



# American Heart Association Cardiovascular Genome-Phenome Study: Foundational Basis and Program

## Citation

Benjamin, Ivor, Nancy Brown, Gregory Burke, Adolfo Correa, Steven R. Houser, Daniel W. Jones, Joseph Loscalzo, Ramachandran S. Vasan, and Gayle R. Whitman. 2014. "American Heart Association Cardiovascular Genome-Phenome Study: Foundational Basis and Program." *Circulation* 131 (1): 100-112. doi:10.1161/CIRCULATIONAHA.114.014190. <http://dx.doi.org/10.1161/CIRCULATIONAHA.114.014190>.

## Published Version

doi:10.1161/CIRCULATIONAHA.114.014190

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:13890585>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

## American Heart Association Cardiovascular Genome-Phenome Study Foundational Basis and Program

Ivor Benjamin, MD; Nancy Brown; Gregory Burke, MD, MSc; Adolfo Correa, MD, MPH, PhD;  
Steven R. Houser, PhD; Daniel W. Jones, MD; Joseph Loscalzo, MD, PhD;  
Ramachandran S. Vasan, MD; Gayle R. Whitman, PhD, RN

The American Heart Association has established the Cardiovascular Genome-Phenome Study (CV-GPS) as a means by which to achieve the goal of using modern genomics and phenotyping optimally to combat cardiovascular disease (CVD). CVD remains the leading cause of death in the United States and has become a major cause of morbidity and mortality worldwide. The course of CVD begins at an early age and evolves throughout life. Typically, risk factor precursors (eg, prehypertension, borderline abnormal lipids) are first observed in adolescence or young adulthood. By middle age, they develop into full-fledged cardiometabolic risk factors and subclinical disease, which usually become clinically apparent in older age. Importantly, risk factors for CVD reflect the complex interplay between genetic and environmental factors and demonstrate the complexity of the multiple determinants of disease.

---

### Editorial see p 17

---

Many of the genetic and environmental determinants of CVD have been identified over the past 50 years, and successful preventive and therapeutic strategies have been developed as a result. However, many potential genetic or acquired disease drivers remain unaccounted for, as illustrated by the observation that the major CVD risk factors combined account for only a fraction of the population-attributable risk.<sup>1</sup> In addition, variations in CVD risk exist among individuals with similar risk profiles. Part of these differences may reflect differences in profiles of known risk factors that have not yet been fully refined but also individual differences in underlying genetic modifiers of such risk factors. Modern genomics, with its ability to provide ideally unbiased analysis of the entire genome, offers an approach to the ascertainment of all genetic determinants of CVD. Furthermore, the increasingly broad range of “omics” methods, including RNA-Seq, modern proteomics, metabolomics, and metabonomics, provide deeper and more refined molecular detail by which to define an individual’s genome and its relationship to (patho)phenotype. Armed with these increasingly powerful molecular methods, modern genomics

has yielded some new genetic targets that may account, in part, for the missing population-attributable risk and unexplained variation in risk among individuals, population subgroups, and aggregate populations.<sup>2</sup> However, the conventional strategy of seeking simple associations between genomic loci or individual gene products and disease phenotype is limited (with rare exception) by the typically small effect sizes of given variants in a population and by a failure to take into consideration the networked complexity of protein-protein interactions and their modification by environmental (epigenetic and posttranslational) modulators.<sup>3</sup>

These limitations of modern genomics are compounded by limitations of conventional phenotyping. Specific cardiovascular phenotypes have been defined on the basis of their mechanistic relevance to clinical atherothrombotic vascular disease (eg, hypertension and hypercholesterolemia), their ease of measurement in individuals and in populations, and their epidemiological tractability. This approach to phenotyping grew in parallel with the evolution of clinical trials that, by their contemporary nature, tend to be diagnostically overinclusive, downplaying (except by predefined subgroup analyses when feasible) differences in phenotype that may better predict outcomes or response to therapy (eg, subpopulation differences in response to specific antihypertensive therapies). In addition, with some important exceptions (*vide infra*), many major epidemiological studies and clinical trials do not collect data over time, which further limits the strength of association between genotype and phenotype, the ability to evaluate the intrasubject and intersubject variations in risk factors over time, and the potential dynamic effects of such variations on genetic susceptibility and the ability to infer causality.

In light of the dramatic expansion of genomic data, it has become clear that traditional phenotyping as conducted in conventional epidemiological cohort studies may limit the ability to elucidate optimally genome-phenome relationships. There is, therefore, a clear need to develop deep phenotyping (ie, traits that are not typically measured or traits that are “orthogonal” to or not directly associated with the clinical

---

From the Medical College of Wisconsin, Milwaukee (I.B.); American Heart Association, Dallas, TX (N.B., G.W.); Wake Forest University School of Medicine, Winston-Salem, NC (G.B.); Mississippi Medical Center, Jackson Heart Study, Jackson (A.C.); Temple University School of Medicine, Philadelphia, PA (S.R.H.); University of Mississippi, Oxford (D.J.); Brigham and Women’s Hospital, Harvard Medical School, Boston, MA (J.L.); and Boston University School of Medicine and Framingham Heart Study, Boston, MA (R.S.V.).

Correspondence to Joseph Loscalzo, Department of Medicine, Brigham and Women’s Hospital, 75 Francis St, Boston, MA 02030. E-mail [jloscalzo@partners.org](mailto:jloscalzo@partners.org)

(*Circulation*. 2015;131:100-112. DOI: 10.1161/CIRCULATIONAHA.114.014190.)

© 2014 The Authors. *Circulation* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial-NoDerivs License, which permits use, distribution, and reproduction in any medium, provided that the Contribution is properly cited, the use is non-commercial, and no modifications or adaptations are made.

*Circulation* is available at <http://circ.ahajournals.org>

DOI: 10.1161/CIRCULATIONAHA.114.014190

phenotype under consideration) and dynamic phenotyping (including responses to perturbations) to capture fully the complexity of human disease. Furthermore, this approach to phenotyping coupled to modern genomics will be essential for achieving a truly individualized approach to the diagnosis, prognosis, and therapy of CVD.

With this background, there are 3 elements of note in CV-GPS. First, CV-GPS aims to provide a platform with which to integrate (virtually) existing and future CVD population studies. Beginning with the Framingham Heart Study (FHS) and the Jackson Heart Study (JHS), CV-GPS will consolidate available phenotypic and diverse omic data sets (genomic, transcriptomic, proteomic, metabolomic, epigenomic), expanding these data sets with the recruitment of other population data sets with information on CVD. This universe of data sets will define the big data platform that will be used to explore testable hypotheses about CVD. Second, CV-GPS will facilitate the development of the information technology infrastructure necessary for handling these big data and for classifying deep phenotypes. Third, CV-GPS will provide core support for methodological analyses (novel omics platforms) and a biorepository (real or virtual) for biological specimens from existing or future population study participants. Through these mechanisms, CV-GPS aspires to establish a national standard for all genome-phenome studies, applying the most cutting-edge approaches to the acquisition, storage, and analysis of information contained within this biophenorepository to deepen our understanding of the determinants of CVD.

## The FHS as a Component of CV-GPS

### Design and Component Cohorts

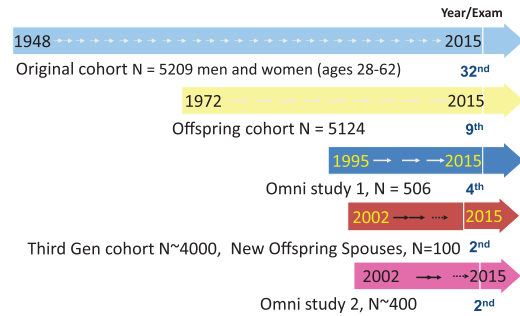
The FHS consists of  $\approx 15\,000$  extensively phenotyped women and men (Figure 1). Details of the constituent cohorts, sample size, DNA availability, ethnic/racial composition, recruitment year and ages, and follow-up are given in the Table. The design and sampling criteria of the FHS cohorts have been published previously.<sup>4-7</sup> The cohorts span a wide range of ages throughout the adult life course, between 20 and 100 years of age. The FHS is funded by the National Heart, Lung, and Blood Institute (NHLBI).

### Suitability of FHS for CV-GPS

Several features of the FHS make it a valuable asset to CV-GPS. These resources are briefly detailed below.

#### Extensive Longitudinal Phenotypic Data

The FHS has routinely collected longitudinal data over the adult life span of each participant in its 6 constituent cohorts, with extensive information on serial measures of risk factors, subclinical disease, and clinical outcomes, including CVD; lung, blood, and neurological conditions (stroke and dementia); and cause-specific mortality. At each FHS examination cycle, data are collected on blood pressure, anthropometry, lipids, smoking, glycemic traits and diabetes mellitus, ECG, and lung function. Across select serial examinations, the FHS has compiled multiple measures of cardiovascular structure and function as measured by echocardiography, cardiac computed tomography (CT), cardiac and brain magnetic resonance



**Figure 1.** Design of the Framingham Heart Study, including its constituent cohorts.

imaging (MRI), carotid ultrasound, conduit artery stiffness, and flow-mediated dilation. Further details on the traits available are detailed in Appendix I, Item 1 and summarized in Figure 2.

#### Extensive Genotypic Data

In the 1990s and early 2000s, DNA samples were collected in the Original, Offspring, Third Generation, and Offspring Spouse cohorts of the FHS to establish an invaluable resource for genetic research. Most of the FHS participants with available genomic and cell line-based DNA are Offspring, Third Generation, and Offspring Spouses. Additionally, FHS collected DNA on  $\approx 1000$  participants from the Original Cohort who were alive at the time of DNA collection; DNA was also extracted on an additional  $\approx 500$  deceased Original Cohort participants from available whole blood as part of the SNP Health Association Resource (SHARe)<sup>8</sup> project (see below). At FHS,  $\approx 9300$  participants across the constituent cohorts underwent genome-wide genotyping of 550 000 single-nucleotide polymorphisms (SNPs) using the Affymetrix platform for the SHARe project, and  $>7500$  individuals underwent genotyping of 50 000 SNPs using the Illumina Cardiochip in the Candidate Gene Association Resource (CARE) project.<sup>9</sup> With the use of extant genotypes, imputation has been completed of all participants to 40 million SNPs using the 1000G Imputation. Additionally, participants in the Offspring and Third Generation cohorts have undergone genotyping of  $\approx 200\,000$  functional exonic variants via Illumina V1.0 Exome Chip. FHS has completed whole-exome sequencing in  $\approx 2975$  participants from several projects (including the NHLBI GO Exome Sequence Project<sup>10</sup> and the CHARGE targeted Sequencing project)<sup>11</sup> and low-pass whole-genome sequencing (WGS) in  $\approx 850$  participants.<sup>12</sup> Of note, the genotyping (including whole-exome sequencing and WGS) that has been conducted to date with the FHS samples in numerous NHLBI-funded genotyping and sequencing programs has produced extremely high-quality sequence data with high call rates and low error rates. Most recently, a subsample of FHS participants has been selected to undergo WGS in a subset of its participants as part of the NHLBI-WGS project.<sup>13</sup> Appendix I, Item 2 tabulates genotypic data available at FHS.

#### Available Omics Data

In addition to genotype data and multiple measures of heart, lung, blood, and sleep phenotypes, the FHS has been actively collecting various omics data, including DNA methylation, transcriptomics (mRNA and microRNA expression in whole

**Table. The Framingham Heart Study and Its Constituent Cohorts: Cohort Characteristics**

Cohort	Total, n	DNA, n	Ancestry, %					Year Recruited	Age, y	Follow-Up		
			EA	AA	HA	AsA	Other			Examinations, n	Intervals, n	Length, y
Original	5209	971	100	0	0	0	0	1948–1953	28–74	32	2	>65
Offspring (and Spouses)	5124	3930	100	0	0	0	0	1971–1975	5–70	9	4–8	43
Third Generation	4095	4077	100	0	0	0	0	2002–2005	19–72	2	6	≈10
Offspring Spouses	103	101	100	0	0	0	0	2003–2005	47–85	2	6	≈10
Omni 1 cohort	506	493	0	28	42	24	6	1994–1998	27–78	4	4–8	≈15–20
Omni 2 cohort	410	407	0	28	42	24	6	2003–2005	20–80	2	6	≈10

AA indicates African American; AsA, Asian American; EA, European American; and HA, Hispanic American.

blood), metabolomic markers, and protein biomarkers, over the past 5 to 10 years. The rich FHS resources with multiple measures of CVD risk factors, key subclinical disease traits and numerous clinical CVD outcomes, extensive genotype data, and a variety of omics data provide an excellent basis for genome-phenome analysis and integrative genomics.

### Family Data

With the recruitment of the Offspring and Third Generation cohorts, the FHS is a full-fledged family study, with 3 generations of participants. DNA is available in 1037 extended families consisting of 5673 individuals who are genetically informative for imputation to a total of 7917 family members across all 3 cohorts. These families will provide a rich resource for CV-GPS that will greatly facilitate the examination of genome-phenome associations across the full spectrum of CVD phenotypes (including risk factors and subclinical and clinical CVD); a wide range of other lung, blood, sleep, and neurological traits; and a comprehensive battery of omics traits (as noted above). In addition, family data will provide an opportunity to track cosegregation of genetic and trait variation within families, strengthening the attribution of genetic causation.

### FHS Biorepositories

The FHS maintains both a genetic and a nongenetic biorepository. FHS maintains in its genetic biorepository whole-blood aliquots, buffy coats, and PAXgene tubes on its cohorts at select examination cycles. There are 8444 unique Epstein-Barr virus—transformed cell lines on FHS participants, and 5823 participants have aliquots of lymphocytes in cryogenic storage. The repository has >25 000 stock DNA samples extracted from either a cell line or a blood source. From those stock DNA

samples, FHS has formulated stock distribution plate sets. Stock DNA sample concentrations are normalized, checked by electrophoresis on an agarose gel for visual confirmation of DNA quality and concentration, and then forensically genotyped and compared with archival forensic data, or known family structures, to validate the identity of each DNA sample.

Currently, the FHS nongenetic biorepository contains ≈1.3 million biospecimens of various sample types, including serum, plasma, buffy coat, red blood cells, and urine. Access to FHS biosamples from its genetic or nongenetic biorepository requires appropriate approvals (institutional review board, laboratory, DNA committee, etc) and a signed data and material distribution agreement and is consistent with the informed consent of its participants.

### FHS Public Databases That Can Be Accessed as Part of CV-GPS

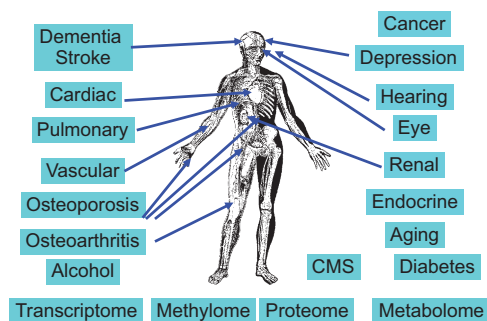
FHS has contributed high quality and high volume to data repository programs such as the Database of Genotypes and Phenotypes (dbGaP), maintained by the National Center for Biotechnology Information, and the Biological Specimen and Data Repositories Information Coordinating Center (BioLINCC) of the NHLBI. FHS data can also be accessed via the parent study. All access to FHS data requires approval by the recipient institutional review board and the relevant FHS committees and must be consistent with the participant informed consent. All FHS data stored at the recipient institution must have appropriate safeguards to protect participant confidentiality.

### FHS dbGaP Phenotypic Data Sets for Genetic Research

Framingham data sets and documentation are formatted according to dbGaP standards before submission. Each data set submission includes a dbGaP-formatted data set, a coding manual, a data collection protocol, related informed consents, an annotated data collection form, and a funding source reference. Both ancillary study data sets and NHLBI FHS-contracted data sets are posted on dbGaP. Well over 365 data sets were submitted by FHS by the first quarter of 2014.

### FHS BioLINCC Data Sets for Nongenetic Research

Since 2000, the BioLINCC data repository has been updated approximately every 2 years with FHS contract—supported and ancillary grant—supported data and documentation. Data sets and documentation are formatted according to BioLINCC standards,



**Figure 2.** Extensive phenotypic characterization of the Framingham Heart Study (FHS) participants. CMS indicates Centers for Medicare and Medicaid Services data.

including data sets, protocols for collection, and coding manuals. We propose to continue these deposits using past experience and expertise. Since 2000, 137 data sets and 442 corresponding documentation files have been submitted to BioLINCC.

### **FHS Summary**

In conclusion, the large community-based cohorts with a 3-generational family structure, accompanying minority cohorts with a parallel examination structure, an extensive catalog of available traits, and the availability of substantial omics resources and biosamples establish the FHS as uniquely suited for the American Heart Association CV-GPS initiative.

## **The JHS as a Component of CV-GPS**

### **Background**

The JHS is a single-site, prospective, cohort study of risk factors of CVD among 5301 blacks living in the Jackson, MS, metropolitan area. The JHS is a collaborative effort among 3 Jackson-area academic institutions, the University of Mississippi Medical Center, Jackson State University, and Tougaloo College. Primary goals of the JHS are to broaden the research on CVD risk factors in a black population, to increase access and participation of black populations and scientists in biomedical research and professions, and to implement outreach activities to increase awareness and to promote healthy lifestyles in the community.<sup>14</sup> Three clinical examinations have been completed, including the baseline examination, examination 1 (2000–2004), examination 2 (2005–2008), and examination 3 (2009–2013), allowing comprehensive assessment of cardiovascular health and disease of the cohort at ≈4-year intervals. Ongoing monitoring of cardiovascular events and deaths among cohort participants was achieved by annual telephone follow-up interviews and surveillance of hospital discharge records and vital records. At present, the JHS is not conducting clinical examinations, but annual telephone follow-up and surveillance of the cohort for CVD events and deaths continue. The JHS is funded by the NHLBI and the National Institute on Minority Health and Health Disparities.

### **Design**

The JHS is a cohort study of extensively phenotyped black women and men. The details of the cohort, including design, sampling, recruitment, examinations, and follow-up, are described elsewhere.<sup>15–17</sup> The JHS cohort spans a wide range of ages throughout the adult life course, between 21 and 101 years of age.

### **Suitability of JHS for CV-GPS**

Several features of the JHS make it a valuable asset to CV-GPS. Among these resources are included extensive phenotypic and genotypic data, which are briefly detailed below.

#### ***Longitudinal Phenotypic Data***

The JHS has collected longitudinal data among its study participants, with extensive information on measures of risk factors, subclinical disease, and clinical outcomes, including CVD, kidney disease, and neurological conditions (stroke and transient ischemic attack), as well as mortality. At each JHS

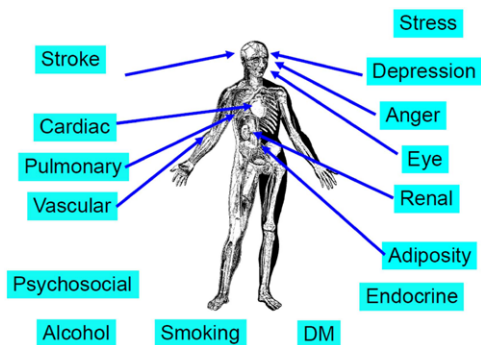
examination cycle, data were collected on blood pressure, anthropometry, lipids, smoking, glycemic status, and diabetes mellitus. Across select serial examinations, the JHS has compiled multiple measures of cardiovascular structure and function as determined by echocardiography, ECG, cardiac CT, cardiac MRI, carotid ultrasound, and conduit artery stiffness. Further details on the phenotypic data available are provided in the Appendix II, Item 1 and are summarized in Figure 3.

#### ***Genotypic Data***

At examinations 1 and 2, DNA samples were collected from consenting study participants to establish an invaluable resource for genetic research. A Family Cohort (see below) includes >1500 participants, including cryptically related individuals identified by genetic analysis, in nearly 300 pedigrees, which vary in structure from sibships to cousin pairs to extended 3-generation families. Through the CARE project and after extensive quality control procedures, 3029 participants have genome-wide genotyping on the Affymetrix 6.0 platform, which interrogates >900000 SNPs and has >900000 probes for copy number variation. These data have been imputed to ≈37 million SNPs using the 1000 Genomes project reference panel (version 3, March 2012 release). In addition, through the CARE project, 2948 individuals have genotyping data for ≈50000 SNPs on the gene-centric Illumina Cardiochip (IBC Array), which provides dense tagging of ≈2100 candidate genes for CVD.<sup>9</sup> Genotyping has been completed in 2790 participants for ≈240000 uncommon and rare, mainly nonsynonymous coding variants and other selected content of the Illumina V1.0 Exome Chip, with support from NHLBI R01HL107816 (principal investigator, S. Kathiresan). Exome sequencing has been completed for all consenting participants through 4 projects: the Exome Sequencing Project (NHLBI), the Minority Health Genomics and Translational Research Bio-Repository Database (NHLBI), the Type 2 Diabetes Genetic Exploration by Next-Generation Sequencing in Multi-Ethnic Samples Project (National Institute of Diabetes and Digestive and Kidney Diseases), and the Broad Institute MI CIP. Joint calling of all exomes was performed by the Broad/Kathiresan laboratory, providing, after extensive quality control measures, a single jointly-called exome data set representing 3237 JHS participants. Prior deep-coverage WGS has been completed for 50 participants, including 25 who met criteria for asthma and 25 control subjects, through the Consortium on Asthma Among African-Ancestry Populations in the Americas Project (NHLBI). Of note, the genotyping and sequencing that have been completed to date using JHS samples have produced extremely high-quality data with high call rates and low error rates. Appendix II, Item 2, also tabulates genotyping and sequence data available at JHS.

#### ***Available Biomarker Data***

In addition to genotype data and multiple measures of heart, lung, blood, and sleep phenotypes, the JHS has been actively collecting various biomarkers data, including protein biomarkers, over the past 5 to 10 years. The rich JHS data set, with multiple measures of CVD risk factors, key subclinical disease traits, several clinical CVD outcomes, and extensive genotype data, provides an excellent basis for genome-phenome analysis and integrative genomics.



**Figure 3.** Extensive phenotypic characterization of the Jackson Heart Study (JHS) participants. DM indicates diabetes mellitus.

### Family Data

With the intentional recruitment of family members of participants selected on the basis of family size and availability, the JHS includes a well-developed, nested family study, with DNA available for 1486 genetically informative individuals in >270 pedigrees (primarily sibships and extended families). These families will support the examination of genome-phenome associations across the full spectrum of CVD phenotypes and a wide range of other lung, blood, sleep, and neurological traits by allowing segregation and other family-based analyses, particularly of uncommon and rare variants.

### JHS Biorepository

The JHS maintains both a genetic and a nongenetic biorepository, including serum, plasma, and urine aliquots, and purified DNA. There are ≈1500 participants with aliquots of mononuclear leukocytes in cryogenic storage. The repository has DNA samples from blood cells of >4700 participants, including ≈3400 whose consent allows sharing of genetic data through controlled-access repositories such as the dbGaP. Currently, the JHS nongenetic biorepository contains biospecimens of various sample types, including serum, plasma, and urine. Access to JHS biosamples from its genetic or nongenetic biorepository requires appropriate approvals (institutional review board, laboratory, genetics committee, etc) and a signed data and material distribution agreement and is subject to stipulations in the informed consent of each participant.

### JHS Public Databases That Can Be Accessed as Part of CV-GPS

JHS has contributed high-quality and high-volume data to repositories such as dbGaP, maintained by the National Center for Biotechnology Information, and BioLINCC of the NHLBI. JHS data can also be accessed via the parent study. All access to JHS data requires approval by the recipient institutional review board and relevant JHS committees and must be consistent with the participant informed consent. All JHS data stored at the recipient institution must have appropriate safeguards to protect participant confidentiality.

### JHS dbGaP Phenotypic Data Sets for Genetic Research

JHS data sets and documentation are formatted according to dbGaP standards before submission. Each data set submission includes a dbGaP-formatted data set, a coding manual, a data

collection protocol, related informed consents, an annotated data collection form, and a funding source reference. NHLBI JHS-contracted data sets are posted on dbGaP.

### JHS BioLINCC Data Sets for Nongenetic Research

Since 2000, the BioLINCC data repository has been updated ≈3 years after completion of the examination with JHS contract—supported and ancillary grant—supported data and documentation. Data sets and documentation are formatted according to BioLINCC standards, including data sets, protocols for collection, and coding manuals. We propose to continue these deposits using past experience and expertise.

### JHS Summary

In conclusion, a large community-based cohort of blacks, the extensive catalog of available traits, and the availability of substantial resources and biosamples establish the JHS as uniquely suited for the American Heart Association CV-GPS initiative.

## Population and Cohort Studies in CVD

### Cohort Studies as Population Laboratories

As illustrated by the FHS and JHS, large, observational, clinical studies provide an exceptional opportunity to understand better the cause of CVD. They have collected a wealth of information on large numbers of participants using rigorous standardized protocols. These and many other cohort and existing databases should be considered population laboratories that provide an opportunity for basic, clinical, and population science colleagues to propose new, cutting-edge science. The advantages of population laboratories for potential CV-GPS (and other) investigators include the following: (1) cost-effectiveness: research infrastructure is in place with cohort recruitment completed; (2) time efficiency: multiple predictor variables and outcome data are immediately available to be leveraged by CV-GPS ancillary study proposals; (3) temporality: cohort studies allow the evaluation of predictors of cardiovascular outcomes because exposures (health behaviors, risk factors, stored specimens, imaging, etc) were evaluated before the development of outcomes; (4) availability of events: validated CVD events are available in many cohorts (eg, myocardial infarction, heart failure, atrial fibrillation, cardiovascular procedures, stroke, and death); (5) large stored specimen repositories: most cohorts have plasma, serum, and DNA available for use; (6) dense genotypes: genetic information has been collected in most cohorts (eg, genome-wide association study [GWAS] data, epigenetic data); (7) community-based sampling: given that cohorts are recruited from the community, risk factors and disease outcomes are more representative of the population at large; and (8) excellent collaborative investigators: an existing cadre of scientists familiar with the study are available to assist with both operational and scientific issues.

### Overview of Current Established Studies

The CV-GPS initiative is fortunate to be able to leverage the excellent research infrastructure resulting from the strategic vision and research funding over many decades by National Institutes of Health/NHLBI colleagues, the hard work of cohort study investigators/staff, and the selfless dedication of study

participants. It should be noted that although the CV-GPS initiative has initially focused on the FHS and JHS, these studies represent only a subset of available cohorts. The larger set of cohort studies provide access to an even broader spectrum of characteristics such as age (ranging from childhood to the oldest old), US race/ethnic group (ie, black, Hispanic, Chinese, Japanese, white, Native American, Alaska Native), phenotypic characteristics (health behaviors, risk factors, subclinical disease, laboratory measures, environmental factors, etc), and CVD states (eg, free of clinical CVD, prevalent CVD, incident CVD).

Some examples of major NHLBI cohorts include the Atherosclerosis Risk in Communities Study<sup>18</sup> (n=15792; age, 45–64 years at baseline examination [1987–1989]; white and black participants); the Coronary Artery Risk Development in Young Adults Study<sup>19</sup> (n=5115; age, 18–30 years at the baseline examination [1985–1986]; white and black participants); the Cardiovascular Health Study<sup>20</sup> (n=5888; age, ≥65 years at baseline examination [1990–1991]; white and black participants); the FHS (described in detail above); the Hispanic Community Health Study/Study of Latinos<sup>21</sup> (n=16400; age, 18–74 years at baseline examination [2008–2011]; participants are Hispanics/Latinos representing different groups of origin [Central Americans, Cubans, Dominicans, Mexicans, Puerto Ricans, and South Americans]); JHS (described in detail above); the Multi-Ethnic Study of Atherosclerosis<sup>22</sup> (n=6814; age, 45–84 years at baseline examination [2000–2002]; participants from multiple race/ethnic groups [black, white, Hispanic, Chinese]); the Strong Heart Study<sup>23</sup> (n=4500 Native American tribal members; age, 45–74 years at baseline examination [1984–1988]); and the Women's Health Initiative Observational Study<sup>24</sup> (n=93 676 postmenopausal women; age, 50–79 years at baseline examination [1991–1994]). Given that these cohorts differ in the populations studied (ie, age, race/ethnicity), phenotypic characteristics evaluated, amount/types of stored specimens, and events observed, CV-GPS investigators should consider the suitability of different cohort studies for the proposed hypotheses they plan to explore.

### Cohort Studies: A Legacy of Discovery

The CV-GPS program builds on an exceptional legacy of discovery in population studies. Although it is clearly beyond the scope of this article to report all significant scientific findings from these studies, some noteworthy findings include the following: (1) identification of major risk factors for CVD in middle-aged and older adults<sup>25,26</sup>; (2) documentation of the ability to measure accurately risk factors in children and of the persistence (tracking) of risk factors from childhood into middle age<sup>27,28</sup>; (3) description of race/ethnic disparities in CVD and identification of some mechanisms for those disparities<sup>29–32</sup>; (4) identification of a number of dietary factors that affect CVD risk (ie, saturated fat, *trans* fats)<sup>33,34</sup>; (5) documentation of associations of active and passive cigarette exposure on CVD risk<sup>35,36</sup>; (6) identification of subclinical disease markers as intermediate markers of CVD risk and predictors of CVD events<sup>37–39</sup>; (7) assessment of the importance of the neighborhood characteristics on CVD risk<sup>40,41</sup>; (8) identification of the importance of inflammatory markers on CVD risk<sup>42–44</sup>; (9) documentation of the importance of the diabetes and obesity epidemic on CVD risk factors and

outcomes<sup>45–47</sup>; (10) monitoring of time trends in medical care, including medication use in the United States<sup>48,49</sup>; (11) identification of genetic markers for outcomes by the integration of dense genotyping<sup>50,51</sup>; and (12) introduction and refinement of risk prediction equations to allow appropriate risk stratification for use in clinical practice.<sup>25,52,53</sup> Although these and many other discoveries have resulted from prior evaluation of cohort study data, it should be noted that tremendous opportunities remain for future innovation using these population resources to advance science through CV-GPS and other research initiatives.

### Power of Large Numbers and Value of Data Set Integration

Often, scientific discovery is inhibited by the lack of an appropriate sample size, an unsuitable study design, or a nonrepresentative population. The next generation of studies will seek to evaluate the importance of technological advances such as genetics, epigenetics, and metabolomics on cardiovascular risk. It is important to note that GWASs have generally observed relatively small associations with risk factors and disease outcomes. Thus, it is reasonable to anticipate that future research efforts will require the evaluation of a vast array of variables that may have small to modest effects on outcomes. This differs substantially from past work that was powered to seek associations with larger effect sizes such as traditional CVD risk factors. Pooling multiple cohort studies provides an opportunity to further increase sample size and study power, which is especially crucial for the investigation of both small effect sizes and CVD outcomes in informative subgroups. A prime example of this pooling is currently being explored in the genetic consortium evaluating GWAS data. Given the relatively small effect size of single genes, it is imperative both to pool cohort data for primary analyses and validation of observed associations and to examine their interactions in the complex molecular networks through which they exert their effects on phenotype. Ensuring standardization of variables across studies is a challenge when data are pooled. However, a major advantage of pooling NHLBI-funded cohort data is that, although not perfectly harmonized, similar methods have been used across studies to evaluate risk factors, survey instruments, physical measurements, and events.

### Integration of Omics and New Technologies Into Population Studies

Recent advances have made large-scale integration of new predictor variables, such as genomics/epigenetics, metabolomics, and the microbiome, more economically and technologically feasible for large-scale implementation. Population studies provide an excellent opportunity for the rapid evaluation of these new technologies as a result of the availability of stored specimens, existing phenotypes, and outcome data. Substantial economies of scale are gained when omics ancillary studies/variables are added to an existing cohort study. Incorporation of new technology not only leverages the wealth of existing data from the parent population study for use by the initiating investigators but also allows the new data to be used as variables or covariates for further discovery by all research colleagues. For example, a novel proposed biomarker for coronary atherosclerosis may also be evaluated as a predictor of heart failure or stroke outcomes. The result of these efforts will be the creation of an enhanced

data set available to a broad community of scientific colleagues that facilitates the efficient evaluation of additional scientific questions. Extramural funding of new science (ie, CV-GPS and other funding sources) that leverage existing cohorts preserves resources and time for omics-based discovery efforts.

### Conclusions

The next generation of scientific discovery certainty requires the creation of large data repositories to ensure adequate sample size to detect small but important effects, computational platforms that seek complex interactions through analysis of interactive genomic and phenomic networks, multiple samples to allow validation of initial study findings, study diversity to

allow the findings to be applicable to a larger universe of the public (age, race/ethnicity, sex, socioeconomic status, etc), and a user-friendly interface that makes these data more accessible to the larger community of researchers. Existing cohorts provide much of the initial data that can be approached as the overall infrastructure of CV-GPS is built; importantly, new prospective study populations and cohorts are encouraged to participate in CV-GPS to ensure that all possible population-based genomic and phenomic data are available in one integrated site for optimal analytic benefit. These data requirements and aspirations for comprehensive participant populations are essential as CV-GPS seeks to identify new determinants of the origin of CVD and, more important, factors that can be translated

## Appendix I

### 1. FHS Phenotypic Data

Repeated measures available on participants across 65 years; 3 generations of participants: Original, Offspring, and Third Generation; entire adult life span from 20 to  $\geq 100$  years

Available phenotype data	Examples
Clinical events (validated and adjudicated)	Coronary heart disease: myocardial infarction, coronary insufficiency, angina, coronary heart disease death, sudden coronary disease death, coronary artery bypass graft surgery, percutaneous transluminal coronary angiography Heart failure Stroke, transient ischemic attack Peripheral artery disease: intermittent claudication, lower-extremity revascularization Atrial fibrillation and electrophysiology procedures Dementia (Alzheimer disease, vascular) Mild cognitive impairment Cancer
Subclinical disease (most measures repeated)	Ultrasound carotid intima-media thickness and carotid stenosis Brachial reactivity/endothelial function Tonometry: arterial stiffness and peripheral arterial tonometry Echocardiographic structure and function (eg, left ventricular mass) CT: coronary artery calcium, abdominal aortic calcium, mitral and aortic valve calcium Cardiac MRI: cardiac structure, cardiac index, and aortic arch plaque Ankle-brachial index Walk test: low-level exercise test Brain MRI: including gray, white, lobar/deep volumes, infarcts, microbleeds, white matter hyperintensities, fractional anisotropy, regional brain volumes
Pulmonary disease and sleep traits (pulmonary function test data available on multiple examinations)	Spirometry and postalbuterol spirometry, diffusion capacity CT lung measures Sleep study and sleep questionnaire Buccal swabs
Traditional risk factors (directly measured)	Systolic and diastolic blood pressures Fasting blood glucose, hemoglobin A <sub>1c</sub> , fasting insulin Fasting lipids
Anthropometry (directly measured)	Height, weight, body mass index Waist, hip, thigh, neck circumference CT measures of regional adipose tissue depots: subcutaneous adipose tissue, visceral adipose tissue, pericardial fat, perithoracic fat Body percent fat (dual-energy x-ray absorptiometry)
Lifestyle	Diet: Willet food frequency questionnaire (calories, supplements) Smoking Exercise (self-report, objective measurement with accelerometry) Alcohol intake
Measures of function	Physical function and mobility Performance: hand grip, walking speed Cognitive function (global and multiple domains) Depression (Center for Epidemiologic Studies Depression Scale) Social network

(Continued)



**1. Continued**

Repeated measures available on participants across 65 years; 3 generations of participants: Original, Offspring, and Third Generation; entire adult life span from 20 to  $\geq 100$  years

Available phenotype data	Examples
Medications	All examinations
Medicare data	Centers for Medicare & Medicaid Services data since 1991, including <i>International Classification of Diseases</i> codes, charges, medications, procedures on individuals $\geq 65$ years of age enrolled in fee for service
Bone health	Fractures, osteoporosis Osteoarthritis
Circulating and urine biomarkers	
Renal function	Creatinine, cystatin C, microalbumin, uric acid
Inflammatory marker panel	Acute-phase reactants: C-reactive protein Cytokines: interleukin-6, tumor necrosis factor- $\alpha$ , tumor necrosis factor receptor 2, osteoprotegerin Selectins: P-selectin, CD40 ligand Cell adhesion: intercellular adhesion molecule Chemokines: monocyte chemoattractant protein-1 Oxidative stress: myeloperoxidase, isoprostanes Lipoprotein-associated phospholipase A2 mass and activity
Markers of hemostasis, thrombosis	Fibrinogen, factor VIIIc, von Willebrand factor, D-dimer, plasminogen activator inhibitor-1, platelet reactivity
Markers of myocardial injury	Troponin I, growth differentiation factor-15, ST-2, brain natriuretic peptide, N-terminal atrial natriuretic peptide
Adipokines	Leptin, leptin receptor $\alpha$ -Fetuin, resistin, ghrelin Retinol binding protein-4 Fatty acid binding protein-4 Adiponectin
Growth factors	Insulin-like growth factor-1, insulin-like growth factor binding protein-3, vascular endothelial growth factor, angiopoietin-2, tie-2, brain-derived neurotrophic factor, nerve growth factor
Molecules interacting with vessel wall and platelets	Matrix remodeling markers: matrix metalloproteinase-9, matrix metalloproteinase-3, tissue inhibitor of matrix metalloproteinase-1, N-terminal propeptide of type III procollagen Plasma homocysteine, asymmetrical dimethyl arginine
Markers of brain injury	$\beta$ -Amyloid, clusterin
Lipid subfractions	Lipoprotein(a), apolipoproteins A1, B48, B100, CI, CII, E, H, J
Hormones and vitamins	Renin-angiotensin-aldosterone pathway Thyroid function (eg, thyroid-stimulating hormone) Sex steroid hormones Vitamin D, folate, B <sub>12</sub> , B <sub>6</sub> , vitamin E

**2. Available FHS Genetic Data**

Framingham Participants With Consent for Genetic Analysis

Variable	ALL Cohorts	Original Cohort	Offspring Cohort	Omni Gr 1	Third Generation	
Sample size	14 271	5079	5013	101	4078	
CARe	7547	647	3023	0	3877	SNPs from $\approx 2100$ candidate genes across a range of cardiovascular, metabolic, and inflammatory syndromes, produced as part of the NHLBI CARe project (phs000282.v12.p8).
CHARGES_TARGET	1095	36	957	0	102	Targeted sequencing produced as part of the CHARGE sequencing consortium (phs000651.v3.p8)
CHARGES_WES_FREEZE1	621	13	487	0	121	First release of whole-exome sequencing produced as part of the CHARGE sequencing consortium (phs000651.v3.p8)
CHARGES_WES_FREEZE3	850	13	715	0	122	Third release of whole-exome sequencing produced as part of the CHARGE sequencing consortium (phs000651.v3.p8)
CHARGES_WES_FREEZE4	1271	45	1107	0	119	Fourth release of whole-exome sequencing produced as part of the CHARGE sequencing consortium (phs000651.v3.p8)
CHARGES_WGS_FREEZE1	320	8	282	0	30	First release of WGS produced as part of the CHARGE sequencing consortium (phs000651.v3.p8)

(Continued)

**2. Continued**

## Framingham Participants With Consent for Genetic Analysis

Variable	ALL Cohorts	Original Cohort	Offspring Cohort	Omni Gr 1	Third Generation	
ESP	464	0	291	0	173	Whole-exome sequence data produced as part of the NHLBI GO-ESP project (phs000401.v6.p8)
MARSH	4112	454	1400	0	2258	Microsatellite markers used in linkage analyses (phs000342.v11.p8)
METHYLATION	2202	0	2202	0	0	Methylation data; Offspring Cohort examination 8 (phs000724.v1.p8)
NHGRI_MEDSEQ	1703	0	1703	0	0	Deep-coverage targeted resequencing and variant identification for 216 genes in the FHS sample collection, produced as part of the NHLBI Medical Resequencing projects. (phs000307.v8.p8)
OMNI5	2472	0	2472	0	0	SNPs from the Illumina HumanOmni5M-4v1 array designed to target variation down to 1% minor allele frequency (phs000342.v11.p8)
ONEK	1342	258	1084	0	0	SNPs from 100 000 Affymetrix GeneChip in a subset of Original Cohort and Offspring participants of the largest 310 pedigrees in the FHS (phs000342.v11.p8)
PERLEGEN	1649	0	1649	0	0	SNPs in inflammatory candidate genes (phs000342.v11.p8)
PGA	1749	0	1749	0	0	SNP data of candidate genes involved in human congenital heart disease and cardiomyopathy (phs000342.v11.p8)
RTPCR	1943	0	1943	0	0	Reverse transcriptase—polymerase chain reaction gene expression, Offspring Cohort examination 8 (phs000325.v1.p5)
SABRE_EXPRESS	5622	0	2442	0	3180	Gene expression profiling of white blood cell—derived RNA to characterize the genomic signatures of atherosclerosis and metabolic syndrome (phs000363.v10.p8)
SABRE_IMMUNO	7315	0	3264	0	4051	Immunoassays of 180 circulating protein biomarkers of atherosclerosis and metabolic syndrome (phs000363.v10.p8)
SABRE_ITRAQ	269	0	269	0	0	iTRAQ Px data set 135 case/control pairs; iTRAQ is used in proteomics to study quantitative changes in the proteome (phs000363.v10.p8)
SABRE_miRNA	5718	0	2484	0	3234	MicroRNA profiling of white blood cell—derived RNA to characterize microRNA regulation of gene expression and the relations of microRNA to clinical traits and diseases (phs000363.v10.p8)
SABRE_MRM	665	0	665	0	0	Targeted multiple reaction monitoring proteomics of 33 targets measured in the CVD study; method for quantitative measurement of target proteins (phs000363.v10.p8)
SHARe	9261	1529	3747	97	3888	SNPs from Affymetrix 500 000 mapping array plus Affymetrix 50 000 supplemental array (phs000342.v11.p8)
SHARe_FOLLOWUP	8070	659	3295	96	4020	SNPs performed as follow-up to the SHARe GWAS (phs000342.v11.p8)
SHARe_IMP	8468	954	3559	95	3860	Imputed SNPs using SHARe genotypes (phs000342.v11.p8)

**Appendix II****1. JHS Phenotypic Data**

JHS phenotype data: repeated measures available on participants across 10 years of follow-up

Available phenotype data	Examples
Clinical events (validated and adjudicated)	Coronary heart disease: myocardial infarction, coronary insufficiency, angina, coronary heart disease death, sudden coronary disease death, coronary artery bypass graft surgery, percutaneous transluminal coronary angiography Heart failure Stroke
Subclinical disease	12-Lead ECGs Ultrasound carotid intima-media thickness and carotid stenosis Echocardiographic structure and function (eg, left ventricular mass) CT: coronary artery calcium, abdominal aortic calcium, mitral and aortic valve calcium Cardiac MRI: cardiac structure, cardiac index, and aortic arch plaque Ankle-brachial index
Pulmonary disease and sleep traits	Spirometry Sleep questionnaire

(Continued)

**1. Continued**


---

 JHS phenotype data: repeated measures available on participants across 10 years of follow-up
 

---

Available phenotype data	Examples
Eye disease	Retinopathy grades
Traditional risk factors (directly measured)	Systolic and diastolic blood pressures 24-Hour ambulatory blood pressure Fasting blood glucose, hemoglobin A <sub>1c</sub> , fasting insulin Fasting lipids
Anthropometry (directly measured)	Height, weight, body mass index Waist, hip, thigh, neck circumference CT measures of regional adipose tissue depots: subcutaneous adipose tissue, visceral adipose tissue, pericardial fat, perithoracic fat
Lifestyle	Diet: food frequency questionnaire (calories, supplements) Smoking Exercise (self-report, objective measurement with accelerometry) Alcohol intake
Measures of function	Physical function and mobility Cognitive function (global and multiple domains) Depression (Center for Epidemiologic Studies Depression Scale) Social network
Medications	All 3 examinations
Medicare data	Centers for Medicare & Medicaid Services data since 1991, including <i>International Classification of Diseases</i> codes, charges, medications, procedures on individuals $\geq 65$ years of age enrolled in fee for service
Circulating and urine biomarkers	
Renal function	Creatinine, cystatin C, microalbumin, uric acid
Inflammatory marker panel	Acute-phase reactants: C-reactive protein Selectins: e-selectin, p-selectin
Markers of myocardial injury	Brain natriuretic peptide
Adipokines	Leptin Ghrelin Adiponectin
Molecules interacting with vessel wall and platelets	Plasma homocysteine
Lipid subfractions	Lipoprotein(a), Vertical Auto Profile lipid panel, apolipoproteins A1 and B
Hormones and vitamins	Renin Aldosterone Cortisol Vitamin D, folate, B <sub>12</sub>
Hematology	Hemoglobin Hematocrit White blood cells Red blood cells
Iron/iron storage	Iron Ferritin Serum total iron-binding capacity Serum unsaturated iron-binding capacity
Psychosocial	
Stress	Global stress Weekly stress Major life event Discrimination
Negative affect	Depressive symptoms Anger Hostility
Coping	Social support Social networks
Socioeconomic status	Education Occupation Income Wealth

---

## 2. Available JHS Genetic Data

Consent for genetic analysis: JHS participants completed a “layered” consent document that allowed them to stipulate whether their data and samples could be used for genetic research, to limit research to certain diseases (eg, CVDs and related diseases and risk factors), to specify whether all qualifying investigators or only those collaborating with JHS investigators could analyze their genotype and phenotype data, and to indicate whether for-profit entities could access their data.

Genetic Data	Number of Participants	Comments
CARe (Illumina ITMAT-Broad-CARe [IBC] Array)	2948	SNPs from ≈2100 candidate genes across a range of cardiovascular, metabolic, and inflammatory syndromes, produced as part of the NHLBI CARe project (phs000499.v2.p1).
CARe GWAS	3029	SNPs from Affymetrix 6.0 array (phs000499.v2.p1)
CARe_IMP	3029	Imputed SNPs using CARe GWAS genotypes (dbGaP upload to be initiated)
ESP	401	Whole-exome sequence data produced as part of the NHLBI GO-ESP project (phs000402.v2.p1)
ESP follow-up*	2790	Approximately 240 000 uncommon and rare mainly nonsynonymous coding variants and other selected content of the Illumina V1.0 Exome Chip, with support from NHLBI R01HL107816 (Kathiresan)
Minority Health Genomics and Translational Research Bio-Repository Database	312	Whole-exome sequence data produced as part of the NHLBI project Minority Health Genomics and Translational Research Bio-Repository Database (dbGaP upload pending)
Type 2 Diabetes Genetic Exploration by Next-Generation Sequencing in Multi-Ethnic Samples Project	1036	Whole-exome sequence data produced as part of the National Institute of Diabetes and Digestive and Kidney Diseases Type 2 Diabetes Genetic Exploration by Next-Generation Sequencing in Multi-Ethnic Samples Project (dbGaP upload pending)
Jointly called exomes*	3237	Exome sequencing has been completed for all consenting participants through 4 projects: (1) the Exome Sequencing Project (NHLBI); (2) the Minority Health Genomics and Translational Research Bio-Repository Database (NHLBI); (3) the Type 2 Diabetes Genetic Exploration by Next-Generation Sequencing in Multi-Ethnic Samples Project (National Institute of Diabetes and Digestive and Kidney Diseases); and (4) the Broad Institute MI CIP. Joint calling of all exomes was performed by the Broad/Kathiresan Laboratory.
JHS_AllelicSpectrum_Seq	1979	Deep-coverage targeted resequencing and variant identification for 219 genes in the JHS sample collection, produced as part of the NHLBI Medical Resequencing projects (phs000498.v2.p1)
MARSHFIELD Marker Set 16	1486	Microsatellite markers used in linkage analyses (dbGaP upload to be initiated)
Consortium on Asthma Among African-Ancestry Populations in the Americas Project*	50	Deep-coverage WGS has been completed for 50 participants, including 25 who met criteria for asthma and 25 control subjects, through the Consortium on Asthma Among African-Ancestry Populations in the Americas Project (NHLBI)

\*Completed through an ancillary study. Data access requires consent of the ancillary study principal investigator and potentially others.

into new and improved diagnostic and treatment pathways for CVD.

## Acknowledgments

We would like to acknowledge the FHS investigators and personnel who contributed to the data displayed in Appendix I, Items 1 and 2, and the JHS investigators and personnel who contributed to data displayed in Appendix II, Items 1 and 2.

## Disclosures

None.

## References

- Cheng S, Claggett B, Correia AW, Shah AM, Gupta DK, Skali H, Ni H, Rosamond WD, Heiss G, Folsom AR, Coresh J, Solomon SD. Temporal trends in the population attributable risk for cardiovascular disease: the Atherosclerosis Risk in Communities Study. *Circulation*. 2014;130:820–828.
- Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitzel NO, Brody JA, Khetarpal SA, Crosby JR, Fornage M, Isaacs A, Jakobsdottir J, Feitosa MF, Davies G, Huffman JE, Manichaikul A, Davis B, Lohman K, Joon AY, Smith AV, Grove ML, Zanon P, Redon V, Demissie S, Lawson K, Peters U, Carlson C, Jackson RD, Ryckman KK, Mackey RH, Robinson JG, Siscovick DS, Schreiner PJ, Mychaleckyj JC, Pankow JS, Hofman A, Uitterlinden AG, Harris TB, Taylor KD, Stafford JM, Reynolds LM, Marioni RE, Dehghan A, Franco OH, Patel AP, Lu Y, Hindy G, Gottesman O, Bottinger EP, Melander O, Orho-Melander M, Loos RJ, Duga S, Merlino PA, Farrall M, Goel A, Asselta R, Girelli D, Martinelli N, Shah SH, Kraus WE, Li M, Rader DJ, Reilly MP, McPherson R, Watkins H, Ardisino D, Zhang Q, Wang J, Tsai MY, Taylor HA, Correa A, Griswold ME, Lange LA, Starr JM, Rudan I, Eiriksdottir G, Launer LJ, Ordovas JM, Levy D, Chen YD, Reiner AP, Hayward C, Polasek O, Deary IJ, Borecki IB, Liu Y, Gudnason V, Wilson JG, van Duijn CM, Kooperberg C, Rich SS, Psaty BM, Rotter JJ, O'Donnell CJ, Rice K, Boerwinkle E, Kathiresan S, Cupples LA; NHLBI GO Exome Sequencing Project. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am J Hum Genet*. 2014;94:223–232.
- Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011;12:56–68.
- Dawber TR, Meadors GF, Moore FE. Epidemiologic approaches to heart disease: the Framingham Study. *Am J Public Health*. 1951;41:279–286.
- Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families: the Framingham Offspring Study. *Am J Epidemiol*. 1979;110:281–290.
- 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.
- 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.
- National Heart, Lung, and Blood Institute. SHARE: SNP Health Association Resource. <http://www.nhlbi.nih.gov/research/resources/genetics-genomics/share.htm>. Accessed October 19, 2014
- Musunuru K, Lettre G, Young T, Farlow DN, Pirruccello JP, Ejebe KG, Keating BJ, Yang Q, Chen MH, Lapchuk N, Crenshaw A, Ziaugra L,

- Rachupka A, Benjamin EJ, Cupples LA, Fornage M, Fox ER, Heckbert SR, Hirschhorn JN, Newton-Cheh C, Nizzari MM, Paltoo DN, Papanicolaou GJ, Patel SR, Psaty BM, Rader DJ, Redline S, Rich SS, Rotter JJ, Taylor HA Jr, Tracy RP, Vasan RS, Wilson JG, Kathiresan S, Fabsitz RR, Boerwinkle E, Gabriel SB; NHLBI Candidate Gene Association Resource. Candidate gene association resource (CARE): design, methods, and proof of concept. *Circ Cardiovasc Genet*. 2010;3:267–275.
10. National Heart, Lung, and Blood Institute. NHLBI Grand Opportunity Exome Sequencing Project (ESP). <https://esp.gs.washington.edu/drupal/>. Accessed October 19, 2014.
  11. Boerwinkle E, Heckbert SR. Following-up genome-wide association study signals: lessons learned from Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Targeted Sequencing Study. *Circ Cardiovasc Genet*. 2014;7:332–334.
  12. Morrison AC, Voorman A, Johnson AD, Liu X, Yu J, Li A, Muzny D, Yu F, Rice K, Zhu C, Bis J, Heiss G, O'Donnell CJ, Psaty BM, Cupples LA, Gibbs R, Boerwinkle E; Cohorts for Heart and Aging Research in Genetic Epidemiology (CHARGE) Consortium. Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. *Nat Genet*. 2013;45:899–901.
  13. National Heart, Lung, and Blood Institute. NHLBI Whole Genome Sequencing Project. <http://grants.nih.gov/grants/guide/notice-files/NOT-HL-14-030.html>. Accessed October 19, 2014.
  14. Taylor HA, Jr. The Jackson Heart Study: an overview. *Ethn Dis*. 2005;15:S6-1–S6-3.
  15. Fuqua SR, Wyatt SB, Andrew ME, Sarpong DF, Henderson FR, Cunningham MF, Taylor HA Jr. Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. *Ethn Dis*. 2005;15(suppl 6):S6–S18.
  16. Wilson JG, Rotimi CN, Ekunwe L, Royal CD, Crump ME, Wyatt SB, Steffes MW, Adeyemo A, Zhou J, Taylor HA Jr, Jaquish C. Study design for genetic analysis in the Jackson Heart Study. *Ethn Dis*. 2005;15(suppl 6):S6–S30.
  17. Keku E, Rosamond W, Taylor HA Jr, Garrison R, Wyatt SB, Richard M, Jenkins B, Reeves L, Sarpong D. Cardiovascular disease event classification in the Jackson Heart Study: methods and procedures. *Ethn Dis*. 2005;15(suppl 6):S6–S62.
  18. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives: the ARIC Investigators. *Am J Epidemiol*. 1989; 129:687–702.
  19. 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.
  20. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, Kuller LH, Manolio TA, Mittelmark MB, Newman A. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1:263–276.
  21. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, Giachello AL, Schneiderman N, Raij L, Talavera G, Allison M, Lavange L, Chambless LE, Heiss G. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. 2010;20:629–641.
  22. 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.
  23. Lee ET, Welty TK, Fabsitz R, Cowan LD, Le N-A, Oopik AJ, Cucchiara AJ, Savage PJ, Howard BV. The Strong Heart Study: a study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol*. 1990;132:1141–1155.
  24. Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Control Clin Trials*. 1998;19:61–109.
  25. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation*. 1998;97:1837–1847.
  26. Odden MC, Shlipak MG, Whitson HE, Katz R, Kearney PM, Defilippi C, Shastri S, Sarnak MJ, Siscovick DS, Cushman M, Psaty BM, Newman AB. Risk factors for cardiovascular disease across the spectrum of older age: the Cardiovascular Health Study. *Atherosclerosis*. 2014;237:336–342.
  27. Webber LS, Cresanta JL, Voors AW, Berenson GS. Tracking of cardiovascular disease risk factor variables in school-age children. *J Chronic Dis*. 1983;36:647–660.
  28. Lauer RM, Connor WE, Leaverton PE, Reiter MA, Clarke WR. Coronary heart disease risk factors in school children: the Muscatine study. *J Pediatr*. 1975;86:697–706.
  29. Kuller L, Fisher L, McClelland R, Fried L, Cushman M, Jackson S, Manolio T. Differences in prevalence of and risk factors for subclinical vascular disease among black and white participants in the Cardiovascular Health Study. *Arterioscler Thromb Vasc Biol*. 1998;18:283–293.
  30. Paramsothy P, Knopp R, Bertoni AG, Tsai MY, Rue T, Heckbert SR. Combined hyperlipidemia in relation to race/ethnicity, obesity, and insulin resistance in the Multi-Ethnic Study of Atherosclerosis. *Metabolism*. 2009;58:212–219.
  31. Daviglus ML, Talavera GA, Avilés-Santa ML, Allison M, Cai J, Criqui MH, Gellman M, Giachello AL, Gouskova N, Kaplan RC, LaVange L, Penedo F, Perreira K, Pirzada A, Schneiderman N, Wassertheil-Smoller S, Sorlie PD, Stamler J. Prevalence of major cardiovascular risk factors and cardiovascular diseases among Hispanic/Latino individuals of diverse backgrounds in the United States. *JAMA*. 2012;308:1775–1784.
  32. Wyatt SB, Akyzbekova EL, Wofford MR, Coady SA, Walker ER, Andrew ME, Keahey WJ, Taylor HA, Jones DW. Prevalence, awareness, treatment, and control of hypertension in the Jackson Heart Study. *Hypertension*. 2008;51:650–656.
  33. Mann GV, Pearson G, Gordon T, Dawber TR. Diet and cardiovascular disease in the Framingham study, I: measurement of dietary intake. *Am J Clin Nutr*. 1962;11:200–225.
  34. Van Horn LV, Ballew C, Liu K, Ruth K, McDonald A, Hilner JE, Burke GL, Savage PJ, Bragg C, Caan B. Diet, body size, and plasma lipids-lipoproteins in young adults: differences by race and sex: the Coronary Artery Risk Development in Young Adults (CARDIA) study. *Am J Epidemiol*. 1991;133:9–23.
  35. Howard G, Burke GL, Szklo M, Tell GS, Eckfeldt J, Evans G, Heiss G. Active and passive smoking are associated with increased carotid wall thickness: the Atherosclerosis Risk in Communities Study. *Arch Intern Med*. 1994;154:1277–1282.
  36. Kannel WB, Castelli WP, McNamara PM. Cigarette smoking and risk of coronary heart disease: epidemiologic clues to pathogenesis: the Framingham study. *Natl Cancer Inst Monogr*. 1968;28:9–20.
  37. Kuller LH, Shemanski L, Psaty BM, Borhani NO, Gardin J, Haan MN, O'Leary DH, Savage PJ, Tell GS, Tracy R. Subclinical disease as an independent risk factor for cardiovascular disease. *Circulation*. 1995;92:720–726.
  38. Detrano R, Guerci AD, Carr JJ, Bild DE, Burke G, Folsom AR, Liu K, Shea S, Szklo M, Bluemke DA, O'Leary DH, Tracy R, Watson K, Wong ND, Kronmal RA. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *N Engl J Med*. 2008;358:1336–1345.
  39. Chambless LE, Folsom AR, Clegg LX, Sharrett AR, Shahar E, Nieto FJ, Rosamond WD, Evans G. Carotid wall thickness is predictive of incident clinical stroke: the Atherosclerosis Risk in Communities (ARIC) study. *Am J Epidemiol*. 2000;151:478–487.
  40. Auchincloss AH, Diez Roux AV, Mujahid MS, Shen M, Bertoni AG, Carnethon MR. Neighborhood resources for physical activity and healthy foods and incidence of type 2 diabetes mellitus: the Multi-Ethnic Study of Atherosclerosis. *Arch Intern Med*. 2009;169:1698–1704.
  41. Henderson C, Diez-Roux AV, Kiefe CI, West DE, Jacobs DR, Williams DR. Neighborhood characteristics, individual-level socioeconomic factors, and depressive symptoms in young adults: the CARDIA Study. *J Epidemiol Community Health*. 2005;59:322–328.
  42. Olson NC, Cushman M, Lutsey PL, McClure LA, Judd S, Tracy RP, Folsom AR, Zakai NA. Inflammation markers and incident venous thromboembolism: the REasons for Geographic And Racial Differences in Stroke (REGARDS) Cohort [published online ahead of print October 7, 2014]. *J Thromb Haemost*. doi: 10.1111/jth.12742. <http://onlinelibrary.wiley.com/doi/10.1111/jth.12742/abstract?jsessionid=ECC98E260A87C4B10C3A6CA07977B031.f02t01>. Accessed October 19, 2014.
  43. Zakai NA, Katz R, Jenny NS, Psaty BM, Reiner AP, Schwartz SM, Cushman M. Inflammation and hemostasis biomarkers and cardiovascular risk in the elderly: the Cardiovascular Health Study. *J Thromb Haemost*. 2007;5:1128–1135.
  44. Fontes JD, Yamamoto JF, Larson MG, Wang N, Dallmeier D, Rienstra M, Schnabel RB, Vasan RS, Keaney JF Jr, Benjamin EJ. Clinical correlates of change in inflammatory biomarkers: the Framingham Heart Study. *Atherosclerosis*. 2013;228:217–223.
  45. Truesdale K, Lewis CE, Loria C, Cai J, Stevens J, Schreiner PJ. Changes in risk factors for cardiovascular disease by baseline weight status in young adults who maintain or gain weight over 15 years: the CARDIA Study. *Int J Obesity*. 2006;30:1397–1407.
  46. Wong ND, Nelson JC, Granston T, Bertoni AG, Blumenthal RS, Carr JJ, Guerci A, Jacobs DR Jr, Kronmal R, Liu K, Saad M, Selvin E, Tracy R, Detrano R. Metabolic syndrome, diabetes, and incidence and progression of coronary calcium: the Multiethnic Study of Atherosclerosis study. *JACC Cardiovasc Imaging*. 2012;5:358–366.

47. Sims M, Diez Roux AV, Boykin S, Sarpong D, Gebreab SY, Wyatt SB, Hickson D, Payton M, Ekunwe L, Taylor HA. The socioeconomic gradient of diabetes prevalence, awareness, treatment, and control among African Americans in the Jackson Heart Study. *Ann Epidemiol*. 2011;21:892–898.
48. Lemaitre RN, Furberg CD, Newman AB, Hulley SB, Gordon DJ, Gottdiener JS, McDonald RH Jr, Psaty BM. Time trends in the use of cholesterol-lowering agents in older adults: the Cardiovascular Health Study. *Arch Intern Med*. 1998;158:1761–1768.
49. Adams NB, Lutsey PL, Folsom AR, Herrington DH, Sibley CT, Zakai NA, Ades S, Burke GL, Cushman M. Statin therapy and levels of hemostatic factors in a healthy population: the Multi-Ethnic Study of Atherosclerosis. *J Thromb Haemost*. 2013;11:1078–1084.
50. Alonso A, Krijthe BP, Aspelund T, Steps KA, Pencina MJ, Moser CB, Sinner MF, Sotoodehnia N, Fontes JD, Janssens AC, Kronmal RA, Magnani JW, Witteman JC, Chamberlain AM, Lubitz SA, Schnabel RB, Agarwal SK, McManus DD, Ellinor PT, Larson MG, Burke GL, Launer LJ, Hofman A, Levy D, Gottdiener JS, Kääb S, Couper D, Harris TB, Soliman EZ, Stricker BH, Gudnason V, Heckbert SR, Benjamin EJ. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc*. 2013;2:e000102.
51. Li J, Lange LA, Duan Q, Lu Y, Singleton AB, Zonderman AB, Evans MK, Li Y, Taylor HA, Willis MS, Nalls M, Wilson JG, Lange EM. Genome-wide admixture and association study of serum iron, ferritin, transferrin saturation and total iron binding capacity in African Americans [published online ahead of print September 15, 2014]. *Hum Mol Genet*. doi: 10.1093/hmg/ddu454. <http://hmg.oxfordjournals.org/content/early/2014/09/29/hmg.ddu454.long>. Accessed October 19, 2014.
52. Matsushita K, Mahmoodi BK, Woodward M, Emberson JR, Jafar TH, Jee SH, Polkinghorne KR, Shankar A, Smith DH, Tonelli M, Warnock DG, Wen CP, Coresh J, Gansevoort RT, Hemmelgarn BR, Levey AS; Chronic Kidney Disease Prognosis Consortium. Comparison of risk prediction using the CKD-EPI equation and the MDRD study equation for estimated glomerular filtration rate. *JAMA*. 2012;307:1941–1951.
53. Miedema MD, Duprez DA, Misialek JR, Blaha MJ, Nasir K, Silverman MG, Blankstein R, Budoff MJ, Greenland P, Folsom AR. Use of coronary artery calcium testing to guide aspirin utilization for primary prevention: estimates from the Multi-Ethnic Study of Atherosclerosis. *Circ Cardiovasc Qual Outcomes*. 2014;7:453–460.

---

KEY WORDS: American Heart Association ■ cardiovascular diseases ■ genome