

**A Genome-wide Association Meta-analysis on Apolipoprotein A-IV Concentrations**

Claudia Lamina<sup>1§</sup>, Salome Friedel<sup>1§</sup>, Stefan Coassin<sup>1</sup>, Rico Rueedi<sup>2,3</sup>, Noha A. Younsri<sup>4,5</sup>,  
 Ilkka Seppälä<sup>6</sup>, Christian Gieger<sup>7,8,9</sup>, Sebastian Schönherr<sup>1</sup>, Lukas Forer<sup>1</sup>, Gertraud Erhart<sup>1</sup>,  
 Barbara Kollerits<sup>1</sup>, Pedro Marques-Vidal<sup>10</sup>, Janina Ried<sup>7</sup>, Gerard Waeber<sup>10</sup>, Sven Bergmann<sup>2,3</sup>,  
 Doreen Dähnhardt<sup>1</sup>, Andrea Stöckl<sup>1</sup>, Stefan Kiechl<sup>11</sup>, Olli T Raitakari<sup>12,13</sup>, Mika Kähönen<sup>14</sup>,  
 Johann Willeit<sup>11</sup>, Ludmilla Kedenko<sup>15</sup>, Bernhard Paulweber<sup>15</sup>, Annette Peters<sup>8,16,17</sup>,  
 Thomas Meitinger<sup>18,19,20</sup>, Konstantin Strauch<sup>7,21</sup>, KORA study group, Terho Lehtimäki<sup>6</sup>,  
 Steven C. Hunt<sup>22,23</sup>, Peter Vollenweider<sup>10</sup>, Florian Kronenberg<sup>1\*</sup>

*§ The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.*

<sup>1</sup> Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, 6020 Innsbruck, Austria

<sup>2</sup> Department of Computational Biology, University of Lausanne, 1015 Lausanne, Switzerland

<sup>3</sup> Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

<sup>4</sup> Department of Physiology and Biophysics, Weill Cornell Medical College – Qatar, Doha, Qatar

<sup>5</sup> Department of Computer and Systems Engineering, Alexandria University, 21526 Alexandria, Egypt

<sup>6</sup> Department of Clinical Chemistry, Fimlab Laboratories and University of Tampere School of Medicine, 33520 Tampere, Finland

<sup>7</sup> Institute of Genetic Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, 85764 Neuherberg, Germany

<sup>8</sup> Institute of Epidemiology II, Helmholtz Zentrum München—German Research Center for Environmental Health, 85764 Neuherberg, Germany

<sup>9</sup> Research Unit of Molecular Epidemiology, Helmholtz Zentrum München—German Research Center for Environmental Health, 85764 Neuherberg, Germany

<sup>10</sup> Department of Medicine, Internal Medicine, Centre Hospitalier Universitaire Vaudois (CHUV), 1015 Lausanne, Switzerland

<sup>11</sup> Department of Neurology, Medical University of Innsbruck, 6020 Innsbruck, Austria

<sup>12</sup> Department of Clinical Physiology, Turku University Hospital, 20520 Turku, Finland

<sup>13</sup> Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, 20520 Turku, Finland

<sup>14</sup> Department of Clinical Physiology, Tampere University Hospital and University of Tampere, 33521 Tampere, Finland

<sup>15</sup> First Department of Internal Medicine, Paracelsus Private Medical University, 5020 Salzburg, Austria

<sup>16</sup> DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, 80802 Munich, Germany

<sup>17</sup> German Center for Diabetes Research (DZD e.V.), 85764 Neuherberg, Germany

<sup>18</sup> Institute of Human Genetics, Technische Universität München, 81675 München, Germany

<sup>19</sup> Institute of Human Genetics, Helmholtz Zentrum München, 85764 Neuherberg, Germany

<sup>20</sup> Munich Cluster for Systems Neurology (SyNergy), 81377 Munich, Germany

<sup>21</sup> Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universität, 81377 Munich, Germany

<sup>22</sup> Cardiovascular Genetics Division, University of Utah School of Medicine, Salt Lake City, UT 84108, USA

<sup>23</sup> Department of Genetic Medicine, Weill Cornell Medicine, Doha, Qatar

**\*Address of correspondence:**

Florian Kronenberg, MD

Division of Genetic Epidemiology

Department of Medical Genetics, Molecular and Clinical Pharmacology

Innsbruck Medical University

Schöpfstr. 41, A-6020 Innsbruck, AUSTRIA

Phone: (+43) 512 9003-70560

Fax: (+43) 512 9003-73560 or -73561

E-mail: Florian.Kronenberg@i-med.ac.at

© The Author 2016. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License

(<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Abstract

Apolipoprotein A-IV (apoA-IV) is a major component of HDL and chylomicron particles and is involved in reverse cholesterol transport. It is an early marker of impaired renal function. We aimed to identify genetic loci associated with apoA-IV concentrations and to investigate relationships with known susceptibility loci for kidney function and lipids. A genome-wide association meta-analysis on apoA-IV concentrations was conducted in five population-based cohorts (n=13,813) followed by two additional replication studies (n=2,267) including approximately 10 M SNPs. Three independent SNPs from two genomic regions were significantly associated with apoA-IV concentrations: rs1729407 near *APOA4* ( $p=6.77 \times 10^{-44}$ ), rs5104 in *APOA4* ( $p=1.79 \times 10^{-24}$ ) and rs4241819 in *KLKB1* ( $p=5.6 \times 10^{-14}$ ). Additionally, a look-up of the replicated SNPs in downloadable GWAS meta-analysis results was performed on kidney function (defined by eGFR), HDL-cholesterol and triglycerides. From these three SNPs mentioned above, only rs1729407 showed an association with HDL-cholesterol ( $p=7.1 \times 10^{-07}$ ). Moreover, weighted SNP-scores were built involving known susceptibility loci for the aforementioned traits (53, 70 and 38 SNPs, respectively) and were associated with apoA-IV concentrations. This analysis revealed a significant and inverse association for kidney function with apoA-IV concentrations ( $p=5.5 \times 10^{-05}$ ). Furthermore, an increase of triglyceride-increasing alleles was found to decrease apoA-IV concentrations ( $p=0.0078$ ). In summary, we identified two independent SNPs located in or next the *APOA4* gene and one SNP in *KLKB1*. The association of *KLKB1* with apoA-IV suggests an involvement of apoA-IV in renal metabolism and/or an interaction within HDL particles. Analyses of SNP-scores indicate potential causal effects of kidney function and by lesser extent triglycerides on apoA-IV concentrations.

## Introduction

Apolipoprotein A-IV (apoA-IV) is an antioxidative glycoprotein that is synthesized primarily in the intestine and to a lesser extent in the liver (1, 2). It is secreted into the lymph as a structural protein of chylomicrons, very-low-density lipoproteins, high-density lipoproteins and participates in reverse cholesterol transport (3, 4). Consequently, it plays an important role in relieving peripheral cells of an overload of cholesterol (5, 6). It has an effect on fat resorption and has been discussed to be a satiety factor and related to diet-induced obesity at least in animal models (2). ApoA-IV shows anti-atherogenic properties (7, 8) and low concentrations were found to be associated with cardiovascular outcomes (9–12). Moreover, it acts as an early marker of impaired renal function and is a predictor of a progression of chronic kidney disease (13–16).

The knowledge on the genetic regulation of apoA-IV is limited. Heritability estimates derived from a family-based study in 119 nuclear families varied between 0% and 67%, depending on the underlying model (17). ApoA-IV is expressed by the *APOA4* gene on chromosome 11. This gene is in close proximity and linkage with *APOA5*, *APOC3* and *APOA1*. This gene region is often referred as the *APOA5-A4-C3-A1* gene cluster. There have been numerous candidate gene studies, which primarily evaluated the non-synonymous variants rs675 (T347S) and rs5110 (Q360H) with e.g. the ability of apoA-IV to bind lipids and to promote cholesterol efflux from cells (18). Association results of these variants with plasma apoA-IV levels (19, 20) as well proposed associations with triglycerides were contradictory (21–23). Variants in the *APOA5-A4-C3-A1* gene cluster have also been found to be associated on a genome-wide scale with lipid phenotypes, primarily with triglyceride and HDL cholesterol (HDL-C) concentrations (24). Up to now, there have been no genome-wide studies (GWAS) investigating apoA-IV concentrations.

The aim of the present study was to identify gene loci that are associated with apoA-IV concentrations based on a hypothesis-free approach. We conducted a genome-wide association meta-analysis using data from five population-based studies followed by a replication step in two additional studies. We also performed gene-based and pathway analyses to shed new light on the functional role of the identified genes and/or apoA-IV. Since the information on the heritability of apoA-IV is limited, we conducted a polygenic analysis to calculate the heritability of apoA-IV concentrations as well as the proportion of phenotypic variance explained by the single nucleotide polymorphisms (SNPs). ApoA-IV is known to be associated with kidney function and lipid phenotypes. Therefore, we also performed look-ups in and from the respective GWAS to elucidate possible causal relationships.

# Results

## Description of cohorts & quality control

Five studies contributed to the discovery stage (n=13,813) and 2 additional studies to the replication phase (n=2,267) (Figure 1), altogether including data from 16,080 participants. Due to the skewed distribution of apoA-IV concentrations, log-transformation of values was performed in all studies, resulting in nearly normal distributions (Supplementary Figure 1). Descriptive characteristics of all studies can be found in Supplementary Table 1. Meta-Analysis and quality control was performed as described in Supplementary Figure 2. The P-Z-plot did not reveal any deviations of the reported p-values and p-values calculated by the beta coefficient and standard error. Genomic inflation factors  $\lambda$  within studies ranged from 1.011 to 1.038 (Supplementary Table 2).

## GWA discovery stage

The GWA meta-analysis (stage 1) resulted in two genome-wide significant gene-regions (Manhattan plot shown in Figure 2, QQ-plot shown in Supplementary Figure 3). In a broad region surrounding the *APOA4* gene, 423 SNPs reached genome-wide significance with the lowest p-value for SNP rs1729407 ( $p=6.00 \times 10^{-40}$ , Figure 3). Additionally, 64 genome-wide significant SNPs in the gene-region around the *KLKB1* gene on chromosome 4 were identified (lowest p-value for SNP rs4241819:  $1.08 \times 10^{-12}$ , Figure 4). Furthermore, one locus on chromosome 5 (*SOWAHA*) reached our predefined level of significance sufficient for replication using the RE model by Han & Eskin (lowest p-value for SNP rs59698941:  $3.76 \times 10^{-07}$ , Supplementary Figure 4).

## SNP selection

Conditional analyses were performed for *APOA4* (chr11: 116177370-117177370, Figure 3), *KLKB1* (chr4: 186657140-187657140, Figure 4) and *SOWAHA* (chr5: 131654912-132654912, Supplementary Figure 4). For *APOA4*, two SNPs were independently associated with apoA-IV concentrations: rs1729407 (p-value single SNP analysis:  $6.00 \times 10^{-40}$ ; p-value conditional analysis:  $2.66 \times 10^{-25}$ ) and rs5104 (p-value single SNP analysis:  $1.24 \times 10^{-22}$ ; p-value conditional analysis:  $4.01 \times 10^{-08}$ , Supplementary Figure 5). After apoA-IV concentrations were adjusted for these two SNPs, no further SNPs remained in the model with an adjusted p-value less than  $1 \times 10^{-6}$  (Supplementary Figure 6). Besides the two SNPs rs1729407 and rs5104 the following missense variants were selected for replication: rs5110 ( $p=9.26 \times 10^{-07}$ ) and rs675 ( $p=0.0021$ ). The latter was selected due to its wide use in the literature. For *KLKB1* and *SOWAHA*, no additional SNP was added by applying the conditional analysis. One missense variant was selected within the *SOWAHA* gene region (rs2292030,  $p=9.24 \times 10^{-07}$  using RE model, within *SHROOM1*). The lead SNP in *KLKB1* (rs4241819) and the *KLKB1*

missense variant rs3733402 that were selected for replication were not accessible to iPLEX genotyping. Therefore, a proxy SNP (rs4253311) in high linkage disequilibrium (LD) with both the lead SNP and the missense variant was chosen for replication ( $p=1.43 \times 10^{-11}$ ,  $r^2=0.932$  with rs4241819,  $r^2=0.994$  with rs3733402, based on 1000 Genomes phase 3 v5; see also Figure 4 for graphical display of LD between the SNPs). Characteristics of all selected SNPs can be found in Supplementary Table 3.

### Replication stage and combined analysis

Altogether, 7 SNPs were genotyped in the two replication studies. The single study results for these SNPs are given in Supplementary Table 4. Of these, 3 SNPs reached a false-discovery rate less than 0.05 on the replication stage and a genome-wide significance level after inclusion of all 7 studies (discovery stage + replication stage, Table 1): rs1729407 near *APOA4* ( $p=6.77 \times 10^{-44}$ ), rs5104 in *APOA4* ( $p=1.79 \times 10^{-24}$ ) and rs4241819 (using rs4253311 as proxy in the replication studies) in *KLKBI* ( $p=5.63 \times 10^{-14}$ ). For these 3 SNPs, effect directions were identical in all studies. For each copy of the minor allele of rs1729407, apoA-IV concentrations decrease by 0.2645 mg/dl. Each minor allele of rs5104 also decreases apoA-IV concentrations by 0.2526 mg/dl. In a joint analysis including both SNPs, both SNPs remain significant ( $p=2.66 \times 10^{-25}$  for rs1729407,  $p=4.01 \times 10^{-08}$  for rs5104) with slightly smaller effect estimates ( $\beta=0.2041$  for rs1729407 and  $\beta=0.1455$  for rs5104). The minor allele of SNP rs4241819/rs4253311 in *KLKBI* increases apoA-IV concentrations by 0.1469 mg/dl.

### Effects in men and women

GWAS stratified for men and women did not reveal any additional genome-wide significant SNPs outside the wider *APOA4* gene region (Supplementary Figures 7 and 8). There was also no genome-wide significant SNP-gender interaction effect (Supplementary Figure 9).

### ApoA-IV variance explained

All SNPs combined within the broad *APOA4* region (lead SNP +/- 500 kB) explained ~3.30% (95% CI: [1.60%; 5.00%]) of the phenotypic variance assuming an additive model, based on both KORA studies. The top SNP rs1729407 alone explained 1.38% in KORAF3/F4 and 1.39% in the SAPHIR study, respectively, and rs5104 alone 1% in KORAF3/F4 and 0.57% in SAPHIR. The *KLKBI* region (lead SNP +/- 500 kB) accounted for 0.67% (95% CI: [0.00%; 1.54%]) in KORAF3/F4. SNP rs4241819/rs4253311 alone explained 0.44% in KORAF3/F4 and 0.19% in the SAPHIR study, respectively. All three SNPs together (two in *APOA4*, one in *KLKBI*) in one model explained 2.2% in the combined KORAF3/F4 dataset. The genome-wide SNP-based explained variance including the entire dataset of available SNPs (genomic heritability) was estimated to be 36.07% in both KORA

studies (95% CI: [18.48%; 53.66%]). The narrow-sense heritability  $h^2$  of apoA-IV, derived from the polygenic model in the family-based FamHS study, was estimated to be 27.45%.

### **Gene-based and Pathway analyses**

The gene-based association scan resulted in 18 significant genes, all of which are located either in the broad *APOA4* or *KLKB1* gene regions (Table 2). The pathway analysis revealed 15 gene sets to be significantly enriched with susceptibility genes, including expected lipid transport and lipoprotein metabolism pathways as well as some additional liver-related pathways (Supplementary Table 5).

### **Look-up in other GWA meta-analysis consortia**

We looked up the two replicated SNPs in the *APOA4* gene region (rs1729407, rs5104), the replicated SNP in *KLKB1* (rs4241819), its proxy used in the replication step (rs4253311) and the correlated missense mutation (rs3733402) in the GWA meta-analysis results on kidney function, HDL-C and triglycerides. Two SNPs (rs1729407, rs4253311) were available in all GWAS consortia. Only one significant result was found: the apoA-IV lead SNP rs1729407 was associated with HDL-C with a p-value of  $7.1 \times 10^{-07}$  (Table 3).

In addition, lead SNPs from the most recent HDL-C (n=70), triglyceride (n=38) and kidney function (n=53) GWAS were selected. The p-values of these partially overlapping 142 SNPs were retrieved from our GWA meta-analysis on log-transformed apoA-IV (Supplementary Tables 6-8). Only one SNP was significantly associated with apoA-IV, rs964184 in *APOA1* (p=0.0001), which is included in the HDL-C as well as in the triglyceride SNP list.

The analyses based on the weighted genetic SNP-scores for kidney function, HDL-C and triglycerides in the combined KORA F3 and F4 dataset yielded two significant results. The weighted kidney function SNP-score was significantly and inversely associated with apoA-IV ( $p=5.5 \times 10^{-05}$ ). That means apoA-IV concentrations increased with an increasing number of GFR-decreasing SNPs. Furthermore, a greater number of triglyceride-increasing alleles was shown to be associated with lower apoA-IV concentrations ( $p=0.0078$ ). The association with HDL-C SNPs was not significant but pointed in the opposite direction as expected: the more HDL-C-increasing alleles, the higher were apoA-IV concentrations ( $p=0.0554$ ).

## Discussion

This study revealed three major findings. First, using genome-wide data from five studies and two independent replication studies we could identify three independent SNPs from two genomic regions (*APOA4* and *KLKB1*), which were significantly associated with apoA-IV concentrations. Second, approximately one third of the phenotypic variability of apoA-IV seems to be genetically regulated. Third, genetic variants that have a significant effect on kidney function and triglyceride concentrations suggest a causal role of these phenotypes on apoA-IV concentrations.

### Genome-wide significant and replicated SNPs in *APOA4* and *KLKB1*

Conditional stepwise regression analysis including all SNPs in the broad *APOA4* gene region (a 1 MB region including the *APOA5-A4-C3-A1* cluster) led to the identification of two SNPs: the lead SNP rs1729407, located between *APOA5* and *APOA4*, and one missense variant (rs5104). So far, the effect of rs5104 has been studied only in some small studies: it was associated with dyslipidemia in Han Chinese (25), with postprandial ApoA-I plasma concentration in healthy young men (26) and with triglyceride response to fenofibrate treatment (27). Conversely, no association between the lead SNP rs1729407 and any phenotype had been shown until now. The effect of other missense variants in *APOA4* (rs675, rs5110), although widely studied before, could not be replicated. However, these previous studies were markedly smaller, showed contradictory results and investigated different inheritance models (19, 20, 23).

Both *APOA4* top hits do not present overt functional effects. The lead SNP rs1729407 is located in an intergenic region (Supplementary Figure 10A) while rs5104 causes a serine to asparagine substitution (Ser147Asn), which is classified as benign by common bioinformatics prediction tools. Of note, the lead SNP is in perfect LD ( $r^2=1$ ) with a SNP located in a large cluster of transcription factor binding sites located approximately 1.5 kb downstream (rs1729405,  $p=9.92E-40$  in our meta-analysis; Supplementary Figure 10B).

Besides the *APOA4* gene, we also identified a locus on chromosome 4 encompassing the three genes *CYP4V2*, *KLKB1* and *F11*. The top hit was in nearly perfect LD with the missense variant rs3733402 in *KLKB1*. *KLKB1* encodes the glycoprotein plasma kallikrein (also known “Fletcher factor” (28)), which acts as a proteolytic activator of several vasoactive and circulating peptides (kinins) (29, 30). Accordingly, SNPs in *KLKB1* showed genome-wide associations with vasoactive peptides (plasma bradykinin (31, 32), active renin (rs3733402) (33), BNP in Blacks (34), aldosterone/renin ratio in Europeans (34), MR-pro-ADM and CT-pro-ET-1 (rs4253238 (35),  $r^2=0.81$  with our tophit). Of note,



*F11* is a paralog of *KLKB1* and codes for the coagulation factor XI. Both are part of the intrinsic pathway (36). However, to our knowledge a mechanism which obviously links apoA-IV to the kinin-kallikrein system or the intrinsic pathway has not been described so far. Therefore replication and functional studies will be required to appraise the significance of this finding. The third gene in the locus, *CYP4V2*, is a nearly ubiquitously expressed omega-hydroxylase, with the phenotype of loss-of-function mutations being restricted to the eye (37, 38) and causing the degenerative ocular disease BCD (39) (OMIM #210370).

Finally, gene-based analysis or pathway-based analysis did not reveal additional novel genes beyond those located in the genomic regions around *APOA4* and *KLKB1*. Since the stepwise conditional analysis resulted in only two independent SNPs located at the *APOA4* or *KLKB1* loci, the observation in the gene-based analysis that multiple genes were significant in each locus could most likely be explained by LD.

### **Variance explained and heritability**

Another aim of this study was the estimation of the heritability of apoA-IV as well as the variation of apoA-IV explained by all included additive-coded SNPs. Both, genomic and also narrow-sense heritability were calculated to be around 30%. Only a relatively small fraction is explained by the two gene regions we have identified which leaves sufficient room for the discovery of other gene regions. In addition, the major extent of apoA-IV concentrations seems to be regulated by non-genetic factors.

### **SNP look-up using results from other GWAS consortia**

Another aspect of this study was the look-up of the identified SNPs in other GWAS consortia. Since variants in the *APOA5-A4-C3-A1* gene cluster have consistently been found to be associated with triglycerides and HDL-C (24, 40, 41), results from the most recent lipids-GWA meta-analysis (24) was used for this look-up. The lead *APOA4*-SNP rs1729407 showed an association with HDL-C ( $p=7.1 \times 10^{-07}$ ). However, this SNP seems to be independent from the lead SNP of the lipid-GWA within that gene region (rs964184, reported gene *APOA1*,  $r^2$  with rs1729407  $< 0.1$ ), which had a p-value of  $6.00 \times 10^{-48}$  in the GWAS on HDL-C and  $7.00 \times 10^{-224}$  in the GWAS on triglycerides (24). SNP rs964184 has also been associated with coronary heart disease on a genome-wide scale, an association possibly triggered by the strong association of rs964184 with triglyceride concentrations (42–44). In our analysis, rs964184 was also associated with apoA-IV ( $p=0.0001$ ). However, this is far away from genome-wide significance. Altogether, it seems that, despite being within the *APOA5-A4-C3-A1* gene



cluster, the SNPs associated with HDL-C and triglycerides are statistically independent from the *APOA4*-SNPs associated with apoA-IV concentrations.

We also performed a look-up to check whether the SNPs detected in our apoA-IV GWAS study were associated with kidney function, defined by eGFR using data from the CKDGen consortium (45). This consortium was chosen because of the already known association of apoA-IV with kidney function and chronic kidney disease (13–16, 46). However, none of the *APOA4* and *KLKB1* lead SNPs showed significant associations with eGFR.

We further applied a look-up approach the other way around: when we selected in total 142 unique SNPs that were retrieved from the kidney- and lipid-GWAS, no single SNP was associated with apoA-IV in our GWAS besides the aforementioned SNP in *APOA1* (TG and HDL-C). However, taken together as weighted SNP-scores, the strongest associations with apoA-IV could be found for the kidney-SNP-score and still significant associations for the triglycerides-SNP-score. These results potentially support a possible causal effect of kidney function on apoA-IV concentrations. This might also be true for a potential causal effect of triglycerides on apoA-IV, but to a lesser extent.

So far, only few studies investigated the association between apoA-IV and triglyceride levels, and the results have been inconsistent: for example, no association could be found in the EARS study (1261 controls and 629 cases) (23), whereas a study conducted in 105 participants reported a significantly positive association between apoA-IV and triglyceride levels (47) concentrations. This finding is contradictory to the direction of correlation we found using a triglyceride-increasing SNP-score.

As part of the HDL particle, apoA-IV plays a role as a mediator in the reverse-cholesterol transport (48). Some epidemiological studies also suggest an association of HDL-C with apoA-IV (23). However, a causal role of HDL-C on apoA-IV could not be shown with our data, but also not ruled out. In Hanniman et al. (49), *APOA4* knockout mice showed decreased HDL-C values, whereas overexpression of *APOA4* led to increase of HDL-C, which suggests a causal role of apoA-IV on HDL-C.

## **Conclusion**

Using data from five population-based studies and two additional replication studies, two independent SNPs located in or next to the *APOA4* gene and one SNP in *KLKB1* gene were significantly associated with apoA-IV levels. These two gene regions alone can only explain a small fraction of the genome-wide explained variance by SNPs which we estimated to be roughly 30%. Therefore, a major part of apoA-IV variability is likely to be regulated by non-genetic factors. Analyses of SNP-scores

explaining kidney function, HDL-C and triglyceride levels indicate a potential causal effect of primarily kidney function and by a lesser extent triglycerides on apoA-IV levels.

## Methods

### Study design

The genome-wide SNP association analysis on apoA-IV is based on a two-stage design with a discovery stage and a replication stage (Figure 1). Genome-wide SNP arrays were available for 5 studies of European ancestry (n=13,813 in total). All independent SNPs and missense variants with a p-value below  $1 \times 10^{-6}$  were taken forward to the replication stage. In addition, one non-synonymous SNP from the *APOA4* gene that did not fulfill the p-value selection criteria was selected for replication (rs675), since it has been widely studied before (18). Altogether, 8 SNPs were then genotyped in both replication studies. Replication of SNPs was achieved, if the following criteria were met: genome-wide significance ( $p < 5 \times 10^{-8}$ ) in the meta-analysis of all 7 studies within the discovery + replication stage (n=16,080 in total), direction of effects in replication studies consistent with the discovery stage and a false-discovery rate (FDR) (50) less than 0.05 on the replication stage.

### GWAs discovery stage: study population, genotyping and imputation

Details on genotyping and imputation for each study can be found in Supplementary Table 2.

The **CoLaus** study is a single-center, cross-sectional study including 6,182 Caucasian subjects aged 35 to 75 years from the city of Lausanne in Switzerland (51). From 5,435 participants, genotypes were imputed using the software minimac (52) and 1000 Genomes (phase 1, version 3), resulting in over 7 million SNPs after filtering. Full phenotype information as well as imputed genotypes are available for n=3,996 participants.

For the **NHLBI Family Heart Study (FamHS)**, 1,200 families (~6,000 individuals) were ascertained in 1992, half randomly sampled, half selected because of an excess of coronary heart disease (CHD) or risk factor abnormalities (53). Study participants belonging to the largest pedigrees were invited for a second clinical exam (2002-2004). GWAS analysis was undertaken for 4135 European American subjects using Illumina arrays. SNP genotypes were subsequently imputed with the software MACH (version 1.0.16) (54) using 1000 genomes phase 1 version 3 (55) as reference, leading to a total of ~7.7 million SNPs after filtering. Both imputed genotype data as well as phenotype information was available for n=1,712 participants.

The **KORA F3** study, conducted in the years 2004/05, is a population-based sample from the general population living in the region of Augsburg, Southern Germany, which has evolved from the WHO MONICA study (Monitoring of Trends and Determinants of Cardiovascular Disease). Genome-wide data are available for all participants (n=3,075 with complete phenotype information) based on Illumina Omni 2.5/Illumina Omni Express. The **KORA F4** survey is an independent non-overlapping sample drawn from the same population in the years 2006/08 (n=2,926 with complete phenotype information). Genome-wide data are available for all participants in the KORA F4 study (Affymetrix Axiom) (40, 56). Both genome-wide genotype data have been imputed with the software IMPUTE using 1000 Genomes phase 1, version 3 (55). After quality control and filtering, about 8.5 M SNPs are available for analyses in both KORA F3 and F4.

The **Cardiovascular Risk in Young Finns Study (YFS)** is a prospective multicenter study from Finland initiated in 1980 (Baseline age 3–18 years) with several follow-ups over 30 years to investigate childhood risk factors for cardiometabolic outcomes (57). For 2443 participants from the 2001 follow-up (ages 24-39 years), high throughput genome wide SNP genotyping using the genome wide Illumina 670K SNP chip was performed at the Wellcome Trust Sanger Centre. Imputation was performed using IMPUTE and the 1000 Genomes Project March 2012 version (phase 1, version 3) as reference, leading to a total of 8.5 million imputed genotypes after filtering. Full phenotype information as well as imputed genotypes is available for n=2,104 participants.

#### **Replication stage: study population and de-novo genotyping**

The **Bruneck** study is a prospective population-based survey designed to investigate the epidemiology and pathogenesis of atherosclerosis (58, 59). The study population was recruited in 1990 as a sex- and age-stratified random sample of all inhabitants of Bruneck, Italy. The attendance rate was 93.6% with complete data in 919 subjects. An intensive phenotyping was done and follow-up data are available for a period of 25 years.

The **SAPHIR** study (Salzburg Atherosclerosis Prevention Program in subjects at High Individual Risk) is an observational study conducted in the years 1999-2002 involving 1,770 unrelated subjects from a healthy working population. Study participants were recruited by health-screening programs in companies in and around the Austrian city of Salzburg (60). Full phenotype and genotype information is available for n=1,454 participants.

In both studies, de-novo genotyping was performed in a multiplex approach using the SEQUENOM MassArray platform and iPLEX Gold chemistry. Full phenotype and genotype information is available for n=802 participants.

## Measurement of apoA-IV

For all participating studies, quantification of plasma apoA-IV was done in the same laboratory (Division of Genetic Epidemiology, Medical University of Innsbruck, Austria). It was based on a double-antibody enzyme-linked immunosorbent assay using an affinity-purified polyclonal rabbit anti-human apoA-IV antibody for coating and the same antibody coupled to horseradish peroxidase for detection. Plasma with a known concentration of apoA-IV was used as the calibration standard (61). Four control sera with different concentrations were run on each plate in double measurements for control purposes throughout the entire project. The intra- and interassay coefficients of variation were 2.7% and 6.0%, respectively (61).

## Statistical methods

### GWAS analysis of single studies & discovery stage meta-analysis

An overview of the quality control and meta-analysis workflow in the discovery stage is given in Supplementary Figure 1. Due to the skewed distribution of apoA-IV concentrations, log-transformation of values was performed in all studies, resulting in nearly normal distributions (Supplementary Figure 2). In each study, each SNP was associated with log-transformed apoA-IV concentrations in an additive genetic model using linear regression, adjusted for age and sex. Additionally, linear regression was performed on the untransformed apoA-IV levels to obtain interpretable effect estimates. Since women have slightly lower apoA-IV levels than men, gender-stratified models have also been applied in all studies (62). Genome-wide analysis in the FamHS study was performed using a linear mixed model accounting for familial dependencies described by a pedigree-based kinship matrix.

Quality control and filtering of SNPs was performed centrally and standardized by the Innsbruck study group using EasyQC (63). SNPs were only included in the analysis if they fulfilled the following criteria: imputation quality  $\geq 0.4$  (e.g. IMPUTE info), minor allele frequency  $\geq 1\%$  and a p-value of the HWE-test  $\geq 1 \times 10^{-6}$ . Additional analyses for quality control were applied on the already filtered datasets, which included a P-Z-plot (63) and calculation of genomic inflation factor  $\lambda$ . The P-Z-plot compares the reported p-values from each study with the p-values calculated from Z-statistics derived from the reported beta coefficient and standard error.

For the meta-analysis over all GWAS studies, METASOFT (64) was used for all imputed SNPs that met imputation and quality control criteria. SNPs were only included in the meta-analysis if they were available in 3 or more studies. Based on the heterogeneity between studies for each SNP, a fixed effects (FE) or optimized random effects model (RE) as proposed by (64), was used as implemented in

METASOFT. The test statistic for this optimized RE model is partitioned into a mean effects and heterogeneity part. To give higher weights to the mean effects, this RE model was only used when the test statistic for the mean effects part was higher than the heterogeneity part and if the test for heterogeneity was significant (p value of Q statistic  $< 0.1$  &  $I^2 \geq 50$ ). The test statistics were corrected for genomic inflation in both the GWAS analysis stage and meta-analysis stage. Based on the gender-stratified analyses, a t-test on effect differences between men and women was performed (62). All regional plots presenting the p-values and LD between SNPs in predefined genomic regions were done using LocusZoom (65).

### **SNP selection for replication**

To detect independently associated SNPs, a conditional stepwise analysis was performed using the program GCTA (version 1.24.7 (66)). For each locus with at least one p-value  $< 10^{-6}$ , the SNP with the lowest p-value on the discovery stage was taken as the lead SNP. All SNPs within a region  $\pm 500$  kb surrounding the lead SNP were included in the conditional analysis. GCTA uses the summary-level statistics of the meta-analysis plus one reference population for LD calculation. As reference population, a combined genotype dataset of KORA F3 and KORA F4 was used (n=6,001). By default, the lead SNP is included in the model. Then, all SNPs in the included gene region are tested for association in addition to the already included SNPs. Finally, all SNPs within a gene region with a p-value of  $< 10^{-6}$  in the conditional analysis were taken forward for replication. Furthermore, all missense mutations with p-values of  $< 10^{-6}$  were selected for replication, irrespective of possible LD with already selected SNPs.

### **Two-stage meta-analysis**

All genotyped SNPs in the replication phase were meta-analyzed in both replication studies separately as well as in a combined analysis of all GWAS and replication-stage studies. Again, METASOFT (64) was used in the same way as in the first stage meta-analysis.

### **Gene-based test and Pathway analysis**

In addition to the analysis of single SNP effects, a gene-based scan and a pathway analysis were performed using KGG version 3.5 (67). Gene regions were defined as the gene  $\pm 20$  kb according to the RefGene database. Using this definition, 66.35% of the available SNPs were included in the analysis. For the gene-based analysis, the extended Simes test (GATES) was used as implemented in KGG (68). To adjust for multiple testing, the Bonferroni-method was applied on the number of tested genes. To calculate LD between the SNPs, the 1000G Phase 1 v3 Reference was used. All pathways that are available in the C2 curated gene set from GSEA

(<http://software.broadinstitute.org/gsea/msigdb/>) were included in the pathway analysis. To test for enrichment of each pathway with significant genes, a hypergeometric test as implemented in KGG was used (69). To adjust for multiple testing, the Bonferroni-method was applied on the number of pathways tested.

### **Variance explained**

The percentage of explained variance for the SNPs that were taken forward for replication was calculated in the SAPHIR study (n=1,465) - as an independent replication cohort - as well as in a combined dataset of both KORA studies (n=6,001). The combined KORA dataset was also used to get an estimate of the proportion of phenotypic variance explained by the regression on additively coded SNPs for a) all SNPs within the *APOA4* and *KLKB1* gene regions, defined as the lead SNPs +/- 500 kB, as in the conditional analysis and b) all available genome-wide imputed SNPs. The latter has been denoted as the genomic heritability (70). Hence, this genomic heritability includes solely the variance attributable to the measured SNP effects. For these analyses, the software GCTA version 1.24.7 was used (66). In the FamHS study, an estimate of the proportion of the additive (polygenic) variance on the phenotypic variance, the narrow-sense heritability  $h^2$ , was obtained using GenABEL's polygenic function, taking the kinship matrix into account. This narrow-sense heritability thus also includes the variance explained by not measured SNPs and other variants (e.g. copy-number-variations). All estimates for the explained variance and heritability refer to log-transformed values of apoA-IV.

### **SNP Look-up**

We performed a look-up of our replicated SNPs in downloadable GWA meta-analysis results on kidney function (defined by eGFR) (71), HDL-C and triglycerides (24). We further looked up lead SNPs identified in these consortia in our apoA-IV GWA meta-analysis. 53 SNPs associated with kidney function, defined by eGFR, were derived from the CKDGen-GWA meta-analysis and 70 SNPs with HDL-C and 38 with triglycerides from the GLGC-GWA meta-analysis. For these SNPs, their respective p-values from the log-transformed analysis on apoA-IV levels on the discovery stage were looked up. Altogether, 143 unique SNPs were included in this analysis, some of them involved in more than one phenotype (especially for HDL-C and triglycerides). Therefore, results are declared significant, when the p-value is lower than  $0.05/143 = 0.00035$ . Since the effect of single SNPs (and therefore the statistical power) is assumed to be low, we also used the imputed genotypes in both KORA studies to create SNP-scores. Weighting and direction of effects were based on the original publication where the SNPs were derived from. All SNPs were scaled in such a way that they are phenotype and/or risk increasing and weighted by the beta-estimate derived from the respective

original study. These weighted genotype scores were then summed up to derive a genetic risk score for each of the phenotypes studied. For these analyses, a combined dataset of KORAF3 and KORAF4 was used (n=6,001). A mixed effects model was performed for this analysis with the study included as a random effects variable. Since three SNP-scores were evaluated, the significance threshold was set to  $0.05/3=0.0167$  for these analyses.

### **Bioinformatic analysis**

Bioinformatic analysis of intergenic variants using ENCODE data was carried out as described before (72). Analysis of the coding variants was performed using tools Polyphen-2 (73), SIFT (74) and MutPred (75). Pairwise LDs were calculated using SNIIPA (76) (<http://snipa.helmholtz-muenchen.de>) using the European 1000 Genomes Phase 3, v5 dataset.

## **Acknowledgments / Funding**

The measurements of apoA-IV were supported by a grant from the "Standortagentur Tirol" to Florian Kronenberg. This work was supported by the Austrian Science Fund (FWF) (Project P 26660-B13) to Claudia Lamina. The NHLBI Family Heart Study was supported by the National Heart, Lung, and Blood Institute cooperative agreement grants U01 HL 67893, U01 HL67894, U01 HL67895, U01 HL67896, U01 HL67897, U01 HL67898, U01 HL67899, U01 HL67900, U01 HL67901, and U01 HL67902. The CoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 33CSO-122661, 33CS30-139468 and 33CS30-148401). The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Kuopio, Tampere and Turku University Hospital Medical Funds (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation of Cardiovascular Research; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; and Yrjö Jahnsson Foundation. Noha A. Yousri was supported by the Biomedical Research Program at Weill



Cornell Medicine-Qatar, funded by the Qatar Foundation. The expert technical assistance in the statistical analyses by Irina Lisinen is gratefully acknowledged.

## **Conflict of Interest Statement**

The authors do not report any conflicts of interest.

## References

1. Utermann,G. and Beisiegel,U. (1979) Apolipoprotein A-IV: a protein occurring in human mesenteric lymph chylomicrons and free in plasma. Isolation and quantification. *Eur. J. Biochem.*, **99**, 333–343.
2. Wang,F., Kohan,A.B., Lo,C.-M., Liu,M., Howles,P. and Tso,P. (2015) Apolipoprotein A-IV: a protein intimately involved in metabolism. *J. Lipid Res.*, **56**, 1403–1418.
3. Steinmetz,A., Barbaras,R., Ghalim,N., Clavey,V., Fruchart,J.C. and Ailhaud,G. (1990) Human apolipoprotein A-IV binds to apolipoprotein A-I/A-II receptor sites and promotes cholesterol efflux from adipose cells. *J. Biol. Chem.*, **265**, 7859–7863.
4. Stein,O., Stein,Y., Lefevre,M. and Roheim,P.S. (1986) The role of apolipoprotein A-IV in reverse cholesterol transport studied with cultured cells and liposomes derived from an ether analog of phosphatidylcholine. *Biochim. Biophys. Acta*, **878**, 7–13.
5. Stoffel,W. (1984) Synthesis, transport, and processing of apolipoproteins of high density lipoproteins. *J. Lipid Res.*, **25**, 1586–1592.
6. Bisgaier,C.L., Sachdev,O.P., Megna,L. and Glickman,R.M. (1985) Distribution of apolipoprotein A-IV in human plasma. *J. Lipid Res.*, **26**, 11–25.
7. Duverger,N., Tresp,G., Caillaud,J.M., Emmanuel,F., Castro,G., Fruchart,J.C., Steinmetz,A. and Denèfle,P. (1996) Protection against atherogenesis in mice mediated by human apolipoprotein A-IV. *Science*, **273**, 966–968.
8. Cohen,R.D., Castellani,L.W., Qiao,J.H., Van Lenten,B.J., Lusic,A.J. and Reue,K. (1997) Reduced aortic lesions and elevated high density lipoprotein levels in transgenic mice overexpressing mouse apolipoprotein A-IV. *J. Clin. Invest.*, **99**, 1906–1916.
9. Kronenberg,F., Stühlinger,M., Trenkwalder,E., Geethanjali,F.S., Pachinger,O., von Eckardstein,A. and Dieplinger,H. (2000) Low apolipoprotein A-IV plasma concentrations in men with coronary artery disease. *J. Am. Coll. Cardiol.*, **36**, 751–757.
10. Manpuya,M.W., Guo,J. and Zhao,Y. (2001) The relationship between plasma apolipoprotein A-IV levels and coronary heart disease. *Chin. Med. J. (Engl.)*, **114**, 275–279.
11. Omori,M., Watanabe,M., Matsumoto,K., Honda,H., Hattori,H. and Akizawa,T. (2010) Impact of serum apolipoprotein A-IV as a marker of cardiovascular disease in maintenance hemodialysis patients. *Ther. Apher. Dial.*, **14**, 341–348.
12. Li,J., Song,M., Qian,D., Lu,W., Wang,J., Jiang,G., Jin,J., Wu,X. and Huang,L. (2013) Decreased plasma apolipoprotein A-IV levels in patients with acute coronary syndrome. *Clin. Investig. Med. Médecine Clin. Exp.*, **36**, E207–E215.
13. Kronenberg,F., Kuen,E., Ritz,E., König,P., Kraatz,G., Lhotta,K., Mann,J.F.E., Müller,G.A., Neyer,U., Riegel,W., *et al.* (2002) Apolipoprotein A-IV serum concentrations are elevated in patients with mild and moderate renal failure. *J. Am. Soc. Nephrol.*, **13**, 461–469.
14. Stangl,S., Kollerits,B., Lamina,C., Meisinger,C., Huth,C., Stöckl,A., Dähnhardt,D., Böger,C.A., Krämer,B.K., Peters,A., *et al.* (2015) Association between apolipoprotein A-IV concentrations and chronic kidney disease in two large population-based cohorts: results from the KORA studies. *J. Intern. Med.*, **278**, 410–423.
15. Lingenhel,A., Lhotta,K., Neyer,U., Heid,I.M., Rantner,B., Kronenberg,M.F., König,P., von Eckardstein,A., Schober,M., Dieplinger,H., *et al.* (2006) Role of the kidney in the metabolism of apolipoprotein A-IV: influence of the type of proteinuria. *J. Lipid Res.*, **47**, 2071–2079.
16. Boes,E., Fliser,D., Ritz,E., König,P., Lhotta,K., Mann,J.F.E., Müller,G.A., Neyer,U., Riegel,W.,

- Riegler, P., *et al.* (2006) Apolipoprotein A-IV predicts progression of chronic kidney disease: the mild to moderate kidney disease study. *J. Am. Soc. Nephrol.*, **17**, 528–536.
17. Zaiou, M., Visvikis, S., Gueguen, R., Steinmetz, J., Parra, H.J., Fruchart, J.C. and Siest, G. (1994) Sources of variability of human plasma apolipoprotein A-IV levels and relationships with lipid metabolism. *Genet. Epidemiol.*, **11**, 101–114.
  18. Gomaraschi, M., Putt, W.E., Pozzi, S., Iametti, S., Barbiroli, A., Bonomi, F., Favari, E., Bernini, F., Franceschini, G., Talmud, P.J., *et al.* (2010) Structure and function of the apoA-IV T347S and Q360H common variants. *Biochem. Biophys. Res. Commun.*, **393**, 126–130.
  19. Larson, I.A., Ordovas, J.M., Sun, Z., Barnard, Lohrmann, J., Feussner, G., Lamon-Fava, S. and Schaefer, E.J. (2002) Effects of apolipoprotein A-IV genotype on glucose and plasma lipoprotein levels. *Clin. Genet.*, **61**, 430–436.
  20. Wong, W.R., Hawe, E., Li, L.K., Miller, G.J., Nicaud, V., Pennacchio, L.A., Humphries, S.E. and Talmud, P.J. (2003) Apolipoprotein AIV gene variant S347 is associated with increased risk of coronary heart disease and lower plasma apolipoprotein AIV levels. *Circ. Res.*, **92**, 969–975.
  21. Talmud, P.J., Hawe, E., Martin, S., Olivier, M., Miller, G.J., Rubin, E.M., Pennacchio, L.A. and Humphries, S.E. (2002) Relative contribution of variation within the APOC3/A4/A5 gene cluster in determining plasma triglycerides. *Hum. Mol. Genet.*, **11**, 3039–3046.
  22. Hubacek, J.A., Skodova, Z., Adamkova, V., Vrablik, M., Horinek, A., Lanska, V., Ceska, R. and Poledne, R. (2003) Apolipoprotein AV gene polymorphisms (T-1131/C and Ser19/Trp) influence plasma triglyceride levels and risk of myocardial infarction. *Exp. Clin. Cardiol.*, **8**, 151–154.
  23. Ehnholm, C., Tenkanen, H., de Knijff, P., Havekes, L., Rosseneu, M., Menzel, H.J. and Tiret, L. (1994) Genetic polymorphism of apolipoprotein A-IV in five different regions of Europe. Relations to plasma lipoproteins and to history of myocardial infarction: the EARS study. European Atherosclerosis Research Study. *Atherosclerosis*, **107**, 229–238.
  24. Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., *et al.* (2013) Discovery and refinement of loci associated with lipid levels. *Nat. Genet.*, **45**, 1274–1283.
  25. Ou, H.J., Huang, G., Liu, W., Ma, X.L., Wei, Y., Zhou, T. and Pan, Z.M. (2015) Relationship of the APOA5/A4/C3/A1 gene cluster and APOB gene polymorphisms with dyslipidemia. *Genet. Mol. Res.*, **14**, 9277–9290.
  26. Delgado-Lista, J., Perez-Jimenez, F., Ruano, J., Perez-Martinez, P., Fuentes, F., Criado-Garcia, J., Parnell, L.D., Garcia-Rios, A., Ordovas, J.M. and Lopez-Miranda, J. (2010) Effects of variations in the APOA1/C3/A4/A5 gene cluster on different parameters of postprandial lipid metabolism in healthy young men. *J. Lipid Res.*, **51**, 63–73.
  27. Liu, Y., Ordovas, J.M., Gao, G., Province, M., Straka, R.J., Tsai, M.Y., Lai, C.-Q., Zhang, K., Borecki, I., Hixson, J.E., *et al.* (2009) Pharmacogenetic association of the APOA1/C3/A4/A5 gene cluster and lipid responses to fenofibrate: the genetics of lipid-lowering drugs and diet network study. *Pharmacogenet. Genomics*, **19**, 161–169.
  28. Wuepper, K.D. (1973) Prekallikrein deficiency in man. *J. Exp. Med.*, **138**, 1345–1355.
  29. Colman, R.W. (1999) Biologic activities of the contact factors in vivo--potentiation of hypotension, inflammation, and fibrinolysis, and inhibition of cell adhesion, angiogenesis and thrombosis. *Thromb. Haemost.*, **82**, 1568–1577.
  30. Schmaier, A.H. (2002) The plasma kallikrein-kinin system counterbalances the renin-angiotensin system. *J. Clin. Invest.*, **109**, 1007–1009.
  31. Shin, S.-Y., Fauman, E.B., Petersen, A.-K., Krumsiek, J., Santos, R., Huang, J., Arnold, M., Erte, I., Forgetta, V., Yang, T.-P., *et al.* (2014) An atlas of genetic influences on human blood metabolites.

*Nat. Genet.*, **46**, 543–550.

32. Suhre, K., Shin, S.-Y., Petersen, A.-K., Mohny, R.P., Meredith, D., Wägele, B., Altmaier, E., Deloukas, P., Erdmann, J., Grundberg, E., *et al.* (2011) Human metabolic individuality in biomedical and pharmaceutical research. *Nature*, **477**, 54–60.
33. Biswas, N., Maihofer, A.X., Mir, S.A., Rao, F., Zhang, K., Khandrika, S., Mahata, M., Friese, R.S., Hightower, C.M., Mahata, S.K., *et al.* (2016) Polymorphisms at the F12 and KLKB1 loci have significant trait association with activation of the renin-angiotensin system. *BMC Med. Genet.*, **17**, 21.
34. Musani, S.K., Fox, E.R., Kraja, A., Bidulescu, A., Lieb, W., Lin, H., Beecham, A., Chen, M.-H., Felix, J.F., Fox, C.S., *et al.* (2015) Genome-wide association analysis of plasma B-type natriuretic peptide in blacks: the Jackson Heart Study. *Circ. Cardiovasc. Genet.*, **8**, 122–130.
35. Verweij, N., Mahmud, H., Mateo Leach, I., de Boer, R.A., Brouwers, F.P., Yu, H., Asselbergs, F.W., Struck, J., Bakker, S.J.L., Gansevoort, R.T., *et al.* (2013) Genome-wide association study on plasma levels of midregional-proadrenomedullin and C-terminal-pro-endothelin-1. *Hypertension*, **61**, 602–608.
36. Björkqvist, J., Jämsä, A. and Renné, T. (2013) Plasma kallikrein: the bradykinin-producing enzyme. *Thromb. Haemost.*, **110**, 399–407.
37. Nakano, M., Kelly, E.J., Wiek, C., Hanenberg, H. and Rettie, A.E. (2012) CYP4V2 in Bietti's crystalline dystrophy: ocular localization, metabolism of  $\omega$ -3-polyunsaturated fatty acids, and functional deficit of the p.H331P variant. *Mol. Pharmacol.*, **82**, 679–686.
38. Nakano, M., Kelly, E.J. and Rettie, A.E. (2009) Expression and characterization of CYP4V2 as a fatty acid omega-hydroxylase. *Drug Metab. Dispos.*, **37**, 2119–2122.
39. Li, A., Jiao, X., Munier, F.L., Schorderet, D.F., Yao, W., Iwata, F., Hayakawa, M., Kanai, A., Shy Chen, M., Alan Lewis, R., *et al.* (2004) Bietti crystalline corneoretinal dystrophy is caused by mutations in the novel gene CYP4V2. *Am. J. Hum. Genet.*, **74**, 817–826.
40. Teslovich, T.M., Musunuru, K., Smith, A. V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., *et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**, 707–713.
41. Waterworth, D.M., Ricketts, S.L., Song, K., Chen, L., Zhao, J.H., Ripatti, S., Aulchenko, Y.S., Zhang, W., Yuan, X., Lim, N., *et al.* (2010) Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.*, **30**, 2264–2276.
42. Dichgans, M., Malik, R., König, I.R., Rosand, J., Clarke, R., Gretarsdottir, S., Thorleifsson, G., Mitchell, B.D., Assimes, T.L., Levi, C., *et al.* (2014) Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke*, **45**, 24–36.
43. Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F.R., Barbalic, M., Gieger, C., *et al.* (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.*, **43**, 333–338.
44. Do, R., Willer, C.J., Schmidt, E.M., Sengupta, S., Gao, C., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., *et al.* (2013) Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat. Genet.*, **45**, 1345–1352.
45. Pattaro, C., Teumer, A., Gorski, M., Chu, A.Y., Li, M., Mijatovic, V., Garnaas, M., Tin, A., Sorice, R., Li, Y., *et al.* (2016) Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.*, **7**, 10023.
46. Kronenberg, F., König, P., Neyer, U., Auinger, M., Pribasnik, A., Lang, U., Reitingner, J., Pinter, G., Utermann, G. and Dieplinger, H. (1995) Multicenter study of lipoprotein(a) and apolipoprotein(a) phenotypes in patients with end-stage renal disease treated by hemodialysis or continuous

ambulatory peritoneal dialysis. *J. Am. Soc. Nephrol.*, **6**, 110–120.

47. Lagrost,L., Gambert,P., Meunier,S., Morgado,P., Desgres,J., d’Athis,P. and Lallemand,C. (1989) Correlation between apolipoprotein A-IV and triglyceride concentrations in human sera. *J. Lipid Res.*, **30**, 701–710.
48. Kohan,A.B., Wang,F., Lo,C.-M., Liu,M. and Tso,P. (2015) ApoA-IV: current and emerging roles in intestinal lipid metabolism, glucose homeostasis, and satiety. *Am. J. Physiol. Gastrointest. Liver Physiol.*, **308**, G472–G481.
49. Hanniman,E.A., Lambert,G., Inoue,Y., Gonzalez,F.J. and Sinal,C.J. (2006) Apolipoprotein A-IV is regulated by nutritional and metabolic stress: involvement of glucocorticoids, HNF-4 alpha, and PGC-1 alpha. *J. Lipid Res.*, **47**, 2503–2514.
50. Benjamini,Y. and Hochberg,Y. (1995) Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.*, **57**, 289–300.
51. Firmann,M., Mayor,V., Vidal,P.M., Bochud,M., Pécoud,A., Hayoz,D., Paccaud,F., Preisig,M., Song,K.S., Yuan,X., *et al.* (2008) The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.*, **8**, 6.
52. Howie,B., Fuchsberger,C., Stephens,M., Marchini,J. and Abecasis,G.R. (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.*, **44**, 955–959.
53. Higgins,M., Province,M., Heiss,G., Eckfeldt,J., Ellison,R.C., Folsom,A.R., Rao,D.C., Sprafka,J.M. and Williams,R. (1996) NHLBI Family Heart Study: objectives and design. *Am. J. Epidemiol.*, **143**, 1219–1228.
54. Li,Y., Willer,C.J., Ding,J., Scheet,P. and Abecasis,G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.
55. Marchini,J., Howie,B., Myers,S., McVean,G. and Donnelly,P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
56. Heid,I.M., Boes,E., Müller,M., Kollerits,B., Lamina,C., Coassin,S., Gieger,C., Döring,A., Klopp,N., Frikke-Schmidt,R., *et al.* (2008) Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based KORA study sheds new light on intergenic regions. *Circ. Cardiovasc. Genet.*, **1**, 10–20.
57. Raitakari,O.T., Juonala,M., Rönkä,T., Keltikangas-Järvinen,L., Räsänen,L., Pietikäinen,M., Hutri-Kähönen,N., Taittonen,L., Jokinen,E., Marniemi,J., *et al.* (2008) Cohort profile: the cardiovascular risk in Young Finns Study. *Int. J. Epidemiol.*, **37**, 1220–1226.
58. Kronenberg,F., Kronenberg,M.F., Kiechl,S., Trenkwalder,E., Santer,P., Oberhollenzer,F., Egger,G., Utermann,G. and Willeit,J. (1999) Role of lipoprotein(a) and apolipoprotein(a) phenotype in atherogenesis: prospective results from the Bruneck study. *Circulation*, **100**, 1154–1160.
59. Kiechl,S., Willeit,J., Mayr,M., Viehweider,B., Oberhollenzer,M., Kronenberg,F., Wiedermann,C.J., Oberthaler,S., Xu,Q., Witztum,J.L., *et al.* (2007) Oxidized phospholipids, lipoprotein(a), lipoprotein-associated phospholipase A2 activity, and 10-year cardiovascular outcomes: prospective results from the Bruneck study. *Arterioscler. Thromb. Vasc. Biol.*, **27**, 1788–1795.
60. Heid,I.M., Wagner,S.A., Gohlke,H., Iglseider,B., Mueller,J.C., Cip,P., Ladurner,G., Reiter,R., Stadlmayr,A., Mackevics,V., *et al.* (2006) Genetic architecture of the APM1 gene and its influence on adiponectin plasma levels and parameters of the metabolic syndrome in 1,727 healthy Caucasians. *Diabetes*, **55**, 375–384.
61. Kronenberg,F., Lobentanz,E.M., König,P., Utermann,G. and Dieplinger,H. (1994) Effect of sample storage on the measurement of lipoprotein[a], apolipoproteins B and A-IV, total and high

- density lipoprotein cholesterol and triglycerides. *J. Lipid Res.*, **35**, 1318–1328.
62. Behrens,G., Winkler,T.W., Gorski,M., Leitzmann,M.F. and Heid,I.M. (2011) To stratify or not to stratify: power considerations for population-based genome-wide association studies of quantitative traits. *Genet. Epidemiol.*, **35**, 867–879.
  63. Winkler,T.W., Day,F.R., Croteau-Chonka,D.C., Wood,A.R., Locke,A.E., Mägi,R., Ferreira,T., Fall,T., Graff,M., Justice,A.E., *et al.* (2014) Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.*, **9**, 1192–1212.
  64. Han,B. and Eskin,E. (2011) Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 586–598.
  65. Pruim,R.J., Welch,R.P., Sanna,S., Teslovich,T.M., Chines,P.S., Gliedt,T.P., Boehnke,M., Abecasis,G.R. and Willer,C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.
  66. Yang,J., Lee,S.H., Goddard,M.E. and Visscher,P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
  67. Li,M.-X., Sham,P.C., Cherny,S.S. and Song,Y.-Q. (2010) A knowledge-based weighting framework to boost the power of genome-wide association studies. *PLoS One*, **5**, e14480.
  68. Li,M.-X., Gui,H.-S., Kwan,J.S.H. and Sham,P.C. (2011) GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am. J. Hum. Genet.*, **88**, 283–293.
  69. Li,M.-X., Kwan,J.S.H. and Sham,P.C. (2012) HYST: a hybrid set-based test for genome-wide association studies, with application to protein-protein interaction-based association analysis. *Am. J. Hum. Genet.*, **91**, 478–488.
  70. de Los Campos,G., Sorensen,D. and Gianola,D. (2015) Genomic heritability: what is it? *PLoS Genet.*, **11**, e1005048.
  71. Pattaro,C., Teumer,A., Gorski,M., Chu,A.Y., Li,M., Mijatovic,V., Garnaas,M., Tin,A., Sorice,R., Li,Y., *et al.* (2016) Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat. Commun.*, **7**.
  72. Lamina,C., Coassin,S., Illig,T. and Kronenberg,F. (2011) Look beyond one’s own nose: combination of information from publicly available sources reveals an association of GATA4 polymorphisms with plasma triglycerides. *Atherosclerosis*, **219**, 698–703.
  73. Adzhubei,I., Jordan,D.M. and Sunyaev,S.R. (2013) Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.*, **Chapter 7**, Unit7.20.
  74. Kumar,P., Henikoff,S. and Ng,P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.*, **4**, 1073–1081.
  75. Li,B., Krishnan,V.G., Mort,M.E., Xin,F., Kamati,K.K., Cooper,D.N., Mooney,S.D. and Radivojac,P. (2009) Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*, **25**, 2744–2750.
  76. Arnold,M., Raffler,J., Pfeufer,A., Suhre,K. and Kastenmüller,G. (2015) SNIIPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics*, **31**, 1334–1336.

## Figure legends

**Figure 1:** Overview of contributing cohorts in the discovery and replication stage

**Figure 2:** Manhattan-plot for the meta-analysis on log-transformed apoA-IV values. Results are based on five discovery cohorts including 13813 individuals.

**Figure 3:** Regional plot showing the genomic region defined by the *APOA4*-lead SNP rs1729407 +/- 500 kB (LD refers to rs1729407, based on 1000G EUR); p-values are derived from the meta-analysis on the five discovery cohorts on log-transformed apoA-IV concentrations.

**Figure 4:** Regional plot showing the genomic region defined by the *KLKB1*-lead SNP rs4241819 +/- 500 kB (LD is based on 1000G EUR and refers to rs4253311 which was used as a proxy for the lead SNP in the replication studies); additionally, the missense variant rs3733402 is marked. P-values are derived from the meta-analysis on the five discovery cohorts on log-transformed apoA-IV concentrations.



**Table 1:** Meta-analysis results of selected SNPs for replication (p-value < 1E-06 in the GWAs stage); the beta estimate and effect direction refer to the minor allele.

SNP	GWAS stage				Replication stage			GWAS + Replication		
	$\beta$ (se)*	Effect Direction <sup>§</sup>	p	I <sup>2</sup>	$\beta$ (se)*	Effect Direction <sup>§</sup>	p / FDR <sup>§</sup>	$\beta$ (se)*	p	I <sup>2</sup>
<b><i>APOA4</i> gene region</b>										
rs1729407	-0.2459 (0.0289)	-----	6.00E-40	0	-0.4895 (0.1003)	--	3.73E-06 / 2.59E-05	-0.2645 (0.0277)	6.77E-44	25.34
rs5104	-0.2399 (0.0367)	-----	1.24E-22	15.63	-0.4533 (0.1460)	--	0.0013 / 0.0046	-0.2526 (0.0356)	1.79E-24	11.05
rs5110	0.2301 (0.0774)	+++++	9.26E-07	0	0.0520 (0.1956)	- +	0.9124 / 0.9124	0.2060 (0.0720)	1.44E-05	1.37
<b><i>APOA4</i> gene region, selected from literature</b>										
rs675	-0.1041 (0.0380)	-----	0.0021	0	-0.1462 (0.1183)	--	0.2931 / 0.5129	-0.1081 (0.0362)	0.0013	0
<b>Other gene regions</b>										
<i>KLKB1</i> : rs4241819/rs4253311	0.1395 (0.0280)	+++++	1.08E-12	45.89	0.2410 (0.1006)	++	0.0093 / 0.0217	0.1469 (0.0270)	5.63E-14	36.03
<i>SOWAHA</i> : rs59698941	-0.3542 (0.1420) <sup>#</sup>	-----	3.76E-07 <sup>&amp;</sup>	68.83	-0.0162 (0.1638) <sup>#</sup>	+ -	0.6637 <sup>&amp;</sup> / 0.7743	-0.2628 (0.1111) <sup>#</sup>	1.75E-06 <sup>&amp;</sup>	64.86
<i>SHROOM1</i> : rs2292030	-0.3502 (0.1407) <sup>#</sup>	-----	9.24E-07 <sup>&amp;</sup>	66.33	-0.0222 (0.1491) <sup>#</sup>	+ -	0.6584 <sup>&amp;</sup> / 0.7743	-0.2629 (0.1090) <sup>#</sup>	4.28E-06 <sup>&amp;</sup>	59.86

\* All effect estimates ( $\beta$  and se) are based on the original scale of apoA-IV, either fixed effect or random effect. Where labeled, all effect estimates refer to the minor allele derived from 1000G, phase 3v5 (see Supplementary Table 3);

<sup>§</sup> Order of included GWA studies: CoLaus, FamHS, KORA F3, KORA F4, YFS. +: positive effect from minor allele on log(apoA-IV) in that specific study, -: inverse effect from minor allele on log(apoA-IV) in that specific study; ?: SNP not available in that study; Order of included studies at replication stages: Bruneck, SAPHIR

<sup>&</sup> P-values are derived from the method proposed by Han and Eskin;

<sup>#</sup> Random effects  $\beta$  and se

<sup>§</sup> False-discovery rate by Benjamini and Hochberg (50)

**Table 2:** Results of gene-based analysis

Gene	Nominal p-value	Bonferroni-corrected p-value	Chromosome	Position	Group
<i>ZPR1</i>	7.93E-39	1.9E-34	11	116649275	protein-coding gene
<i>APOA5</i>	2.14E-38	5.38E-34	11	116660085	protein-coding gene
<i>APOA4</i>	4.06E-38	1.02E-33	11	116691417	protein-coding gene
<i>APOC3</i>	7.2E-38	1.81E-33	11	116700623	protein-coding gene
<i>APOA1</i>	1.23E-29	3.09E-25	11	116706468	protein-coding gene
<i>APOA1-AS</i>	1.53E-29	3.84E-25	11	116706832	non-coding RNA
<i>SIK3</i>	2.78E-29	6.99E-25	11	116714117	protein-coding gene
<i>BUD13</i>	5.57E-23	1.40E-18	11	116618885	protein-coding gene
<i>PAFAH1B2</i>	1.32E-18	3.32E-14	11	117014999	protein-coding gene
<i>SIDT2</i>	1.08E-16	2.71E-12	11	117049938	protein-coding gene
<i>TAGLN</i>	1.69E-16	4.25E-12	11	117070039	protein-coding gene
<i>LOC100652768</i>	1.85E-16	4.65E-12	11	117066328	unknown
<i>PCSK7</i>	3.59E-16	9.02E-12	11	117075786	protein-coding gene
<i>KLKB1</i>	5.86E-11	1.47E-06	4	187148671	protein-coding gene
<i>RNF214</i>	3.24E-10	8.14E-06	11	117103451	protein-coding gene
<i>F11</i>	9.03E-10	2.27E-05	4	187187117	protein-coding gene
<i>CYP4V2</i>	1.21E-09	3.04E-05	4	187112673	protein-coding gene
<i>FLJ38576</i>	3.35E-08	8.42E-04	4	187110185	unknown

**Table 3:** Look-up of lead apoA-IV SNPs in other consortia, providing p-values

SNPs	P-value of association with...		
	TG	HDL-C	eGFR
rs1729407 ( <i>APOA4</i> )	0.96	7.1E-07	0.73
rs4253311 ( <i>KLKB1</i> )	0.07	0.54	0.17

**Table 4:** Association of weighted SNP-scores including susceptibility SNPs for kidney function, HDL cholesterol and triglycerides (TG) with log-transformed apoA-IV levels (age-and sex-adjusted) in the combined KORAF3 and F4 dataset using a mixed effects model.

Weighted SNP scores	beta	se	p-value	Explained variance in %
Kidney function SNP-score (53 SNPs)	-0.4068	0.1008	5.50E-05	0.33%
HDL cholesterol SNP-score (70 SNPs)	0.0246	0.0130	0.0575	0.06%
Triglycerides SNP-score (38 SNPs)	-0.0430	0.0161	0.0078	0.12%

## Abbreviations

apoA-IV	apolipoprotein A-IV
eGFR	estimated glomerular filtration rate
GWAS	genome-wide association study
GWA	genome-wide analysis
HDL	high-density lipoprotein
HDL-C	high-density lipoprotein cholesterol
SNP	single nucleotide polymorphism
LD	linkage disequilibrium

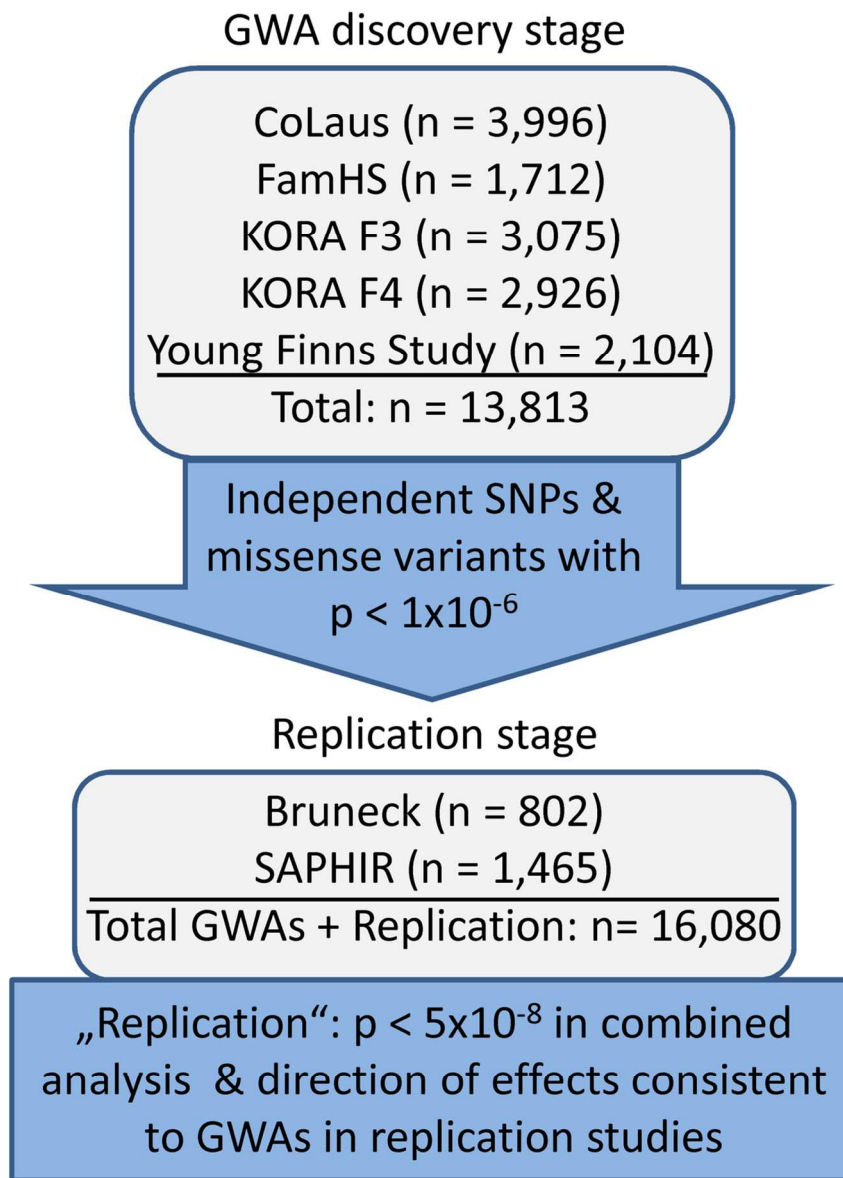


Figure 1 Overview of contributing cohorts in the discovery and replication stage  
108x147mm (300 x 300 DPI)

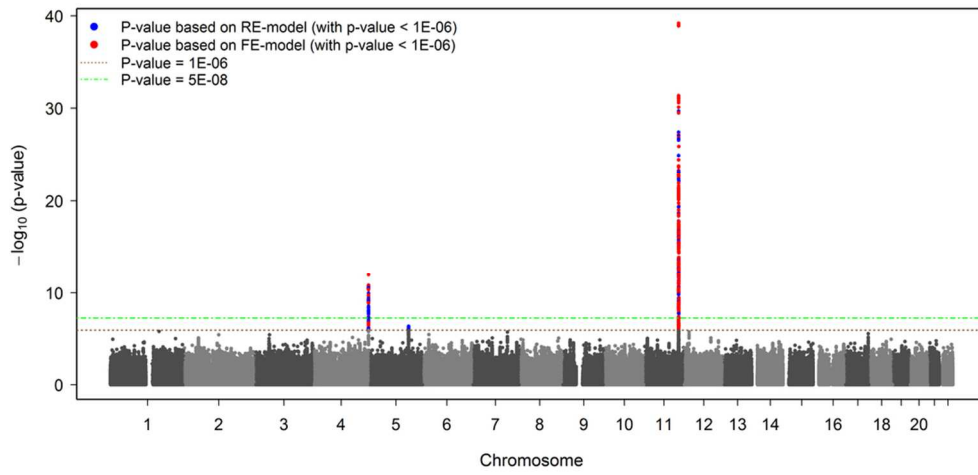


Figure 2 Manhattan-plot for the meta-analysis on log-transformed apoA-IV values. Results are based on five discovery cohorts including 13813 individuals.  
95x47mm (300 x 300 DPI)



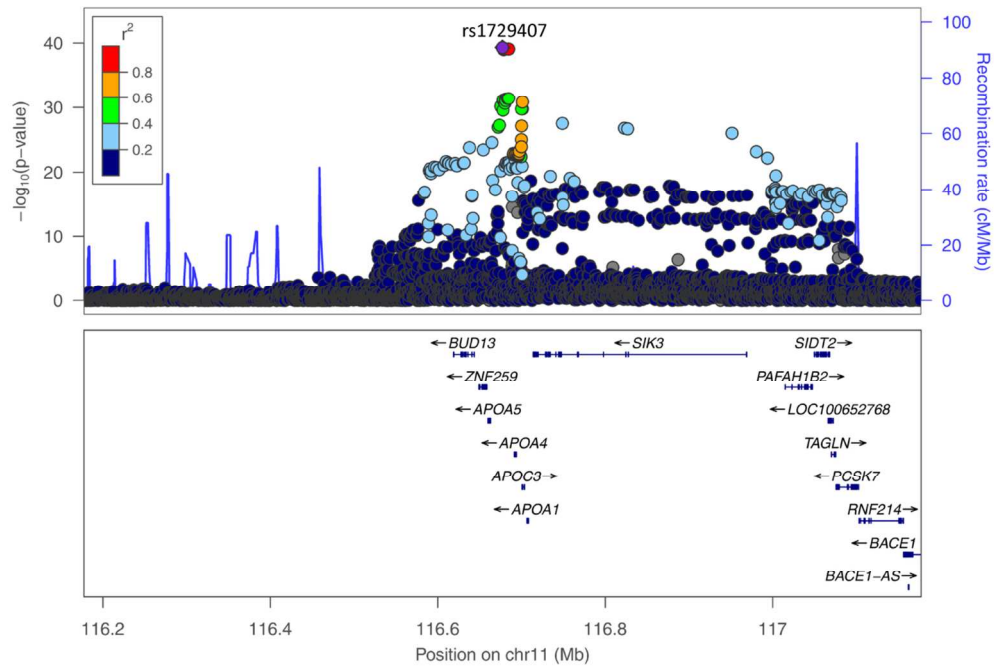


Figure 3 Regional plot showing the genomic region defined by the APOA4-lead SNP rs1729407 +/- 500 kb (LD refers to rs1729407, based on 1000G EUR); p-values are derived from the meta-analysis on the five discovery cohorts on log-transformed apoA-IV concentrations.  
119x79mm (300 x 300 DPI)

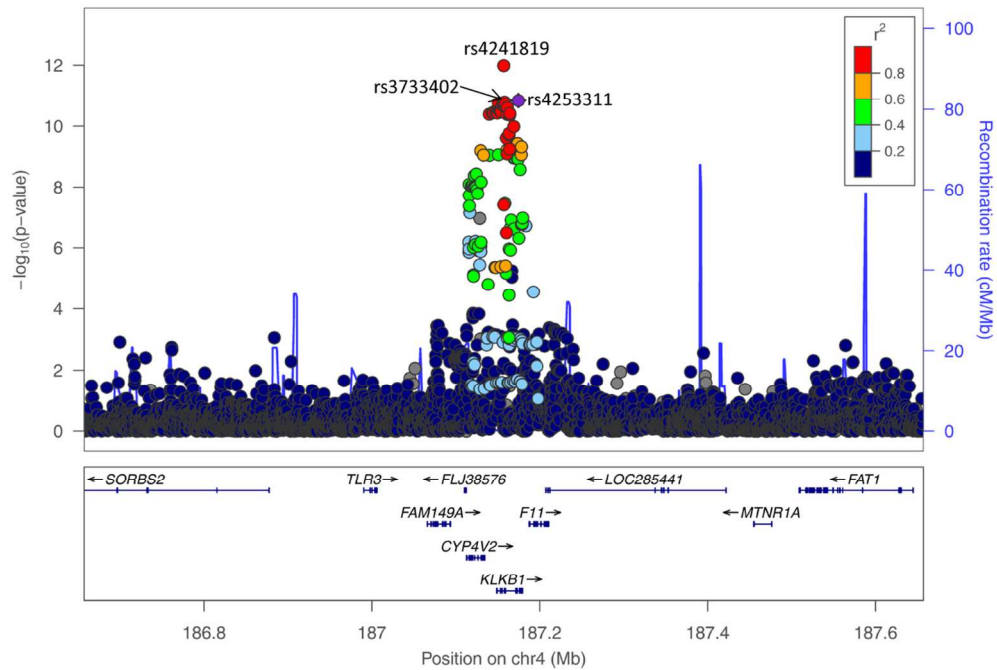


Figure 4 Regional plot showing the genomic region defined by the KLKB1-lead SNP rs4241819 +/- 500 kB (LD is based on 1000G EUR and refers to rs4253311 which was used as a proxy for the lead SNP in the replication studies); additionally, the missense variant rs3733402 is marked. P-values are derived from the meta-analysis on the five discovery cohorts on log-transformed apoA-IV concentrations.  
 119x80mm (300 x 300 DPI)