Circulation
Cardiovascular Genetics
JOURNAL OF THE AMERICAN HEART ASSOCIATION
American Heart Association
Learn and Live

Candidate Gene Association Resource (CARe) : Design, Methods, and Proof of Concept

Kiran Musunuru, Guillaume Lettre, Taylor Young, Deborah N. Farlow, James P. Pirruccello, Kenechi G. Ejebe, Brendan J. Keating, Qiong Yang, Ming-Huei Chen, Nina Lapchyk, Andrew Crenshaw, Liuda Ziaugra, Anthony Rachupka, Emelia J. Benjamin, L. Adrienne Cupples, Myriam Fornage, Ervin R. Fox, Susan R. Heckbert, Joel N. Hirschhorn, Christopher Newton-Cheh, Marcia M. Nizzari, Dina N. Paltoo, George J. Papanicolaou, Sanjay R. Patel, Bruce M. Psaty, Daniel J. Rader, Susan Redline, Stephen S. Rich, Jerome I. Rotter, Herman A. Taylor, Jr, Russell P. Tracy, Ramachandran S. Vasan, James G. Wilson, Sekar Kathiresan, Richard R. Fabsitz, Eric Boerwinkle, Stacey B. Gabriel and for the NHLBI Candidate Gene Association Resource

Circ Cardiovasc Genet 2010;3;267-275; originally published online April 17, 2010;
DOI: 10.1161/CIRCGENETICS.109.882696
Circulation: Cardiovascular Genetics is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514
 Print ISSN: 1942-325X. Online ISSN: 1942-3268

Candidate Gene Association Resource (CARe)
Design, Methods, and Proof of Concept

Kiran Musunuru, MD, PhD, MPH; Guillaume Lettre, PhD; Taylor Young, MA; Deborah N. Farlow, PhD; James P. Pirruccello, BS; Kenechi G. Ejebe, BA; Brendan J. Keating, PhD; Qiong Yang, PhD; Ming-Huei Chen, PhD; Nina Lapchyk, MS; Andrew Crenshaw, MS; Liuda Ziaugra, MS; Anthony Rachupka, BS; Emelia J. Benjamin, MD, ScM; L. Adrienne Cupples, PhD; Myriam Fornage, PhD; Ervin R. Fox, MD, MPH; Susan R. Heckbert, MD, MPH, PhD; Joel N. Hirschhorn, MD, PhD; Christopher Newton-Cheh, MD; Marcia M. Nizzari, MS; Dina N. Paltoo, PhD, MPH; George J. Papanicolaou, PhD; Sanjay R. Patel, MD, MS; Bruce M. Psaty, MD, PhD; Daniel J. Rader, MD; Susan Redline, MD, MPH; Stephen S. Rich, PhD; Jerome I. Rotter, MD; Herman A. Taylor, Jr, MD, MPH; Russell P. Tracy, PhD; Ramachandran S. Vasan, MD, DM; James G. Wilson, MD; Sekar Kathiresan, MD; Richard R. Fabsitz, PhD; Eric Boerwinkle, PhD; Stacey B. Gabriel, PhD; for the NHLBI Candidate Gene Association Resource

Background—The National Heart, Lung, and Blood Institute's Candidate Gene Association Resource (CARe), a planned cross-cohort analysis of genetic variation in cardiovascular, pulmonary, hematologic, and sleep-related traits, comprises >40 000 participants representing 4 ethnic groups in 9 community-based cohorts. The goals of CARe include the discovery of new variants associated with traits using a candidate gene approach and the discovery of new variants using the genome-wide association mapping approach specifically in African Americans.

Methods and Results—CARe has assembled DNA samples for >40 000 individuals self-identified as European American, African American, Hispanic, or Chinese American, with accompanying data on hundreds of phenotypes that have been standardized and deposited in the CARe Phenotype Database. All participants were genotyped for 7 single-nucleotide polymorphisms (SNPs) selected based on prior association evidence. We performed association analyses relating each of these SNPs to lipid traits, stratified by sex and ethnicity, and adjusted for age and age squared. In at least 2 of the ethnic groups, SNPs near *CETP*, *LIPC*, and *LPL* strongly replicated for association with high-density lipoprotein cholesterol concentrations, *PCSK9* with low-density lipoprotein cholesterol levels, and *LPL* and *APOA5* with serum triglycerides. Notably, some SNPs showed varying effect sizes and significance of association in different ethnic groups.

Conclusions—The CARe Pilot Study validates the operational framework for phenotype collection, SNP genotyping, and analytic pipeline of the CARe project and validates the planned candidate gene study of ≈2000 biological candidate loci in all participants and genome-wide association study in ≈8000 African American participants. CARe will serve as a valuable resource for the scientific community. **(*Circ Cardiovasc Genet*. 2010;3:267-275.)**

Key Words: genetics ■ lipids ■ epidemiology

A key goal of biomedical research is to understand how genetic variation contributes to interindividual differences in risk for disease. Despite many years of effort, the DNA sequence variants and underlying genes that affect complex genetic diseases such as myocardial infarction or type 2 diabetes and risk factors such as blood lipoprotein levels or body weight in humans remain mostly unknown. Critical questions that remain largely unanswered include

Received June 20, 2009; accepted February 12, 2010.

From the Broad Institute (K.M., G.L., T.Y., D.N.F., J.P.P., K.G.E., N.L., A.C., L.Z., A.R., J.N.H., C.H.N.-C., S.K., S.B.G.), Cambridge, Mass; Massachusetts General Hospital (K.M., J.P.P., K.G.E., C.H.N.-C., S.K.), Boston, Mass; Harvard Medical School (K.M., C.H.N.-C., J.N.H., S.K.), Boston, Mass; Johns Hopkins University School of Medicine (K.M., J.P.P.), Baltimore, Md; Montreal Heart Institute (G.L.), Université de Montréal, Montreal, Canada; Boston University (Q.Y., M.-H.C., E.J.B., L.A.C., R.S.V.), Boston, Mass; The National Heart, Lung, Blood Institute's Framingham Heart Study (Q.Y., M.-H.C., E.J.B., L.A.C., R.S.V.), Framingham, Mass; The University of Texas Health Science Center at Houston (M.F., E.B.); University of Mississippi Medical Center (E.R.F., H.A.T., J.G.W.), Jackson, Miss; University of Washington (S.R.H., B.M.P.), Seattle, Wash; Children's Hospital (J.N.H.), Boston, Mass; National Heart, Lung, and Blood Institute (D.N.P., G.J.P., R.R.F.), National Institutes of Health, Bethesda, Md; University Hospitals, Case Medical Center and Case Western Reserve University (S.R.P., S.R.), Cleveland, Ohio; Group Health Research Institute (B.M.P.), Group Health Cooperative, Seattle, Wash; University of Virginia School of Medicine (S.S.R.), Charlottesville, Va; Medical Genetics Institute (J.I.R.), Cedars-Sinai Medical Center, Los Angeles, Calif; and The University of Vermont College of Medicine (R.P.T.), Burlington, Vt.

The online Data Supplement is available at http://circgenetics.ahajournals.org/cgi/content/full/CIRCGENETICS.109.882696/DC1.

Guest Editor for this article was Donna K. Arnett, PhD.

Dr Kiran Musunuru, Dr Guillaume Lettre, Taylor Young, and Dr Deborah N. Farlow contributed equally to this work.

Correspondence to Stacey B. Gabriel, PhD, Broad Institute, 5 Cambridge Center, Cambridge, MA 02142. E-mail stacey@broadinstitute.org

Circ Cardiovasc Genet **is available at http://circgenetics.ahajournals.org** **DOI: 10.1161/CIRCGENETICS.109.882696**

which biological pathways are altered in patients in a manner that contributes causally to disease and might therefore be the best targets for interventions and therapies.

Clinical Perspective on p 275

Genetic association studies, both genome-wide association and those based on biological candidates, have proven to be robust tools to discover genes associated with disease processes, potentially leading to novel therapeutics, personalized medicine, and preventive programs. With genotyping efforts and association studies being conducted at an increasing number of institutions, it has become critical to establish collaborations or a centralized database where shared resources for analyses of the association of genotypes with phenotypes relevant to the biomedical community are located. The majority of large-scale genetic studies that have been completed have focused on whites of European ancestry and have had limited representation of other ethnic groups; specifically, there has been a paucity of large-scale population-based studies that have included African Americans, Hispanics, and Chinese Americans. The application of candidate gene and genome-wide association approaches to a large number of individuals from a variety of ethnic groups promises to comprehensively advance our understanding of the heritability and biology of many diseases and traits.

The Candidate Gene Association Resource (CARe), a unique initiative of the National Heart, Lung, and Blood Institute (NHLBI), seeks to assemble well-characterized phenotypic data from 9 NHLBI cohorts to generate new genotype data across candidate genes and/or the whole genome and to perform genotype-phenotype association analyzes across >100 cardiovascular, pulmonary, hematologic, and sleep-related traits available from the different cohorts. The 9 cohorts include the Atherosclerosis Risk in Communities (ARIC) study,[1] the Coronary Artery Risk Development in Young Adults study,[2] the Cleveland Family Study (CFS),[3] the Cardiovascular Health Study,[4] the Cooperative Study of Sickle Cell Disease,[5] the Framingham Heart Study (FHS),[6–8] the Jackson Heart Study,[9] the Multi-Ethnic Study of Atherosclerosis,[10] and the Sleep Heart Health Study[11] (Table 1).

The overall goals of CARe include (1) the discovery of new variants associated with traits using a candidate gene approach; (2) the discovery of new variants using the genome-wide association mapping approach in ≈8000 African Americans across the cohorts; (3) characterization of validated variants across clinical subgroups stratified by ethnicity, sex, or clinical covariates; and (4) exploration of gene-environment interactions. CARe comprises 2 major components—candidate gene studies in the combined population of >40 000 individuals and genome-wide association studies in ≈8000 African American participants.

We conducted the CARe Pilot Study to validate the operational framework of CARe with respect to the phenotype collection, single-nucleotide polymorphism (SNP) genotyping, and construction of an analytic pipeline for genotype-phenotype association studies by (1) studying a pilot set of SNPs originally identified in European-derived cohorts and (2) assessing whether previously reported associations of SNPs with select traits in European Americans would replicate in other ethnic groups, namely African Americans, Hispanics, and Chinese Americans. The latter exercise will set the stage for multiethnic genetic association studies (CARe-wide and beyond). In the CARe Pilot Study, we analyzed 7 SNPs with prior evidence of association (Table 2) with respect to plasma levels of high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), or triglycerides.

Methods

CARe Study Conception and Design

The CARe Study was initiated by the NHLBI in 2006. DNA samples and phenotypic information from the 9 NHLBI cohorts—ARIC, Coronary Artery Risk Development in Young Adults study, CFS, Cardiovascular Health Study, Cooperative Study of Sickle Cell Disease, FHS, Jackson Heart Study, Multi-Ethnic Study of Atherosclerosis, and Sleep Heart Health Study—were sent to the Broad Institute of Massachusetts Institute of Technology and Harvard and deposited in the CARe Phenotype Database. The overall CARe procedures include (1) assembly of phenotypes into a single database; (2) genotyping of assembled DNA samples using 3 platforms: Sequenom for the Pilot Study, the ITMAT-Broad-CARe array for candidate gene studies, and the Affymetrix 6.0 array for genome-wide association studies; (3) quality control (QC) procedures on genotype data; (4) establishment of a secure data repository to allow approved researchers to access phenotype and genotype data to pursue hypothesis testing (while meticulously maintaining full confidentiality for the study participants); (5) statistical modeling of the phenotypes of interest; (6) statistical analyses to identify associations between genotypes and phenotypes of interest; and (7) use of existing, public, Internet-based resources to disseminate summary data from the genome-wide association studies (GWAS) and candidate gene studies to the wider scientific community (Figure 1). The availability of these data to the scientific community will (a) maximize the scientific utility to be gained from the investment in sample collection, genotyping, and phenotyping, (b) facilitate the replication and extension of CARe-derived results in other cohorts, and (c) foster the development of additional analytic and computational methods that can be tested on a large-scale genetic data set.

Governance

The CARe project is overseen by a Steering Committee comprising representatives from each of the 9 NHLBI cohorts as well as the Broad Institute of Massachusetts Institute of Technology and Harvard and the University of Pennsylvania; under the steering committee are a number of subcommittees and working groups that are responsible for the execution of various aspects of the CARe project (supplemental Figure I). NHLBI staff members participate in the meetings and teleconferences of the various committees and groups. An Oversight Committee appointed by the NHLBI monitors the progress of the CARe project.

Phenotype Database Construction

The phenotypes that were assembled for CARe include phenotypes collected at the baseline and follow-up examinations of participants in each of the 9 NHLBI cohorts. Each CARe cohort was contacted to contribute the phenotypes, and once each cohort data set was received, they were deposited in the CARe Phenotype Database. Available phenotypes that have been cataloged in the CARe Phenotype Database range from hundreds to many thousands, depending on the cohort. Descriptions of phenotype groupings and examples of phenotypes are available in Table 3 and at the CARe website: http://www.broad.mit.edu/gen_analysis/care/index.php/Main_Page. When they are requested for use by investigators, phenotypes are to be standardized according to the Clinical Data Interchange Standard Consortium Study Data Tabulation Model available at the website: http://www.cdisc.org/models/sds/v3.1/index.html.

Three of these phenotypes were used for the Pilot Study—HDL-C, LDL-C, and triglycerides. Values of these traits (except LDL-C, which was calculated; see below) determined at baseline visits for each cohort

Table 1. Study Design and Participant Characteristics*

Study	ARIC		JHS	MESA				CSSCD	CFS		FHS				CHS		CARDIA	
Ancestry composition	AA	EA	AA	AA	EA	HIS	CHI	AA	AA	EA	Gen. 1 EA	Gen. 2 EA	Gen. 3 EA	All EA	AA	EA	AA	EA
n (phenotypes)	4141	11 412	2702	1754	2537	1475	778	3205	721	722	2511	3765	4082	10 358	842	4653	2637	2478
Ascertainment scheme	Community-based		Community-based	Population-based				Patient-based	Family-based		Community-based	Community-based	Community-based	Community-based	Community-based		Community-based	
Recruitment period	1987–1998		2000–2004	2000–2007				1979–1988†	1990–2005		1948–1953	1971–1975	2002–2005		1989–1993		1985–1986	
Mean age (y)	54±6	54±6	52±13	62±10	63±10	61±10	62±10	17±20	39±20	42±20	69±8	44±10	40±9	49±15	73±6	73±6	24±4	25±3
Female sex, %	62	53	62	55	52	52	51	51	57	52	60	52	53	54	63	56	56	53
Total cholesterol, mg/dL	215±45	215±50	197±39	190±36	196±35	198±37	192±31	. . .	171±42	181±40	229±42	204±39	189±35	204±41	210±40	211±39	177±34	176±32
LDL cholesterol, mg/dL	137±43	138±40	125±36	116±33	117±30	120±33	115±29	. . .	96±33	101±30	. . .	130±35	112±32	120±34	129±37	130±36	110±32	108±30
HDL cholesterol, mg/dL	55±18	50±17	51±15	52±15	52±16	48±13	49±12	. . .	48±15	46±13	50±16	48±13	54±16	51±15	60±16	53±16	54±13	51±13
Triglycerides, mg/dL	114±81	138±93	109±98	105±69	133±90	157±101	143±85	. . .	102±62	143±118	. . .	. . .	116±90	112±95	116±63	144±79	67±38	79±57
Systolic BP, mm Hg	129±22	119±17	126±18	131	124±21	127±22	125±22	106±16	123±18	121±16	137±19	122±17	117±14	124±18	142±23	136±22	111±11	109±11
Hypertension, %	58	33	58	62	44	44	41	4	38	29	56	23	11	26	78	64	5	3
Type 2 diabetes mellitus, %*	20	9	18	18	6	18	14	0.3	14	9	9	3	3	4	25	15	1	.5
Body mass index, kg/m^2	29.6±6	27.0±5	31.9±7	30.0±6	27.7±5	29.4±5	23.9±3	18.9±9	32.0±10	30.1±9	26.2±4	25.6±4	26.9±6	26.0±5	28.0±6	26.3±4	25.2±6	23.5±4
Lipid-lowering therapy, %	1	3	11	16	18	13	14	. . .	8	8	3	1	7	4	7	5	. . .	. . .
Antihypertensive therapy, %	44	26	49	50	33	33	29	1	32	23	32	10	. . .	19	63	45	1	1
Reference for study design	1		9	10				5	3		6 to 8				4		2	

Values are presented as mean±SD unless otherwise indicated. Hypertension was defined as systolic blood pressure ≥140 mm Hg, diastolic blood pressure ≥90 mm Hg, or use of antihypertensive therapy. AA indicates African American; EA, European American; HIS, Hispanic; CHI, Chinese American; CARDIA, Coronary Artery Risk Development in Young Adults study; CHS, Cardiovascular Health Study; CSSCD, Cooperative Study of Sickle Cell Disease; JHS, Jackson Heart Study; MESA, Multi-Ethnic Study of Atherosclerosis; and SHHS, Sleep Heart Health Study.

*The SHHS[11] is not represented in this table because the subgroup of the SHHS cohort with genotype data comprises individuals originally recruited from ARIC, CHS, and FHS (and who are included in this table under these other cohorts). A number of individuals are participants in both ARIC and JHS; in this table, they are included only in ARIC.

†1979 to 1981, >6 mo of age; 1979 to 1988, <6 mo of age.

Table 2. Pilot SNPs and Associated Phenotypes Tested in Design Study

Gene	Symbol	Polymorphism	Associated Phenotype	rs No.	CEU MAF	YRI MAF	Reference
Apolipoprotein A5	*APOA5*	Ser19Trp	Triglycerides	rs3135506	0.06	0.05	12
Cholesteryl ester transfer protein	*CETP*	C–1337T	HDL-C	rs17231506	0.37	0.03	13
Cholesteryl ester transfer protein	*CETP*	G–971A	HDL-C	rs4783961	0.46	0.42	14
Hepatic lipase	*LIPC*	C–480T	HDL-C	rs1800588	0.26	0.47	15
Lipoprotein lipase	*LPL*	Ser447X/Ser474X	HDL-C, triglycerides	rs328	0.13	0.03	16
Proprotein convertase subtilisin/kexin type 9	*PCSK9*	Arg46Leu	LDL-C	rs11591147	0.00	0.00	17
Proprotein convertase subtilisin/kexin type 9	*PCSK9*	Glu670Gly	LDL-C	rs505151	0.04	0.31	18

CEU indicates European descent (HapMap); YRI, Yoruban (HapMap).

were used. Of note, the subgroup of the Sleep Heart Health Study cohort with genotype data comprises individuals originally recruited from ARIC, Cardiovascular Health Study, and FHS and subsequently evaluated for sleep phenotypes. In this article, Sleep Heart Health Study participants appear under their parent cohorts.

Genotyping and QC

CARe Pilot (Sequenom)

After appropriate Material Transfer Agreements were in place, samples were shipped from each cohort's central genetics laboratory to the Broad Institute of Massachusetts Institute of Technology and Harvard. DNA concentration was determined by the Picogreen assay (Invitrogen, Carlsbad, Calif) before storage in 2D bar-coded 0.75 mL Matrix tubes at −20°C in the SmaRTStore (RTS, Manchester, UK) automated sample handling system. Seven SNPs were selected for the CARe Pilot Study based on previously published evidence of association (see Table 2 for SNPs and references). These SNPs were genotyped using the Sequenom MassArray System platform (Sequenom, San Diego, Calif). All DNA samples passing initial quality checks were plated at a concentration of 5 ng/μL for processing on the platform. Sequenom SNP genotyping uses beadless and label-free primer-extension chemistry in a multiplex format to generate allele-specific products with distinct masses that are distinguished by mass spectroscopy. The results were automatically loaded into a database and then scored using SpectroTyper 4.0 Software (Sequenom) and uploaded to the laboratory information management system and data storage system.

Several QC procedures were performed on the genotype data separately for each cohort (supplemental Table I). Sample duplicates were identified using sample IDs. For each set of duplicates or monozygotic twins, data from the sample with the highest genotyping success rate were retained. Reported sex and genotype-inferred sex (2 independent Sequenom assays for each sample) were compared for concordance. All discordant samples and samples for which no sex information was available were resolved in consultation with the relevant cohort or excluded. SNPs with a missing data rate >10% and samples with a genotyping success rate <90% were removed. Only samples with available phenotypic information were used for association studies. In the cohorts with available family information—CFS and FHS—data of descendants from families accounting for the most mendelian errors in the data set were excluded. Because several different ethnic groups were represented, with the expectation of differing genotype frequencies and admixture, no filters were applied for minor allele frequency (MAF) or Hardy-Weinberg probability values. All QC analyses were performed in PLINK.[19]

Candidate Gene Studies (Illumina iSelect–ITMAT-Broad-CARe Chip)

The design of the ITMAT-Broad-CARe Chip, a custom 50K SNP genotyping array, has been described recently.[20] The SNPs (49 320

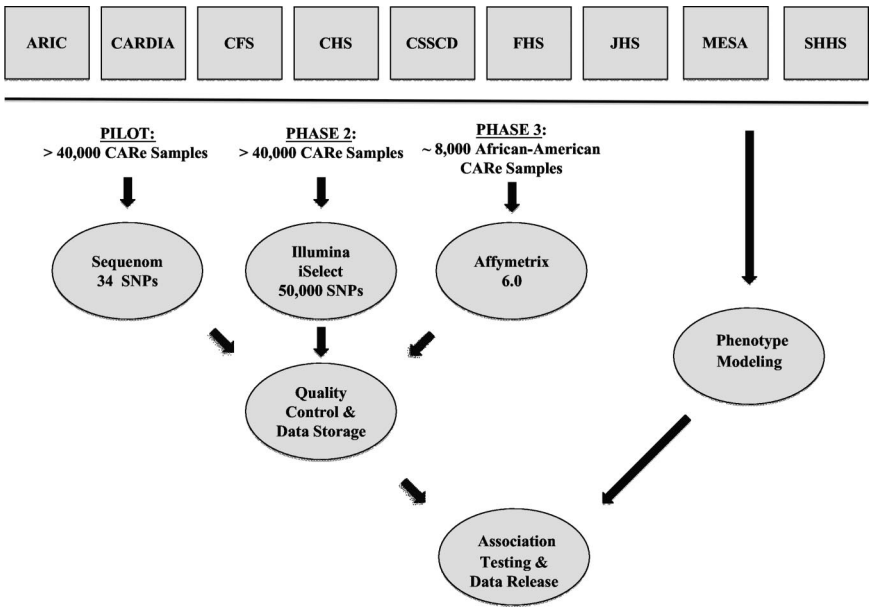

Figure 1. Operational framework of CARe.

Table 3. Phenotype Categories Available in CARe Database

Category	Example Phenotypes
Aging	Age at examination
	Age at menopause
	Age at birth of first child
	Age at birth of last child
Anthropometry	Height
	Weight
	Body mass index
	Waist circumference
Atrial fibrillation/ electrocardiography	Atrial fibrillation
	PR interval
	QT interval
Blood biomarkers	C-reactive protein
	MCP-1
	IL-6
	Fibrinogen
	Factor VII
	Blood cell counts
Blood pressure/hypertension	Systolic blood pressure
	Diastolic blood pressure
	Pulse pressure
	Mean arterial pressure
	Hypertension
	Hypertension medications
Coronary heart disease	Coronary heart disease
Diabetes mellitus/glucose	Type 2 diabetes mellitus
	Age at diabetes diagnosis
	Diabetes medications
	Diabetic retinopathy
	Hemoglobin A1C
	Fasting blood glucose
	Fasting insulin
	Insulin use
Echocardiography/congestive heart failure	Heart failure
	Left ventricular mass
	Left ventricular ejection fraction
	Right ventricular ejection fraction
Kidney disease	Creatinine
	Estimated glomerular filtration rate
	Cystatin C
	Urinary albumin:creatinine ratio
Lipids	Total cholesterol
	HDL
	LDL
	Triglycerides
	Apo B
	Apo A-I
	Lipid-lowering therapy
Musculoskeletal	Gout
	Serum urate
Peripheral arterial disease	Peripheral arterial disease
	Ankle-brachial index
Pulmonary function	FEV_1
	FVC
	$FEF_{25–75}$
Sleep	Sleep apnea
	Snoring
	Epworth Sleepiness Scale
	Excessive daytime sleepiness
Smoking	Smoking status
	Pack-year history
	Smoking intensity
Stroke	Incident stroke
	Ischemic stroke
Subclinical atherosclerosis	Coronary artery calcification
	Carotid intima-media thickness

total) were chosen to densely map about 2000 candidate gene loci deemed to be relevant to phenotypes available in the CARe Phenotype Database. All DNA samples passing initial quality checks were interrogated with the ITMAT-Broad-CARe chip. Analyses with this genotype data are not included in this article but rather will be the focus of future CARe studies.

GWAS (Affymetrix 6.0 Array Set)

Approximately 8000 African American participants from 5 of the CARe cohorts—ARIC, Jackson Heart Study, Coronary Artery Risk Development in Young Adults study, CFS, and Multi-Ethnic Study of Atherosclerosis—were genotyped with the Affymetrix 6.0 (“million-SNP”) Array Set (Affymetrix, Santa Clara, Calif), typing more than 906 600 SNPs and 946 000 probes for copy number variation across the genome. The genotype data so obtained will be used for GWAS on phenotypes of interest. Analyses with this genotype data are not included in this article but rather will be the focus of future CARe studies.

Data Management

A key goal of CARe is to prepare a comprehensive genotype and phenotype data set that serves as a scientific resource that is broadly accessible to the research community. This was performed taking care to protect the confidentiality and interests of study participants and consistent with the informed consent procedures in each of the cohorts. The institutional review boards of each CARe cohort (ie, the institutional review boards for each cohort’s field centers, coordinating center, and laboratory center) have reviewed the cohort’s interaction with CARe. CARe itself has been approved by the Committee on the Use of Humans as Experimental Subjects of the Massachusetts Institute of Technology. Identifiers were removed and codes were assigned to any protected health information transmitted to the CARe Data Repository, with a Certificate of Confidentiality issued by the National Institutes of Health. The Data Repository will release limited data sets to qualifying investigators whose projects have been approved by their local institutional review boards and who have completed a CARe Data Distribution Agreement. Each such data set will have its own unique, randomly generated set of participant identifiers.

Of note, at the time of study, not all subjects had provided specific consent for data to be made available to non-CARe investigators. Accordingly, on request, public access to the data will be provided to the extent that the informed consent process allows.

Phenotype Modeling

For the CARe Pilot Study, we modeled the pilot phenotypes in the following ways. LDL-C was calculated according to the Friedewald

formula: LDL-C=total cholesterol−HDL-C−(triglycerides/5). If a triglyceride value was >400 mg/dL, LDL-C was treated as a missing value. For individuals on lipid-lowering therapy, the LDL-C value was multiplied by 1.42 to model a 30% reduction in LDL-C on therapy. This represents the average expected reduction in LDL-C with a first-generation statin, the most commonly used lipid-lowering medication during the study periods of most of the cohorts.[21] Triglyceride values were log(10) transformed. Sex-specific phenotype residuals were constructed within strata of cohort and ethnicity with adjustment for age and age squared in each individual stratum. Each set of residuals was standardized to a mean of 0 and SD of 1. The standardized residual served as the phenotype in genotype-phenotype association analyses.

Generation of residuals was performed with the R statistical package (The R Foundation for Statistical Computing, Vienna, Austria).

Association Testing

Cohorts were divided into subgroups by ethnicity; association analyses were performed for each subgroup, followed by meta-analysis of the subgroups for each ethnicity. For cohorts in which individuals were largely unrelated or when family information was not available (ARIC, Coronary Artery Risk Development in Young Adults Study, Cardiovascular Health Study, Cooperative Study of Sickle Cell Disease, Jackson Heart Study, and Multi-Ethnic Study of Atherosclerosis) we used linear regression to test SNP-phenotype associations assuming an additive genetic model. These association analyses were performed in PLINK. For the 2 cohorts for which there were significant numbers of related individuals, and for which family information was available at the time of the Pilot Study (CFS and FHS) we used a linear mixed effects (LME) model to analyze the traits, with the SNP genotype treated as a fixed effect and a random effect according to the degree of relatedness within a family.[22] Genotype-phenotype associations within each ethnic group were assessed by variance-weighted meta-analyses, and heterogeneity within each ethnic group or between ethnic groups was assessed using the Cochran Q statistic and/or the I^2 inconsistency metric.[23]

For future GWAS and candidate gene studies, association analyses will include procedures to account for the effects of local and global ancestry, particularly with regard to African American subjects. Given the small number of SNPs addressed in the Pilot Study, such procedures were not feasible.

The authors had full access to the data and take responsibility for its integrity. All authors have read and agree to the article as written.

Results

Baseline Characteristics of CARe Cohorts

Characteristics of each individual cohort's study participants with respect to the pilot phenotypes are presented in Table 1. (See also supplemental Table I.) At the time of the Pilot Study, the combined sample that had been received and processed at the Broad Institute and undergone successful genotyping and association analyses included 40 324 total individuals of whom 26 647 were European Americans, 11 550 African Americans, 1410 Hispanics, and 717 Chinese Americans. The cohorts varied with regard to the mean age, the proportions with T2D, and the body mass index of the participants. Mean lipid and systolic blood pressure values were similar across cohorts. Of note, lipid phenotypes were not available for Cooperative Study of Sickle Cell Disease individuals and thus these individuals were not included in analyses for HDL-C, LDL-C, and triglycerides.

Evidence for Admixture in African American Cohorts

We compared MAFs for each of the pilot SNPs in the African American cohorts and European American cohorts (supplemental Table II) with the MAFs for the SNPs in the Yoruba and European-descended (CEU) groups in the International HapMap Project (Table 2).[24] These comparisons suggest that all African American cohorts have a significant degree of admixture of African and European chromosomes. For example, for rs17231506 in the *CETP* locus, the MAFs range from 10% to 16% in the African American cohorts, whereas MAFs for the European American cohorts range from 31% to 33%. The European American cohorts' MAFs are consistent with the CEU MAF of 37%, whereas the African American cohorts' MAFs are intermediate between those of CEU and Yoruba (3%). This is generally true for all the pilot SNPs whose MAFs differ greatly between the CEU and Yoruba groups.

Pilot Phenotype-Genotype Association Results

We performed association analyses for each of the 7 pilot SNPs against each of the relevant pilot phenotypes (ie, the phenotypes with which they had previously been shown to have association; Table 2) in each ethnic group within each cohort, assuming additive genetic models. The results of the analyses are presented in Figure 2 and supplemental Table II. To account for multiple testing, we considered a probability value of 1×10^{-4} to represent the threshold of statistical significance. We performed meta-analyses and heterogeneity analyses for each SNP-phenotype combination (supplemental Table II).

An SNP in the *CETP* gene, rs17231506, was associated with HDL-C in all 4 ethnic groups: African Americans, European Americans, Hispanics, and Chinese Americans. Another SNP in *CETP*, rs4783961, was associated with HDL-C in African Americans, European Americans, and Chinese Americans. Other HDL-C–associated SNPs were rs1800588 in *LIPC* and rs328 in *LPL*, both in African Americans and European Americans. Similarly, triglyceride-related SNPs were replicated in both African Americans and European Americans: rs328 in *LPL* and rs3135506 in *APOA5*. LDL-C–related SNPs were associated in either African Americans or European Americans but not both; rs505151 in *PCSK9* was associated in African Americans but not in European Americans, whereas rs11591147 in *PCSK9* was associated with LDL-C in European Americans but not in African Americans.

We found that within each ethnicity, there was low heterogeneity of effect of each SNP on each trait (supplemental Table II). For example, for the SNP rs4783961 in *CETP* and HDL-C, the direction of the association was consistent across all the cohorts, with the G allele representing higher HDL-C levels. The effect sizes (β coefficients) associated with this allele were remarkably consistent across the African American cohorts (ranging from 0.17 to 0.24) and across the European American cohorts (0.09 to 0.15). Formal heterogeneity analyses showed very low heterogeneity among the 2 sets of cohorts, with the I^2 inconsistency metric being 0% for each set. Indeed, for all the statistically significant SNP-phenotype associations, in many cases, the *I2* metric was 0%, in no case exceeding 50%; similarly, in every case, the Cochran Q probability value for heterogeneity was nonsignificant ($P>0.05$).

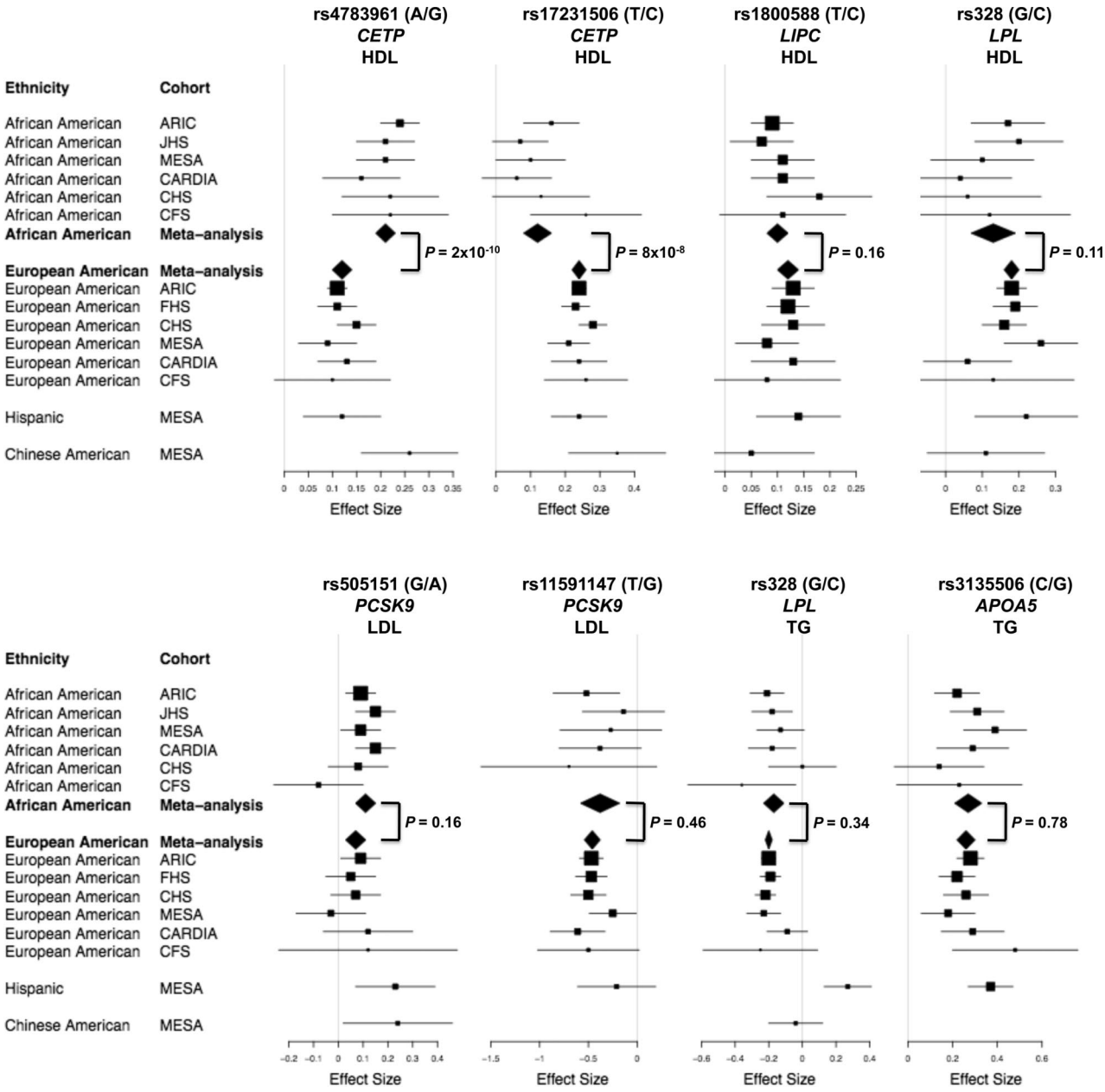

Figure 2. Forest plots for SNP-phenotype effects, with Cochran *Q* probability values.

There were notable differences in effect sizes across ethnic groups for some of the SNPs (Figure 2). For example, at rs4783961 in *CETP* and HDL-C, the effect size was uniformly larger in African Americans (0.17 to 0.24) than in European Americans (0.09 to 0.15; $P=2\times10^{-10}$ by Cochran *Q* heterogeneity test). The effect size for this variant in Hispanics (0.12) was more similar to European Americans and in Chinese Americans (0.26) with African Americans. Unlike rs4783961, the HDL-C–associated SNP rs17231506, also in *CETP*, had larger effect sizes in European Americans and Hispanics (0.21 to 0.28) compared with African Americans (0.06 to 0.26; $P=8\times10^{-8}$ between European Americans and African Americans), although smaller than in Chinese Americans (0.35). Thus, even though the same gene locus (*CETP*) was highly associated with HDL-C across the ethnicities, there were differences between ethnicities in the contributions of individual SNPs at the locus to interindividual variation in HDL-C levels.

We observed interethnic differences with other gene loci vis-à-vis other phenotypes. The rs505151 SNP in the *PCSK9* locus had stronger statistical association with LDL-C in African Americans than in European Americans, although this is attributable to the higher MAF in African Americans (≈25%) than in European Americans (≈4%).

Other SNPs seem to affect phenotypes to similar degrees across ethnic groups: rs1800588 in *LIPC* with HDL-C; rs3135506 in *APOA5* with triglycerides; and rs328 in *LPL* with both HDL-C and triglycerides.

Discussion

The CARe Pilot Study was designed to evaluate the operational framework established for the resource, including

phenotype collection and integration, routing of a large number of DNA samples for genotyping analyses, QC procedures on all the collected data, data analysis, and synthesis of the results. Our ability to obtain robust phenotype-genotype associations for SNPs with strong prior evidence in the published literature validates the CARe study framework and sets the stage for the candidate gene and GWAS discovery phases of CARe that are in progress. The Pilot Study also afforded the scientific opportunity to evaluate the effects of DNA variants on clinically important traits in the largest group of African American individuals with genotype information assembled to date. The critical findings from the pilot phenotype-genotype associations were that (1) there was low heterogeneity for highly associated SNPs within cohorts of the same ethnicity; (2) for some associated SNPs, there were interethnic differences in effect size; and (3) each gene replicated in multiple ethnic groups, although not necessarily through the same SNPs.

The finding of low heterogeneity with the lipid traits suggests that despite the expected variation in phenotypes among the cohorts because of the use of different assays or different disease definitions, at least some of the phenotypes that have been collected in CARe from different cohorts can be standardized and, on appropriate analysis, yield meaningful scientific results.

We found that for several SNPs in replicated gene loci—most notably in *CETP* and *PCSK9*—the associations with phenotype differed between African Americans and European Americans. For example, although 1 *CETP* SNP (rs4783961) had a larger effect size in African American cohorts, another SNP in the same locus (rs17231506) showed a larger effect size in European American cohorts. One possible explanation for this observation is that African Americans and European Americans share the same causal variant in a gene, but because of ethnicity-specific differences in the major and minor allele frequencies, an SNP may have differing strength of correlation with the causal variant, which manifests as varying effect sizes and the degree of association. In some cases, the interethnic differences are observed despite similar allele frequencies (eg, rs4783961 in African Americans and European Americans). An alternative explanation is that interethnic differences in the linkage disequilibrium patterns in the gene locus result in SNPs having differing correlations with the casual variants. A third possibility is that different causal variants in the same gene predominate in the ethnic groups, with different SNPs in the locus linked with the variants.

This last possibility seems to be the case for *PCSK9* in which the SNP that is strongly associated with LDL-C in European Americans (rs11591147) has a weak association in African Americans and vice versa (rs505151). Both SNPs are coding variants that are likely to be casual for the effect on LDL-C, and the explanation for the strong association in the one ethnic group and the weak association in the other group is that the SNP has a lower MAF in the latter group. For example, rs505151 has a MAF of about 25% in the African American cohorts but only 4% in the European American cohorts. This finding with *PCSK9* and LDL-C suggests that for some proportion of lipid-associated loci, the specific SNPs related to lipids will differ among ethnic groups—in contrast with *CETP*, *LIPC*, *LPL*, and *APOA5*—for which the same SNPs replicated in multiple ethnic groups in this study. Future CARe analyses, particularly large-scale candidate gene studies, will be useful in assessing the prevalence of this phenomenon.

Conclusion

CARe represents the successful assembly of DNA samples and phenotype data from more than 40 000 participants in 9 NHLBI cohort studies into a unique, valuable, publicly available resource to test an array of genes for a variety of phenotypes. The Pilot Study validates operational framework of CARe and provides an initial evaluation of interethnic differences for selected SNP-phenotype relationships. The ongoing large-scale candidate gene and genome-wide association analyses in CARe will explore the contribution of genetic variation to interindividual, interethnic, age-related, and cohort-specific differences in cardiovascular, pulmonary, hematologic, and sleep-related phenotypes. Thus, CARe should serve as a valuable resource for the scientific community.

Acknowledgments

We thank the National Heart, Lung, and Blood Institute and the research institutions, study investigators, and field staff for their support in creating this resource for biomedical research. We also thank the study participants, without whom this endeavor would not have been possible.

Sources of Funding

The following 9 parent studies, funded by the listed National Institutes of Health grants, have contributed parent study data, ancillary study data, and DNA samples through the Broad Institute (N01-HC-65226) to create this genotype/phenotype database for wide dissemination to the biomedical research community: Atherosclerotic Risk in Communities: University of North Carolina at Chapel Hill (N01-HC-55015, N01-HC-55018), Baylor Medical College (N01-HC-55016), University of Mississippi Medical Center (N01-HC-55021), University of Minnesota (N01-HC-55019), Johns Hopkins University (N01-HC-55020), University of Texas, Houston (N01-HC-55022); Cardiovascular Health Study: University of Washington (N01-HC-85079, N01-HC-55222, U01-HL-080295), Wake Forest University (N01-HC-85080), Johns Hopkins University (N01-HC-85081, N01-HC-15103), University of Pittsburgh (N01-HC-85082), University of California, Davis (N01-HC-85083), University of California, Irvine (N01-HC-85084), New England Medical Center (N01-HC-85085), University of Vermont (N01-HC-85086), Georgetown University (N01-HC-35129), University of Wisconsin (N01-HC-75150); Cleveland Family Study: Case Western Reserve University (R01-HL-46380, M01-RR-00080); Cooperative Study of Sickle Cell Disease: University of Illinois (N01-HB-72982, N01-HB-97062), Howard University (N01-HB-72991, N01-HB-97061), University of Miami (N01-HB-72992, N01-HB-97064), Duke University (N01-HB-72993), George Washington University (N01-HB-72994), University of Tennessee (N01-HB-72995, N01-HB-97070), Yale University (N01-HB-72996, N01-HB-97072), Children's Hospital-Philadelphia (N01-HB-72997, N01-HB-97056), University of Chicago (N01-HB-72998, N01-HB-97053), Medical College of Georgia (N01-HB-73000, N01-HB-97060), Washington University (N01-HB-73001, N01-HB-97071), Jewish Hospital and Medical Center of Brooklyn (N01-HB-73002), Trustees of Health and Hospitals of the City of Boston, Inc (N01-HB-73003), Children's Hospital-Oakland (N01-HB-73004, N01-HB-97054), University of Mississippi (N01-HB-73005), St Luke's Hospital-New York (N01-HB-73006), Alta Bates-Herrick Hospital (N01-HB-97051), Columbia University (N01-HB-97058), St Jude's Children's Research Hospital (N01-HB-97066), Research Foundation, State University of New York-Albany (N01-HB-97068, N01-HB-97069), New England

Research Institute (N01-HB-97073), Interfaith Medical Center-Brooklyn (N01-HB-97085); Coronary Artery Risk in Young Adults: University of Alabama at Birmingham (N01-HC-48047, N01-HC-95095), University of Minnesota (N01-HC-48048), Northwestern University (N01-HC-48049), Kaiser Foundation Research Institute (N01-HC-48050), Tufts-New England Medical Center (N01-HC-45204), Wake Forest University (N01-HC-45205), Harbor-UCLA Research and Education Institute (N01-HC-05187), University of California, Irvine (N01-HC-45134, N01-HC-95100); Framingham Heart Study: Boston University (N01-HC-25195, R01-HL-092577, R01-HL-076784, R01-AG-028321); Jackson Heart Study: Jackson State University (N01-HC-95170), University of Mississippi (N01-HC-95171), Tougaloo College (N01-HC-95172); Multi-Ethnic Study of Atherosclerosis: University of Washington (N01-HC-95159), University of California, Los Angeles (N01-HC-95160), Columbia University (N01-HC-95161), Johns Hopkins University (N01-HC-95162, N01-HC-95168), University of Minnesota (N01-HC-95163), Northwestern University (N01-HC-95164), Wake Forest University (N01-HC-95165), University of Vermont (N01-HC-95166), New England Medical Center (N01-HC-95167), Harbor-UCLA Research and Education Institute (N01-HC-95169), Cedars-Sinai Medical Center (R01-HL-071205), University of Virginia (subcontract to R01-HL-071205); Sleep Heart Health Study: Johns Hopkins University (U01-HL-064360), Case Western University (U01-HL-063463), University of California, Davis (U01-HL-053916), University of Arizona (U01-HL-053938, U01-HL-053934), University of Pittsburgh (U01-HL-077813), Boston University (U01-HL-053941), MedStar Research Institute (U01-HL-063429), Johns Hopkins University (U01-HL-053937).

Disclosures

None.

References

1. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives: the ARIC Investigators. *Am J Epidemiol.* 1989;129:687–702.
2. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Hulley SB, Jacobs DR Jr, Liu K, Savage PJ. CARDIA: study design, recruitment, and some characteristics of the examined subjects. *J Clin Epidemiol.* 1988;41:1105–1116.
3. Buxbaum SG, Elston RC, Tishler PV, Redline S. Genetics of the apnea hypopnea index in Caucasians and African Americans, I: segregation analysis. *Genet Epidemiol*. 2002;22:243–253.
4. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A, O'Leary DH, Psaty B, Rautaharju P, Tracy RP, Weiler PG. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1:263–276.
5. Gaston M, Smith J, Gallagher D, Flournoy-Gill Z, West S, Bellevue R, Farber M, Grover R, Koshy M, Ritchey AK. Recruitment in the Cooperative Study of Sickle Cell Disease (CSSCD). *Control Clin Trials.* 1987;8(4 suppl):131S–140S.
6. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health*. 1951;41:279–281.
7. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study: design and preliminary data. *Prev Med*. 1975;4:518–525.
8. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB Sr, Fox CS, Larson MG, Murabito JM, O'Donnell CJ, Vasan RS, Wolf PA, Levy D. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am J Epidemiol*. 2007;165:1328–1335.
9. Taylor HA Jr. The Jackson Heart Study: an overview. *Ethn Dis.* 2005;15(4 suppl 6):S6-1–S6-3.
10. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR Jr, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156:871–881.
11. Quan SF, Howard BV, Iber C, Kiley JP, Nieto FJ, O'Connor GT, Rapoport DM, Redline S, Robbins J, Samet JM, Wahl PW. The Sleep Heart Health Study: design, rationale, and methods. *Sleep*. 1997;20:1077–1085.
12. Pennacchio LA, Olivier M, Hubacek JA, Krauss RM, Rubin EM, Cohen JC. Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. *Hum Mol Genet*. 2002;11:3031–3038.
13. Frisdal E, Klerkx AH, Le Goff W, Tanck MW, Lagarde JP, Jukema JW, Kastelein JJ, Chapman MJ, Guerin M. Functional interaction between -629C/A, -971G/A and -1337C/T polymorphisms in the CETP gene is a major determinant of promoter activity and plasma CETP concentration in the REGRESS Study. *Hum Mol Genet*. 2005;14:2607–2618.
14. Le Goff W, Guerin M, Nicaud V, Dachet C, Luc G, Arveiler D, Ruidavets JB, Evans A, Kee F, Morrison C, Chapman MJ, Thillet J. A novel cholesteryl ester transfer protein promoter polymorphism (-971G/A) associated with plasma high-density lipoprotein cholesterol levels: interaction with the TaqIB and -629C/A polymorphisms. *Atherosclerosis*. 2002;161:269–279.
15. Guerra R, Wang J, Grundy SM, Cohen JC. A hepatic lipase (LIPC) allele associated with high plasma concentrations of high density lipoprotein cholesterol. *Proc Natl Acad Sci USA*. 1997;94:4532–4537.
16. Hata A, Robertson M, Emi M, Lalouel JM. Direct detection and automated sequencing of individual alleles after electrophoretic strand separation: identification of a common nonsense mutation in exon 9 of the human lipoprotein lipase gene. *Nucleic Acids Res*. 1990;18:5407–5411.
17. Kotowski IK, Pertsemlidis A, Luke A, Cooper RS, Vega GL, Cohen JC, Hobbs HH. A spectrum of PCSK9 alleles contributes to plasma levels of low-density lipoprotein cholesterol. *Am J Hum Genet*. 2006;78:410–422.
18. Chen SN, Ballantyne CM, Gotto AM Jr, Tan Y, Willerson JT, Marian AJ. A common PCSK9 haplotype, encompassing the E670G coding single nucleotide polymorphism, is a novel genetic marker for plasma low-density lipoprotein cholesterol levels and severity of coronary atherosclerosis. *J Am Coll Cardiol*. 2005;45:1611–1619.
19. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575.
20. Keating BJ, Tischfield S, Murray SS, Bhangale T, Price TS, Glessner JT, Galver L, Barrett JC, Grant SF, Farlow DN, Chandrupatla HR, Hansen M, Ajmal S, Papanicolaou GJ, Guo Y, Li M, Derohannessian S, de Bakker PI, Bailey SD, Montpetit A, Edmondson AC, Taylor K, Gai X, Wang SS, Fornage M, Shaikh T, Groop L, Boehnke M, Hall AS, Hattersley AT, Frackelton E, Patterson N, Chiang CW, Kim CE, Fabsitz RR, Ouwehand W, Price AL, Munroe P, Caulfield M, Drake T, Boerwinkle E, Reich D, Whitehead AS, Cappola TP, Samani NJ, Lusis AJ, Schadt E, Wilson JG, Koenig W, McCarthy MI, Kathiresan S, Gabriel SB, Hakonarson H, Anand SS, Reilly M, Engert JC, Nickerson DA, Rader DJ, Hirschhorn JN, Fitzgerald GA. Concept, design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. *PLoS One*. 2008;3:e3583.
21. Kapur NK, Musunuru K. Clinical efficacy and safety of statins in managing cardiovascular risk. *Vasc Health Risk Manag*. 2008;4:341–353.
22. Chen MH, Yang Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics*. 2010;26:580–581.
23. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557–560.
24. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449:851–861.

CLINICAL PERSPECTIVE

The National Heart, Lung, and Blood Institute's Candidate Gene Association Resource (CARe), a planned cross-cohort analysis of genetic variation in cardiovascular, pulmonary, hematologic, and sleep-related traits, comprises more than 40 000 participants representing 4 ethnic groups in 9 community-based cohorts. The CARe study is anticipated to provide many new insights into the relationship of genetics and various clinical traits and diseases.

SUPPLEMENTAL MATERIAL

Supplemental Table 1. Quality control pipeline*

Individuals with phenotypes	51,186								
	7,305 individuals' samples not sent to Broad due to lack of patient consent, lack of DNA, or DNA quality issues								
Total DNA samples sent to Broad Institute for genotyping (including controls and duplicates)	46,560								
	2,679 control samples or duplicates removed								
Study	**ARIC**	**CARDIA**	**CFS**	**CHS**	**CSSCD**	**FHS**	**JHS**	**MESA**	**Total**
Samples genotyped (not including controls or duplicates)	14,996	3,520	1,481	5,485	1,707	7,992	2,156	6,544	43,881
Twins removed						23			23
Removed for incomplete family information						436			436
Removed for Mendel errors			7			26			33
Removed for low genotyping rate	287	124	48	183	57	130	36	249	1,114
Final number of genotyped unique individuals (% of samples genotyped)	14,709 (98.1%)	3,396 (96.5%)	1,426 (96.3%)	5,302 (96.7%)	1,650 (96.7%)	7,377 (92.3%)	2,120 (98.3%)	6,295 (96.2%)	42,275 (96.3%)
Self-identified solely as *AA*, with available pilot phenotypes	3,435	1,613	693	794			2,118	1,713	10,366
Self-identified solely as *EA*, with available pilot phenotypes	10,504	1,743	691	4,362		6,914		2,433	26,647
Self-identified solely as *HIS*, with available pilot phenotypes								1,410	1,410
Self-identified solely as *CHI*, with available pilot phenotypes								717	717
Total analyzed[†] (% of final number of genotyped individuals)	13,939 (94.8%)	3,356 (98.8%)	1,384 (97.1%)	5,156 (97.2%)		6,914 (93.7%)	2,118 (99.9%)	6,273 (99.7%)	39,140 (92.6%)

AA = African American; EA = European American or Caucasian; HIS = Hispanic; CHI = Chinese American.

* The Sleep Heart Health Study (SHHS)[11] is not represented in this table because the subgroup of the SHHS cohort with genotype data comprises individuals originally recruited from ARIC, CHS, and FHS (and who are included in this table under these other cohorts).

[†] Because no lipid phenotypes were available for the CSSCD cohort, no analyses were performed for this cohort in the Pilot Study.

Supplemental Table 2. Genotype-phenotype associations for lipid traits

Trait	Gene	SNP (minor/major alleles)	*African American*							*European American*							*Hispanic*	*Chinese American*
			ARIC N=3435	JHS N=2118	MESA N=1713	CARDIA N=1613	CHS N=794	CFS N=693	Meta-analysis N=10366	ARIC N=10504	FHS N=6914	CHS N=4362	MESA N=2433	CARDIA N=1743	CFS N=691	Meta-analysis N=26647	MESA N=1410	MESA N=717
HDL	*CETP*	rs4783961 A/G	0.24 (0.02) 4.2E-22 (43%) 0.49	0.21 (0.03) 8.5E-12 (44%) 0.12	0.21 (0.03) 9.6E-10 (44%) 0.59	0.16 (0.04) 6.8E-06 (44%) 0.84	0.22 (0.05) 1.2E-05 (45%) 0.94	0.22 (0.06) 5.7E-05 (41%) 0.16	0.21 (0.01) 1.8E-52 0.64 0%	0.11 (0.01) 1.6E-16 (49.8%) 0.41	0.11 (0.02) 4.1E-10 (49.6%) 0.34	0.15 (0.02) 4.6E-12 (49%) 0.49	0.09 (0.03) 0.003 (49.8%) 0.44	0.13 (0.03) 1.9E-04 (49.7%) 0.67	0.10 (0.06) 0.10 (46%) 1.0	0.12 (0.01) 1.6E-40 0.59 0%	0.12 (0.04) 0.001 (49.5%) 1.0	0.26 (0.05) 1.2E-05 (24%) 0.18
HDL	*CETP*	rs17231506 T/C	0.16 (0.04) 6.9E-06 (14%) 1.0	0.07 (0.04) 0.11 (15%) 0.34	0.10 (0.05) 0.04 (15%) 0.85	0.06 (0.05) 0.18 (16%) 0.35	0.13 (0.07) 0.06 (16%) 0.30	0.26 (0.08) 0.001 (10%) 1.0	0.12 (0.02) 2.5E-09 0.21 30%	0.24 (0.01) 1.7E-60 (32%) 0.71	0.23 (0.02) 9.8E-33 (32%) 0.83	0.28 (0.02) 1.0E-35 (33%) 0.70	0.21 (0.03) 2.9E-11 (31%) 0.78	0.24 (0.04) 5.1E-11 (32%) 0.74	0.26 (0.06) 4.1E-05 (32%) 0.86	0.24 (0.01) 2.0E-148 0.42 0%	0.24 (0.04) 4.7E-09 (29%) 0.40	0.35 (0.07) 3.9E-07 (17%) 1.0
HDL	*LIPC*	rs1800588 T/C	0.09 (0.02) 5.9E-04 (48%) 1.0	0.07 (0.03) 0.03 (49.8%) 0.93	0.11 (0.03) 0.002 (49.9%) 0.81	0.11 (0.03) 0.002 (49.5%) 0.40	0.18 (0.05) 2.1E-04 (45%) 0.22	0.11 (0.06) 0.05 (46%) 0.86	0.10 (0.01) 2.3E-12 0.48 0%	0.13 (0.02) 2.3E-14 (21%) 0.03	0.12 (0.02) 5.7E-08 (21%) 0.93	0.13 (0.03) 2.0E-06 (21%) 0.21	0.08 (0.03) 0.02 (22%) 0.19	0.13 (0.04) 0.001 (21%) 0.14	0.08 (0.07) 0.31 (18%) 0.31	0.12 (0.01) 2.8E-29 0.86 0%	0.14 (0.04) 1.3E-04 (47%) 0.01	0.05 (0.06) 0.35 (35%) 0.12
HDL	*LPL*	rs328 G/C	0.17 (0.05) 5.0E-04 (7%) 0.68	0.20 (0.06) 5.2E-04 (8%) 0.44	0.10 (0.07) 0.13 (7%) 1.0	0.04 (0.07) 0.50 (8%) 0.86	0.06 (0.10) 0.55 (7%) 0.79	0.12 (0.11) 0.28 (7%) 0.48	0.13 (0.03) 1.1E-06 0.46 0%	0.18 (0.02) 2.9E-15 (10%) 0.26	0.19 (0.03) 1.1E-09 (10%) 0.16	0.16 (0.03) 8.9E-06 (11%) 0.38	0.26 (0.05) 2.0E-08 (11%) 0.46	0.06 (0.06) 0.27 (10%) 0.44	0.13 (0.11) 0.21 (9%) 0.63	0.18 (0.01) 1.8E-33 0.13 41%	0.22 (0.07) 0.001 (8%) 0.48	0.11 (0.08) 0.17 (12%) 1.0
LDL	*PCSK9*	rs505151 G/A	0.09 (0.03) 0.002 (25%) 0.26	0.15 (0.04) 3.7E-05 (26%) 0.86	0.09 (0.04) 0.02 (24%) 0.52	0.15 (0.04) 1.6E-04 (25%) 0.01	0.08 (0.06) 0.17 (24%) 0.01	-0.08 (0.09) 0.37 (28%) 1.0	0.11 (0.02) 1.1E-10 0.17 35%	0.09 (0.04) 0.02 (4%) 0.11	0.05 (0.05) 0.36 (4%) 1.0	0.07 (0.05) 0.17 (4%) 0.07	-0.03 (0.07) 0.71 (4%) 1.0	0.12 (0.09) 0.17 (4%) 1.0	0.12 (0.18) 0.49 (4%) 1.0	0.07 (0.02) 0.004 0.76 0%	0.23 (0.08) 0.004 (6%) 0.22	0.24 (0.11) 0.03 (6%) 0.76
LDL	*PCSK9*	rs11591147 T/G	-0.52 (0.17) 0.002 (0.5%) 0.07	-0.14 (0.21) 0.50 (0.5%) 1.0	-0.27 (0.26) 0.30 (0.4%) 1.0	-0.38 (0.21) 0.08 (0.6%) 0.06	-0.70 (0.45) 0.12 (0.3%) 1.0	— — — — —	-0.38 (0.10) 1.8E-04 0.62 0%	-0.47 (0.06) 1.9E-17 (1.6%) 0.53	-0.47 (0.08) 6.5E-10 (1.5%) 1.0	-0.50 (0.09) 3.7E-08 (1.4%) 0.60	-0.25 (0.12) 0.05 (1.4%) 1.0	-0.61 (0.14) 1.6E-05 (1.5%) 1.0	-0.50 (0.26) 0.05 (1.1%) 1.0	-0.46 (0.04) 2.3E-37 0.50 0%	-0.21 (0.20) 0.30 (1.0%) 1.0	— — — — —
TG	*LPL*	rs328 G/C	-0.21 (0.05) 3.5E-05 (7%) 0.68	-0.18 (0.06) 0.003 (8%) 0.44	-0.13 (0.07) 0.07 (7%) 1.0	-0.18 (0.07) 0.007 (8%) 0.86	0.00 (0.10) 0.99 (7%) 0.79	-0.36 (0.16) 0.02 (7%) 0.48	-0.17 (0.03) 1.3E-09 0.35 10%	-0.20 (0.02) 1.1E-17 (10%) 0.26	-0.19 (0.03) 8.5E-09 (10%) 0.16	-0.22 (0.03) 2.8E-10 (11%) 0.38	-0.23 (0.05) 4.9E-07 (11%) 0.46	-0.09 (0.06) 0.10 (10%) 0.44	-0.25 (0.17) 0.14 (9%) 0.63	-0.20 (0.01) 2.7E-39 0.41 0.5%	0.27 (0.07) 7.23E-05 (8%) 0.48	-0.04 (0.08) 0.59 (12%) 1.0
TG	*APOA5*	rs3135506 C/G	0.22 (0.05) 2.3E-05 (6%) 0.37	0.31 (0.06) 1.6E-06 (6%) 0.57	0.39 (0.07) 4.0E-08 (6%) 1.0	0.29 (0.08) 1.2E-04 (6%) 0.70	0.14 (0.10) 0.16 (7%) 0.26	0.23 (0.14) 0.09 (5%) 1.0	0.27 (0.03) 2.6E-20 0.30 17%	0.28 (0.03) 2.6E-24 (7%) 0.94	0.22 (0.04) 1.1E-08 (7%) 1.0	0.26 (0.05) 1.1E-08 (6%) 0.19	0.18 (0.06) 0.002 (6%) 0.61	0.29 (0.07) 5.3E-05 (6%) 0.38	0.48 (0.14) 9.4E-04 (7%) 0.59	0.26 (0.02) 1.3E-45 0.36 8.9%	0.37 (0.05) 1.2E-11 (13%) 0.09	— — — — —

For each study cohort/ethnicity with each phenotype, the five numbers are (1) effect size (beta-coefficient), (2) standard error for the effect size, (3) *P* value for genotype-phenotype association, (4) allele frequency for the allele designated in the *SNP* column as the minor allele, and (5) the Hardy-Weinberg *P* value for the SNP. For the meta-analyses, the five numbers are (1) the overall effect size, (2) the overall standard error, (3) the overall *P* value for the ethnicity, (4) the Cochran's Q *P* value for heterogeneity among the cohorts for the ethnicity, and (5) the I^2 inconsistency metric for heterogeneity. The numbers of DNA samples successfully genotyped and passing QC criteria for each cohort are indicated in the top row. HDL = high-density lipoprotein cholesterol; LDL = low-density lipoprotein cholesterol; TG = triglyceride

SUPPLEMENTAL FIGURE LEGENDS

Supplemental Figure 1. Governance structure of CARe.

Supplemental Figure 1. Governance structure of CARe