

Genome-Wide Linkage Analysis of Cardiovascular Disease Biomarkers in a Large, Multigenerational Family

Daniel Nolan¹✉, William E. Kraus²✉, Elizabeth Hauser¹, Yi-Ju Li¹, Dana K. Thompson², Jessica Johnson¹, Hsiang-Cheng Chen³, Sarah Nelson⁴, Carol Haynes¹, Simon G. Gregory¹, Virginia B. Kraus², Svati H. Shah^{1,2*}

1 Center for Human Genetics, Duke University, Durham, North Carolina, United States of America, **2** Department of Medicine, Duke University, Durham, North Carolina, United States of America, **3** Division of Rheumatology, Immunology and Allergy, Tri-Service General Hospital, Taipei, Taiwan, **4** Department of Biostatistics, University of Washington, Seattle, Washington, United States of America

Abstract

Given the importance of cardiovascular disease (CVD) to public health and the demonstrated heritability of both disease status and its related risk factors, identifying the genetic variation underlying these susceptibilities is a critical step in understanding the pathogenesis of CVD and informing prevention and treatment strategies. Although one can look for genetic variation underlying susceptibility to CVD *per se*, it can be difficult to define the disease phenotype for such a qualitative analysis and CVD itself represents a convergence of diverse etiologic pathways. Alternatively, one can study the genetics of intermediate traits that are known risk factors for CVD, which can be measured quantitatively. Using the latter strategy, we have measured 21 cardiovascular-related biomarkers in an extended multigenerational pedigree, the CARRIAGE family (Carolinas Region Interaction of Aging, Genes, and Environment). These biomarkers belong to inflammatory and immune, connective tissue, lipid, and hemostasis pathways. Of these, 18 met our quality control standards. Using the pedigree and biomarker data, we have estimated the broad sense heritability (H^2) of each biomarker (ranging from 0.09–0.56). A genome-wide panel of 6,015 SNPs was used subsequently to map these biomarkers as quantitative traits. Four showed noteworthy evidence for linkage in multipoint analysis (LOD score ≥ 2.6): paraoxonase (chromosome 8p11, 21), the chemokine RANTES (22q13.33), matrix metalloproteinase 3 (MMP3, 17p13.3), and granulocyte colony stimulating factor (GCSF, 8q22.1). Identifying the causal variation underlying each linkage score will help to unravel the genetic architecture of these quantitative traits and, by extension, the genetic architecture of cardiovascular risk.

Citation: Nolan D, Kraus WE, Hauser E, Li Y-J, Thompson DK, et al. (2013) Genome-Wide Linkage Analysis of Cardiovascular Disease Biomarkers in a Large, Multigenerational Family. PLoS ONE 8(8): e71779. doi:10.1371/journal.pone.0071779

Editor: Xiping Cui, University of California, Riverside, United States of America

Received: October 3, 2012; **Accepted:** July 3, 2013; **Published:** August 2, 2013

Copyright: © 2013 Nolan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funding provided by National Institutes of Health (NIH) National Institutes on Aging grant P30AG028716-01 (Cohen), R01HL095987-01 (SHS), Hartford Center award (SHS), Mary Duke Biddle Foundation, and the Trent Foundation, NIH NHLBI P01-HL036587 (Williams). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: svati.shah@duke.edu

✉ These authors contributed equally to this work.

Introduction

Cardiovascular disease (CVD) is the leading cause of death, accounting for over 500,000 deaths per year in the United States and greater than seven million deaths worldwide. It has been estimated that over 82 million Americans are afflicted with at least one form of CVD and over 16 million Americans are afflicted with clinically significant CVD [1]. There are many well accepted CVD risk factors including common metabolic conditions (hypertension, dyslipidemia, metabolic syndrome, diabetes), behavioral factors (smoking, sedentary lifestyle), and non-modifiable factors (sex, age). Thus CVD is etiologically complex with independent genetic influences on CVD

susceptibility as well as genetic influences on CVD-related risk factors.

Given the importance of CVD to public health and the demonstrated heritability of both disease status and its related clinical risk factors, identifying the genetic variation underlying these susceptibilities is a critical step in understanding the pathogenesis of CVD to inform prevention and treatment strategies. Although many studies have examined genetic variation underlying susceptibility to CVD, it can be difficult to define the disease phenotype for such a qualitative analysis because CVD likely represents an end convergence of diverse biological etiologic pathways, reflects genetic factors operating through these different pathways which may be different than

factors influencing CVD directly, and results from an incompletely understood interaction of genetic and environmental factors. Alternatively, one can study the genetics of CVD through a quantitation approach that uses intermediate traits that are known risk factors for CVD and that can be measured quantitatively; we have used this method successfully for identification of osteoarthritis genes [2].

As quantitative traits we measured 21 biomarkers with relevance to CVD. We present herein the subsequent quantitative trait loci (QTL) mapping of these disease-related intermediate traits in an extended multigenerational pedigree as part of the CARRIAGE (Carolinas Region Interaction of Aging, Genes, and Environment) study. The CARRIAGE family is one of the most extensively pedigreed families in the U.S., comprising 10 generations and 3,327 individuals descended from a single founding couple born in the 18th century, and composed of primarily African and Native American ethnic origin. The potential for reduced genetic heterogeneity within extended pedigrees, such as this one, facilitates linkage analysis [3]. This family has a reduced prevalence of CVD relative to the general population with 15% of studied family members having at least one cardiovascular condition compared to approximately 33% in the US population [1]. Thus the genetic effects identified for the CVD-risk biomarkers reflect effects in a sample not ascertained for CVD conditions.

Materials and Methods

Study population: the CARRIAGE family

The CARRIAGE family data collection has been described previously [2]. For this study, we have used a detailed ascertainment of 350 family members from whom blood samples were available, as previously reported [2]. Written informed consent was obtained from each participant, and the study was approved by the Duke Institutional Review Board. All information and work was conducted under a Federal Certificate of Confidentiality to ensure the privacy of each participating member's clinical and genetic data.

Selection and measurement of serum biomarkers

The biomarkers selected belong to important known pathways involved in CVD risk: (1) inflammatory and immune (C reactive protein [hsCRP], monocyte chemoattractant protein one [MCP1], Regulated upon Activation, Normal T-cell Expressed, and Secreted [RANTES], granulocyte colony stimulating factor [GCSF], interleukin eight [IL-8], TNF-related apoptosis-inducing ligand [TRAIL], interleukin-six [IL-6], interleukin-two [IL-2], interleukin-one beta [IL-1 β], interleukin-one receptor antagonist [IL1RA], tumor necrosis factor receptor two [TNFR2], tumor necrosis factor receptor one [TNFR1], and tumor necrosis factor alpha [TNF α]); (2) connective tissue (vascular endothelial growth factor [VEGF], matrix metalloproteinase three [MMP3], and brain derived neurotrophic factor [BDNF]); (3) lipid (paraoxonase, adiponectin, and leptin); (4) hemostasis (D-dimer); and (5) metabolic pathways (glycated albumin [GSP]).

The procedures for biomarker quantification are described below (unless otherwise stated, all kits were used to analyze serum according to the manufacturer's instructions). Total

adiponectin was measured by enzyme-linked immunosorbent assay (ELISA) using a kit from ALPCO Diagnostics (Salem, NH), with samples diluted 1:1000. D-dimer was measured by ELISA using a kit from American Diagnostics (Stamford, CT), with plasma samples diluted 1:50. IL-6 was measured using a high sensitivity immunoassay from MesoScale Discovery (Gaithersburg, Maryland). Leptin was measured by ELISA using a kit from Millipore Corporation (Billerica, MA), with 25 μ l of sample and standards used. MMP-3 was measured by ELISA using a kit from Invitrogen Corporation. Paraoxonase activity was measured (nmol product formed/min/ml) via a kit from Invitrogen Corporation; serum samples were diluted 1:100 and the reaction was stopped after 60 minutes and fluorescence measured with an excitation wavelength of 360nm and an emission wavelength of 465nm. TRAIL was measured by ELISA using a kit from Invitrogen. High sensitivity CRP was detected by solid-phase ELISA (MAGIWEL; UBI, Mountain View, CA). GSP was detected via the specific enzymatic method using reagents from DIAZYME (Poway, CA). For individual assays, samples were measured in duplicate. The inter- and intra-assay percent coefficients of variation for the individual assays are as follows: adiponectin (2.31, 8.84), D-dimer (4.32, 9.35), IL-6 (4.36, 16.9), leptin (5.61, 6.46), MMP-3 (2.59, 11.04), paraoxonase (2.58, 15.9), and TRAIL (3.74, 18.0). A Luminex Panel from Invitrogen Corporation (Carlsbad, CA) was used to measure BDNF, GCSF, IL-1RA, IL-1 β , IL-2, IL-8, MCP1, RANTES, TNF α , TNFR1, TNFR2, and VEGF. For Luminex, samples were assayed in individual wells, from which at least one hundred beads were analyzed. Biomarkers with greater than 25% of the measured sample concentrations at or below the limit of quantification (LOQ) were not further analyzed (BDNF, IL2, and IL1 β). For the remaining biomarkers, samples with concentrations below LOQ (representing \leq 2% of samples for any given biomarker, Table S1) were assigned a value of $\frac{1}{2}$ LOQ.

Genome-wide genotyping

DNA was isolated and quantified according to standard protocols, as previously described [2]. Genome-wide genotyping was performed using the Infinium HumanLinkage-12 Genotyping BeadChip (Illumina, San Diego, CA). This assay includes 6,090 single-nucleotide polymorphism (SNP) markers with an average marker density of one per 0.58 centiMorgans (cM). The quality control measures included: genotyping two control samples from the Centre d'Étude du Polymorphisme Humain (CEPH) and requirement of a genotype call-rate \geq 98% for each SNP. A total of 6,015 of the 6,090 SNPs (98.8%) met this quality control standard. Two individuals were excluded from the analysis due to overall low call rates (<95%). In addition, Mendelian inconsistencies, Hardy-Weinberg equilibrium and errors in sex assignment were examined. Specifically, the genetic analysis package RELPAIR [4] was used to verify the reported family relationships. Twenty-four unrelated individuals were excluded from any further analysis. One individual was removed due to gender error using PLINK [5]. The genetic analysis package VITESSE [6] was used to identify Mendelian errors in genotyping across

generations. Of the 2+ million genotypes across all SNPs typed on these individuals, there were a total of 188 Mendelian errors. These genotypes were excluded from further analysis. SNPs out of deviation with Hardy-Weinberg equilibrium ($p < 0.0001$) were removed and not further analyzed. The deCODE genetic map was used to position the markers (deCODE Genetics, Reykjavik, Iceland).

Statistical analysis

Heritability. The broad-sense heritability (H^2) of each biomarker was estimated using the Sequential Oligogenic Linkage Analysis Routines (SOLAR) [7] with an adjustment for age and sex. Broad-sense heritability includes the aggregate genetic variance resulting from additive, epistatic, dominant, maternal, and paternal genetic effects. As the assumption of normality is very important in variance components analysis several transformations were used to achieve approximate normality [8]. All biomarker levels were log transformed to approximate a normal distribution. After log transformation, there was residual kurtosis for nine of the biomarkers (D-dimer, GSP, IL1RA, IL-6, RANTES, TNF α , TNFR1, TNFR2, and TRAIL). Extreme outliers were sequentially removed for these biomarkers, starting with the removal of any values greater than or equal to four standard deviations from the mean. If the marker distribution still had significant kurtosis, then any values greater than or equal to three standard deviations from the mean were removed. This process resulted in residual kurtosis remaining for two markers (TNF α and IL-6) and these markers were analyzed using the “lodadj” command implemented in SOLAR to compensate for this lack of normality. Three traits (adiponectin, TNF α , and TNFR1) required rescaling to approximate normality and in those cases the trait value was multiplied by a factor ranging from 2.3–5.8. A polygenic variance components model was fitted and used as the foundation for subsequent linkage analysis.

Quantitative trait linkage analysis. Two-point and multipoint genome-wide linkage scans were performed for all autosomes using 18 CVD-related biomarkers as quantitative traits. As recommended for multigenerational pedigrees, linkage between each of the biomarker traits and marker loci was tested by maximum-likelihood methods using a variance components model [9]. The size and complexity of the CARRIAGE pedigree necessitated first computing the identity-by-descent (IBD) probabilities for each pair of individuals at each marker using the Markov-Chain Monte-Carlo (MCMC) algorithm implemented in the Loki analysis package [10]. Linkage was interpreted as significant if the logarithm of odds (LOD) score was ≥ 3.0 , “noteworthy” if the LOD score was ≥ 2.6 , and scores ≥ 2.0 were identified as “interesting”.

Results

Baseline clinical characteristics of the study population are presented in Table 1. Mean levels for each of the 18 biomarkers are listed in Table 2, and are similar to those seen in the general American population, consistent with the fact that this population was not ascertained for any particular disease or metabolic phenotype. Eleven of the 18 biomarkers showed

Table 1. Baseline clinical characteristics of the CARRIAGE cohort.

Clinical characteristic	Mean (SD)	Percent
Sex (% female)		66%
Hypertension		42%
Diabetes		15%
Smoker		78%
Body mass index	31.1 (6.8)	
Age, years	54.1 (15.3)	
Low density lipoprotein (LDL) cholesterol, mg/dL	112.2 (36.5)	
High density lipoprotein (HDL) cholesterol, mg/dL	48.0 (14.1)	
Triglycerides, mg/dL	137.2 (90.9)	

*. Quantitative traits presented as mean (standard deviation); discrete traits presented as percent prevalence.

Table 2. Summary statistics for CVD biomarkers.

Biomarker	Units	Mean (SD)	Max	Min
Adiponectin	ng/mL	13044.2 (6404.3)	52794.0	4462.7
hsCRP	ng/mL	8.1 (1.4)	11.4	3.8
DDIMER	ng/mL	703.7 (1015.0)	9057.0	78.3
GCSF	pg/mL	314.6 (388.5)	3225.0	20.6
GSP	μ mol/L	227.3 (58.1)	675.1	106.0
IL1RA	pg/mL	4502.1 (10758.7)	110666.0	158.0
IL6	pg/mL	1.5 (5.4)	80	.03
IL8	pg/mL	85.1 (50.8)	542.6	12.2
Leptin	ng/mL	30.4 (26.2)	304.1	0.3
MCP1	pg/mL	2601.5 (1601.2)	11840.6	151.7
MMP3	ng/mL	7.2 (7.3)	79.1	1.0
Paraoxonase	nmol/min/L	11.6 (4.1)	27.4	2.3
RANTES	pg/mL	9584.7 (17167.6)	243369.0	819.7
TNF α	pg/mL	29.5 (88.8)	1240.7	5.9
TNFR1	pg/mL	5799.2 (3444.5)	25533.2	533.0
TNFR2	pg/mL	2485.7 (1482.1)	13936.6	136.0
TRAIL	pg/mL	491.2 (195.2)	2012.9	86.9
VEGF	pg/mL	227.6 (233.2)	1685.6	32.9

The unit of measurement, mean, standard deviation, maximum, and minimum values for each biomarker are given.

nominally statistically significant heritability ($p < 0.05$, Figure 1), with heritability estimates ranging from 0.33 (leptin, $p = 0.02$, $SE = 0.18$) to 0.56 (hsCRP, $p = 0.00006$, $SE = 0.15$).

Two point linkage analysis revealed that the highest LOD score was for VEGF (chromosome 4p14, SNP rs790142, $LOD = 2.6$). In addition, eight additional SNPs were identified as interesting based on evidence of linkage with $LOD \geq 2.0$ for five biomarkers (VEGF, hsCRP, MCP-1, D-dimer, and paraoxonase, Table 3). The maximum two point LOD score obtained for each biomarker per chromosome is presented in Table S2.

In multipoint analyses, we identified noteworthy ($LOD \geq 2.6$) QTL for four of the 18 CVD biomarkers: paraoxonase, RANTES, MMP3, and GCSF (Figure 2a–d, Table 4, Figure S1a–d, Figure S2a–n). Specifically, paraoxonase had

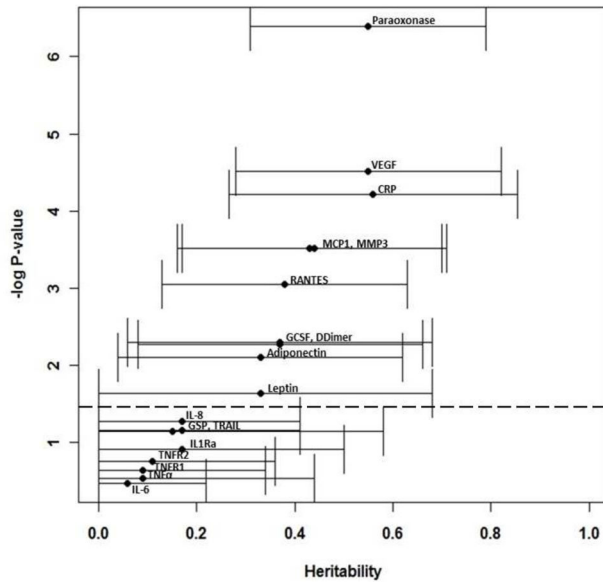


Figure 1. Heritabilities of measured CVD biomarkers. Presented is the distribution of heritability and its corresponding p-value for all 18 biomarkers, with the heritability estimate on the X-axis and the $-\log$ base 10 of the associated p-value on the Y-axis. The 95% confidence interval is represented by a horizontal error bar. The threshold for significance is represented by a dashed line.

doi: 10.1371/journal.pone.0071779.g001

Table 3. Results for genome-wide linkage, two-point LOD scores.

Biomarker	LOD	SNP	Chromosome	Gene*
hsCRP	2.10	RS1419607	7	Intergenic (<i>POT1</i> , <i>GRM8</i>)
	2.02	RS728919	11	<i>NAA40</i>
	2.49	RS7240966	18	Intergenic (<i>CTIF</i>)
D-dimer	2.18	RS2780701	9	<i>SYK</i>
MCP1	2.14	RS857819	1	Intergenic (<i>OR6N1</i>)
Paraoxonase	2.10	RS491603	1	Intergenic (<i>EIF2C3</i>)
	2.00	RS726455	13	Intergenic (<i>SOX1</i>)
	2.25	RS234	7	Intergenic (<i>CDHR3</i>)
VEGF	2.56	RS790142	4	Intergenic (<i>NSUN7</i>)

Results for genomic regions with noteworthy (LOD \geq 2.6) or interesting (LOD \geq 2.0) two-point LOD scores are presented. The biomarker name is given, followed by the two-point LOD score, SNP rs number, chromosome and gene name.

*if SNP is intergenic, the closest gene(s) is listed in parentheses.

noteworthy or interesting evidence for linkage at three locations: 8p11.21 (multipoint LOD [MLOD] 2.8), 7q22.1 (MLOD 2.5), and 19q12 (MLOD 2.1). RANTES had evidence for linkage at one location, 22q13.33 (MLOD 2.8). MMP3 had evidence for linkage at three locations: 17p13.3 (MLOD 2.6), 6p22.3 (MLOD 2.5), and 5q12.3 (MLOD 2.5). Finally, GCSF had evidence for linkage at one location, 8q22.1 (MLOD 2.6). Additionally, six of the remaining 14 biomarkers had interesting

results based on evidence for linkage with a LOD \geq 2.0 (Table 4).

Discussion

Using a large, multiethnic, multigenerational extended family, we have successfully identified QTL using CVD-related inflammatory and metabolic biomarkers. These QTL may harbor genes for genetic susceptibility for CVD mediated through these biological pathways. Most of the biomarkers were also found to be significantly heritable ($p < 0.05$); we believe this to be a novel finding for GCSF and MMP3, while the other heritable biomarkers have support from the previous literature [11–13]. Our unique study design, employing a single extended multigenerational family not ascertained on CVD and thus with a burden of CVD and related risk factors similar to or less than the United States population, facilitates extension of these findings to the general population with a common burden of risk factors.

The strongest QTL was for paraoxonase, an enzyme associated with high density lipoprotein (HDL) cholesterol that inhibits oxidation of low density lipoprotein (LDL) cholesterol. Oxidized LDL is important in the atherosclerotic process [14], and it has been shown that low paraoxonase levels are associated with increased risk of myocardial infarction [15]. In our study, there were several interesting QTL for paraoxonase levels, the strongest of which was on chromosome 8p11, 21, a genomic region which contains genes encoding several transcription factors (*ZMAT4*, *NKX6-3*, *IKBKB*, *THAP1*, and *RNF170*) and an enzyme involved in post-translational modification (*FNTA*); all of these could be plausible candidates for regulators *in trans* of paraoxonase levels. Previously published genome-wide linkage studies have reported linkage to other parts of the genome, including the physical locus for the paraoxonase gene cluster (*PON1*, *PON2*, *PON3*), on chromosome 7q, but those studies did not report evidence at any of the other loci detected in our study [16]. The linkage peak we identified at 7q22.1 occurs at the paraoxonase gene cluster and serves as a proof of principle for the accuracy of our analyses, even in this complex family-based study. The paraoxonase gene cluster regulates paraoxonase levels and *PON1* and *PON3* genetic variants are associated with CVD risk [17,18], supporting the concept of using intermediate disease-related markers as quantitative genetic traits for disease gene mapping. The regions linked to paraoxonase levels are also linked to other disease traits. For example, the 7q22 region, that contains the genes *reelin* (*RELN*) and *leptin* (*LEP*), is linked to several conditions including: osteoarthritis [19], autism [20], body mass index [21,22], and dilated cardiomyopathy [23]. The region on 19q12, that contains the gene for the ryanodine receptor [*RYR1*] [24] (a class of intracellular calcium channels found primarily in cardiac muscle), are linked to paraoxonase levels as well as linked to waist circumference [25], BMI [26], resistance to muscle fatigue [27], essential hypertension [28], prostate cancer [29], maturity onset diabetes of the young (MODY) [30], and malignant hyperthermia [31].

The chemotactic cytokine RANTES recruits T-cells, eosinophils, and basophils to sites of inflammation and is

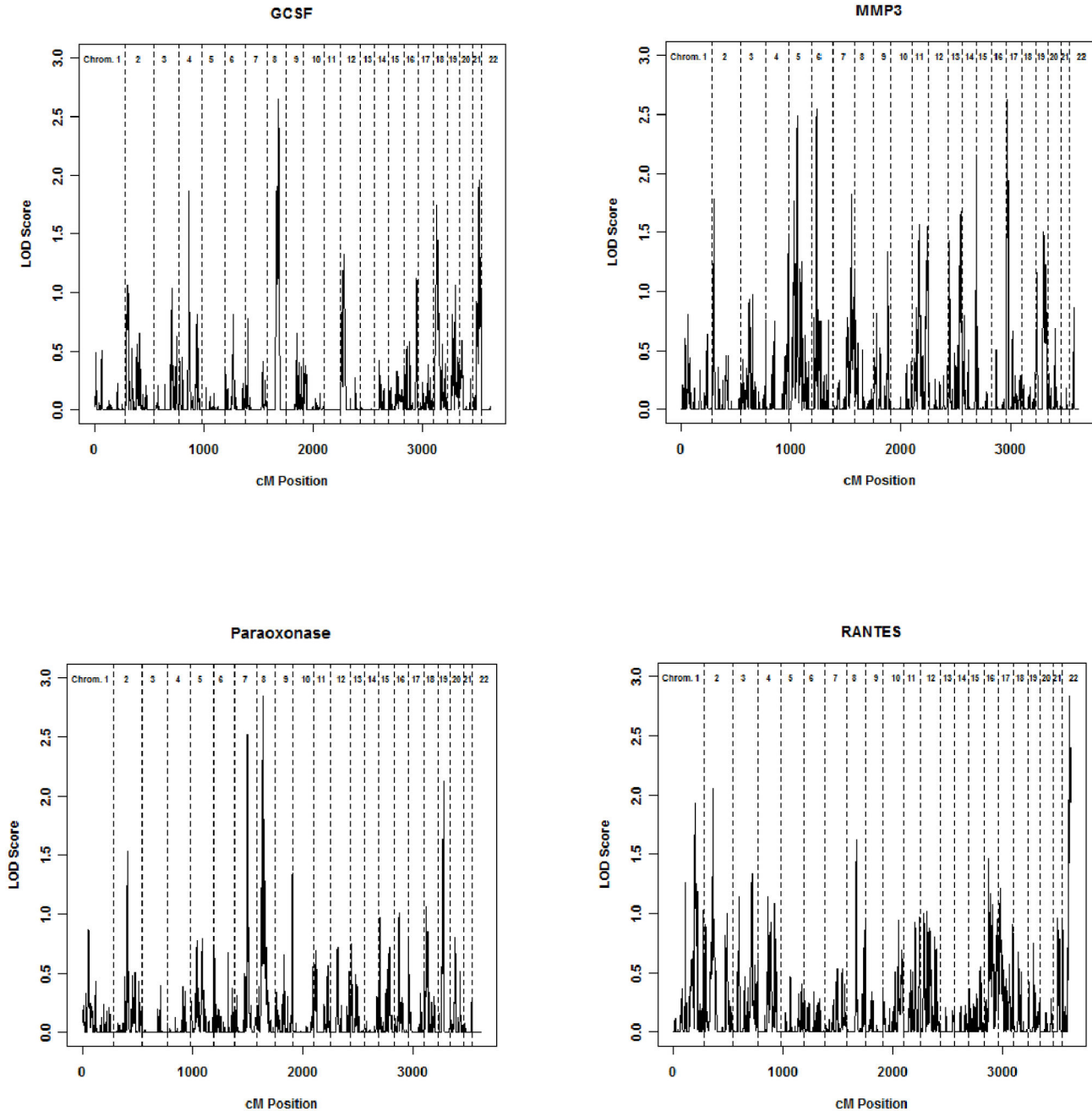


Figure 2. Multipoint genome-wide linkage scans for GCSF, MMP3, paraoxonase, and RANTES. Vertical dashed lines represent boundaries between chromosomal regions and the cumulative cM position is indicated on the X axis.

doi: 10.1371/journal.pone.0071779.g002

therefore a likely participant in CVD through the contribution of inflammation to atherosclerotic plaque formation and response to plaque rupture. The region around the LOD score peak for RANTES (22q13) contains the gene *IL17REL* (IL17 Receptor E Like), which was recently identified in a genome-wide association study (GWAS) for ulcerative colitis [32]. The ligands for this receptor are unknown but, due to its high sequence homology to *IL17RE*, it is very likely that it binds a

cytokine and has an immunologic role and thus could affect levels of RANTES. This locus is linked to several traits including pulse pressure [33], bone mineral density [34], serum creatinine [35], rheumatoid arthritis [36], schizophrenia [37], height [38], and breast cancer [39].

Matrix metalloproteinase (MMP3) plasma concentrations are linked to increased risk of plaque rupture and, thus, to myocardial infarction [40]. In our study, the strongest evidence

Table 4. Chromosomal locations for highest multipoint LOD scores.

Biomarker	Region	SNPs*	Region	SNPs*
	LOD 2.6-3.0		LOD 2.0-2.5	
Paraoxonase	8p11.21(2.8)	RS8685, RS749540	7q22.1(2.5) 19q12(2.1)	RS1229540, RS234 RS7250192, RS2194198
RANTES	22q13.33(2.8)	RS7410750, RS10451		
MMP3	17p13.3(2.6)	RS216219, RS12746	6p22.3(2.5) 5q12.3(2.5)	RS965037, RS1264451 RS1020661, RS164561
GCSF	8q22.1(2.6)	RS1051624, RS951826		
GSP			13q13.2(2.4)	RS306395, RS668103
TRAIL			1q44(2.3)	RS164561, RS2027432
hsCRP			7q32.3(2.2)	RS12217, RS1371463
MCP1			9q34.3(2.1) 1q22(2.0)	RS3132332, RS7357733 RS13320, RS10918078
D-dimer			9q34.2(2.1)	RS10818768, RS7860423
Adiponectin			11p15.5(2.1)	RS741737, RS879114

Presented is a summary of all multipoint LOD score results greater than 2.0 obtained in the genome scan (multipoint LOD score in parentheses), with the SNPs flanking the 1 LOD down interval.

*. SNPs flanking 1 LOD down interval around peak marker in QTL.

for linkage to MMP3 levels was found within a relatively gene-dense region of chromosome 17, which contains several interesting candidate genes including a scavenger receptor for LDL (*SCARF1*) and a highly conserved intracellular signaling protein (*YWHAE*). In addition, this region is linked to ventricular hypertrophy [41], childhood obesity [42], and rheumatoid arthritis [43], among others. The chromosome 6p22 region we found linked to MMP3 levels is linked to sarcoidosis [44], schizophrenia [45], reading ability [45], pulse pressure [46], and early onset myocardial infarction [47]. The third peak for MMP3 in the 5q12 region is linked to low density lipoprotein size [48] and stroke [49].

GCSF is a growth factor and a cytokine which, in addition to its obvious role in promoting the growth of granulocytes, also promotes the growth of stem cells and their release from the bone marrow and has been implicated in the response of vascular endothelial cells to oxidative stress [50]. We observed the strongest evidence for linkage to GCSF levels at chromosome 8q22, which contains the hematopoietic transcription factor *RUNX1T1*. Although there are no data specifically suggesting this, one might postulate regulatory

networks in hematopoiesis wherein alterations in *RUNX1T1* function or expression impact GCSF levels. Other studies have reported linkage at 8q22 for hypertension [51], dihydrotestosterone levels [52], and Tourette's syndrome [53], although the actual genes have not yet been identified.

The significant heritabilities of these CVD biomarkers are not necessarily surprising, as they are potential predictors or risk factors of CVD, and CVD itself has a relatively strong genetic component (i.e. heritability of 0.38-0.57 [54]). A significant heritability means genetic components can explain part of the variation in the trait. Such components can, in theory, be mapped; however, it does not necessarily mean that the underlying genetic model will allow those components to be easily detected via current techniques. For example, as is the case for human height, a trait may have a very high heritability that is due to the additive behavior of many genes, each of which contributes a small amount to the trait variability; such a trait would be very difficult to map with current QTL methods. Thus, it is interesting to note that of the 11 biomarkers with significant heritability point estimates, eight showed interesting results with evidence for linkage to one or more genomic regions, (LOD \geq 2.0). This suggests that the genetic architecture governing levels of these biomarkers may be amenable to mapping and potentially eventual positional cloning. Interestingly, the biomarker TRAIL did not have significant heritability estimates and yet had some interesting results for evidence for linkage. This discrepancy is likely to result from a higher intra-assay coefficient of variation which may impact the estimate of the heritability as well as reduce power of the linkage analysis.

We can make some specific inferences about this genetic architecture by examining the position of linkage peaks relative to those loci that directly encode a biomarker. For example, the enzyme paraoxonase is encoded by a cluster of genes (*PON1-3*) on chromosome 7q. In our study, we had linkage to this region on 7q, suggesting regulation *in cis* and possibly allelic variation in the *PON* genes themselves, as has been previously observed [18]. However, the strongest LOD score related to paraoxonase in our study was not linked to the paraoxonase locus on 7q but was found on 8p11. This suggests that regulation *in trans* could contribute in some significant way to variability in paraoxonase levels. The role of regulation *in trans* is underscored by the fact that, of the remaining LOD scores \geq 2.6 (MMP3, RANTES, and GCSF), none were coincident with the physical loci encoding the biomarker. Thus, just by examining the locations of our LOD scores relative to the loci directly encoding the biomarker in question, we were able to unravel some of the genetic architecture of the trait. Interestingly, we did not find overlap between the only interesting QTL for hsCRP in our study (chromosome 7q32, LOD 2.2) with other published genomic regions linked to and/or associated with hsCRP levels (1p22 [55], 1q23 [55], 2q14 [55], 10q21 [55], 11p11-p13 [56], 11q14 [55], 12p11 [57], 12q15 [57], 19q12 [55], and 20q13 [58]); CRP levels (1p22 [55], 1q23 [55], 2q14 [55], 10q21 [55], 11p11-p13 [56], 11q14 [55], 12p11 [57], 12q15 [57], 19q12 [55], and 20q13 [58]); this may be due to the fact that our population was not ascertained based on disease status, locus heterogeneity,

and/or the nonspecific nature of hsCRP as a biomarker of inflammation.

There are limitations to the current approach. Namely, the study was conducted in a population with a specific ancestry, primarily African and Native American and therefore the genetic loci detected in our genome screen may not be applicable to other ethnicities and populations. However, in some ways this 'limitation' is also a great strength, as African Americans are an understudied population at high risk for CVD. Furthermore, it is not known to what extent a primarily African American sample would be expected to share causal variation underlying biomarker levels with another ethnic group. In addition, it is possible that the presence of linkage disequilibrium (LD) between SNPs can inflate LOD scores and, as our population is of mixed African American descent, there exists the potential for the significant LD inherent in recently admixed populations. However, the SNPs selected for the genome scan were designed to limit LD between markers and there was no significant LD ($r^2 > 0.4$) between any of the SNPs in our most significant linkage peaks (data not shown). The genome scans themselves were conducted under the assumption that the biomarkers were not correlated and that each genome scan was a set of independent tests. However, the level of inter-marker correlations between biomarkers could invalidate that assumption and imposes a multiple testing burden that could influence the significance interpretation of our LOD scores. Identifying all MLODs greater than or equal to two as "interesting" resulted in 15 loci, translating to only 3.7% of the MLODS across all biomarkers and the 22 autosomes. Even so, any of the 15 highlighted results could be type I errors. However, our intent was to highlight the regions providing the most evidence for linkage in our study. Finally, there is no single method that allows for maximal power while using all measurements of the quantitative trait, including extreme outliers. In the current study, we have elected to remove extreme values at either range of the quantitative measure, thus creating distributions that are closer to normality. In so doing, it is possible that some biologically meaningful information has been lost. Other analytic methods (such as the lodadj option in SOLAR) would allow those values to be included, but could also introduce compromises in power, particularly when the overall data does correspond well to a normal distribution.

The results presented here can advance our understanding of CVD in several ways depending on how the biomarker in question relates to the pathogenesis of CVD. First, if the biomarker is a biochemical mediator of CVD, i.e. the biomarker is causal in CVD pathogenesis, then identifying the genes that modify levels of this biomarker could identify targets for development of therapeutic targeting of those genes as well as serving as CVD biomarkers themselves. Second, if the biomarkers are a result of CVD, then the genetic variants

responsible for their levels should be genetic risk factors for disease severity. In that case, identifying those variants will advance efforts to predict CVD burden or risk of progression by genotyping. Each of the four biomarkers with LOD scores ≥ 2.6 has been implicated in some way with the pathogenesis or risk of CVD. Fine-mapping of these QTL to identify the responsible gene(s), subsequent evaluation of those genetic markers for association with CVD, and validation in further cohorts are necessary. The identification of the putative causal variants underlying the linkage results for these four biomarkers will not only advance our understanding of cardiovascular risk but hopefully serve as a model for the study of other complex diseases via the genetic dissection of intermediate traits.

Supporting Information

Table S1. Percent of samples measured as below lower limits of quantification for a given biomarker assay.
(DOCX)

Table S2. Displayed are the maximum two-point LOD score at theta equal zero (MaxLOD) for each biomarker by chromosome, with the name of the probe (RS number), the centimorgan (cM) position of the probe, and the gene annotation for the SNP (or the closest gene for intergenic SNPs).
(DOCX)

Figure S1. Chromosome linkage plots for the most significant multipoint linkage peaks.
(DOCX)

Figure S2. Genome-wide autosomal multipoint linkage results for each biomarker (for those not already presented in main manuscript).
(DOCX)

Acknowledgements

We are grateful to the CARRIAGE family members for their participation in this research. We also thank Elaine Dowdy, Norine Hall and Milton Campbell for their assistance with data and biological sample collection.

Author Contributions

Conceived and designed the experiments: DN WEK ERH YJL HCC CH SGG VBK SHS. Performed the experiments: DKT JJ HCC. Analyzed the data: DN ERH YJL SN CH SHS. Contributed reagents/materials/analysis tools: VBK WEK SHS. Wrote the manuscript: DN ERH SHS.

References

1. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD et al. (2012) Heart disease and stroke statistics--2012 update: a report from the American Heart Association. *Circulation* 125: e2-e220. doi:10.1161/CIR.0b013e31823ac046. PubMed: 22179539.
2. Chen HC, Kraus VB, Li YJ, Nelson S, Haynes C et al. (2010) Genome-wide linkage analysis of quantitative biomarker traits of osteoarthritis in a large, multigenerational extended family. *Arthritis Rheum* 62: 781-790. doi:10.1002/art.27288. PubMed: 20187133.
3. Ober C, Abney M, McPeck MS (2001) The genetic dissection of complex traits in a founder population. *Am J Hum Genet* 69: 1068-1079. doi:10.1086/324025. PubMed: 11590547.
4. Epstein MP, Duren WL, Boehnke M (2000) Improved inference of relationship for pairs of individuals. *Am J Hum Genet* 67: 1219-1231. doi:10.1086/321195. PubMed: 11032786.
5. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575. doi:10.1086/519795. PubMed: 17701901.
6. O'Connell JR, Weeks DE (1995) The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. *Nat Genet* 11: 402-408. doi:10.1038/ng1295-402. PubMed: 7493020.
7. Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62: 1198-1211. doi: 10.1086/301844. PubMed: 9545414.
8. Blangero J, Williams JT, Almasy L (2001) Variance component methods for detecting complex trait loci. *Adv Genet* 42: 151-181. doi: 10.1016/S0065-2660(01)42021-9. PubMed: 11037320.
9. Amos CI (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 54: 535-543. PubMed: 8116623.
10. Heath SC (1997) Markov chain Monte Carlo methods for radiation hybrid mapping. *J Comput Biol* 4: 505-515. doi:10.1089/cmb.1997.4.505. PubMed: 9385542.
11. Bielinski SJ, Pankow JS, Miller MB, Hopkins PN, Eckfeldt JH et al. (2007) Circulating MCP-1 levels shows linkage to chemokine receptor gene cluster on chromosome 3: the NHLBI family heart study follow-up examination. *Genes Immun* 8: 684-690. doi:10.1038/sj.gene.6364434. PubMed: 17917677.
12. Rainwater DL, Rutherford S, Dyer TD, Rainwater ED, Cole SA et al. (2009) Determinants of variation in human serum paraoxonase activity. *Heredity (Edinb)* 102: 147-154. doi:10.1038/hdy.2008.110. PubMed: 18971955.
13. Dupuis J, Larson MG, Vasan RS, Massaro JM, Wilson PW et al. (2005) Genome scan of systemic biomarkers of vascular inflammation in the Framingham Heart Study: evidence for susceptibility loci on 1q. *Atherosclerosis* 182: 307-314. doi:10.1016/j.atherosclerosis.2005.02.015. PubMed: 16159603.
14. Navab M, Berliner JA, Watson AD, Hama SY, Territo MC et al. (1996) The Yin and Yang of oxidation in the development of the fatty streak. A review based on the 1994 George Lyman Duff Memorial Lecture. *Arterioscler Thromb Vasc Biol* 16: 831-842. doi:10.1161/01.ATV.16.7.831. PubMed: 8673557.
15. McElveen J, Mackness MI, Colley CM, Peard T, Warner S et al. (1986) Distribution of paraoxon hydrolytic activity in the serum of patients after myocardial infarction. *Clin Chem* 32: 671-673. PubMed: 3006944.
16. Winnier DA, Rainwater DL, Cole SA, Dyer TD, Blangero J et al. (2006) Multiple QTLs influence variation in paraoxonase 1 activity in Mexican Americans. *Hum Biol* 78: 341-352. doi:10.1353/hub.2006.0050. PubMed: 17216806.
17. Bhattacharyya T, Nicholls SJ, Topol EJ, Zhang R, Yang X et al. (2008) Relationship of paraoxonase 1 (PON1) gene polymorphisms and functional activity with systemic oxidative stress and cardiovascular risk. *JAMA* 299: 1265-1276. doi:10.1001/jama.299.11.1265. PubMed: 18349088.
18. Sanghera DK, Aston CE, Saha N, Kamboh MI (1998) DNA polymorphisms in two paraoxonase genes (PON1 and PON2) are associated with the risk of coronary heart disease. *Am J Hum Genet* 62: 36-44. doi:10.1086/301669. PubMed: 9443862.
19. Evangelou E, Valdes AM, Kerkhof HJ, Styrkarsdottir U, Zhu Y et al. (2011) Meta-analysis of genome-wide association studies confirms a susceptibility locus for knee osteoarthritis on chromosome 7q22. *Ann Rheum Dis* 70: 349-355. doi:10.1136/ard.2010.132787. PubMed: 21068099.
20. Cukier HN, Skaar DA, Rayner-Evans MY, Konidari I, Whitehead PL et al. (2009) Identification of chromosome 7 inversion breakpoints in an autistic family narrows candidate region for autism susceptibility. *Autism Res* 2: 258-266. doi:10.1002/aur.96. PubMed: 19877165.
21. Jiang Y, Wilk JB, Borecki I, Williamson S, DeStefano AL et al. (2004) Common variants in the 5' region of the leptin gene are associated with body mass index in men from the National Heart, Lung, and Blood Institute Family Heart Study. *Am J Hum Genet* 75: 220-230. doi: 10.1086/422699. PubMed: 15197684.
22. Wu X, Cooper RS, Borecki I, Hanis C, Bray M et al. (2002) A combined analysis of genomewide linkage scans for body mass index from the National Heart, Lung, and Blood Institute Family Blood Pressure Program. *Am J Hum Genet* 70: 1247-1256. doi:10.1086/340362. PubMed: 11923912.
23. Schönberger J, Kühler L, Martins E, Lindner TH, Silva-Cardoso J et al. (2005) A novel locus for autosomal-dominant dilated cardiomyopathy maps to chromosome 7q22.3-31.1. *Hum Genet* 118: 451-457. doi: 10.1007/s00439-005-0064-2. PubMed: 16228230.
24. MacLennan DH, Duff C, Zorzato F, Fujii J, Phillips M et al. (1990) Ryanodine receptor gene is a candidate for predisposition to malignant hyperthermia. *Nature* 343: 559-561. doi:10.1038/343559a0. PubMed: 1967823.
25. Voruganti VS, Diego VP, Haack K, Cole SA, Blangero J et al. (2011) A QTL for Genotype by Sex Interaction for Anthropometric Measurements in Alaskan Eskimos (GOCADAN Study) on Chromosome 19q12-13. *Obesity (Silver Spring)*, 20: 1122-6. PubMed: 22016090.
26. Dai F, Sun G, Aberg K, Keighley ED, Indugula SR et al. (2008) A whole genome linkage scan identifies multiple chromosomal regions influencing adiposity-related traits among Samoans. *Ann Hum Genet* 72: 780-792. doi:10.1111/j.1469-1809.2008.00462.x. PubMed: 18616661.
27. Thomas MA, De Mars G, Windelinckx A, Peeters MW, Huygens W et al. (2010) Genome-wide linkage scan for resistance to muscle fatigue. *Scand J Med Sci Sports*, 21: 580-8. PubMed: 20459472.
28. Bell JT, Wallace C, Dobson R, Wiltshire S, Mein C et al. (2006) Two-dimensional genome-scan identifies novel epistatic loci for essential hypertension. *Hum Mol Genet* 15: 1365-1374. doi:10.1093/hmg/ddl058. PubMed: 16543358.
29. Liu X, Cheng I, Plummer SJ, Suarez BK, Casey G et al. (2010) Fine-mapping of prostate cancer aggressiveness loci on chromosome 7q22-35. *Prostate*, 71: 682-9. PubMed: 20945404.
30. Kim SH, Ma X, Weremowicz S, Ercolino T, Powers C et al. (2004) Identification of a locus for maturity-onset diabetes of the young on chromosome 8p23. *Diabetes* 53: 1375-1384. doi:10.2337/diabetes.53.5.1375. PubMed: 15111509.
31. McCarthy TV, Healy JM, Heffron JJ, Lehane M, Deufel T et al. (1990) Localization of the malignant hyperthermia susceptibility locus to human chromosome 19q12-13.2. *Nature* 343: 562-564. doi: 10.1038/343562a0. PubMed: 2300206.
32. Franke A, Balschun T, Sina C, Ellinghaus D, Häsler R et al. (2010) Genome-wide association study for ulcerative colitis identifies risk loci at 7q22 and 22q13 (IL17REL). *Nat Genet* 42: 292-294. doi:10.1038/ng.553. PubMed: 20228798.
33. Aberg K, Dai F, Viali S, Tuitele J, Sun G et al. (2009) Suggestive linkage detected for blood pressure related traits on 2q and 22q in the population on the Samoan islands. *BMC Med Genet* 10: 107. doi: 10.1186/1471-2350-10-107. PubMed: 19852796.
34. Peacock M, Koller DL, Lai D, Hui S, Frouud T et al. (2009) Bone mineral density variation in men is influenced by sex-specific and non sex-specific quantitative trait loci. *Bone* 45: 443-448. doi:10.1016/j.bone.2009.05.002. PubMed: 19427925.
35. Pattaro C, Aulchenko YS, Isaacs A, Vitart V, Hayward C et al. (2009) Genome-wide linkage analysis of serum creatinine in three isolated European populations. *Kidney Int* 76: 297-306. doi:10.1038/ki.2009.135. PubMed: 19387472.
36. Barton A, Thomson W, Ke X, Eyre S, Hinks A et al. (2008) Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nat Genet* 40: 1156-1159. doi:10.1038/ng.218. PubMed: 18794857.
37. Condra JA, Neibergs H, Wei W, Brennan MD (2007) Evidence for two schizophrenia susceptibility genes on chromosome 22q13. *Psychiatr Genet* 17: 292-298. doi:10.1097/YPG.0b013e3181ac2345. PubMed: 17728668.
38. Sammalisto S, Hiekkalinna T, Suviolahti E, Sood K, Metzidis A et al. (2005) A male-specific quantitative trait locus on 1p21 controlling human stature. *J Med Genet* 42: 932-939. doi:10.1136/jmg.2005.031278. PubMed: 15827092.
39. Hartikainen JM, Tuhkanen H, Kataja V, Dunning AM, Antoniou A et al. (2005) An autosome-wide scan for linkage disequilibrium-based association in sporadic breast cancer cases in eastern Finland: three

- candidate regions found. *Cancer Epidemiol Biomarkers Prev* 14: 75-80. PubMed: 15668479.
40. Abilleira S, Bevan S, Markus HS (2006) The role of genetic variants of matrix metalloproteinases in coronary and carotid atherosclerosis. *J Med Genet* 43: 897-901. doi:10.1136/jmg.2006.040808. PubMed: 16905683.
 41. Mayosi BM, Avery PJ, Farrall M, Keavney B, Watkins H (2008) Genome-wide linkage analysis of electrocardiographic and echocardiographic left ventricular hypertrophy in families with hypertension. *Eur Heart J* 29: 525-530. doi:10.1093/eurheartj/ehn028. PubMed: 18276622.
 42. Meyre D, Lecoq C, Delplanque J, Francke S, Vatin V et al. (2004) A genome-wide scan for childhood obesity-associated traits in French families shows significant linkage on chromosome 6q22.31-q23.2. *Diabetes* 53: 803-811. doi:10.2337/diabetes.53.3.803. PubMed: 14988267.
 43. Jawaheer D, Seldin MF, Amos CI, Chen WV, Shigeta R et al. (2003) Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. *Arthritis Rheum* 48: 906-916. doi:10.1002/art.10989. PubMed: 12687532.
 44. Rybicki BA, Levin AM, McKeigue P, Datta I, Gray-McGuire C et al. (2011) A genome-wide admixture scan for ancestry-linked genes predisposing to sarcoidosis in African-Americans. *Genes Immun* 12: 67-77. doi:10.1038/gene.2010.56. PubMed: 21179114.
 45. Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y et al. (2009) Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 460: 753-757. PubMed: 19571809.
 46. Zintzaras E, Kitsios G, Kent D, Camp NJ, Atwood L et al. (2007) Genome-wide scans meta-analysis for pulse pressure. *Hypertension* 50: 557-564. doi:10.1161/HYPERTENSIONAHA.107.090316. PubMed: 17635856.
 47. Zintzaras E, Kitsios G (2006) Identification of chromosomal regions linked to premature myocardial infarction: a meta-analysis of whole-genome searches. *J Hum Genet* 51: 1015-1021. doi:10.1007/s10038-006-0053-x. PubMed: 17024316.
 48. Bossé Y, Pérusse L, Després JP, Lamarche B, Chagnon YC et al. (2003) Evidence for a major quantitative trait locus on chromosome 17q21 affecting low-density lipoprotein peak particle diameter. *Circulation* 107: 2361-2368. doi:10.1161/01.CIR.0000065577.60129.F5. PubMed: 12732599.
 49. Gretarsdottir S, Sveinbjörnsdottir S, Jonsson HH, Jakobsson F, Einarsdottir E et al. (2002) Localization of a susceptibility gene for common forms of stroke to 5q12. *Am J Hum Genet* 70: 593-603. doi: 10.1086/339252. PubMed: 11833004.
 50. Kojima H, Otani A, Oishi A, Makiyama Y, Nakagawa S et al. (2010) Granulocyte colony-stimulating factor attenuates oxidative stress-induced apoptosis in vascular endothelial cells and exhibits functional and morphological protective effect in oxygen-induced retinopathy. *Blood*.
 51. Ciullo M, Bellenguez C, Colonna V, Nutile T, Calabria A et al. (2006) New susceptibility locus for hypertension on chromosome 8q by efficient pedigree-breaking in an Italian isolate. *Hum Mol Genet* 15: 1735-1743. doi:10.1093/hmg/ddl097. PubMed: 16611673.
 52. Ukkola O, Rankinen T, Gagnon J, Leon AS, Skinner JS et al. (2002) A genome-wide linkage scan for steroids and SHBG levels in black and white families: the HERITAGE Family Study. *J Clin Endocrinol Metab* 87: 3708-3720. doi:10.1210/jc.87.8.3708. PubMed: 12161500.
 53. Simonic I, Nyholt DR, Gericke GS, Gordon D, Matsumoto N et al. (2001) Further evidence for linkage of Gilles de la Tourette syndrome (GTS) susceptibility loci on chromosomes 2p11, 8q22 and 11q23-24 in South African Afrikaners. *Am J Med Genet* 105: 163-167. doi:10.1002/ajmg.1192. PubMed: 11304830.
 54. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI et al. (2002) Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med* 252: 247-254. doi: 10.1046/j.1365-2796.2002.01029.x. PubMed: 12270005.
 55. Benjamin EJ, Dupuis J, Larson MG, Lunetta KL, Booth SL et al. (2007) Genome-wide association with select biomarker traits in the Framingham Heart Study. *BMC Med Genet* 8 Suppl 1: S11. doi: 10.1186/1471-2350-8-S1-S11. PubMed: 17903293.
 56. Fox ER, Benjamin EJ, Sarpong DF, Rotimi CN, Wilson JG et al. (2008) Epidemiology, heritability, and genetic linkage of C-reactive protein in African Americans (from the Jackson Heart Study). *Am J Cardiol* 102: 835-841. doi:10.1016/j.amjcard.2008.05.049. PubMed: 18805107.
 57. Ruchat SM, Després JP, Weisnagel SJ, Chagnon YC, Bouchard C et al. (2008) Genome-wide linkage analysis for circulating levels of adipokines and C-reactive protein in the Quebec family study (QFS). *J Hum Genet* 53: 629-636. doi:10.1007/s10038-008-0291-1. PubMed: 18414778.
 58. Lakka TA, Rankinen T, Rice T, Leon AS, Rao DC et al. (2006) Quantitative trait locus on chromosome 20q13 for plasma levels of C-reactive protein in healthy whites: the HERITAGE Family Study. *Physiol Genomics* 27: 103-107. doi:10.1152/physiolgenomics.00054.2005. PubMed: 16822830.