npg

ARTICLE

Meta-analysis of four new genome scans for lipid parameters and analysis of positional candidates in positive linkage regions

Bastiaan T Heijmans*,[1], Marian Beekman[1], Hein Putter[2], Nico Lakenberg[1],
Henk Jan van der Wijk[2], John B Whitfield[3,4], Daniëlle Posthuma[5], Nancy L Pedersen[6],
Nicholas G Martin[4], Dorret I Boomsma[5] and P Eline Slagboom[1]

[1]Molecular Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands; [2]Medical Statistics, Leiden University Medical Centre, Leiden, The Netherlands; [3]Department of Clinical Biochemistry, Royal Prince Alfred Hospital, Sydney, Australia; [4]Queensland Institute of Medical Research, Brisbane, Australia; [5]Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [6]Department of Medical Epidemiology, Karolinska Institute, Stockholm, Sweden

Lipid levels in plasma strongly influence the risk for coronary heart disease. To localise and subsequently identify genes affecting lipid levels, we performed four genome-wide linkage scans followed by combined linkage/association analysis. Genome-scans were performed in 701 dizygotic twin pairs from four samples with data on plasma levels of HDL- and LDL-cholesterol and their major protein constituents, apolipoprotein AI (ApoAI) and Apolipoprotein B (ApoB). To maximise power, the genome scans were analysed simultaneously using a well-established meta-analysis method that was newly applied to linkage analysis. Overall LOD scores were estimated using the means of the sample-specific quantitative trait locus (QTL) effects inversely weighted by the standard errors obtained using an inverse regression method. Possible heterogeneity was accounted for with a random effects model. Suggestive linkage for HDL-C was observed on 8p23.1 and 12q21.2 and for ApoAI on 1q21.3. For LDL-C and ApoB, linkage regions frequently coincided (2p24.1, 2q32.1, 19p13.2 and 19q13.31). Six of the putative QTLs replicated previous findings. After fine mapping, three maximum LOD scores mapped within 1 cM of major candidate genes, namely APOB (LOD = 2.1), LDLR (LOD = 1.9) and APOE (LOD = 1.7). APOB haplotypes explained 27% of the QTL effect observed for LDL-C on 2p24.1 and reduced the LOD-score by 0.82. Accounting for the effect of the LDLR and APOE haplotypes did not change the LOD score close to the LDLR gene but abolished the linkage signal at the APOE gene. In conclusion, application of a new meta-analysis approach maximised the power to detect QTLs for lipid levels and improved the precision of their location estimate.
European Journal of Human Genetics (2005) 13, 1143–1153. doi:10.1038/sj.ejhg.5201466;
published online 13 July 2005

Keywords: linkage analysis; meta-analysis; haplotype; lipid metabolism

*Correspondence: Dr BT Heijmans, Molecular Epidemiology, Leiden University Medical Centre, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands. Tel: +31 71 527 1914; Fax: +31 71 527 1985;
E-mail: b.t.heijmans@lumc.nl

Introduction

The risk of coronary heart disease is influenced by genetic factors. Identification of the underlying variation in genes and other genomic sequences may provide valuable insights into the biological pathways contributing to coronary heart disease. Achieving this goal by investigating the disease itself in human genetic studies is challenging.

Patients diagnosed with the same disease will frequently have been exposed to genetic disturbances of different biological pathways. A powerful alternative approach is the analysis of quantitative intermediate phenotypes correlating with disease risk.[1,2] For coronary heart disease, LDL- and HDL-cholesterol are well-characterised intermediate phenotypes.[1]

LDL particles contribute to coronary heart disease by accumulating in the vessel wall thus forming a lipid core and initiating various cellular responses that promote atherosclerosis.[3] LDL particles contain one molecule of the protein apolipoprotein B (ApoB), which serves as a ligand for receptors such as the LDL-receptor. Hence, levels of LDL-C and ApoB in plasma are highly correlated. The function of HDL particles is to transport cholesterol from peripheral tissues, including from LDL particles trapped in the vessel wall, back to the liver.[4] The major protein constituent of HDL, apolipoprotein AI (ApoAI), mediates the uptake of cholesterol. HDL particles contain a varying number of ApoAI molecules.[5] Therefore, levels of HDL-C and ApoAI are correlated although less strong than LDL-C and ApoB. Importantly, interindividual variation in plasma levels of LDL-C, HDL-C and their main apolipoproteins is considerably heritable,[6,7] which provides the rationale for studies aimed at identifying the underlying genetic variation.

A commonly adopted strategy to disentangle the genetic basis of complex phenotypes is localisation using a genome-wide linkage scan in families followed by identification using association approaches to scrutinise genetic variation in positional candidate genes. For lipid research, the power of this approach is illustrated by the recent identification of the *USF1* as a gene involved in familial combined hyperlipidaemia.[8] Notwithstanding successes like this and others,[2] genome-wide scans have attracted criticism for their relatively low statistical power and consequently high rate of false negative results.[9] An advantage of genome-wide scans, however, is that the design allows detecting the effects of multiple, individually rare allelic variants at a locus. Alternative approaches, like (genome-wide) association studies that rely on common single-nucleotide polymorphisms (SNPs) to tag haplotypes, depend on the validity of the 'common disease, common variant' hypothesis, which may not apply to all genetic variation influencing complex traits. Indeed, evolutionary models predicted rare variants to make up a large part of the genetic risk for common diseases,[10] and recently multiple rare variants of candidate genes were shown to significantly contribute to variation in HDL-C in the general population.[11] Hence, despite the concern of low power, genome-wide linkage scans may remain one of the major tools to map loci contributing to complex traits. Therefore, developing new analysis methods is vital to overcome the relatively low power of the study design. Meta-analysis methods that allow the simultaneous analysis of multiple genome scans are among the most promising solutions.

We performed four autosomal genome-wide linkage scans to localise quantitative trait loci (QTLs) influencing LDL-C and HDL-C and their major apolipoproteins, ApoB and ApoAI. To maximise power, data from the four twin populations were combined using a well-established meta-analysis method[12,13] newly applied to linkage analysis. The method estimates an overall LOD score using the mean of the sample-specific QTL effects weighted by the precision of these estimates (ie inversely weighted by their standard errors (SEs)). Next, the influence of haplotypes of positional candidates was tested within the same framework.

Subjects and methods
Subjects
The current study was performed using twin studies from three countries:[7] Australia (Semi-Structured Assessment for the Genetics of Alcoholism Study), The Netherlands (Netherlands Twin Registry) and Sweden (Adoption/Twin Study of Aging); from The Netherlands two twin samples were used: an adolescent and an adult sample. All twins were recruited without regard to any trait or disease status and may thus be regarded as representing the general population. Zygosity was initially determined using standard questionnaires on physical similarity and the degree to which others confused them (Australia, Netherlands Adult) or serological markers (Netherlands adolescent, Sweden). Subsequently, the relationship of dizygotic (DZ) twin pairs was checked using short tandem repeat markers measured in the genome scan with the software GRR.[14] Phenotypes and genotypes were available for 457 Australian, 83 adolescent Dutch, 117 adult Dutch and 44 Swedish DZ twin pairs. In addition, 76 parents and 75 siblings of Australian DZ twins without phenotypic data were genotyped to improve the estimate of the proportion of alleles shared identical-by-descent (IBD). One adult Dutch family and five Australian families had two DZ twin pairs and were analysed as such.

Phenotyping
Phenotyping was performed as described previously.[7] Briefly, blood samples were taken after an overnight fast except for the Australian twins. HDL-C levels were measured in plasma after precipitation of ApoB-containing lipoproteins. Plasma ApoB and ApoAI were measured using a Behring nephelometer (Australia), radial immunodiffusion (Netherlands adolescent), Beckman Instruments' Array Protein System (Netherlands adult) or a radio-immunoassay (Sweden). LDL-C was calculated using the Friedewald formula (LDL-C = total cholesterol−HDL-C−(triglycerides/2.2); unit mmol/l) and set to missing if plasma triglyceride concentration exceeded 4.52 mmol/l. Plasma levels of total cholesterol and triglycerides were determined using standard enzymatic methods.

Genotyping

Short tandem repeat polymorphisms A 395 marker genome scan (~10 cM spacing) was performed in Dutch, Swedish and half of the Australian DZ pairs (*n* = 250) by the Molecular Epidemiology Section, Leiden University Medical Centre, The Netherlands. The scan was performed in several phases. First, chromosome 19 was scanned using in-house markers at an ~8 cM spacing using an ALFexpress automated sequencer (Amersham Pharmacia Biotech) as described previously.[15] Next, chromosomes 1, 2, 6, 7, 8, 11, 15, 16 and 17 were scanned at an 18 cM spacing on the basis of early power calculations[16] with markers from Weber screening set 8 using an ALFexpress automated sequencer. While the scan was ongoing, results from other genome scans showed that an 18 cM spacing is too sparse. Therefore, additional markers were measured on these chromosomes to decrease the spacing to ~10 cM. These markers were taken from the Human Linkage Set v2.5 MD10 and HD5 (Applied Biosystems) and measured using an ABI Prism DNA Analyzer 3700 (Applied Biosystems). The remaining chromosomes were finished with the 10 cM Human Linkage Set v2.5 MD10 and the same equipment. Genotyping was performed according to the manufacturer except that the amount of primer pairs for the markers was reduced up to five-fold and duplex PCR reactions were designed if possible to reduce costs, time expense and amount of genomic DNA used. Lastly, fine mapping was performed for promising linkage regions on chromosomes 2 and 19 with 12 and 9 markers, respectively, from the Human Linkage Set v2.5 MD10 and HD5. Genotyping was carried out using Fragment Analyser 1.02 (Amersham Pharmacia Biotech) when using ALFexpress equipment and Genemapper 2.0 and 3.0 (Applied Biosystems) when using an ABI Prism DNA Analyzer 3700. For the purpose of quality control, ~5% of samples were genotyped twice and the results compared. These comparisons indicated that markers D5S416, D9S158, D11S4175, D12S86, D14S68, D18474 and D19S209 from the Human Linkage Set v2.5 MD10 could not be genotyped reliably in our hands due to unstable ratio of the signal of the long and short allele, low success rate or unclear 1-bp alleles.

An additional 730 marker genome scan (~5 cM spacing) was carried out in 354 Australian DZ pairs by the Mammalian Genotyping Service, Center for Medical Genetics, Marshfield, USA. Two marker sets were used: screening set 13 and 52. Marker ATA27C11 had 1-bp alleles and showed an excessive noncorrespondence rate when genotypes were compared for eight samples that were genotyped in duplicate. This marker was removed prior to statistical analysis. There was an overlap of 147 Australian pairs between the scans carried out in Leiden and in Marshfield so that the total number of scanned Australian DZ pairs is 457.

Mendelian errors were detected using PEDSTATS and unlikely double recombinants using MERLIN.[17] A locally developed SQL server database was used to store genotypic data, compare repeated genotypings, integrate the Leiden and Marshfield genome scans and generate files for linkage analysis. The location of the markers was taken from an integrated genetic map with interpolated genetic map positions (http://www2.qimr.edu.au/davidD/). The position is in Decode cM, estimated via locally weighted linear regression (lo(w)ess) from the Build 34.3 physical map positions and published Decode and Marshfield genetic map positions.

Single-nucleotide polymorphisms Tagging SNPs in the genes *APOB*, *LDLR* and *APOE* were selected from literature[18–20] and associated databases on the web (*APOB*: rs512535 (−837G/A; minor allele frequency (MAF) in Australian sample = 0.48), rs934197 (−516C/T; MAF = 0.34), rs693 (T2488T; MAF = 0.46), rs676210 (P2712L; MAF = 0.22), rs1801701 (R3611Q; MAF = 0.10), rs1042031 (E4154K; MAF = 0.18); *LDLR*: rs6511720 (MAF = 0.12), rs8102912 (MAF = 0.24), rs6511721 (MAF = 0.49), rs12980593 (MAF = 0.47), rs5925 (*Ava*II; MAF = 0.45), rs2738459 (MAF = 0.46), rs13306506 (*Msp*Ia; MAF = 0.24), rs1433099 (*Msp*Ib; MAF = 0.27); *APOE*: rs449647 (−491A/T; MAF = 0.19), rs405509 (−219G/T; MAF = 0.42), rs769450 (MAF = 0.49), rs429358 (ε3/ε4; MAF = 0.14), rs7412 (ε3/ε2; MAF = 0.08)). The two *APOE* SNPs underlying the ε2/ε3/ε4 isoforms were measured using PCR followed by *Hha*I digestion. For the remaining 17 SNPs, Assay Designer software (Sequenom) was used to design multiplex genotyping assays.[21] Genotyping was performed using the MassArray platform according to the protocols of the manufacturer (Sequenom). Briefly, after PCR, a primer extension reaction was performed to introduce mass-differences between alleles and, after removing salts by adding a resin, ~15 nl of the product was spotted onto a target chip with 384 patches containing matrix. Mass differences were detected using matrix-assisted laser desorption\ionisation time-of-flight mass spectrometry (MALDI-TOF) and genotypes were assigned real-time using Typer 3.0 software (Sequenom). As quality control, cluster plots were made of the low and high mass allele signals as detected using mass spectrometry. Calls falling outside a cluster and no-calls falling inside a cluster were checked manually. In addition, 5–10% of samples were genotyped in duplicate and no inconsistencies were observed.

Statistical analysis

Prior to analysis, phenotypes were transformed by the natural logarithm if appropriate (this was the case for all phenotypes but LDL-C in the Australian and LDL-C and ApoB in the adult Dutch sample), adjusted for sex and age and standardised, for each study population separately, so that all phenotypes had a mean of 0 and a standard deviation of 1. Phenotypic values of individuals using statins were included in the analysis without correction (11 Australian DZ twins from 10 pairs and four adult Dutch DZ twins from four pairs). To check the validity of this

approach, the linkage analysis was repeated after excluding statin users, which did not materially alter the results for LDL-C and ApoB levels.

Linkage analysis was performed using an inverse regression method[22] implemented as MERLIN-REGRESS in the software MERLIN. This method regresses the estimated proportion of alleles shared IBD on the squared sums and squared differences of standardised trait values of the relative pairs and yields an estimate of the QTL effect and its SE. Although designed for selected samples, this inverse regression method has asymptotically equivalent power to variance components linkage analysis for random samples,[23] but has advantages that it is fast and less sensitive to the assumption of multivariate normality than variance components analysis. Inverse regression needs the heritability of the phenotype to be specified, which has the advantage that an appropriate upper limit for the estimation of QTL effects is obtained and the risk of overestimation is reduced.[24] Heritabilities of standardised and age and sex adjusted phenotypes were estimated for each study population separately using variance components analysis as implemented in the software QTDT.[25] For this analysis, data on both monozygotic and DZ twins were used. Only the genetic (heritability) and unique environmental variances were estimated since previous work showed that neither dominance nor common environment affected the variance.[7] Heritabilities of lipid phenotypes were greater than 0.50 in all samples generally being somewhat lower among the Australian twins and highest among adolescent Dutch twins.[7] Sample-specific heritabilities and allele frequencies of short tandem repeat polymorphisms (for estimation of the proportion of alleles shared IBD) were used in the linkage analysis.

After analysing the four twin samples separately on a common 1 cM grid using the inverse regression method, the four QTL effects and their corresponding SEs were combined into a single QTL effect estimate and SE using meta-analytic tools that account for heterogeneity.[12,13] The single QTL effect estimate was calculated as the mean of the sample-specific QTL effects weighted by their precision (ie inversely weighted by their SE). Meta-analysis under heterogeneity acknowledges the fact that the true QTL effect may be different in different samples, and attempts to estimate the mean QTL effect around which these sample-specific QTL effects vary (using a random effects model). If no heterogeneity exists, the power of the meta-analysis method accounting for heterogeneity is the same as the method assuming homogeneity (ie without random effects model). The overall LOD score for each location was calculated using the weighted mean QTL effect (b) and its standard error ($SE(b)$), which can be converted into a Z-score ($Z = b/SE(b)$) and subsequently into a LOD score ($LOD = Z^2/2{*}\ln 10$). The meta-analysis method has been implemented in R (http://www.r-project.org/) and is available from the authors. A manuscript describing details on how this meta-analysis method is applied to linkage analysis is in preparation.

For putative QTLs that harboured well-known candidate genes, it was tested whether genetic variation in these genes contributed to the LOD score observed. If so, this would lent support to the linkages observed and the applied meta-analysis method. This analysis was performed in three steps for every twin sample separately. First, the expectation-maximisation algorithm implemented in SNPHAP was used to assign diplotypes (ie haplotype pairs) to individuals (http://www-gene.cimr.cam.ac.uk/clayton/software/). Diplotypes were assigned only if the probability of the most likely haplotype was greater than 50%. Second, the effect of haplotypes on lipid phenotypes was assessed using a maximum likelihood method that simultaneously estimates haplotype frequencies and haplotype–phenotype associations implemented in the software THESIAS.[26]

Table 1 Characteristics of four twin samples under investigation

	Australia	Netherlands adolescent	Netherlands adult	Sweden
Age	45.9 (11.7)	16.7 (2.0)	44.2 (6.6)	65.6 (8.4)
Male (%)	33.5	50.2	49.2	40
DZ pairs (*n*)	457	83	117	44
HDL-C (mmol/l)	1.44 (0.39)[1]	1.30 (0.27)	1.23 (0.35)	1.48 (0.43)[6]
ApoAI (g/l)	1.44 (0.28)[2]	1.39 (0.20)	1.72 (0.42)	1.40 (0.28)
LDL-C (mmol/l)	3.43 (0.97)[3]	2.64 (0.71)	3.67 (0.97)[5]	4.63 (1.23)[7]
ApoB (g/l)	0.98 (0.26)[4]	0.79 (0.17)	1.24 (0.34)	1.10 (0.24)
HDL/ApoAI				
Total correlation	0.70	0.77	0.71	0.73
Genetic correlation	0.78	0.87	0.77	0.77
LDL-C/ApoB				
Total correlation	0.79	0.80	0.88	0.82
Genetic correlation	0.87	0.83	0.94	0.85

Data are number, percentage or mean (SD).
1–7. Number of DZ twin pairs with incomplete data was 41 (for 1), 19 (for 2), 66 (25/66 because triglycerides >4.52 mmol/l; for 3), 18 (for 4), 1 (1/1 because triglycerides >4.52 mmol/l; for 5), 8 (for 6) and 6 (2/6 because triglycerides >4.52 mmol/l; for 7).
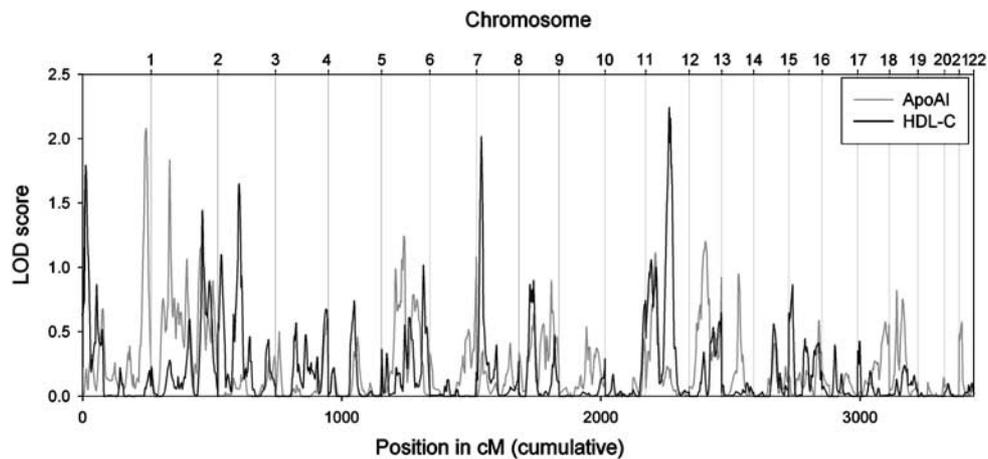
Figure 1 Result of a meta-analysis of four genome-wide scans for plasma levels of HDL-C and ApoAI.

Table 2 Chromosomal regions yielding a LOD score >1.5 in the simultaneous analysis of Australian, adolescent Dutch, adult Dutch and Swedish twins

Chromosome	Position in cM (Decode map)	Marker	LOD score	
			HDL-C	ApoAI
1p36.23	13	D1S1612	1.8	0.2
1q21.3	246	AAT243–AATA011	0.1	2.1
2p21	71	D2S391	0.3	1.7
3p14.2	81	GATA10H05	1.6	0.1
8p23.1	18	D8S1130	2.0	0.3
12q21.2	91	D12S326	2.2	0.2
			LDL-C	ApoB
2p25.1	25	D2S423	1.9	2.0
2p25.1[a]	25	D2S423	2.0	2.0
2p24.1	45	AAT263	1.6	2.1
2p24.1[a]	45	AAT263/D2S2150	2.1	2.1
2q31.2	184	D2Ss335–D2S1391	2.1	1.7
2q32.1[a]	187	D2S1391	2.0	1.7
5p15.33	0	D5S1981	1.6	0.3
19p13.12	37	D19S714	1.9	0.6
19p13.2[a]	33	D19S221	1.9	0.7
19q13.32	76	APOC2–D19S246	1.7	1.0
19q13.31[a]	73	APOC2	1.7	0.7

[a]After fine mapping.

Third, the estimated regression-coefficients associated with the various haplotypes were used to adjust the phenotypic value for the DZ twins and these haplotype-adjusted values were used for linkage analysis. Prior to haplotype analysis, QTDT was used to test for the presence of population stratification by comparing the within- and between-twin pair association of SNPs with the phenotypes under study.

Results
In total, 701 DZ twin pairs were studied from four twin samples: an Australian, an adolescent Dutch, an adult Dutch and a Swedish sample (Table 1). There was a strong correlation between levels of LDL-C and ApoB as well as between levels of HDL-C and ApoAI in all four samples (Table 1). Consistent with biology, the correlation between HDL-C and ApoAI (>0.70) was somewhat lower than that between LDL-C and ApoB (>0.79).

Genome-wide scans were performed on the four twin samples and were analysed simultaneously using a meta-analysis method. The principle of the meta-analysis method is that it combines effect estimates from the four twin samples inversely weighted by their SEs and is robust to heterogeneity. The result of the genome scan for the correlated lipid phenotypes HDL-C and ApoAI is shown in
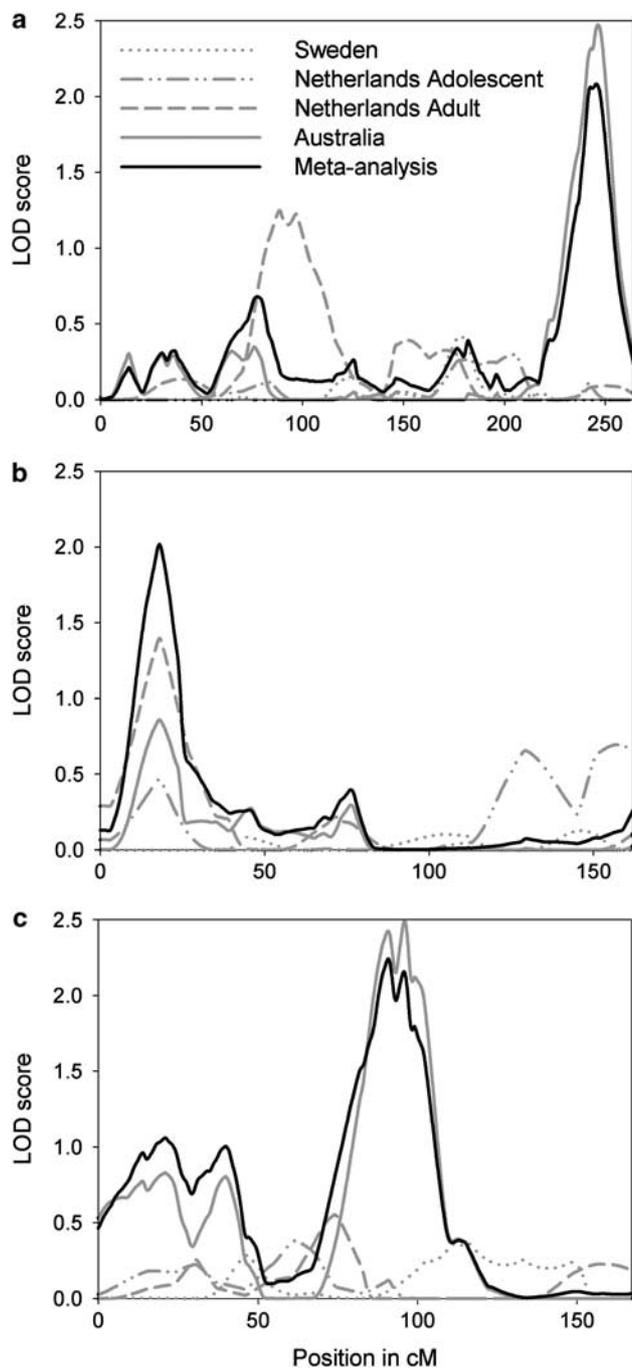
Figure 2 HDL-C and ApoAI: result of meta-analysis and analysis individual samples. Linkage of ApoAI with chromosome 1 (**a**), HDL-C with chromosome 8 (**b**) and HDL-C with chromosome 12 (**c**).

Figure 1. LOD scores greater than 1.5 were found on chromosomes 1, 3, 8 and 12 for HDL-C and on 1 and 2 for ApoAI (Table 2). There was no overlap in regions yielding a LOD score greater than 1.5 between the two lipid phenotypes.

Figure 2 shows the contribution of the four separate twin samples to the LOD scores greater than 2. Linkage with ApoAI on 1q21.3 (LOD = 2.1) and with HDL-C on 12q21.2 (LOD = 2.2) was mainly dependent on the Australian sample (sample-specific LODs on 1p36.23 and 12q21.2 both 2.5). In contrast, three out of four twin samples contributed to the LOD score of 2.0 found for HDL-C on 8p23.1 with the highest sample-specific LOD score being 1.4 in the adult Dutch sample.

Linkage signals observed for LDL-C and ApoB levels generally coincided (Figure 3 and Table 2). LOD scores greater than 2 were observed for both LDL-C and ApoB on 2p and 2q. LOD scores between 1.5 and 2 were observed for LDL-C on 19p and 19q; the corresponding LOD scores for ApoB varied between 0.5 and 1.0. Typing additional markers did not appreciably change the height of these LOD scores but shifted their position up to 4 cM (Table 2). Linkages with LDL-C and ApoB on chromosome 2 were mainly attributable to the Australian sample except for 2p24.1, which was also influenced by the adolescent Dutch and Swedish sample (Figure 4). In contrast, all four samples contributed to the linkage with LDL-C on chromosome 19 (Figure 4). These sample-specific linkage signals were distributed over the whole chromosome. The simultaneous analysis, however, indicated the presence of two distinct linkage signals on the p- and the q-arm.

After fine mapping, the position of maximum LOD scores observed for LDL-C on chromosome 2p24.1 (45 cM), 19p13.2 (33 cM) and 19q13.31 (73 cM) were close (~1 cM) to three main candidate genes *APOB* (44 cM), *LDLR* (32 cM) and *APOE* (72 cM). To assess the possible contribution of genetic variation in these genes to the linkages observed, tagging SNPs were measured in the twin pairs (six in *APOB*, eight in *LDLR* and five in *APOE*) and haplotype analysis was performed. Genotype distributions of all SNPs were in agreement with Hardy–Weinberg equilibrium and population stratification could be excluded by testing for differences in between and within twin pair association of SNPs with lipid phenotypes in the four samples. Diplotypes were assigned to DZ twins only if diplotype probability exceeded 50% (96.5% of individuals) and only part of the Australian sample was genotyped (250 DZ pairs). Nevertheless, linkage signals for LDL-C obtained using this subset of DZ pairs were similar as compared to those observed in the complete data set (Figure 5). Accounting for the association of *APOB* haplotypes in the linkage analysis reduced the LOD score for LDL-C from 1.8 to 1.0 and the QTL effect from 44 to 32% near the *APOB* gene. The strongest association of *APOB* haplotypes with LDL-C was observed in the Australian sample, namely for haplotypes ACCCGA (frequency, 7.2%; P = 0.0018) and ACCCGG (frequency, 6.4%; P = 0.012; see Subjects and methods for SNPs measured). *APOB* haplotypes did not influence the linkage signal observed 20 cM p-ter on the same chromosome. Accounting for the effect of the *LDLR*
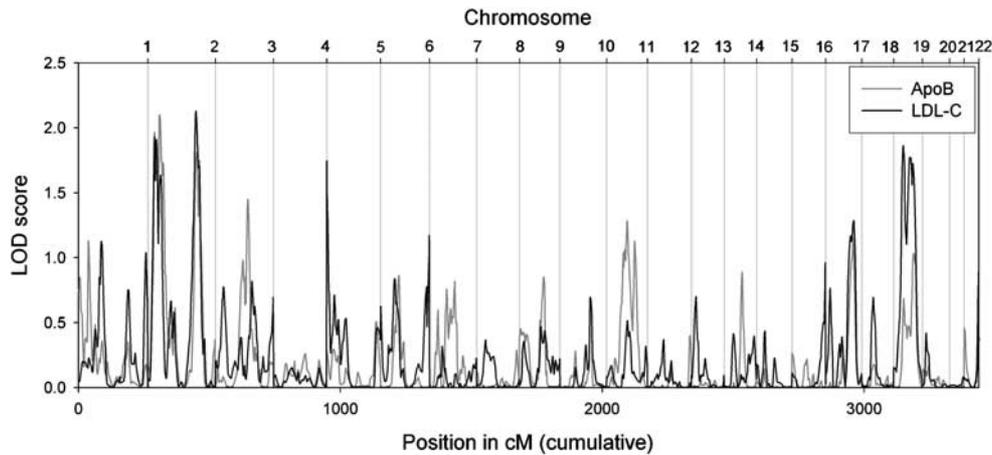
Figure 3 Result of a meta-analysis of four genome-wide scans for plasma levels of LDL-C and ApoB.
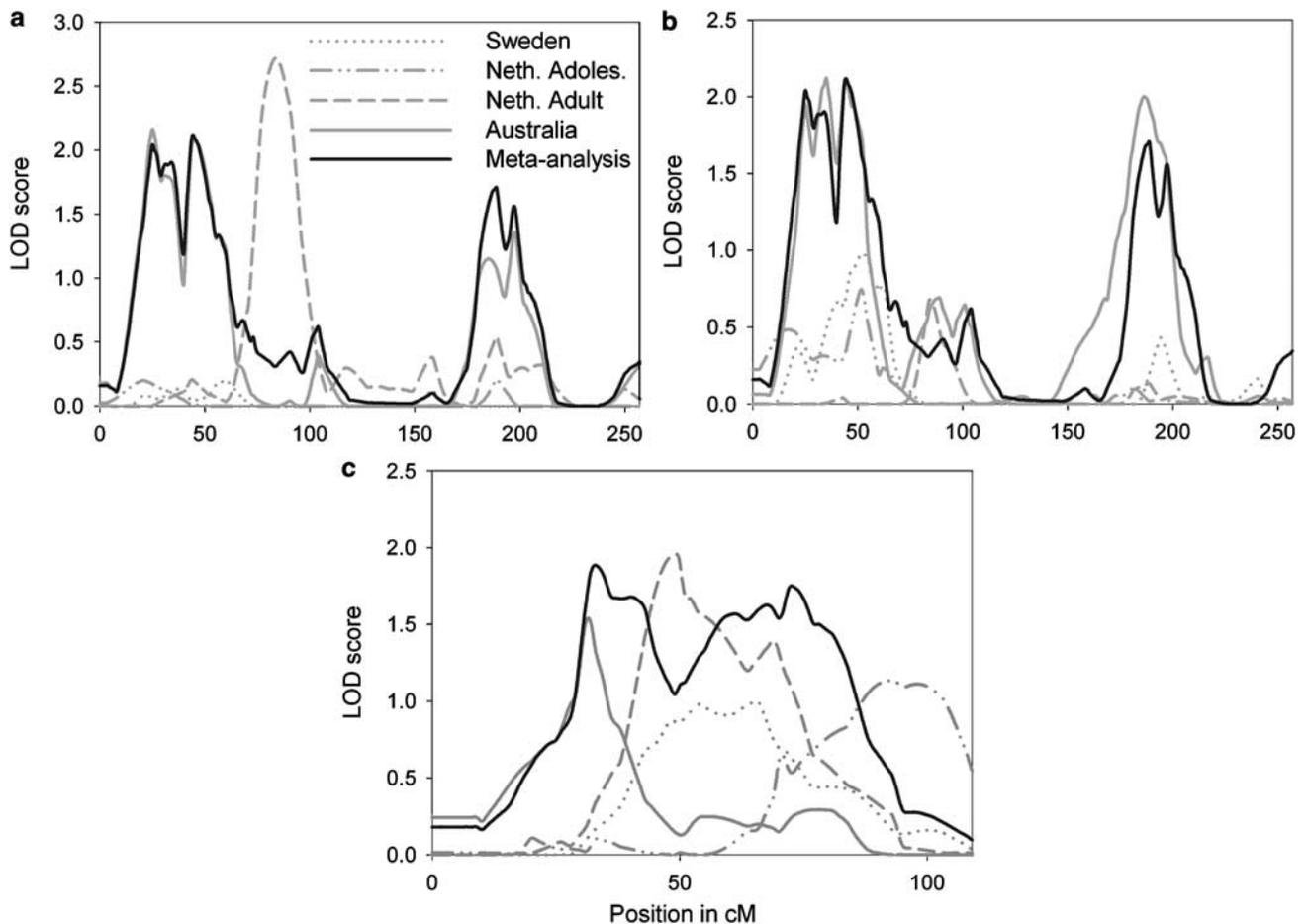

Figure 4 LDL-C and ApoB: result of meta-analysis and analysis individual samples. Linkage of ApoB with chromosome 2 (**a**), LDL-C with chromosome 2 (**b**) and LDL-C with chromosome 19 (**c**).

and *APOE* haplotypes did not change the LOD score close to the *LDLR* gene but abolished the linkage signal at the *APOE* gene. Moreover, the QTL effect was reduced from 39 to 9%, thus the haplotypes tested explained the bulk of the LDL-C QTL on 19q. As expected, especially *APOE* haplotypes harbouring the ε2 and, to a lesser extent, the ε4 allele
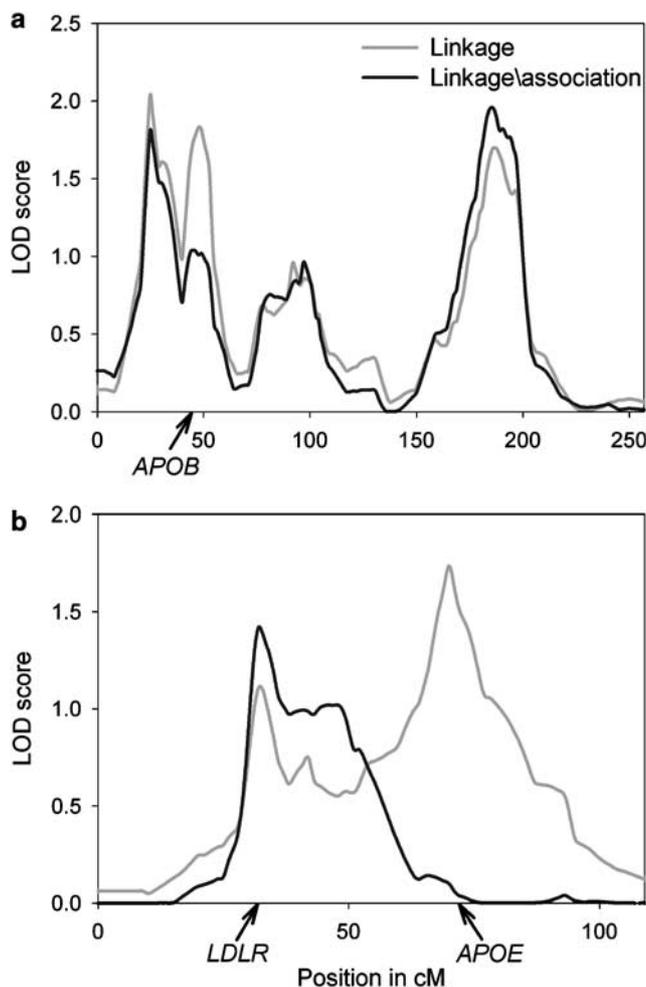
Figure 5 Effect on linkage with LDL-C of accounting for *APOB* haplotypes for chromosome 2 (**a**) and both *LDLR* and *APOE* haplotypes for chromosome 19 (**b**). Accounting for the effect of *APOB* haplotypes reduced the LDL-C QTL effect on 2p24.1 from 44 to 32% (**a**). Accounting for the effect of *LDLR* and *APOE* haplotypes reduced the LDL-C QTL effect on 19q13.31 from 39 to 9% (**b**).

were strongly associated with LDL-C levels in all four twin samples (data not shown).

Discussion

To map QTLs for HDL-C, LDL-C and their main apolipo-proteins, we performed a meta-analysis of four new genome-wide linkage scans on DZ twins from Australia, The Netherlands and Sweden. For HDL-C, suggestive linkage was observed on 8p23.1 and 12q21.2, whereas for its protein component ApoAI this was observed on 1q21.3. The explanation for the lacking overlap in linkage signals for HDL-C and ApoAI is not clear but may be related to differences in underlying biology[5] as described in the Introduction. Also, the power of our study may play a role since the QTL effects observed for ApoAI at putative HDL-C QTLs were greater than 10%, and thus compatible with the presence of genetic variation influencing both HDL-C and ApoAI.

For LDL-C and its protein component ApoB, suggestive linkage signals were coinciding on 2p25.1, 2p24.1 and 2q32.1. LOD scores between 1.7 and 1.9 for LDL-C with lower levels of evidence for ApoB were observed on 19p13.2 and 19q13.31. Haplotype analysis of the well-known candidate genes *APOB*, *LDLR* and *APOE* showed that genetic variation in these positional candidates was responsible for linkage of LDL-C with 2p24.1 and 19q13.31.

Thresholds for significant linkage in genome-wide scans remain a matter of debate (suggested LOD score thresholds include 3.6,[27] 3.0[28] and 2.0[29]). Additional data may assist in discriminating true from false positive results. Within our study, linkage of LDL-C level with 2p24.1 and 19q13.31 could be attributed to variation in positional candidates adding credibility to our linkage results and the meta-analysis method applied. Moreover, previous studies also

Table 3 Putative quantitative traits loci influencing lipid levels from the current study that are supported by previous studies

Chromosome	cM[a]	LOD	Phenotype	Population and reference
2p	25	2.0, 2.0	LDL-C, ApoB	Twins (this study)
	20	2.2	Total cholesterol	European-American sibling pairs[37]
2p	45	2.1, 2.1	LDL-C, ApoB	Twins (this study)
	40	2.2	LDL-C	Old order Amish[38]
2q	187	2.0, 1.7	LDL-C, ApoB	Twins (this study)
	164	3.7	LDL-C	Sardinian isolate[39]
12q	91	2.2	HDL-C	Twins (this study)
	74	4.1	HDL-C	Quebeçois families[32]
	115	2.1	Unesterified HDL-C	Mexican-American families[40]
19p	33	1.9	LDL-C	Twins (this study)
	33	2.1	LDL-C	Old order Amish[38]
	20	3.9	Total cholesterol	Pima Indians[41]
	35	2.3	LDL size fraction	Mexican-American families[42]
19q	73	1.7	LDL-C	Twins (this study)
	72	3.6	LDL-C	Quebeçois families[32]

[a]Position in cM originally reported transformed to our Decode genetic map.

observed an indication for linkage at these two loci (Table 3). Likewise, four other putative QTLs mapped in our study are backed by previous genome scans (Table 3). Such replication increases the probability that our and previous findings are true positives and warrant efforts to identify the underlying genetic variation.

Comparing different genome-wide scans, however, has limitations and interpretation of its result is not straight-forward. For example, various analysis methods are used, studies may focus solely on significant evidence for linkage not reporting potentially interesting suggestive linkages and, importantly, there are no criteria for the maximum distance between two mapped QTLs that can still be considered replication. Thus, comparison of published genome scans remains problematic and individual genome scans generally lack power, so producing many false negatives. Therefore, combining the available original data rather than the published results is imperative to obtain a complete map of QTLs for lipid phenotypes. Meta-analysis, using our method or others',[30] will be crucial to achieve this goal. A meta-analysis based map of QTLs for lipid phenotypes will provide valuable information, for example, on significant linkages previously reported for LDL-C[31,32] and HDL-C[33] that have not been replicated possibly due to lack of power of single genome scans. Also, such a map may shed light on the suggestion that the majority of the genetic variance of lipid phenotypes can be attributed to well-known candidate genes.[34]

The meta-analysis approach presented in this study is based on well-established methods developed for the analysis of clinical trials.[12,13] We newly applied it to genome scans analysed using an inverse regression method[22] that has asymptotically equivalent power to variance components linkage analysis,[23] and produces the SEs for QTL effects as is necessary for meta-analysis. However, any other linkage analysis providing effect estimates and SEs may be used. Importantly, the meta-analysis method provides valid estimates for the QTL effect and LOD score in the presence of heterogeneity without compromising power in the presence of homogeneity. Causes of heterogeneity between genome-wide scans include differences in QTL effect, overall heritability, phenotyping assay, differences in allele frequencies of short tandem repeat markers and different marker sets. Hence, heterogeneity between studies is a common phenomenon. All causes of heterogeneity mentioned were present in our study. The effect of heterogeneity in QTL effect is illustrated for the suggestive linkage with HDL-C on 12q21.2. Linkage was supported by the Australian sample but not by the other samples, and the overall LOD score was lower than that in the Australian sample alone. The increased power of meta-analysis if multiple samples support linkage was illustrated by the 8p23.1 result. None of the individual samples showed suggestive linkage and this putative QTL would have been missed. When analysed

simultaneously, however, the LOD score increased to 2.0.

Meta-analysis may not only provide more insight into the strength of the statistical evidence for linkage but may also allow a more precise mapping of QTLs in some cases. For instance, individual samples indicated the presence of linkage with LDL-C distributed over chromosome 19 whereas meta-analysis supported the presence of two distinct QTLs: one on the p-arm close to LDLR and one on the q-arm close to APOE. Interestingly, our previous combined analysis of chromosome 19 using a large part of the currently presented data analysed within a variance components framework did not resolve the two QTLs as our current method does.[15] A drawback of our meta-analysis method is that either raw data are needed or, logistically less complicated, different research groups need to reanalyse their data according to the same protocol and share QTL effects and SEs. However, initiatives have been launched to assemble existing genome scans.[35] Large-scale application of meta-analysis methods may thus become feasible in the near future.

Gene identification should follow linkage mapping of QTLs. We were able to ascribe two putative QTLs to haplotypes of positional candidates using a combined analysis of linkage and association. To our knowledge this is the first time that this was carried out successfully using unphased haplotypes. APOB haplotypes explained part of the linkage with LDL-C found at the APOB locus itself on 2p24.1. Factors that may have contributed to the observation that not all linkage was explained include the presence of additional relevant but undetected genetic variation either in APOB itself or in neighbouring genes, the possibility that the remaining LOD score simply constitutes a false positive signal and limitations of the statistical method used. As to the latter, future statistical approaches that take into account family relations when assigning haplotypes and use complete distributions of possible haplotypes for individuals may have higher power.

Haplotypes of positional candidates completely explained linkage with LDL-C at the APOE gene on 19q13.31. Interestingly, linkage with LDL-C at the LDLR gene was not explained by LDLR-haplotypes. Apart from limitations of our statistical method, it may be that genetic variation in the LDLR gene did indeed not contribute to the linkage result and unknown genes are involved. Alternatively, our haplotype analysis using common SNPs may not have been successful in capturing the relevant genetic variation. The latter is supported by extensive data showing that rare variants with severe or milder effects in the LDLR gene affect LDL-C levels[36] and would be compatible with the notion that linkage is able to detect the effects of multiple rare variants.[2] The same statistical framework may also be applied to the remaining replicated HDL-C and LDL-C QTLs to reveal the underlying genetic variation. In these instances, however, a positional candi-

date approach may not suffice since the linkage regions lack strong candidates in which case the limited current biological knowledge may be a poor guide to select genes for further study.

In conclusion, meta-analysis of multiple genome scans maximises power to detect QTLs and, in addition, may improve the precision of their location estimate. In conjunction with combined linkage/association analysis to test haplotypes, meta-analysis provides a powerful approach to disentangle complex traits. Using this approach we mapped multiple putative QTLs for lipid levels and identified genes underlying two of them.

Acknowledgements

References

1 Vogler GP, McClearn GE, Snieder H *et al*: Genetics and behavioral medicine: risk factors for cardiovascular disease. *Behav Med* 1997; **22**: 141–149.

2 Blangero J: Localization and identification of human quantitative trait loci: king harvest has surely come. *Curr Opin Genet Dev* 2004; **14**: 233–240.

3 Libby P: Inflammation in atherosclerosis. *Nature* 2002; **420**: 868–874.

4 Assmann G, Gotto Jr AM: HDL cholesterol and protective factors in atherosclerosis. *Circulation* 2004; **109**: III-8-14.

5 Tall AR, Breslow JL, Rubin EM: Genetic disorders affecting high-density lipoproteins; in: Scriver CR, Beaudet AL, Valle D, Sly WS, Vogelstein B, Childs B, Kinzler KW (eds): *The Metabolic and Molecular Bases of Inherited Disease*. New York: McGraw-Hill, chapter 121 (http://genetics.accessmedicine.com/). 2002.

6 Hamsten A, Iselius L, Dahlen G, de Faire U: Genetic and cultural inheritance of serum lipids, low and high density lipoprotein cholesterol and serum apolipoproteins A-I, A-II and B. *Atherosclerosis* 1986; **60**: 199–208.

7 Beekman M, Heijmans BT, Martin NG *et al*: Heritabilities of apolipoprotein and lipid levels in three countries. *Twin Res* 2002; **5**: 87–97.

8 Pajukanta P, Lilja HE, Sinsheimer JS *et al*: Familial combined hyperlipidemia is associated with upstream transcription factor 1 (USF1). *Nat Genet* 2004; **36**: 371–376.

9 Risch NJ: Searching for genetic determinants in the new millennium. *Nature* 2000; **405**: 847–856.

10 Pritchard JK: Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001; **69**: 124–137.

11 Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH: Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004; **305**: 869–872.

12 DerSimonian R, Laird N: Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–188.

13 van Houwelingen HC, Arends LR, Stijnen T: Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002; **21**: 589–624.

14 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: GRR: graphical representation of relationship errors. *Bioinformatics* 2001; **17**: 742–743.

15 Beekman M, Heijmans BT, Martin NG *et al*: Evidence for a QTL on chromosome 19 influencing LDL cholesterol levels in the general population. *Eur J Hum Genet* 2003; **11**: 845–850.

16 Beekman M, Lakenberg N, Cherny SS *et al*: A powerful and rapid approach to human genome scanning using small quantities of genomic DNA. *Genet Res* 2001; **77**: 129–134.

17 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.

18 Fullerton SM, Clark AG, Weiss KM *et al*: Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. *Am J Hum Genet* 2000; **67**: 881–900.

19 Crawford DC, Carlson CS, Rieder MJ *et al*: Haplotype diversity across 100 candidate genes for inflammation, lipid metabolism, and blood pressure regulation in two populations. *Am J Hum Genet* 2004; **74**: 610–622.

20 Tahri-Daizadeh N, Tregouet DA, Nicaud V, Poirier O, Cambien F, Tiret L: Exploration of multilocus effects in a highly polymorphic gene, the apolipoprotein (APOB) gene, in relation to plasma apoB levels. *Ann Hum Genet* 2004; **68**: 405–418.

21 Jurinke C, van den BD, Cantor CR, Koster H: The use of MassARRAY technology for high throughput genotyping. *Adv Biochem Eng Biotechnol* 2002; **77**: 57–74.

22 Sham PC, Purcell S, Cherny SS, Abecasis GR: Powerful regression-based quantitative-trait linkage analysis of general pedigrees. *Am J Hum Genet* 2002; **71**: 238–253.

23 Lebrec J, Putter H, Houwelingen JC: Score test for detecting linkage to complex traits in selected samples. *Genet Epidemiol* 2004; **27**: 97–108.

24 Beekman M, Heijmans BT, Martin NG *et al*: Two-locus linkage analysis applied to putative quantitative trait loci for lipoprotein(a) levels. *Twin Res* 2003; **6**: 322–324.

25 Abecasis GR, Cardon LR, Cookson WO: A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 2000; **66**: 279–292.

26 Tregouet DA, Barbaux S, Escolano S *et al*: Specific haplotypes of the P-selectin gene are associated with myocardial infarction. *Hum Mol Genet* 2002; **11**: 2015–2023.

27 Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; **11**: 241–247.

28 Morton NE: Significance levels in complex inheritance. *Am J Hum Genet* 1998; **62**: 690–697.

29 Rao DC, Province MA: The future of path analysis, segregation analysis, and combined models for genetic dissection of complex traits. *Hum Hered* 2000; **50**: 34–42.

30 Etzel CJ, Guerra R: Meta-analysis of genetic-linkage analysis of quantitative-trait loci. *Am J Hum Genet* 2002; **71**: 56–65.

31 Coon H, Eckfeldt JH, Leppert MF *et al*: A genome-wide screen reveals evidence for a locus on chromosome 11 influencing variation in LDL cholesterol in the NHLBI Family Heart Study. *Hum Genet* 2002; **111**: 263–269.

32 Bossé Y, Chagnon YC, Despres JP *et al*: Genome-wide linkage scan reveals multiple susceptibility loci influencing lipid and lipoprotein levels in the Quebec Family Study. *J Lipid Res* 2004; **45**: 419–426.

33 Peacock JM, Arnett DK, Atwood LD *et al*: Genome scan for quantitative trait loci linked to high-density lipoprotein choles-

terol: The NHLBI Family Heart Study. *Arterioscler Thromb Vasc Biol* 2001; **21**: 1823–1828.

34 Knoblauch H, Bauerfeind A, Toliat MR *et al*: Haplotypes and SNPs in 13 lipid-relevant genes explain most of the genetic variance in high-density lipoprotein and low-density lipoprotein cholesterol. *Hum Mol Genet* 2004; **13**: 993–1004.

35 Peltonen L: GenomEUtwin: a strategy to identify genetic influences on health and disease. *Twin Res* 2003; **6**: 354–360.

36 Ose L: An update on familial hypercholesterolaemia. *Ann Med* 1999; **31** (Suppl 1): 13–18.

37 Coon H, Leppert MF, Eckfeldt JH *et al*: Genome-wide linkage analysis of lipids in the Hypertension Genetic Epidemiology Network (HyperGEN) Blood Pressure Study. *Arterioscler Thromb Vasc Biol* 2001; **21**: 1969–1976.

38 Pollin TI, Hsueh WC, Steinle NI, Snitker S, Shuldiner AR, Mitchell BD: A genome-wide scan of serum lipid levels in the Old Order Amish. *Atherosclerosis* 2004; **173**: 89–96.

39 Falchi M, Forabosco P, Mocci E *et al*: A genomewide search using an original pairwise sampling approach for large genealogies identifies a new locus for total and low-density lipoprotein cholesterol in two genetically differentiated isolates of Sardinia. *Am J Hum Genet* 2004; **75**: 1015–1031.

40 Almasy L, Hixson JE, Rainwater DL *et al*: Human pedigree-based quantitative-trait-locus mapping: localization of two genes influencing HDL-cholesterol metabolism. *Am J Hum Genet* 1999; **64**: 1686–1693.

41 Imperatore G, Knowler WC, Pettitt DJ *et al*: A locus influencing total serum cholesterol on chromosome 19p: results from an autosomal genomic scan of serum lipid concentrations in Pima Indians. *Arterioscler Thromb Vasc Biol* 2000; **20**: 2651–2656.

42 Rainwater DL, Almasy L, Blangero J *et al*: A genome search identifies major quantitative trait loci on human chromosomes 3 and 4 that influence cholesterol concentrations in small LDL particles. *Arterioscler Thromb Vasc Biol* 1999; **19**: 777–783.