

RESEARCH ARTICLE

# Genome-Wide Linkage Scan Identifies Two Novel Genetic Loci for Coronary Artery Disease: In GeneQuest Families

Hanxiang Gao<sup>1,2,9</sup>, Lin Li<sup>2,9</sup>, Shaoqi Rao<sup>2,3,9</sup>, Gongqing Shen<sup>2</sup>, Quansheng Xi<sup>2</sup>, Shenghan Chen<sup>2</sup>, Zheng Zhang<sup>1</sup>, Kai Wang<sup>4</sup>, Stephen G. Ellis<sup>4</sup>, Qiuyun Chen<sup>2</sup>, Eric J. Topol<sup>5\*</sup>, Qing K. Wang<sup>2,4,6,7,8\*</sup>



CrossMark  
click for updates

 OPEN ACCESS

**Citation:** Gao H, Li L, Rao S, Shen G, Xi Q, et al. (2014) Genome-Wide Linkage Scan Identifies Two Novel Genetic Loci for Coronary Artery Disease: In GeneQuest Families. *PLoS ONE* 9(12): e113935. doi:10.1371/journal.pone.0113935

**Editor:** Ralf Krahe, University of Texas MD Anderson Cancer Center, United States of America

**Received:** July 29, 2014

**Accepted:** October 30, 2014

**Published:** December 8, 2014

**Copyright:** © 2014 Gao et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that, for approved reasons, some access restrictions apply to the data underlying the findings. Due to ethical restrictions, the data underlying the findings described in the manuscript are available on request and upon approval by the Cleveland Clinic Institutional Review Board on Human Subject Research. Data requests can be made to Dr. Qing K. Wang at [wangq2@ccf.org](mailto:wangq2@ccf.org).

**Funding:** This work was supported by the National Institute of Health grants (R01 HL121358, R01 HL094498 and P50 HL77107), the American Heart Association grants (0630193N, 11SDG5510001 and 11IRG5570046), the National Natural Science Foundation of China (grant nos. 31071166, 81373085 to SR) and Science and Technology Planning Project of Guangdong Province (grant no. 2009A030301004 to SR). 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.

1. Heart Center, the First Affiliated Hospital, Lanzhou University, Lanzhou, Gansu 730000, P. R. China, 2. Center for Cardiovascular Genetics, Department of Molecular Cardiology, Lerner Research Institute, Cleveland Clinic, Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, 9500 Euclid Ave., Cleveland, Ohio, 44195, United States of America, 3. Institute of Medical Systems Biology and School of Public Health, Guangdong Medical College, Dongguan, Guangdong, 523808, P. R. China, 4. Center for Cardiovascular Genetics, Department of Cardiovascular Medicine, Sydell and Arnold Miller Family Heart and Vascular Institute, Cleveland Clinic, 9500 Euclid Ave., Cleveland, Ohio, 44195, United States of America, 5. Scripps Translational Science Institute, Scripps Research Institute, Scripps Clinic, La Jolla, California, 92037, United States of America, 6. Center for Sleep Medicine, Neurological Institute, Cleveland Clinic, 9500 Euclid Ave., Cleveland, Ohio, United States of America, 7. Department of Genetics and Genome Sciences, Case Western Reserve University, 9500 Euclid Ave., Cleveland, Ohio, 44195, United States of America, 8. Key Laboratory of Molecular Biophysics of the Ministry of Education, College of Life Science and Technology and Center for Human Genome Research, Huazhong University of Science and Technology, Wuhan, P.R. China

\*[etopol@scripps.edu](mailto:etopol@scripps.edu) (EJT); [wangq2@ccf.org](mailto:wangq2@ccf.org) (QKW)

These authors contributed equally to this work.

## Abstract

Coronary artery disease (CAD) is the leading cause of death worldwide. Recent genome-wide association studies (GWAS) identified >50 common variants associated with CAD or its complication myocardial infarction (MI), but collectively they account for <20% of heritability, generating a phenomena of “missing heritability”. Rare variants with large effects may account for a large portion of missing heritability. Genome-wide linkage studies of large families and follow-up fine mapping and deep sequencing are particularly effective in identifying rare variants with large effects. Here we show results from a genome-wide linkage scan for CAD in multiplex GeneQuest families with early onset CAD and MI. Whole genome genotyping was carried out with 408 markers that span the human genome by every 10 cM and linkage analyses were performed using the affected relative pair analysis implemented in GENEHUNTER. Affected only nonparametric linkage (NPL) analysis identified two novel CAD loci with highly significant evidence of linkage on chromosome 3p25.1 (peak NPL =5.49) and 3q29 (NPL =6.84). We also identified four loci with suggestive linkage on 9q22.33, 9q34.11, 17p12, and

21q22.3 (NPL = 3.18–4.07). These results identify novel loci for CAD and provide a framework for fine mapping and deep sequencing to identify new susceptibility genes and novel variants associated with risk of CAD.

## Introduction

Coronary artery disease (CAD) is the leading cause of death globally, and accounts for >13% of all deaths ([www.who.int](http://www.who.int)) [1]. The number of people who die from heart disease is expected to increase to 9.8 million by 2030 ([www.who.int](http://www.who.int)). Long-term prospective clinical and epidemiological studies have identified several major risk factors for development of CAD, including smoking history, older age, male gender, high fat diet, personal history of angina pectoris, family history of myocardial infarction (MI), obesity, diabetes mellitus, high blood pressure, increased plasma total and low-density lipoprotein cholesterol, increased plasma triglycerides, and decreased plasma high-density lipoprotein cholesterol [2–4]. Genetic-epidemiologic studies suggest that family history is the most significant independent risk factor, which accounts for approximately 40% to 60% of the risk for CAD (i.e. an estimated heritability of 0.40–0.60) [5–12]. Identification of the genetic factors responsible for CAD might unravel novel biological pathways involved in development of CAD.

There are two major genome-wide, systematic, comprehensive and unbiased approaches to identify genes and genomic variants for a human disease, including genome-wide linkage studies (GWLS) using family samples and genome-wide association studies (GWAS) using population samples. There are advantages and disadvantages for both. GWAS are more powerful than GWLS to detect common alleles at a locus, but less powerful if the phenotypes of interest are due to the segregation of low-frequency or rare alleles [13]. Furthermore, whereas associations identified by GWAS can be due to spurious causes, especially heterogeneity/population stratification, family-based GWLS are not subject to such type 1 errors [13]. Recent GWAS by many large groups including our GeneID/Cardio-X team have identified about 50 loci for CAD [14–26]. However, it is estimated that the risk variants identified by GWAS may, in aggregate, capture only a small fraction (<20%) of the overall heritability of CAD [8, 27]. Thus, much of the heritability remains missing for CAD.

The major disadvantage of GWAS is that its discovering power is limited to common variants with relatively low risk (odds ratio or OR of 1.2 on average) [13]. In contrast, GWLS are potentially more powerful in detection of low frequency or rare variants, but with large effects on risk of the disease [13]. Therefore, it is highly likely that some of the remaining heritability for CAD may be associated with rare variants but with large effects. Because of the genetic heterogeneity of CAD/MI, GWLS in families could be much more powerful in identifying such rare variants. We consider a family study design to be more

optimal for identification of rare genetic variants with large effects because the same mutation can be observed in many family members and the cause-effect relationship can be easily determined by co-segregation with the disease in the families. Genome-wide linkage scan is a traditional and classical approach to identify new regions of the genome that predispose to CAD. These genome-wide linkage scans with hundreds of sibling pairs have identified several major genetic susceptibility loci for CAD or MI, including loci on chromosome 1p34–36, 1q25, 2p13.1, 2q21.2–22, 2q36–q37.3, 3q13, 5q31, 7p14, 8p22, 13q12–13, 14q32.2, 15q26.2, 16q13, and Xq23–26 [3, 4, 28–30]. The loci identified by linkage analyses were much fewer than GWAS, suggesting that many more linkage loci remain to be identified with new CAD and MI families. Linkage analysis and GWAS have totally different mechanisms in detecting susceptibility loci, and it is highly possible that the linkage loci represent some of the missing heritability of CAD.

In this study, we report results from a genome-wide linkage scan from a well-characterized U.S. population with 428 nuclear families with early onset CAD and myocardial infarction (MI) (each family has at least two affected sibs and one unaffected sib, GeneQuest, mean pedigree size = 5). Multiple new susceptibility genetic loci were identified for CAD, which provide new genetic bases of CAD and a framework to identify underlying susceptibility genes at each locus.

## Subjects and Methods

### GeneQuest Cohort of Families with Early-Onset CAD

The study subjects have been recruited in the United States through the Center for Cardiovascular Genetics and Department of Cardiovascular Medicine at the Cleveland Clinic [31–33]. Important demographic and epidemiological features are shown in [Table 1](#) and as in previously published reports [31–33]. The present study has been approved by the Cleveland Clinic Institutional Review Board on Human Subject Research, and written informed consent was acquired from all participants.

The GeneQuest study was the first large-scale prospective collection of families with early-onset CAD and MI for identifying the underlying genes. The GeneQuest cohort consists of 1,613 subjects from 428 multiplex families with early-onset CAD and MI. Each family has at least two affected sibs, and the majority of families also have one unaffected sibling. The details, including diagnostic criteria and ascertainment of the patients and their families have been described elsewhere [31–33]. In brief, the presence or absence of CAD was assessed according to a previous diagnosis of MI (based on the existence of at least two of the following: chest pain of  $\geq 30$  minutes duration, ECG patterns consistent with acute MI, or significant elevation of cardiac enzymes), history of revascularization procedures such as angioplasty or coronary artery bypass surgery (CABG), or patients undergoing treatment for angina pectoris. For patients in this cohort, the distribution of patients classified by the diagnostic criteria was 54.9% MI, 15.4% CABG, 14.9% angiography with  $>70\%$  of stenosis, 12.2% PTCA

**Table 1.** Clinical and Demographic Features of the Study Populations (GeneQuest)\*.

Feature	GeneQuest
No. male/no. female	1152/878
Age at onset (mean $\pm$ SD)	44.4 $\pm$ 9.7 years
No. affected with CAD/MI	974
Ethnicity:	
Caucasian	91.7%
African American	2.3%
Native American or Alaskan	1.7%
Hispanic	1.4%
Asian or Pacific Islander	0.8%
Mixed ethnicity	0.8%
Unknown ethnicity	1.3%
Pedigree structure:	
No. of pedigrees	428
Pedigree size (mean $\pm$ SD)	4.7 $\pm$ 1.3
No. of relative pairs:	
Sibling/sibling	1,303
Sister/sister	258
Brother/brother	476
Brother/sister	569
Half sibling/half sibling	25

\*From Wang et al and Shen et al [31, 33].

doi:10.1371/journal.pone.0113935.t001

(coronary artery bypass graft), and 2.6% other causes. The GeneQuest cohort had a total of 974 affected individuals recruited, and the average age at onset was 44.4  $\pm$  9.7 years. The epidemiological and demographic features were similar for both affected individuals and unaffected individuals. GeneQuest was characterized by predominantly males (n=1,152) and white (91.7% for all the participants and 93.6% for those affected).

Patients with hypercholesterolemia, insulin-dependent diabetes, childhood hypertension, and congenital heart disease were excluded from this study.

### Isolation of Human Genomic DNA and Genotyping

Whole blood samples were drawn from each participant, and genomic DNA was isolated from the blood by use of the commercial Puregene Kits (Gentra).

Genotyping of microsatellite STR markers was performed by the National Heart, Lung, and Blood Institute (NHLBI) Mammalian Genotyping Service (directed by Dr. James L. Weber), Center for Medical Genetics at Marshfield Clinic by use of the Screening Set 11 consisting of 408 markers that span the entire human genome at every 10 cM [34]. DNA samples were genotyped for 1,163 well-characterized participants from 428 families of the GeneQuest cohort.

## Statistical Analysis

Before statistical analysis of genome-wide genotyping data, obvious pedigree errors, data errors, genotyping errors, and locus-order errors that commonly occur with a large-scale linkage analysis were corrected as described previously [13, 33, 35]. Allele frequencies for all markers used in the analysis were estimated in the GeneQuest population by a maximum likelihood method using the SAGE program *FREQ* [36]. We used *RELTEST* to analyze pedigree relationships and excluded 27 of 428 pedigrees with uncorrectable errors from further linkage analysis. *RELTEST* uses a Markov process model of allele sharing along the chromosome and classifies pairs of pedigree members according to their true relationship by use of genome-scan data [37]. We then used the S.A.G.E. program *MARKERINFO* to detect any Mendelian inheritance inconsistencies and excluded three families with inconsistent Mendelian inheritance from further analysis. Moreover, three pairs of monozygotic twins were identified and excluded from further statistical analysis. Only 388 Caucasian families were analyzed for linkage. To map new susceptibility loci for CAD, genome-wide nonparametric linkage (NPL) analysis was performed by use of the affected relative pair (ARP) analysis implemented in *GENEHUNTER* [38]. Like other ARP methods, the NPL statistic measures allele sharing among the affected individuals within a pedigree [38]. The scoring function statistic was used to evaluate, simultaneously, allele sharing among all those affected in a nuclear family, in contrast to pairwise comparison. Both two-point and multipoint NPL analyses were performed. This scoring function was asymptotically distributed as the *Z* statistic. The significance for linkage inference was determined by use of the criteria proposed previously [39, 40]. Specifically for the sib pair design, the NPL scores for suggestive, significant and highly significant linkage are 3.18 (corresponding to *P* value of  $7.4 \times 10^{-4}$  or LOD score of 2.2), 4.08 (*P* value of  $2.2 \times 10^{-5}$  or LOD score of 3.6) and 4.99 (*P* value of  $3 \times 10^{-7}$  or LOD score of 5.4), respectively.

## Results

Genome-wide NPL analysis of affected sibling pairs in the GeneQuest families was conducted using *GENEHUNTER* (<http://www-genome.wi.mit.edu/ftp/distribution/software/genehunter/>). The genomic regions with a peak two-point or multiple point NPL score that meets the suggestive linkage threshold (i.e.  $NPL \geq 3.18$ ) are listed in [Table 2](#) and shown in [Figs. 1–3](#). Six markers on four different chromosomes (chromosomes 3, 9, 17 and 21) produced an either two-point or multi-point NPL score of  $>3.18$ . The most significant linkage was identified for marker MFD427-AAA on chromosome 3q29 with a two-point NPL score of 6.54 and a multipoint NPL score of 6.84, which reached the genome-wide highly significant level ([Figs. 1–3](#), [Table 2](#)). The second highly significant genome-wide linkage peak was also obtained on chromosome 3 (3p25.2), and had a two-point NPL score of 5.49 and a multipoint NPL score of 5.14 ([Figs. 1–3](#), [Table 2](#)).

**Table 2.** Novel Genetic Loci for CAD Identified by GWLS in the GeneQuest Families.

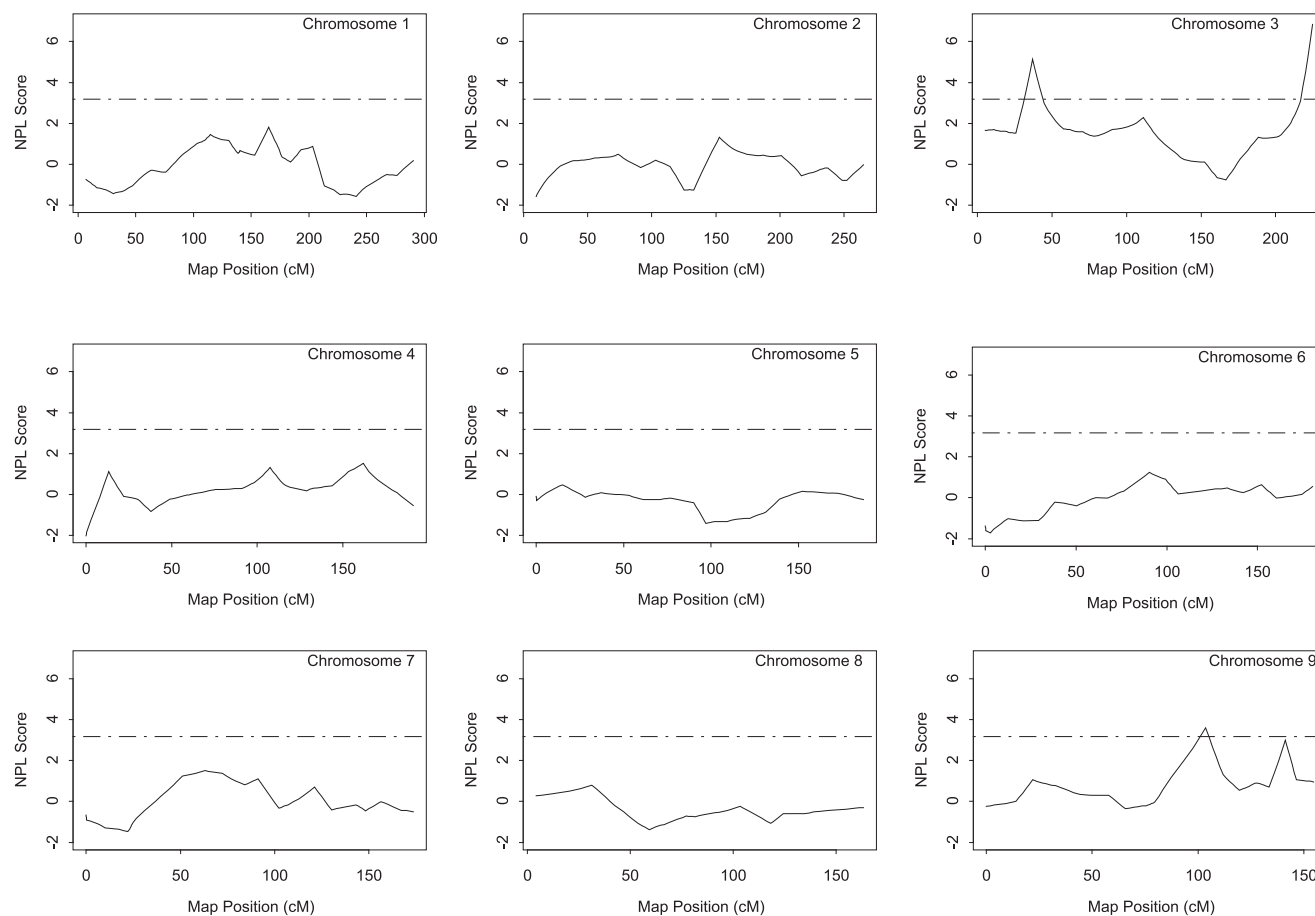
Marker	Chromosomal Location	Map Position (cM) <sup>a</sup>	Physical Position (Mb) <sup>b</sup>	Two Point NPL	Multipoint NPL
D3S2403	3p25.1	37.2	13.17	<b>5.49</b>	<b>5.14</b>
MFD427-AAA	3q29	224.0	197.65	<b>6.54</b>	<b>6.84</b>
D9S910	9q22.33	104.0	101.62	3.72	3.61
AGAT125	9q34.11	141.8	131.95	3.69	2.99
ATA78D02	17p12	32.0	13.12	3.34	2.50
D21S1446	21q22.3	57.8	48.04	3.73	3.26

<sup>a</sup>Based on Marshfield Medical Genetics marker set 11.

<sup>b</sup>Based on either UCSC Genome Browser database or Marshfield Medical Genetics marker set 11.

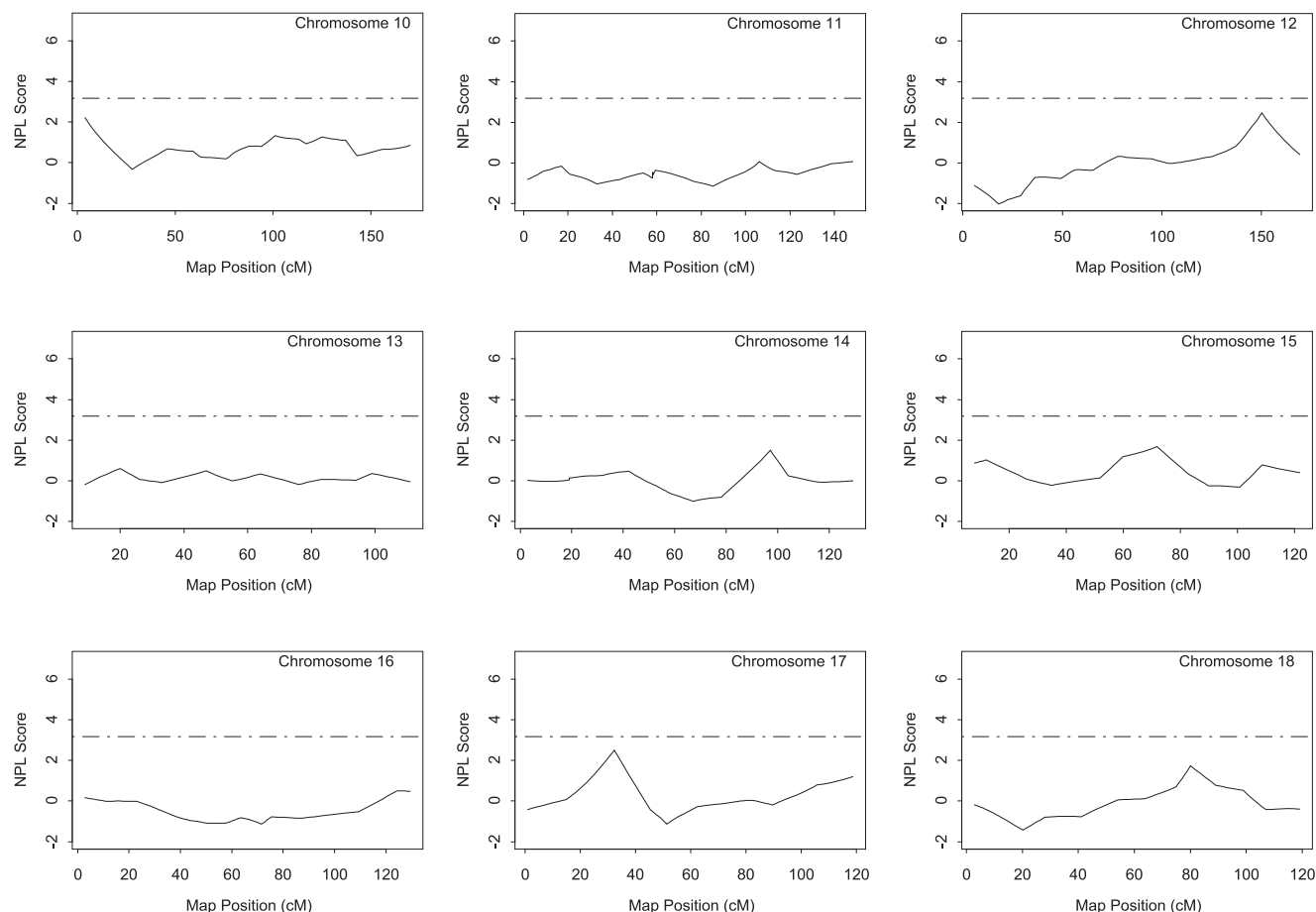
Note that a marker with a two-point or multipoint NPL score of  $>4.99$  ( $P < 3 \times 10^{-7}$ ) was considered to be linked to CAD highly significantly; A two-point or multipoint NPL score between 3.18 and 4.08 indicates suggestive linkage [38, 53].

doi:10.1371/journal.pone.0113935.t002



**Figure 1. Genome-wide NPL scan for CAD-susceptibility loci on chromosomes 1–9 in the GeneQuest families.** The vertical axis of each plot denotes NPL scores, and the horizontal axis represents map positions in cM from the telomere of the p arm to the telomere of the long arm of each chromosome. The horizontal solid line in each plot corresponds to an NPL score of 3.18 ( $P = 7.4 \times 10^{-4}$ ), the criterion for suggestive linkage.

doi:10.1371/journal.pone.0113935.g001



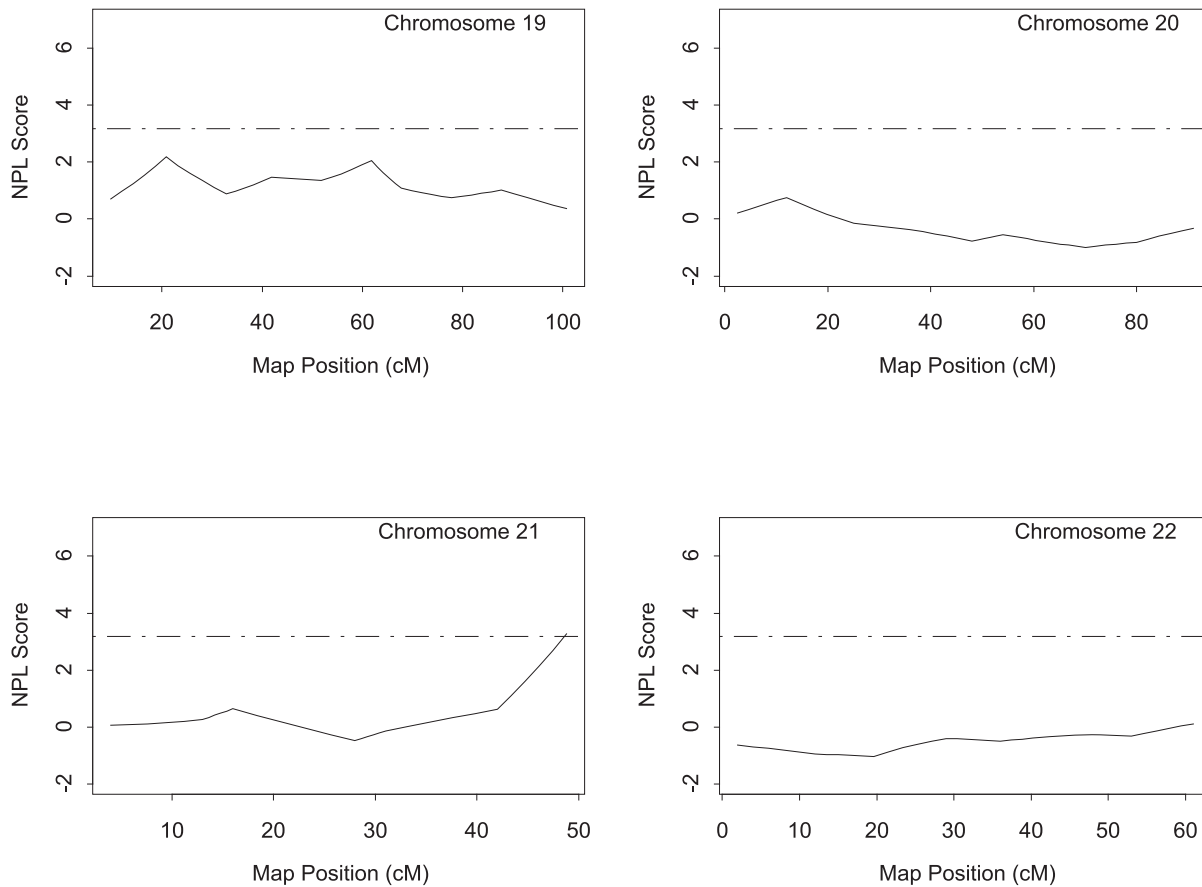
**Figure 2. Genome-wide NPL scan for CAD-susceptibility loci on chromosomes 10–18 in the GeneQuest families.**

doi:10.1371/journal.pone.0113935.g002

The remaining four chromosomal regions on 9q22.33, 9q34.11, 17p12, and 21q22.3 generated evidence of suggestive linkage for CAD with either a two-point or multipoint NPL score of  $\geq 3.18$  (Table 2, Figs. 1–3).

## Discussion

In this study, we identified a list of new genetic loci linked to CAD using the whole-genome genotyping data from the GeneQuest nuclear families. The new CAD loci include two loci showing a high level of significance on the genome-wide level (3p25.2 and 3q29; NPL score of 4.99 or above) and four loci with suggestive linkage (9q22.33, 9q34.11, 17p12, and 21q22.3; NPL score of 3.18 to 4.07) (Tables 2, Figs. 1–3). These “potentially interesting” CAD linkages are novel. These results provide a framework for future fine mapping and deep sequencing for identifying new genomic variants and new susceptibility genes associated with risk of CAD. For example, future analysis of next generation



**Figure 3. Genome-wide NPL scan for CAD-susceptibility loci on chromosomes 19–22 in the GeneQuest families.**

doi:10.1371/journal.pone.0113935.g003

sequencing data from CAD families may focus on the linkage regions first to identify genomic variants with potentially large effects on risk of familial CAD.

Using genotyping data from the same GeneQuest families, we previously identified a novel linkage for HDL-cholesterol, which spanned a 65 cM region, showed a bi-phasic pattern, and peaked at markers MFD433 and GATA131D09 on chromosome 3p25–26 [13]. The CAD locus represented by D3S2403 is on 3p25.2 and partially overlaps with the second, smaller peak of the HDL-C locus. Because a lower HDL-cholesterol level is a significant risk factor for CAD, it is possible that the same susceptibility gene may be responsible for HDL-cholesterol and CAD at the partially overlapping locus. The *PPARG* gene encoding peroxisome proliferator-activated receptor gamma is located within the CAD locus, and is a short distance of 697 kb from D3S2403. A variant in *PPARG* (rs1801282, P12A) was found to be associated with type 2 diabetes [41–43], which is a risk factor for CAD. A large-scale gene-centric meta-analysis showed that SNP rs12631819 in *PPARG* was associated with HDL-cholesterol levels [44]. The association between variants in *PPARG* and CAD was controversial because many studies reported inconsistent results. Meta-analysis showed that the H161T



variant in *PPARG* was associated with risk of CAD under a fixed-effect model, but the association became non-significant under a random-effect model; the P12A variant was not associated with CAD [45]. Future studies are needed to assess whether *PPARG* is the gene responsible for CAD at the 3p25.2 locus. Moreover, the nearest gene from the CAD marker D3S2403 is *IQSEC1*, which encodes a protein with an IQ motif and a Sec domain. Interestingly, marker MFD427-AAA with the peak NPL score for the 3q29 CAD locus is located within the *IQCG* gene encoding another protein with an IQ motif. Therefore, *IQSEC1* and *IQCG* became candidate genes for CAD for the 3p25.2 and 3q29 loci for CAD, respectively.

The 3q29 locus for CAD was supported by several other studies showing positive but not significant linkage. For instance, meta-analysis of four genome-wide linkage studies for CAD indicated that the genetic region of 3q26-27 might contain susceptibility genes for CAD ( $P=0.0001$ ) [46]. Chromosome 3q27 was also suggested to be a CAD susceptible locus in Indo-Mauritians (LOD = 2.13,  $p=0.0009$ ) [47]. The 3q29 CAD locus identified in this study may not be the same locus as the 3q26-27 locus because the two linkage peaks for the two loci are separated by 14 Mb and do not appear to overlap (Fig. 1 and Francke et al [47]). Chromosome 3q29 was found to be linked to atherogenic dyslipidemia [48]. Using the same genotyping data as in the present study, we previously identified a locus at 3q28-29 for HDL-cholesterol with LOD scores passed the threshold for suggestive significance [13]. Together, the tip of the long arm of chromosome 3 may harbor one or more novel susceptibility genes for CAD and its risk factors. In addition to the *IQCG* gene, the 3q29 CAD locus harbors several other possible candidate genes for CAD. The *LMLN* gene encodes an M8 family of metalloproteinase associated with lipid droplets and playing a role in lipid storage, generation of ROS, cell migration and invasion [49]. The *CEP19* gene encodes a 19 kDa centrosomal protein involved in obesity, glucose intolerance and insulin resistance [50]. The second interesting locus is the 9q22.33 CAD locus represented by marker D9S910. The finding of this locus was supported by careful comparison to linkage analysis in 42 families in a French Canadian population, which showed that our 9q22.33 CAD locus had a positive, but not significant, NPL score of 2.35 in the French Canadian population (Fig. 1 in [30]).

The linkages on chromosome 17 are of particular interest. A suggestive linkage for MI was previously reported on chromosome 17 at a position of 50.7 cM in a cohort with 739 affected sibpairs (LOD = 2.85). The entire putative MI linkage spanned from marker D17S921 at 17q11.2 (40 cM) to D17S787 at 17q21 (79 cM) [28], and overlaps with the 17p12 locus identified in our study (Fig. 2).

We compared our linkage results to GWAS data reported to date. Interestingly, we found that SNP rs579459 in the *ABO* gene identified as a risk variant for CAD by GWAS is located 4.20 Mb from marker AGAT125 at the 9q34.11 region showing peak linkage to CAD in this study. It is important to note that in other CAD loci identified by GWAS in this study, GWAS did not identify any risk variants for CAD. There are several potential explanations. First, GWAS signals may represent variants or susceptibility genes for familial CAD, but not for sporadic CAD as used in GWAS. Second, GWAS signals may represent low

frequency or rare variants with large effects on risk of CAD, which were generally ignored by GWAS. Considering the fact that larger meta-GWAS may not be easily practical for CAD now, re-focus on family-based studies may be the priority in the short future to identify new risk variants for CAD in our opinion.

The identification of new CAD linkage loci might have been facilitated by our strict criteria for pedigree recruitment, which focused on white families enriched with well-diagnosed early-onset CAD, and pre-empted common risk factors, such as hypercholesterolemia and insulin-dependent diabetes. Furthermore, compared with other reported genome-wide linkage scans for CAD and MI, our patient population has the youngest age at onset, <45 years in males and <50 years in females, which is expected to significantly increase the genetic component involved in CAD. Family history was another factor for us to recruit families, with aim to identify inheritable CAD cases and to exclude environmental ones. Seeking for families with strong and multi-generational CAD history remains the most accessible way of increasing the inherited component for the recruited familiar cohort [51]. Clinical studies suggest that individuals with a family history of CAD form a potential target groups for early aggressive primary prevention strategies [52].

Despite a great success in identifying new genetic loci for CAD in this study, several technical limitations associated with non-parametric linkage analysis should be well-recognized. First, the families were ascertained through probands with a strong family history of CAD, and by the unique selection strategy, which focused on white families enriched with well-diagnosed early-onset CAD. Such family recruitment may be subjected to some selection biases and some parameter estimates for CAD risk may not straightforwardly be extended to the general population. Therefore, a precaution is clearly needed for interpreting the risk estimate when translated into clinical practice. Second, due to very nature of non-parametric linkage analysis, our analysis did not model person-specific covariates to remove the effects of potentially confounding factors. Although some measures such as pre-empting common risk factors like hypercholesterolemia and insulin-dependent diabetes were taken during the family recruitment, some sporadic or environmental cases might be included in the analysis, causing identity-by-decent values between affected sib pairs downward-estimated, particular when gene-environment interactions are present. Third, due to limited sample sizes, we did not perform a stratified analysis by different subtypes of CAD and hence only genomic regions linked to general definition of CAD were identified. Fourth, this study is limited by inadequate marker density provided by microsatellite markers, fine mapping and analyses of genetic variants by use of high-density SNP markers or new-generation DNA sequencing technologies might be a better design in our future studies. Finally, there is discrepancy between our previous report of linkage to chromosome 1p34-36 [31–33] and novel linkages identified in this study, although both utilized the original 438 GeneQuest nuclear families. The discrepancy may be related to the different statistics involved in the SAGE program used in the previous report. The SAGE program in our previous study [33] analyzes all available information from the GeneQuest pedigrees, including

both affected and unaffected family members, but the Genehunter program used in the present study analyzes only affected members in GeneQuest families. Furthermore, methodological differences might explain the discrepancies. In the previous report [33], we used a Haseman-Elston regression-based method implemented in the SAGE program, which models the squared phenotypic difference of a sib pair as function of genetic variance and linkage parameter. In the present study, the Genehunter program computes the averaged identical-by-descent sharing at a marker locus between affected sibs, and then the linkage inference is made by comparing this nonparametric statistic with its expectation. Therefore, it is not surprising that the statistical power for the two methods might be different under different scenarios. Moreover, the Haseman-Elston regression-based method can take into account of epidemiological covariates, as we did in the previous report [33], while the nonparametric method suffers from this limitation. Nevertheless, affected only analysis has a merit of more accurate phenotyping as these affected individuals are often subject to a strict diagnosis. In contrast, phenotyping for normal individuals for a late onset disease like CAD can be a difficult task as the presently unaffected individuals can become affected later in their life.

In conclusion, the present study has identified two novel genetic loci for CAD that passed the threshold of highly significant and four genetic loci with suggestive evidence of linkage. Future fine mapping and deep sequencing should facilitate the identification of an underlying major (or minor) gene for this complex disorder at each locus. Identification of the CAD gene(s) should uncover the biological pathways and molecular mechanisms for the pathogenesis of CAD.

## Acknowledgments

Hanxiang Gao is a registered graduate student at Lanzhou University and also a Predoctoral Graduate Fellow at Cleveland Clinic. We wish to thank the patients and their relatives to enthusiastically participate in the study.

## Author Contributions

Conceived and designed the experiments: QKW EJT SGE QC. Performed the experiments: LL SR GS QX SC. Analyzed the data: HG LL SR QC QKW. Contributed reagents/materials/analysis tools: SR LL GS QX SC. Wrote the paper: HG SGE QC SR LL ZZ KW EJT QKW.

## References

1. **Roger VL, Go AS, Lloyd-Jones DM, Adams RJ, Berry JD, et al.** (2011) Heart disease and stroke statistics—2011 update: a report from the American Heart Association. *Circulation* 123: e18–e209.
2. **Topol EJ, Smith JC, Plow EF, Wang Q** (2006) Genetic susceptibility to myocardial infarction and coronary artery disease. *Hum Mol Genet* 15: R117–123.
3. **Wang Q** (2005) Molecular genetics of coronary artery disease. *Curr Opin Cardiol* 20: 182–188.

4. **Wang Q** (2005) Advances in the genetic basis of coronary artery disease. *Curr Atheroscler Rep* 7: 235–241.
5. **Wienke A, Holm NV, Skytthe A, Yashin AI** (2001) The heritability of mortality due to heart diseases: a correlated frailty model applied to Danish twins. *Twin Res* 4: 266–274.
6. **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.
7. **Marenberg ME, Risch N, Berkman LF, Floderus B, De Faire U** (1994) Genetic susceptibility to death from coronary heart disease in a study of twins. *N Engl J Med* 330: 1041–1046.
8. **Girelli D, Martinelli N, Peyvandi F, Olivieri O** (2009) Genetic architecture of coronary artery disease in the genome-wide era: implications for the emerging "golden dozen" loci. *Semin Thromb Hemost* 35: 671–682.
9. **Nora JJ, Lortscher RH, Spangler RD, Nora AH, Kimberling WJ** (1980) Genetic—epidemiologic study of early-onset ischemic heart disease. *Circulation* 61: 503–508.
10. **Colditz GA, Stampfer MJ, Willett WC, Rosner B, Speizer FE, et al.** (1986) A prospective study of parental history of myocardial infarction and coronary heart disease in women. *Am J Epidemiol* 123: 48–58.
11. **Colditz GA, Rimm EB, Giovannucci E, Stampfer MJ, Rosner B, et al.** (1991) A prospective study of parental history of myocardial infarction and coronary artery disease in men. *Am J Cardiol* 67: 933–938.
12. **Schildkraut JM, Myers RH, Cupples LA, Kiely DK, Kannel WB** (1989) Coronary risk associated with age and sex of parental heart disease in the Framingham Study. *Am J Cardiol* 64: 555–559.
13. **Yang R, Li L, Seidemann SB, Shen GQ, Sharma S, et al.** (2010) A genome-wide linkage scan identifies multiple quantitative trait loci for HDL-cholesterol levels in families with premature CAD and MI. *J Lipid Res* 51: 1442–1451.
14. **Erdmann J, Grosshennig A, Braund PS, Konig IR, Hengstenberg C, et al.** (2009) New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat Genet* 41: 280–282.
15. **Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, et al.** (2007) A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* 316: 1491–1493.
16. **Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissino D, et al.** (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 41: 334–341.
17. **McPherson R, Pertsemlidis A, Kavassari N, Stewart A, Roberts R, et al.** (2007) A common allele on chromosome 9 associated with coronary heart disease. *Science* 316: 1488–1491.
18. **Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, et al.** (2007) Genomewide association analysis of coronary artery disease. *N Engl J Med* 357: 443–453.
19. **Samani NJ, Deloukas P, Erdmann J, Hengstenberg C, Kuulasmaa K, et al.** (2009) Large scale association analysis of novel genetic loci for coronary artery disease. *Arterioscler Thromb Vasc Biol* 29: 774–780.
20. **Tregouet DA, Konig IR, Erdmann J, Munteanu A, Braund PS, et al.** (2009) Genome-wide haplotype association study identifies the SLC22A3-LPAL2-LPA gene cluster as a risk locus for coronary artery disease. *Nat Genet* 41: 283–285.
21. **Coronary Artery Disease (CAD) Genetics Consortium** (2011) A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet* 43: 339–344.
22. **CARDIoGRAM Consortium** (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 43: 333–338.
23. **Wang F, Xu CQ, He Q, Cai JP, Li XC, et al.** (2011) Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nat Genet* 43: 345–349.
24. **Lette G, Palmer CD, Young T, Ejebe KG, Allayee H, et al.** (2011) Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARE Project. *PLoS Genet* 7: e1001300.

25. **Reilly MP, Li M, He J, Ferguson JF, Stylianou IM, et al.** (2011) Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* 377: 383–392.
26. **Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al.** (2013) Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* 45: 25–33.
27. **Ouwehand WH, Bloodomics and Cardiogenics Consortia** (2009) The discovery of genes implicated in myocardial infarction. *J Thromb Haemost (Suppl 1)*: 305–307.
28. **Farrall M, Green FR, Peden JF, Olsson PG, Clarke R, et al.** (2006) Genome-wide mapping of susceptibility to coronary artery disease identifies a novel replicated locus on chromosome 17. *PLoS Genet* 2: e72.
29. **Samani NJ, Burton P, Mangino M, Ball SG, Balmforth AJ, et al.** (2005) A genomewide linkage study of 1,933 families affected by premature coronary artery disease: The British Heart Foundation (BHF) Family Heart Study. *Am J Hum Genet* 77: 1011–1020.
30. **Engert JC, Lemire M, Faith J, Brisson D, Fujiwara TM, et al.** (2008) Identification of a chromosome 8p locus for early-onset coronary heart disease in a French Canadian population. *Eur J Hum Genet* 16: 105–114.
31. **Shen GQ, Li L, Girelli D, Seidelmann SB, Rao S, et al.** (2007) An LRP8 variant is associated with familial and premature coronary artery disease and myocardial infarction. *Am J Hum Genet* 81: 780–791.
32. **Shen GQ, Girelli D, Li L, Olivieri O, Martinelli N, et al.** (2013) Multi-allelic haplotype association identifies novel information different from single-SNP analysis: a new protective haplotype in the LRP8 gene is against familial and early-onset CAD and MI. *Gene* 521: 78–81.
33. **Wang Q, Rao S, Shen GQ, Li L, Moliterno DJ, et al.** (2004) Premature Myocardial Infarction Novel Susceptibility Locus on Chromosome 1P34–36 Identified by Genomewide Linkage Analysis. *Am J Hum Genet* 74: 262–271.
34. **Weber JL, Broman KW** (2001) Genotyping for human whole-genome scans: past, present, and future. *Adv Genet* 42: 77–96.
35. **Seidelmann SB, Li L, Shen GQ, Topol EJ, Wang QK** (2008) Identification of a novel locus for triglyceride on chromosome 1p31-32 in families with premature CAD and MI. *J Lipid Res* 49: 1034–1038.
36. **S.A.G.E.** (2006) *Statistical Analysis for Genetic Epidemiology* 5.3. Cork, Ireland.
37. **Olson JM** (1999) Relationship estimation by Markov-process models in a sib-pair linkage study. *Am J Hum Genet* 64: 1464–1472.
38. **Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES** (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58: 1347–1363.
39. **Altmuller J, Palmer LJ, Fischer G, Scherb H, Wjst M** (2001) Genomewide scans of complex human diseases: true linkage is hard to find. *Am J Hum Genet* 69: 936–950.
40. **Lander E, Kruglyak L** (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results [see comments]. *Nat Genet* 11: 241–247.
41. **Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, et al.** (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316: 1341–1345.
42. **Deeb SS, Fajas L, Nemoto M, Pihlajamaki J, Mykkanen L, et al.** (1998) A Pro12Ala substitution in PPARgamma2 associated with decreased receptor activity, lower body mass index and improved insulin sensitivity. *Nat Genet* 20: 284–287.
43. **Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, et al.** (2000) The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26: 76–80.
44. **Asselbergs FW, Guo Y, van Iperen EP, Sivapalaratnam S, Tragante V, et al.** (2012) Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *Am J Hum Genet* 91: 823–838.
45. **Ding S, Liu L, Zhuge QC, Yu Z, Zhang X, et al.** (2012) The meta-analysis of the association of PPARG P12A, C161T polymorphism and coronary heart disease. *Wien Klin Wochenschr* 124: 671–677.

46. **Chiodini BD, Lewis CM** (2003) Meta-analysis of 4 coronary heart disease genome-wide linkage studies confirms a susceptibility locus on chromosome 3q. *Arterioscler Thromb Vasc Biol* 23: 1863–1868.
47. **Francke S, Manraj M, Lacquemant C, Lecoœur C, Lepretre F, et al.** (2001) A genome-wide scan for coronary heart disease suggests in Indo- Mauritians a susceptibility locus on chromosome 16p13 and replicates linkage with the metabolic syndrome on 3q27. *Hum Mol Genet* 10: 2751–2765.
48. **Yip AG, Ma Q, Wilcox M, Panhuysen CI, Farrell J, et al.** (2003) Search for genetic factors predisposing to atherogenic dyslipidemia. *BMC Genet* 4 Suppl 1: S100.
49. **Di CF, Duca E, Dunbar DR, Cagney G, Heck MM** (2013) Invadolysin, a conserved lipid-droplet-associated metalloproteinase, is required for mitochondrial function in *Drosophila*. *J Cell Sci* 126: 4769–4781.
50. **Shalata A, Ramirez MC, Desnick RJ, Friedigkeit N, Buettner C, et al.** (2013) Morbid obesity resulting from inactivation of the ciliary protein CEP19 in humans and mice. *Am J Hum Genet* 93: 1061–1071.
51. **Banerjee A** (2012) A review of family history of cardiovascular disease: risk factor and research tool. *Int J Clin Pract* 66: 536–543.
52. **Pandey AK, Blaha MJ, Sharma K, Rivera J, Budoff MJ, et al.** (2014) Family history of coronary heart disease and the incidence and progression of coronary artery calcification: Multi-Ethnic Study of Atherosclerosis (MESA). *Atherosclerosis* 232: 369–376.
53. **Kruglyak L** (1996) Thresholds and sample sizes. *Nat Genet* 14: 132–133.