

RESEARCH ARTICLE

# Genetics of Coronary Artery Disease in Taiwan: A CardiometaboChip Study by the Taichi Consortium

Themistocles L. Assimes<sup>1</sup>\*, I.-T. Lee<sup>2</sup>\*, Jyh-Ming Juang<sup>3</sup>, Xiuqing Guo<sup>4</sup>, Tzung-Dau Wang<sup>3</sup>, Eric T. Kim<sup>4</sup>, Wen-Jane Lee<sup>5</sup>, Devin Absher<sup>6</sup>, Yen-Feng Chiu<sup>7</sup>, Chih-Cheng Hsu<sup>7</sup>, Lee-Ming Chuang<sup>8</sup>, Thomas Quertermous<sup>1</sup>, Chao A. Hsiung<sup>6</sup>, Jerome I. Rotter<sup>4</sup>, Wayne H.-H. Sheu<sup>2</sup>, Yii-Der Ida Chen<sup>4</sup>, Kent D. Taylor\*<sup>4</sup>



**1** Department of Medicine, Stanford University School of Medicine, Stanford, California, United States of America, **2** Division of Endocrine and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan, **3** Cardiovascular Center and Division of Cardiology, Department of Internal Medicine, National Taiwan University Hospital, National University College of Medicine, Taipei, Taiwan, **4** Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, and Department of Pediatrics, University of California Los Angeles, Los Angeles, California, United States of America, **5** Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan, **6** HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, **7** Institute of Population Health Sciences, Division of Biostatistics and Bioinformatics, National Health Research Institutes, Zhunan Town, Miaoli County, Taiwan, **8** Division of Endocrine and Metabolism, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

\* These authors contributed equally to this work.  
\* [ktaylor@labiomed.org](mailto:ktaylor@labiomed.org)

OPEN ACCESS

**Citation:** Assimes TL, Lee I-T, Juang J-M, Guo X, Wang T-D, Kim ET, et al. (2016) Genetics of Coronary Artery Disease in Taiwan: A CardiometaboChip Study by the Taichi Consortium. PLoS ONE 11(3): e0138014. doi:10.1371/journal.pone.0138014

**Editor:** Kelvin Yuen Kwong Chan, Hospital Authority, CHINA

**Received:** April 20, 2015

**Accepted:** August 24, 2015

**Published:** March 16, 2016

**Copyright:** © 2016 Assimes 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 Statement:** Data are held on dbGAP. Once the dbGAP provides identification for the dataset, the authors will post a public comment on the article. In the meantime, please contact the corresponding author for data requests.

**Funding:** The TAICHI study was supported by a grant from NHLBI grant HL087647. TLA is supported by an NIH career development award K23DK088942. The Taiwan Dragon was supported by grants from National Science Council (NSC 98-2314-B-075A-002-MY3) and Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1013001C; TCVGH-

## Abstract

By means of a combination of genome-wide and follow-up studies, recent large-scale association studies of populations of European descent have now identified over 46 loci associated with coronary artery disease (CAD). As part of the TAICHI Consortium, we have collected and genotyped 8556 subjects from Taiwan, comprising 5423 controls and 3133 cases with coronary artery disease, for 9087 CAD SNPs using the CardioMetaboChip. We applied penalized logistic regression to ascertain the top SNPs that contribute together to CAD susceptibility in Taiwan. We observed that the 9p21 locus contributes to CAD at the level of genome-wide significance (rs1537372, with the presence of C, the major allele, the effect estimate is -0.216, standard error 0.033, p value  $5.8 \times 10^{-10}$ ). In contrast to a previous report, we propose that the 9p21 locus is a single genetic contribution to CAD in Taiwan because: 1) the penalized logistic regression and the follow-up conditional analysis suggested that rs1537372 accounts for all of the CAD association in 9p21, and 2) the high linkage disequilibrium observed for all associated SNPs in 9p21. We also observed evidence for the following loci at a false discovery rate >5%: SH2B3, ADAMTS7, PHACTR1, GGCX, HTRA1, COL4A1, and LARP6-LRRC49. We also took advantage of the fact that penalized methods are an efficient approach to search for gene-by-gene interactions, and observed that two-way interactions between the PHACTR1 and ADAMTS7 loci and between the SH2B3 and COL4A1 loci contribute to CAD risk. Both the similarities and differences between the significance of these loci when compared with significance of loci in studies of

1013002D). The TCAD was supported by grants from National Science Council (NSC 98-2314-B-075A-002-MY3) and Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1013001C;TCVGH-1013002D). The SAPPHIRe was supported by grants from National Health Research Institutes (BS-094-PP-01 & PH-100-PP-03). HALST was supported by grants from National Health Research Institutes (PH-100-SP-01). The funders had no role in study design, data collection and analysis, decision to publish or preparation of manuscript.

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

populations of European descent underscore the fact that further genetic association of studies in additional populations will provide clues to identify the genetic architecture of CAD across all populations worldwide.

## Introduction

Recent large-scale association studies of populations of European descent have now identified over 46 loci associated with coronary artery disease (CAD)[1–3]. In order to achieve this result, the CARDIoGRAM consortium performed meta-analyses on numerous genome-wide association datasets with over 22,000 CAD cases and over 64,000 controls, with subsequent meta-analyses of follow-up studies with over 63,000 CAD cases and 130,000 controls employing the Illumina CardioMetaboChip. The CARDIoGRAM Consortium contributed the top loci from their genome-wide analyses to the design of this chip and, along with other consortia studying cardio-metabolic traits, were able to achieve these large numbers because of the “economy of scale” made possible by the price reduction available due to the large number of chips manufactured for this combined effort [4]. Since the initial genome-wide association studies of CAD in 2007 [5, 6], the greatest contribution to CAD remains the 9p21 locus. The features of this locus include the CDKN2A and CDKN2B genes, a CDKN2B-antisense RNA known as ANRIL, microRNA 384, and numerous binding sites for transcription factors as identified by ENCODE [7]. While a recent sequencing study of the 9p21 region in the Framingham Cohort demonstrated associations with higher risk for myocardial infarction, higher coronary calcium scores, and larger aortic diameters, no obviously functional polymorphisms were found; however, variants were associated with gene expression of a short form of ANRIL and for the gene CDK2NB [8].

With respect to populations of Chinese descent, a meta-analysis of two genome-wide association studies of Han subjects from Beijing by the CARDIoGRAM consortium (over 1500 cases and over 5000 controls), followed by a replication study (over 15,000 cases and over 11,000 controls), replicated four of the European loci, 9p21, PHACTR1, TCF21, and C12orf51 (now known as HECTD4), and identified four additional four loci, TTC32-WDR35, GUCY1A3, the MHC, and ATP2B1 [9]. This study proposed that there were two association signals in the 9p21 locus because: 1) a conditional analysis demonstrating additional association with several SNPs after controlling for the main effect, and 2) differences in the linkage disequilibrium pattern of these associated SNPs were observed between their samples and samples of European descent from the International HapMap Project.[10]

As amply demonstrated by these and other studies of non-Mendelian or complex genetic traits, the genome-wide association study represents one of the most powerful tools for the identification of the genetic contribution of individual SNPs to specific phenotypes. Most of these studies have been limited to analyses of the association of phenotype with individual genetic loci or single-nucleotide polymorphisms (SNPs), taken one at a time. This is because applying multivariate regression to such a large dataset, with the goals of investigating the contribution of all SNPs to phenotype and of constructing models of the genetic architecture of the phenotype, becomes computationally impossible with the large number of SNPs to be considered; the requisite matrix inversion procedures or linear equations scale with the number of SNPs to the power of three [11]. One means to circumvent these problems is to employ the use of a regression penalty that shrinks most of the coefficients (“betas”) to zero during the selection of the model. Such a penalty may be estimated by cross-validation methods. Furthermore, penalized methods make it possible to explore SNP by SNP interactions and so may provide clues to the genetic architecture of the phenotype [12].

For the study of the genetics of CAD in Taiwan, we took advantage of the fact that the CAD SNPs identified in populations of European descent were incorporated into the design of the Illumina CardioMetaboChip [4]. This design provided a cost-effective means to genotype known CAD SNPs. Through our academic collaboration, we have collected phenotypic data and DNA from a large cohort of subjects with coronary artery disease and controls without disease across the island of Taiwan. We genotyped this cohort with the CardioMetaboChip and applied the method of penalized logistic regression; using this approach, we were able to identify gene-by-gene interactions contributing to CAD in this population.

## Methods

### Subjects

The TAICHI Consortium is a collaborative study in Taiwan, with the aims to genotype a large number of Taiwanese Chinese subjects with the Illumina CardioMetaboChip and to identify genetic determinants of atherosclerosis- and metabolic-related traits [13–20]. Academic centers participating include Taichung Veteran’s General Hospital, Tri-Service General Hospital and the National Taiwan University Hospital, and the National Health Research Institute in Taiwan for subject ascertainment and phenotyping. A total of 8556 subjects, 3133 cases with coronary artery disease, 5423 controls without disease, were recruited in this report (defined below; see also Table 1).

Institutional Review Boards at all participating centers have approved this study: Taichung Veteran’s General Hospital, Tri-Service General Hospital, the National Taiwan University Hospital, and the National Health Research Institute of Taiwan. Written informed consent was obtained from each participant and is on file in Taiwan. All data and samples that are transferred to the Los Angeles Biomedical Research Institute in the US are identified by codes so that all personal identifying information is kept in Taiwan. The Institutional Review Board (IRB) of the Los Angeles Biomedical Research Institute has approved collection and handling of data and samples from Taiwan as well as molecular studies of the samples including genotyping, all data analyses, and publication of results. The IRB’s of the Hudson-Alpha Research Institute and Stanford University have given approval to this project as exempt from the regulations at 45 CFR 46 or 21 CFR56.

### Phenotype

**CAD.** Patients admitted to the several Hospital Cardiovascular Departments were enrolled. Subjects were defined as CAD positive with one of the following criteria: 1) history of

**Table 1. Characteristics of Subjects.**

	No CAD		CAD	
	Male	Female	Male	Female
Number of subjects	2728	2695	2325	808
Age (mean, yr)	66	65	65	67
BMI (mean, kg/m <sup>2</sup> )	24.7	24.8	25.6	25.5
SBP (mean, mmHg)	128	130	130	134
DBP (mean, mmHg)	74	72	75	74
Total cholesterol (mean, mg/dl)	185	197	179	190
LDL-cholesterol (mean, mg/dl)	114	118	107	111
HDL-cholesterol (mean, mg/dl)	45	52	42	45
Triglycerides (mean, mg/dl)	143	141	149	172

Age, BMI, blood pressure, and serum lipids are significantly different between case/control status and between gender at p<0.001 by analysis of variance.

doi:10.1371/journal.pone.0138014.t001

myocardial infarction; 2) history of coronary artery bypass graft (CABG) or percutaneous coronary angioplasty (PTCA) and or stenting; or 3) at least 1 major vessel with stenosis  $\geq 50\%$  demonstrated by angiography. Age was defined as age of event, graft or angioplasty, or angiography.

**Controls.** Subjects were enrolled from an epidemiology study and from the several Hospital Endocrinology and Metabolism Departments either as outpatients or as their family members. Subjects with history of CAD were excluded. Age was defined as age of recruitment.

## Genotypes

**Chip.** Genotyping was performed using the Illumina CardioMetaboChip v1.0 (Illumina, San Diego, CA) following the manufacturer's protocol [21, 22] in a total of 11,582 subjects using DNA purified by column from whole blood (Qiagen, Valencia, CA). The chip has been described elsewhere [4]. Calling of genotypes was performed using the Illumina Genome Studio software and both samples and SNPs were removed with call rates less than 98%.

**Quality Control.** Because the set of subjects includes family members, subjects were removed for relatedness and cryptic relatedness with PI-Hat values  $>0.125$ . Population outliers were defined as those subjects with eigenvalues from principal components 1, 2, and 3 greater than 3 standard deviations from the mean and were removed after calculating principal components using standard methods [23]. Only subjects with age between 40 and 90 years were then retained. Additional SNPs were removed if the minor allele frequency  $< 5\%$  or if the *p*-value from the Hardy-Weinberg test of allele frequencies in controls  $<0.05$ . The results from 8556 subjects are therefore included in this report.

**CAD SNPs.** "CAD SNPs" were selected from the 23 fine-mapping regions and GWAS hits contributed by the Cardiogram Consortium to the original design of the CardioMetaboChip [4]. This list and other CardioMetaboChip annotation files are available from the Abecasis group at <http://csg.sph.umich/kang/MetaboChip>. The results from 9087 SNPs are reported here.

## Statistical Analyses

**Penalized logistic regression.** Penalized logistic regression was employed using the software MENDEL v14 [24] available from <http://www.genetics.ucla.edu/software/mendel> with age, gender, and eigenvalues from principal components 1 and 2 added as covariates. As applied to this study, lasso penalized regression is a method for continuous model selection on all of the CAD SNPs and covariates, together and analyzed at the same time, and is of particular value when the number of SNPs exceeds the number of subjects, as in the standard GWAS design.[11] The order that each SNP is added to the statistical model is correlated with its individual significance, and the lasso method eliminates SNPs that enter with low explanatory power. The result of a lasso penalized regression of this dataset may be considered to have studied the association of all CAD SNPs on the CardioMetaboChip.

Two-way interactions were also explored by penalized lasso logistic regression within the top 100 associated SNPs in order to reduce the number of interactions to be computed using the Mendel software.

**Other tests.** As implemented in R software, the Fisher exact test, calculation of Odds Ratios, logistic regression, and the Cochran-Mantel-Haenszel test were used to explore the effects of gender in the association of CAD, both within and across gender, and across genotypes [25, 26]. Conditional analysis was performed using PLINK v1.9 [27, 28] with covariates as listed in the legends to the tables. "Locus Zoom" plots were drawn using the website at the University of Michigan [29]. Linkage disequilibrium was assessed using Haploview v4.[30]

**Table 2. Association of SNPs with Coronary Artery Disease by Penalized Logistic Regression.**

SNP ID	Locus	Chr	Position	P Value	Effect	Standard Error	Major Allele	Allele Freq. Cases	Allele Freq. Controls
rs1537372	9p21	9	22,093,183	5.7x10 <sup>-10</sup>	-0.216	0.033	C	0.50	0.54
rs79105258	SH2B3	12	110,202,614	2.5x10 <sup>-5</sup>	-0.140	0.036	C	0.70	0.72
rs79265682	ADAMTS7	15	76,829,811	3.8x10 <sup>-5</sup>	0.138	0.034	G	0.62	0.58
rs9349379	PHACTR1	6	13,011,943	4.4x10 <sup>-5</sup>	0.153	0.037	G	0.73	0.70
rs6738645	GGCX	2	85,636,639	6.3x10 <sup>-5</sup>	-0.130	0.034	A	0.60	0.63
rs2268344	HTRA1	10	124,234,959	0.00012	-0.129	0.034	A	0.58	0.61
rs2289800	COL4A1	13	109,622,824	0.00019	-0.132	0.036	G	0.68	0.71
rs11072221	LARP6	15	68,919,620	0.00023	-0.121	0.033	C	0.50	0.53

Effect estimates in null model: Grand Mean, -0.68; Age, 0.049; Gender, -0.52. Eigenvalues from Principal Components 1, 2, and 3 were included as covariates. False discovery rate < 5% corresponds to a p-value of 0.0004. Positions of SNPs are GRCh37/hg19.

doi:10.1371/journal.pone.0138014.t002

Control data from the Thousand Genomes Project were used as indicated.[31] Power for detecting an association was computed using QUANTO (available from <http://biostats.usc.edu/Quanto.html>).

## Results

### Dataset

Herein we report observations from 8556 Taiwan subjects (3133 cases and 5423 controls) and 9087 coronary artery disease (CAD) SNPs genotyped using the Illumina CardioMetaboChip.

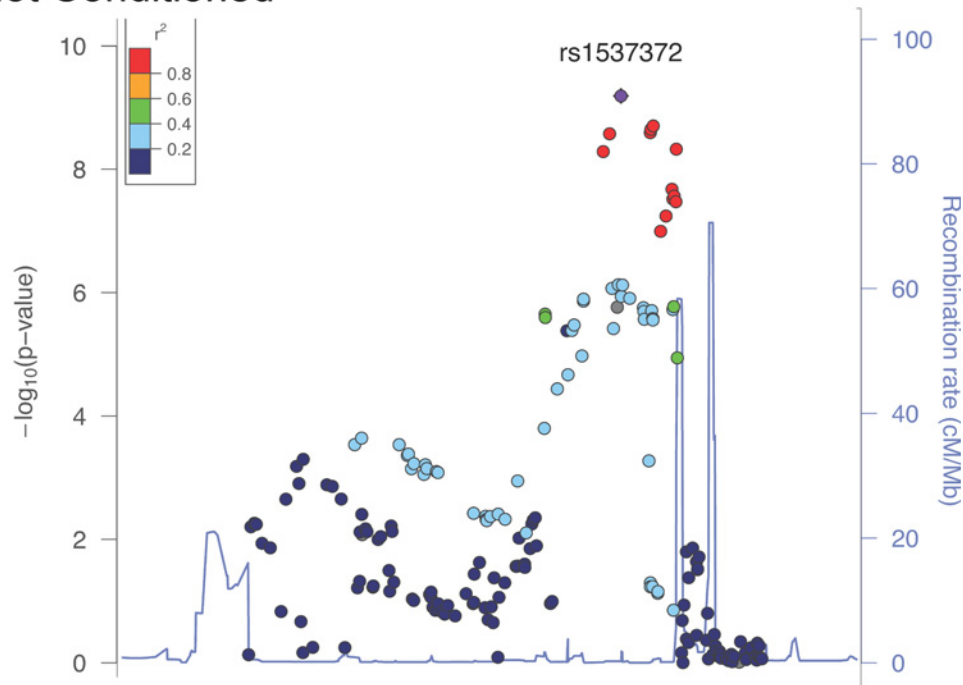
Characteristics of the subjects by gender and by case/control status are given in Table 1. As expected, cases and controls were significantly different for body mass index (BMI), serum lipids, and blood pressure (p anova <0.001).

### Association Study

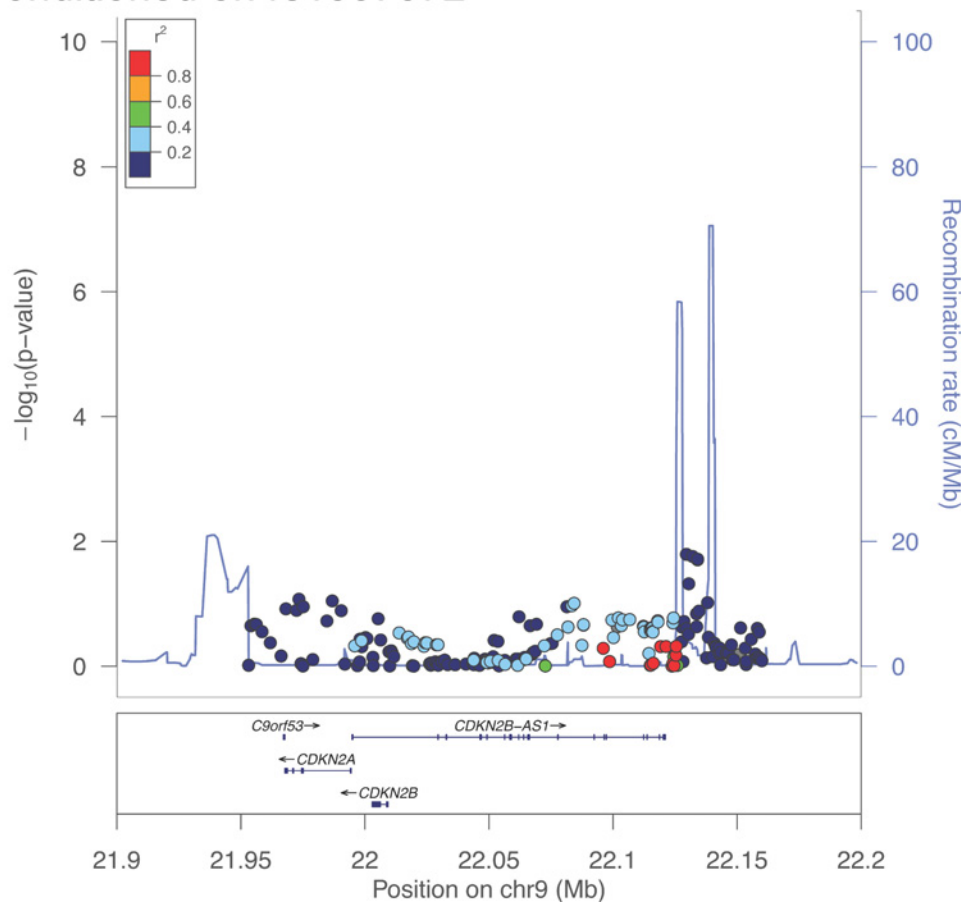
**Penalized logistic regression.** The 9087 coronary artery disease (CAD) SNPs identified by the CARDIoGRAM Consortium for populations of European descent were tested for association with CAD in Taiwan using a penalized logistic regression method. This method allows consideration of the contribution of all SNPs together and is appropriate for a dataset with more SNPs than subjects. With age, gender, and the eigenvalues from principal components 1 and 2 as covariates, the p-values and effect estimates are given in Table 2. Loci associated with CAD with a significance less than a false discovery rate of 5% were: 9p21, SH2B3, ADAMTS7, PHACTR1, GGCX, HTRA1, COL4A1, and LARP6-LRRC49. Furthermore, the 9p21 locus was significantly associated with CAD by the “genome-wide” criteria of 5x10<sup>-8</sup> (rs1537372, with the presence of C, the major allele, the effect estimate is -0.216, standard error 0.033, p value 5.8x10<sup>-10</sup>).

**Conditional analysis.** As a check of the result of the penalized logistic regression, we investigated the fact that only a single SNP in the 9p21 region was retained. A “conventional,” or single-SNP logistic regression of the 9p21 region was conducted with age, gender, and the eigenvalues from principal components 1 and 2 as covariates; the results without and with conditioning on the 9p21 SNP that was retained by the penalized logistic regression, rs1537372, are compared in Fig 1 (“Locus Zoom” plot; without rs1537372 as a covariate, Fig 1A; with rs1537372 as an additional covariate, Fig 1B). No additional association was detectable in the 9p21 region when conditioning on rs1537372, thus providing additional support for the retention of one 9p21 SNP in the penalized logistic regression. Furthermore, SNPs with significant

a) Not Conditioned



b) Conditioned on rs1537372



**Fig 1. Conditional Analysis of the 9p21 Region.** “Locus Zoom” plots [29] of the 9p21 region demonstrating conditional analysis of the top SNP rs1537372 using logistic regression adjusted for age, gender, and the eigenvalues from principal components 1 and 2. Linkage disequilibrium of SNPs with rs1537372 is shown by the color of the points and is based on data from the Asian subjects in the Thousand Genomes Project. a) Association of SNP with coronary disease; b) Association of SNP with coronary disease conditioned on rs1537372 (rs1537372 added as an additional covariate).

doi:10.1371/journal.pone.0138014.g001

associations in this conventional logistic regression are in high linkage disequilibrium with each other as determined in data from Asian subjects in the Thousand Genomes Project (Fig 1, color of data points; detail in Fig).

In order to place our work within the context of the previous study of CAD in a Chinese population,[9] we show a detail of this conditional analysis for the previous reported SNPs rs10757274, rs1333049, rs9632884, and rs1333042 (S1 Table). The association of each of these four SNPs is abolished by conditioning on the top SNP in this report, rs1537372. We also calculated the power for an association with 3130 cases and 5423 controls at a significance level of 0.05 using the program QUANTO. At the observed minor allele frequency (MAF) of rs1333042 (0.32) and of rs10757274 (0.48), we have 80% or 85% power, respectively, to detect an association with a relative risk of 1.10, and at the MAF of rs9632884 (0.27), we have 76% power to detect an association with a relative risk of 1.10. Thus, we had sufficient power to detect additional associations in 9p21 if such had existed in our dataset.

### Interaction Study

Since penalized methods are also efficient for exploring two-way interactions, we explored possible interactions among the 100 SNPs with the highest association with CAD (Table 3). Please note that this is an analysis separate and secondary to the analysis presented in Table 2. Differences in p values between the two tables result from the absence or presence of two-way interactions. The result suggested the presence of SNP by SNP interactions between PHACTR1 and ADAMTS7, and between COL4A1 and SH2B3.

These interactions were studied further using frequency tables (Tables 4 and 5). For the interaction between PHACTR1 and ADAMTS7, the Odds Ratio for CAD was highest for subjects with both the PHACTR1 GG genotype and the ADAMTS7 GG or GA genotype (Table 4). A Cochran-Mantel-Haenszel test across all genotype combinations confirmed the interaction between these two loci that was observed in the penalized logistic regression analysis (CMH p value = 0.00015). The highest Odds Ratio for CAD was therefore observed in those subjects homozygous for the major allele at both loci. For the interaction between SH2B3 and COL4A1, the Odds Ratio for CAD was highest for subjects with both the SH2B3 AA genotype and the COL4A1 GG genotype (Table 5; Cochran-Mantel-Haenszel test p value = 0.00047). The

**Table 3. Possible SNP by SNP Interactions Contributing to Coronary Artery Disease in Taiwan by Penalized Logistic Regression.**

Individual SNP or SNP x SNP Interaction	P Value	Effect Estimate	SNPs
9p21	8.5x10 <sup>-10</sup>	-0.208	rs1537372
PHACTR1 x ADAMTS7	1.9x10 <sup>-7</sup>	-0.104	rs9349379      rs79265862
COL4A1 x SH2B3	2.5x10 <sup>-7</sup>	-0.137	rs2289800      rs79105258
PHACTR1	2.8x10 <sup>-5</sup>	0.131	rs9349379
ADAMTS7	3.0x10 <sup>-5</sup>	0.055	rs79265862
GGCX1	5.4x10 <sup>-5</sup>	-0.128	rs6738645
HTRA1	0.00015	-0.122	rs2268344

Effect estimates in null model and covariates included in the analysis, are the same as in legend to Table 1.

doi:10.1371/journal.pone.0138014.t003

**Table 4. Interaction of PHACTR1 and ADAMTS7.**

PHACTR1 rs9349379 Genotype	ADAMTS7 rs79265862 Genotype														
	GG					GA					AA				
	Control		Case		OR	Control		Case		OR	Control		Case		OR
	N*	freq	N	freq		N	freq.	N	freq		N	freq	N	freq	
GG	898	0.487	654	0.550	1.67	1285	0.493	753	0.509	1.42	467	0.481	228	0.494	0.952
GA	759	0.411	454	0.382	1.22	1070	0.410	623	0.421	1.00	425	0.438	195	0.421	1.064
AA	188	0.102	82	0.069	**	253	0.097	104	0.070	**	78	0.080	40	0.086	**
Fisher Test for 3x3	2.7 x 10 <sup>-4</sup>					0.014					ns				

\* N, number; freq, frequency

\*\* assigned OR = 1

ns: Not significant at p<0.05.

Cochran-Mantel-Haenszel test across the three genotypes, p = 0.00015.

doi:10.1371/journal.pone.0138014.t004

highest Odds Ratio for CAD was therefore seen in those subjects homozygous for the minor allele for the SH2B3 locus and the major allele for the COL4A1 locus.

### Discussion

In this study we applied penalized logistic regression to a dataset composed of Taiwanese CAD subjects and controls genotyped for CAD SNPs identified by the CARDIoGRAM Consortium and included in the design of the Illumina CardioMetaboChip. We observed that the major genetic contribution to CAD in the Taiwanese is represented by a single association at the 9p21 locus. This association was of “genome-wide” significance at  $<5 \times 10^{-8}$  in our dataset (Table 2). Thus, the strongest genetic signal for CAD in Taiwanese is very similar or identical to the strongest GWAS signal observed in populations of European descent [1] and to the signal identified by a candidate gene approach [15]. The conclusion that this association represents a single genetic contribution was suggested by our observations that: 1) the penalized logistic regression retained only one SNP from this locus (Table 2), 2) SNPs with significant associations in a conventional logistic regression analysis were in high linkage disequilibrium with

**Table 5. Interaction of SH2B3 and COL4A1.**

SH2B3 rs79105258 Genotype	COL4A1 rs2289800 Genotype														
	GG					GA					AA				
	Control		Case		OR	Control		Case		OR	Case	Control		OR	
	N*	Freq	N	Freq		N	Freq	N	Freq		N	Freq	N	freq	
CC	1379	0.508	660	0.454	**	1109	0.497	631	0.469	**	208	0.438	149	0.449	**
CA	1128	0.415	631	0.434	1.17	912	0.409	573	0.426	1.10	215	0.453	150	0.458	0.99
AA	210	0.077	164	0.113	1.61	210	0.094	142	0.105	1.18	52	0.109	33	0.099	0.88
Fisher Test for 3x3	6.9x10 <sup>-5</sup>					ns					ns				

\* N, number; freq, frequency

\*\* assigned OR = 1

ns: Not significant at p<0.05.

Cochran-Mantel-Haenszel test across the three genotypes, p = 0.00047.

doi:10.1371/journal.pone.0138014.t005



each other (Fig 1A and S1 Fig), and 3) conditioning on this single SNP reduced all other associations in this region to non-significance (Fig 1B), and 4) we had sufficient power to detect additional association if such had existed in our dataset.

Our observations supporting a single genetic contribution to CAD susceptibility in the 9p21 region contrast with the previous report of a CAD GWAS in a Chinese population [9]. That study ascertained subjects from Beijing and from northern provinces in China and reported: 1) a difference in linkage disequilibrium (LD) structure between their sample and populations of European descent, and 2) a conditional analysis supporting two independent genetic signals from SNPs rs1333042 and rs10757274. Our dataset contains, in addition to our top SNP, rs1537372, high quality genotypes for rs1333042 and rs10757274, plus two others mentioned in the previous report, rs1333049 and rs9632884. We observe all five SNPs to be in high linkage disequilibrium with each other with an  $r^2$  value  $> 0.87$  in data from Asian subjects in the Thousand Genomes Project (Panel A in S1 Fig), in data from European subjects in the Thousand Genomes Project (Panel B in S1 Fig) as well as in our own data from Taiwan (Panel C in S1 Fig). We therefore do not observe the differences in linkage disequilibrium structure reported previously. Furthermore, we observed no additional association upon conditioning on our top SNP, rs1537372 (S1 Table). We therefore think that the contrast between our results and those of Lu and co-workers are due to differences in the populations studied: the previous work combined subjects from Beijing and the northern provinces of China, while our work studied subjects from the island of Taiwan.

We identified additional suggestive loci, significant at a false discovery rate less than 0.05; in order of importance these were: SH2B3, ADAMTS7, PHACTR1, GGCX, HTRA1, COL4A1, LARP6-LRRC49.

By allowing two-way interactions in the penalized logistic regression, we observed an interaction between the major alleles of the PHACTR1 and ADAMTS7 loci contributes to CAD in Taiwan (Table 4). In addition to association with CAD, PHACTR1 (phosphatase and actin regulator 1; dbGENE 221692) has been associated with cervical artery dissection association [32] and shown to play a role in regulation of endothelial cells [33]. ADAMTS7 (a disintegrin and metalloproteinase with thrombospondin type 1 motif; dbGENE 11173) has been recently shown to affect the migration of smooth muscle cells [34, 35]. While PHACTR1 is also expressed in smooth muscle cells [35], further work will be required to unravel the relationship between these two genes in CAD risk.

We also observed that the SH2B3 and COL4A1 contribute to CAD through an interaction, suggesting that the minor allele of SH2B3 contributes to CAD only in the presence of the major allele of the COL4A1 locus (Table 5). SH2B3 (SH2B adaptor protein 3; dbGENE 10019) is involved in cytokine signaling pathways and COL4A1 (collagen type 4 alpha 1; dbGENE 1282) has also been implicated in cerebral small vessel disease [36]. Both genes are up-regulated over 5-fold as endothelial cells change from normal to the atherosclerotic disease state in a mouse model of atherosclerosis [35].

While the mathematical advantages of penalized logistic regression methods have been discussed elsewhere, [11, 12, 37] we have compared the results of the penalized logistic regression reported here with a standard logistic regression in order to draw practical conclusions on the use of this method (S2 Table; age, gender, and the eigenvalues from principal components 1 and 2 as covariates; top 100 results are listed; Plink v1.90). First, the p-values for a given SNP are of the same magnitude and very close in value between the two analyses. Thus, p-values reported using penalized logistic regression will be comparable to those obtained with standard logistic regression. Second, the penalized logistic regression result readily suggested that all of the 50 9p21 associations represented a single association signal, since no other 9p21 loci were

identified by this method. Third, the interaction analysis identified two SNP by SNP interactions within these top 100 SNPs.

Our conclusions on the genetic architecture of coronary artery disease in Taiwan are limited by several factors. First, our sample size of 3133 cases and 5423 controls, while among the largest reported for a population of Chinese descent, yet remains modest when compared with the sample sizes reported for populations of European descent by the CARDIoGRAM consortium. Aside from the genome-wide significance of  $5 \times 10^{-8}$  observed for the 9p21 locus, the results with a significance of a false discovery rate less than 5% should be considered tentative for Taiwan. However, it should be noted in this context that all loci tested in this report were already established CAD loci and this fact further supports the conclusions presented here. The next step to be taken by us will be to collaborate with other groups studying CAD genetics of East Asians with the aim of confirming the conclusions presented here. Second, since our SNPs were selected based on GWAS results from European populations as incorporated in the CardioMetaboChip and not based on coverage of the entire genome, it is likely that there are genetic effects contributing to the genetic architecture of CAD in Taiwan not yet detected. The next step to be taken by us will be to genotype a SNP set selected for genome-wide coverage of East Asian populations, in addition to collaborating with other investigators. Third, while we did identify 2 SNP by SNP interactions, the interaction analysis was limited to the study of two-way interactions. We expect to explore higher order complexities contributing to CAD in Taiwan in future studies with larger samples and more genotypes.

While all SNPs tested in this report were selected from meta-analyses by the CARDIoGRAM consortium for the design of the CardioMetaboChip, two of these loci, HTRA1 and LARP6-LRRC49, were not listed in the final tables of GWAS studies of CAD associations in Europeans [1] nor in Chinese from Beijing [9]. Furthermore, we did not observe an association with the SORT1 or ZC3HC1 loci, though both have been reported with effects equal or greater in significance to the PHACTR1 and ADAMTS7 loci in Europeans. While this observation could be the effect of our smaller sample size or differences in subject ascertainment, there may also be differences in the magnitude of effect size such that different loci are most important for CAD in different populations. These observations, however, underscore the value of continuing genome-wide association studies in populations of other than European descent; comparisons between populations as well as within different populations will provide findings relevant to the genetics of CAD in all populations.

## Supporting Information

**S1 Fig. Linkage Disequilibrium between Major SNPs in the 9p21 Region.** Linkage disequilibrium between the top SNP in this report and the four coronary artery disease SNPs reported in Lu et al. [9] Figures prepared using Haploview v4. [30] Panel A) Data for subjects of Chinese descent from the Thousand Genomes Project; [31] Panel B) Data for subjects of European descent from the Thousand Genomes Project; Panel C) Data for Taiwan subjects in this report. (TIFF)

**S1 Table. Conditional Analysis of Major SNPs in the 9p21 Region, Detail.** Association assessed by logistic regression with age, gender, and the eigenvalues from Principal Components 1 and 2 as covariates; this table provides data from Fig 1 for SNPs reported by Lu et al. [9]. (XLSX)

**S2 Table. Comparison of Standard Logistic Regression with Penalized Logistic Regression.** (XLSX)

**S3 Table. Comparison of standard logistic regression with penalized logistic regression.** (XLSX)

## Acknowledgments

The TAICHI study was supported by a grant from NHLBI grant HL087647. TLA is supported by an NIH career development award K23DK088942. The Taiwan Dragon was supported by grants from National Science Council (NSC 98-2314-B-075A-002-MY3) and Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1013001C; TCVGH-1013002D). The TCAD was supported by grants from National Science Council (NSC 98-2314-B-075A-002-MY3) and Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1013001C; TCVGH-1013002D). The SAPPHiRE was supported by grants from National Health Research Institutes (BS-094-PP-01 & PH-100-PP-03). HALST was supported by grants from National Health Research Institutes (PH-100-SP-01). The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

## Author Contributions

Conceived and designed the experiments: JR YDC WS CAH TQ KT TA. Performed the experiments: DA KT IL JJ EK WL CAH LC TQ. Analyzed the data: JR YDC KT. Contributed reagents/materials/analysis tools: IL JJ TW EK WL DA YFC TQ LC CCH JR WS YC XG KT. Wrote the paper: KT JR YDC TA WS XG.

## References

1. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nature genetics*. 2013; 45(1):25–33. Epub 2012/12/04. doi: [10.1038/ng.2480](https://doi.org/10.1038/ng.2480) PMID: [23202125](https://pubmed.ncbi.nlm.nih.gov/23202125/); PubMed Central PMCID: [PMC3679547](https://pubmed.ncbi.nlm.nih.gov/PMC3679547/).
2. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature genetics*. 2011; 43(4):333–8. Epub 2011/03/08. doi: [10.1038/ng.784](https://doi.org/10.1038/ng.784) PMID: [21378990](https://pubmed.ncbi.nlm.nih.gov/21378990/); PubMed Central PMCID: [PMC3119261](https://pubmed.ncbi.nlm.nih.gov/PMC3119261/).
3. Makinen VP, Civelek M, Meng Q, Zhang B, Zhu J, Levian C, et al. Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease. *PLoS genetics*. 2014; 10(7): e1004502. Epub 2014/07/18. doi: [10.1371/journal.pgen.1004502](https://doi.org/10.1371/journal.pgen.1004502) PMID: [25033284](https://pubmed.ncbi.nlm.nih.gov/25033284/); PubMed Central PMCID: [PMC4102418](https://pubmed.ncbi.nlm.nih.gov/PMC4102418/).
4. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS genetics*. 2012; 8(8):e1002793. Epub 2012/08/10. doi: [10.1371/journal.pgen.1002793](https://doi.org/10.1371/journal.pgen.1002793) PMID: [22876189](https://pubmed.ncbi.nlm.nih.gov/22876189/); PubMed Central PMCID: [PMC3410907](https://pubmed.ncbi.nlm.nih.gov/PMC3410907/).
5. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, et al. Genomewide association analysis of coronary artery disease. *The New England journal of medicine*. 2007; 357(5):443–53. Epub 2007/07/20. doi: [10.1056/NEJMoa072366](https://doi.org/10.1056/NEJMoa072366) PMID: [17634449](https://pubmed.ncbi.nlm.nih.gov/17634449/); PubMed Central PMCID: [PMC2719290](https://pubmed.ncbi.nlm.nih.gov/PMC2719290/).
6. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447(7145):661–78. Epub 2007/06/08. doi: [10.1038/nature05911](https://doi.org/10.1038/nature05911) PMID: [17554300](https://pubmed.ncbi.nlm.nih.gov/17554300/); PubMed Central PMCID: [PMC2719288](https://pubmed.ncbi.nlm.nih.gov/PMC2719288/).
7. Chen HH, Almontashiri NA, Antoine D, Stewart AF. Functional genomics of the 9p21.3 locus for atherosclerosis: clarity or confusion? *Current cardiology reports*. 2014; 16(7):502. Epub 2014/06/05. doi: [10.1007/s11886-014-0502-7](https://doi.org/10.1007/s11886-014-0502-7) PMID: [24893939](https://pubmed.ncbi.nlm.nih.gov/24893939/).
8. Johnson AD, Hwang SJ, Voorman A, Morrison A, Peloso GM, Hsu YH, et al. Resequencing and clinical associations of the 9p21.3 region: a comprehensive investigation in the Framingham heart study.

- Circulation. 2013; 127(7):799–810. Epub 2013/01/15. doi: [10.1161/circulationaha.112.111559](https://doi.org/10.1161/circulationaha.112.111559) PMID: [23315372](https://pubmed.ncbi.nlm.nih.gov/23315372/); PubMed Central PMCID: PMCPCmc3686634.
9. Lu X, Wang L, Chen S, He L, Yang X, Shi Y, et al. Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease. *Nature genetics*. 2012; 44(8):890–4. Epub 2012/07/04. doi: [10.1038/ng.2337](https://doi.org/10.1038/ng.2337) PMID: [22751097](https://pubmed.ncbi.nlm.nih.gov/22751097/); PubMed Central PMCID: PMCPCmc3927410.
  10. HapMap Consortium. The International HapMap Project. *Nature*. 2003; 426(6968):789–96. Epub 2003/12/20. doi: [10.1038/nature02168](https://doi.org/10.1038/nature02168) PMID: [14685227](https://pubmed.ncbi.nlm.nih.gov/14685227/).
  11. Wu TT, Chen YF, Hastie T, Sobel E, Lange K. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics (Oxford, England)*. 2009; 25(6):714–21. Epub 2009/01/30. doi: [10.1093/bioinformatics/btp041](https://doi.org/10.1093/bioinformatics/btp041) PMID: [19176549](https://pubmed.ncbi.nlm.nih.gov/19176549/); PubMed Central PMCID: PMCPCmc2732298.
  12. Lange K, Papp JC, Sinsheimer JS, Sobel EM. Next generation statistical genetics: Modeling, penalization, and optimization in high-dimensional data. *Annual review of statistics and its application*. 2014; 1(1):279–300. Epub 2014/06/24. doi: [10.1146/annurev-statistics-022513-115638](https://doi.org/10.1146/annurev-statistics-022513-115638) PMID: [24955378](https://pubmed.ncbi.nlm.nih.gov/24955378/); PubMed Central PMCID: PMCPCmc4062304.
  13. Kuo JZ, Sheu WH, Assimes TL, Hung YJ, Absher D, Chiu YF, et al. Trans-ethnic fine mapping identifies a novel independent locus at the 3' end of CDKAL1 and novel variants of several susceptibility loci for type 2 diabetes in a Han Chinese population. *Diabetologia*. 2013; 56(12):2619–28. doi: [10.1007/s00125-013-3047-1](https://doi.org/10.1007/s00125-013-3047-1) PMID: [24013783](https://pubmed.ncbi.nlm.nih.gov/24013783/); PubMed Central PMCID: PMCPCMC3825282.
  14. Sheu WH, Kuo JZ, Lee IT, Hung YJ, Lee WJ, Tsai HY, et al. Genome-wide association study in a Chinese population with diabetic retinopathy. *Hum Mol Genet*. 2013; 22(15):3165–73. doi: [10.1093/hmg/ddt161](https://doi.org/10.1093/hmg/ddt161) PMID: [23562823](https://pubmed.ncbi.nlm.nih.gov/23562823/); PubMed Central PMCID: PMCPCMC3699066.
  15. Lee IT, Goodarzi MO, Lee WJ, Rotter JI, Chen YD, Liang KW, et al. The chromosome 9p21 variant not predicting long-term cardiovascular mortality in Chinese with established coronary artery disease: an eleven-year follow-up study. 2014; 2014:626907. doi: [10.1155/2014/626907](https://doi.org/10.1155/2014/626907) PMID: [24804228](https://pubmed.ncbi.nlm.nih.gov/24804228/).
  16. Chang YC, Chiu YF, Liu PH, Shih KC, Lin MW, Sheu WH, et al. Replication of genome-wide association signals of type 2 diabetes in Han Chinese in a prospective cohort. *Clinical endocrinology*. 2012; 76(3):365–72. Epub 2011/07/20. doi: [10.1111/j.1365-2265.2011.04175.x](https://doi.org/10.1111/j.1365-2265.2011.04175.x) PMID: [21767287](https://pubmed.ncbi.nlm.nih.gov/21767287/).
  17. Chang YC, Chiu YF, Shih KC, Lin MW, Sheu WH, Donlon T, et al. Common PCSK1 haplotypes are associated with obesity in the Chinese population. *Obesity (Silver Spring, Md)*. 2010; 18(7):1404–9. Epub 2009/10/31. doi: [10.1038/oby.2009.390](https://doi.org/10.1038/oby.2009.390) PMID: [19875984](https://pubmed.ncbi.nlm.nih.gov/19875984/).
  18. Liang KW, Sheu WH, Lee WL, Lee IT, Lin SY, Ting CT, et al. Shorter GT repeats in the heme oxygenase-1 gene promoter are associated with a lower severity score in coronary artery disease. *Journal of the Chinese Medical Association: JCMA*. 2013; 76(6):312–8. Epub 2013/04/23. doi: [10.1016/j.jcma.2013.03.005](https://doi.org/10.1016/j.jcma.2013.03.005) PMID: [23602216](https://pubmed.ncbi.nlm.nih.gov/23602216/).
  19. Sheu WH, Chen YD, Yu CY, Guo X, Lee TC, Lee WJ, et al. C-reactive protein gene polymorphism 1009A>G is associated with serum CRP levels in Chinese men: a TCVGHAGE study. *Clinica chimica acta; international journal of clinical chemistry*. 2007; 382(1–2):117–23. Epub 2007/05/22. doi: [10.1016/j.cca.2007.04.013](https://doi.org/10.1016/j.cca.2007.04.013) PMID: [17511977](https://pubmed.ncbi.nlm.nih.gov/17511977/).
  20. Tsai HJ, Hsiao CF, Ho LT, Chuang LM, He CT, Curb JD, et al. Genetic variants of human urea transporter-2 are associated with metabolic syndrome in Asian population. *Clinica chimica acta; international journal of clinical chemistry*. 2010; 411(23–24):2009–13. Epub 2010/08/28. doi: [10.1016/j.cca.2010.08.025](https://doi.org/10.1016/j.cca.2010.08.025) PMID: [20797394](https://pubmed.ncbi.nlm.nih.gov/20797394/).
  21. Gunderson KL. Whole-genome genotyping on bead arrays. *Methods in molecular biology (Clifton, NJ)*. 2009; 529:197–213. Epub 2009/04/22. doi: [10.1007/978-1-59745-538-1\\_13](https://doi.org/10.1007/978-1-59745-538-1_13) PMID: [19381978](https://pubmed.ncbi.nlm.nih.gov/19381978/).
  22. Fan JB, Gunderson KL, Bibikova M, Yeakley JM, Chen J, Wickham Garcia E, et al. Illumina universal bead arrays. *Methods in enzymology*. 2006; 410:57–73. Epub 2006/08/30. doi: [10.1016/s0076-6879\(06\)10003-8](https://doi.org/10.1016/s0076-6879(06)10003-8) PMID: [16938546](https://pubmed.ncbi.nlm.nih.gov/16938546/).
  23. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006; 38(8):904–9. Epub 2006/07/25. doi: [ng1847 \[pii\] doi: 10.1038/ng1847](https://doi.org/10.1038/ng1847) PMID: [16862161](https://pubmed.ncbi.nlm.nih.gov/16862161/).
  24. Lange K, Papp JC, Sinsheimer JS, Sripracha R, Zhou H, Sobel EM. Mendel: the Swiss army knife of genetic analysis programs. *Bioinformatics (Oxford, England)*. 2013; 29(12):1568–70. Epub 2013/04/24. doi: [10.1093/bioinformatics/btt187](https://doi.org/10.1093/bioinformatics/btt187) PMID: [23610370](https://pubmed.ncbi.nlm.nih.gov/23610370/); PubMed Central PMCID: PMCPCmc3673222.
  25. Huber W, Carey VJ, Gentleman R, Anders S. Orchestrating high-throughput genomic analysis with Bioconductor. 2015; 12(2):115–21. doi: [10.1038/nmeth.3252](https://doi.org/10.1038/nmeth.3252) PMID: [25633503](https://pubmed.ncbi.nlm.nih.gov/25633503/).
  26. Crawley MJ. *The R Book*. Hoboken, NJ: John Wiley; 2007.
  27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*.

- 2007; 81(3):559–75. Epub 2007/08/19. doi: [10.1086/519795](https://doi.org/10.1086/519795) PMID: [17701901](https://pubmed.ncbi.nlm.nih.gov/17701901/); PubMed Central PMCID: [PMC1950838](https://pubmed.ncbi.nlm.nih.gov/PMC/PMC1950838/).
28. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015; 4:7. Epub 2015/02/28. doi: [10.1186/s13742-015-0047-8](https://doi.org/10.1186/s13742-015-0047-8) PMID: [25722852](https://pubmed.ncbi.nlm.nih.gov/25722852/); PubMed Central PMCID: [PMC4342193](https://pubmed.ncbi.nlm.nih.gov/PMC/PMC4342193/).
  29. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010; 26(18):2336–7. Epub 2010/07/17. doi: [10.1093/bioinformatics/btq419](https://doi.org/10.1093/bioinformatics/btq419) PMID: [20634204](https://pubmed.ncbi.nlm.nih.gov/20634204/); PubMed Central PMCID: [PMC2935401](https://pubmed.ncbi.nlm.nih.gov/PMC/PMC2935401/).
  30. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics (Oxford, England)*. 2005; 21(2):263–5. Epub 2004/08/07. doi: [10.1093/bioinformatics/bth457](https://doi.org/10.1093/bioinformatics/bth457) PMID: [15297300](https://pubmed.ncbi.nlm.nih.gov/15297300/).
  31. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491(7422):56–65. Epub 2012/11/07. doi: [10.1038/nature11632](https://doi.org/10.1038/nature11632) PubMed Central PMCID: [PMC3498066](https://pubmed.ncbi.nlm.nih.gov/PMC/PMC3498066/). PMID: [23128226](https://pubmed.ncbi.nlm.nih.gov/23128226/)
  32. Debette S, Kamatani Y, Metso TM, Kloss M, Chauhan G, Engelter ST, et al. Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection. *Nat Genet*. 2015; 47(1):78–83. Epub 2014/11/25. doi: [10.1038/ng.3154](https://doi.org/10.1038/ng.3154) PMID: [25420145](https://pubmed.ncbi.nlm.nih.gov/25420145/).
  33. Jarray R, Allain B, Borriello L, Biard D, Loukaci A, Larghero J, et al. Depletion of the novel protein PHACTR-1 from human endothelial cells abolishes tube formation and induces cell death receptor apoptosis. *Biochimie*. 2011; 93(10):1668–75. Epub 2011/07/30. doi: [10.1016/j.biochi.2011.07.010](https://doi.org/10.1016/j.biochi.2011.07.010) PMID: [21798305](https://pubmed.ncbi.nlm.nih.gov/21798305/).
  34. Pu X, Xiao Q, Kiechl S, Chan K, Ng FL, Gor S, et al. ADAMTS7 cleavage and vascular smooth muscle cell migration is affected by a coronary-artery-disease-associated variant. *American journal of human genetics*. 2013; 92(3):366–74. Epub 2013/02/19. doi: [10.1016/j.ajhg.2013.01.012](https://doi.org/10.1016/j.ajhg.2013.01.012) PMID: [23415669](https://pubmed.ncbi.nlm.nih.gov/23415669/); PubMed Central PMCID: [PMC3591856](https://pubmed.ncbi.nlm.nih.gov/PMC/PMC3591856/).
  35. Erbilgin A, Civelek M, Romanoski CE, Pan C, Hagopian R, Berliner JA, et al. Identification of CAD candidate genes in GWAS loci and their expression in vascular cells. *Journal of lipid research*. 2013; 54(7):1894–905. Epub 2013/05/15. doi: [10.1194/jlr.M037085](https://doi.org/10.1194/jlr.M037085) PMID: [23667179](https://pubmed.ncbi.nlm.nih.gov/23667179/); PubMed Central PMCID: [PMC3679391](https://pubmed.ncbi.nlm.nih.gov/PMC/PMC3679391/).
  36. Lemmens R, Maugeri A, Niessen HW, Goris A, Tousseyn T, Demaerel P, et al. Novel COL4A1 mutations cause cerebral small vessel disease by haploinsufficiency. *Hum Mol Genet*. 2013; 22(2):391–7. Epub 2012/10/16. doi: [10.1093/hmg/dds436](https://doi.org/10.1093/hmg/dds436) PMID: [23065703](https://pubmed.ncbi.nlm.nih.gov/23065703/); PubMed Central PMCID: [PMC3606013](https://pubmed.ncbi.nlm.nih.gov/PMC/PMC3606013/).
  37. Zhou H, Alexander DH, Sehl ME, Sinsheimer JS, Sobel EM, Lange K. Penalized regression for genome-wide association screening of sequence data. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing*. 2011:106–17. Epub 2010/12/02. PMID: [21121038](https://pubmed.ncbi.nlm.nih.gov/21121038/).