Whole-genome linkage scan combined with exome sequencing identifies novel candidate genes for carotid intima-media thickness

Dina Vojinovic, Maryam Kavousi, Mohsen Ghanbari, Rutger W.W. Brouwer, Jeroen G.J. van Rooij, Mirjam C.G.N. van den Hout, Robert Kraaij, Wilfred F.J. van IJcken, Andre G. Uitterlinden, Cornelia M. van Duijn, Najaf Amin

This chapter is accepted for publication in Frontiers in Genetics.

The supplemental information for this paper is available at https://drive.google.com/drive/folders/13m2uhJ5MJ2kjsvaH5CqoyNHCV-IIAcWu?usp=sharing

ABSTRACT

Carotid intima-media thickness (cIMT) is an established heritable marker for subclinical atherosclerosis. In this study, we aim to identify rare variants with large effects driving differences in cIMT by performing genome-wide linkage analysis of individuals in the extremes of cIMT trait distribution (> 90th percentile) in a large family-based study from a genetically isolated population in the Netherlands. Linked regions were subsequently explored by fine-mapping using exome sequencing. We observed significant evidence of linkage on chromosomes 2p16.3 (rs1017418, heterogeneity LOD (HLOD) = 3.35), 19q13.43 (rs3499, HLOD = 9.09), 20p13 (rs1434789, HLOD = 4.10) and 21q22.12 (rs2834949, HLOD = 3.59). Fine-mapping using exome sequencing data identified a non-coding variant (rs62165235) in *PNPT1* gene under the linkage peak at chromosome 2 that is likely to have a regulatory function. The variant was associated with quantitative cIMT in the family-based study population (effect = 0.27, *p*-value = 0.013). Furthermore, we identified several genes under the 21q22 linkage peak highly expressed in tissues relevant for atherosclerosis. To conclude, our linkage analysis identified four genomic regions significantly linked to cIMT. Further analyses are needed to demonstrate involvement of identified candidate genes in development of atherosclerosis.

INTRODUCTION

Cardiovascular diseases, including heart and cerebrovascular diseases, are listed among the leading causes of death in developed countries. [1] The underlying pathology in the majority of cases is atherosclerosis.[2] Carotid intima-media thickness (cIMT), a quantitative measure of carotid artery wall thickening, is a marker for subclinical atherosclerosis that has been shown to predict future cardiovascular events in large epidemiological studies.[3-5] cIMT is determined by both traditional cardiovascular risk factors, such as aging, blood pressure, body mass index, plasma lipid levels, diabetes mellitus or smoking, and genetic factors.[6] Genetic factors play a key role in the etiology of cIMT with heritability estimates ranging from 30-60%.[7,8] Several genome-wide linkage studies of quantitative cIMT published up to date, reported significant and suggestive evidence of linkage on chromosomes 2q33-q35, 6p12-p22, 7p, 11q23, 12q24, 13q32-q33, and 14q31.[8-11] The largest genome-wide association study (GWAS) of cIMT, including 42,484 individuals, identified only three genomic regions of common non-coding genetic variation on 8q24 (near *ZHX2*), 19q13 (near *APOC1*) and 8q23.1 (*PINX1*) and an additional suggestive region on 6p22 (near *SLC17A4*).[12] In addition, an exome-wide association study in 52,869 individuals identified the association of protein-coding variants in *APOE* with cIMT.[13] The identified variants provide valuable insights into the genetic architecture of cIMT but explain a small proportion of the trait variance.[12] A previous sequencing study of cIMT candidate regions in population-based cohorts yielded inconclusive results due to limited power.[14] A more powerful approach for uncovering the role of rare variants is a family-based study design due to the higher frequency of the rare variants.[15] The chances of success for family-based studies are even higher in genetic isolates since rare variants become more frequent due to founder effect, genetic drift and inbreeding.[15-17]

In this study, we hypothesized that there may be rare variants with large effects driving differences in cIMT independently of traditional cardiovascular risk factors and that these variants are enriched in the extremes of the cIMT distribution. To the best of our knowledge, no study to date explored extremes of quantitative cIMT. However, this approach has been demonstrated as successful for some other quantitative traits. Following the same approach as described in our study, Amin *et al.* successfully identified a rare variant of large effect in large extended families.[18] To discover such variants in the extremes of cIMT distribution, we performed affected-only genome-wide linkage analysis of cIMT followed by fine-mapping using exome sequencing in a large family-based study from a genetically isolated population in the Netherlands.

MATERIAL AND METHODS

Study population

Our discovery population consisted of participants from Erasmus Rucphen Family (ERF) study. ERF is a family-based cohort that includes around 3,000 inhabitants of a genetically isolated community in the South-West of the Netherlands.[19] The community was constituted as a religious isolate at the middle of the 18th century by a limited number of founders.[19] The population has remained in isolation with minimal immigration rate and high inbreeding.[19,20] All ERF participants are living descendants of a limited number of founders living in the 19th century. The Medical Ethical Committee of the Erasmus University Medical Center, Rotterdam approved the study. Written informed consent was obtained from all participants.

Phenotypes

Participants from ERF underwent extensive clinical examination between 2002 and 2005. cIMT was measured using high-resolution B-mode ultrasonography with a 7.5-MHz linear array transducer (ATL UltraMark IV). Maximum cIMT was measured on the 3 still, longitudinal, two-dimensional ultrasound images of the near and far wall from both left and right arteries, as described previously.[21] The mean value of these measurements was used for the analyses.

Information on covariates for both studies included age, sex, and smoking status. Body mass index (BMI) was defined as weight divided by the square of height (kg/m^2) and waist-hip ratio (WHR) was computed by dividing the waist and hip circumferences with each other. Hypertension was defined as systolic blood pressure above 140 mmHg, diastolic blood pressure above 90 mmHg or use of medication for treatment of hypertension. Dyslipidemia was defined as total cholesterol above 6.2 mmol/L or use of lipid-lowering medication, whereas diabetes was defined as fasting plasma glucose levels above 7 mmol/L, random plasma glucose above 11.1 mmol/L or use of medication indicated for treatment of diabetes.

Genotyping

Genotyping on the Illumina 6K Array

Genomic DNA was extracted from peripheral venous blood of all study participants using the salting out procedure.[22] Genotyping was performed using the 6K Illumina Linkage IV Panels (Illumina, San Diego, CA, USA) at the Centre National de Genotypage in France. Markers with a minor allele frequency (MAF) < 5%, call rate < 98% or which failed an exact test of Hardy-Weinberg equilibrium (HWE) (p-value < 10^{-8}) were removed

during the quality control process. In total 5,250 autosomal variants were available for analysis.

Exome sequencing

The exomes of randomly selected participants from the ERF study were sequenced at the Cell Biology Department of the Erasmus MC, The Netherlands. Sequencing was performed at a median depth of 57× using the Agilent version V4 capture kit on an Illumina Hiseq2000 sequencer using the TruSeq Version 3 protocol.[23,24] After quality control, we retrieved 528,617 single nucleotide variants (SNVs) in 1,308 individuals, of which 1,046 had cIMT data available. Annotation of the SNVs was performed using the SeattleSeq annotation database (http://snp.gs.washington.edu/SeattleSeqAnnotation138/). To further assess the functionality of the variants, we used RegulomeDB database that annotates SNVs with known and predicted regulatory elements and Combined Annotation Dependent Depletion (CADD) tool for scoring the deleteriousness of variants.[25,26] The ERF data is available in the European Genome-phenome Archive (EGA) public repository with ID number EGAS00001001134.

Statistical analysis

Linkage analysis

We performed affected only genome-wide multipoint non-parametric linkage (NPL) analysis in MERLIN 1.1.2 using individuals from the ERF study.[27] Individuals that scored above the 90th percentile of the distribution of the residuals from the regression of cIMT onto age, age^2, sex, smoking status, BMI, WHR, diabetes, dyslipidemia, and hypertension were set as affected (N = 103). Descriptive characteristics of the selected individuals are presented in **Table 1**. The selected individuals were older and higher cIMT measurements compared to all ERF study participants (**Table 1**). They also had a higher prevalence of hypertension, dyslipidemia, and diabetes than all ERF study participants, whereas the body mass index and waist to hip ratio were comparable (**Table 1**). These 103 affected individuals were connected to each other in a large pedigree consisting of 5,083 individuals. To facilitate linkage analysis, the 103 affected individuals were clustered into 21 smaller non-overlapping sub-pedigrees with a maximum bit size of 24 using the PEDCUT software version 1.19.[28] Bit size value is used to characterize the maximal number of subjects of interest who share a common ancestor.[28] The number of affected subjects of interest in the sub-pedigrees ranged from two to eight. MEGA2 software tool version 4.4[29] was used to create input files for MERLIN. Mendelian inconsistencies were set to missing within the whole sub-pedigree. There were 543 Mendelian inconsistencies observed among 5,250 autosomal variants. After they were set to missing, 4,707 autosomal variants were used in the linkage analysis. We also performed affected only parametric

linkage analysis under the dominant and recessive models assuming incomplete penetrance of 0.5 and a disease allele frequency of 0.01 using MERLIN. Marker allele frequencies were calculated from all genotyped individuals in the pedigrees. Subsequently, we carried out per family analyses in order to identify families that were contributing predominantly to the linkage signals, henceforth referred to as 'contributing families'. Additionally, we performed variance component linkage analysis in MERLIN using quantitative cIMT in the total study population. To facilitate analysis PEDCUT software was used to cluster individuals into 116 non-overlapping sub-pedigrees. The number of subjects of interest in the sub-pedigrees ranged from two to eighteen. To determine the significance of each test, the logarithm of the odds (LOD) score was calculated as the log10 of the likelihood ratio. The LOD score of 3.3 or higher was considered to represent genome-wide significance threshold, whereas the LOD score of 1.9 was used to declare genome-wide suggestive threshold.[30]

Identification of variants under the linkage peaks using exome sequencing

We used exome sequence data to identify variants that could explain observed linkage peaks. To this end, we looked for variants that were shared among the majority of affected individuals from the contributing families within the respective linkage peak. We only considered variants with MAF < 5% or absent in 1000 Genome Project (1kG) and MAF < 5% in the ERF controls which were defined as individuals who scored below the mean of the distribution of the residuals from the regression of cIMT onto age, age^2, sex, smoking status, BMI, WHR, diabetes, dyslipidemia, and hypertension. The

Table 1. Descriptive statistics of study populations including ERF cases (N=103) selected for the linkage analysis and ERF overall.

Characteristics	ERF cases	ERF overall
Age, mean (sd)	53.6 (13.6)	48.3 (14.2)
Gender, % of males	45.6%	40.2%
IMT (mm), mean (sd)	1.1 (0.2)	0.8 (0.2)
Smoking, % of ever smokers	44.7%	41.9%
BMI (kg/m^2), mean (sd)	26.9 (3.7)	26.7 (4.4)
WHR, mean (sd)	0.9 (0.1)	0.9 (0.1)
Hypertension, % of cases with hypertension	63.1%	48.7%
Dyslipidemia, % of cases with dyslipidemia	51.5%	36.2%
Diabetes, % of patients with diabetes	6.8%	4.5%

Abbreviations: IMT - intima-media thickness, BMI - body mass index, WHR - waist to hip ratio; Hypertension: systolic blood pressure above 140 mmHg, diastolic blood pressure above 90 mmHg or use of medication for treatment of hypertension; Dyslipidemia - total cholesterol above 6.2 mmol/L or use of lipid-lowering medication; Diabetes - fasting plasma glucose levels above 7 mmol/L, random plasma glucose above 11.1 mmol/L or use of medication indicated for the treatment of diabetes;

MAF of variants absent in 1kG project was checked in NHLBI Exome Sequencing Project (http://evs.gs.washington.edu/EVS/). Candidate variants were subjected to quantitative trait association analysis with cIMT in the ERF under the same model as in the sharing analysis (additive, dominant, recessive) using the RVtests software.[31] Inverse normalized residuals from the regression of cIMT onto age, age^2, sex, smoking status, BMI, WHR, diabetes, dyslipidemia, and hypertension were used in the association analysis. To take into account multiple tests, we first calculated a number of independent tests using the method of Li and Ji.[32] Subsequently, Bonferroni corrected *p*-value was calculated based on number of independent tests. GTEx portal (https://www.gtexportal.org/home/) was used to check for gene expression.

RESULTS

The results of affected only genome-wide non-parametric and parametric linkage scans are illustrated in **Figure 1**. Regions with significant (LOD > 3.3) evidence of linkage in either the non-parametric or the parametric analyses are shown in **Table 2**. Significant evidence of linkage for cIMT was observed to chromosomes 2p16.3, 19q13.43, 20p13, and 21q22.12 in the parametric linkage analysis under the dominant model, and to chromosomes 19q13.43 and 20p13 in the parametric linkage analysis under the recessive model. The families contributing predominantly to these linkage peaks and the distribution of their per-family heterogeneity LOD (HLOD) scores are shown in **Supplementary Figure 1-4**.

Table 2. Genome-wide significant results of linkage analyses for cIMT. Start and end SNV are reported for base to base linkage regions.

Region*	Start SNV	End SNV	Start position**	End position**	SNV with max LOD	Dominant (HLOD)	Recessive (HLOD)	Non-parametric (LOD)
2p16.3	rs1447107	rs1017267	45272197	56785785	rs1017418	**3.35**	2.88	1.40
19q13.43	rs897783	rs3499	52031162	59093484	rs3499	**7.17**	**9.09**	**3.73**
20p13	rs1434789	rs241605	137900	3915064	rs1434789	**4.10**	**3.87**	**3.34**
21q22.12	rs762173	rs2836301	33832675	40351780	rs2834949	**3.59**	1.86	2.14

Abbreviations: SNV - Single Nucleotide Variant; HLOD - heterogeneity LOD;
*Regions with significant evidence of linkage are in bold;
**Start and end positions correspond to genetic position of SNV at the start or end of the base to base linkage region according to hg19 assembly;

We next determined to what extent the affected members in these families shared rare variants under the linkage peaks. Sharing analyses under the base to base linkage peak at 2p16 (family specific HLOD = 3.63) identified intronic and coding-synonymous vari-

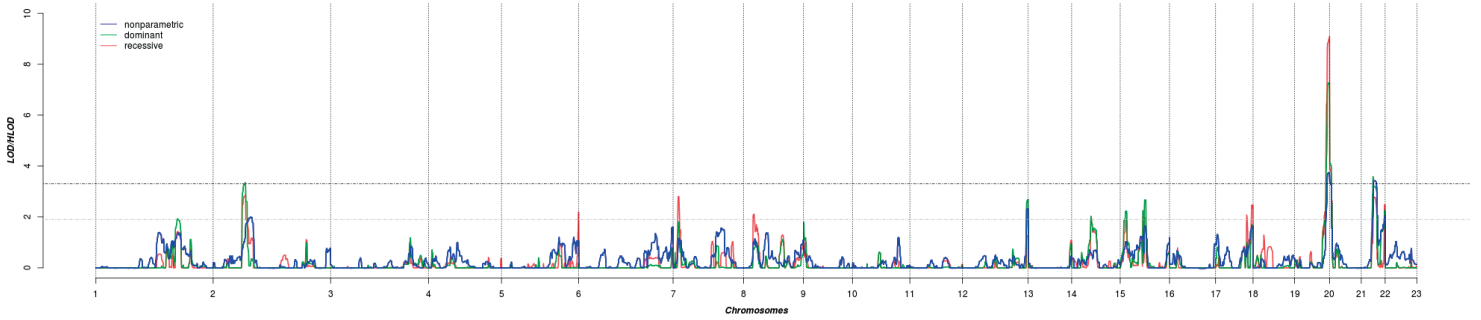


Figure 1. The results of genome-wide linkage scan for non-parametric (blue) and parametric analyses under the dominant (green) and recessive (red) models. The x-axis shows 22 autosomal chromosomes, whereas the y-axis shows the LOD scores for non-parametric model and heterogeneity LOD (HLOD) scores for dominant and recessive model. Black dotted line depicts the genome-wide significant threshold, whereas gray dotted line shows the suggestive threshold.

Table 3. Variants shared among the affected family members of the family that predominantly contributed to the LOD score at 2p16

Name*	Function	Gene	MAF ERF controls**	MAF 1kG***	CADD	Regulome DB	Association analysis in ERF			
							Beta	$Beta_{untransformed}$	SE	P
rs375801385	intron	*FBXO11*	0.031	0.056	17.18	6	0.023	0.014	0.145	0.876
2:48848294	intron	*GTF2A1L*	0.031	NA	5.82	-	0.023	0.014	0.145	0.876
rs149304214	coding-synonymous	*SPTBN1*	0.016	0.002	15.95	-	0.311	0.059	0.182	0.087
rs62165235	intron	*PNPT1*	0.044	0.038	4.413	2b	0.265	0.037	0.107	0.013
rs114706375	intron	*USP34*	0.026	0.010	0.122	5	0.124	0.036	0.145	0.393
rs144629927	intron	*XPO1*	0.021	0.003	4.595	3a	0.131	0.041	0.159	0.409

Abbreviations: MAF - minor allele frequency; 1kG - 1000 Genomes; CADD - Combined Annotation Dependent Depletion score; Regulome DB - Regulome DB score; Beta - effect estimate; Betauntransformed - effect estimate from association analysis in which untransformed cIMT was used; SE - standard error of Beta; P - p-value; ERF - Erasmus Rucphen Family study;

*The variants are ordered based on their genomic position (hg19 assembly). First four variants are shared by 6 of 8 affected family members in a family contributing predominately to the linkage peak, and last two variants are shared by 5 of 8 affected family members;

**ERF controls were defined as individuals who scored below the mean of the distribution of the residuals from the regression of cIMT onto age, age^2, sex, smoking status, BMI, WHR, diabetes, dyslipidemia, and hypertension;

***If MAF was unknown in 1kG, MAF reported from Exome Sequencing Project if available;

Table 4. Functional annotation of rs62165235 variant and variants that are in LD ($r^2 > 0.6$) using HaploReg 4.1[69]

Variant	LD (r^2)	Ref	Alt	GERP cons*	Promoter histone marks	Enhancer histone marks	DNAse	Proteins bound	Motifs changed	Selected eQTL hits	RefSeq genes	function
rs7591128	0.65	T	G	No	-	BLD	-	-	6 altered motifs	1 hit	39kb 5′ of *CCDC88A*	intergenic
rs78928997	0.75	T	A	Yes	-	-	-	-	7 altered motifs	-	*SMEK2*	3′-UTR
rs62165193	0.77	C	T	No	-	2 tissues	-	-	GR	-	*SMEK2*	intronic
rs62165227	0.61	G	A	No	-	-	-	-	-	-	*PNPT1*	intronic
rs62165231	0.95	C	T	No	-	-	HRT	-	Rad21,Tgif1	1 hit	*PNPT1*	intronic
rs62165235	1	T	C	No	24 tissues	-	15 tissues	E2F6	7 altered motifs	-	*PNPT1*	intronic
rs62165236	0.96	T	C	No	-	-	-	-	ZID	-	2.1kb 5′ of *PNPT1*	intergenic
rs79873145	0.76	T	C	No	-	GI	-	-	4 altered motifs	1 hit	30kb 5′ of *PNPT1*	intergenic

Abbreviations: Ref - reference allele; Alt - Alternative allele; GERP - conservation score;*GERP conservation score indicates whether the element is conserved or not according to the algorithm; The variant highlighted in red was identified in the sharing and association analysis in the ERF;

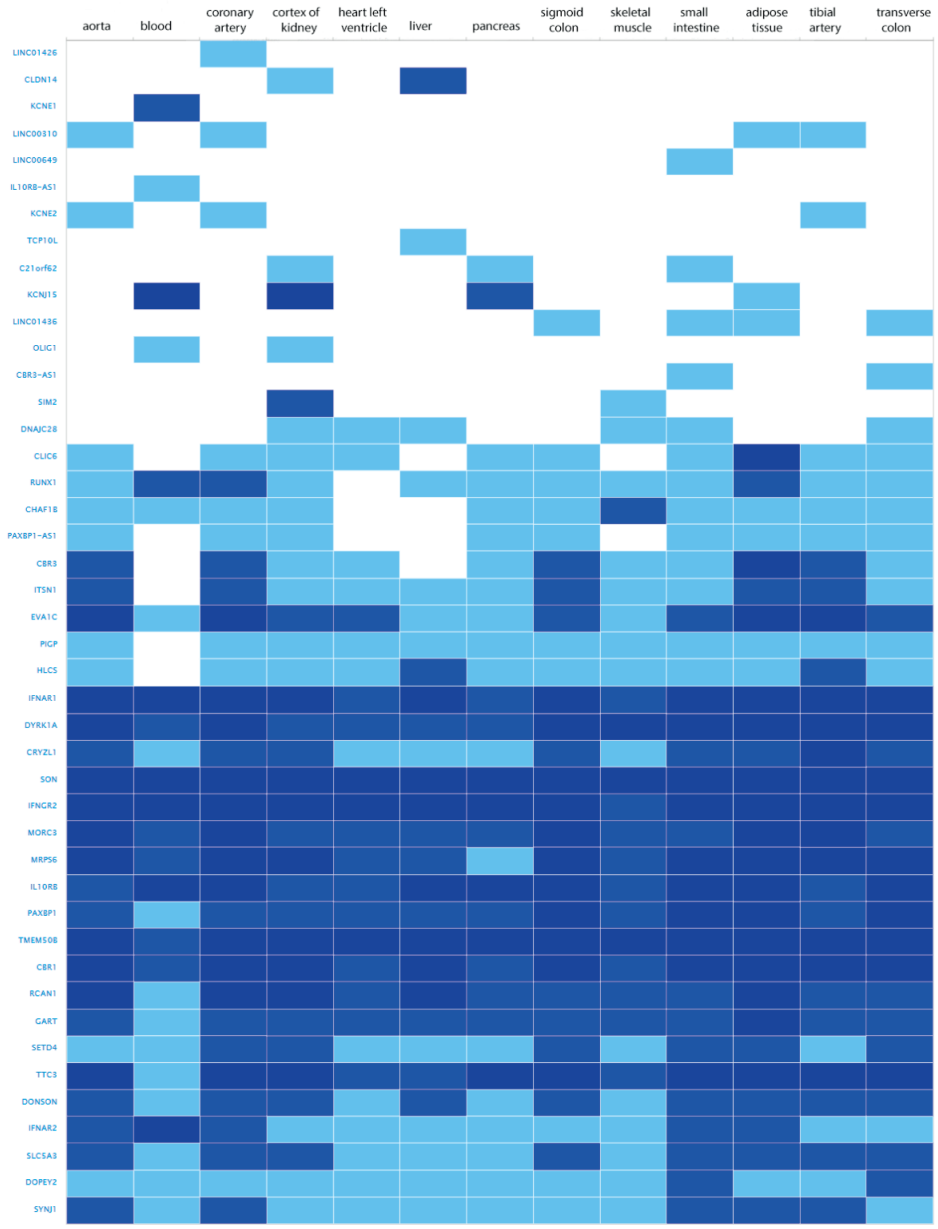


Figure 2. Genes under the base to base linkage peak at 21q22 and their expression levels in tissues relevant for atherosclerosis according to GTEx database. Gene names are shown on y-axis and tissues on x-axis. Colour depicts expression level estimated as fragments per kilobase of exon model per million reads mapped (FPKM). The grey box stands for expression level below cutoff (0.5 FPKM), light blue box stands for low expression level (between 0.5 to 10), medium blue for medium expression level (between 11 to 1000 FPKM), and dark blue for high expression level (more than 1000 FPKM or more than 1000 TPM) and white for no data available.

ants (**Table 3**). The most interesting finding is a variant (rs62165235) with MAF 0.038 in 1kG mapping to *PNPT1*. The variant, shared by 6 out of 8 affected relatives, is likely having a regulatory function and affecting transcription factor binding and matched DNase Footprinting and DNase sensitivity (Category 2b Regulome DB score; **Table 3-4**). The variant was sequenced at a read depth of 37x and it showed association with quantitative cIMT in the ERF (effect = 0.27, *p*-value = 0.013, **Table 3**) after applying Bonferroni correction (*p*-value = 0.05/3 independent tests = 0.017). The effect estimate of the minor allele C on untransformed cIMT suggested a mean increase of 0.04 mm for each minor allele (0.04 mm for heterozygote C/T carriers and 0.08 for homozygous C/C carriers) (**Table 3**). This variant explained 0.3% of variation in the ERF.

The search for shared variants within the linkage region 19q13.43, 20p13 and 21q22.12 identified several variants to be shared among the affected family members, however, none of the variants showed significant association with quantitative cIMT (**Supplementary Table 1-3**). There are, however, several potentially interesting candidate genes for atherosclerosis in each of these linked regions, for instance, among the genes under the base to base 19q13 and 20p13 linkage peaks, several genes have been implicated in the pathogenesis of cardiovascular disease, including *FCAR, TNNT1, OSCAR, FPR2* under the 19q13 and *ADAM33, TRIB3, HSPA12B* under the 20p13 peak. The 21q22 region harbors several genes that are highly expressed in tissues relevant for atherosclerosis, including *IFNAR1, DYRK1A, SON, IFNGR2, MORC3, MRPS6, IL10RB, TMEM50B, CBR1, RCAN1,* and *TTC3* (**Figure 2**). According to the Ingenuity Pathway Analysis (IPA) tool (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis), which exposes possible functional relationship between the genes by expanding upstream analysis to include regulators that are not directly connected to targets in the dataset, these genes connected to a network illustrated in **Supplementary Figure 5**.

There were several regions that showed suggestive evidence of linkage, including 1q31.1, 5q35.3, 7p21.3, 8p22, 12q24.33, 14q22.2, 15q21.3 and 17q25.3. As the 12q24 region has previously been linked to cIMT, we have explored it further (**Figure 1**). Search for shared variants within 12q24 identified no variants that can explain linkage signal.

The results of linkage analysis when using cIMT as a quantitative outcome are shown in **Supplementary Figure 6**.

DISCUSSION

In this study, we have identified genomic regions at 2p16.3, 19q13.43, 20p13 and 21q22.12 with significant evidence of linkage to cIMT. These regions have not been reported before. Identification of variants under the linkage peaks using exome sequencing revealed a variant with likely regulatory function mapping to *PNPT1* gene at chromosome 2 and several candidate genes at 21q22.

As the present study targets genes with relatively large effects, we studied the extremes of cIMT distribution in the discovery population. Even though extreme trait approach neglects much of the overall distribution of the trait and some rare variants with the moderate effects may be missed, it has been shown that the power to detect rare variants can be increased due to an excess of rare variants in the upper tails of the distribution.[15,34,35] Comparison of the results obtained in the linkage analysis of the extremes of cIMT distribution and those using cIMT as a quantitative trait revealed no overlap, highlighting the power of the approach we have followed.

When comparing the results of several genome-wide linkage studies of cIMT that have been conducted so far, we noticed that a region with suggestive evidence of linkage in our study, 12q24, has previously been linked to cIMT through the linkage scan.[10] Similarly to the previous study which identified 12q24, we did not identify variants by sharing analysis that could explain the linkage signal. Even though linkage findings from prior studies already showed limited generalizability across the reported linkage peaks due to selected nature of the cohorts, the overlap of our finding with the literature suggests that our study population is representative of the general population. However, the linkage single at chromosome 12q24 is relatively weak in our population.

Among the regions with significant evidence of linkage to cIMT in our study, we identified 2p16.3 which gave significant linkage signal under the dominant model and suggestive signal under the recessive model. This region has previously been associated with polycystic ovary syndrome (PCOS) and primary open angle glaucoma (POAG).[36,37]Interestingly, women with PCOS are at a greater risk of premature atherosclerosis,[38] whereas atherosclerosis associated with vascular conditions that are correlated with POAG.[39] The region has further been implicated in body mass index[40] and glycated hemoglobin levels.[41] Both obesity and poor glycemic control are risk factors for variety of diseases including atherosclerosis and cardiovascular diseases.

Identification of variants under the base to base peak at 2p16 using exome sequence data revealed a variant that lies in DNase sites, promote histone marks and protein

binding regions and changes regulatory motifs based on the variant allele change. The variant is mapped to intron 1 of Polyribonucleotide Nucleotidyltransferase 1 (*PNPT1*) gene which encodes a protein predominantly localized in the mitochondrial intermembrane space and is involved in import of RNA to mitochondria (http://www.genecards.org/cgi-bin/carddisp.pl?gene=PNPT1). *PNPT1* has been characterized as a type I interferon-inducible early response gene.[42,43] Type I interferons promote atherosclerosis by enhancing macrophage-endothelial cell adhesion and promoting leukocyte attraction to atherosclerosis-prone sites in animal models.[44] Even though our finding supports a role of *PNPT1* as a candidate gene in atherosclerosis, we acknowledge that *PNPT1* variant is unlikely to be causal and cannot explain the linkage signal at 2p16.3 to cIMT. Furthermore, we attempted to replicate the association of this variant with cIMT in the Rotterdam Study, a population-based cohort study (detailed information provided in **Supplementary Methods** and **Supplementary Table 4**). However, the variant was not available in the exome sequencing data of the Rotterdam Study.

The other interesting region includes 21q22, which is also known as a Down critical region. Interestingly, persons with Down syndrome are protected against atherosclerosis, in spite of increases in metabolic disturbances and obesity in Down syndrome.[45] Even though identification of variants using exome sequencing did not identify a causal variant, this region contains several plausible candidate genes which are highly expressed in relevant tissues[46], including *IFNAR1, DYRK1A, SON, IFNGR2, MORC3, MRPS6, IL10RB, TMEM50B, CBR1, RCAN1,* and *TTC3*. *DYRK1A* signaling pathway is linked to homocysteine cycle which is associated with an increased risk of atherosclerosis.[47,48] *IFNGR2* and *RCAN1* also play a role in atherosclerosis.[49,50] Notably, IPA analysis revealed that those genes are connected in one network, and directly or indirectly linked to *TP53*. *TP53* encodes a tumor suppressor gene p53 involved in regulation of cell proliferation and apoptosis. Numerous studies implicated p53 in development of atherosclerosis and vascular smooth muscle cell apoptosis.[51-55] Higher plasma p53 levels were also associated with an increased cIMT.[56] However, it is important to note that network analysis is based on the knowledge databases that are always evolving and new discoveries happen all time.

Furthermore, we identified 19q13.43 and 20p13 regions with significant evidence of linkage to cIMT. Several genes under the linkage peak have previously been implicated in the pathogenesis of cardiovascular disease. The base to base peak at 19q13 encompassed *FCAR* and *TNNT1* genes associated with coronary heart disease[57,58] and *OSCAR* and *FPR2* genes associated with atherosclerosis plaque phenotype,[59,60] whereas the base to base peak at 20p13 encompassed *ADAM33* and *TRIB3* associated with extent and promotion of atherosclerosis[61-65] and *HSPA12B* which is found to be enriched in atherosclerotic lesions.[66]

Our study presents the linkage analysis using extreme phenotype approach that was designed to capture region with genetic variants that have large effects on cIMT. Combination of linkage analysis in a large family-based study and exome sequence data provide a unique opportunity to explore the variants in the linkage regions. However, despite these distinct advantages, we were able to identify a genetic variant for only one of the several linked genomic regions, for which, there may be several reasons including structural variants, and intronic or intergenic single-nucleotide variants that were not evaluated in the current study. Interestingly, the 19q13.43, 20p13 and 21q22.12 linkage peaks were previously associated with various phenotypes in our study population including personality traits and depressive symptoms.[67,68]

To conclude, our linkage analysis identified four genomic regions at 2p16.3, 19q13.43, 20p13 and 21q22.12 for cIMT. The significant linkage regions contain several plausible candidate genes. Further analyses are needed to demonstrate their involvement in atherosclerosis.

REFERENCES

1. Xu, J., Murphy, S.L., Kochanek, K.D. & Bastian, B.A. Deaths: Final Data for 2013. *Natl Vital Stat Rep* **64**, 1-119 (2016).
2. Falk, E. Pathogenesis of atherosclerosis. *J Am Coll Cardiol* **47**, C7-12 (2006).
3. Lorenz, M.W., Markus, H.S., Bots, M.L., Rosvall, M. & Sitzer, M. Prediction of clinical cardiovascular events with carotid intima-media thickness: a systematic review and meta-analysis. *Circulation* **115**, 459-67 (2007).
4. Polak, J.F. *et al.* Carotid-wall intima-media thickness and cardiovascular events. *N Engl J Med* **365**, 213-21 (2011).
5. Den Ruijter, H.M. *et al.* Common carotid intima-media thickness measurements in cardiovascular risk prediction: a meta-analysis. *JAMA* **308**, 796-803 (2012).
6. Lusis, A.J. Genetics of atherosclerosis. *Trends Genet* **28**, 267-75 (2012).
7. Fox, C.S. *et al.* Genetic and environmental contributions to atherosclerosis phenotypes in men and women: heritability of carotid intima-media thickness in the Framingham Heart Study. *Stroke* **34**, 397-401 (2003).
8. Sacco, R.L. *et al.* Heritability and linkage analysis for carotid intima-media thickness: the family study of stroke risk and carotid atherosclerosis. *Stroke* **40**, 2307-12 (2009).
9. Wang, D. *et al.* A genome-wide scan for carotid artery intima-media thickness: the Mexican-American Coronary Artery Disease family study. *Stroke* **36**, 540-5 (2005).
10. Fox, C.S. *et al.* Genomewide linkage analysis for internal carotid artery intimal medial thickness: evidence for linkage to chromosome 12. *Am J Hum Genet* **74**, 253-61 (2004).
11. Kuipers, A.L. *et al.* Genetic epidemiology and genome-wide linkage analysis of carotid artery ultrasound traits in multigenerational African ancestry families. *Atherosclerosis* **231**, 120-3 (2013).
12. Bis, J.C. *et al.* Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque. *Nat Genet* **43**, 940-7 (2011).
13. Natarajan, P. *et al.* Multiethnic Exome-Wide Association Study of Subclinical Atherosclerosis. *Circ Cardiovasc Genet* **9**, 511-520 (2016).
14. Bis, J.C. *et al.* Sequencing of 2 subclinical atherosclerosis candidate regions in 3669 individuals: Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Targeted Sequencing Study. *Circ Cardiovasc Genet* **7**, 359-64 (2014).
15. Auer, P.L. & Lettre, G. Rare variant association studies: considerations, challenges and opportunities. *Genome Med* **7**, 16 (2015).
16. Stacey, S.N. *et al.* A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet* **43**, 1098-103 (2011).
17. Gudmundsson, J. *et al.* A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nature Genetics* **44**, 1326-1329 (2012).
18. Amin, N. *et al.* A rare missense variant in RCL1 segregates with depression in extended families. *Molecular Psychiatry* **23,** 1120-1126 (2018).
19. Pardo, L.M., MacKay, I., Oostra, B., van Duijn, C.M. & Aulchenko, Y.S. The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet* **69**, 288-95 (2005).
20. Aulchenko, Y.S. *et al.* Linkage disequilibrium in young genetically isolated Dutch population. *Eur J Hum Genet* **12**, 527-34 (2004).
21. Sayed-Tabatabaei, F.A. *et al.* Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study. *Stroke* **36**, 2351-6 (2005).

22. Miller, S.A., Dykes, D.D. & Polesky, H.F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**, 1215 (1988).
23. Amin, N. *et al.* Exome-sequencing in a large population-based study reveals a rare Asn396Ser variant in the LIPG gene associated with depressive symptoms. *Mol Psychiatry* (2016).
24. Amin, N. *et al.* Nonsynonymous Variation in NKPD1 Increases Depressive Symptoms in European Populations. *Biol Psychiatry* **81**, 702-707 (2017).
25. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research* **22**, 1790-1797 (2012).
26. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* **46**, 310 (2014).
27. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* **30**, 97-101 (2002).
28. Liu, F., Kirichenko, A., Axenovich, T.I., van Duijn, C.M. & Aulchenko, Y.S. An approach for cutting large and complex pedigrees for linkage analysis. *Eur J Hum Genet* **16**, 854-60 (2008).
29. Baron, R.V., Kollar, C., Mukhopadhyay, N. & Weeks, D.E. Mega2: validated data-reformatting for linkage and association analyses. *Source Code Biol Med* **9**, 26 (2014).
30. Ott, J., Wang, J. & Leal, S.M. Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet* **16**, 275-84 (2015).
31. Zhan, X.W., Hu, Y.N., Li, B.S., Abecasis, G.R. & Liu, D.J.J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423-1426 (2016).
32. Li, J. & Ji, L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity* **95,** 221-227 (2005).
33. Ramsey, S.A., Gold, E.S. & Aderem, A. A systems biology approach to understanding atherosclerosis. *Embo Molecular Medicine* **2,** 79-89 (2010).
34. Lamina, C. Digging into the extremes: a useful approach for the analysis of rare variants with continuous traits? *BMC Proc* **5** Suppl 9, S105 (2011).
35. Coassin, S. *et al.* Investigation and functional characterization of rare genetic variants in the adipose triglyceride lipase in a large healthy working population. *PLoS Genet* **6**, e1001239 (2010).
36. Mutharasan, P. *et al.* Evidence for chromosome 2p16.3 polycystic ovary syndrome susceptibility locus in affected women of European ancestry. J *Clin Endocrinol Metab* **98**, E185-90 (2013).
37. Liu, Y.T., Qin, X.J., Schmidt, S., Allingham, R.R. & Hauser, M.A. Association between chromosome 2p16.3 variants and glaucoma in populations of African descent. *Proceedings of the National Academy of Sciences of the United States of America* **107**, E61-E61 (2010).
38. Meyer, M.L., Malek, A.M., Wild, R.A., Korytkowski, M.T. & Talbott, E.O. Carotid artery intima-media thickness in polycystic ovary syndrome: a systematic review and meta-analysis. *Human Reproduction Update* **18**, 112-126 (2012).
39. Belzunce, A. & Casellas, M. [Vascular risk factors in primary open angle glaucoma] Factortes de riesgo vascular en el glaucoma primario de angulo abierto. *An Sist Sanit Navar* **27**, 335-44 (2004).
40. Akiyama, M. *et al.* Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat Genet* **49**, 1458-1467 (2017).
41. Wheeler, E. *et al.* Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS Med* **14**, e1002383 (2017).

42. Leszczyniecka, M. *et al.* Identification and cloning of human polynucleotide phosphorylase, hPNPase old-35, in the context of terminal differentiation and cellular senescence. *Proc Natl Acad Sci U S A* **99**, 16636-41 (2002).
43. Leszczyniecka, M., Su, Z.Z., Kang, D.C., Sarkar, D. & Fisher, P.B. Expression regulation and genomic organization of human polynucleotide phosphorylase, hPNPase(old-35), a Type I interferon inducible early response gene. *Gene* **316**, 143-56 (2003).
44. Goossens, P. *et al.* Myeloid type I interferon signaling promotes atherosclerosis by stimulating macrophage recruitment to lesions. *Cell Metab* **12**, 142-53 (2010).
45. Colvin, K.L. & Yeager, M.E. What people with Down Syndrome can teach us about cardiopulmonary disease. *European Respiratory Review* **26** (2017).
46. Ramsey, S.A., Gold, E.S. & Aderem, A. A systems biology approach to understanding atherosclerosis. *EMBO Mol Med* **2**, 79-89 (2010).
47. Noll, C. *et al.* DYRK1A, a novel determinant of the methionine-homocysteine cycle in different mouse models overexpressing this Down-syndrome-associated kinase. *PLoS ONE* **4**, e7540 (2009).
48. Tlili, A. *et al.* Hepatocyte-specific Dyrk1a gene transfer rescues plasma apolipoprotein A-I levels and aortic Akt/GSK3 pathways in hyperhomocysteinemic mice. *Biochim Biophys Acta* **1832**, 718-28 (2013).
49. Mendez-Barbero, N. *et al.* A major role for RCAN1 in atherosclerosis progression. *EMBO Mol Med* **5**, 1901-17 (2013).
50. Voloshyna, I., Littlefield, M.J. & Reiss, A.B. Atherosclerosis and interferon-gamma: new insights and therapeutic targets. *Trends Cardiovasc Med* **24**, 45-51 (2014).
51. Boesten, L.S. *et al.* Macrophage p53 controls macrophage death in atherosclerotic lesions of apolipoprotein E deficient mice. *Atherosclerosis* **207**, 399-404 (2009).
52. Tabas, I. p53 and atherosclerosis. *Circ Res* **88**, 747-9 (2001).
53. Mayr, M., Hu, Y., Hainaut, H. & Xu, Q. Mechanical stress-induced DNA damage and rac-p38MAPK signal pathways mediate p53-dependent apoptosis in vascular smooth muscle cells. *FASEB J* **16**, 1423-5 (2002).
54. Mercer, J., Figg, N., Stoneman, V., Braganza, D. & Bennett, M.R. Endogenous p53 protects vascular smooth muscle cells from apoptosis and reduces atherosclerosis in ApoE knockout mice. *Circ Res* **96**, 667-74 (2005).
55. Varela, A. *et al.* Elevated expression of mechanosensory polycystins in human carotid atherosclerotic plaques: association with p53 activation and disease severity. *Sci Rep* **5**, 13461 (2015).
56. Chen, W. *et al.* p53 Levels positively correlate with carotid intima-media thickness in patients with subclinical atherosclerosis. *Clin Cardiol* **32**, 705-10 (2009).
57. Iakoubova, O.A. *et al.* Asp92Asn polymorphism in the myeloid IgA Fc receptor is associated with myocardial infarction in two disparate populations - CARE and WOSCOPS. *Arteriosclerosis Thrombosis and Vascular Biology* **26**, 2763-2768 (2006).
58. Guay, S.P. *et al.* Epigenetic and genetic variations at the TNNT1 gene locus are associated with HDL-C levels and coronary artery disease. *Epigenomics* **8**, 359-371 (2016).
59. Goettsch, C. *et al.* The Osteoclast-Associated Receptor (OSCAR) Is a Novel Receptor Regulated by Oxidized Low-Density Lipoprotein in Human Endothelial Cells. *Endocrinology* **152**, 4915-4926 (2011).
60. Petri, M.H. *et al.* The role of the FPR2/ALX receptor in atherosclerosis development and plaque stability. *Cardiovascular Research* **105**, 65-74 (2015).
61. Figarska, S.M., Vonk, J.M., van Diemen, C.C., Postma, D.S. & Boezen, H.M. ADAM33 gene polymorphisms and mortality. A prospective cohort study. *PLoS One* **8**, e67768 (2013).

62. Holloway, J.W. *et al.* ADAM33 expression in atherosclerotic lesions and relationship of ADAM33 gene variation with atherosclerosis. *Atherosclerosis* 211, 224-30 (2010).
63. Formoso, G. *et al.* The TRIB3 R84 variant is associated with increased carotid intima-media thickness in vivo and with enhanced MAPK signalling in human endothelial cells. *Cardiovascular Research* **89**, 184-192 (2011).
64. Wang, Z.H. *et al.* Silence of TRIB3 Suppresses Atherosclerosis and Stabilizes Plaques in Diabetic ApoE(-/-)/LDL Receptor(-/-) Mice. *Diabetes* **61**, 463-473 (2012).
65. Prudente, S. *et al.* Infrequent TRIB3 coding variants and coronary artery disease in type 2 diabetes. *Atherosclerosis* **242**, 334-339 (2015).
66. Han, Z.H., Truong, Q.A., Park, S. & Breslow, J.L. Two Hsp70 family members expressed in atherosclerotic lesions. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 1256-1261 (2003).
67. Amin, N. *et al.* A rare missense variant in RCL1 segregates with depression in extended families. *Mol Psychiatry (*2017).
68. Amin, N. *et al.* A genome-wide linkage study of individuals with high scores on NEO personality traits. *Mol Psychiatry* **17**, 1031-41 (2012).
69. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Research* **40,** D930-D934 (2012).

SUPPLEMENTARY METHODS AND TABLES

Supplementary Methods

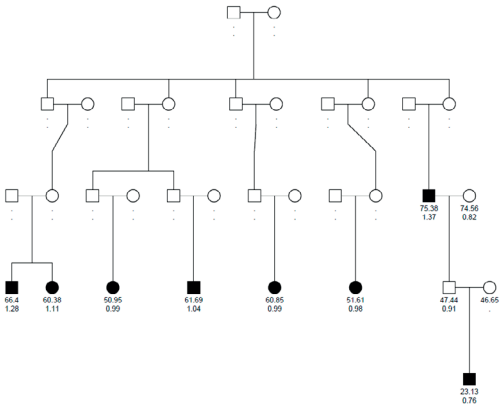
Supplementary Table 1. Variants shared by majority of affected family members in the family contributing the most to the linkage peak at chromosome 19 and their association with cIMT in the ERF.

Supplementary Table 2. Variants shared by majority of affected family members in the family contributing the most to the linkage peak at chromosome 20 and their association with cIMT in the ERF.

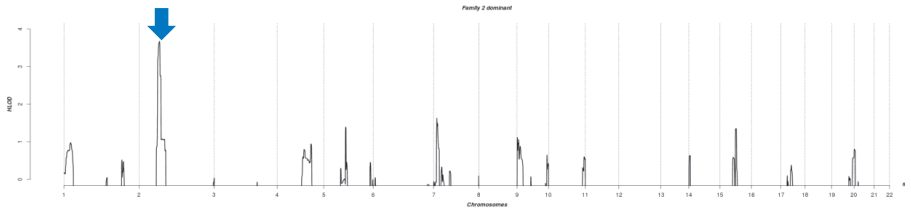
Supplementary Table 3. Variants shared by majority of affected family members in the family contributing the most to the linkage peak at chromosome 21 and their association with cIMT in the ERF.

Supplementary Table 4. Descriptive statistics of the Rotterdam Study.

(A)

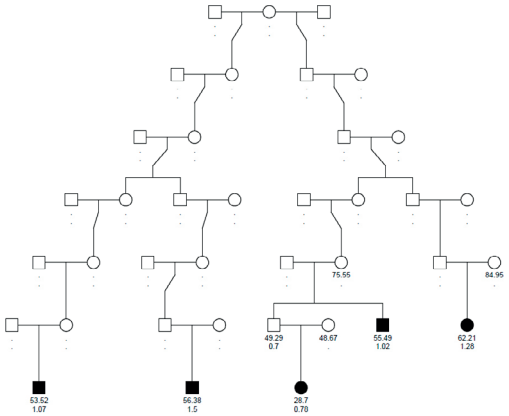


(B)

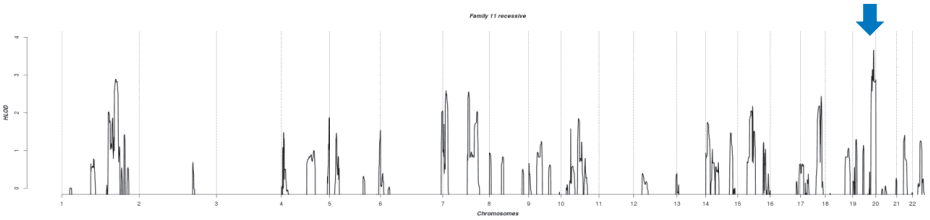


Supplementary Figure 1. The pedigree of highest HLOD score contributing family and the results of parametric per-family linkage analysis under the dominant model for the chromosome 2p16.3. (A) Squares represent males and circles females. Solid symbols depict affected family members. These family members were used in the linkage analysis. Age of individual and carotid intima-media thickness are displayed on the pedigree. Open symbols denote unaffected individuals (carotid intima-media thickness is displayed underneath the symbol) or individuals with no data available for analysis (carotid intima-media thickness is missing). (B) The x-axis shows 22 autosomal chromosomes, and the y-axis shows the heterogeneity LOD (HLOD) scores for the dominant model.

(A)

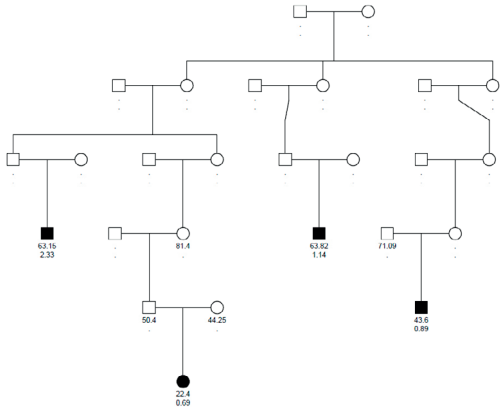


(B)

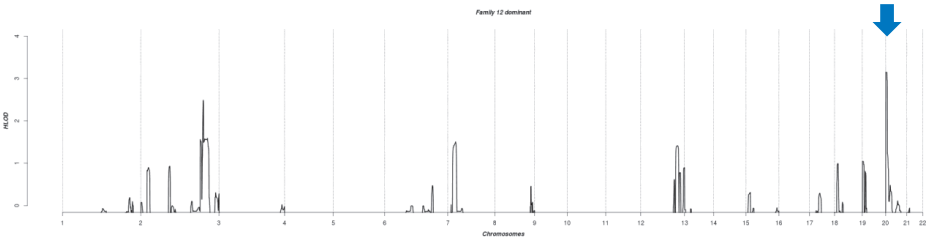


Supplementary Figure 2. The pedigree of highest HLOD score contributing family and the results of parametric per-family linkage analysis under the recessive model for the chromosome 19q13.43. (A) Squares represent males and circles females. Solid symbols depict affected family members. These family members were used in the linkage analysis. Age of individual and carotid intima-media thickness are displayed on the pedigree. Open symbols denote unaffected individuals (carotid intima-media thickness is displayed underneath the symbol) or individuals with no data available for analysis (carotid intima-media thickness is missing). (B) The x-axis shows 22 autosomal chromosomes, and the y-axis shows the heterogeneity LOD (HLOD) scores for recessive model.

(A)

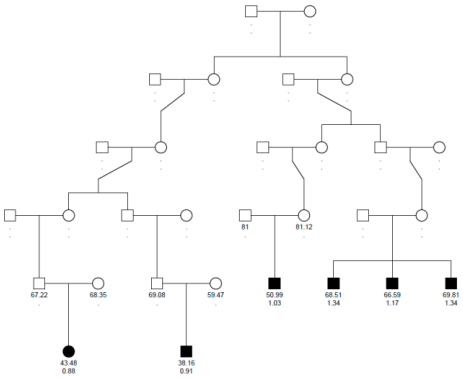


(B)

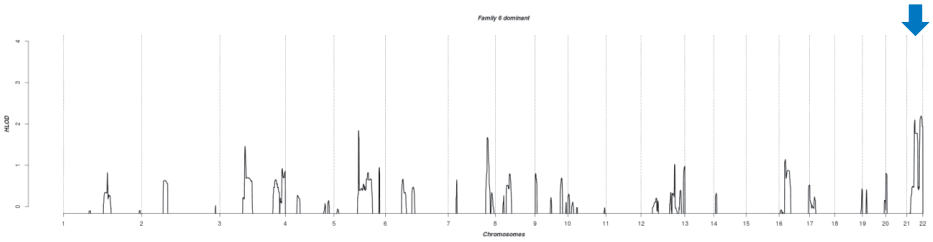


Supplementary Figure 3. The pedigree of highest HLOD score contributing family and the results of parametric per-family linkage analysis under the dominant model for the chromosome 20p13. (A) Squares represent males and circles females. Solid symbols depict affected family members. These family members were used in the linkage analysis. Age of individual and carotid intima-media thickness are displayed on the pedigree. Open symbols denote unaffected individuals (carotid intima-media thickness is displayed underneath the symbol) or individuals with no data available for analysis (carotid intima-media thickness is missing). (B) The x-axis shows 22 autosomal chromosomes, and the y-axis shows the heterogeneity LOD (HLOD) scores for the dominant model.

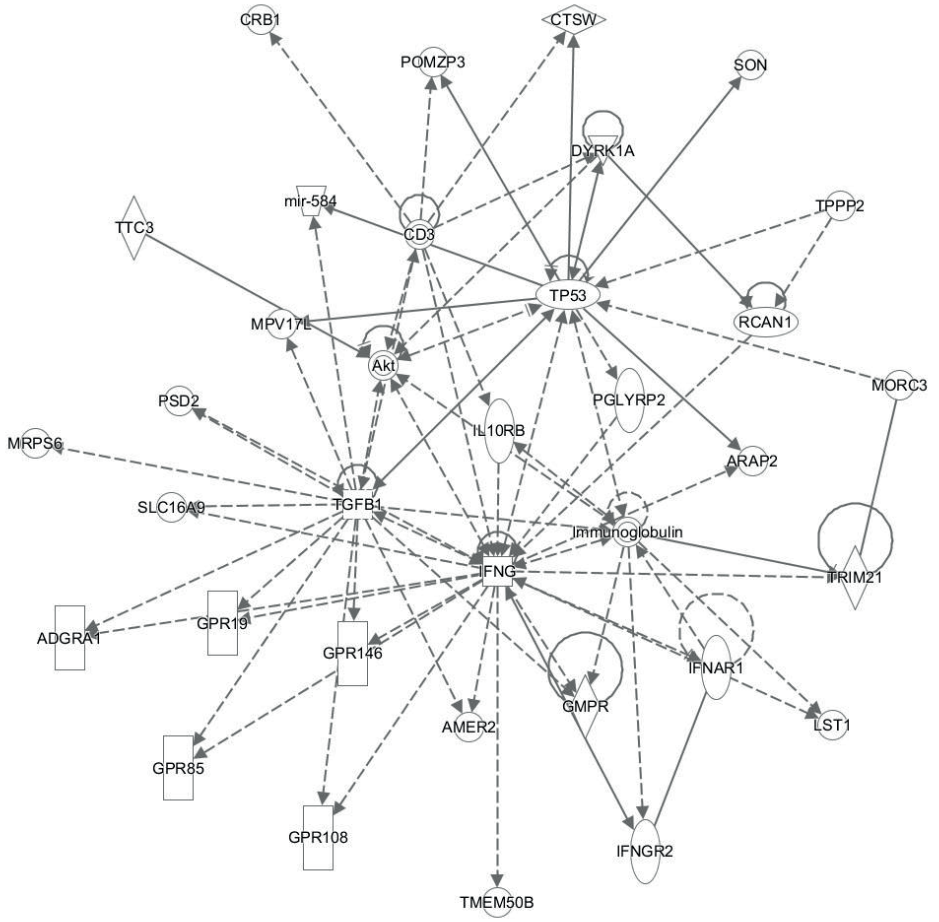
(A)



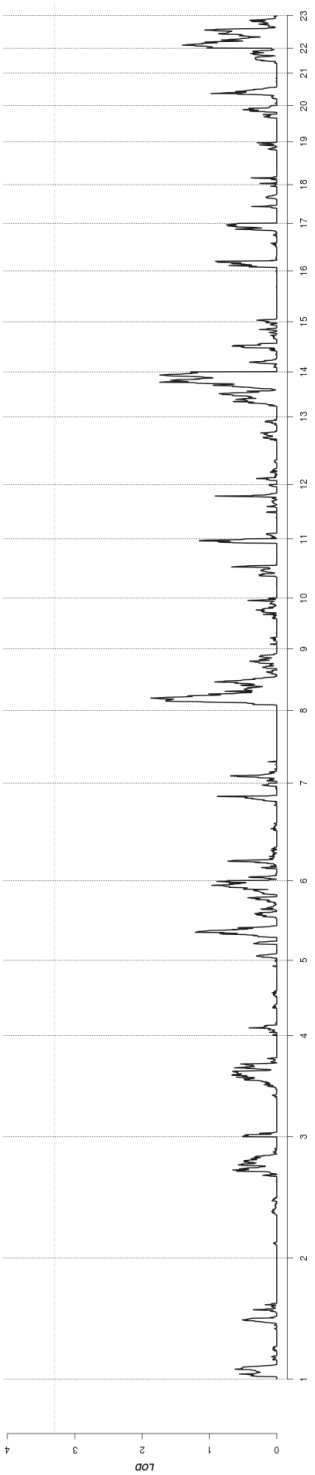
(B)



Supplementary Figure 4. The pedigree of highest HLOD score contributing family and the results of parametric per-family linkage analysis under the dominant model for the chromosome 21q22.12. (A) Squares represent males and circles females. Solid symbols depict affected family members. These family members were used in the linkage analysis. Age of individual and carotid intima-media thickness are displayed on the pedigree. Open symbols denote unaffected individuals (carotid intima-media thickness is displayed underneath the symbol) or individuals with no data available for analysis (carotid intima-media thickness is missing). (B) The x-axis shows 22 autosomal chromosomes, and the y-axis shows the heterogeneity LOD (HLOD) scores for the dominant model.



Supplementary Figure 5. Ingenuity Pathway Analysis tool network of genes under the linkage peak at 21q22 which are highly expressed in tissues relevant for atherosclerosis. Lines without arrow indicate interactions (chemical-chemical, protein-protein, chemical-protein, RNA-RNA, correlation) while lines with an arrow indicate activation, causation, expression, localization, transcription, molecular cleavage, membership, modification, phosphorylation, protein-DNA and/or protein-RNA interactions. Solid lines indicate direct interaction while dashed lines indicate indirect interactions. Diamond molecule shape denotes enzyme, rhombus - peptidase, inverted triangle - kinase, inverted trapezium - microRNA, circle in a circle - complex/group, vertical rectangle - G-protein coupled receptor, horizontal rectangle - ligand-dependent nuclear receptor, vertical ellipse - transmembrane receptor, horizontal ellipse - transcription regulator, square - cytokine, and circle - other.



Supplementary Figure 6. The results of quantitative trait linkage analysis using variance component linkage.