Data in Brief 17 (2018) 820–829

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data article

Data on genetic associations of carotid atherosclerosis markers in Mexican American and European American rheumatoid arthritis subjects

Rector Arya [a,b,*], Agustin Escalante [c], Vidya S. Farook [a,b], Jose F. Restrepo [c], Daniel F. Battafarano [d], Marcio Almeida [a,b], Mark Z. Kos [a,b], Marcel J. Fourcaudot [e], Srinivas Mummidi [a,b], Satish Kumar [a,b], Joanne E. Curran [a,b], Christopher P. Jenkinson [a,b], John Blangero [a,b], Ravindranath Duggirala [a,b], Inmaculada del Rincon [c]

[a] *Department of Human Genetics, School of Medicine, the University of Texas Rio Grande Valley, Edinburg/Brownsville, Texas, USA*
[b] *South Texas Diabetes and Obesity Institute, School of Medicine, the University of Texas Rio Grande Valley, Edinburg/Brownsville, Texas, USA*
[c] *Department of Medicine, Division of Rheumatology and Clinical Immunology, the University of Texas Health, San Antonio, Texas, USA*
[d] *San Antonio Military Medical Center, Fort Sam Houston, Texas, USA*
[e] *Division of Diabetes, Department of Medicine, the University of Texas Health, San Antonio, Texas, USA*

ARTICLE INFO

Article history:
Received 12 December 2017
Received in revised form
1 February 2018
Accepted 2 February 2018
Available online 8 February 2018

ABSTRACT

Carotid Intima-media thickness (CIMT) and plaque are well established markers of subclinical atherosclerosis and are widely used for identifying subclinical atherosclerotic disease. We performed association analyses using Metabochip array to identify genetic variants that influence variation in CIMT and plaque, measured using B-mode ultrasonography, in rheumatoid arthritis (RA) patients. Data on genetic associations of common variants associated with both CIMT and plaque in RA subjects involving Mexican Americans (MA) and European Americans (EA) populations are presented in this article. Strong associations were observed after adjusting for covariate effects including baseline clinical characteristics and statin use. Susceptibility loci and genes and/or nearest genes associated with CIMT in MAs and EAs with

DOI of original article: https://doi.org/10.1016/j.atherosclerosis.2017.11.024
* Corresponding author.
E-mail address: rector.arya@utrgv.edu (R. Arya).

https://doi.org/10.1016/j.dib.2018.02.006
2352-3409/© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

RA are presented. In addition, common susceptibility loci influencing CIMT and plaque in both MAs and EAs have been presented. Polygenic Risk Score (PRS) plots showing complementary evidence for the observed CIMT and plaque association signals are also shown in this article. For further interpretation and details, please see the research article titled "A Genetic Association Study of Carotid Intima-Media Thickness (CIMT) and Plaque in Mexican Americans and European Americans with Rheumatoid Arthritis" which is being published in Atherosclerosis (Arya et al., 2018) [1]. (Arya et al., in press) Thus, common variants in several genes exhibited significant associations with CIMT and plaque in both MAs and EAs as presented in this article. These findings may help understand the genetic architecture of subclinical atherosclerosis in RA populations.

© 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Specifications table

Subject area	*Genetics, Genomics and Molecular Biology*
More specific subject area	*CIMT and plaque Genetic Association Data in RA subjects*
Type of data	*Tables, figures and text file*
How data was acquired	*ORALE study data were collected from 11 private and public rheumatology outpatient clinics in San Antonio, Texas.*
Data format	*Original and analyzed data set*
Experimental factors	*Recruited patients who met the 1987 criteria for RA*
Experimental features	*Metabochip array-based Association Analyses*
Data source location	*Division of Rheumatology and Clinical Immunology, Department of Medicine, UTHSCSA, San Antonio, TX USA.*
Data accessibility	*Data are available with this article and/or upon request*

Value of the data.

- Genetic association data for CIMT and plaque in subjects with rheumatoid arthritis are valuable to understand the genetic architecture of the carotid atherosclerosis markers in RA patients.
- Observed association signals for carotid atherosclerosis markers in both Mexican American and European American cohorts would give more insight into population differences as well as trait-specific and common genetic determinants.
- This data could also be potentially used for replication of genetic association findings for atherosclerosis markers in other populations.

1. Data

A comprehensive association results of the transformed CIMT after adjustment for covariate effects in MAs and EAs are shown in Tables 1 and 2, respectively. As presented in Arya et al. [1], a total of 24 SNPs from 11 chromosomes exhibited association with CIMT in MAs at $p < 1 \times 10^{-4}$, and the p values ranged from 9.95×10^{-5} to 3.80×10^{-6}, while 12 SNPs from 7 chromosomes were associated with

Table 1
Susceptibility loci associated with carotid IMT in Mexican Americans.

Chr.	SNP	Position, bp[a]	Gene/nearest gene	Loc.	A1[b]	MAF	BETA	SE	P[c]
6	rs17526722	26,026,834	*SLC17A2*	I	A	0.10	−0.8377	0.1797	3.80E-06
6	rs36014129	25,992,498	*SLC17A3 | SLC17A2*	IG	A	0.10	−0.85	0.184	4.63E-06
6	rs13212534	26,090,989	*TRIM38*	I	A	0.10	−0.8607	0.1891	6.36E-06
6	rs13213957	25,894,205	*SLC17A1*	I	C	0.10	−0.7806	0.1762	1.10E-05
6	rs55912630	25,974,914	*SLC17A3*	I	C	0.10	−0.7806	0.1762	1.10E-05
6	rs13191296	25,792,585	*SCGN*	I	T	0.08	−0.7771	0.176	1.19E-05
2	rs4894108	180,112,792	*ZNF385B*	I	G	0.19	−0.2501	0.05751	1.59E-05
6	rs13211947	25,972,797	*SLC17A3*	I	T	0.10	−0.8179	0.189	1.75E-05
6	rs41266779	26,129,851	*HIST1H3A | HIST1H4A*	IG	T	0.10	−0.8178	0.1891	1.78E-05
17	rs2672901	76,411,261	*KIAA1303*	I	A	0.32	−0.1929	0.04489	2.00E-05
13	rs11619113	109,716,661	*COL4A1*	I	G	0.20	−0.3918	0.09165	2.21E-05
13	rs12873154	109,718,853	*COL4A1*	I	G	0.20	−0.3918	0.09165	2.21E-05
13	rs11619038	109,721,800	*COL4A1*	I	T	0.20	−0.3918	0.09165	2.21E-05
6	rs13202688	26,101,448	*TRIM38 | HIST1H1A*	IG	G	0.10	−0.7505	0.1762	2.37E-05
10	rs61850526	63,190,012	*C10orf107*	I	T	0.02	−0.4624	0.1126	4.56E-05
6	rs34043431	25,983,063	*SLC17A3 | SLC17A2*	IG	C	0.10	−0.7012	0.173	5.68E-05
11	rs76599700	61,401,283	*FADS3*	I	T	0.26	−0.7317	0.1814	6.17E-05
16	rs11860529	82,330,023	*CDH13*	I	T	0.08	0.4112	0.1028	7.06E-05
1	rs17436982	219,622,889	*HLX | DUSP10*	IG	T	0.22	0.2827	0.07173	8.98E-05
5	rs250216	50,317,115	*PARP8| LOC642366*	IG	C	0.07	0.2322	0.05899	9.20E-05
7	rs11761467	27,828,130	*TAX1BP1*	I	T	0.10	−0.3244	0.08256	9.44E-05
13	rs11620140	109,730,512	*COL4A1*	I	C	0.11	−0.3696	0.09417	9.62E-05
6	rs11966018	12,317,214	*HIVEP1 | EDN1*	IG	C	0.03	0.691	0.1763	9.82E-05
15	rs7177074	97,357,475	*LOC145814*	I	A	0.05	−0.5138	0.1312	9.95E-05

Chr. = chromosome; SNP = single nucleotide polymorphism; Loc = Location; I = Intron; IG = Intergenic; MAF = minor allele frequency.

[a] Based on National Center for Biotechnology Information (NCBI) Build36 (hg18).
[b] A1 = minor allele.
[c] *p* values ranked from low to high.

Table 2
Susceptibility loci associated with carotid IMT in European Americans.

Chr.	SNP	Position, bp[a]	Gene/nearest gene	Loc.	A1[b]	MAF	BETA	SE	P[c]
15	rs1867148	73,127,038	*PPCDC*	I	C	0.44	−0.2804	0.06064	5.11E-06
15	rs7163636	73,133,813	*PPCDC / C15orf39*	IG	C	0.48	−0.2748	0.06053	7.51E-06
13	rs323453	104,232,379	*LOC728183 | DAOA*	IG	G	0.34	−0.2724	0.06346	2.23E-05
15	rs3812945	73,076,775	*SCAMPS*	I	G	0.45	0.2626	0.06346	2.23E-05
1	rs4846566	217,797,888	*LOC728510 | ZC3H11B*	IG	T	0.01	−1.186	0.2777	2.46E-05
2	rs12987042	38,518,674	*ARL6IP2|RPLPO-like*	IG	A	0.36	−0.2608	0.0619	3.13E-05
1	rs17006057	217,790,022	*LOC728510/ZC3H11B*	IG	A	0.01	−1.225	0.2915	3.29E-05
15	rs6495122	72,912,698	*CPLX3 | ULK3*	IG	A	0.42	−0.2525	0.06149	4.90E-05
1	rs2645091	2,214,505	*SKI*	I	T	0.15	−0.3511	0.08741	7.07E-05
16	rs17821532	52,504,199	*FTO*	I	A	0.05	−0.7006	0.1751	7.58E-05
6	rs7742814	119,185,974	*C6orf204|ASF1A*	IG	G	0.36	0.2555	0.06462	9.09E-05
11	rs4387380	5,824,023	*OR52E6|OR52E8*	IG	C	0.03	−0.8226	0.2085	9.45E-05

Chr. = chromosome; SNP = single nucleotide polymorphism; Loc = Location; I = Intron; IG = Intergenic; MAF = minor allele frequency.

[a] Based on National Center for Biotechnology Information (NCBI) Build36 (hg18).
[b] A1 = minor allele.
[c] *p* values ranked from low to high.

Table 3
Common susceptibility loci associated with carotid IMT and plaque in Mexican Americans.

Chr.	SNP	Position, bp[a]	Gene/nearest gene	Loc.	A1[b]	MAF	BETA	SE	P[c] CIMT	P[c] Plaque
6	rs17526722	26,026,834	*SLC17A2*	I	A	0.10	−0.8377	0.1797	3.80E-06	0.04112
2	rs4894108	180,112,792	*ZNF385B*	I	G	0.19	−0.2501	0.05751	1.59E-05	0.00823
17	rs2672901	76,411,261	*KIAA1303*	I	A	0.32	−0.1929	0.04489	2.00E-05	0.01672
13	rs12873154	109,718,853	*COL4A1*	I	G	0.20	−0.3918	0.09165	2.21E-05	0.04923
10	rs61850526	63,190,012	*C10orf107*	I	T	0.02	−0.4624	0.1126	4.56E-05	0.001367
11	rs76599700	61,401,283	*FADS3*	I	T	0.26	−0.7317	0.1814	6.17E-05	0.08027
16	rs11860529	82,330,023	*CDH13*	I	T	0.08	0.4112	0.1028	7.06E-05	0.02431
1	rs17436982	219,622,889	*HLX | DUSP10*	IG	T	0.22	0.2827	0.07173	8.98E-05	0.006092
5	rs250216	50,317,115	*PARP8 | LOC642366*	IG	C	0.07	0.2322	0.05899	9.20E-05	0.0262
7	rs11761467	27,828,130	*TAX1BP1*	I	T	0.10	−0.3244	0.08256	9.44E-05	0.01331
6	rs11966018	12,317,214	*HIVEP1 | EDN1*	IG	C	0.03	0.691	0.1763	9.82E-05	0.009176
15	rs7177074	97,357,475	*LOC145814*	I	A	0.05	−0.5138	0.1312	9.95E-05	0.001647

Chr. = chromosome; SNP = single nucleotide polymorphism; Loc = Location; I = Intron; IG = Intergenic; MAF = minor allele frequency.

[a] Based on National Center for Biotechnology Information (NCBI) Build36/130 (hg18).
[b] A1 = minor allele.
[c] *p* values ranked from low to high.

Table 4
Common susceptibility loci associated with plaque and CIMT in Mexican Americans.

Chr.	SNP	Position, bp[a]	Gene/nearest gene	Loc	A1[b]	MAF	OR	SE	P[c] Plaque	P[c] CIMT
13	rs496916	109,649,015	*COL4A1*	I	C	0.41	0.514	0.1428	3.15E-06	0.03256
15	rs9806753	46,953,709	*SHC4 | EID1*	IG	A	0.26	1.739	0.1229	6.71E-06	0.1274
6	rs9463110	12,890,588	*PHACTR1*	I	G	0.44	1.699	0.1222	1.45E-05	0.02229
22	rs2092179	38,364,539	*CACNA1I*	I	C	0.25	0.5883	0.1289	3.85E-05	0.02903
9	rs7869506	98,127,033	*SLC35D2*	I	T	0.26	1.688	0.1296	5.41E-05	0.005388
1	rs6667860	36,730,800	*CSF3R|GRIK3*	IG	C	0.48	1.658	0.1298	9.76E-05	0.0002729

Chr. = chromosome; SNP = single nucleotide polymorphism; Loc = Location; I = Intron; IG = Intergenic; MAF = minor allele frequency.

[a] Based on National Center for Biotechnology Information (NCBI) Build36 (hg18); OR = Odds Ratio.
[b] A1 = minor allele.
[c] *p* values ranked from low to high.

CIMT in EAs and the *p* values ranged from 9.45×10^{-5} to 5.11×10^{-6}. The best associated SNPs are different in both populations.

Top SNPs that are associated with CIMT and plaque were different in significance levels but exhibited associations with either phenotype at nominal significance levels ($p < 0.05$) as shown in Tables 3–6. It is well known that some variants exhibit unique associations with a given phenotype (CIMT or plaque) while other variants exhibit common associations with both phenotypes (CIMT and plaque). Despite the correlation between CIMT and plaque, different trait-specific genetic determinants can be expected due to their differing pathobiology and associated phenotypic severity of plaque as shown by our earlier study as well as other studies [2,3]. In addition, genetic differences are expected between the populations of European background and admixed populations such as the Mexican Americans that have both European and Native American ancestries. Furthermore, results from previous studies also support our findings [4,5].

As shown in Fig. 1, Quantile-Quantile (Q-Q) plots of the transformed CIMT and plaque in both MAs and EAs have been generated using EPACTS. Q-Q plots exhibited a roughly straight line through the origin with a unit slope indicating almost no inflation.

To further investigate the genetic architecture of CIMT, PRS analysis was conducted, with scores representing summations of CIMT-, and plaque-associated alleles from the Metabochip array. A PRS is

Table 5
Common susceptibility loci associated with CIMT and plaque in European Americans.

Chr.	SNP	Position, bp[a]	Gene/Nearest Gene	Loc.	A1[b]	MAF	BETA	SE	P[c] **CIMT**	P[c] **plaque**
15	rs1867148	73,127,038	*PPCDC*	I	C	0.44	−0.2804	0.06064	5.11E-06	0.01242
13	rs323453	104,232,379	*LOC728183 | DAOA*	IG	G	0.34	−0.2724	0.06346	2.23E-05	0.2049
1	rs4846566	217,797,888	*LOC728510 | ZC3H11B*	IG	T	0.01	−1.186	0.2777	2.46E-05	0.0676
2	rs12987042	38,518,674	*ARL6IP2|RPLPO-like*	IG	A	0.36	−0.2608	0.0619	3.13E-05	0.0005153
15	rs6495122	72,912,698	*CPLX3 | ULK3*	IG	A	0.42	−0.2525	0.06149	4.90E-05	0.005717
1	rs2645091	2,214,505	*SKI*	I	T	0.15	−0.3511	0.08741	7.07E-05	0.03331
16	rs17821532	52,504,199	*FTO*	I	A	0.05	−0.7006	0.1751	7.58E-05	0.03347
6	rs7742814	119,185,974	*C6orf204|ASF1A*	IG	G	0.36	0.2555	0.06462	9.09E-05	0.3764
11	rs4387380	5,824,023	*OR52E6|OR52E8*	IG	C	0.03	−0.8226	0.2085	9.45E-05	0.06833

Chr. = chromosome; SNP = single nucleotide polymorphism; Loc = Location; I = Intron; IG = Intergenic; MAF = minor allele frequency.

[a] Based on National Center for Biotechnology Information (NCBI) Build36 (hg18).
[b] A1 = minor allele.
[c] *p* values ranked from low to high.

Table 6
Common susceptibility loci associated with carotid plaque and CIMT in European Americans.

Chr.	**SNP**	**Position, bp[a]**	**Gene**	**Loc.**	**A1[b]**	**MAF**	**OR**	**SE**	**P[c] Plaque**	**P[c] CIMT**
12	rs515291	38,689,259	*SLC2A13*	I	G	0.32	0.4987	0.167	3.09E-05	0.0007138
11	rs10501399	68,979,039	*MYEOV|CCND1*	IG	T	0.09	0.2938	0.2952	3.34E-05	0.001367
15	rs692390	39,274,282	*EXDL1*	I	A	0.19	0.4138	0.2157	4.30E-05	0.0001269
5	rs6887230	148,706,339	*GRPEL2*	I	G	0.30	2.02	0.1746	5.61E-05	0.009851
17	rs2070776	59,361,230	*CD79B*	C (ns)	T	0.33	0.5033	0.1722	6.68E-05	0.02914
16	rs3785233	7,607,511	*A2BP1*	I	C	0.16	0.419	0.2229	9.49E-05	0.0001409
6	rs10948573	50,800,310	*TFAP2D*	I	G	0.37	2.044	0.1834	9.68E-05	0.2071

Chr. = chromosome; SNP = single nucleotide polymorphism; Loc = Location; I = Intron; IG = Intergenic; MAF = minor allele frequency.

[a] Based on National Center for Biotechnology Information (NCBI) Build36 (hg18); OR = Odds Ratio.
[b] A1 = minor allele.
[c] *p* values ranked from low to high.

the sum of trait-associated alleles across many genetic loci, and was calculated in an independent population i.e., EA population, using the genome-wide association results from the MA population, a discovery population for a given trait (i.e., CIMT) to detect shared genetic etiology among traits, to establish the presence of a genetic signal in underpowered studies, and to infer the genetic architecture of a trait [6]. Scoring routines were determined from the association test results for the MA cohort, with risk alleles identified based on varying p-value thresholds (1,000 different p-value (Pts) thresholds, representing increments of $p = 0.001$), each weighted by their estimated effect sizes on CIMT or plaque. Scores were then computed in the independent EA cohort, and evaluated as predictors of CIMT or plaque via regression models (covariates age, sex, PCs 1 and 2, RA duration, statin use, and htn) as shown in Fig. 2 (A and B). This work was performed using PRSice v.1.23, a polygenic risk score software.

2. Experimental design, subjects and methods

2.1. Subjects

We used existing samples/data from the ORALE (**O**utcome of **R**heumatoid **A**rthritis **L**ongitudinal **E**valuation) study, involving 700 unrelated MAs and unrelated 415 EAs. From 1996 to 2009, we

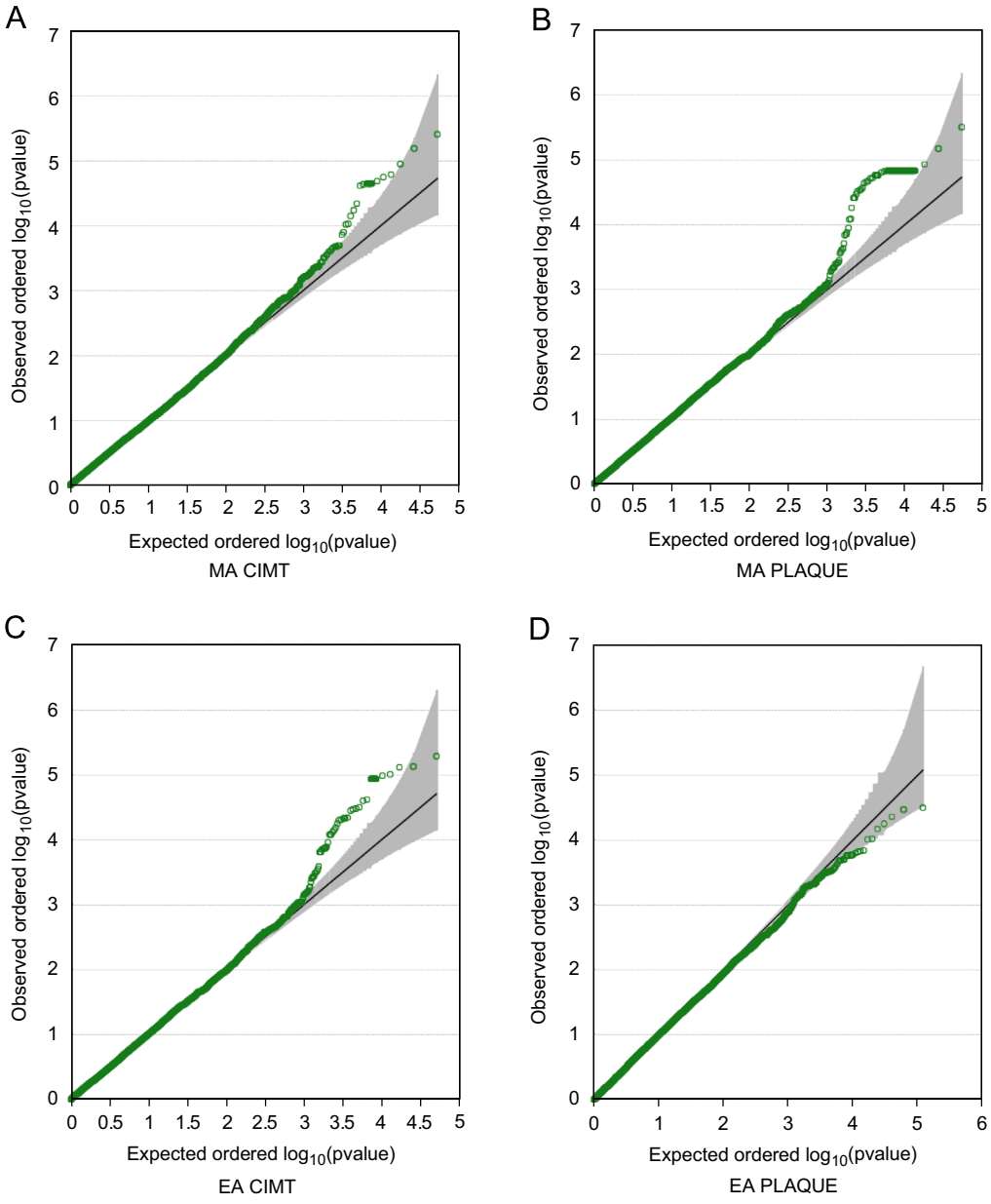
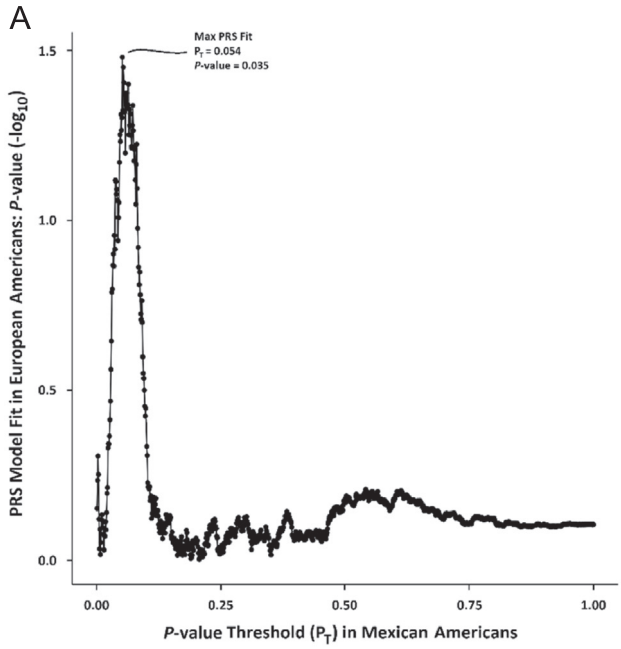


Fig. 1. Q-Q plots for CIMT and plaque association p-values in MAs and EAs.

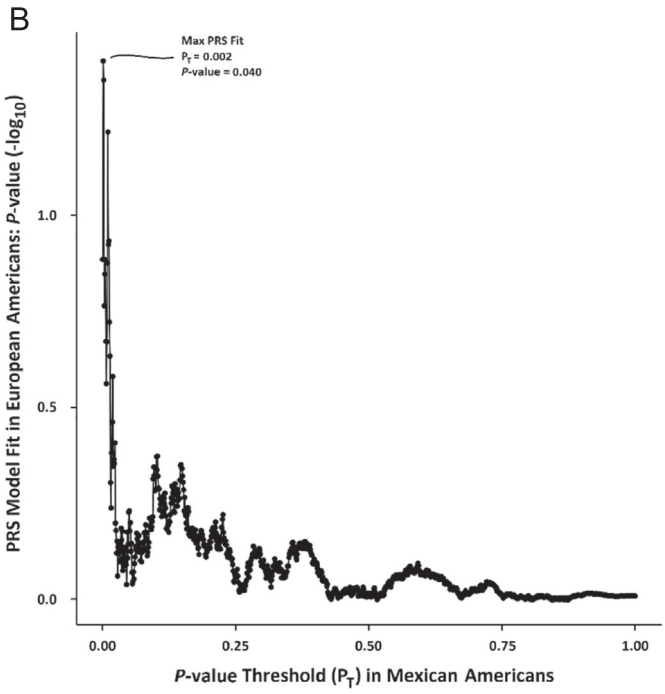
recruited consecutive patients who met the 1987 criteria for RA [7] from 11 private and public rheumatology outpatient clinics in San Antonio, Texas.

2.2. Phenotyping

CIMT and plaque were measured using carotid ultrasound. A single technician performed a duplex scan of the carotid arteries in all patients, following a standardized vascular protocol as implemented



High-resolution Polygenic Risk Score Model Fit in EA for CIMT with theoretical expectations across range of P_T in MA



High-resolution Polygenic Risk Score Model Fit in EA for Plaque with theoretical expectations across a range of P_T in MA

Fig. 2. High-resolution Polygenic Risk Score Model Fit in EA for CIMT (A) and Plaque (B) with theoretical expectations across a range of P_T in MA. A. High-resolution Polygenic Risk Score Model Fit in EA for CIMT with theoretical expectations across a range of P_T in MA. B. High-resolution Polygenic Risk Score Model Fit in EA for Plaque with theoretical expectations across a range of P_T in MA.

in an ATL HDI-3000 High Resolution Imaging machine with a L7-4 Transducer (Philips Medical Systems North America Company, Bothell, WA). For CIMT, we measured the end diastole at each of the near and far walls of the right and left common carotid arteries, and the anterior oblique, lateral and posterior oblique views of the internal carotid artery, for a total of 16 CIMT measurements per person. The maximal CIMT of the common and internal carotid arteries were obtained by averaging the maximal measurement from the near and far walls at each projection, from the right and left sides. Then the composite maximal CIMT was calculated by averaging the common and internal carotid maximal CIMT values. The result is one CIMT value per person, expressed in millimeters. Carotid plaque was identified as a discrete projection of 50% or more from the adjacent wall into the vessel lumen.

2.3. Genotyping

The Metabochip (Illumina) is a custom BeadChip targeting 196,725 genetic variants. Common and less common genetic variants were chosen from among the first iteration of the 1000 Genomes Project and represent index GWAS-identified variants regardless of disease or phenotype as of 2009 [8]. As previously described [8], it was primarily designed for fine mapping of metabolic and cardiovascular disease-related loci, and replication of susceptibility loci for specific GWAS-identified regions associated with cardio-metabolic disease and related phenotypes. Several studies have used this platform to successfully identify genetic risk factors influencing complex disease phenotypes [8,9].

Briefly, Metabochip was a custom Illumina iSelect genotyping array designed to test ~200,000 SNPs identified through genome-wide meta-analyses for metabolic and atherosclerotic/ cardiovascular diseases and traits in a cost-effective manner. It was designed by representatives of the following GWAS meta-analysis Consortia: CARDIoGRAM (coronary artery disease), DIAGRAM (type 2 diabetes), GIANT (height and weight), MAGIC (glycemic traits), Lipids (lipids), ICBP-GWAS (blood pressure), and QT-IGC (QT interval). It supports genotyping of SNPs selected according to five sets of criteria: (1) individual SNPs displaying evidence for association in GWA meta-analyses to diseases and traits relevant to metabolic and atherosclerotic-cardiovascular endpoints, (2) detailed fine mapping of loci validated at genome-wide significance from these meta-analyses, (3) all SNPs associated at genome-wide significance with any human trait, (4) "wildcards" selected by each meta-analysis Consortium for Consortium-specific purposes, and (5) other useful content, including SNPs that tag common CNPs, SNPs in the HLA region, SNPs marking the X and Y chromosomes and mtDNA, and for sample fingerprinting (common SNPs represented on major genome-wide array products from both Illumina and Affymetrix) [8,10]. After merging and pruning the lists (to remove redundant SNPs), a total of 217,697 SNPs representing 245,243 bead types was submitted to Illumina for manufacturing on August 19, 2009. The final chip consisted of genotypes of ~200,000 SNPs per sample. We performed the genotyping according to the Illumina protocol and initial data handling and analysis was performed using Genome Studio v1.7.4 (Illumina).

2.4. Sample and SNP quality control measures

Several quality control measures were applied to the genotypic data of each ethnic group, and only the autosomal SNPs that passed QC were considered for this study. Subjects with low call rates ($<$ 0.95) were removed (MA $=$ 13 and EA $=$ 0). To identify and exclude highly related individuals or duplicate samples, we performed the relationship inference analytical procedure as implemented in the computer program KING [Kinship-based Inference for Genome-wide association studies, [11] and identified related individuals. Subsequently, using the program PLINK [12] and the identity-by-descent (IBD) analysis, closely related individuals up to 3rd degree relatives (IBD $>$ 0.185) were removed (MA $=$ 17 and EA $=$ 3). To detect ethnic outliers, we used EIGENSTRAT c3.0 software package [13] to employ principal components analysis to a subset of autosomal SNPs in our data that were in low LD (r^2 $<$ 0.2) and the HapMap samples as reference for the ethnic groups. Plots were generated using the first two principal components (PCs) for visual inspection. Using our data by ethnic group, samples were identified as population outliers, defined by 4SD from the mean of each of the 2 PCs

that explained the majority of variation in the data, and were subsequently removed (MA = 2 and EA = 0). SNPs with a genotyping call rate less than 95% were removed using PLINK [12]. In addition, SNPs with Hardy-Weinberg Equilibrium (HWE) values of $p < 10^{-4}$ [(MA = 236 (CIMT) and 120 (plaque); and EA = 114 (CIMT), and 38 (plaque)] and with minor allele frequency (MAF) < 0.01 (MA = 57, 323 and EA = 60, 168) were removed from the analysis. After filtering and genotyping pruning, 122,549 SNPs from 668 MAs and 120,827 from 415 EAs were remained in the association analyses.

2.5. Quantile-quantile (Q-Q) plots

Q-Q plots are probability plots, which are useful to compare two probability distributions, sample quantile distribution of the observed chi-squared values (y –axis) versus the quantile distribution of expected (normal or theoretical) chi-squared values (x –axis) graphically by plotting their quantiles against each other. Q-Q plots were done using Efficient and Parallelizable Association Container Toolbox software [EPACTS, http://genome.sph.umich.edu/wiki/EPACTS]. Association p values were adjusted for multiple testing using the conservative Bonferroni correction: 4.08×10^{-7} for MA and 4.14×10^{-7} for EA.

2.6. Statistical genetic analyses

We performed association analyses between the transformed CIMT (as a quantitative trait) and plaque (as a discrete trait) and SNP genotypes in both MA and EA samples after QCs, using PLINK software version 1.07 [12]. Principal Components (PCs) were derived using EIGENSTRAT principal component analysis [13] to adjust for potential population stratification influences. A linear regression additive genetic model (SNPs coded as 0,1, or 2 based on the minor allele dosage) adjusted for the effects of covariates age, sex, RA duration, medication status (statin use, and hypertension, [htn, medication]), and the first two PC1 and PC2, was used for association testing of CIMT, a quantitative trait. Association statistics for plaque, a discrete trait, were calculated using logistic regression assuming an additive model.

2.7. Polygenic risk score (PRS)

PRS for an individual is a summation of their genotypes at variants genome-wide, weighted by effect sizes on a trait of interest i.e. CIMT. Although effect sizes are usually estimated from a GWAS study, we used our Mexican American cohort association results for weighting. Thus, a sum of trait-associated alleles across many genetic loci, has been calculated in an independent population i.e., European American population, using the genome-wide association results from the Mexican-American population, a discovery population for a given trait (i.e., CIMT) to detect shared genetic etiology among traits, to establish the presence of a genetic signal in underpowered studies, and to infer the genetic architecture of a trait [6].

Using the array-wide association results for the MA samples for CIMT (n = 122,549 SNPs), PRS routines were designed and computed in the EA cohort, revealing significant associations with the target phenotype. As shown in Figs. 1 and 2, the PRS model for $P_t < 0.054$ yielded the best fit for CIMT in EAs, but 1000 different P_t thresholds were examined, representing increments of $P = 0.001$. The computed PRSs are then used as predictors of a targeted phenotype in an independent European data set using regression models (i.e., linear or logistic based on the target phenotype). Furthermore, SNPs in the association results for the Mexican-American samples were pruned using PLINK's clumping methodology based on linkage disequilibrium (LD), distance, and association P-values (see http://pngu.mgh.harvard.edu/~purcell/plink/clump.shtml). We used the standard settings (r^2 = 0.1 and 250 Kb), which reduced the number of SNPs actually used in the scoring routines from 122,549 to 36,630. This work was performed using PRSice v.1.23, a polygenic risk score software [6].

Acknowledgements

We thank the participants of the ORALE study. This work was supported by grants from the US National Institutes of Health (R01-HL-085742, R01-HD-037151, and UL1-RR-25767).

Transparency document. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dib.2018.02.006.

References

[1] R. Arya, A. Escalante, V.S. Farook, J.F. Restrepo, D.F. Battafarano, M. Almeida, M.Z. Kos, M.J. Fourcaudot, S. Mummidi, S. Kumar, J.E. Curran, C.P. Jenkinson, J. Blangero, R. Duggirala and I. del Rincon, A Genetic Association Study of carotid intima-media thickness (CIMT) and plaque in Mexican Americans and European Americans with rheumatoid arthritis, *Atherosclerosis* 270, 2018, 1–10. http://dx.doi.org/10.1016/j.atherosclerosis.2017.11.024. (in press).
[2] J.L. Bjorkegren, J.C. Kovacic, J.T. Dudley, E.E. Schadt, Genome-wide significant loci: how important are they? Systems genetics to understand heritability of coronary artery disease and other common complex disorders, J Am. Coll. Cardiol. 65 (2015) 830–845.
[3] K.J. Hunt, R. Duggirala, H.H. Goring, J.T. Williams, L. Almasy, J. Blangero, D.H. O'Leary, et al., Genetic basis of variation in carotid artery plaque in the San Antonio Family Heart Study, Stroke 33 (2002) 2775–2780.
[4] J.C. Bis, C.C. White, N. Franceschini, J. Brody, X. Zhang, D. Muzny, J. Santibanez, et al., Sequencing of 2 subclinical atherosclerosis candidate regions in 3669 individuals: cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Targeted Sequencing Study, Circ. Cardiovasc. Genet. 7 (2014) 359–364.
[5] L. Zhang, P. Buzkova, C.L. Wassel, M.J. Roman, K.E. North, D.C. Crawford, J. Boston, et al., Lack of associations of ten candidate coronary heart disease risk genetic variants and subclinical atherosclerosis in four US populations: the Population Architecture using Genomics and Epidemiology (PAGE) study, Atherosclerosis 228 (2013) 390–399.
[6] J. Euesden, C.M. Lewis, P.F. O'Reilly, PRSice: polygenic Risk Score software, Bioinformatics 31 (2015) 1466–1468.
[7] F.C. Arnett, S.M. Edworthy, D.A. Bloch, D.J. McShane, J.F. Fries, N.S. Cooper, L.A. Healey, et al., The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis, Arthritis Rheum. 31 (1988) 315–324.
[8] B.F. Voight, H.M. Kang, J. Ding, C.D. Palmer, C. Sidore, P.S. Chines, N.P. Burtt, et al., The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits, PLoS Genet. 8 (2012) e1002793.
[9] S. Buyske, Y. Wu, C.L. Carty, I. Cheng, T.L. Assimes, L. Dumitrescu, L.A. Hindorff, et al., Evaluation of the metabochip genotyping array in African Americans and implications for fine mapping of GWAS-identified loci: the PAGE study, PLoS One 7 (2012) e35651.
[10] N. Franceschini, Y. Hu, A.P. Reiner, S. Buyske, M. Nalls, L.R. Yanek, Y. Li, et al., Prospective associations of coronary heart disease loci in African Americans using the MetaboChip: the PAGE study, PLoS One 9 (2014) e113203.
[11] A. Manichaikul, J.C. Mychaleckyj, S.S. Rich, K. Daly, M. Sale, W.M. Chen, Robust relationship inference in genome-wide association studies, Bioinformatics 26 (2010) 2867–2873.
[12] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, et al., PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (2007) 559–575.
[13] A.L. Price, N.J. Patterson, R.M. Plenge, M.E. Weinblatt, N.A. Shadick, D. Reich, Principal components analysis corrects for stratification in genome-wide association studies, Nat. Genet. 38 (2006) 904–909.