crucially enable further targeted fine localization in the search for the critical causal variant for the observed association in our population.

MATERIALS AND METHODS

Patients and controls

This study was approved by the Lothian Research and Ethics Committee (LREC) and written consent was obtained from all patients. A total of 249 patients with UC and 179 patients with CD were recruited from the Lothian region, Scotland, UK. Diagnosis of IBD was determined by standard clinical, radiological, endoscopic and histological criteria. Phenotypic details were extracted prior to genotyping. A total of 260 HC were also recruited from the Lothian region between 2000 and 2002.

Table 6 summarizes the clinical characteristics of studied patients. The median ages of diagnosis for UC and CD were 35.8 years [interquartile range (IQR) 26.4–50.2] and 27.3 years (IQR 21.0–42.7), respectively. The ethnicity of our study population in both cases and controls were Scottish-Caucasians. There were more males in the UC cohort (54.5%) and fewer in CD (44.5%). These differences were not significant when compared with controls. The overall gender distribution was 124 males:136 females; median age of recruitment was 34.1 years (IQR 25.1–45.0). The extent of disease was documented at time of latest follow-up. We defined extensive disease as disease extending beyond splenic flexure and left-sided colitis as disease extending to the splenic flexure as determined by histological and macroscopic evidence. In discordant cases, the histological evidence was used. Patients who had developed acute severe attack of UC which satisfied the Truelove and Witts criteria. For CD, disease location and behaviour were classified according to the Vienna Classification (50). Disease behaviour was taken from time of diagnosis. Both groups comprised Scottish-Caucasians.

Genotyping

Genotyping of the subjects in the study was performed using TaqMan technology (tSNP1,2,3,4) (ABI, San Diego, CA, USA) or re-sequencing of PCR products (rs3789243, tSNP5,6) using ABI 3700 Big Dye Terminators. Sequences for tags 11 and 12 were scored twice independently using the Sequencher software (Gene Codes Corporation, Ann Arbour, MI, USA).

Selection of tSNPs

The strategy and performance criteria for the selection of the set of tSNPs for ABCB1/MDR1 used in this study have been described elsewhere (24). A set of six tSNPs that provide a coefficient of determination of at least 0.80 in predicting all the known SNPs were identified (31,32). The criteria for the choice of tSNPs were (i) average $r^2$ weighted by allele frequency between the tSNPs and all SNPs was $\geq 0.80$ and (ii) minimum $r^2$ between the tSNPs and each individual SNPs in the gene was 0.80.

Data analysis

Genotype and allelic frequencies between cases and controls were compared using a $2 \times 2$ contingency table and Fisher’s exact test. ORs were given with 95% confidence intervals and two-sided $P$-values. All calculations were performed using the Graph Pad Instat programme (Graph Pad Software, San Diego, CA, USA).

Haplotypes frequencies of the ABCB1/MDR1 tSNPs were inferred using the expectation–maximization algorithm implemented in SNPHAP (48). Haplotype associations with disease were assessed in two ways. (i) a log-likelihood ratio analysis was implemented in the EH and PM programmes (bioinformatic programmes available and accessed via the Medical Research Council-Rosalind Franklin Centre of Genomic Research website: www.rfcgr.mrc.ac.uk) (48,49). This involves the calculation of the log-likelihoods of inferred haplotypes. Significance for association is calculated using the test statistic $2*\ln(L_{case}) + \ln(L_{control}) - \ln(L_{case/ control})$, which has a $\chi^2$ distribution with $n - 1$ degrees of freedom (where $n$ = number of inferred haplotypes). (ii) Contingency tests were also used to test the association for all haplotypes showing a frequency difference between cases and controls. There were no differences in the estimated haplotype frequencies derived from either SNPHAP or EH+ (which we have used complementarily to detect subtle differences). We measured the LD between SNPs using Cocaphase Software (also accessed via the Medical Research Council-Rosalind Franklin Centre of Genomic Research website: www.rfcgr.mrc.ac.uk).
ACKNOWLEDGEMENTS

Dr G.-T.H. was supported by the Chief Scientist’s Office, Scottish Executive, UK (2001–04). Dr A.T. is funded by Cancer Research UK grant C348/A3758. Dr E.R.N. is supported by program grant from the Wellcome Trust, C27289/Z/03/Z.

Conflict of Interest statement. None declared.

REFERENCES


Conflict of Interest statement. None declared.


