Identification of QTLs for serum lipid levels in a female sib-pair cohort: a novel application to improve the power of two-locus linkage analysis

Mario Falchi¹,²,*, Toby Andrew¹, Harold Snieder¹,³, Ramasamyiyer Swaminathan⁴, Gabriela L. Surdulescu¹ and Tim D. Spector¹

¹Twin Research and Genetic Epidemiology Unit, St Thomas’ Hospital, London, UK, ²Medical Genetics Unit, Department of Mother and Child, University of Modena and Reggio Emilia, Modena, Italy, ³Department of Pediatrics, Georgia Prevention Institute, Medical College of Georgia, Augusta, GA, USA and ⁴Department of Chemical Pathology, Guys and St Thomas NHS Foundation Trust, London, UK

Received June 14, 2005; Revised August 8, 2005; Accepted August 24, 2005

Using a novel approach for a two-locus model that provides a greatly increased power to detect multiple quantitative trait loci (QTLs) in simulated data, we identified in a sample of 961 female sib-pairs, three genome-wide significant QTLs for apolipoprotein A1 on chromosomes 8p21.1–q13.1 (LOD score 3.71), 9q21.32–33.1 (LOD score 3.28) and 10p15.1–p13 (LOD score 5.51), two for lipoprotein (a) on chromosomes 6q25.2–q27 (LOD score 10.18) and 21q21.1–q21.3 (LOD score 4.57) and two for triglycerides on chromosomes 4q28.3–32.1 (LOD score 3.71) and 5q23.1–q32 (LOD score 3.60). The two-locus ordered-subset analysis has led to the confirmation of known and likely identification of novel regions linked to serum lipid levels that would have otherwise been missed and deserves wider application in linkage analyses of quantitative traits. Given the relative lack of power for the sample sizes commonly used in human genetics linkage studies, minor QTL effects often go undetected and those that are detected will be upwardly biased. We show through simulation that the discrepancy between the real and estimated QTL-effects is often likely to generate an unpredictable source of false-negative errors, using multi-locus models, reducing the power to detect multiple QTLs through oligogenic linkage analysis. The successful simultaneous modelling of the identified QTLs in a multi-locus context helps to eliminate false positives and increases the power to detect linkages, adding compelling evidence that they are likely to be reliable QTLs for these lipid traits.

INTRODUCTION

Complex traits are likely to be influenced by multiple loci with effects of different size. Under these circumstances, the application of a multi-locus model could help the dissection of genetic determinants and could provide more accurate estimates of gene location. Moreover, while performing a genome-wide scan (GWS) for the identification of novel genetic determinants for a complex trait, inclusion of established quantitative trait loci (QTLs) in the same model can potentially provide additional information as to the reliability of identified novel QTLs.

Variance component (VC) models can be extended to simultaneously evaluate linkage of a QTL to multiple loci in a conditional oligogenic linkage analysis (1,2). This article investigates the properties of an additive two-locus VC model of QTL linkage for sib-pairs. Our investigation was prompted by the observation that insufficiently powered studies will produce upwardly biased estimates of the QTL-effect size in single-locus models (3), but that this problem...
can be exacerbated for the multi-locus model. We will show that multi-locus model fitting problems, which arise from underpowered sample sizes commonly used for human linkage studies, may lead to an additional source of Type II error.

In general, for a given quantitative trait, some families might be informative for linkage at one QTL, others informative for a second and some might be informative for both. To this end, tests to identify linkage heterogeneity have been successfully applied in both parametric and non-parametric linkage analyses. When some families are informative for linkage at one QTL whose effect size has been overestimated, their residual genetic variance will be underestimated. Under these conditions, even if the same families might be informative for a second QTL, the lack of residual genetic variance hampers the simultaneous modelling of both QTL-effects, as if the family was unlinked to the second QTL.

The application of a linkage heterogeneity method can be used to offset this problem and potentially increasing the power of a VC analysis to detect multiple QTLs. We illustrate the utility of an ordered-subset analysis (OSA) via simulation and by application to serum lipid data in 961 dizygotic twin pairs of the TwinsUK Adult Twin registry. Using a two-locus linkage analysis, we identified seven genomic regions associated with lipid levels at genome-wide significance levels.

RESULTS

Simulation results

Using simulations, we assessed the relative contribution to the overall power and the false-positive rate of a subset analysis applied to the two-locus VC linkage analysis of sib-pair data. The relative contribution of OSA (at the 0.01 nominal level) to the overall power was expressed as the fraction of power due to the application of the OSA method (Fig. 1).

The subset analysis in the lowest-powered samples greatly improved the power of a two-locus VC to detect a minor QTL (D1) and two QTLs of equal effect size (D2). A minor relative improvement in power due to OSA was observed when the polygenic variance was large enough to accommodate the effects of both QTLs in the multi-locus model (D3). The improvement was negligible when two equal-effect QTLs were analysed in the larger sample (D4). The absolute power of the different datasets to identify the second QTL was obviously dependent on power to show a minimal linkage signal for both the QTLs in the single-locus model. Therefore, the absolute power ranged from a low of 7% for the D1 dataset to 60% for the D4 dataset. When both QTLs were identified as significant in the single-locus models, the OSA contribution to the overall power of the two-locus model followed the same trend as observed in Figure 1. Interestingly, in this case, the D2 simulation for 1000 sib-pairs increased absolute power from 40 to 94% using the OSA method, which is comparable to the 96% power seen in the D4 simulation with 3000 sib-pairs.

Figure 2 illustrates the influence of the QTL heritability upward bias estimation on the two-locus VC analysis. The sum of the two single-locus QTL heritabilities was plotted against the LOD score obtained in the two-locus analysis for the D2 simulation (upper scatter plot for each graph) and the D4 simulation (lower scatter plot for each graph). For a large number of the simulated D2 datasets, the residual polygenic variance estimated in the alternative model of linkage in the two-locus analyses tended toward zero. Black points highlight those simulations where the estimated residual polygenic variance, according to its distribution among all the simulations, lie below the first quartile.

When significant linkage was observed for a QTL, the median (first to third quartiles) of the estimated QTL-effect in the D2 dataset was 32.96 (31.08–35.47) and in the D4 dataset it was 21.47 (19.46–24.06), close to the expected value of 20%. The overall magnitude of the heritability estimates for both QTLs grew linearly with the significance of linkage observed in the two-locus model. This effect was particularly marked in the D2 simulation, in which greater upward bias may be observed by comparing the linear regression lines for both datasets in Figure 2A. Here, the linear relationship between estimated QTL-effects and LOD score is truncated because the genetic variance is not sufficient to accommodate both of the QTL estimates in the two-locus model. This ceiling effect is evidenced in Figure 2A by the dense number of two-locus analyses with residual polygenic variance tending toward zero and constrained to lie in the area of low evidence of linkage.

Figure 2B shows the effect of the OSA approach on the distribution of the two-locus analyses results. A large number of analyses, likely to have encountered a model-fitting fault, now reach nominal significance of linkage. Even if the lower powered D2 dataset causes a mean overestimation of the QTL-effects, represented by the different linear regression intercept values in Figure 2B, the rate of increased QTL effect sizes (i.e. the slope) is nearly equal for the two datasets after the application of the OSA method, resulting in similar power for the two-locus model applied to both datasets (i.e. 94% for D2 and 96% for D4).
The false-positive rate attributable to the OSA method, over and above the standard VC two-locus model, was negligible, reaching a maximum of 0.3% for the D2 simulation.

**Single-locus linkage analysis results**

A single-locus model GWS was used to identify the more significant loci (Fig. 3) for triglycerides (TG) on chromosome 4q28.3–32.1 (peak LOD score 3.71 on marker D4S420), apolipoprotein A1 (ApoA1) on chromosome 9q21.32–33.1 (peak LOD score 3.28 on marker D9S1776) and lipoprotein (a) [Lp(a)] on chromosome 6q25.2–q27 (peak LOD score 4.05 on marker D6S2112).
Suggestive evidence for two QTLs influencing variation in apolipoprotein B was found on chromosome 5, at marker D5S408 (LOD score 2.89) and marker D5S647 (LOD score 2.30), but we did not investigate these loci further here. No suggestive evidence for linkage was found for the other lipid variables [total cholesterol, LDL-cholesterol (LDL-C) and HDL-cholesterol (HDL-C)].

Two-locus linkage analysis results

In the light of the single-locus analysis results, we performed a conditional two-locus GWS for TG, ApoA1 and Lp(a), comparing the results obtained during the first GWS and the results obtained through the evaluation of the two-locus model before and after the application of the OSA method. The results are shown in Figure 4 and further details of the model are provided in Subjects and Methods section. For these data, the standard two-locus model results (thin line) were always less significant than the linkage signal obtained in the first single-locus GWS (dotted line). In contrast, the OSA linkage signals, expressed as the maximum ordered subset LOD scores (thick line), were always the most significant with empirical $P$-value $< 10^{-5}$ for the subset-specific LOD score peaks. OSA identified significant linkage for TG on chromosome 5q23.1–q32 (peak LOD score 3.60 on marker D5S2115), ApoA1 on chromosomes 8p21.1–q13.1 and 10p15.1–p13 (peak LOD scores of 3.71 on marker D8S601 and 5.51 on marker D10S1649, respectively) and Lp(a) on chromosome 21q21.1–q21.3 (peak LOD score 4.57 on marker D21S1914). The size of the sample identified by OSA, expressed as a proportion of the total number of sibs, was 0.81, 0.86, 0.85 and 0.49 for the respective identified linkage signals.

Using the single-locus analysis, all the identified QTLs (excluding chromosome 5 for TG at LOD score 2.64) also showed significant linkage. However, when analysed using a standard two-locus VC model, the evidence of linkage for all these signals dropped below the genome-wide significance level. For these regions, the residual genetic variance in the two-locus alternative model was always close to zero. The narrow heritability for ApoA1 was estimated to be 68%, and yet the QTL effects estimated for the significant regions using single-locus models on chromosomes 8, 9 and 10 were 38, 71 and 37, respectively. Consequently, the two-locus
model of linkage was unlikely to accommodate the effects of both QTL when the results for ApoA1 on chromosomes 8 and 10 were conditioned on chromosome 9. Similarly, the narrow heritability for Lp(a) was estimated to be 80% with the QTL heritability on chromosome 6 estimated to be 71%, and the narrow heritability for TG estimated to be 60% with the QTL heritability on chromosome 4 estimated to be 48%. We did not apply a multiple testing correction for scanning for the second QTL because all the linkage signals highlighted by OSA were genome-wide significant in the single point GWS, with the exception of chromosome 5 for TG which was suggestive. We would have applied a multiple testing correction if new loci had been detected only by the multi-locus model. However, in this particular context, a multiple testing correction should be over-conservative.

DISCUSSION

Multi-locus models of linkage in principle are more suited for the genetic study of complex traits, for which many genes of unknown function combine with environmental influences to control trait variation. Indeed, the successful simultaneous modelling of QTLs in a multi-locus model is expected to improve the power of the analysis, at the same time reducing the false-positive rate. Here, we show that the upward bias of estimated QTL-effects, arise from the relatively underpowered human linkage studies (3), is likely to generate an unpredictable source of false-negative errors in multi-locus model evaluation. The biased estimated QTL-effects reflect an inflated estimation of the genetic variance attributable to the QTL. In a multi-locus model, if the estimated variance attributable to the identified QTLs exceeds the trait genetic variance, the likelihood of the linkage model, and consequently the LOD score, is likely to be deflated giving rise to increased Type II error rates.

Therefore, most QTLs will fail to be detected or interpreted as false-positive results, when simultaneously modelled, even if significant evidence was observed for the single-locus model. This is particularly true when novel QTLs are conditioned upon a known QTL for the trait using a multi-locus model. We suggest the application of a linkage heterogeneity model to avoid this problem, showing, through simulations, a greatly increased power to detect QTLs in a two-locus model using sib-pairs. The approach has been applied to the analysis of serum lipid data in 961 female sib-pairs from the TwinsUK registry. A first GWS identified the most significant results, at genome-wide significance level, for ApoA1 on chromosome 9q21.32–33.1, Lp(a) on 6q25.2–q27 and TG on 4q28.3–32.1. The scan of the genome was then repeated for the second QTL because all the linkage signals highlighted by OSA were genome-wide significant in the single point GWS, with the exception of chromosome 5 for TG which was suggestive. We would have applied a multiple testing correction if new loci had been detected only by the multi-locus model. However, in this particular context, a multiple testing correction should be over-conservative.

5q23.1–q32, which in the OSA analysis increased to a significant linkage of 3.6. However, all these QTLs were missed using a standard two-locus linkage analysis. In the region surroundings the QTLs, the residual genetic variance in the two-locus alternative model was always close to zero. Therefore, the lack of significance can be ascribed to fitting problems of the two-locus model. This observation has important implications for claims made about the ability of the oligogenic model to increase power to detect minor QTLs using conditional likelihood models (1,2). While true in principle, if sample sizes are sufficiently large—in practice, Type II errors are likely to increase using oligogenic models applied to insufficiently powered samples (3).

This approach has led to the replication of known and identification of novel regions linked to serum lipid levels. The LPA gene on chromosome 6q27 [MIM 152200] has been estimated to account for a substantial proportion of the phenotypic variation in Lp(a) plasma levels (4). Through a two-loci GWS conditioned to the locus on 6q25.2–q27, we identified a novel, and perhaps major, locus that contributes to variation in serum Lp(a) levels on 21q21.1–q21.3. Three previous GWS studies have identified genome-wide significant loci for Lp(a), which were on chromosome 1 (5), chromosome 18 (6) and chromosome 19 (7). None of these loci appears to have a strong effect in our sample. The identified 16 cM region did not appear to harbour any obvious candidate genes.

To our knowledge, only one previous GWS has identified a significant locus for ApoA1 on 3p25.2 (8). The identified region for ApoA1 at 9q21.32–33.1 harbour a gene known to affect the trait levels, ABCA1 [MIM 600046], that lies about 2 cM away from the second most significant marker D9S1690 (LOD score 3.16). Although this decreases the number of potential genes, the confidence interval region still needs to be narrowed before initiating positional cloning studies. The same is true for the other ApoA1 regions, 8p21.1–q13.1 and 10p15.1–p13 regions which span 42 and 32 cM, respectively, and need to be fine-mapped. Interestingly, the peak LOD score on chromosome 8 includes marker D8S601, which lies about 1 cM from the retinitis pigmentosa 1 (RP1) gene [MIM 603937]. RP1 has been shown (9) to be a likely candidate determinant for plasma TG and HDL-C metabolism.

We report two novel loci that contribute to variation in serum TG levels: one on 4q28.3–32.1 and another on 5q23.1–q32. Other genome screens in different, randomly ascertained population samples reported evidence for significant linked loci on chromosomes 2p, 9q (10) and 15q (11). Significant loci were detected in families ascertained for familial combined hyperlipidemia on chromosome 4p (12) and in a familial type 2 diabetes sample on chromosome 19q (13). A QTL on chromosome 15q was identified through the analysis of a composite TG—HDL-C phenotype (14) and a bivariate TG—LDL-C study identified a QTL on chromosome 21q (15). Overall, there is little consensus among previous studies or between our study and the previously reported genome scans.

We did not identify any obvious candidate genes in the chromosome 4 region. For chromosome 5, marker D5S2115 at the linkage peak is about 1 cM from the SARA2 gene [MIM 607690], associated with disorders of fat
malabsorption (16). However, just as for ApoA1, the regions detected on chromosomes 4 and 5 for TG, which span 26 and 18 cm, respectively, need further fine mapping.

This OSA application can readily be extended to irregular family data structures by either accepting the P-values would be approximate or by applying some sort of weighting scheme. This is because for the simplified situation of sib-pairs, as for our study, the observations are equally informative for each family and are therefore interchangeable. When all families are of the same size and structure, this assumption is justified.

The described approach relies only on the additive effects of two unlinked and non-interacting QTLs. However, the same complications arise using a multi-locus model that includes epistatic (gene-to-gene) interactions. It is well known that power to detect epistasis in a QTL linkage framework is low (17,18). However, an additive VC model in some circumstances may be adequate for the detection of QTLs that in reality act epistatically (19). In this context, the estimates of additive variances may be inflated due to the unmodelled epistatic effects and are therefore more likely to be accommodated in a two-locus model using the OSA approach. We believe the two-locus approach outlined here could be a useful power-enhancing tool for many linkage analysis studies.

SUBJECTS AND METHODS

Multi-locus VC linkage analysis

The multi-locus model can be employed for different purposes. First, it can be used to increase power to detect QTLs. Using a sequential strategy, an initial GWS identifies the chromosomal location that yields the largest LOD score by fitting a single-locus model. The location of the putative QTL is subsequently retained in the VC model and the genome is scanned again. The procedure can be continued successively including newly identified QTLs in the model, until there is no further evidence of significant linkage in the genome. In principle, this strategy has the potential to increase the power to detect linkage, because the conditional tests have the effect of maximizing the relative signal-to-noise ratio of true QTLs (20). The same procedure can be adopted by a priori by conditioning the analysis on a well-known QTL for the trait (21).

Secondly, multi-locus models can also be used to identify false-positive results, by jointly modelling the effects of two or more loci that are initially determined to be significant on the basis of a genome scan that first considers each locus separately. This procedure is often adopted when one of the linkage signals is considered to be reliable (due to previous replication or other evidence) and its location is used in a conditional VC analysis to evaluate the reliability of the other linkage signals (22).

Under the simplified scenario of a two-locus VC analysis, the overall phenotypic variance $\sigma^2_P$ is modelled as the sum of the phenotypic variances due to additive effects of the two QTLs ($\sigma^2_{QTL1}$, $\sigma^2_{QTL2}$), an additive polygenic effect $\sigma^2_R$, assumed to be due to a large number of unlocalized loci acting additively and a random environmental deviation $\sigma^2_E$.

While taking into account the effect of the conditioning QTL1, we can test the null hypothesis that the $\sigma^2_{QTL2}$ additive genetic variance equals zero, by comparing the likelihood of this model with that of an alternative model in which $\sigma^2_{QTL2}$ is estimated.

$$\frac{\text{max} L(\sigma^2_{QTL1}, \sigma^2_{QTL2}, \sigma^2_R, \sigma^2_E)}{\text{max} L(\sigma^2_{QTL1}, \sigma^2_{QTL2}=0, \sigma^2_R, \sigma^2_E)}$$

Source of bias and heterogeneity approach

Assuming that the mode of inheritance for the two QTLs is correctly specified and complete information on chromosomal segregation is available, the hypothesis of linkage can still be incorrectly rejected, giving rise to a Type II error with probability $\beta$. The power $(1 - \beta)$ for a given dataset to detect QTLs is a function of both the overall heritability of the trait and the QTL heritability (23). As a function of the estimated VC's in the maximum likelihood framework, the QTL-effect size (or QTL heritability) of the $j$th QTL is expressed by the ratio:

$$k^2_{QTLj} = \frac{\sigma^2_{QTLj}}{\sigma^2_P}$$

It has previously been shown that QTL-effect size estimates obtained by commonly used methods can be upwardly biased for several reasons (3). For insufficiently powered datasets that are commonly used for linkage studies, the upward bias is particularly severe when the observation is conditional upon a highly stringent significance threshold (e.g. due to that imposed by multiple testing) and/or the true genetic effect is small.

In this context, after a GWS has been used to identify significant evidence of linkage for a trait, an upward biased estimate is likely to be observed for more than one of the identified QTLs. As a result, the observed overall trait genetic variance may be smaller than the sum of the estimated variances of the individual QTLs. Indeed, in some circumstances, the residual polygenic variance estimated by maximum likelihood tends towards 0 when the $j$th QTL is added to the model (24). Under such conditions the $(j+1)$th QTL cannot be accommodated in the model, because the VCs are usually constrained to remain non-negative in the maximum likelihood estimation framework.

If the estimated variance attributable to the identified QTLs exceeds the trait genetic variance, the likelihood of the linkage model, and consequently the LOD score, is likely to be deflated, giving rise to increased Type II error rates. An additional complication is that one is unlikely to be able to identify from the GWS results alone, which of the identified QTLs is more biased, and to what extent. Assuming that the linkage signal is not due to sampling error, failure to detect significant evidence of linkage for a multi-locus model may arise for two reasons, which are difficult to distinguish in practice. The first is loss of power due to a minor QTL effect being masked by conditioning upon a larger QTL. The second is a problem of statistical model fit due to the residual correlation.
or polygenic variance tending towards zero. Both these situations will result in some families failing to show evidence of linkage for a standard multi-locus model. By applying a linkage heterogeneity approach to the two-locus model, an increase in sample homogeneity for the identified subset of families is, therefore, expected to improve the power, provided the decline in sample size is offset by the increase in the observed QTL effect size.

A promising approach to deal with sample heterogeneity is through an OSA (25). OSA identifies the subset of families that provide the maximal evidence for linkage and also provides empirical P-values that reflect the significance of the increase in the LOD score in the identified ordered subset compared with the baseline LOD score in all families.

OSA can use a variety of covariates to define the optimal subset of families, including evidence for linkage at other loci (26). We used the evidence of linkage at the major QTL, as evaluated in a single-locus model, to identify the subset of families giving maximal evidence of linkage in the multi-locus model. For the standard two-locus model, individual families that show stronger evidence of linkage with the major conditioning QTL are less likely to be informative for the minor QTL or may have insufficient residual correlation to accommodate both estimated effects in the model.

As a result, subsets of informative families may be identified by ordering the family LOD scores for the two-locus model upon the rank of the single-locus model LOD score for the (apparently) larger QTL.

**Simulation study**

Quantitative trait phenotypes were simulated under four different scenarios. The trait genetic component was attributable to two unlinked QTLs acting additively and to a non-localized additive polygenic component. The four datasets were simulated to explore the effects of sample size, QTL heritability and extent of residual polygenic variance upon our results. The characteristics of the four simulated datasets are reported in Table 1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of sib-pairs</th>
<th>( h_1^2 )</th>
<th>( h_2^2 )</th>
<th>Total ( h^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>1000</td>
<td>0.3</td>
<td>0.15</td>
<td>0.6</td>
</tr>
<tr>
<td>D2</td>
<td>1000</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>D3</td>
<td>1000</td>
<td>0.2</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>D4</td>
<td>3000</td>
<td>0.2</td>
<td>0.2</td>
<td>0.5</td>
</tr>
</tbody>
</table>

\( h^2 \), additive heritability.

The empirical significance of a larger LOD score observed in a subset of N sib-pairs was assessed by randomly sampling \( N \) sib-pairs 100 000 times and evaluating the subset specific cumulative LOD score. The P-value for the subset-specific LOD score was estimated from the proportion of randomly sampled sib-pairs giving a LOD score equal or greater than that observed: 
\[ P\text{-value} = (\text{number of significant subsets} + 1)/(\text{number of samplings} + 1). \]

In the simulation study, the evaluation of a two-locus analysis was constrained by the requirement that a significant linkage signal was observed in at least one of the single-locus analyses. Therefore, the number of two-locus analyses performed, depended on the power of the sample under the specified genetic model or by its false-positive rate. To reach a consistent number of observations, thousands of simulations were performed to evaluate; for each dataset, not less than 2000 two-locus analyses for the power study and not less than 1000 two-locus analyses for the false-positive rate study.

**Subjects and clinical evaluation**

Twins from the registry were ascertained from the general population through national media campaigns in the UK for a wide variety of studies of common diseases and traits (27). The study was approved by the St Thomas’ Hospital Research Ethics Committee. Details concerning the lipid measurements are described in more details elsewhere (28). The phenotypic characteristics of the sample are summarized in Table 2. The body mass index (BMI) was calculated as weight divided by the square of height (kg/m\(^2\)). Extreme outliers (i.e. trait value below the mean − 3 SD or above the mean + 3 SD) were excluded from the analysis. Before the GWS, Lp(a) and TG were transformed by natural logarithm. Age and BMI were incorporated in the analysis as covariates.

**Genotyping**

The genome scan was based on DNA extracted from venous blood samples of the study subjects. Scans involved the use of standard fluorescence based genotyping for the analysis of 583 highly polymorphic microsatellite markers (approximate average spacing 10 cM) from the ABI Prism linkage mapping set (Applied Biosystems) and the Généthon Genetic Linkage Map (29), as has been described elsewhere (30). Map distances are based on the Marshfield map (http://research.marshfieldclinic.org/genetics). Allele frequencies were estimated from the whole sample of genotyped subjects.
Table 2. Median (25–75% quartiles) of age, BMI and lipid levels for the 1922 dizygotic twins of the TwinsUK Twin registry used in the analysis

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Median (25–75% quartiles)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>49.1 (40.2–57.5)</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>24.2 (21.9–27.0)</td>
</tr>
<tr>
<td>TC (mmol/l)</td>
<td>5.5 (4.7–6.3)</td>
</tr>
<tr>
<td>LDL-C (mmol/l)</td>
<td>3.4 (2.7–4.2)</td>
</tr>
<tr>
<td>HDL-C (mmol/l)</td>
<td>1.5 (1.3–1.8)</td>
</tr>
<tr>
<td>TG (mmol/l)</td>
<td>1.1 (0.8–1.5)</td>
</tr>
<tr>
<td>ApoA1 (g/l)</td>
<td>1.6 (1.5–1.9)</td>
</tr>
<tr>
<td>ApoB (g/l)</td>
<td>1.1 (0.9–1.4)</td>
</tr>
<tr>
<td>Lp(a) (mg/dl)</td>
<td>15 (7.0–39.0)</td>
</tr>
</tbody>
</table>

BMI, body mass index; TC, total cholesterol; LDL-C, low density lipoprotein cholesterol; HDL-C, high density lipoprotein cholesterol; TG, triglycerides; ApoA1, Apolipoprotein A1; ApoB, Apolipoprotein B; Lp(a), lipoprotein(a).

Linkage analysis of sib-pairs lipid levels

In light of the encouraging simulation results, we applied the OSA approach to the analysis of real data using lipid parameters from the TwinsUK cohort. The analyses were performed in two stages. In the first stage, a single-locus GWS identified the most significant results for the analysed traits. In the second stage, the GWS was repeated through a two-locus model for the traits that displayed significant results, conditioning the analysis on the QTL identified in the first stage. The analyses compared the results obtained through the evaluation of the two-locus model before and after the application of the OSA method. The empirical significance of the maximum ordered subsets LOD scores was assessed by sampling 100,000 permutations for each dataset. We chose as thresholds for significant linkage in the two-locus models, a P-value less than the nominal one-degree of freedom P-value associated with a LOD score of 3 (P = 0.0001). We further required the two-locus model to be an improvement over the nested models of lower dimension.

ELECTRONIC-DATABASE INFORMATION

Accession numbers and URLs for data in this article are as follows: Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/omim (for ABCA1, LPA, RP1 and SARA2).

ACKNOWLEDGEMENTS

We thank the Twin Research Unit staff and the Twins for their continued participation. We thank two anonymous reviewers for their valuable suggestions and comments on the original manuscript. This work has been supported by EU funding for the Euroclot and GenomEutwin FP6 (Ref: LSHM-CT-2004-005268, QLK2-CT-2002-01254) programs. The TwinsUK project receives support from the Wellcome Trust.

Conflict of Interest statement. None declared.

REFERENCES


