Association of single nucleotide polymorphisms in the *LPA* gene region with serum Lp(a) levels and myocardial infarction

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.) DER NATURWISSENSCHAFTLICHEN FAKULTÄT III - BIOLOGIE UND VORKLINISCHE MEDIZIN - DER UNIVERSITÄT REGENSBURG

vorgelegt von

Katharina Neureuther

aus Regensburg

Februar 2008

Association of single nucleotide polymorphisms in the *LPA* gene region with serum Lp(a) levels and myocardial infarction

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.) der naturwissenschaftlichen Fakultät III - Biologie und vorklinische Medizin - der Universität Regensburg

vorgelegt von
Katharina Neureuther

aus Regensburg
Februar 2008

Promotionsgesuch eingereicht am: 06.02.2008
Die Arbeit wurde angeleitet von: PD Dr. Thomas Langman

Prüfungsauschuss:

Vorsitzender:	Prof. Dr. Stephan Schneuwly
Erstgutachter:	PD Dr. Thomas Langmann
Zweitgutachter:	Prof. Dr. med. Christian Hengstenberg
Dritter Prüfer:	Prof. Dr. Gernot Längst

Quidquid agis prudenter agas et respice finem.

List of abbreviations

AA	amino acid
apo(a)	apolipoprotein (a)
asPCR	allele-specific PCR
BMI	body mass index
bp	base pair
CAD	coronary artery disease
CHD	coronary heart disease
CI	confidence interval
cM	centiMorgan
Da	Dalton
dbSNP	public single nucleotide database
EDTA	Ethylendiamintetraacetate
FAM	6-carboxyfluoresceine
FBAT	Family-Based Association Tests in genetic analyses
GWA	genome-wide association
HDL	high density lipoprotein
HW	Hardy-Weinberg
KIV-2	kringle IV-like type 2
KV	kringle V-like
LD	linkage disequilibrium
LDL	low density lipoprotein
LOD	logarithm of the odds
MAF	minor allele frequency
MGB	minor groove binder
MI	myocardial infarction
NCBI	National Center for Biotechnology Information
OD	optical density
ON	oligonucleotide, primer
OR	odds ratio
PCR	polymerase chain reaction
QTL	quantitative trait locus
RefSeq	reference sequence
SD	standard deviation
SNP	single (or: simple) nucleotide polymorphism
Taq	*Thermus Aquaticus*
TDT	Transmission Disequilibrium Test
Tm	melting temperature
TRIS	Tris(hydroxymethyl)aminomethane
UCSC	University of California Santa Cruz
VIC	a proprietary fluorescent dye produced by Applied Biosystems
WHO	World Health Organization

List of Figures

1.1 Structural composition of the Lp(a) particle 2
1.2 Kringle structure 2
1.3 Structures of apo(a) 3

2.1 Principle of allele-specific PCR 18
2.2 Amplification plot 19
2.3 Distribution of SNPs in the *LPA* gene region investigated by asPCR on DNA pools 20
2.4 5' nuclease assay 24
2.5 Allelic discrimination 25
2.6 Cluster plot 25
2.7 GeneChip Human Mapping 500K Array Set from Affymetrix 27

3.1 Schematic representation of the *LPA* gene structure and its LD pattern 36
3.2 Comparison of mean Lp(a) levels between control and MI/CAD study populations and influence of KIV-2 repeat number on Lp(a) levels in each study population 44
3.3 Interaction between KIV-2 repeat number and rs11751605 genotype and the influence on Lp(a) levels 45
3.4 Odds ratios from logistic regression analysis for MI and CAD according to rs11751605 genotypes 46
3.5 Results from the 500K genome-wide association with square-root transformed Lp(a) levels using PLINK 49
3.6 Results of association for markers of the *LPA* gene region from the 500K chip with square-root transformed Lp(a) levels using PLINK . . 50
3.7 Distribution of tagging SNPs 51
3.8 Results for chromosome 6 from genome-wide association of Lp(a) levels, adjusted for *LPA* tagging SNPs 53

3.9 Results from genome-wide association of Lp(a) levels adjusted for *LPA* tagging SNPs . . . 54

List of Tables

2.1 Oligonucleotides for allele-specific PCR 21
2.2 Standard asPCR reaction 22
2.3 Thermocycling conditions of asPCR runs on DNA pools 22
2.4 SNP characteristics 26

3.1 Results from asPCR on contributor and non-contributor pools 35
3.2 SNP characteristics 37
3.3 Clinical characteristics of the family sample 38
3.4 Distribution of genotypes in the family sample 38
3.5 Results from family-based association testing with square-root transformed Lp(a) levels as a quantitative trait 39
3.6 Lp(a) QTL linkage analysis on chromosome 6q27 with SNP rs11751605 and/or KIV-2 included as covariates 40
3.7 Distribution of genotypes in case-control samples 41
3.8 Clinical characteristics of case-control study populations 42
3.9 SNP genotypes and the corresponding median Lp(a) levels (mg/dl) with 25% to 75% quantile range in all populations 43
3.10 Clinical characteristics of the Cardiogenics sub-sample 48
3.11 List of tagging SNPs 52

Contents

1 Introduction **1**

1.1 Structures of the Lp(a) particle . . . 1

1.1.1 The LDL component . . . 1

1.1.2 The apo(a) component . . . 1

1.2 Structure-function relationships in Lp(a) components . . . 3

1.3 Function of Lp(a) . . . 5

1.4 Evolution of Lp(a) . . . 6

1.5 Elevated Lp(a) levels as a cardiovascular risk factor . . . 6

1.6 Metabolism of Lp(a) . . . 7

1.7 Heritability of Lp(a) levels . . . 8

1.8 Genetic polymorphisms of the *LPA* gene . . . 9

1.8.1 Repeat polymorphisms . . . 9

1.8.2 Single nucleotide polymorphisms . . . 10

1.9 Aim of this investigation . . . 12

2 Materials and methods **13**

2.1 DNA processing . . . 13

2.1.1 Isolating genomic DNA from blood . . . 13

2.1.2 DNA preparation for pooled DNA samples . . . 14

2.1.3 DNA preparation for individual genotyping . . . 15

2.2 Allele-specific PCR on pooled DNA . . . 16

2.2.1 Quantitative real-time PCR . . . 16

2.2.2 Allele-specific PCR . . . 17

2.2.3 Selection of SNP markers for allele-specific PCR . . . 19

2.2.4 Selection and pooling of DNA samples . . . 20

2.2.5 asPCR procedures . . . 22

2.2.6 Analyzing the PCR data . . . 22

2.3 Genotyping with TaqMan Assays . . . 23

2.3.1 The 5’ nuclease assay . . . 23
2.3.2 SNP selection and genotyping of individual DNA samples . . . 26
2.3.3 Standard TaqMan Assay . . . 26
2.4 Genotyping with microarrays for genome-wide association . . . 27
2.5 Sample selection . . . 28
2.5.1 Measurement of Lp(a) levels . . . 30
2.5.2 Phenotypic apo(a) information and information on the pentanucleotide repeat polymorphism . . . 30
2.5.3 Definitions . . . 30
2.6 Statistical analysis . . . 31
2.6.1 Family-based association testing . . . 31
2.6.2 Case-control studies . . . 32
2.6.3 Genome-wide association testing . . . 32

3 Results 34

3.1 Allele-specific PCR on pooled DNA samples . . . 34
3.1.1 SNP selection and genotyping of individual DNA samples . . . 34
3.2 Results from individual genotyping in the MI family study . . . 37
3.2.1 Characteristics of the family sample . . . 37
3.2.2 Family-based association testing . . . 38
3.2.3 Contribution to the evidence of linkage . . . 39
3.3 Case-control association samples . . . 40
3.3.1 Genotyping in the population-based sample . . . 40
3.3.2 Individual genotyping in the KORA MI sample . . . 41
3.3.3 Genotype distribution in the subsamples from the family sample for association testing . . . 41
3.3.4 Clinical characteristics of the case-control study populations . 42
3.3.5 Association of individual SNP markers with Lp(a) levels . . . 42
3.3.6 Interaction between SNP markers and KIV-2 repeats on Lp(a) levels . . . 44
3.3.7 Association of SNP markers with MI . . . 45
3.4 Linkage disequilibrium analysis and Hardy-Weinberg proportions . . . 47
3.5 Genome-wide association testing . . . 47
3.5.1 Characteristics of the subsample with GWA data . . . 47
3.5.2 Quantitative association analysis . . . 48

3.5.3 Results for the SNPs from the *LPA* gene region 49
3.5.4 Adjustment of the GWA data for the significant SNPs from the *LPA* gene region 50
3.5.5 Regions of interest for further association analysis 53

4 Discussion 55
4.1 Association of single SNP markers with MI and Lp(a) levels 55
4.1.1 Influence of previously described rare SNP markers 56
4.1.2 Possible influences of rs11751605 on functional properties of Lp(a) 59
4.2 Genome-wide association data 60

5 Summary 63

Bibliography 75

1 Introduction

In 1963, Berg first reported the identification of lipoprotein (a) [Lp(a)] in human plasma as an antigen associated with low-density lipoprotein (LDL) [1]. He also performed a family study and reported in the same year an autosomal genetic inheritance of the presence of Lp(a) [2]. Further investigation revealed that Lp(a) was a macromolecule, comprised of LDL and a glycoprotein component [3]. In the 1970s the first studies suggested an association of plasma Lp(a) levels with atherosclerotic vascular disease [4, 5]. Though intense research on Lp(a) for decades, many aspects on the function, regulation of plasma concentration and genetics of Lp(a) still have to be revealed [6].

1.1 Structures of the Lp(a) particle

The Lp(a) particle is made up by a LDL particle and the glycoprotein component apolipoprotein a [apo(a)] [3]. These two structures are linked covalently by a single disulfide bond and are further stabilized by non-covalent interactions [3]. Figure 1.1 shows the structural composition of the Lp(a) particle.

1.1.1 The LDL component

Cholesteryl esters and triglycerides build up the lipid core of the LDL component [7]. The lipid core is surrounded by a monolayer of unesterified cholesterol, phospholipids and, as the protein component, one apolipoprotein B-100 (apo B-100) per particle, as shown in figure 1.1 [7]. The apo(a) component surrounds the particle and is linked to the apoB-100 molecule [7].

1.1.2 The apo(a) component

The apo(a) protein component, encoded by the gene *LPA*, has a remarkable structure [8]. Several cysteine residues lead to the formation of intramolecular disulfide bonds that give the molecule the look of nodules in the display of the two-dimensional

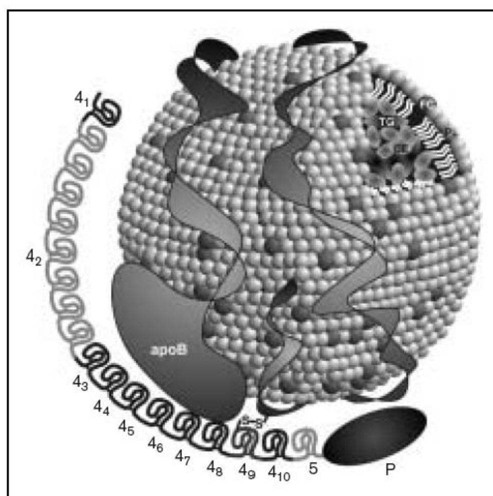

Figure 1.1: Structural composition of the Lp(a) particle, as described in the text. Figure from [7].

amino acid chain [9]. These nodules, better known as kringles, are the structural specialty of apo(a). Apo(a) has a strong homology to plasminogen, a proenzyme of the fibrinolytic system [10]. Apo(a) and plasminogen contain both a series of motifs called kringles [10]. The plasminogen gene contains coding sequences for 5 different kringle domains, kringle I to kringle V [11]. The kringle structures of apo(a) resemble some of the kringle structures of plasminogen, and are therefore divided into kringle IV- like and kringle V-like [12]. Figure 1.2 displays one of the kringles of apo(a).

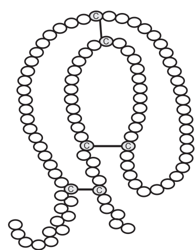

Figure 1.2: Amino acids of a kringle structure, organized into three loops by disulfide bonds between the cysteine residues (C), modified after [13].

Each kringle contains about 77 to 79 amino acids, the kringles are connected by linker regions of 26 to 36 amino acids [14, 10]. The kringles are each coded by two separate exons with introns inserted at homologous positions as in the plasminogen gene (in the middle and at both ends of each kringle) [15]. The number of KIV-2 repeats is encoded by corresponding numbers of 5.5 kb DNA units, each consisting of two exons (of 342 bp length in total), separated by a 4 kb intron [14]. The KIV-2 structures are separated by 1.2 kb introns [14].

Each apo(a) molecule contains only one kringle V-like structure, but several variants of kringle IV-like [16, 14], as depicted in figure 1.3.

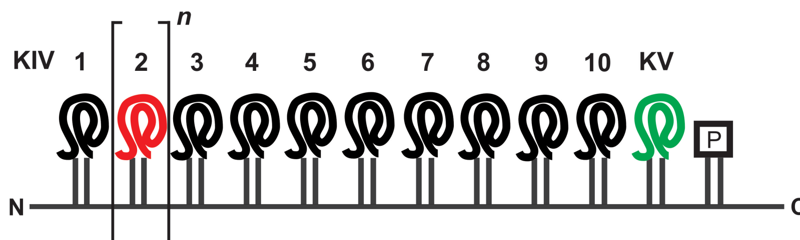

Figure 1.3: Apolipoprotein (a) is made of 10 different types of kringle IV (KIV), followed by a kringle V (KV) and a nonfunctional protease domain (P). Kringle IV type 2 (KIV-2) is variable in its repeat number (indicated by n). Modified after [17].

One of these kringle IV-like stuctures, kringle IV type 2 (KIV-2) can be present in several repeats [16, 14]. The number of KIV-2 repeats is highly variable, 3 to more than 40 copies have been found, therefore the molecular mass of apo(a) is highly polymorphic, with a range between 300 and 800 kDa [16, 14]. The other kringle IV like domains (KIV-1 and KIV-3 to KIV-10) are present in single copy [18, 16, 14].

Apo(a) is a glycoprotein, and sites for N-linked glycosylation are present within the core of each KIV motif, whereas a minimum of six O-linked glycosylation sites are present within the linker sequences that join individual kringles [10, 19, 20]. The hydrophilic O-linked glycans are mainly composed of monosialylated core structures [21]. Apo(a) and thus, Lp(a), is heterogeneous in its glycosylation, but the influence on the *in vivo* properties is still unclear [22].

1.2 Structure-function relationships in Lp(a) components

Epidemiological studies identified elevated serum Lp(a) concentrations as an independent risk factor for coronary heart disease (CHD) (see section 1.5). The high degree of homology between apo(a) and plasminogen has evoked speculations about a role of Lp(a) in thrombosis, while the LDL-like component of Lp(a) proposed a possible role in atherosclerosis [10]. Several relationships between structure and function of apo(a) have been investigated, and specific functions have been mapped to discrete structural units in apo(a).

According to a concept based on homology between apo(a) and plasminogen,

Lp(a) would interfere with the process of plasmin generation and, in consequence, lead to decreased fibrinolysis and increased accumulation of fibrin in the atherosclerotic plaque.

But, unlike plasminogen, the protease-like domain of apo(a) can not be activated by tissue plasminogen activator, urokinase or streptokinase and is catalytically inactive [23, 24]. Therefore, it has been speculated that Lp(a) might act as a competitive inhibitor for plasminogen activation and inhibit plasmin-mediated thrombus degradation by binding to fibrin. The first investigations about Lp(a) aimed at this possible role in thrombogenesis and showed interference with several steps of the fibrinolytic pathway [23, 25].

The different kinds of kringle IV-like structures have been investigated and different functions have been revealed mostly by *in vitro* experiments [26, 27].

KIV-2, as already mentioned, is responsible for the isoform size heterogeneity and the number of the *LPA* KIV-2 repeats determines about 50% of the variability of Lp(a) levels with an inverse correlation between number of KIV-2 repeats and Lp(a) levels [28, 29, 30]. One explanation for this inverse correlation was reduced secretion efficiency of larger apo(a) isoforms, leading to their increased intracellular degradation [31]. Other investigations have shown that the number of KIV-2 has an influence on the conformational status of the apo(a) component, and therefore influences the efficiency of covalent Lp(a) assembly [32, 33]. It has been shown that shorter apo(a) isoforms tend to stay in a open conformation, which favors efficient Lp(a) assembly, whereas in longer, more flexible, apo(a) isoforms intramolecular interaction can lead to a closed conformation that hinders efficient Lp(a) assembly [32].

Lp(a) is cleaved *in vivo* between the KIV-4 and the KIV-5 domain by elastases, dividing apo(a) in two parts, with the C-terminal domain spanning from KIV-5 to the protease domain [26]. Recent studies in transgenic mice indicate that this C-terminal apo(a) component mediates Lp(a) clearance from plasma by the liver [34].

The domains KIV-5 to KIV-8 have weak lysine binding sites, whereas KIV-10 has a strong lysine binding site [27]. These lysine binding sites have been shown to mediate the non-covalent step of Lp(a) assembly [35].

The part from KIV-5 to the protease domain at the C-terminus, might contribute to the preferential retention of Lp(a) in atherosclerotic lesions. Binding to laminin and other components of the extracellular matrix has been shown for these parts of

apo(a) [36, 37].

KIV-6 and KIV-7 are the domains that are recognized by the foam cell receptor, and Lp(a) may participate in macrophage foam cell formation [38].

In the KIV-9 domain, an unpaired cysteine residue mediates the formation of the single disulfide bond between the apo(a) and the apoB-100 component of LDL [39]. Therefore, KIV-9 domain is essential for the formation of stable Lp(a) particles [40].

The formation of a Lp(a) particle is a two step process; initial non-covalent interactions between apo(a) and apoB-100 precede the formation of a disulfide bond [41]. It has been shown that the conformation of the apo(a) molecule is important for efficient assembly of the Lp(a) particle [32, 33]. A closed as well as an open conformation status of apo(a) have been identified, and can *in vitro* be changed by lysine analogues [9]. Furthermore the apo(a) molecule has one KV like domain, as already mentioned, and a protease domain.

Edelstein et al. found that lysine residues of the KV domain of apo(a) mediate binding with oxidized phosphatidylcholine and postulated that apo(a) might act as a scavenger and prevent LDL from oxidation [13].

Recent findings by Tsimikas et al. showed that Lp(a) levels strongly correlate with the oxidized phospholipid/apoB-100 ratio in plasma and that both factors are associated with the extend of angiographically documented coronary artery disease [42]. In conjunction with the finding that the oxidized phospholipids are physically associated with Lp(a) via the KV domain of apo(a) [13], this supports the idea that in a setting of enhanced oxidative stress the atherogenic properties of Lp(a) might be enhanced through pro-inflammatory oxidized phospholipids.

1.3 Function of Lp(a)

The physiological role of the complex macromolecule Lp(a) in the organism is still unknown. Individuals with near absence of Lp(a) in plasma have been identified and characterized, but no deficiency syndrome or any kind of disease was found [40]. A pathophysiological role of elevated Lp(a) levels in the context of atherosclerotic and cardiovascular disease seems quite clear and many possible pathological functions of Lp(a) have been proposed, as mentioned above. However, a concrete pathological mechanism in human disease has not yet been shown.

1.4 Evolution of Lp(a)

Apo(a) protein (and the Lp(a) particle) has been detected in several species of nonhuman primates and Old World monkeys (as chimpanzee, orangutan, gorilla, rhesus monkey and baboon), although most mammals lack apo(a) [43, 44, 45, 46]. In most of these species it has been shown that the protein is polymorphic [45].

Apo(a) belongs to an extended gene superfamily that contains domains that resemble the serine protease trypsin [47]. One branch of the family tree that contains only kringle and protease-like domains includes plasminogen, apo(a) and hepatocyte growth factor [47]. *LPA* evolved from a duplicated plasminogen gene during recent primate evolution [10].

The *LPA* gene is located on human chromosome 6q26-27 in close proximity to the plasminogen gene from which it evolved during primate evolution by duplication, deletions, gene conversions and mutations [10, 48, 43].

Furthermore Lp(a) is also found in the European hedgehog. However, the hedgehog version of apo(a) appears to have evolved separately from the human version, as it shows strong differences in composition to primate apo(a) [44]. The hedgehog apo(a) lacks sequences that resemble the K-IV, K-V and protease domain, but instead is composed of tandem repeats of sequences homolog to the K-III domain of plasminogen [44].

Kraft et al. suggested that Lp(a) is an old African trait, as they found the highest Lp(a) concentrations in Khoi San from southern Africa, one of the oldest living populations [49]. The distribution of Lp(a) levels in different ethnic groups (Khoi San, South African Blacks, Hong Kong Chinese and Caucasians) was found to be heterogeneous, with a broad distribution in Africans and a highly skewed distribution (towards lower levels) in Caucasians [49].

1.5 Elevated Lp(a) levels as a cardiovascular risk factor

Several studies on Lp(a) and its role in cardiovascular disease have been performed. The majority of studies performed on the on the impact of elevated Lp(a) levels on the development of cardiovascular disease provided evidence for Lp(a) as a independent predictor of myocardial infarction (MI) [50, 51, 52], coronary artery disease (CAD) [53, 54, 55], and extent of the disease process [56, 57, 58, 59]. Some stud-

ies have shown weak [60] or no association with various cardiovascular phenotypes [61, 62, 63, 64, 65, 66]. A meta-analysis of 27 prospective studies has shown a clear association between Lp(a) and CHD including information from 5,436 CHD cases that were observed during a mean follow-up time of 10 years [55]. Risk of CHD was at 70% increased in people from the general population with Lp(a) levels in the top third of baseline measurement compared to those in the bottom third [55]. Other studies showed that high Lp(a) levels potentiate the risk rising from high LDL, it has therefore been discussed controversly, weather Lp(a) is a independent risk factor for CHD [67, 68, 69, 57]. However, from the insights in structure-function relationships of Lp(a) is has become clear that a complex macromolecule as Lp(a) can have a influence in multiple stages of the atherosclerotic process through several different mechanisms.

The progress of atherosclerosis can finally lead to the complication of myocardial infarction, one of the most common causes of death in Western societies [70].

Epidemiologic studies have identified risk factors for the development of atherosclerosis [70]. Elevated levels of LDL, low levels of HDL, hypertension, diabetes mellitus, male gender, elevated levels of homocysteine, metabolic syndrome, insulin resistance, obesity and family history of cardiovascular disease have been identified apart from elevated Lp(a) levels as factors contributing to the development of atherosclerosis [70]. Also environmental factors as smoking, lack of exercise and high fat diet can promote atherosclerosis [70].

While many risk factors for atherosclerosis can be influenced by medication (for example hypertension or high LDL cholesterol) or changes in lifestyle (smoking, obesity), no applicable therapy is available for elevated Lp(a) levels [70, 22].

1.6 Metabolism of Lp(a)

Lp(a) levels may vary 1000-fold between individuals, but they exhibit a high degree of stability within persons [71]. Correlation between two measurements taken some years apart in the same individual is 0.9 (self-correlation) [72]. This high inter-individual variability is largely determined by the *LPA* gene on chromosome 6q27 and several polymorphisms of *LPA* have been shown to influence the size and amount of apo(a), as well as Lp(a) plasma levels [8, 6, 73, 28, 29].

It has been shown that the subunits of Lp(a), apo(a) and apoB-100 containing lipoproteins are independently processed and secreted by liver cells [74, 75, 76]. It

is suggested that the Lp(a) particle forms in the extracellular compartment. Very recently, a study has provided evidence that the disulfide bond formation between apo(a) and apoB-100 can be markedly enhanced by a putative Lp(a) oxidase activity secreted by a human liver cell line [77]. The identification of a putative, extracellular Lp(a) oxidase could provide a new modulator of Lp(a) levels.

Lp(a) concentrations have been found to be determined at the level of synthesis rather than catabolism [78].

The pathway of Lp(a) catabolism *in vivo* remains unclear, and no major site of catabolism of Lp(a) particles has been identified in humans [22]. Experiments in transgenic mice indicate that Lp(a) is cleared from the plasma primarily by the liver in a process mediated by apo(a) [34]. This apo(a) clearance in transgenic mice has been shown to be independent from LDL receptor, apoE or the asialoglycoprotein receptor and is supposed to run via a receptor that is responsible for the hepatic clearance of cholesterol-rich remnant lipoproteins (RLP) [34, 79].

Lp(a) concentrations are relatively resistant to alteration by traditional pharmacologic and non-pharmacologic approaches [22]. In contrast to plasma LDL concentrations, Lp(a) concentrations are thought to be relatively resistent to diet or exercise [22]. Few drugs have been reported to be capable of reducing Lp(a) concentrations *in vivo*, as niacin in high doses, aspirin and others [80, 22]. For the influence of statin therapy an Lp(a) concentrations, results are unclear, a modest reduction as well as a modest increase of Lp(a) concentrations were reported for different statins [22]. The most efficient method to decrease Lp(a) concentrations is LDL or Lp(a) apheresis. This costly procedures can decrease Lp(a) concentrations by 50% [22].

1.7 Heritability of Lp(a) levels

In fact, the variation of the Lp(a) plasma levels is under strong genetic influence with a heritability of >90% [81, 82]. Twin, family and sib-pair linkage studies have revealed that the intra-individual variability in plasma concentration is under strong genetic control and almost entirely explained by variations at the *LPA* gene locus [83].

1.8 Genetic polymorphisms of the *LPA* gene

Several genetic polymorphisms of the *LPA* gene have been described, including single nucleotide polymorphisms and different kind of repeat polymorphisms.

1.8.1 Repeat polymorphisms

As explained above, the most prominent polymorphism of the protein is the variable number of KIV-2 repeats. An other repeat polymorphism is located in the promotor region of gene.

Variable number of KIV-2 repeats

The number of the *LPA* KIV-2 repeats appears to determine about 50% of the variability of Lp(a) levels and the number of KIV-2 repeats are negatively correlated with Lp(a) levels [28, 29, 30]. Though the number of KIV-2 repeats has a strong influence on Lp(a) levels in most populations, the variance in Lp(a) levels between individuals with alleles for same number of KIV-2 repeats still differs by a factor of 200 [49, 84]. The heterozygosity for the apo(a) KIV-2 alleles is high, approximately 94% in Caucasians [49]. Most studies investigating the relation between the number of KIV-2 repeats and Lp(a) levels, used the method of apo(a) phenotyping by immunoblotting of serum, therefore providing information about the number of expressed KIV-2 repeats. Some studies have also investigated in parallel the number of repeats on gene level, and revealed that so called "null" -alleles exist [85]. These not expressed apo(a) alleles were found to be distributed over the whole range of KIV-2 repeat number in different populations [49]. In Caucasians, the relative frequency of non-expressed alleles increased with KIV-2 repeat number [49]. "Null" -alleles were found to be more frequent in Caucasians (18.9%) than in other populations [49].

Variable number of pentanucleotide repeats

A repeat polymorphism of the pentanucleotide sequence TTTTA was found at -1371 upstream of the apo(a) translational start site of *LPA* [47]. This pentanucleotide repeat (PNR) polymorphism in the promotor region may account for 3-14 % of the inter-individual variations in plasma concentrations in Caucasians [50, 29]. The PNR polymorphism, usually varies from 8 to 11 repeats, depending on the population and has been associated with Lp(a) levels repeatedly in different populations [56, 29, 86,

87, 88]. Lower numbers of PNR repeats have been associated with higher Lp(a) levels in Caucasians [50].

1.8.2 Single nucleotide polymorphisms

Other polymorphisms of the *LPA* gene have been identified and revealed small effects on Lp(a) plasma concentrations, however, in total, they do not fully explain the genetic variation of Lp(a) levels [87, 89, 90, 91].

SNP markers in the 5-prime region

Several SNPs in the 5-prime flanking region of the *LPA* gene have been investigated: Brazier et al. investigated three SNPs (-914 G/A, -49 C/T and -21 G/A) in the promotor region of the gene and their association with Lp(a) levels and MI in a French and Irish population [50]. They genotyped 594 MI patients, and 683 controls, and found no association of the SNPs with MI or Lp(a) levels, but revealed a strong LD between these SNP markers and the PNR polymorphism [50].

Ichinose et al. identified three polymorphisms in the *LPA* 5-prime flanking region: -772 A/G, +93 C/T and +121 G/A, relative to the transcription start site [92].

Suzuki et al. investigated the influence of these three SNPs on gene expression [87]. Deletion analysis in an *in vitro* experiment revealed that a T allele at position +93 led to negative regulation in gene expression and the A allele at position +121 led to positive regulation of gene expression [87]. They also measured Lp(a) levels in Japanese individuals and found significantly higher Lp(a) level in homozygous carriers of the +93 C allele and the +121 A allele, compared with homozygous carriers of the +93 T allele and the +121 G allele [87].

Kraft et al. showed that the +93 C/T polymorphism has no impact no Lp(a) concentrations in Caucasians, as this SNP is in strong linkage disequilibrium with the PNR polymorphism [89].

For Caucasian populations, the SNPs in the 5-prime flanking region of *LPA*, have been shown to be in linkage disequilibrium (LD) with the PNR [50]. In summary, for none of the SNP markers in the 5-prime region an independent influence on Lp(a) levels could be shown in Caucasian populations [50].

SNP markers in exon regions

The other SNP markers that have been investigated so far are mainly located in exon regions of the *LPA* gene. Ogorelkova et al. identified 14 exon SNPs in apo(a) KIV-6, KIV-8, KIV-9, and KIV-10, but found no sequence variants common to Africans and Caucasians, when they investigated healthy individuals [91]. A substitution in KIV-6 and another in KIV-8 were associated with Lp(a) levels significantly below average in Africans [91]. In contrast, a substitution in KIV-9, which occurred with a frequency of 8% in Khoi San Africans, resulted in a significantly increased Lp(a) concentration [91]. For the SNPs that were found only in Caucasians, no association with Lp(a) levels could be detected, as the SNPs were to rare for statistical analysis [91].

Recently, Luke et al. investigated a population of Americans with European descent and reported an association between Lp(a) levels as well as CAD and an exonic SNP, rs3798220, located in the protease domain of *LPA* [93]. This is the first described SNP marker of *LPA* significantly associated with MI, independent from Lp(a) levels [93].

SNP markers and not expressed apo(a) isoforms

Two SNPs have been reported to be associated with not expressed apo(a) isoforms, also named “null”-alleles.

Ogorelkova et al. described a G to A substitution at the +1 donor splice site of the apo(a) KIV-8 intron that results in alternative splicing of the intron and thus encodes a truncated apo(a) form [40]. This SNP marker was identified in a Caucasian sample (n=239) with an allele frequency of 6% [40].

Parson et al. described a C/T SNP at nucleotide 61 in exon 1 of the KIV-2 domain, with this nonsense mutation producing a stop codon (R21X) the resulting protein is predicted to be truncated [90]. An allele frequency of 0.02 was detected for this SNP [90].

These described SNP markers are rare and, therefore, account only for a small proportion of number of not expressed apo(a) isoforms.

1.9 Aim of this investigation

In this investigation different methods were used to explore the influence of additional SNP markers located in the *LPA* gene region on the Lp(a) levels and risk for myocardial infarction. With a pooling approach using information from a previous linkage analysis, relevant SNP markers should be identified. Replication of these relevant SNP markers in a large sample with a different technique and a detailed analysis of the association should confirm the findings. The replication of findings in different population samples should provide further evidence for a robust finding, and allow a more generalized statement.

2 Materials and methods

Phenotypic differences between individuals of a species are mostly caused by genetic variants [94]. In the human genome, millions of polymorphisms have been identified so far, the vast majority are changes of a single nucleotide, known as single nucleotide polymorphism (SNPs) [94]. The Human Genome Project enhanced the identification of DNA variations, and different methods of typing DNA variations (genotyping) were developed [94]. Two different PCR-based techniques for the purpose of genotyping were performed in this work and data resulting from a third technique were analyzed. The first technique of allele-specific PCR was used to analyze the differences in allele frequency between two samples. The second technique of TaqMan genotyping allowed the exact determination of genotypes for a large set of individual samples. Through an additional technique, the Affymetrix GeneChip Human Mapping 500K Array Set, a huge set of SNP markers for a large set of individual samples was obtained.

2.1 DNA processing

2.1.1 Isolating genomic DNA from blood

DNA from all participants of the German MI Family Study was extracted from EDTA blood samples (peripheral blood) in the laboratory at the University of Regensburg. For this purpose, a commercially available PureGene Genomic DNA Purification Kit (Gentra System Inc. Minneapolis, USA) and the according standard protocol was used. After photometric concentration and purity measurement, DNA aliquots with a concentration of about 100 ng/μl were stored at -20°C in a tube storage system (ABgene, Hamburg, Germany). After sample selection (see 2.2.2 and 2.3.2) the DNA tubes were identified in the ABgene tube storage system and fitted into 96-well plate format for further processing.

2.1.2 DNA preparation for pooled DNA samples

The aim of the allele-specific PCR on pooled DNA samples was to reveal differences in allele frequencies between two samples. For pooling of DNA samples an exact quantification of the single DNA samples is crucial, as differences in the DNA content of the individual samples directly influence the amount of allele brought into the sample pool.

DNA quantification and dilution

For every single DNA sample, a volume of 100 μl was removed from the ABgene tube using the automated pipetting system TECAN Gemini Freedom 150 (Tecan, Crailsheim, Germany). This robot allows to retain DNA from the storage tubes without opening the tubes. To ensure the quality of the DNA samples and to prevent contamination of DNA samples a manual removal of DNA from the storage tubes is not possible. The DNA samples were transferred into 96-well deep-well plates (Corning, NY, USA). Then a 5 μl aliquot was pre-diluted with water (1:10), and from this dilution, a 5 μl aliquot was used for DNA quantification with PicoGreen (Molecular Probes, Eugene, OR, USA), a DNA intercalating fluorescent dye. The resulting DNA-dye-complex absorbs blue light (λmax = 485 nm) and emits green light (λmax = 535 nm). In a 96-well black measurement plate (Nunc, Wiesbaden, Germany) the 5 μl DNA solution were diluted with 95 μl 1xTE buffer (10mM Tris, 1mM EDTA) and mixed thoroughly. Per measurement plate a DNA standard dilution series with 8 different concentrations of λ DNA between 0 and 1 ng/μl (0; 0.001; 0.01; 0.125; 0.25; 0.5, and 1 ng/μl, respectively, each in duplicate) was used to create an 8 point calibration curve. The PicoGreen stock solution was then diluted 1:200 with 1xTE buffer, and 100 μl of this solution were added to each well of the measurement plate. After 5 min incubation of the plate, fluorescence was measured in the Tecan Spectra Fluor Plus (Tecan, Crailsheim, Germany) with a excitation wavelength of 485 nm and an emission wavelength of 535nm; the software Xfluor4 allows the quantification of the DNA samples using the 8 point calibration curve of standardized DNA. According to this measured DNA concentration, samples were then diluted with 0.1xTE buffer (containing 1mM Tris and 0.1mM EDTA) to the concentration of 8 ng/μl. The DNA concentration of these diluted samples was then confirmed twice by photometrical measurement (Gene Quant Pro, Amersham (now GE Health Care), Freiburg, Germany) without further dilution. In photometric concentration

measurement the optical density (OD) of a DNA solution was measured at 260 nm. An OD of 1 at 260 nm corresponds to a concentration of 50 ng/μl double-stranded DNA [95]. DNA samples were then diluted to the exact concentration of 8.0 ng/μl, and once again measured in the photometer. According to this measured concentration a further dilution to 4 ng/μl was made in TE buffer. This 4 ng/μl DNA solutions were used for pooling.

Pooling of DNA samples

Equal amounts of DNA samples (at 4 ng/μl) from 42 individuals contributing to the linkage signal (see sample selection) were used to establish two pools, each pool containing 21 individual samples. This procedure was replicated twice, resulting in two "contributor" pools, each in triplicate. DNA from 42 individuals that were identified as not contributing to the linkage signal was also joined in two "non-contributor" pools (each as triplicate) of 21 individuals. 5 μl of each pool, corresponding to 20 ng of DNA, was pipetted in 384-well plates and dried overnight before the application of 15 μl of PCR master mix.

2.1.3 DNA preparation for individual genotyping

DNA quantification, dilution, normalization

DNA samples were transferred into 96-well plates using the automated TECAN pipetting system and the Gemini 3.2 software (Tecan, Crailsheim, Germany). DNA samples were 1:10 diluted with water. Out of this diluted plate a 5 μl aliquot was removed for DNA quantification with PicoGreen assay (Molecular Probes, Eugene, OR, USA) as described above. After the concentration measurement, worklists for automated pipetting of normalized DNA masterplates (2 ng) were prepared using Microsoft Excel macros delivered with the automated pipetting system. Pipetting of the normalized plates was automated and performed with the Tecan pipetting system and its DNA normalization software.

Pipetting of 384-well DNA microplates

Four 96-well plates (with 2 ng DNA / μl) were used to create one 384-well plate with 10 ng of DNA per well. Every 384-well plate contained the minimum of 4 empty wells serving as no template controls in PCR. DNA was dried overnight at room temperature and then stored at -20°C until usage.

2.2 Allele-specific PCR on pooled DNA

The allele-specific PCR is an application of quantitative real-time PCR. When performed on pooled DNA, the allele-frequencies of a SNP marker in the pool can be determined [96].

2.2.1 Quantitative real-time PCR

Before starting a PCR reaction the sequence to be amplified (target sequence) must be identified, and oligonucleotide primers that are complementary to DNA sequences located on opposite DNA strands and flanking the target sequence must be designed. PCR consists of cycles of denaturation, annealing of primers and DNA synthesis. During the first step, the denaturation, the PCR mixture is heated to 90-95 °C and the DNA double strands separate, as the inter-molecular hydrogen bonds are disrupted. In the next step the PCR mixture is cooled to down to about 50-65 °C. During this cooling process the short molecules of the oligonucleotide primers quickly find their complementary DNA sequences flanking the target sequence and build a short double strand with a free 3' end (annealing). In the extension step (60-72°C), heat stable DNA polymerase starts the synthesis of a DNA strand complementary to the target sequence (using the four desoxynucleotides dATP, dGTP, dCTP and dTTP) at the free 3' end of the primer, incorporating the primer molecule in the new DNA strand. After 30 cycles the target sequence with the flanking primer sequences has been massively amplified and is the predominant DNA species in the reaction. Quantitative real-time PCR is based on regular PCR technique but offers a quantification of the PCR products at every PCR cycle. This quantification is possible through addition of a fluorescent dye (for example: SYBR Green I), which preferentially/specifically binds to double-stranded DNA and the use of a fluorescence-detecting thermocycler machine. DNA binding enhances the fluorescence of the dye and thus allows the correlation between increase of fluorescence and an increase of amplified PCR product in every cycle. In the presence of SYBR Green a fluorescence signal can only be detected at the annealing/extension stage of the PCR cycle, when double-stranded DNA molecules are created that allow intercalation of the dye.

After PCR with SYBR Green is completed, the products can be analyzed in a real-time PCR machine using a melting curve. In this procedure, the PCR mixture is heated slowly in steps of 0.5 degree from 60°C to about 95°C and fluo-

rescence is detected at every temperature step. Reaching the temperature point where the double-stranded PCR product melts and DNA strands separate as the inter-molecular hydrogen bonds are disrupted, a strong decrease in fluorescence is detected, because the SYBR Green dye is released through separation of strands. This decrease in fluorescence can be monitored in a derivative dissociation curve as a sharp peak at a specific temperature, if the PCR has led to a strong enrichment of a product of defined length and therefore has worked properly. If products of different length or unspecific PCR products were amplified, diffuse or multiple peaks can be detected in the derivative dissociation curve. Melting curves are used to check the specificity of the amplification reaction. To determine the length of the amplification product a gel electrophoresis using a DNA standard must be performed.

If two PCR reaction are compared, as in the application of allele-specific PCR, it is crucial to ensure that the efficiency of the two reactions is equal and ideally both reactions have an efficiency of 100%. The efficiency of a PCR reaction can be revealed using a dilution series of template DNA (standard curve).

2.2.2 Allele-specific PCR

Allel-specific PCR (asPCR) allows the selective amplification of only one allele of an existing SNP, by choosing an allele-specific primer whose last 3' base is complementary to the selected SNP allele. To look at both alleles of a SNP, three primers in summary have to be designed. The location of the both allele-specific primers (only different in the last base) is determined by the SNP. The other primer can be used for both reactions. The use of real-time PCR with SYBR Green and allele-specific primers, allows to detect differences in allele content. The principle of asPCR is depicted in figure 2.1.

For every DNA sample two reactions have to be carried out in parallel, as every allele must be amplified separately for differentiation of allele content. The differences in allele content can be monitored in real-time PCR. Figure 2.2 displays an amplification plot of two asPCR reactions. A single DNA heterozygous for the investigated SNP has been used as template and shows parallel amplification of both alleles, which indicates equal amounts of both alleles, as expected. This amplification plot also shows a comparable PCR efficiency of both reactions, as amplification curves rise simultaneously.

To ensure a good amplification efficiency, the ideal length of the amplicon is between 90 and 140 bp. The melting temperatures of all the primers as well as their

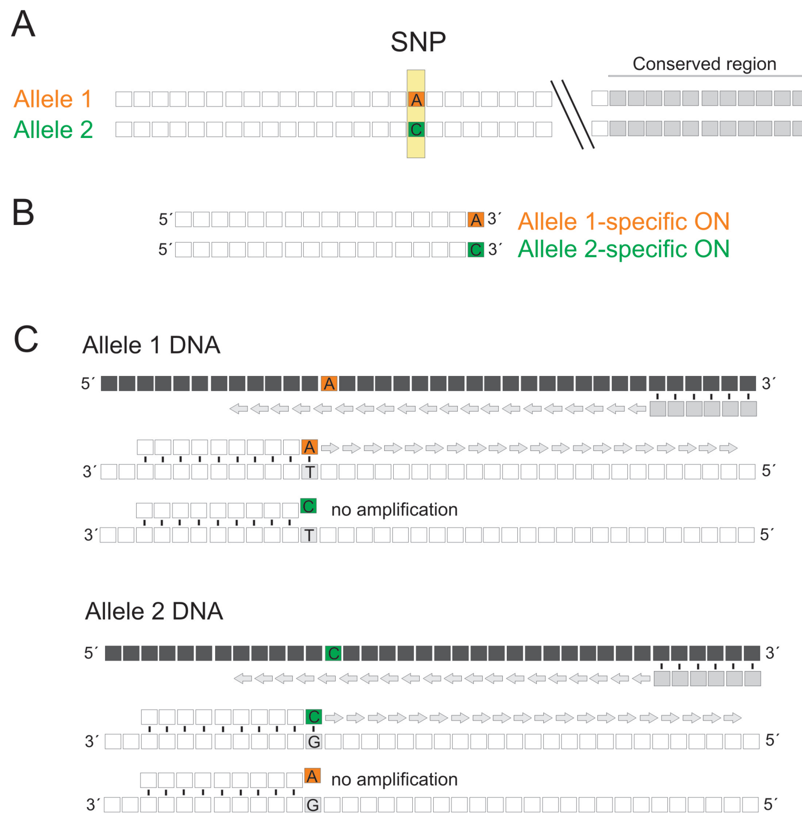

Figure 2.1: Principle of allele-specific PCR. (A) In appropriate distance to the SNP a conserved region without SNPs is identified and from this region the third, common oligonucleotide (ON) is chosen. (B) Allel-specific ON are designed according to the SNP position, varying in the 3'-terminating base. (C) In a PCR reaction with all three ON, the allele 1-specific ON will perfectly bind to the complementary strand of the allele 1 sequence, permitting amplification with the ON in the conserved region. However, the 3'-terminal C of the allele 2-specific ON mismatches the T of the allele 1 sequence, making amplification impossible. In case of the allele 2 sequence, only allele 2-specific ON can initiate amplification. Modified after [94], p.128.

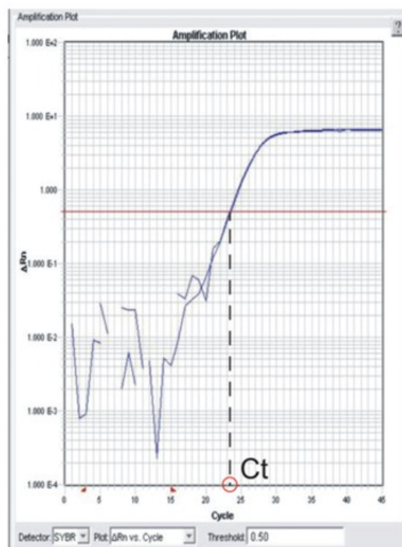

Figure 2.2: Real-time PCR amplification plot. On the y-axis the increase in detected fluorescence as a measure for the created PCR product is plotted in a logarithmical scale. The x-axis shows the number of the corresponding PCR cycle. The red line (threshold) must be set in the exponential PCR phase (automatical or manual) and allows to compare two samples using the crossing point of the threshold line and the amplification plot, with the Ct value representing the corresponding PCR cycle. In the depicted amplification plot, two asPCR amplification curves of a single DNA sample heterozygous for the investigated SNP marker are shown. Both PCR reactions show parallel amplification and Ct values are identical, indicating equal amounts of both alleles as expected for a heterozygous sample.

GC-content should be comparable. The intention of applying asPCR to DNA pools, is to quickly identify SNPs with relevant differences in allele frequency between the pools. The asPCR on DNA pools can not provide genotypes but reveals differences in allele frequencies between the pools, and therefore allows a relatively quick identification of relevant SNPs. The asPCR conditions were established on test DNA samples and test DNA pools from healthy volunteers to save patient DNA material.

2.2.3 Selection of SNP markers for allele-specific PCR

The position of the investigated SNP markers is depicted in figure 2.3. The region for the SNP selection comprised the *LPA* gene as well as the intergenic regions and the neighboring genes *LPAL2* encoding soluable lipoprotein lipase 2 and the *PLG* gene, encoding plasminogen, a protein similar to Lp(a).

SNP markers were selected for the following criteria: 1. suitability of the surrounding sequence for asPCR and 2. rough distribution in and around the *LPA* gene 3. validated SNP marker with a MAF above 10% in Caucasians. Primers as well as the resulting amplicons were checked for specificity by a BLAST search on the NCBI site. Furthermore secondary structure of the amplicons under PCR conditions (annealing temperature and ionic conditions) was checked to sort out candidates with a presumably bad PCR efficiency due to strong hairpin structures (using mfold version 3.2) [97]. Of 3,210 validated SNPs in the *LPA* gene region (from dbSNP build 125) 12 SNPs were selected for asPCR on pooled DNA. All used

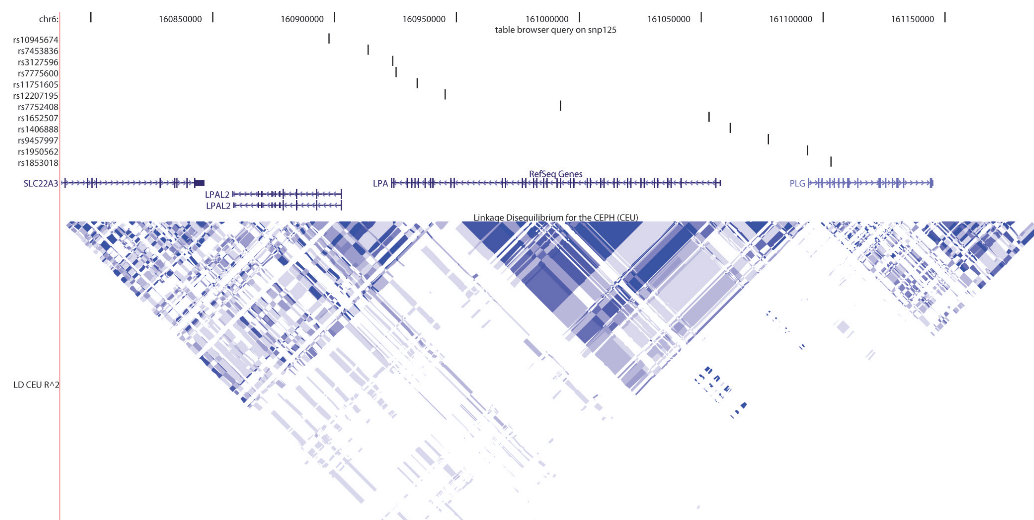

Figure 2.3: Distribution of SNPs in the *LPA* gene region investigated by asPCR on DNA pools. The position of the SNPs is depicted above the schematic representation of the 300 kb region on human chromosome 6q27 NCBI build 35: 160,850,000-161,150,000 containing the *LPA* gene and its neighboring genes, *LPAL2* and *PLG*. Underneath the LD structure in the *LPA* gene region is represented, pairwise r^2-values between markers from CEU samples of HapMap project phase II release 21 (http://www.hapmap.org) are shown with darker regions representing higher LD.

primer trios are listed in table 2.1. Prior to the experiments with pools, the primer trios were tested on different DNA samples (from healthy volunteers) to optimize PCR conditions and evaluate the specificity (melting curve) and efficiency (standard curve) of the PCR reactions. Runs on test pools with defined allele content were then performed to analyze the limits of resolution in allele frequency differences.

2.2.4 Selection and pooling of DNA samples

For 499 families with 1,298 individuals of the German MI family study, linkage data from a total genome scan using 394 microsatellite markers were available from a previous investigation [81]. The analysis of the microsatellite markers with the Lp(a) serum levels led to a linkage signal at chromosome 6q27 with a LOD score of 26.99 [81]. Of these families 42 were identified as contributors to this linkage signal (per family LOD score >0.4) and 42 families without contribution to this linkage signal (per family LOD score <-0.2). DNA pools of MI patients that contributed to this QTL on Lp(a) plasma levels on chromosome 6q27 and DNA pools of MI patients that did not contribute to this QTL were investigated.

Table 2.1: Oligonucleotides for allele-specific PCR

dbSNP rs number (bulid 125)	location	position on Chr6 (may 2004 assembly)	alleles	oligo short name	oligo name	sequence (5´-3´)	Tm (°C)	length (bp)	length of amplicon (bp)
rs10945674	Lp-PLA2	160,897,331	G/C	A ON-1	ON-rs1094674-1	TCTCAGAGGCTTTGTTCATTTTT	59.42	23	106
				A ON-2	ON-rs1094674-2	GTAGTTAAGGAAAGAATCTCAGACCTTc	59.69	28	
				A ON-3	ON-rs1094674-3	AGTTAAGGAAAGAATCTCAGACCTTg	59.75	26	
rs7453836	intergen.	160,913,302	T/A	B ON-1	ON-rs7453836-1	GGCCTAGCGAGTGACCCTTA	61.65	20	122
				B ON-2	ON-rs7453836-2	AGAAATTCACAAGTGCACAACGTt	62.13	24	
				B ON-3	ON-rs7453836-3	GTAGAAATTCACAAGTGCACAACGTa	61.61	26	
rs3127596	LPA	160,923,446	G/A	C ON-1	ON-rs3127596-1	GTGGTTTTGTTGAATAGACTTTGC	59.14	24	98
				C ON-2	ON-rs3127596-2	ATTTTCCCAGTAATTCAAATCAGAAa	60.42	26	
				C ON-3	ON-rs3127596-3	ATTTTCCCAGTAATTCAAATCAGAAg	60.12	26	
rs7775600	LPA	160,925,020	A/G	D ON-1	ON-rs7775600-1	TTCCAACCTTCCACATGTTT	57.89	20	90
				D ON-2	ON-rs7775600-2	AATATGTCTATATAATTATGGGAACTGAGAc	58.15	31	
				D ON-3	ON-rs7775600-3	AATATGTCTATATAATTATGGGAACTGAGAt	57.79	31	
rs11751605	LPA	160,933,641	C/T	E ON-1	ON-rs-11751605-1	CAAGCAAACTGAACTTACGAATTG	60.21	24	137
				E ON-2	ON-rs-11751605-2	CATCTATTCACAAACACATGAACAAg	59.87	26	
				E ON-3	ON-rs-11751605-3	CATCTATTCACAAACACATGAACAAa	60.20	26	
rs12207195	LPA	160,944,989	A/G	F ON-1	ON-rs12207195-1	CGTCCAACATCAAAGATAACAAG	58.69	23	134
				F ON-2	ON-rs12207195-2	TTATCACTTTAAAGATGTTATGCCATTa	58.68	28	
				F ON-3	ON-rs12207195-3	TCACTTTAAAGATGTTATGCCATTg	59.48	25	
rs7752408	LPA	160,992,284	C/G	G ON-1	ON-rs-7752408-1	ACAGGTTGCACCAAAAATCACAC	63.28	23	131
				G ON-2	ON-rs-7752408-2	ATACCCCCACTGACGCACAAc	64.31	21	
				G ON-3	ON-rs-7752408-3	ATACCCCCACTGACGCACAAg	64.4	21	
rs1652507	LPA	161,052,872	G/A	H ON-1	ON-rs1652507-1	TCTCAGGAGGTCAGGACAGG	60.39	20	102
				H ON-2	ON-rs1652507-2	CACTGCAGAGAACATGTGTTTGTAc	61.03	25	
				H ON-3	ON-rs1652507-3	CACTGCAGAGAACATGTGTTTGTAt	60.51	25	
rs1406888	intergen	161,062,004	G/A	I ON-1	ON-rs1406888-1	CAGAGGTGCTGGTGAGGACA	62.51	20	136
				I ON-2	ON-rs1406888-2	CTGGCTAGTGGCAGAGTCAGAAc	62.72	23	
				I ON-3	ON-rs1406888-3	CTGGCTAGTGGCAGAGTCAGAAt	62.17	23	
rs9457997	intergen	161,077,429	A/G	J ON-1	ON-rs9457997-1	GCGCTGGTATTGTCCAAGG	61.59	19	92
				J ON-2	ON-rs9457997-2	CATACTTAATGTGGAAAGGTTGCTTc	61.35	26	
				J ON-3	ON-rs9457997-3	ACATACTTAATGTGGAAAGGTTGCTTt	61.52	27	
rs1950562	PLG	161,093,596	A/G	K ON-1	ON-rs1950562-1	TGAGAGTAAACGCATCCACA	57.84	20	117
				K ON-2	ON-rs1950562-2	GTTGTAAGGACTCTTTATTCAAGGTAAt	58.03	28	
				K ON-3	ON-rs1950562-3	TTGTAAGGACTCTTTATTCAAGGTAAc	57.73	27	
rs1853018	PLG	161,103,215	G/A	L ON-1	ON-rs1853018-1	CTCTTTTGCTCCACAATTTGAGTC	61.47	24	103
				L ON-2	ON-rs1853018-2	CAAATCTGAGAAAAGATCAAAGATGACTAc	62.07	30	
				L ON-3	ON-rs1853018-3	CAAATCTGAGAAAAGATCAAAGATGACTAt	61.64	30	

2.2.5 asPCR procedures

All PCRs were performed on an Applied Biosystems 7900HT Real-time PCR system. The composition of a standard asPCR reaction is listed in table 2.2.

Table 2.2: Standard asPCR reaction

	total volume: 15 μl
PCR Master Mix (ABI or Qiagen)	7.5 μl
ON-1 (10 μM)	0.45 μl
ON-2 or ON-3 (10 μM)	0.45 μl
water	6.6 μl

Two different commercial PCR master mix, containing SYBR Green, were used: Qiagen SYBR Mix (Qiagen, Hilden, Germany) and Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA). Three PCR runs were performed on the DNA pools with different annealing/extension temperatures (listed in table 2.3).

Table 2.3: Thermocycling conditions of asPCR runs on DNA pools

asPCR run	enzyme	PCR cycles (40)		
	activation	denaturation	annealing	extension
run 1				
temperature(°C)	95	94	58	72
time (s)	900	15	15	20
run 2				
temperature(°C)	95	94	57	72
time (s)	900	15	15	20
run 3				
temperature(°C)	95	94	58	60
time (s)	900	15	15	20

Allele-specific PCR conditions for rs1652507 and rs1406888 could not be sufficiently optimized and therefore patient pools were not analyzed for these two markers.

2.2.6 Analyzing the PCR data

Allele frequencies of the pooled samples were determined according to Germer et al. [96]. The number of PCR product (y) in a PCR reaction can be obtained using the following equation:

$$y = x \cdot (1 + E)^n$$

with x being the initial target copy number, E the efficiency of the reaction and n the number of PCR cycles. At 100% efficiency of the PCR reaction the equation becomes:

$$y = x \cdot 2^n$$

which means that in a fully efficient PCR reaction the amount of PCR product doubles at every cycle. Thus, a 2-fold dilution series of template DNA results in a ΔCt of 1 between the samples. For DNA pools the difference in Ct between two different PCR reactions (ΔCt) is a measure of the allele frequency. In the first step the inter-allelic Ct difference is determined for every pool (C= contributor pool, N= non-contributor pool):

$$Ct_{\text{Allele 1 of Pool C}} - Ct_{\text{Allele 2 of Pool C}} = \Delta Ct_{\text{Pool C}}$$

The derived ΔCt, can be either positive or negative depending on which allele exhibits the lower Ct. A ΔCt of one cycle, means that the ratio of one allele to the other is 1:2, or in general:

$$1 : 2^{\Delta Ct}$$

The difference in Ct can also be expressed in allele frequency, instead of ratios. To convert a ratio to a frequency, the numerator is added to the denominator, which results in the following formula:

$$\text{frequency of allele 1} = \frac{1}{(2^{\Delta Ct} + 1)}$$

The allele frequencies are calculated for the contributor and the non-contributor pools and compared.

2.3 Genotyping with TaqMan Assays

2.3.1 The 5' nuclease assay

The TaqMan technique combines in a single step PCR amplification and detection by the use of fluorogenic probes in a 5' nuclease assay (Figure 2.4) [98].

The fluorogenic probes consist of an oligonucleotide labeled with a fluorescent reporter at its 5' end and a quencher dye at its 3' end. In the intact probe the proximity of the quencher dye reduces the fluorescence signal from the reporter

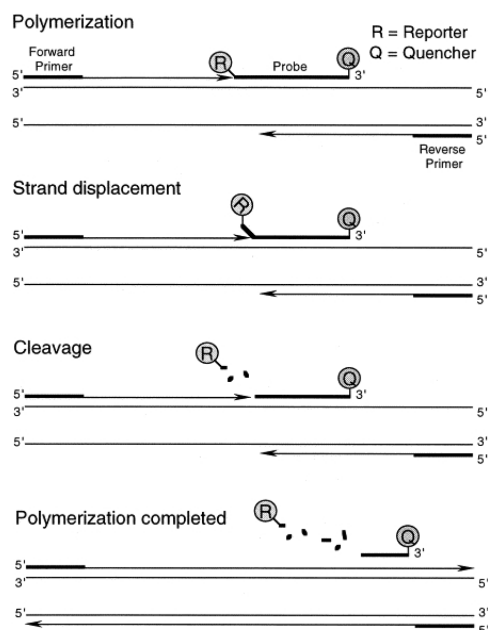

Figure 2.4: 5' nuclease assay. Stepwise representation of the forklike-structure-dependent, polymerization-associated, 5' to 3' nuclease activity of *Taq* DNA polymerase acting on a fluorogenic probe during one extension phase of PCR. Two dyes, a fluorescent reporter (R) and a quencher (Q), are attached to the probe. When both dyes are attached to the probe reporter dye emission is quenched. During each extension cycle the *Taq* DNA polymerase cleaves the reporter dye from the probe. Once separated from the quencher, the reporter dye emits its characteristic fluorescence [98].

dye. The probes are added to the PCR amplification reaction together with the forward and reverse primer. During the annealing/extension step of PCR the labeled probe hybridizes to the target sequence. Starting at the 3' end of the primer, *Taq* polymerase extends the amplicon until reaching the bound probe. The 5' nuclease activity of the *Taq* polymerase cleaves the probe at the 5' end and liberates the reporter dye from the probe. Uncoupling of the reporter dye leads to an increase in fluorescence signal, as the proximity to the quencher dye is lost. Thus, an increase of fluorescence indicates that the probe specific target has been amplified.

The cleavage of the probe occurs only if the probe is specifically hybridized to its target sequence, as a mismatch promotes displacement of the probe rather than cleavage. Therefore, the 5' nuclease assay allows to discriminate alleles that differ by a single base substitution, with the use of differently labeled specific probes for each allele (Figure 2.5) [98].

As the read-out of the experiment is fluorescence, a measurement of the fluorescence spectrum is required after PCR. This post PCR fluorescence data detected for every well of the PCR plate is than automatically processed by the Sequence Detection System software (SDS) to make genotype calls. Ideally, a sample homozygous for the allele detected by a FAM labeled probe should emit the fluorescence spectrum of the FAM dye, a sample homozygous for the allele detected by a VIC labeled probe should emit the fluorescence spectrum of the VIC dye. A heterozygous sample should emit the spectra of both dyes in equal intensity. In a cluster plot

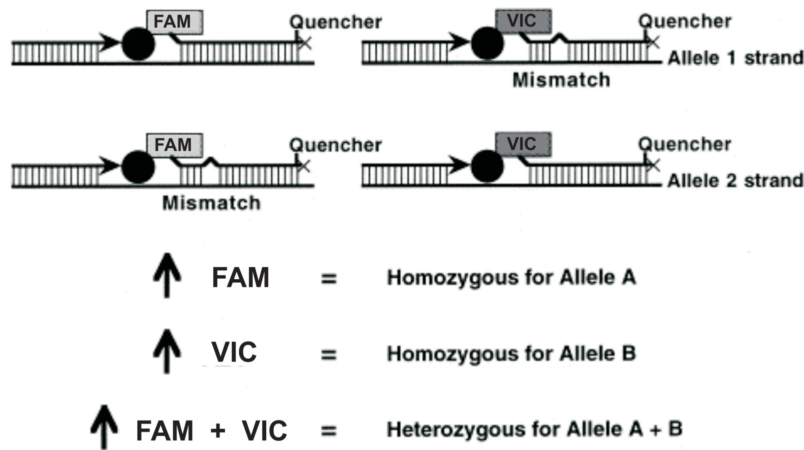

Figure 2.5: Allelic discrimination assay design strategy with fluorogenic probes in the 5' nuclease assay. The presence of a mismatch between probe and target reduces the efficiency of probe cleavage. Modified after [98].

the fluorescence intensity of the different dyes is plotted for every well of the PCR plate and the software identifies the different clusters, representing the three clouds of possible genotypes (Figure 2.6). The TaqMan technique is especially suitable for genotyping of large sample sets with few assays.

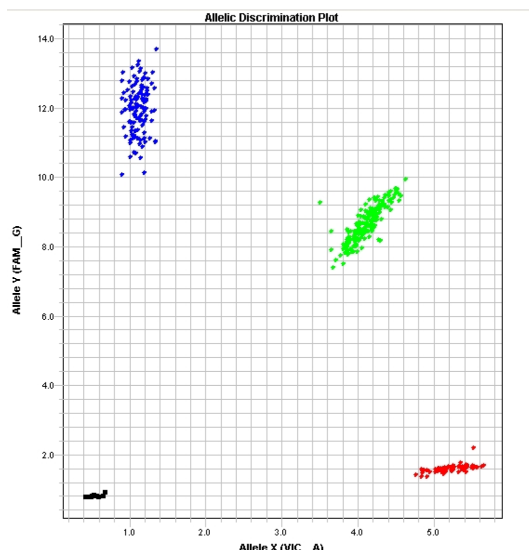

Figure 2.6: Cluster plot as result of allelic discrimination with fluorogenic probes in the 5' nuclease assay. The plot shows four clouds, the blue cloud represents the samples homozygous for G allele detected by the FAM labeled probe, the green cloud represents the heterozygous samples, and the red cloud represents the samples homozygous for the A allele detected by the VIC labeled probe. The black cloud in the lower left corner of the plot represents the no template control samples.

2.3.2 SNP selection and genotyping of individual DNA samples

In total six SNPs of the *LPA* gene were genotyped with the TaqMan assays, the SNP characteristics are given in table 2.4.

Table 2.4: SNP characteristics

SNP	absolute position*	*LPA* Location	Alleles	ABI-assay ID
rs7752408	160,992,284	Intron 19	G>C	C_1550878_10
+79170 G>A	160,976,489	Intron 26	G>A	-
rs6923877	160,967,208	Intron 28	G>A	C_103952_10
rs12207195	160,944,989	Intron 30	G>A	C_407852_10
rs11751605	160,933,641	Intron 34	T>C	C_282793_10
rs3798220	160,931,548	Exon 36	T>C	C_25930271_10

$*$ on chromosome 6, NCBI build 35

2.3.3 Standard TaqMan Assay

Individual DNA samples were genotyped using 5´ nuclease TaqMan technology (Applied Biosystems, Foster City, CA, USA) with differently fluorescence labeled probes including non-fluorescence quencher and minor groove binder (MGB). Available TaqMan assays were ordered from Applied Biosystems, the assay for +79170 G>A was designed using the Primer Express software (Applied Biosystems), producing a 98 bp-amplicon with primer sequences for left primer: 5´CATCGGAGGATCCCATTAT 3´ and right primer: 5´CATGGGCCAGCTAAGAGA 3´. The following MGB probes were used: 5´6FAM-ATAGACATACGCATTTG 3´MGB detecting the G allele and 5´VIC-TAGACATATGCATTTGG 3´ MGB detecting the A allele. For each genotyping reaction 10 ng of genomic DNA were applied in a 384-well plate and dried. Then a 5 μl volume containing 2.5 μl 2xTaqMan Genotyping Master Mix (Applied Biosystems) and the recommended amount of assay was applied to each well. Thermocycling was performed under recommended conditions (10 min of initial denaturation at 95°C, followed by 40 cycles of 15 sec denaturation at 95°C and a combined annealing/extension step of 1 min at 60°C) in the Applied Biosystems 7900HT Real-time PCR system or an external PCR system (GeneAmp PCR 9700, Applied Biosystems). The post PCR endpoint plate read was carried out using the Applied Biosystems 7900HT Real-time PCR system. Automated genotype calling (applying the allelic discrimination test) was performed with the Sequence Detection

System (SDS) software 2.2 from Applied Biosystems. To check for correct genotyping 20% of DNA samples were genotyped twice and no genotyping discrepancies between the duplicates were detected. The call rates were between 98.6 and 99.4%.

2.4 Genotyping with microarrays for genome-wide association

A microarray is a miniature array of different DNA or oligonucleotide sequences on a glass surface that is intended to be used in hybridization assay [94]. The different steps of the genotyping procedure with the Affymetrix GeneChip Human Mapping 500K Array Set are schematically depicted in figure 2.7. The GeneChip Human Mapping 500K Array Set is comprised of two arrays [99]. One array uses the *NspI* restriction enzyme and is capable of genotyping about 262,000 SNPs, the other array uses the *StyI* restriction enzyme and is capable of genotyping about 238,000 SNPs [99]. The median physical distance between SNPs is 2.5 kb and the average distance between SNPs is 5.8 kb [99]. Per SNP, 24 to 40 different 25 bp- oligonucleotides are present in the array set, the oligonucleotides are synthesized in parallel by a photolithographic procedure [99].

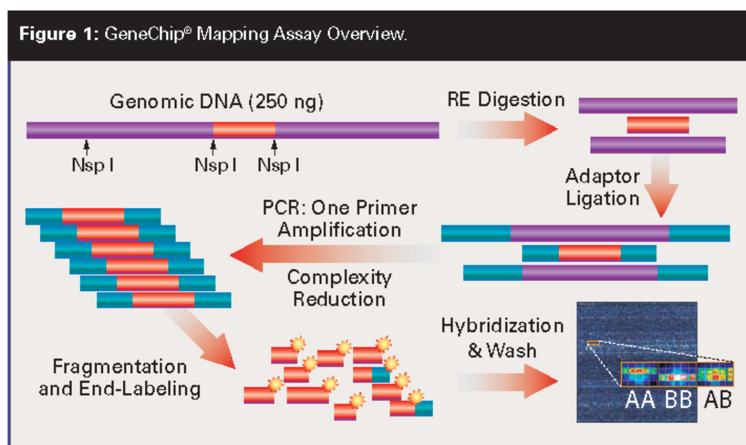

Figure 2.7: Genotyping with an Affymetrix SNP Chip. After digestion of the genomic DNA with the restriction enzyme *NspI* (or *StyI*), adapter molecules are ligated to the DNA. Then an amplification step, a labeling step and finally the hybridization step follow. The fluorescence intensities of the hybridized probes are read out in a scanner, and transferred into genotype calls with a calling algorithm [99].

From a previous investigation for a subsample of patients from the German MI Family Study (n=875) data from a genome-wide SNP genotyping using 500,000 SNP markers on the GeneChip Human Mapping 500K Array Set from Affymetrix was available [100]. The genotyping of the German samples was performed at the GSF National Research Centre in Neuherberg (now Helmholtz Zentrum Munich, German Research Center for Environmental Health).

2.5 Sample selection

Three different population samples were used for individual genotyping with TaqMan assays.

Family sample from the German MI family study

The MI families were ascertained through index patients, who were identified by screening 200,000 patient charts in 14 cardiac in-hospital rehabilitation centers distributed throughout Germany [81]. Index patients had all suffered from MI before the age of 60 years [81]. If at least one first-degree relative had suffered from MI or had severe coronary artery disease (percutaneous transluminal coronary angioplasty (PTCA) or bypass surgery (CABG) before the age of 70 years), the nuclear family was contacted and invited to participate in the study [81]. From all 7,573 persons included in the study those with measured Lp(a) levels were selected for the present investigation. In total 2,292 individuals from 687 families with measured Lp(a) levels were identified in the database and the DNA samples were used for TaqMan genotyping. As this sample allowed only family based association testing, two subsamples (case samples) were deduced from this family sample for a case-control association study with an external control population. The first case sub-population, referred to as 'Index MI patients', consisting of 595 Index MI patients (as defined above) with measured Lp(a) plasma levels were available for genotyping in the present study. A second case population consisting of single siblings of index MI patients with documented CAD was used for verification (n=471, referred to as 'affected CAD siblings'). The study protocol was approved by Ethics Committee of the University of Regensburg, and all participants gave informed consent.

Population-based controls

As a population-based control sample, controls from the MONICA Echocardiographic substudy were used. Subjects participated in the Echocardiographic Substudy, of the third MONICA (MONitoring of Trends and Determinants In CArdiovascular Disease) Augsburg Survey, which is now continued in the framework of KORA (Cooperative Health Research in the Augsburg Area) [101]. This study population originated from a sex- and age-stratified cluster sample of all German residents of the Augsburg study area [102, 101]. The study design, sampling frame, and data collection have been described in detail elsewhere [101]. The third survey represents individuals aged 25 to 74 years (mean: 51.8±13.8, including 851 (50.7%) women and 827 (49.3%) men and 300 subjects for each 10-year increment (total n=1,674). From the 1,674 subjects, both DNA samples and Lp(a) data were available in 975 subjects, in the following referred to as 'MONICA controls'.

Replication MI case sample

In 1996 to 1997, MI patients from the population-based MONICA MI registry Augsburg who had experienced their first MI before 60 years of age were examined in a study center (n=609). MI had been verified according to standard MONICA criteria [101]. Of these, 450 blood samples with measured Lp(a) levels were available for genotyping in the present study and are referred to as 'KORA MI cases'. These subjects served as an additional independent case population consisting of incidental non-familial MI cases for replication of the results.

Cardiogenics sample for genome-wide association

In 2007 a genome-wide association analysis of coronary artery disease, with about 500,000 SNP markers per individual investigated, was performed in 1,926 case subjects and 2,938 controls from the Wellcome Trust Case Control Consortium (WTCCC) and replicated in the German MI Family Study (with 875 cases and 1,644 controls) [100]. Genotyping in both studies was performed with the use of the GeneChip Human Mapping 500K Array Set from Affymetrix [100]. For 487 of the investigated cases from the German MI Family Study Lp(a) levels were available from a previous investigation [52]. With this subsample of Cardiogenics cases a genome-wide association of Lp(a) levels using the data from the Affymetix SNP genotyping was performed.

2.5.1 Measurement of Lp(a) levels

The Lp(a) levels used in this investigation were obtained in a previous investigation [52]. The method of measurement used in this previous investigation yielded valid data including the result of 0 mg/dl Lp(a). Square-root transformation was therefore used to normalize the skewed distribution of Lp(a) values. A logarithmical transformation could better harmonize the extreme high values, but also would loose the data points at 0 mg/dl, and therefore a logarithmical transformation was not eligible for the existing data. A new measurement of Lp(a) in all samples was also not possible, as serum was not accessible from all samples and the age of samples could strongly influence the measurement, though in the mean time Lp(a) level measurements have been refined.

2.5.2 Phenotypic apo(a) information and information on the pentanucleotide repeat polymorphism

Phenotypic size of the expressed apo(a) (corresponding to the number of KIV-2 repeats) obtained by immunoblotting of serum was available for 406 Index MI patients, 355 affected CAD siblings, 678 MONICA control samples and 172 KORA MI samples from a previous investigation [52]. Additionally, information on the number of pentanucleotide repeats in the 5' region of the *LPA* gene obtained by PCR was available for 346 Index MI patients, 301 affected CAD siblings, 693 MONICA control samples and 414 KORA MI samples from a previous investigation [52].

2.5.3 Definitions

Cardiovascular risk factors were defined using standard criteria. Systemic arterial hypertension was defined as a systolic blood pressure of >140 mmHg or diastolic blood pressure >90 mmHg or the use of antihypertensive medication. Actual hypertension was defined as systolic blood pressure of >140 mmHg or diastolic blood pressure >90 mmHg. Hypercholesterolemia was defined as a documented low density cholesterol value >130 mg/dl or the treatment with lipid lowering drugs. Diabetes was defined as history of diabetes mellitus or the use of antidiabetic medication. Smoking status was defined as current or former cigarette smoking on a regular basis. Body mass index (BMI) was calculated as body weight in kilograms divided by height in meters squared. Age is given as age at MI for Index MI patients and

KORA MI patients, as age at onset of coronary disease for the affected CAD siblings and as age at examination for the MONICA controls. Continuous variables in tables are given as mean values with standard deviation unless stated otherwise.

2.6 Statistical analysis

2.6.1 Family-based association testing

Two different approaches were used for association testing in the family sample with the quantitative trait Lp(a). Square-root transformed Lp(a) levels were used in both tests. First the program PLINK was used, all analysis were performed with version v0.99s [103] (http://pngu.mgh.harvard.edu/purcell/plink/). In PLINK the QFAM procedure allows family-based tests of association with quantitative phenotypes. PLINK performs a simple linear regression of phenotype on genotype, but then uses a special permutation procedure to correct for family structure. The QFAM procedure was performed with 100,000 permutations. The other method used is implemented in the program FBAT (Family-Based Association Tests in genetic analyses) [104]. Similar in spirit to a classical TDT (Transmission Disequilibrium Test), in which alleles transmitted to affected offspring are compared with the expected distribution of alleles among offspring, the approach compares the genotype distribution observed in the 'cases' to its expected distribution under the null hypothesis. In the case of linkage the null hypothesis is: no association, in the presence of linkage. Under the null hypothesis of "no association in the presence of linkage" sibling marker genotypes are correlated and nuclear families can no longer be treated as independent [105]. However, Lake et al. showed that a valid association test in the presence of linkage is performed using the mean of the test statistic computed via the Rabinowitz-Laird algorithm under the null hypothesis of "no association and no linkage", by using an empirical variance-covariance estimator that adjusts for the correlation among sibling marker genotypes and for different nuclear families within a single pedigree [106]. FBAT offers the -e option for association analysis in the presence of linkage, this tools provides an option to calculate this empirical correction to the variance [106, 104]. The -e option of FBAT was used for the association analysis in the family sample.

2.6.2 Case-control studies

All parameters were compared between the MONICA control group and each case group separately, using JMP 5.0.1 (SAS Institute, Cary, NC, USA). Clinical characteristics were compared with a student's t test for normally distributed continuous variables, and the χ^2 statistics for dichotomized variables. Because of skewness of Lp(a) levels, the nonparametric Mann-Whitney U test was used for comparison of Lp(a) levels. Lp(a) levels by genotypes were compared by the non-parametric Kruskal-Wallis test. For each of the SNPs, the observed allele frequencies were tested for deviation from Hardy-Weinberg proportion. Linkage disequilibrium (LD) was assessed as r^2 and D' between all pairs of SNPs. PLINK and Haploview 4.0 were used to assess r^2 between SNPs and D' was assessed applying the standard definition of D' [103, 107, 108]. Crude and adjusted logistic regression analysis was used to assess the association of the risk allele 2 with MI and/or CAD (genotype 12 or 22 vs. 11). Adjustments have been made for age, gender, BMI, systolic blood pressure, diabetes, smoking, the total cholesterol/HDL cholesterol ratio, and in separate models additionally for square root transformed Lp(a) levels. Moreover, MI patients were also matched with control subjects by age ($\pm$10 years) and gender by use of an automated, randomized selection of control subjects, in order to avoid systematic differences in genetic composition between the two groups. Using this strategy, an appropriate control subject could be found for 473 MI patients. In this matched case-control sample, conditional logistic regression analysis was used for testing the association between case status and the risk allele. To test whether a particular SNP, either directly or through LD with the QTL underlying mutation, contributes to the evidence for linkage, each SNP under a dominant model was entered into the QTL linkage analysis and the change in the magnitude of the LOD score calculated. Testing for evidence of linkage to the Lp(a)-QTL on chr6q27 was done in the Index MI cases and their affected siblings included in the present study, using the variance component approach implemented in the SOLAR software package (version 2.1.4) [109, 110, 111]. A two-sided P value $<$0.05 was considered significant.

2.6.3 Genome-wide association testing

The data from the GeneChip Human Mapping 500K Array Set were analyzed with PLINK. Square-root transformed Lp(a) levels were used for quantitative association tests with linear regression, in this analysis 262,314 SNP markers from the *NspI*

array and 189,456 SNP markers from the *StyI* array were used. The results were graphically displayed using Haploview 4.0. As a lot of SNP markers in the *LPA* gene region gave strong association results with Lp(a) levels, an adjustment for these highly significant SNPs was performed, also using PLINK. A linear model association was calculated with square-root transformed Lp(a) levels, using the selected tagging SNPs in a condition list. Tagging SNPs were selected using HapMap data of build 36 and the "pairwise tagging" option in Haploview 4.0.

3 Results

3.1 Allele-specific PCR on pooled DNA samples

Of 3,210 validated SNP markers in the *LPA* gene region (from dbSNP build 125) 12 SNPs were selected for asPCR on pooled DNA (see 2.2.3). The allele frequencies of the investigated SNP markers in the pool were determined according to [96], as described in 2.2.6. Allele-specific PCR on pooled DNA samples revealed that three SNP markers showed a difference in allele content between the contributor and non-contributor pools. Table 3.1 shows the results of asPCR. The three SNP markers, rs11751605, rs12207195 and rs7752408, all located within the *LPA* gene, showed a marked difference in allele distribution between DNA pools of MI patients that contributed (per family lod score >0.5) to a previously reported QTL on Lp(a) plasma levels on chromosome 6q27 and (Pools A + B) MI patients that did not contribute (per family lod score <0.5) to this QTL (Pools C + D) [81]. A difference in allele frequency of more than 10% between the pools was considered significant. This rather conservative relevant difference was chosen as the compared pools represent opposite extremes of samples. For rs12207195 an exception was made, the allele frequency difference was only 4.4%, but as the detected allele had a low frequency (8.7% in the contributor pool and 13.1% in the non-contributor pool), and the absolute difference in allele content was 1.5 fold, therefore the difference was considered relevant.

3.1.1 SNP selection and genotyping of individual DNA samples

To verify the results from asPCR on pooled samples, an individual genotyping using TaqMan technology was performed in the MI family study. Additionally to the three SNP markers, rs11751605, rs12207195 and rs7752408 that had shown a marked difference in the pooled samples, three other SNPs within the *LPA* gene were chosen. One SNP, rs6923877, was identified on the Applied Biosystems homepage

Table 3.1: Results from asPCR on contributor and non-contributor pools

SNP marker	run	Pool	mean of inter-allelic Ct difference per pool (ΔCt) as	frequency of allele 1 (calculated for contributors (A+B)and non-contributors(C+D) $1/(2^{\Delta Ct}+1)$)	difference between contributor and non-contributor pools greater than 10%?
rs10945674	3	A	1.426	0.276	no
		B	1.360		
		C	1.877	0.255	
		D	1.221		
rs7453836	1	A	0.262	0.374	no
		B	1.223		
		C	0.843	0.313	
		D	1.422		
rs3127596	1	A	0.290	0.443	no
		B	0.374		
		C	-0.169	0.444	
		D	0.492		
rs7775600	3	A	-1.686	0.780	no
		B	-1.965		
		C	-1.599	0.756	
		D	-1.670		
rs11751605	2	A	-0.273	0.567	yes
		B	-0.504		
		C	0.828	0.396	
		D	-0.219		
rs12207195	1	A	3.424	0.087	no
		B	3.349		
		C	2.676	0.131	
		D	2.786		
rs7752408	1	A	-1.718	0.800	yes
		B	-2.284		
		C	-0.894	0.625	
		D	-0.580		
rs9457997	1	A	-1.930	0.790	no
		B	-1.893		
		C	-1.809	0.791	
		D	-2.034		
rs1950562	1	A	0.902	0.386	no
		B	0.436		
		C	0.743	0.426	
		D	0.115		
rs1853018	1	A	-0.132	0.453	no
		B	0.401		
		C	0.223	0.455	
		D	0.300		

as appropriate for TaqMan genotyping. Two more SNP markers were chosen from the literature: rs3798220 was reported to be associated with plasma Lp(a) levels and severe CAD and a G to A substitution at the +1 donor slice site of KIV-8 intron, with the position 79,170 relative to ATG, and the absolute position on Chr6: 160,976,489 on NCBI build 35 (in the following referred to as '+79170G>A' was reported to have a strong decreasing effect on Lp(a) plasma concentration [93, 40]. The SNP characteristics are given in Table 3.2 and their position in the *LPA* gene is depicted in Figure 3.1.

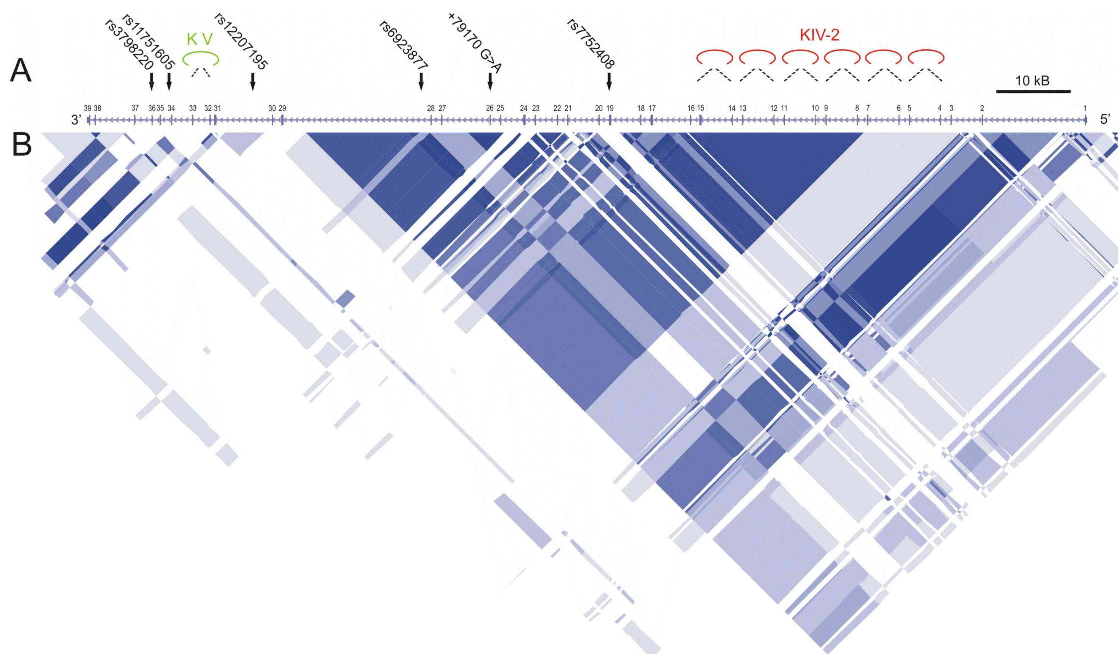

Figure 3.1: Schematic representation of the *LPA* gene structure and its LD pattern.(A) The 132.8 kb region on human chromosome 6q27 NCBI build 35: 160,922,927-161,055,702 containing the *LPA* gene (RefSeq NM 005577.2) with 39 exons (vertical blue lines) oriented from right to left, SNP positions (black arrows) and repeats of KIV-2 (red kringles) are given, as well as the kringle V domain (in green). (B) Representation of LD structure in the *LPA* gene. Pairwise r^2-values between markers from CEU samples of HapMap project phase II release 21 (http://www.hapmap.org) are shown with darker regions representing higher LD.

Table 3.2: SNP characteristics

SNP	absolute position*	Location	Alleles	MAF† (%)
rs7752408	160,992,284	Intron 19	G>C	37.2
+79170 G>A	160,976,489	Intron 26	G>A	2.8
rs6923877	160,967,208	Intron 28	G>A	34.7
rs12207195	160,944,989	Intron 30	G>A	13.8
rs11751605	160,933,641	Intron 34	T>C	15.1
rs3798220	160,931,548	Exon 36	T>C	2.2

∗ on chromosome 6, NCBI build 35

† minor allele frequency calculated for all 2,491 individually genotyped samples from the case/control investigation

3.2 Results from individual genotyping in the MI family study

3.2.1 Characteristics of the family sample

In total 2,292 individuals from 760 families (with up to 11 members) with measured Lp(a) levels were identified in the database and the DNA samples used for TaqMan genotyping.

Baseline characteristics

The clinical characteristics of the family sample are given in table 3.3. The median of Lp(a) levels was 14.4 mg/dl, with a 25% - 75% quantile range of 5.9 to 54.0 mg/dl. The median of square-root transformed Lp(a) levels was 3.80 mg/dl, with a 25% - 75% quantile range of 2.43 to 7.35 mg/dl.

Genotype distribution

The distribution of genotypes obtained by TaqMan genotyping in the family sample is given in table 3.4.

To use the full information from 2,292 genotypes for an association test with Lp(a) levels, the method of family-based association testing had to be applied. Here, the family structure of the sample and therefore the dependence of genotypes, is taken into account. Otherwise only a subsample comprised of one individual per family could have been analyzed.

Table 3.3: Clinical characteristics of the family sample

Variables	Family sample (n=2,292)
Age, y (SD)	58.9 ± 10.7
Women, n (%)	35.3
Body Mass Index, kg/m^2 (SD)	27.0 ± 3.5
Systolic BP, mm Hg (SD)	141.6 ± 21.3
Diastolic BP, mm Hg (SD)	83.8 ± 11.1
Arterial hypertension, n (%)	88.1
Actual hypertension, n (%)	45.4
Diabetes mellitus, n (%)	13.6
Former or current smoking, n (%)	63.8
LDL, mg/dl (SD)	163.4 ± 45.7
Total/HDL cholesterol ratio (SD)	5.0 ± 2.1
Hypercholesterolemia, n (%)	92.1
Myocardial infarction, n (%)	52.1

Table 3.4: Distribution of genotypes in the family sample

SNP	Alleles (1>2)	11 (%)	12 (%)	22 (%)
rs7752408	G>C	40.0	46.6	14.4
+79170 G>A	G>A	95.1	4.8	0.1
rs6923877	G>A	42.4	45.0	12.6
rs12207195	G>A	72.7	25.2	2.2
rs11751605	T>C	69.7	28.2	2.1
rs3798220	T>C	94.7	5.2	0.04

11: homozygous for the major allele, 12: heterozygous, 22: homozygous for the minor allele

3.2.2 Family-based association testing

Two different statistic programs were used to analyze the association between SNP markers and Lp(a) levels in the family sample. The Lp(a) levels were square-root transformed to harmonize the skewed distribution. First the genotype data and Lp(a) levels were analyzed using FBAT (Family-Based Association Tests in genetic analyses)[106]. In this program the -e option allows to test for association in the presence of linkage (see 2.6.1), and this option was used as linkage of the *LPA* gene region and Lp(a) levels had already been shown: The analysis of microsatellite markers (394 microsatellite markers investigated on 499 families with 1,298 individuals of the German MI family study) with the Lp(a) serum levels led to a linkage signal at chromosome 6q27 with a LOD score of 26.99 [81]. Results of the family-based association testing using FBAT are shown in table 3.5. All investigated markers showed significant association with Lp(a) levels. To confirm the FBAT results the program PLINK was used [103]. In PLINK the QFAM procedure allows family-based tests of

association with quantitative phenotypes (see 2.6.1). The results from the PLINK analysis are also shown in table 3.5. With PLINK all investigated SNP markers showed significant association with Lp(a) levels, and P values are well comparable to the FBAT results in size of significance. The significant results from the family-

Table 3.5: Results from family-based association testing with square-root transformed Lp(a) levels as a quantitative trait

SNP marker	FBAT P values *	PLINK P values †
rs3798220	3.4 E-6	1.0 E-5
rs11751605	3.1 E-6	1.0 E-5
rs12207195	8.8 E-3	2.2 E-2
rs6923877	1.8 E-5	1.0 E-5
+79170 G>A	5.6 E-3	3.6 E-3
rs7752408	2.3 E-6	1.0 E-5

* from empirical variance statistic using -*e* option
† empirical P values of QFAM after permutation

based association analysis of SNP markers with Lp(a) levels gave further hints that variations influencing the Lp(a) levels had been identified. To confirm this finding, the SNP genotypes were used in a contribution to linkage analysis, where also the strongest known factor modifying the Lp(a) levels, the KIV-2 repeat polymorphism, was taken into account.

3.2.3 Contribution to the evidence of linkage

To assess the contribution to linkage with Lp(a) levels, individual SNPs were used as covariates in variance-component linkage analysis of 471 sibpairs. The largest change in the maximum LOD score (MLS) occurred with SNP rs11751605. Initially, the peak for Lp(a) had an MLS of 8.9 at 161 cM on chromosome 6q27 in our subset of affected siblings genotyped for this SNP. After adjustment for SNP rs11751605, the LOD score change was a decrease to an MLS of 7.5 following a dominant model (table 3.6).

This decrease did also reach statistical significance ($P<0.05$). The proportion of variance due to SNP rs11751605 as covariate was 5% compared to 44% when KIV-2 repeats were included in the model. As expected inclusion of KIV-2 repeats led to a marked decrease in the MLS to 4.3. In a model that included both SNP rs11751605 and KIV-2 repeats, both covariates were independently associated with Lp(a) levels leading to a decrease in the MLS of 3.6.

Table 3.6: Lp(a) QTL linkage analysis on chromosome 6q27 with SNP rs11751605 and/or KIV-2 included as covariates

Model	Position (cM)	MLS*	Proportion variance due to covariates	*P* Value
crude	161	8.9	-	-
rs11751605 adjusted	160	7.5	5%	<0.00001
KIV-2 adjusted	162	4.3	44%	<0.00001
rs11751605 and KIV-2 adjusted	160	3.6	45%	0.015 and <0.00001

*MLS, maximum LOD score

From this analysis SNP marker rs11751605 had shown the strongest effect on Lp(a) levels of all investigated SNP markers, but further evidence from independent samples was needed. First a sample from the general population was investigated, this allowed also to perform a case-control study on the association of MI and the SNP markers. Second, an independent sample of incidental, non-familial MI cases from the Augsburg area was investigated.

3.3 Case-control association samples

In total four samples were used for case-control association studies. Three case populations were compared to the population-based control sample. KORA MI patients were used as a independent sample of incidental, non-familial MI cases. Two case population were deduced from the Regensburg MI family sample: one case sub-population, referred to as 'Index MI patients', consisted of 595 Index MI patients, the second case population (n=471), referred to as 'affected CAD siblings', consisted of single siblings of index MI patients with documented CAD. This sample of siblings was used for verification.

3.3.1 Genotyping in the population-based sample

As a population-based control sample, subjects from the MONICA Echocardiographic substudy were used (see 2.5). From the 1,674 subjects, both DNA samples and Lp(a) data were available in 975 subjects, which were genotyped. The distribution of genotypes is given in table 3.7.

Table 3.7: Distribution of genotypes in case-control samples

	Alleles (1>2)	MONICA controls n=975 %	Index MI patients n=595 %	affected CAD siblings n=471 %	KORA MI cases n=450 %
rs7752408	G>C				
11		37.7	41.6	40.3	-
12		48.5	44.4	45.9	-
22		13.8	14.0	13.8	-
+79170 G>A	G>A				
11		93.2	95.6	95.1	94.4
12		6.8	4.4	4.7	5.6
22		0.0	0.0	0.002	0.0
rs6923877	G>A				
11		40.7	43.8	44.1	-
12		47.4	44.0	44.8	-
22		11.9	12.2	11.1	-
rs12207195	G>A				
11		75.4	76.1	71.8	-
12		22.0	22.2	26.1	-
22		2.6	2.2	2.1	-
rs11751605	T>C				
11		76.2	68.3	69.2	71.6
12		22.2	29.5	28.4	25.1
22		1.7	2.2	2.4	3.3
rs3798220	T>C				
11		97.4	93.3	93.4	97.5
12		2.6	6.5	6.6	2.3
22		0.0	0.2	0.0	0.2

11: homozygous for the major allele, 12: heterozygous, 22: homozygous for the minor allele

3.3.2 Individual genotyping in the KORA MI sample

This sample comprised MI patients (n=609) from the population-based MONICA MI registry Augsburg who had experienced their first MI before 60 years of age (see 2.5). Of these, 450 blood samples with measured Lp(a) levels were available for genotyping in the present study. In this sample only rs11751605, rs3798220 and +79170 G>A were investigated and genotypes are listed in table 3.7.

3.3.3 Genotype distribution in the subsamples from the family sample for association testing

The genotype distribution in the subsamples of Index MI patients and affected CAD siblings is given in table 3.7. From table 3.7, a marked difference in allele distribution between the samples can be seen for the markers: +79170 G>A, rs11751605, and rs3798220.

Table 3.8: Clinical characteristics of case-control study populations

Variables	MONICA controls (n=975)	Index MI patients (n=595)	affected CAD siblings (n=471)	KORA MI patients (n=450)
Age, y (SD)	49.7 ±14.1	50.0 ±8.4	57.8 ±9.5‡	56.6 ±7.1‡
Women, n (%)	49.7	13.1‡	20.8‡	11.1‡
Body Mass Index, kg/m^2 (SD)	26.6 ±4.0	27.2 ±3.3†	26.9±3.2	28.3 ±3.7‡
Systolic BP, mm Hg (SD)	133.7 ±19.7	139.0 ±19.1‡	144.4 ±21.6‡	132.8 ±17.1
Diastolic BP, mm Hg (SD)	80.4 ±11.6	83.3 ±10.6‡	84.1 ±11.5‡	84.4 ±10.4‡
Arterial hypertension, n (%)	43.7	92.9‡	93.3‡	98.9‡
Actual hypertension, n (%)	37.9	53.7‡	59.2‡	40.8
Diabetes mellitus, n (%)	3.8	14.7‡	15.3‡	15.1‡
Former or current smoking, n (%)	56.9	77.3‡	71.1‡	78.4‡
LDL, mg/dl (SD)	142.3 ±44.0	161.8 ±45.6‡	164.9 ±45.1‡	134.2 ±38.4‡
Total/HDL cholesterol ratio (SD)	4.7 ±1.7	5.2 ±1.7‡	5.0 ±1.7‡	5.3 ±2.1‡
Hypercholesterolemia, n (%)	58.4	95.4‡	93.6‡	75.6‡

† $P < 0.01$ vs. controls;
‡ $P < 0.001$ vs. controls

3.3.4 Clinical characteristics of the case-control study populations

The clinical characteristics of study populations are shown in table 3.8. The relative frequency of female subjects was higher in MONICA controls than in all other study populations. Furthermore, MONICA controls were slightly younger than Index MI patients, and significantly younger than affected CAD siblings and KORA MI patients. In each study population, Lp(a) levels were comparable in men and women. As expected, cardiovascular risk factor distribution was more pronounced in cases than controls. Mean Lp(a) levels were significantly lower in controls (median 11.8 mg/dl, 25% - 75% quantile range: 0.0 to 26.4) compared to Index MI patients (median: 14.8 mg/dl, 25% - 75% quantile range: 5.9 to 64.6), affected CAD siblings (median: 14.5 mg/dl, 25% - 75% quantile range: 5.9 to 51.0) and incident KORA MI patients (median: 16.0 mg/dl, 25% - 75% quantile range: 5.5 to 58.0) (depicted in figure 3.2).

As reported before, the number of expressed KIV-2 repeats showed a strong association with Lp(a) levels across all populations studied, when the number of KIV-2 was dichotomized to ≤21 and >21 KIV-2 repeats, respectively (figure 3.3) [52].

3.3.5 Association of individual SNP markers with Lp(a) levels

A strong association between individual SNP genotypes and Lp(a) levels could be seen. Consistent association in each population was observed for rs11751605, rs3798220 and the +79170 G>A SNP (table 3.9). As SNP markers rs3798220 and +79170 G>A were rare, the maximum number of one homozygous carrier of the rare allele was observed per population. The rare SNP markers rs3798220 and +79170 G>A showed a strong association. The most robust association was revealed for

Table 3.9: SNP genotypes and the corresponding median Lp(a) levels (mg/dl) with 25% to 75% quantile range in all populations

MONICA controls

	11	*12*	*22*
rs7752408	13.5 [1.8-37.6]	11.3 [1.8-23.1]	10.7 [6.4-15.1]
*P Value**		*0.0189*	
+79170 G>A	12.2 [1.8-26.9]	1.8 [0-14.4]	-
P Value		*0.0004*	
rs6923877	11.5 [6.6-16.8]	11.3 [1.8-23.2]	13.1 [1.8-35.1]
P Value		*0.2238*	
rs12207195	11.8 [1.8-28.8]	12.0 [1.8-21.1]	9.1 [0-17.85]
P Value		*0.1383*	
rs11751605	11.5 [1.8-21.5]	13.4 [1.8-48.4]	46.6 [2.5-80.0]
P Value		*0.0024*	
rs3798220	11.6 [1.8-24.7]	101.0 [9.35-133.1]	-
P Value		*<0.0001*	

Index MI patients

	11	*12*	*22*
rs7752408	26.0 [7.3-84.4]	13.9 [5.9-60.5]	9.0 [5.1-15.4]
*P Value**		*<0.0001*	
+79170 G>A	16.0 [6.3-66.8]	5.6 [1.2-16.3]	-
P Value		*0.0002*	
rs6923877	9.7 [5.3-18.5]	13.5 [5.9-60.1]	23.2 [6.2-81.6]
P Value		*0.0004*	
rs12207195	16.0 [6.3-68.3]	12.6 [5.2-47.9]	15.4 [9.2-28.5]
P Value		*0.1874*	
rs11751605	12.4 [5.6-41.3]	48.9 [7.1-80.8]	60.3 [7.9-89.6]
P Value		*<0.0001*	
rs3798220	13.7 [5.6-55.0]	109.5 [85.7-134.1]	168.6 [168.6-168.6]
P Value		*<0.0001*	

affected CAD siblings

	11	*12*	*22*
rs7752408	21.4 [7.0-77.9]	13.7 [5.5-43.6]	9.3 [5.1-16.7]
*P Value**		*0.0004*	
+79170 G>A	14.5 [6.0-51.8]	8.0 [0.6-36.5]	7.4 [7.4-7.4]
P Value		*0.1540*	
rs6923877	9.8 [5.8-18.3]	13.9 [5.2-45.8]	19.1[6.7-74.7]
P Value		*0.0055*	
rs12207195	15.3 [6.1-60.0]	13.7 [4.7-34.3]	7.6 [3.7-33.0]
P Value		*0.1586*	
rs11751605	13.4 [5.8-33.2]	23.7 [6.1-72.5]	87.3 [34.5-109.2]
P Value		*0.0001*	
rs3798220	14.2 [5.7-39.1]	120.6 [87.2-160.0]	-
P Value		*<0.0001*	

KORA MI patients

	11	*12*	*22*
rs7752408	n.d.	n.d.	n.d.
*P Value**			
+79170 G>A	17.0 [6.0-63.5]	2.0 [1.0-13.5]	-
P Value		*<0.0001*	
rs6923877	n.d.	n.d.	n.d.
P Value			
rs12207195	n.d.	n.d.	n.d.
P Value			
rs11751605	14.0 [5.0-34.8]	55.5 [7.0-84.3]	71.0 [29.0-119.0]
P Value		*<0.0001*	
rs3798220	15.0 [5.0-55.0]	118.0 [111.5-156.0]	323.0 [323.0-323.0]
P Value		*<0.0001*	

**P* values from non-parametric Kruskal-Wallis-Test for comparison of Lp(a)levels by genotypes

11: homozygous for the major allele, 12: heterozygous, 22: homozygous for the minor allele, n.d.: not determined

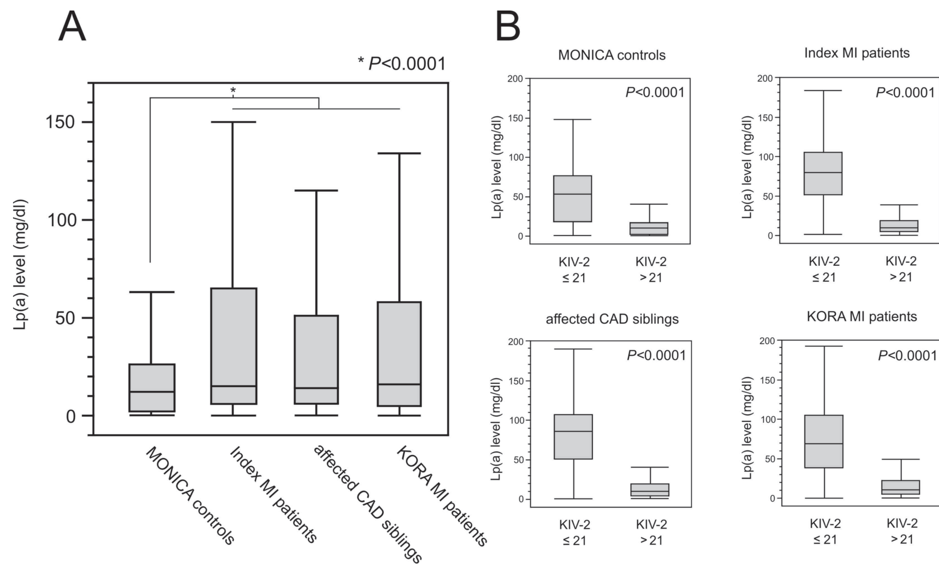

Figure 3.2: A: Comparison of mean Lp(a) levels between control and MI/CAD study populations. Median Lp(a) levels (mg/dl) with 25% to 75% quantile range are shown for each study population. B: Influence of KIV-2 repeat number on Lp(a) levels in each study population. Median Lp(a) levels with 25% to 75% quantile range are given according to KIV-2 repeats ($\leq$ 21 or $>$ 21 repeats).

rs11751605. Specifically, a marked increase of Lp(a) levels with the number of C alleles of rs11751605 could be observed. This increase was consistent in all populations but at higher levels in subjects with MI/CAD.

3.3.6 Interaction between SNP markers and KIV-2 repeats on Lp(a) levels

A test for interaction between the identified SNP markers and the number of KIV-2 repeats was performed (figure 3.3).

In individuals with $\leq$21 KIV-2 repeats (high Lp(a) levels) homozygosity for the C allele of SNP rs11751605 leads to a further increase of Lp(a) levels. This effect was consistent in subjects from the general population as well as in Index MI patients. Interaction analysis for the rare SNP markers rs3798220 and +79170 G>A was not possible, due to the small number of heterozygous individuals and the single observations of homozygous carriers of the rare allele.

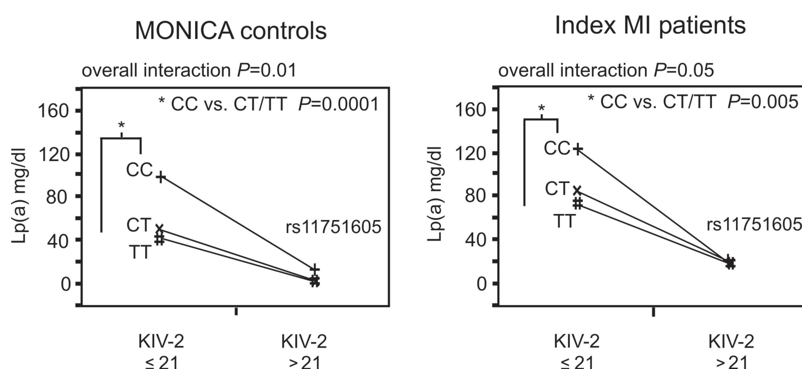

Figure 3.3: Interaction between KIV-2 repeat number and rs11751605 genotype and the influence on Lp(a) levels. The interaction plots display Lp(a) levels of subjects with ≤ 21 and > 21 KIV-2 repeats on the x-axis with separate lines for the means of each genotype of SNP rs11751605 (CC genotype indicated by '+', CT by '×' and TT by '♯'). Significant overall interactions between the SNP and the KIV-2 repeat number could be revealed in controls from the MONICA general population (left) and in Index MI cases (right). Homozygosity for the C allele of SNP rs11751605 leads to a significant increase in Lp(a) levels compared to CT or TT allele carriers when ≤ 21 KIV-2 repeats are present, also in both populations.

3.3.7 Association of SNP markers with MI

SNP marker rs11751605 was tested for association with MI and/or CAD. A significant difference (P=0.001) in the allele frequency of the minor allele (C allele) was observed between healthy subjects from the general MONICA population (12.8%) and Index MI patients (17.0%), with an odds ratio (OR) of 1.4; 95% confidence interval (CI) 1.14 to 1.70. Using a dominant model, a strong association of MI with C alleles of rs11751605 (OR, 1.48; 95% CI, 1.18 to 1.86, P=0.0008) was found. When adjusting for the potential confounders, age, sex, BMI, systolic blood pressure, diabetes mellitus, smoking, and total/HDL cholesterol ratio, the association remained significant (OR, 1.58; 95% CI, 1.22 to 2.05, P=0.0006). After additional inclusion of the square root transformed Lp(a) levels, a trend for association remained, though the level of significance could not be reached (OR, 1.28;95% CI, 0.97 to 1.68, P=0.079, Figure 3.4). Due to age and gender differences between Index MI patients and MONICA controls additionally an age- and sex matched sample was analyzed, including 473 Index MI patients and 473 MONICA controls. Here the risk increase by the C-allele of rs11751605 was comparable to the overall sample (OR, 1.65; 95% CI, 1.24 to 2.22, P=0.0008) using conditional logistic regression. Similar to the overall sample the risk increase for MI remained when accounting for cardiovascular risk factors and square root transformed Lp(a) levels (OR, 1.36; 95% CI, 0.99 to

1.89). These effects could be confirmed using KORA MI patients and affected CAD siblings (figure 3.4).

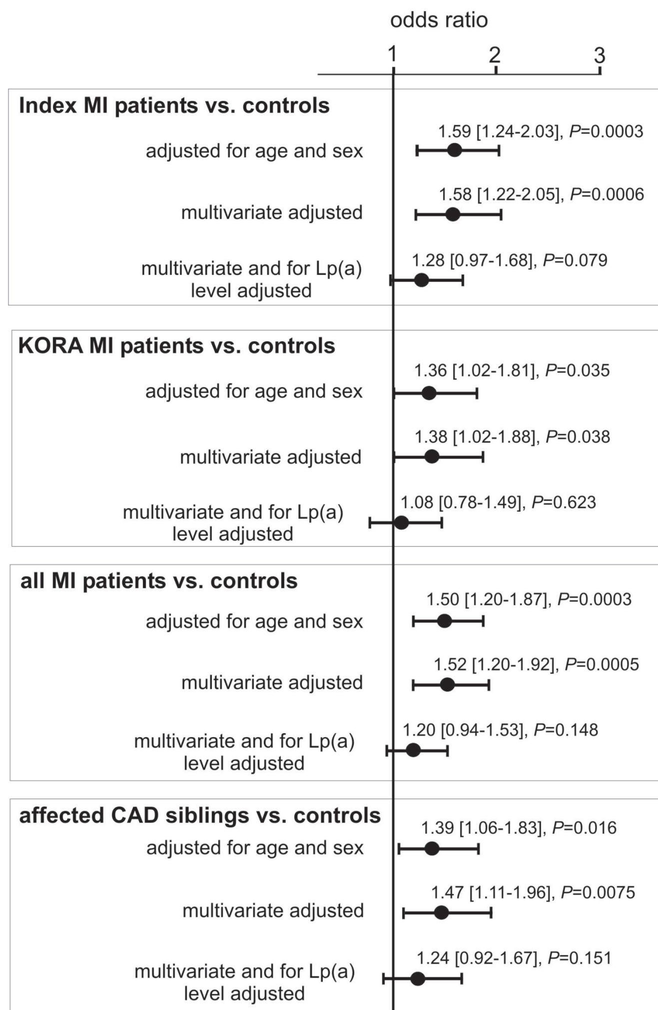

Figure 3.4: Odds ratios from logistic regression analysis for MI and CAD according to rs11751605 genotypes. Odds ratios with 95% CI and P values are given as three different adjustments of the dominant model (CC+CT vs. TT): adjustment for sex and age, multivariate adjustment (age, sex, BMI, systolic blood pressure, diabetes mellitus, smoking, total cholesterol/HDL ratio), and multivariate adjustment additionally including square root transformed Lp(a) levels.

For marker rs3798220 an association with MI could be detected in allele frequency of the rare C allele between MONICA controls and Index MI patients (P=0.0002). This association remained significant after adjustment for the risk factors, age, sex, BMI, systolic blood pressure, diabetes mellitus, smoking, and total/HDL cholesterol ratio (OR, 2.74; 95% CI, 1.54 to 4.98, P=0.0007). After additional inclusion of the square root transformed Lp(a) levels, the association did not remain, and the level

of significance was clearly failed (OR, 1.28; 95% CI, 0.68 to 2.44, P=0.43). The association of rs3798220 with CAD could also be seen in comparison of MONICA controls with affected CAD siblings (P=0.0004; OR, 2.66; 95% CI, 1.56 to 4.60) and the association remained significant (P=0.01; OR, 2.26; 95% CI, 1.12 to 4.23) after risk factor adjustment (risk factors as given above). When KORA MI cases were compared with MONICA controls, rs3798220 failed to show association with MI (P=0.9; OR, 0.96; 95% CI, 0.45 to 1.92). Therefore, no consistent association of rs3798220 genotype with MI could be shown.

3.4 Linkage disequilibrium analysis and Hardy-Weinberg proportions

SNP markers are transmitted in blocks, a measure of this block structure is LD, the linkage disequilibrium [112]. Between two SNP markers this can be expressed by the correlation coefficient r^2 [112]. LD between the investigated SNP markers was analyzed (in terms of r^2) and a two block structure was found, that reflects the LD structure of HapMap in CEPH samples, as shown in figure 3.1 [113, 114]. These data indicate that rs7752408, SNP +79170 G>A, and rs6923877 are in LD with the KIV-2 repeat polymorphism. In contrast, there was only marginal LD between rs11751605, rs12207195 and the KIV-2 repeat polymorphism. All SNPs were tested for Hardy-Weinberg proportions and all genotypes fulfilled Hardy-Weinberg expectations. Deviations from HW equilibrium can indicate inbreeding, population stratification or problems in genotyping [115].

3.5 Genome-wide association testing

3.5.1 Characteristics of the subsample with GWA data

For a subsample (n=487) of the Cardiogenics patients with GWA data, Lp(a) levels were available. These samples had been genotyped with GeneChip Human Mapping 500K Array Set from Affymetrix (comprised of two arrays: *StyI* and *NspI*). The baseline characteristics of this Cardiogenics sub-sample are given in table 3.10. The sample was comprised of MI patients only. The median Lp(a) level was 16.0 mg/dl, with a 25% - 75% quantile range of 6.8 to 65.9 mg/dl.

Table 3.10: Clinical characteristics of the Cardiogenics sub-sample

Variables	Cardiogenics sub-sample (n=487)
Age, y (SD)	50.2 ± 7.9
Women, n (%)	33.5
Body Mass Index, kg/m^2 (SD)	27.3 ± 3.5
Systolic BP, mm Hg (SD)	142.3 ± 21.0
Diastolic BP, mm Hg (SD)	84.2 ± 10.5
Arterial hypertension, n (%)	92.3
Actual hypertension, n (%)	47.8
Diabetes mellitus, n (%)	16.1
Former or current smoking, n (%)	70.4
LDL, mg/dl (SD)	162.2 ± 46.2
Total/HDL cholesterol ratio (SD)	5.1 ± 1.7
Hypercholesterolemia, n (%)	96.1

3.5.2 Quantitative association analysis

The genome-wide data were analyzed using PLINK. A quantitative association analysis was performed using square-root transformed Lp(a) levels. From the 262,314 SNP markers of the *NspI* array, 12,725 SNPs failed the missingness test (excluding SNP with more than 10% missing genotypes) and 33,105 SNPs failed the frequency test (excluding SNPs with a minor allele frequency below 0.01). The remaining 216,546 SNP markers from the *NspI* array were used for the quantitative association analysis in PLINK, the total genotyping rate was 97.8%. For the 238,354 markers of the *StyI* array, 19,212 were excluded for low genotyping rate and 26,694 SNPs were excluded for low minor allele frequency. For the analysis, 189,456 markers from the *StyI* array were used, resulting in a genotyping rate of 97.2%.

Figure 3.5 displays graphically the results from the quantitative association of the SNP markers with square-root transformed Lp(a) levels. The plot was created using Haploview 4.0.

For chromosome 6 a peak of SNP markers that is highly associated with square-root transformed Lp(a) levels can be seen. These strong association signals emerge from the region around the *LPA* gene. Marker SNP_A-4285662 (= rs2048327), located within the *SLC22A3* gene, a neighbor of the *LPA* gene, showed the strongest association with square-root transformed Lp(a) levels providing a *P* value of 2.4 E-10. This *P* value can be considered as genome-wide significant, when using the Bonferroni approach of dividing the commonly used *P* value of 0.05 by the number of tested markers, in this case: 500,000. When a threshold of genome-wide significance of 1.0 E-7 is applied, eight SNP markers, all markers located in the *LPA* gene region, show genome-wide significant association with square-root transformed Lp(a) levels.

In total 3,759 SNP markers showed association with square-root transformed

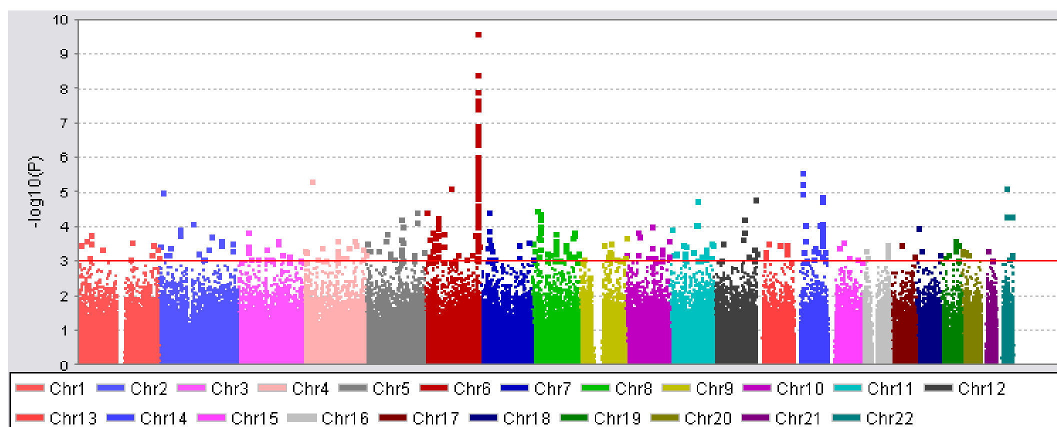

Figure 3.5: P values on the y-axis are displayed as minus log P, chromosome number is given on the x-axis.

Lp(a) levels with a P value below 0.05. Of these markers, 379 SNP markers showed association with a P value below 0.001. 99 of these 379 markers are located on chromosome 6, but also on several other chromosomes, as chromosome 8 (36 markers) and chromosome 11 (23 markers) associated markers were identified.

3.5.3 Results for the SNPs from the *LPA* gene region

A detailed view on the results from the genome-wide association for the *LPA* gene region is given in figure 3.6.

Only 13 SNPs located within the *LPA* gene show association with square-root transformed Lp(a) levels, of 18 SNPs (from both arrays) located within the *LPA* gene. Of these 18 SNPs only 11 SNPs had good cluster plots, which means that the genotype clusters could be distinguished properly. Therefore, SNP markers from the GeneChip Human Mapping 500K Array Set did not strongly improve the SNP coverage within the *LPA* gene itself.

The strongest associated SNPs are located in three haploblocks (indicated by $*$ in figure 3.6). These three haploblocks originate in the *LPA* gene itself, as indicated by the HapMap data. LD analysis of the SNP markers from the GeneChip Human Mapping 500K Array Set in the *LPA* gene region showed a distribution of haploblocks equivalent to the HapMap data. Furthermore, an interaction analysis between SNP markers of the 500K array and the KIV-2 repeats revealed significant interaction of SNP markers in the middle of the indicated ($*$) blocks and the KIV-2 repeat polymorphism (figure 3.6). No interaction was found between SNPs located

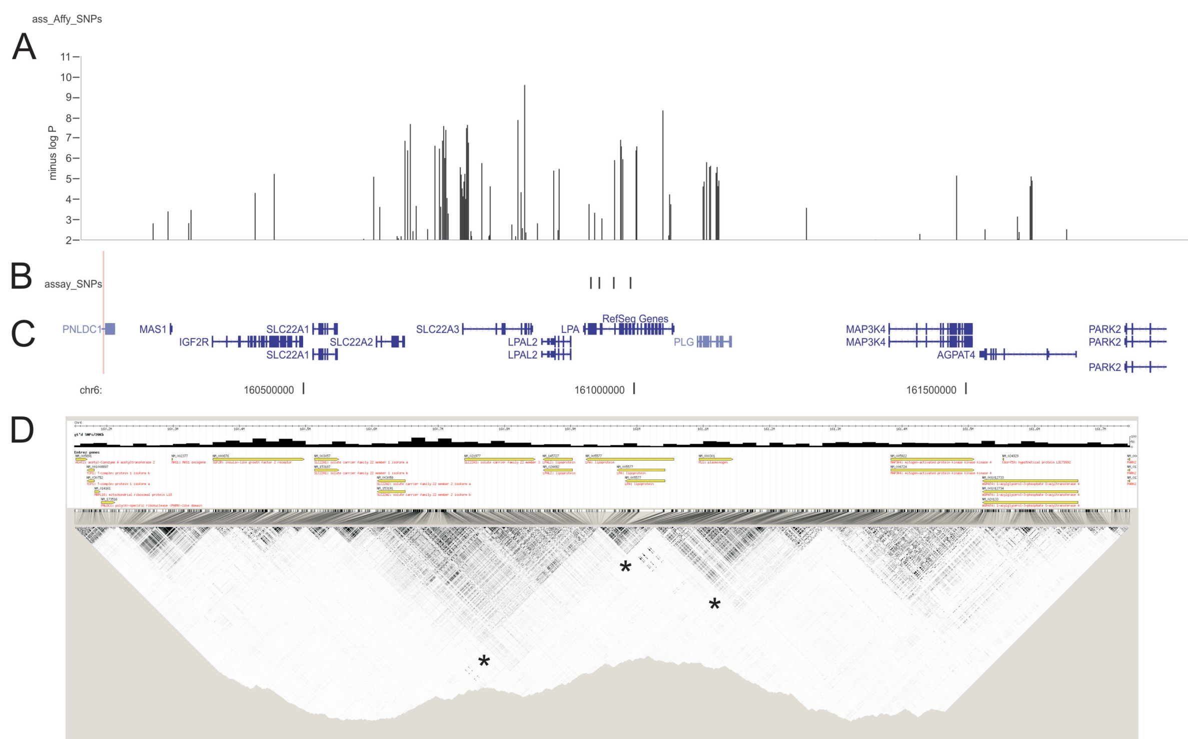

Figure 3.6: A: P values of association for markers of the *LPA* gene region from the 500K chip with square-root transformed Lp(a) levels are displayed as minus log P on the y-axis. B: Position of SNP markers investigated with TaqMan assays: rs11751605, rs12207195, rs6923877 and rs7752408 (from left to right). C: Position of RefSeq genes in the *LPA* gene region, displayed in UCSC March 2006 assembly, Chr6: 160,000,000-162,000,000. D: Representation of LD structure in the *LPA* gene region, pairwise r^2 values between markers from CEU samples of HapMap project II phase release 21 are shown with darker regions representing higher LD. The three haploblocks of the *LPA* gene are indicated by $*$

in the left of the indicated haploblocks and the KIV-2 repeat polymorphism (figure 3.6).

3.5.4 Adjustment of the GWA data for the significant SNPs from the *LPA* gene region

Linkage analysis identified several other loci of the genome that have been linked to Lp(a) levels [81, 82]. Previous linkage data from our family study indicate that a locus on chromosome 1 (closest micro-satellite marker D1S1679, with a LOD score of 3.81) is also linked to Lp(a) levels [81]. Other, multiple QTL (apart from *LPA*) influencing the serum Lp(a) concentration have been identified on chromosomes 11,

13, 15 and 19 in an other genome-wide linkage screen [82]. These QTL regions were identified after taking into account the strong effect detected on chromosome 6 [81, 82]. For this reason an adjustment for those SNP markers giving the strong association result in the *LPA* gene region was performed on the genome-wide SNP data. To allow a computable adjustment of the huge genome-wide data set, a subset of tagging SNPs (representing the three haploblocks in the *LPA* gene region) had to be selected.

Identification of tagging SNPs in the *LPA* gene region

A tagging SNP represents the information from the neighboring SNP markers when they are transmitted together in a block of markers. The selection of tagging SNPs was performed from HapMap data of build 36. The genotype data of the *LPA* gene region on chr6: 160,650,001-161,280,000 on NCBI build 36 were downloaded and analyzed in Haploview, where the three LD block structure (shown in figure 3.6) appeared. In Haploview, those SNP markers with an LD of $r^2 > 0.8$ were selected with pairwise tagging. The list of 708 SNP markers was exported and matched with a list of HapMap SNP markers present on the GeneChip Human Mapping 500K Array Set and 88 possible tagging SNPs remained. With this large number of SNPs an adjustment using PLINK was not possible and the number of SNPs was manually reduced to 31 SNPs, 13 from the *NspI* array and 18 from the *StyI* array, listed in table 3.11. The distribution of selected tagging SNPs is depicted in figure 3.7.

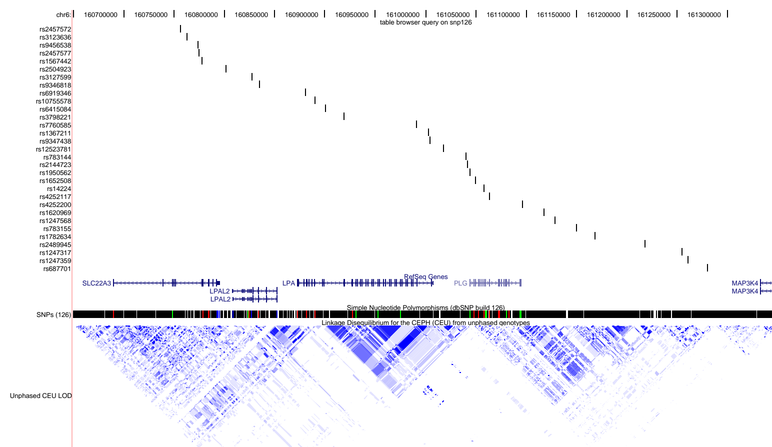

Figure 3.7: Distribution of tagging SNPs. The position of 31 selected tagging SNPs, that capture the three LD blocks underlying the *LPA* gene are shown in UCSC genome browser on build 36.

Table 3.11: List of tagging SNPs

Affymetrix SNP identifier	500K Chip	absolute position on Chr 6*	rs number†
SNP_A−2183575	*StyI*	160,756,396	rs2457572
SNP_A−4299085	*StyI*	160,762,527	rs3123636
SNP_A−1803733	*StyI*	160,773,383	rs9456538
SNP_A−2248891	*StyI*	160,774,128	rs2457577
SNP_A−1988377	*NspI*	160,777,305	rs1567442
SNP_A−2259393	*NspI*	160,801,498	rs2504923
SNP_A−1961463	*StyI*	160,827,124	rs3127599
SNP_A−1826864	*StyI*	160,834,450	rs9346818
SNP_A−2272818	*StyI*	160,880,349	rs6919346
SNP_A−1795611	*NspI*	160,889,728	rs10755578
SNP_A−2264009	*StyI*	160,900,320	rs6415084
SNP_A−4255889	*StyI*	160,918,138	rs3798221
SNP_A−2237752	*StyI*	160,990,860	rs7760585
SNP_A−2180871	*NspI*	161,002,685	rs1367211
SNP_A−2055441	*NspI*	161,003,623	rs9347438
SNP_A−2254455	*StyI*	161,017,268	rs12523781
SNP_A−1824552	*NspI*	161,039,229	rs783144
SNP_A−1831715	*NspI*	161,041,380	rs2144723
SNP_A−4266065	*StyI*	161,043,175	rs1950562
SNP_A−4227819	*NspI*	161,048,937	rs1652508
SNP_A−4201898	*NspI*	161,057,769	rs14224
SNP_A−2012234	*StyI*	161,063,366	rs4252117
SNP_A−1898098	*StyI*	161,095,711	rs4252200
SNP_A−2244463	*NspI*	161,117,058	rs1620969
SNP_A−2100761	*StyI*	161,128,376	rs1247568
SNP_A−1830392	*StyI*	161,149,787	rs783155
SNP_A−4212106	*NspI*	161,168,127	rs1782634
SNP_A−1943498	*StyI*	161,217,714	rs2489945
SNP_A−2128791	*StyI*	161,254,404	rs1247317
SNP_A−2124265	*NspI*	161,260,672	rs1247359
SNP_A−2054021	*NspI*	161,279,548	rs687701

*UCSC March 2006 assembly
† dbSNP build 126

The adjustment procedure was performed with PLINK. A linear model association was calculated with square-root transformed Lp(a) levels, using the selected 31 tagging SNPs in a condition list. The results were displayed using Haploview 4.0 and are depicted in figure 3.8 for chromosome 6 and genome-wide in figure 3.9.

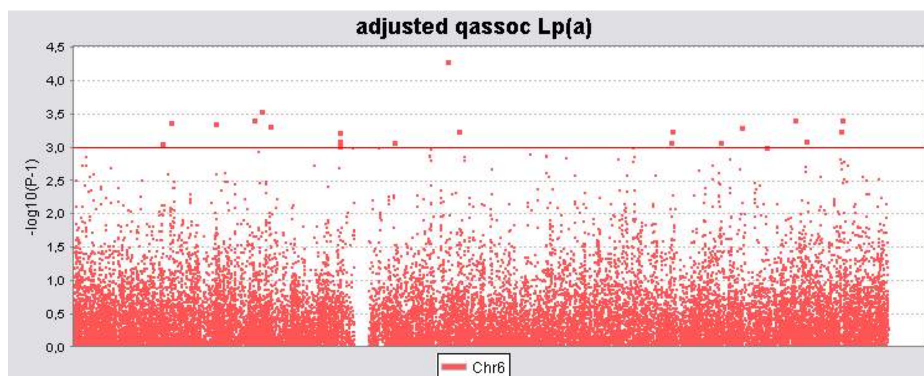

Figure 3.8: Results for chromosome 6 from genome-wide association of Lp(a) levels, adjusted for *LPA* tagging SNPs

For chromosome 6 the significant association signals for association with square-root transformed Lp(a) levels from the *LPA* gene region have vanished. After adjustment the *P* values for genome-wide association with square-root transformed Lp(a) levels dropped markedly, reaching the maximum of 2.16 E-5 for the marker SNP_A-2219110 (= rs10017198) on chromosome 4. This marker does not map to a gene, the closest gene is *DHX15*, a member of the DEAD box protein family and putative ATP-dependent RNA helicase. In total 388 SNP markers provided association signals with *P* values below 0.001. Most of these SNPs (40) are located on chromosome 8. The number of markers on the chromosomes with described linkage is lower: only 24 of these SNP markers are located on chromosome 1, 15 markers are on chromosome 11, 11 markers on chromosome 13, 8 markers on chromosome 15 and 5 markers on chromosome 19.

3.5.5 Regions of interest for further association analysis

For the chromosome 1 region of the previously identified QTL on serum Lp(a) levels, no enhancement of association results could be obtained through the adjustment. In this region (chromosome 1, 156,380,000-161,800,000 on NCBI build 35), five SNP markers were associated (*P* between 0.001 and 0.0001) before the adjustment and only two SNP markers (*P* between 0.001 and 0.0001) after the adjustment. One of

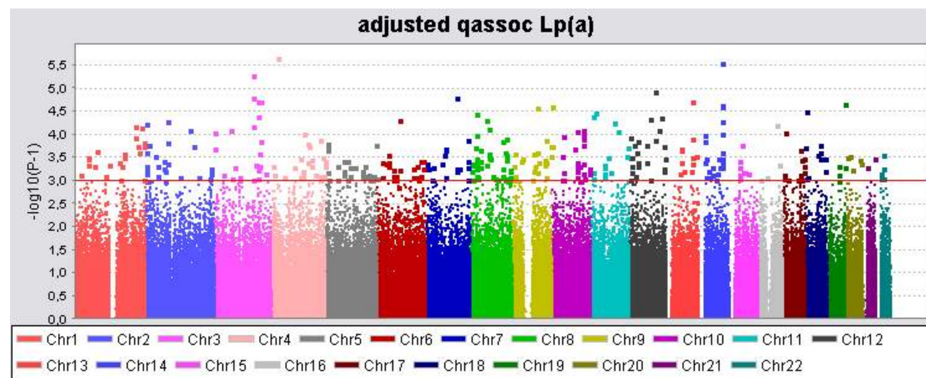

Figure 3.9: Results from genome-wide association of Lp(a) levels adjusted for *LPA* tagging SNPs

the markers, SNP_A-1860563 (= rs400214), showed association in the unadjusted analysis as well as in the adjusted analysis. This marker is located near the *CD1D* gene, encoding a member of the CD1 family of transmembrane glycoproteins. These association results have not been replicated so far. Other promising regions, including markers located in regions with previously described linkage (on chromosomes 11, 13, 15 and 19) have to be analyzed in detail.

4 Discussion

Lp(a) levels are highly heritable (heritability >90%[81]) and the apo(a) locus on chromosome 6q27 has been proven in several linkage studies as the major determinant of Lp(a) serum levels [81, 82]. Elevated Lp(a) levels have been identified as a risk factor for atherosclerosis and myocardial infarction [70]. Consequently, polymorphisms of the *LPA* gene have been identified and explain one part of the variation in Lp(a) levels in Caucasians [73]. Nevertheless, a significant proportion of the genetic variation of Lp(a) levels remains unexplained.

4.1 Association of single SNP markers with MI and Lp(a) levels

For complex diseases, investigators mapped the phenotype of interest first by linkage analysis, followed by association analysis in regions with evidence for linkage [116, 117, 118]. A pooling strategy has been proposed as a way to identify meaningful allele frequency differences in regions of linkage [96, 119], and was successfully used in the present investigation to identify SNP markers associated with Lp(a) serum levels. A positive replication of the findings from the pooling approach in different samples underlines the robustness of the results. Therefore, it can be stated, that a pooling approach using identical-by-descent (IBD)-sharing information from a linkage analysis is a cost effective and practical way to identify relevant SNP markers by differences in allele frequency.

In particular, association of the SNP rs11751605 could be detected with elevated Lp(a) serum levels in patients with a history of premature MI, in MI patients without such a history, and in subjects from a healthy population based sample as well. Moreover, a strong association of the C allele of rs11751605 and MI and CAD independently from traditional cardiovascular risk factors across these populations was found.

The influence of the KIV-2 repeat polymorphism was considered in the present in-

vestigation, by genotying individuals with known apo(a) isoform size (corresponding to the number of expressed KIV-2 repeats), as this polymorphism has been described to account for about 50% of the variability of Lp(a) levels [120, 28]. It turned out, that both polymorphisms (KIV-2 and rs11751605) act somewhat independently as revealed by LD mapping, interaction analysis, and the observation of an additive contribution to the evidence of linkage. On the other hand, the present study clearly corroborates the KIV-2 repeat polymorphism as the major contributor of Lp(a) concentrations.

4.1.1 Influence of previously described rare SNP markers

An association of the rare A allele of the +79170 G>A SNP was observed with lower Lp(a) levels. Only 5.6 % of the individuals were heterozygous for this SNP and only one individual out of 2,491 (0.04%) was a homozygous carrier of the A allele. From the literature higher allele frequencies had been expected. Ogorelkova et al. described a G to A substitution at the +1 donor splice site of the apo(a) KIV-8 intron that results in alternative splicing of the intron and thus encodes a truncated apo(a) form [40]. The shorter apo(a) form is missing the KIV-9 domain (as well as KIV-10, KV and the protease domain) which is essential to form a stable covalent Lp(a) complex. The authors concluded from their genotyping results in a Tyrolean (n=113) and a Finn (n=126) population that approximately 11% of the Europeans are heterozygous and 0.3% are homozygous for the G to A splice site mutation; the one homozygous individual had an extremely low Lp(a) level (0.03 mg/dl) [40]. In the present investigation of 2,491 individuals only one homozygous individual was found (=0.04%), with this individual having a Lp(a) level of 7.4 mg/dl. 5.6% of the individuals were heterozygous and had a lower (mean Lp(a) 15.6 mg/dl) Lp(a) level than individuals homozygous for the frequent allele (mean Lp(a) 32.2 mg/dl). While an association of the minor A allele with lower Lp(a) levels could be confirmed, this mutation can not explain the variability of Lp(a) levels on the population level, due to its very low allele frequency.

For the +79170 G>A SNP, but not for any other SNP analyzed, a strong association ($P < 0.001$) in all populations with the presence of only one detectable apo(a) isoform size in serum was found. The KIV-2 repeat polymorphism was not genotyped but the apo(a) isoform size in serum was measured. The number of expressed KIV-2 repeats was estimated with a standard [52]. In the present study 48% of the persons analyzed had only one detectable apo(a) isoform size. The mean Lp(a) level

of individuals with only one expressed isoform of 29.2 mg/dl is significantly lower than the mean Lp(a) of 33.5 mg/dl for individuals with two different sized apo(a) isoforms, but the distribution of KIV-2 repeat number ranged randomly from 2 to 45 repeats in both groups. In the group of individuals where only apo(a) isoform of one size could be detected the bigger part is due the expression of only one *LPA* allele. However, this group includes homozygotes for the same number of alleles, as well as individuals with one very large isoform, where concentration is low and blotting of this large molecules is often incomplete and thus detection difficult. But again this very rare SNP (minor allele frequency 2.8%) can only explain a small proportion of these "not expressed alleles". The +79170 G>A SNP was investigated to exclude an influenc of this SNP on the effects seen for rs11751605 on the Lp(a) level. The influence of rs11751605 on the Lp(a) level for the different +79170 G>A SNP genotypes was analyzed in all populations and higher Lp(a) levels were found for rs11751605 CC/CT allele carriers independent from the +79170 G>A SNP genotype. This effect reached significance only in the +79170 G homozygous as the number of heterozygous individuals was small.

The other investigated rare SNP marker rs3798220, was identified in a search for genetic variants associated with severe coronary artery disease in a population of Americans with European descent [93]. The variant was identified from 12,077 putative functional SNPs in more than 7,000 genes in a three stage design (with 1,806 cases and 1,274 controls in total) and was the only investigated SNP marker that showed significant association with severe CAD in that study [93]. The alleles of the SNP marker rs3798220 were given by Luke at al. as A>G [93]. They detected a MAF of 5.2% in the (third) case sample and a MAF of 2.7% in the (third) control sample [93]. These allele frequencies are well comparable to the MAF determined in the present investigation, though nomenclature differed. Here, the alleles of rs3798220 are given according to dbSNP as T>C. For the both samples from the MI family study (Index MI patients and affected CAD siblings) a MAF (for the C allele) of 6.6% was detected. The MONICA control sample showed a MAF of 2.6% and a nearly identical MAF of 2.5% was found in the KORA MI cases sample. According to the similar MAFs, also the obtained odds ratios were comparable. Luke at al. observed an odds ratio for severe CAD of 3.14 (95% CI 1.51 to 6.56, P=0.005) for carriers of the minor allele after adjusting for traditional risk factors (age, sex, smoking, hypertension, diabetes, dyslipidemia and BMI) [93]. In the family samples an adjusted odds ratio of 2.26 (95% CI 1.21 to 4.23 and P=0.01) was obtained

for CAD, and an odds ratio of 2.74 (95% CI, 1.54 to 4.98, P=0.0007) for MI was obtained for carriers of the minor allele when compared to MONICA controls. In contrast to these findings, no association of rs3798220 could be detected in the sample of incidental KORA MI cases, when compared to the control sample. No obvious reasons (e.g. bad genotyping performance) can explain this non-replication in the KORA sample.

Therefore, no consistent replication of the association of rs3798220 with MI/CAD could be shown in the present investigation.

Luke et al. investigated also association of rs3798220 with plasma Lp(a) levels in 646 subjects and found significantly higher Lp(a) levels (P=0.003) in carriers of the minor allele [93]. Furthermore, they demonstrated in a sub-sample of these subjects (n=122) that this effect was not due to apo(a) size differences, as the association with Lp(a) levels (P=0.002, median Lp(a) level 5.9 fold higher for rare allele carriers) remained significant (P=0.01, median Lp(a) level 3.7 fold higher for rare allele carriers) after adjustment for apo(a) size [93].

In the present investigation, the association of rs3798220 with Lp(a) levels could be confirmed in all MI/CAD samples and in the general population sample (MONICA controls). The effect of rs3798220 on Lp(a) levels was well comparable to the effect size reported by Luke et al. In the present investigation, carriers of the minor allele had in all populations about 8 fold higher median Lp(a) levels than non-carriers.

However, as in the case of +79170 G>A, rs3798220 can not explain the variability of Lp(a) levels on the population level, due to its low allele frequency.

A mechanism for the influence of rs3798220 on Lp(a) levels and CAD could not be shown so far. However, Luke et al. proposed a possible mechanism for the influence of rs3798220 on Lp(a) levels and CAD risk. From the location of the SNP rs3798220 in the protease domain of *LPA*, and the putative resulting amino acid change from isoleucine (I) to methionine (M), Luke et al. proposed a possible role of this amino acid change on Lp(a) catabolism and oxidation status [93]. They speculated that the oxidation of this methionine residue could alter the properties of apo(a) and Lp(a), in analogue to apolipoprotein A-I, where it has been shown that the oxidation of methionine residues can alter sites and rates of the proteolytic cleavage of apolipoprotein A-I [93].

Taken together, the finding that two neighboring SNPs, rs11751605 (located in intron 34) and rs3798220 (located in exon 36), show association with Lp(a) levels and with MI/CAD, give further hints that mutations in the protease domain of the

LPA gene could have a strong influence on the catabolism of Lp(a), as well as on its atherogenic properties.

4.1.2 Possible influences of rs11751605 on functional properties of Lp(a)

The mechanism by which rs11751605 influences Lp(a) concentrations, thereby acting detrimental on the process of atherosclerosis, remains speculative. A direct causative effect of the SNP appears unlikely due to its genomic location in intron 34 (figure 3.1). Potentially, the marker rs11751605 tags another functional mutation in one of the neighboring domains of KV domain (encoded by exons 32 and 33) or the serin protease domain (encoded by exons 34 to 39) that alters the properties of the apo(a) molecule. An influence on the receptor binding properties of the molecule might be conceivable, thus increasing the levels of circulating Lp(a) particles. Lp(a) is degraded in vivo by elastases, dividing apo(a) in two parts, with the C-terminal domain spanning from KIV-5 to the protease domain [26]. Recent studies in transgenic mice indicate that this C-terminal apo(a) component mediates Lp(a) clearance from plasma by the liver [34]. The apo(a) clearance has been shown to be independent from LDL receptor or apoE and is supposed to run via a receptor that is responsible for the hepatic clearance of cholesterol-rich remnant lipoproteins (RLP) [34, 79]. Enhanced levels of RLP might further increase atherosclerosis [79]. Edelstein et al. found that lysine residues of the KV domain of apo(a) mediate binding with oxidized phosphatidylcholine and postulated that apo(a) might act as a scavenger and prevent LDL from oxidation [13]. Recent findings by Tsimikas et al. showed that Lp(a) levels strongly correlate with the oxidized phospholipid/apoB-100 ratio in plasma and that both factors are associated with the extend of angiographically documented coronary artery disease [42]. In conjunction with the finding that the oxidized phospholipids are physically associated with Lp(a) via the KV domain of apo(a)[13], this supports the idea that in a setting of enhanced oxidative stress the atherogenic properties of Lp(a) might be enhanced through pro-inflammatory oxidized phospholipids. Prolonged circulation of Lp(a) due to delayed clearance in a setting of enhanced oxidative stress might therefore strongly promote atherosclerosis. SNP rs11751605 is located in the vicinity of the KV domain and linkage disequilibrium with this functional site exists according to HapMap data.
On the other hand, alternative hypotheses to explain the detrimental properties of

Lp(a) on atherosclerosis exist [121, 122, 37, 123]. Both, independent confirmation of the association results and functional studies are certainly warranted.
A limitation of the present study is that the Lp(a) levels have not been measured with an assay that is independent for size. However, this issue can considered less relevant as the expressed apo(a) sizes were determined. Furthermore in the present study the influence of polymorphisms of the promoter region of the *LPA* gene was not considered. Promotor polymorphisms (for example the pentanucleotide repeat (PNR) polymorphism) have been repeatedly associated with Lp(a) concentrations [56, 87, 50]. For Caucasian populations the promoter polymorphisms have been shown to be in strong LD with the KIV-2 polymorphism [50, 89, 28] and thus, might themselves not contribute independently to the overall variation of Lp(a). Furthermore it has to be stated that all SNP markers in the present investigation were analyzed only in Caucasian populations and therefore the described associations with Lp(a) levels and MI/CAD are restricted to this ethnicity.

4.2 Genome-wide association data

The genome-wide association of SNP markers from the GeneChip Human Mapping 500K Array Set with Lp(a) levels in the present investigation, clearly proved the huge influence of the *LPA* gene region on the Lp(a) levels. An association of SNP markers within the *LPA* gene region with square-root transformed Lp(a) levels providing P values of up to 2.4 E-10 could be shown. P values in this order of magnitude can be considered as genome-wide significant, even when the threshold of 1.0 E-7 resulting from the conservative Bonferroni approach is applied [124]. The Bonferroni approach is widely used and accepted to address multiple testing in association studies. While for an association analysis testing 5 SNP markers the Bonferroni approach changes the initial level of significance of 0.05 to 0.01, the level of significance drops to 1.0 E-7 for the testing of 500,000 SNP markers [124]. The level of significance in GWA studies has been matter of discussion, as the power to detect P values of 1.0 E-7 can be limited in studies with a small sample size [124]. Therefore a specific result from a GWA study is more emphasized by consistent findings in several replication studies that by the mere strength of the P value in the initial study [124]. A robust replication strategy thereby also reduces the false positive results [124].

Quite disappointing, the GWA analysis could not enhance SNP coverage of the *LPA* gene itself. Though the number of 500,000 SNP markers represented on the chip

used for genotyping is astonishing, on gene levels the SNP coverage shows gaps and a substantial proportion of the known SNPs are not captured [124]. This problem is widely known and manufacturers have answered with the next generation chips, providing information on nearly 1 million SNPs [124]. Furthermore the gaps in SNP coverage are strongly embroidered when SNP markers have to be excluded from the analysis due to a high missingness rate. The quality of the genotyping procedure is the determining factor and the suboptimal call rates of 97.8% and 97.2% in the present investigation as well as the low intensity cluster plots surely do not prove a perfect quality of the genotyping procedure. It should be noted that in the present investigation the data of a already performed GWA analysis were re-analyzed and therefore, the quality of the genotyping procedure could not be influenced.

Though GWA analysis offers the possibility to identify genes previously unsuspected of being related to a phenotype or disease [124], in the present investigation of factors influencing the Lp(a) levels a different strategy could be more promising.

Linkage analysis has led to the initial finding of the locus on chromosome 6q27 as the major determinant of Lp(a) levels. However, several other loci with significant LOD scores were identified when the effect of the chromosome 6 locus was taken into account [81, 82].

An investigation of the genome-wide SNP data in a region of chromosome 1 that had previously shown linkage with Lp(a) levels [81] failed to give clear evidence of a second “hot spot” of SNP markers associated with Lp(a) levels. Also an adjustment for the highly significant associated SNPs in the *LPA* gene region did not improve the association signals for the chromosome 1 region. In part these findings may be explained by gaps in the SNP coverage. A recent investigation by Barlera et al. investigated linkage of Lp(a) in a large sample of 4,012 individuals from 1,812 families. The sample was comprised of European CAD patients and families collected through the PROCARDIS coronary heart disease study [82]. They identified several loci with significant (chromosome 13q22-31 and 11p15) and tentative (chromosome 15q23-25 and 19q13.4) linkage, after the effect of the chromosome 6 locus had been considered [82]. This study also could show linkage for the locus on chromosome 1q23 (previously identified by Broeckel et al.), but with a not significant LOD score of 1.5 [82].

Though the first attempt to identify SNP markers associated with Lp(a) levels from the GWA analysis in regions of linkage has failed, the idea of narrowing the huge genomic regions underlying QTL regions with the use of genome-wide SNP

data to identify appropriate candidate genes may still be promising and should be applied in the future work on identification of modulators of Lp(a) levels.

An other aspect that could be considered in the search for modulators of the Lp(a) levels was recently provided by the findings of Becker et al. [77]. They reported the existence of a extracellular Lp(a) oxidase that catalyzes the covalent Lp(a) assembly [77]. Variances of this Lp(a) oxidase may have a profound influence on Lp(a) levels. Therefore, a screening for genes harboring a oxidase motif in the regions of linkage or association might be an alternative strategy.

5 Summary

Elevated serum Lp(a) levels are a risk factor for atherosclerosis and myocardial infarction. Lp(a) serum levels are highly heritable, and to a great extend determined by the *LPA* locus on chromosome 6q27. Polymorphisms influencing the Lp(a) levels have already been identified, but a proportion of the variance in Lp(a) levels remains to be explained. In this investigation different methods were used to explore the influence of SNP markers located in the *LPA* gene region on Lp(a) levels. The methods comprised a pooling strategy that made use of data from a previous linkage analysis, and served as a screening tool for probably relevant SNP markers. For several SNP markers an association with serum Lp(a) levels could be shown. The identified SNP markers were genotyped in individuals of a large family sample and findings were replicated for all investigated markers. In a third step, a sample of the general population was investigated to underline the robustness of findings. Furthermore, for SNP markers showning a strong association with Lp(a) levels, the association with MI/CAD was investigated. Particularly one SNP marker in the *LPA* gene could be identified. This marker (rs11751605) showed a strong influence on Lp(a) levels in multiple populations and is significantly associated with both familial and incidental MI. Moreover, it contributes to the evidence of linkage to the Lp(a) locus on chromosome 6q27. Association with Lp(a) levels and with CAD could be replicated for a SNP marker (rs3798220) recently described in the literature. Furthermore data from a genome-wide association analysis using 500,000 SNP markers were analyzed. The results provide the basis for the identification of factors (apart from markers in *LPA* gene region) modulating Lp(a) levels.

Bibliography

[1] K. Berg. A new serum type system in man–the lp system. *Acta Pathol.Microbiol.Scand.*, 59:369–382, 1963.

[2] K. Berg and J. Mohr. Genetics of the lp system. *Acta Genet.Stat.Med.*, 13:349–360, 1963.

[3] B.R. Gabel and M.L. Koschinsky. Sequences within apolipoprotein(a) kringle iv types 6-8 bind directly to low-density lipoprotein and mediate noncovalent association of apolipoprotein(a) with apolipoprotein b-100. *Biochemistry*, 37(21):7892–7898, May 1998.

[4] G. Dahlén, K. Berg, T. Gillnäs, and C. Ericson. Lp(a) lipoprotein/pre-beta1-lipoprotein in swedish middle-aged males and in patients with coronary heart disease. *Clin Genet*, 7(4):334–341, Apr 1975.

[5] J.J. Albers, J.L. Adolphson, and W.R. Hazzard. Radioimmunoassay of human plasma lp(a) lipoprotein. *J Lipid Res.*, 18(3):331–338, May 1977.

[6] G. Utermann. The mysteries of lipoprotein(a). *Science*, 246(4932):904–910, November 1989.

[7] M.L. Koschinsky and S.M. Marcovina. Structure-function relationships in apolipoprotein(a): insights into lipoprotein(a) assembly and pathogenicity. *Curr.Opin.Lipidol.*, 15(2):167–174, April 2004.

[8] S.L. Frank, I. Klisak, R.S. Sparkes, T. Mohandas, J.E. Tomlinson, J.W. McLean, R.M. Lawn, and A.J. Lusis. The apolipoprotein(a) gene resides on human chromosome 6q26-27, in close proximity to the homologous gene for plasminogen. *Hum.Genet.*, 79(4):352–356, August 1988.

[9] J.W. Weisel, C. Nagaswami, J.L. Woodhead, A.A. Higazi, W.J. Cain, S.M. Marcovina, M.L. Koschinsky, D.B. Cines, and K. Bdeir. The structure of lipoprotein(a) and ligand-induced conformational changes. *Biochemistry*, 40(35):10424–10435, September 2001.

[10] J.W. McLean, J.E. Tomlinson, W.J. Kuang, D.L. Eaton, E.Y. Chen, G.M. Fless, A.M. Scanu, and R.M. Lawn. cdna sequence of human apolipoprotein(a) is homologous to plasminogen. *Nature*, 330(6144):132–137, November 1987.

[11] C.P. Ponting, J.M. Marshall, and S.A. Cederholm-Williams. Plasminogen: a structural review. *Blood Coagul.Fibrinolysis*, 3(5):605–614, October 1992.

[12] A.M. Scanu and C. Edelstein. Kringle-dependent structural and functional polymorphism of apolipoprotein (a). *Biochim.Biophys.Acta*, 1256(1):1–12, April 1995.

[13] C. Edelstein, D. Pfaffinger, J. Hinman, E. Miller, G. Lipkind, S. Tsimikas, C. Bergmark, G.S. Getz, J.L. Witztum, and A.M. Scanu. Lysine-phosphatidylcholine adducts in kringle v impart unique immunological and potential pro-inflammatory properties to human apolipoprotein(a). *J.Biol.Chem.*, 278(52):52841–52847, December 2003.

[14] C. Lackner, J.C. Cohen, and H.H. Hobbs. Molecular definition of the extreme size polymorphism in apolipoprotein(a). *Hum.Mol.Genet.*, 2(7):933–940, July 1993.

[15] A. Ichinose. Characterization of the apolipoprotein(a) gene. *Biochem.Biophys.Res.Commun.*, 209(1):365–371, April 1995.

[16] Y.Y. van der Hoek, M.E. Wittekoek, U. Beisiegel, J.J. Kastelein, and M.L. Koschinsky. The apolipoprotein(a) kringle iv repeats which differ from the major repeat kringle are present in variably-sized isoforms. *Hum.Mol.Genet.*, 2(4):361–366, April 1993.

[17] A.M. Scanu. Lp(a) lipoprotein–coping with heterogeneity. *N.Engl.J.Med.*, 349(22):2089–2090, November 2003.

[18] C. Haibach, H.G. Kraft, S. Kochl, A. Abe, and G. Utermann. The number of kringle iv repeats 3-10 is invariable in the human apo(a) gene. *Gene*, 208(2):253–258, February 1998.

[19] W. Sangrar, S.M. Marcovina, and M.L. Koschinsky. Expression and characterization of apolipoprotein(a) kringle iv types 1, 2 and 10 in mammalian cells. *Protein Eng*, 7(5):723–731, May 1994.

[20] H. Kratzin, V.W. Armstrong, M. Niehaus, N. Hilschmann, and D. Seidel. Structural relationship of an apolipoprotein (a) phenotype (570 kda) to plasminogen: homologous kringle domains are linked by carbohydrate-rich regions. *Biol.Chem Hoppe Seyler*, 368(12):1533–1544, December 1987.

[21] B. Garner, A.H. Merry, L. Royle, D.J. Harvey, P.M. Rudd, and J. Thillet. Structural elucidation of the n- and o-glycans of human apolipoprotein(a): role of o-glycans in conferring protease resistance. *J Biol.Chem*, 276(25):22200–22208, June 2001.

[22] S.M. Marcovina, M.L. Koschinsky, J.J. Albers, and S. Skarlatos. Report of the national heart, lung, and blood institute workshop on lipoprotein(a) and cardiovascular disease: recent advances and future directions. *Clin.Chem.*, 49(11):1785–1796, November 2003.

[23] S.J. Mao and M.A. Tucci. Lipoprotein(a) enhances plasma clot lysis in vitro. *FEBS Lett.*, 267(1):131–134, July 1990.

[24] B.R. Gabel and M.L. Koschinsky. Analysis of the proteolytic activity of a recombinant form of apolipoprotein(a). *Biochemistry*, 34(48):15777–15784, December 1995.

[25] K.A. Hajjar, D. Gavish, J.L. Breslow, and R.L. Nachman. Lipoprotein(a) modulation of endothelial cell surface fibrinolysis and its potential role in atherosclerosis. *Nature*, 339(6222):303–305, May 1989.

[26] A.M. Scanu, M.M. Atzeni, C. Edelstein, G. Tonolo, M. Maioli, and O. Klezovitch. Lipoprotein(a): identification of subjects with a superbinding capacity for fibrinogen. *Clin.Genet.*, 52(5):367–370, November 1997.

[27] A. Ernst, M. Helmhold, C. Brunner, A. Petho-Schramm, V.W. Armstrong, and H.J. Muller. Identification of two functionally distinct lysine-binding sites in kringle 37 and in kringles 32-36 of human apolipoprotein(a). *J Biol.Chem*, 270(11):6227–6234, March 1995.

[28] O. Rosby, P. Alestrom, and K. Berg. Sequence conservation in kringle iv-type 2 repeats of the lpa gene. *Atherosclerosis*, 148(2):353–364, February 2000.

[29] M. Trommsdorff, S. Kochl, A. Lingenhel, F. Kronenberg, R. Delport, H. Vermaak, L. Lemming, I.C. Klausen, O. Faergeman, G. Utermann, and . A pentanucleotide repeat polymorphism in the 5' control region of the apolipoprotein(a) gene is associated with lipoprotein(a) plasma concentrations in caucasians. *J.Clin.Invest*, 96(1):150–157, July 1995.

[30] D. Gavish, N. Azrolan, and J.L. Breslow. Plasma ip(a) concentration is inversely correlated with the ratio of kringle iv/kringle v encoding domains in the apo(a) gene. *J.Clin.Invest*, 84(6):2021–2027, December 1989.

[31] C. Brunner, E.M. Lobentanz, A. Petho-Schramm, A. Ernst, C. Kang, H. Dieplinger, H.J. Muller, and G. Utermann. The number of identical kringle iv repeats in apolipoprotein(a) affects its processing and secretion by hepg2 cells. *J.Biol.Chem.*, 271(50):32403–32410, December 1996.

[32] L. Becker, P.M. Cook, and M.L. Koschinsky. Identification of sequences in apolipoprotein(a) that maintain its closed conformation: a novel role for apo(a) isoform size in determining the efficiency of covalent lp(a) formation. *Biochemistry*, 43(31):9978–9988, August 2004.

[33] L. Becker, P.M. Cook, T.G. Wright, and M.L. Koschinsky. Quantitative evaluation of the contribution of weak lysine-binding sites present within apolipoprotein(a) kringle iv types 6-8 to lipoprotein(a) assembly. *J Biol.Chem*, 279(4):2679–2688, January 2004.

[34] W.J. Cain, J.S. Millar, A.S. Himebauch, U.J.F. Tietge, C. Maugeais, D. Usher, and D.J. Rader. Lipoprotein [a] is cleared from the plasma primarily by the liver in a process mediated by apolipoprotein [a]. *J Lipid Res*, 46(12):2681–2691, Dec 2005.

[35] B.R. Gabel, L.F. May, S.M. Marcovina, and M.L. Koschinsky. Lipoprotein(a) assembly. quantitative assessment of the role of apo(a) kringle iv types 2-10 in particle formation. *Arterioscler.Thromb.Vasc.Biol.*, 16(12):1559–1567, December 1996.

[36] Y.Y. van der Hoek, W. Sangrar, G.P. Cote, J.J. Kastelein, and M.L. Koschinsky. Binding of recombinant apolipoprotein(a) to extracellular matrix proteins. *Arterioscler.Thromb.*, 14(11):1792–1798, November 1994.

[37] A. D'Angelo, D. Geroldi, M.A. Hancock, V. Valtulina, A.I. Cornaglia, C.A. Spencer, E. Emanuele, A. Calligaro, M.L. Koschinsky, P. Speziale, and L. Visai. The apolipoprotein(a) component of lipoprotein(a) mediates binding to laminin: contribution to selective retention of lipoprotein(a) in atherosclerotic lesions. *Biochim.Biophys.Acta*, 1687(1-3):1–10, February 2005.

[38] G.A. Keesler, B.R. Gabel, C.M. Devlin, M.L. Koschinsky, and I. Tabas. The binding activity of the macrophage lipoprotein(a)/apolipoprotein(a) receptor is induced by cholesterol via a post-translational mechanism and recognizes distinct kringle domains on apolipoprotein(a). *J.Biol.Chem.*, 271(50):32096–32104, December 1996.

[39] M.L. Koschinsky, G.P. Cote, B. Gabel, and Y.Y. van der Hoek. Identification of the cysteine residue in apolipoprotein(a) that mediates extracellular coupling with apolipoprotein b-100. *J Biol.Chem*, 268(26):19819–19825, September 1993.

[40] M. Ogorelkova, A. Gruber, and G. Utermann. Molecular basis of congenital lp(a) deficiency: a frequent apo(a) 'null' mutation in caucasians. *Hum.Mol.Genet.*, 8(11):2087–2096, October 1999.

[41] V.N. Trieu and W.J. McConathy. A two-step model for lipoprotein(a) formation. *J Biol.Chem*, 270(26):15471–15474, June 1995.

[42] S. Tsimikas, E.S. Brilakis, E.R. Miller, J.P. McConnell, R.J. Lennon, K.S. Kornman, J.L. Witztum, and P.B. Berger. Oxidized phospholipids, lp(a) lipoprotein, and coronary artery disease. *N.Engl.J Med.*, 353(1):46–57, July 2005.

[43] J.E. Tomlinson, J.W. McLean, and R.M. Lawn. Rhesus monkey apolipoprotein(a). sequence, evolution, and sites of synthesis. *J Biol.Chem*, 264(10):5957–5965, April 1989.

[44] R.M. Lawn, N.W. Boonmark, K. Schwartz, G.E. Lindahl, D.P. Wade, C.D. Byrne, K.J. Fong, K. Meer, and L. Patthy. The recurring evolution of lipoprotein(a). insights from cloning of hedgehog apolipoprotein(a). *J Biol.Chem*, 270(41):24004–24009, October 1995.

[45] J.E. Hixson, M.L. Britten, G.S. Manis, and D.L. Rainwater. Apolipoprotein(a) (apo(a)) glycoprotein isoforms result from size differences in apo(a) mrna in baboons. *J Biol.Chem*, 264(11):6013–6016, April 1989.

[46] R.L. Nachman, D. Gavish, N. Azrolan, and T.B. Clarkson. Lipoprotein(a) in diet-induced atherosclerosis in nonhuman primates. *Arterioscler.Thromb.*, 11(1):32–38, January 1991.

[47] D.P. Wade, J.G. Clarke, G.E. Lindahl, A.C. Liu, B.R. Zysow, K. Meer, K. Schwartz, and R.M. Lawn. 5' control regions of the apolipoprotein(a) gene and members of the related plasminogen gene family. *Proc.Natl.Acad.Sci.U.S.A*, 90(4):1369–1373, February 1993.

[48] L. Patthy. Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. *Cell*, 41(3):657–663, July 1985.

[49] H.G. Kraft, A. Lingenhel, R.W. Pang, R. Delport, M. Trommsdorff, H. Vermaak, E.D. Janus, and G. Utermann. Frequency distributions of apolipoprotein(a) kringle iv repeat alleles and their effects on lipoprotein(a) levels in caucasian, asian, and african populations: the distribution of null alleles is non-random. *Eur.J.Hum.Genet.*, 4(2):74–87, 1996.

[50] L. Brazier, L. Tiret, G. Luc, D. Arveiler, J.B. Ruidavets, A. Evans, J. Chapman, F. Cambien, and J. Thillet. Sequence polymorphisms in the apolipoprotein(a) gene and their association with lipoprotein(a) levels and myocardial infarction. the ectim study. *Atherosclerosis*, 144(2):323–333, June 1999.

[51] M. Seed, K.L. Ayres, S.E. Humphries, and G.J. Miller. Lipoprotein (a) as a predictor of myocardial infarction in middle-aged men. *Am.J.Med.*, 110(1):22–27, January 2001.

[52] S.R. Holmer, C. Hengstenberg, H.G. Kraft, B. Mayer, M. Poll, S. Kurzinger, M. Fischer, H. Lowel, G. Klein, G.A. Riegger, and H. Schunkert. Association of polymorphisms of the apolipoprotein(a) gene with lipoprotein(a) levels and myocardial infarction. *Circulation*, 107(5):696–701, February 2003.

[53] G. Luc, J.M. Bard, D. Arveiler, J. Ferrieres, A. Evans, P. Amouyel, J.C. Fruchart, and P. Ducimetiere. Lipoprotein (a) as a predictor of coronary heart disease: the prime study. *Atherosclerosis*, 163(2):377–384, August 2002.

[54] A.R. Sharrett, C.M. Ballantyne, S.A. Coady, G. Heiss, P.D. Sorlie, D. Catellier, and W. Patsch. Coronary heart disease prediction from lipoprotein cholesterol levels, triglycerides, lipoprotein(a), apolipoproteins a-i and b, and hdl density subfractions: The atherosclerosis risk in communities (aric) study. *Circulation*, 104(10):1108–1113, September 2001.

[55] J. Danesh, R. Collins, and R. Peto. Lipoprotein(a) and coronary heart disease. meta-analysis of prospective studies. *Circulation*, 102(10):1082–1085, September 2000.

[56] H. Amemiya, T. Arinami, S. Kikuchi, K. Yamakawa-Kobayashi, L. Li, H. Fujiwara, M. Hiroe, F. Marumo, and H. Hamaguchi. Apolipoprotein(a) and pentanucleotide repeat polymorphisms are associated with the degree of atherosclerosis in coronary heart disease. *Atherosclerosis*, 123(1-2):181–191, June 1996.

[57] F. Kronenberg, M.F. Kronenberg, S. Kiechl, E. Trenkwalder, P. Santer, F. Oberhollenzer, G. Egger, G. Utermann, and J. Willeit. Role of lipoprotein(a) and apolipoprotein(a) phenotype in atherogenesis: prospective results from the bruneck study. *Circulation*, 100(11):1154–1160, September 1999.

[58] M. Hartmann, C. von Birgelen, G.S. Mintz, M.G. Stoel, H. Eggebrecht, H. Wieneke, M. Fahy, T. Neumann, J. van der Palen, H.W. Louwerenburg, P.M.J. Verhorst, and R. Erbel. Relation between lipoprotein(a) and fibrinogen and serial intravascular ultrasound plaque progression in left main coronary arteries. *J Am Coll Cardiol*, 48(3):446–452, August 2006.

[59] E. Zorio, C. Falco, M.A. Arnau, F. Espana, A. Osa, L.A. Ramon, R. Castello, L. Almenar, M.A. Palencia, and A. Estelles. Lipoprotein (a) in young individuals as a marker of the presence of ischemic heart disease and the severity of coronary lesions. *Haematologica*, 91(4):562–565, April 2006.

[60] E.J. Schaefer, S. Lamon-Fava, J.L. Jenner, J.R. McNamara, J.M. Ordovas, C.E. Davis, J.M. Abolafia, K. Lippel, and R.I. Levy. Lipoprotein(a) levels and risk of coronary heart disease in men. the lipid research clinics coronary primary prevention trial. *JAMA*, 271(13):999–1003, April 1994.

[61] G. Alfthan, J. Pekkanen, M. Jauhiainen, J. Pitkaniemi, M. Karvonen, J. Tuomilehto, J.T. Salonen, and C. Ehnholm. Relation of serum homocysteine and lipoprotein(a) concentrations to atherosclerotic disease in a prospective finnish population based study. *Atherosclerosis*, 106(1):9–19, March 1994.

[62] P.M. Ridker, C.H. Hennekens, and M.J. Stampfer. A prospective study of lipoprotein(a) and the risk of myocardial infarction. *JAMA*, 270(18):2195–2199, November 1993.

[63] R. Guerra, Z. Yu, S. Marcovina, R. Peshock, J.C. Cohen, and H.H. Hobbs. Lipoprotein(a) and apolipoprotein(a) isoforms: no association with coronary artery calcification in the dallas heart study. *Circulation*, 111(12):1471–1479, March 2005.

[64] B. Cantin, J.P. Despres, B. Lamarche, S. Moorjani, P.J. Lupien, P. Bogaty, J. Bergeron, and G.R. Dagenais. Association of fibrinogen and lipoprotein(a) as a coronary heart disease risk factor in men (the quebec cardiovascular study). *Am J Cardiol*, 89(6):662–666, March 2002.

[65] M. Jauhiainen, P. Koskinen, C. Ehnholm, M.H. Frick, M. Manttari, V. Manninen, and J.K. Huttunen. Lipoprotein (a) and coronary heart disease risk: a nested case-control study of the helsinki heart study participants. *Atherosclerosis*, 89(1):59–67, July 1991.

[66] S. Sunayama, H. Daida, H. Mokuno, H. Miyano, H. Yokoi, Y.J. Lee, H. Sakurai, and H. Yamaguchi. Lack of increased coronary atherosclerotic risk due to elevated lipoprotein(a) in women $>$ or $=$ 55 years of age. *Circulation*, 94(6):1263–1268, September 1996.

[67] V.W. Armstrong, P. Cremer, E. Eberle, A. Manke, F. Schulze, H. Wieland, H. Kreuzer, and D. Seidel. The association between serum lp(a) concentrations and angiographically assessed coronary atherosclerosis. dependence on serum ldl levels. *Atherosclerosis*, 62(3):249–257, December 1986.

[68] V.M. Maher, B.G. Brown, S.M. Marcovina, L.A. Hillger, X.Q. Zhao, and J.J. Albers. Effects of lowering elevated ldl cholesterol on the cardiovascular risk of lipoprotein(a). *JAMA*, 274(22):1771–1774, December 1995.

[69] P.N. Hopkins, L.L. Wu, S.C. Hunt, B.C. James, G.M. Vincent, and R.R. Williams. Lipoprotein(a) interactions with lipid and nonlipid risk factors in early familial coronary artery disease. *Arterioscler.Thromb.Vasc.Biol.*, 17(11):2783–2792, November 1997.

[70] C. K. Glass and J. L. Witztum. Atherosclerosis. the road ahead. *Cell*, 104(4):503–516, Feb 2001.

[71] J.J. Albers and W.R. Hazzard. Immunochemical quantification of human plasma lp(a) lipoprotein. *Lipids*, 9(1):15–26, January 1974.

[72] R. Clarke, M. Shipley, S. Lewington, L. Youngman, R. Collins, M. Marmot, and R. Peto. Underestimation of risk associations due to regression dilution in long-term follow-up of prospective studies. *Am J Epidemiol.*, 150(4):341–353, August 1999.

[73] H.G. Kraft, S. Kochl, H.J. Menzel, C. Sandholzer, and G. Utermann. The apolipoprotein (a) gene: a transcribed hypervariable locus controlling plasma lipoprotein (a) concentration. *Hum.Genet.*, 90(3):220–230, November 1992.

[74] T. Demant, K. Seeberg, A. Bedynek, and D. Seidel. The metabolism of lipoprotein(a) and other apolipoprotein b-containing lipoproteins: a kinetic study in humans. *Atherosclerosis*, 157(2):325–339, August 2001.

[75] H. Vu, K. Cianflone, Z. Zhang, D. Kalant, and A.D. Sniderman. Characterization and modulation of lp(a) in human hepatoma hepg2 cells. *Biochim.Biophys.Acta*, 1349(2):97–108, November 1997.

[76] A.L. White, D.L. Rainwater, and R.E. Lanford. Intracellular maturation of apolipoprotein[a] and assembly of lipoprotein[a] in primary baboon hepatocytes. *J Lipid Res.*, 34(3):509–517, March 1993.

[77] L. Becker, M.E. Nesheim, and M.L. Koschinsky. Catalysis of covalent lp(a) assembly: Evidence for an extracellular enzyme activity that enhances disulfide bond formation. *Biochemistry*, 45(32):9919–9928, August 2006.

[78] D.J. Rader, W. Cain, K. Ikewaki, G. Talley, L.A. Zech, D. Usher, and Jr. Brewer, H.B. The inverse association of plasma lipoprotein(a) concentrations with apolipoprotein(a) isoform size is not due to differences in lp(a) catabolism but to differences in production rate. *J Clin Invest*, 93(6):2758–2763, June 1994.

[79] C.M. Devlin, S.J. Lee, G. Kuriakose, C. Spencer, L. Becker, I. Grosskopf, C. Ko, L.S. Huang, M.L. Koschinsky, A.D. Cooper, and I. Tabas. An apolipoprotein(a) peptide delays chylomicron remnant clearance and increases plasma remnant lipoproteins and atherosclerosis in vivo. *Arterioscler.Thromb.Vasc.Biol.*, 25(8):1704–1710, August 2005.

[80] A. Gurakar, J.M. Hoeg, G. Kostner, N.M. Papadopoulos, and Jr. Brewer, H.B. Levels of lipoprotein lp(a) decline with neomycin and niacin treatment. *Atherosclerosis*, 57(2-3):293–301, November 1985.

[81] U. Broeckel, C. Hengstenberg, B. Mayer, S. Holmer, L.J. Martin, A.G. Comuzzie, J. Blangero, P. Nurnberg, A. Reis, G.A. Riegger, H.J. Jacob, and H. Schunkert. A comprehensive linkage analysis for myocardial infarction and its related risk factors. *Nat Genet*, 30(2):210–214, February 2002.

[82] S. Barlera, C. Specchia, M. Farrall, B.D. Chiodini, M.G. Franzosi, S. Rust, F. Green, E.B. Nicolis, J. Peden, G. Assmann, R. Collins, A. Hamsten, G. Tognoni, H. Watkins, and P. R. O. C. A. R. D. I. S. Consortium. Multiple qtl influence the serum lp(a) concentration: a genome-wide linkage screen in the procardis study. *Eur J Hum Genet*, 15(2):221–227, Feb 2007.

[83] E. Boerwinkle, C.C. Leffert, J. Lin, C. Lackner, G. Chiesa, and H.H. Hobbs. Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. *J Clin Invest*, 90(1):52–60, July 1992.

[84] J.C. Cohen, G. Chiesa, and H.H. Hobbs. Sequence polymorphisms in the apolipoprotein (a) gene. evidence for dissociation between apolipoprotein(a) size and plasma lipoprotein(a) levels. *J Clin Invest*, 91(4):1630–1636, April 1993.

[85] D.P. Wade, B.L. Knight, K. Harders-Spengel, and A.K. Soutar. Detection and quantitation of apolipoprotein(a) mrna in human liver and its relationship with plasma lipoprotein(a) concentration. *Atherosclerosis*, 91(1-2):63–72, November 1991.

[86] J.H. Wu and I.N. Lee. Studies of apolipoprotein (a) promoter from subjects with different plasma lipoprotein (a) concentrations. *Clin.Biochem.*, 36(4):241–246, June 2003.

[87] K. Suzuki, M. Kuriyama, T. Saito, and A. Ichinose. Plasma lipoprotein(a) levels and expression of the apolipoprotein(a) gene are dependent on the nucleotide polymorphisms in its 5'-flanking region. *J.Clin.Invest*, 99(6):1361–1366, March 1997.

[88] V. Mooser, F.P. Mancini, S. Bopp, A. Petho-Schramm, R. Guerra, E. Boerwinkle, H.J. Muller, and H.H. Hobbs. Sequence polymorphisms in the apo(a) gene associated with specific levels of lp(a) in plasma. *Hum.Mol.Genet.*, 4(2):173–181, February 1995.

[89] H.G. Kraft, M. Windegger, H.J. Menzel, and G. Utermann. Significant impact of the +93 c/t polymorphism in the apolipoprotein(a) gene on lp(a) concentrations in africans but not in caucasians: confounding effect of linkage disequilibrium. *Hum.Mol.Genet.*, 7(2):257–264, February 1998.

[90] W. Parson, H.G. Kraft, H. Niederstatter, A.W. Lingenhel, S. Kochl, F. Fresser, and G. Utermann. A common nonsense mutation in the repetitive kringle iv-2 domain of human apolipoprotein(a) results in a truncated protein and low plasma lp(a). *Hum.Mutat.*, 24(6):474–480, December 2004.

[91] M. Ogorelkova, H.G. Kraft, C. Ehnholm, and G. Utermann. Single nucleotide polymorphisms in exons of the apo(a) kringles iv types 6 to 10 domain affect lp(a) plasma concentrations and have different patterns in africans and caucasians. *Hum.Mol.Genet.*, 10(8):815–824, April 2001.

[92] A. Ichinose and M. Kuriyama. Detection of polymorphisms in the 5'-flanking region of the gene for apolipoprotein(a). *Biochem.Biophys.Res.Commun.*, 209(1):372–378, April 1995.

[93] M.M. Luke, J.P. Kane, D.M. Liu, C.M. Rowland, D. Shiffman, J. Cassano, J.J. Catanese, C.R. Pullinger, D.U. Leong, A.R. Arellano, C.H. Tong, I. Movsesyan, J. Naya-Vigne, C. Noordhof, N.T. Feric, M.J. Malloy, E.J. Topol, M.L. Koschinsky, J.J. Devlin, and S.G. Ellis. A polymorphism in the protease-like

domain of apolipoprotein(a) is associated with severe coronary artery disease. *Arterioscler.Thromb.Vasc.Biol.*, 27(9):2030–2036, September 2007.

[94] T. Strachan and A.P. Read. *Human Molecular Genetics, Third edition.* Garland Science, New York, 2004.

[95] J. Sambrook and D.W. Russell. *Molecular Cloning: A Laboratory Manual, Third edition.* Cold Spring Harbor Laboratory Press, Cold Spring Habor, New York, 2001.

[96] S. Germer, M. J. Holland, and R. Higuchi. High-throughput snp allele-frequency determination in pooled dna samples by kinetic pcr. *Genome Res*, 10(2):258–266, Feb 2000.

[97] M. Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31 (13):3406–15, 2003.

[98] K. J. Livak. Allelic discrimination using fluorogenic probes and the 5' nuclease assay. *Genetic Analysis: Biomolecular Engineering*, 14(5-6):143–149, February 1999.

[99] *GeneChip Human Mapping 500K Array Set data sheet, Part No. 702087 Rev.4, Affymetrix, Santa Clara, CA, USA.*

[100] N.J. Samani, J. Erdmann, A.S. Hall, C. Hengstenberg, M. Mangino, B. Mayer, R.J. Dixon, T. Meitinger, P. Braund, H.E. Wichmann, J.H. Barrett, I.R. Konig, S.E. Stevens, S. Szymczak, D.A. Tregouet, M.M. Iles, F. Pahlke, H. Pollard, W. Lieb, F. Cambien, M. Fischer, W. Ouwehand, S. Blankenberg, A.J. Balmforth, A. Baessler, S.G. Ball, T.M. Strom, I. Braenne, C. Gieger, P. Deloukas, M.D. Tobin, A. Ziegler, J.R. Thompson, and H. Schunkert. Genomewide association analysis of coronary artery disease. *N.Engl.J Med.*, 357(5):443–453, August 2007.

[101] The world health organization monica project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. who monica project principal investigators. *J Clin Epidemiol.*, 41(2):105–114, 1988.

[102] U. Keil, J. Stieber, A. Doring, L. Chambless, U. Hartel, B. Filipiak, H.W. Hense, M. Tietze, and J.G. Gostomzyk. The cardiovascular risk factor profile in the study area augsburg. results from the first monica survey 1984/85. *Acta Med.Scand.Suppl*, 728:119–128, 1988.

[103] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. de Bakker, M.J. Daly, and P.C. Sham. Plink: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum.Genet.*, 81(3):559–575, September 2007.

[104] N. M. Laird, S. Horvath, and X. Xu. Implementing a unified approach to family-based tests of association. *Genet Epidemiol*, 19 Suppl 1:S36–S42, 2000.

[105] S. Horvath, X. Xu, and N.M. Laird. The family based association test method: strategies for studying general genotype–phenotype associations. *Eur.J Hum.Genet.*, 9(4):301–306, April 2001.

[106] S.L. Lake, D. Blacker, and N.M. Laird. Family-based tests of association in the presence of linkage. *Am J Hum.Genet.*, 67(6):1515–1525, December 2000.

[107] J.C. Barrett, B. Fry, J. Maller, and M.J. Daly. Haploview: analysis and visualization of ld and haplotype maps. *Bioinformatics.*, 21(2):263–265, January 2005.

[108] R.C. Lewontin. The interaction of selection and linkage. ii. optimum models. *Genetics*, 50:757–782, October 1964.

[109] L. Almasy and J. Blangero. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum.Genet.*, 62(5):1198–1211, May 1998.

[110] J. T. Williams and J. Blangero. Power of variance component linkage analysis to detect quantitative trait loci. *Ann Hum Genet*, 63(Pt 6):545–563, Nov 1999.

[111] J. Blangero, J. T. Williams, and L. Almasy. Variance component methods for detecting complex trait loci. *Adv Genet*, 42:151–181, 2001.

[112] S.B. Gabriel, S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S.N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E.S. Lander, M.J. Daly, and D. Altshuler. The structure of haplotype blocks in the human genome. *Science*, 296(5576):2225–2229, Jun 2002.

[113] The international hapmap project. *Nature*, 426(6968):789–796, December 2003.

[114] A haplotype map of the human genome. *Nature*, 437(7063):1299–1320, October 2005.

[115] J.E. Wigginton, D.J. Cutler, and G.R. Abecasis. A note on exact tests of hardy-weinberg equilibrium. *Am J Hum Genet*, 76(5):887–893, May 2005.

[116] U. Styrkarsdottir, J.-B. Cazier, A. Kong, O. Rolfsson, H. Larsen, E. Bjarnadottir, V.D. Johannsdottir, M.S. Sigurdardottir, Y. Bagger, C. Christiansen, I. Reynisdottir, S.F.A. Grant, K. Jonasson, M.L. Frigge, J.R. Gulcher, G. Sigurdsson, and K. Stefansson. Linkage of osteoporosis to chromosome 20p12 and association to bmp2. *PLoS Biol*, 1(3):E69, Dec 2003.

[117] X. Zhu, R. Cooper, D. Kan, G. Cao, and X. Wu. A genome-wide linkage and association study using coga data. *BMC Genet*, 6 Suppl 1:S128, Dec 2005.

[118] M. Alarcón, B.S. Abrahams, J.L. Stone, J.A. Duvall, J.V. Perederiy, J.M. Bomar, J. Sebat, M. Wigler, C.L. Martin, D.H. Ledbetter, S.F. Nelson, R.M. Cantor, and D.H. Geschwind. Linkage, association, and gene-expression analyses identify cntnap2 as an autism-susceptibility gene. *Am J Hum Genet*, 82(1):150–159, Jan 2008.

[119] P. Sham, J.S. Bader, I. Craig, M. O'Donovan, and M. Owen. Dna pooling: a tool for large-scale association studies. *Nat Rev Genet*, 3(11):862–871, Nov 2002.

[120] H.G. Kraft, A. Lingenhel, S. Kochl, F. Hoppichler, F. Kronenberg, A. Abe, V. Muhlberger, D. Schonitzer, and G. Utermann. Apolipoprotein(a) kringle iv repeat number predicts risk for coronary heart disease. *Arterioscler.Thromb.Vasc.Biol.*, 16(6):713–719, June 1996.

[121] M.A. Hancock, M.B. Boffa, S.M. Marcovina, M.E. Nesheim, and M.L. Koschinsky. Inhibition of plasminogen activation by lipoprotein(a): critical domains in apolipoprotein(a) and mechanism of inhibition on fibrin and degraded fibrin surfaces. *J Biol.Chem*, 278(26):23260–23269, June 2003.

[122] M.B. Boffa, S.M. Marcovina, and M.L. Koschinsky. Lipoprotein(a) as a risk factor for atherosclerosis and thrombosis: mechanistic insights from animal models. *Clin.Biochem.*, 37(5):333–343, May 2004.

[123] S.N. Sotiriou, V.V. Orlova, N. Al-Fakhri, E. Ihanus, M. Economopoulou, B. Isermann, K. Bdeir, P.P. Nawroth, K.T. Preissner, C.G. Gahmberg, M.L. Koschinsky, and T. Chavakis. Lipoprotein(a) in atherosclerotic plaques recruits inflammatory cells through interaction with mac-1 integrin. *FASEB J.*, 20(3):559–561, March 2006.

[124] D.J. Hunter and P. Kraft. Drinking from the fire hose - statistical issues in genomewide association studies. *N. Engl. J. Med.*, 357:436–439, 2007.

Acknowledgement
I have enjoyed working on this PhD thesis and I am grateful for all the ideas and insights provided by many people who have worked with me. The friendly and supportive atmosphere inherent to the whole group contributed essentially to the final outcome of my studies.
I would like to thank, first and foremost, my supervisor, Prof. Dr. med. Christian Hengstenberg, for the patient guidance, encouragement and advice he has provided throughout my time as a PhD student. At many stages in the course of this research project I benefited from his advice, particularly so when exploring new ideas. His positive outlook and confidence in my research inspired me and gave me confidence. This thesis is a result of continuous fruitful interaction and many helpful discussions in the last years.
Special thanks also to Prof. Dr. med. Günter Riegger, the director of the Clinic of Internal Medicine II, University hospital of Regensburg, whose support I appreciate.
I am also extremely grateful to PD Dr. Thomas Langmann, for his input and interest in this research and his very prompt replies to ANY question.
I am deeply indebted to Dr. med Marcus Fischer, my supporting mentor, who kindly induced me into statistics. Moreover, many of his innovative ideas strongly influenced this work.
Furthermore, I thank Prof. Dr. med Stephan Holmer for his work on Lp(a) that provided the starting point for this thesis.
My thanks go to all people in the lab, in particular to:
Dr. Klaus Stark who has been a close research partner and friend in the past years. I am grateful for the opportunity to work with him. Special thanks go to Martina Köhler, Michaela Vöstner and Josef Simon, for their support and the good working atmosphere in the lab.
Moreover, I thank Silke Wiedmann for the friendship, support and motivation.
I would also like to thank all the members of the coordinating center who were involved in recruiting patients and managing the data. And beyond, I appreciate the invaluable contribution of participants of the German MI Family study, KORA Register and WHO MONICA survey.
Finally, I would like to mention that I was awarded a generous scholarship (Graduiertenförderung nach dem Bayerischen Eliteförderungsgesetz) which made this work possible.
I dedicate this work to my beloved husband Christoph. His great support and encouragement were an inestimable help for me.

Die vorliegende Arbeit wurde an der Klinik und Poliklinik für Innere Medizin II der Universität Regensburg in der Zeit von Februar 2005 bis Februar 2008 unter der Betreuung von PD Dr. Thomas Langmann angefertigt.

Ich versichere, dass ich diese Arbeit selbst verfasst habe und keine anderen als die angegebenen Hilfsmittel benutzt habe.

Katharina Neureuther