8 | Summary and General Discussion

Summary and discussion of results

Increased plasma levels of LDL cholesterol, total cholesterol, TG, apoB and apoE, as well as decreased plasma levels of HDL cholesterol, apoA1 and apoA2 are intermediate phenotypes of CVD. These quantitative traits in the lipid metabolism pathway are considered to be among the main risk factor for development of atherosclerosis[1,2], the primary cause of CVD. Over the last years, many candidate gene studies and genome-wide scans have been performed to determine the contribution of genetic loci to the clinical endpoints or intermediate phenotypes of CVD. In general, these studies differ in samples size, sample ethnicity, statistical methods and the way the study subjects and families in those studies were selected for various clinical endpoints of CVD. These differences in design, samples, and approaches often make both association studies and whole genome scan results difficult to replicate[3,4].

The project reported in this thesis is aimed at the mapping of major genes involved in the regulation of the quantitative traits of lipid metabolism in the general population. Phenotypes and marker data were collected in four population-based samples; an adolescent and adult twin pair sample from the Netherlands, one adult twin sample from Sweden and an adult twin sample from Australia. These four samples can be used to investigate replication, as well as to test QTL effects simultaneously.

Heritability of lipid and apolipoprotein levels

To investigate the influence of genes and environment on the variation of apolipoprotein and lipid levels, the heritability of these traits were estimated in the four twin pair samples (Chapter 2). Between 48-87 % of the variance in lipid and apolipoprotein levels are attributable to genetic factors in the four twin pair samples. Also the heritabilities of apoE and apoA2 levels, which have rarely been reported before, were found to be very high, comparable to other lipid and apolipoprotein levels. Common environment does not influence the variances and there are no sex differences in the heritabilities of these intermediate phenotypes. Overall, the heritability estimates were the highest in the Dutch samples and the lowest in the Australian sample. These differences in heritability were significant, indicating a sample specific effect on heritability of lipid and apolipoprotein levels or on the influence of unique environmental factors, which was overall the largest in the Australian sample. Higher unique environmental effects lead concomitantly to lower estimates of genetic factors. This larger influence of unique environmental factors could be explained by larger environmental effects, interactions with genetic factors or measurement errors of the trait. Overall, these twin cohorts from three different countries thus provide a unique sample for QTL mapping of the genes involved.

Genome scanning protocol

Since genetic variation seems to play an important role in lipid and apolipoprotein levels, a genome-wide scan is a logical subsequent analysis to localise the chromosomal regions harbouring putative quantitative trait loci (QTL). In our study, a total number of 1,000 DZ subjects need to be genotyped. Regularly, 10 cM spaced marker sets, including about 400 STRs, are used for genome scanning, which would result in roughly 400,000 genotypings in our genome scan. In Chapter 3, power simulations are carried out to estimate the optimal inter marker distance, resulting in an optimal balance between power to detect major genes and genotyping effort. From these simulations resulted that using an average spacing of 20 cM, there is 80% power to detect a QTL effect of 25% at the significance level of 5%. Because of the large samples size of the study, a specific and practically feasible protocol was developed at the start of the genome scan based on the simulations. Therefore, a screening set of 229 STRs with an average inter marker distance of 18.3 cM was chosen. A complete genome scan using this protocol requires 80 multiplex PCR reactions, which can all be carried out using one set of conditions and would consume 2.5 μg of genomic DNA, which is an exceptionally small amount for a genome wide scan. Current automated methodology allows for much higher throughput. In spite of the simulation study we would now prefer a much denser marker spacing, for example 5 or 10 cM inter-marker distance, which will not only gain the power in linkage analyses, to obtain genome wide significance, but also enables a more accurate search for possible genotyping errors and taking into account that not all markers will be completely informative.

QTL mapping

Linkage analyses were carried out on the data obtained from chromosomes 1, 2, 6, 7, 8, 11, 15, 16, 17 and 19. Before the linkage analyses, all family relationships were checked and an estimate was obtained of the genotyping error rate. For the current study, this is especially important since only few parental data were available, so Mendelian inheritance of marker alleles cannot be checked. All genotypings were used as input for the GRR software[5], which calculates the

proportion of shared alleles for each twin pair across all markers. The linkage analyses were finally performed in 83 DZ pairs for the adolescent Dutch, 117 for the adult Dutch, 44 for the Swedish and 249 for the Australian twins. Furthermore, marker data were checked for unusual double recombinants using SIBMED[6], the occurrence of which may be due to genotyping errors. After running SIBMED and checking the possible genotyping errors with the raw data, approximately 0.01% of the total genotypings appeared to be erroneous. In this way, only a small proportion of the potential genotyping errors can be found. The large inter-marker distance in our study increases the probability of a double recombination. The erroneous genotypes were corrected when possible or otherwise set to missing. Linkage analyses were performed with these cleaned data sets and reported in Chapter 4.

To investigate whether the genome scan with 18.3 cM spacing had indeed sufficient power to detect major loci, the *LPA* locus on chromosome 6q27, which is known to explain approximately 90% of the variance in Lp(a) levels, was tested for linkage. In all four samples linkage was found at chromosome 6q; in the adolescent Dutch and Swedish samples there was suggestive linkage[3] (MLS between 2.2 and 3.6) with MLS of 2.75 and 2.28 respectively. Additionally, in the adult Dutch twins as well as in the Australian twins significant linkage was found with MLSs of 4.40 and 5.13 respectively. Consequently, we have confirmed linkage according to the criteria of Lander and Kruglyak[3] in independent samples, which gives strong evidence that we located a QTL influencing Lp(a) levels. Even more, in a simultaneous variance components linkage analysis, in which the proportion of the QTL effect was equated over the four samples, the MLS increased to 9.8.

To investigate whether we could detect QTLs with a moderate effect using our genome scan protocol, the *APOE* locus was tested for linkage. The *APOE* gene is the structural coding gene for apolipoprotein E and is located on chromosome 19q13.2. Between 9 to 20% of the total variation in apoE levels has been ascribed to this candidate gene[7-9], a QTL effect much smaller then assumed in the simulation analyses (QTL effect of 25%) that formed the basis for the choice of the marker spacing. At this locus we found indications for linkage in the adolescent Dutch twins (MLS = 1.0) and in the adult Dutch twins (MLS = 0.6). No evidence for linkage was found in the Australian twins. The heritability in this population, however, was also reasonably low, raising doubts as to the quality of the apoE phenotypings performed in this twin sample. Since apoE levels were not measured in the Swedish twins, no linkage analysis with apoE levels could be performed in this sample. In the simultaneous analysis of the three twin samples, no evidence for heterogeneity was present and the MLS was 1.0. Since a MLS of 1.0 is expected to occur approximately 8 times at random in a genome scan[10], we conclude that we are unable to detect this QTL with a moderate effect on apoE plasma levels. With the 18 cM spacing in our studies, we cannot, as expected, detect a QTL explaining approximately 15% of the variance in apoE levels at a significant level in three twin pair samples totalling 449 pairs. The *APOE* example indeed illustrates that in our genome scanning approach, only QTLs with a major effect (>25%) on the trait can be detected. Consequently, we will miss as false negatives QTLs with smaller effects. Reconsidering, we are convinced that denser marker spacing would have improved the study since effects of 25% for a single QTL may be somewhat unrealistically high.

Analysing the 10 finished chromosomes for suggestive or significant linkage, we detected 11 putative QTLs; 4 loci influencing Lp(a) levels, 3 loci influencing apoA1 levels, 2 loci influencing apoB levels and 2 loci influencing LDL cholesterol levels. Two loci, in addition to the *LPA* locus, showed significant linkage (MLS>3.6) with Lp(a) levels. The significance criterion may

be too stringent for detecting QTLs in the separate twin samples, while only part of the analyses has the power to actually reach MLS higher than 3.6. This is why we consider that suggestive linkage results (MLS between 2.2 and 3.6) may also be valuable. Of the eleven loci showing significant or suggestive linkage, three loci could also be replicated in an other twin sample with MLS higher than 1.4. According to Lander & Kruglyak, a point-wise p-value of 0.01, corresponding with a LOD score of 1.4, is needed for a 20 cM interval-wide significance level of 0.05[3]. Two loci, on chromosome 1 and 2, show replicated linkage with Lp(a) levels, and chromosome 19 shows replicated linkage with LDL cholesterol levels. Although, total cholesterol, HDL cholesterol, triglycerides and apoE levels are all highly heritable, no significant or suggestive linkage results have been found. This might indicate that the major gene loci are not present on these 10 chromosomes, or that the QTL effects on these traits are smaller that 25%.

Replicated linkage with Lp(a) levels

Among the replicated loci are two loci influencing Lp(a) levels, on chromosome 1 and 2. In the region on chromosome 2, the *APOB* gene is located, encoding apolipoprotein B, which is the major constituent of the Lp(a) particle. It, therefore, could be hypothesised that mutations in the APOB gene might affect Lp(a) levels. However, in the two Dutch samples we found no evidence yet for linkage of chromosome 2 with Lp(a) levels.

Our genome scan results showed also a putative QTL influencing Lp(a) levels on chromosome 1. The Swedish and the Australian sample show MLS of 2.6 and 1.8 for Lp(a) levels, respectively. Broeckel *et al*[11] performed a genome scan for myocardial infarction and its quantitative risk factors. From this genome scan was concluded that on chromosome 1 a second QTL was located influencing Lp(a) levels. In attempt to confirm the result from Broeckel, we carried out advanced linkage analyses (Chapter 5). Simultaneous analyses of the four twin samples resulted in a MLS of 1.6 for Lp(a) levels on chromosome 1. This QTL would explain 44% of the total variance in Lp(a) levels, although the LPA gene already explained 82% of the variance. It is a common phenomenon that the QTL effect sizes are overestimated[12], resulting in more than 100% of the total trait variance explained by several QTLs found in one genome scan. Therefore, a two-locus analysis was conducted to analyse the chromosome 1 QTL together with the *LPA* locus in one model. It could be concluded that 82% of the variance in Lp(a) levels can be explained by the *LPA* locus and that the putative QTL on chromosome 1 was a false positive result. This may point at low power. Alternatively, it may indicate that two-locus analyses constitute a valuable tool to detect false positive linkage results.

Fine mapping of chromosome 19

On chromosome 19, the adult Dutch and Australian twin samples show linkage with LDL cholesterol levels with MLS of 2.2 at 54.5 cM from pter and 1.7 at 34.3 cM from pter, respectively. Also in the Swedish sample, we found positive linkage in the same chromosomal region, which makes this locus even more interesting. Previous studies reported evidence for linkage of chromosome 19 with total cholesterol levels and LDL-C levels in Pima Indians[13] and Hutterites[14], respectively. Our study could extend these findings in genetic isolates to the general, Caucasian population. However, the chromosomal region of linkage we found is very broad. To narrow down this region and also to gain power in the linkage analysis in this region, additional markers were genotyped in all twin samples, resulting in an inter-marker spacing of 6 cM in the

Dutch twin samples and 8 cM in the Swedish and Australian twin samples. MLS increased to 4.5, 1.7 and 2.1 in the adult Dutch, Swedish and Australian samples, respectively (Chapter 6). No evidence for linkage was found in the adolescent Dutch twins. After fine mapping, the linkage region is still very broad and was not narrowed down by the genotyping of additional STRs of additional family members (data not shown). From QTL mapping in *Drosophila melanogaster* and *Saccharomyces cerevisiae*, it appeared that linkage results are often caused by several polymorphisms with a small effect, which could be interpreted as one gene with a major effect[15,16]. This might explain that the position of the MLS in the Australian sample is different from that detected in the Dutch and Swedish samples. Possibly, two loci at 35 and 55 cM play a role in LDL cholesterol levels, and the former plays a major role in Australians and the latter in the European samples.

A simultaneous analysis of the three adult samples showing linkage on chromosome 19 with LDL cholesterol levels increased the MLS to 5.7 at 60 cM. In the Dutch sample, this QTL explained approximately 90% of the total variance of LDL cholesterol levels, in the Swedish 80% and in the Australian sample 40%. Hence, we located a major QTL influencing LDL cholesterol levels on chromosome 19. Since apoB is a constituent of LDL particles, apoB levels are highly correlated to LDL cholesterol levels. We performed bivariate linkage analysis of chromosome 19 with both parameters. The putative LDL cholesterol QTL on chromosome 19, indeed contributes also to the variance in apoB levels. Considering the peak-width even after fine mapping and bivariate analysis, various candidate loci may exert combined and pleiotropic effects on LDL cholesterol and apoB levels.

Previously, evidence for linkage of chromosome 19 with total cholesterol and LDL cholesterol levels were reported in Pima Indians[13] and Hutterites[14], respectively. Our study extends these findings in genetic isolates to the general, Caucasian population. Genome scans in the San Antonio Heart Study[17] and the Rochester Family Heart Study[18] were also performed in Caucasians from the general population. In Mexican Americans, linkage with LDL1 and LDL 2 cholesterol levels were found with MLS of 2.4 and 1.9[17], respectively, and in non-Hispanic whites, linkage with total cholesterol levels was found with MLS of 1.1[18]. Although the reported LOD scores for chromosome 19 were not significant in these studies, the findings are compatible with our findings. No suggestion for linkage with total cholesterol or LDL-C levels, however, was found in selected samples of myocardial infarction patients, type 2 diabetes patients and patients with familial combined hyperlipidemia[11,19-22]. The LDL-C QTL on chromosome 19 constitutes one of the most replicated result from linkage studies, virtually ruling out the possibility that it is a false positive observation. Our study in three twin samples strongly suggest that a locus influencing LDL cholesterol levels in both isolated and out bred populations is to be found.

Most obvious candidate genes that might explain the linkage on chromosome 19 are the *LDLR* gene, the *APOE/C1/C4/C2* gene cluster and *LRP3* gene. The gene product of the *LDLR* gene and the *APOE/C1/C4/C2* gene cluster are known to play a role in lipid metabolism and the *LRP3* gene is a member of the *LDLR* gene family, which leads to the speculation of a role in lipid metabolism[23]. The contribution of these loci to the positive linkage obtained could be investigated by combined association and linkage approach as described in Chapter 7. Since mutations in the LDLR gene are known to cause hypercholesterolaemia, the complete coding sequence of this gene was searched for mutations in the three sib pairs contributing most to the linkage of chromosome 19 with LDL cholesterol levels. These pairs were extreme discordant for

LDL cholesterol levels. In one sib pair a rare mutation was found, which was previously identified in an English patient with mild symptoms of FH[24]. However, the mutation was found in both members of the discordant sib pair, indicating that it is not the causative mutation. In the other two pairs, no *LDLR* mutations were found.

The contribution of the APOEε2/ε3/ε4 polymorphism to LDL cholesterol levels for example was investigated in a combined association and linkage approach as described in Chapter 7. The APOEε2/ε3/ε4 polymorphism explained part of the total variance in LDL cholesterol levels, decreasing the MLS from 5.7 to 3.5 at 60 cM from pter. So, LDL cholesterol levels are influenced by the APOEε2/ε3/ε4 polymorphism, but still other genetic variants remain to be found on chromosome 19 affecting LDL cholesterol levels (data not shown).

Combined association and linkage analysis

After the fine mapping on chromosome 19, we reanalysed linkage for apoE levels at this chromosome. The MLS in adolescent Dutch twins was still 1.0 and 0.1 in the Australian twins, but in the adult Dutch it increased to 2.6. The three twin samples could not be analysed simultaneously, since the analysis showed evidence for heterogeneity (data not shown). Therefore, only the adolescent and adult Dutch samples where analysed simultaneously, showing no evidence for heterogeneity. The MLS in the Dutch simultaneous analysis increased to 3.1 at the *APOE* locus, indicating suggestive linkage with apoE levels.

We then investigated what proportion of the linkage result of the apoE levels could be explained by the *APOE*ε2/ε3/ε4 polymorphism. A combined linkage-association approach was used. The combined association and linkage study in the two Dutch samples, showed that the linkage in the adolescent Dutch twins could entirely be explained by the APOEε2/ε3/ε4 polymorphism, indicating that probably no other polymorphisms on chromosome 19 influences apoE levels in adolescent Dutch twins. Only polymorphisms in complete linkage disequilibrium (LD) with this APOEε2/ε3/ε4 polymorphism can not be excluded to play a role in plasma levels of apoE. In the adult Dutch twins, however, only part of linkage could be explained by the APOEε2/ε3/ε4 polymorphism, indicating that other genetic variation in the QTL region influences apoE levels. Other polymorphisms that might play a role in determining apoE levels are, for example, the functional *APOE* promoter variants -291G/T and -491A/T[25].

Although the Australian twins did not show any linkage, association of the APOEε2/ε3/ε4 polymorphism with apoE plasma levels could be detected. This illustrates well that association studies have more power than linkage studies. However, this polymorphism explained merely 5% of the variance in Australian apoE levels, which is much less than known from literature. These results in the Australian sample, together with the much lower heritability of apoE levels (57% vs 86-87% Dutch samples), might suggest that the apoE measurements may not be accurate.

Additionally, the finding that the *APOE*ε2/ε3/ε4 polymorphism is the only variant in the APOE gene influencing apoE levels in adolescents, not taking into account variants in complete LD, is different from the findings in adults. Thus, the genetic architecture of adolescent apoE levels seems to differ from that of adult levels. This is consistent with a previous quantitative genetic studies, describing that some lipid and apolipoprotein levels are influenced by partly different genes in adolescence and adulthood[26], which might also be the case for apoE levels.

Conclusion and Future perspectives

The studies in this thesis indicate that lipid and apolipoprotein levels are highly heritable in Caucasian samples and that thus the genes involved in these traits have to be identified. We were able to localise several putative QTLs for lipid and apolipoprotein levels and we identified one major LDL cholesterol QTL on chromosome 19, probably also influencing apoB levels. It is expected that in our approach, QTLs with smaller effect have been missed. Since high throughput genotypings are now feasible, the remaining chromosomes will be scanned with a 10 cM inter-marker distance. In the fine mapping of positive linkage regions, we have applied several statistic tools of which the combined association and linkage approach seems to be the most valuable tool. Using this approach, genes might be identified to have their influence on quantitative parameters of the human lipid metabolism.

References

1. Khot UN, Khot MB, Bajzer CT, Sapp SK, Ohman EM *et al.* Prevalence of conventional risk factors in patients with coronary heart disease. *JAMA* 2003; 290: 898-904.
2. Greenland P, Knoll MD, Stamler J, Neaton JD, Dyer AR *et al.* Major risk factors as antecedents of fatal and nonfatal coronary heart disease events. *JAMA* 2003; 290: 891-897.
3. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; 11: 241-247.
4. Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M. Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 2001; 69: 936-950.
5. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. GRR: graphical representation of relationship errors. *Bioinformatics* 2001; 17: 742-743.
6. Douglas JA, Boehnke M, Lange K. A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. *Am J Hum Genet* 2000; 66: 1287-1297.
7. Larson IA, Ordovas JM, DeLuca C, Barnard JR, Feussner G, Schaefer EJ. Association of apolipoprotein (Apo)E genotype with plasma apo E levels. *Atherosclerosis* 2000; 148: 327-335.
8. Lambert JC, Brousseau T, Defosse V, Evans A, Arveiler D *et al.* Independent association of an APOE gene promoter polymorphism with increased risk of myocardial infarction and decreased APOE plasma concentrations-the ECTIM study. *Hum Mol Genet* 2000; 9: 57-61.
9. Kaprio J, Ferrell RE, Kottke BA, Kamboh MI, Sing CF. Effects of polymorphisms in apolipoproteins E, A-IV, and H on quantitative traits related to risk for cardiovascular disease. *Arterioscler Thromb* 1991; 11: 1330-1348.
10. Wiltshire S, Cardon LR, McCarthy MI. Evaluating the results of genomewide linkage scans of complex traits by locus counting. *Am J Hum Genet* 2002; 71: 1175-1182.
11. Broeckel U, Hengstenberg C, Mayer B, Holmer S, Martin LJ *et al.* A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat Genet* 2002; 30: 210-214.
12. Göring HH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-specific effects from genomewide scans. *Am J Hum Genet* 2001; 69: 1357-1369.
13. Imperatore G, Knowler WC, Pettitt DJ, Kobes S, Fuller JH *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.
14. Ober C, Abney M, McPeek MS. The genetic dissection of complex traits in a founder population. *Am J Hum Genet* 2001; 69: 1068-1079.
15. Barton NH, Keightley PD. Understanding quantitative genetic variation. *Nat Rev Genet* 2002; 3: 11-21.
16. Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* 2002; 416: 326-330.
17. Rainwater DL, Almasy L, Blangero J, Cole SA, VandeBerg JL *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.
18. Klos KL, Kardia SL, Ferrell RE, Turner ST, Boerwinkle E, Sing CF. Genome-wide linkage analysis reveals evidence of multiple regions that influence variation in plasma lipid and apolipoprotein levels associated with risk of coronary heart disease. *Arterioscler Thromb Vasc Biol* 2001; 21: 971-978.
19. Reed DR, Nanthakumar E, North M, Bell C, Price RA. A genome-wide scan suggests a locus on chromosome 1q21-q23 contributes to normal variation in plasma cholesterol concentration. *J Mol Med* 2001; 79: 262-269.
20. Aouizerat BE, Allayee H, Cantor RM, Davis RC, Lanning CD *et al.* A genome scan for familial combined hyperlipidemia reveals evidence of linkage with a locus on chromosome 11. *Am J Hum Genet* 1999; 65: 397-412.

21. Francke S, Manraj M, Lacquemant C, Lecoeur C, Lepretre F *et al*. A genome-wide scan for coronary heart disease suggests in Indo-Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. *Hum Mol Genet* 2001; 10: 2751-2765.
22. Elbein SC, Hasstedt SJ. Quantitative trait linkage analysis of lipid-related traits in familial type 2 diabetes: evidence for linkage of triglyceride levels to chromosome 19q. *Diabetes* 2002; 51: 528-535.
23. Ishii H, Kim DH, Fujita T, Endo Y, Saeki S, Yamamoto TT. cDNA cloning of a new low-density lipoprotein receptor-related protein and mapping of its gene (LRP3) to chromosome bands 19q12-q13. 2. *Genomics* 1998; 51: 132-135.
24. Sun XM, Patel DD, Knight BL, Soutar AK. Comparison of the genetic defect with LDL-receptor activity in cultured cells from patients with a clinical diagnosis of heterozygous familial hypercholesterolemia. The Familial Hypercholesterolaemia Regression Study Group. *Arterioscler Thromb Vasc Biol* 1997; 17: 3092-3101.
25. Stengard JH, Clark AG, Weiss KM, Kardia S, Nickerson DA *et al*. Contributions of 18 additional DNA sequence variations in the gene encoding apolipoprotein e to explaining variation in quantitative measures of lipid metabolism. *Am J Hum Genet* 2002; 71: 501-517.
26. Snieder H, van Doornen LJ, Boomsma DI. The age dependency of gene expression for plasma lipids, lipoproteins, and apolipoproteins. *Am J Hum Genet* 1997; 60: 638-650.