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

Citation

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.

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. 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 metagenomics, 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. 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 (epi)genetic and posttranslational modulators.

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...
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 methodological analyses (novel omics platforms) and a biorepository (real or virtual) for biological specimens from existing or future population study participants. These resources are briefly detailed below.

**The FHS as a Component of CV-GPS**

**Design and Component Cohorts**

The FHS consists of ≈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.4–7 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 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 ≈1000 participants from the Original Cohort who were alive at the time of DNA collection; DNA was also extracted on an additional ≈500 deceased Original Cohort participants from available whole blood as part of the SNP Health Association Resource (SHARe) project (see below). At FHS, ≈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.9 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 ≈200,000 functional exonic variants via Illumina V1.0 Exome Chip. FHS has completed whole-exome sequencing in ≈2975 participants from several projects (including the NHLBI GO Exome Sequence Project10 and the CHARGE targeted Sequencing project)11 and low-pass whole-genome sequencing (WGS) in ≈850 participants.12 Of note, the genotyping (including whole-exome sequencing and WGS) that has been conducted have 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.13 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
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,
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. 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 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. 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 >900 000 SNPs and has >900 000 probes for copy number variation. These data have been imputed to 37 million SNPs using the 100 Genomes project reference panel (version 3, March 2012 release). In addition, through the CARe project, 2948 individuals have genotyping data for 50 000 SNPs on the gene-centric Illumina Cardiochip (IBC Array), which provides dense tagging of 2100 candidate genes for CVD. Genotyping has been completed in 2790 participants for 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 (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 genotypic data, provides an excellent basis for genome-phenome analysis and integrative genomics.
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 sibs 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 (n=15,792; age, 45–64 years at baseline examination [1987–1989]; white and black participants); the Coronary Artery Risk Development in Young Adults Study (n=5115; age, 18–30 years at the baseline examination [1985–1986]; white and black participants); the Cardiovascular Health Study (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 (n=16,400; 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 (n=93,676 postmenopausal women; age, 50–79 years at baseline examination [1991–1994]).

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 genomics, 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.

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**

<table>
<thead>
<tr>
<th>Available phenotype data</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical events (validated and adjudicated)</td>
<td>Coronary heart disease: myocardial infarction, coronary insufficiency, angina, coronary heart disease death, sudden coronary disease death, coronary artery bypass graft surgery, percutaneous transluminal coronary angioplasty, 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</td>
</tr>
</tbody>
</table>

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 glucose, hemoglobin A1c, 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 ≥100 years

<table>
<thead>
<tr>
<th>Available phenotype data</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medications</td>
<td>All examinations</td>
</tr>
<tr>
<td>Medicare data</td>
<td>Centers for Medicare &amp; Medicaid Services data since 1991, including International Classification of Diseases codes, charges, medications, procedures on individuals ≥65 years of age enrolled in fee for service</td>
</tr>
<tr>
<td>Bone health</td>
<td>Fractures, osteoporosis, Osteoarthritis</td>
</tr>
<tr>
<td>Circulating and urine biomarkers</td>
<td></td>
</tr>
<tr>
<td>Renal function</td>
<td>Creatinine, cystatin C, microalbumin, uric acid</td>
</tr>
<tr>
<td>Inflammatory marker panel</td>
<td>Acute-phase reactants: C-reactive protein</td>
</tr>
<tr>
<td></td>
<td>Cytokines: interleukin-6, tumor necrosis factor-α, tumor necrosis factor receptor 2, osteoprotegerin</td>
</tr>
<tr>
<td></td>
<td>Selectins: P-selectin, CD40 ligand</td>
</tr>
<tr>
<td></td>
<td>Cell adhesion: intercellular adhesion molecule</td>
</tr>
<tr>
<td></td>
<td>Chemokines: monocyte chemoattractant protein-1</td>
</tr>
<tr>
<td></td>
<td>Oxidative stress: myeloperoxidase, isoprostanes</td>
</tr>
<tr>
<td></td>
<td>Lipoprotein-associated phospholipase A2 mass and activity</td>
</tr>
<tr>
<td>Markers of hemostasis, thrombosis</td>
<td>Fibrinogen, factor Vllc, von Willebrand factor, D-dimer, plasminogen activator inhibitor-1, platelet reactivity</td>
</tr>
<tr>
<td>Markers of myocardial injury</td>
<td>Troponin I, growth differentiation factor-15, ST-2, brain natriuretic peptide, N-terminal atrial natriuretic peptide</td>
</tr>
<tr>
<td>Adipokines</td>
<td>Leptin, leptin receptor</td>
</tr>
<tr>
<td></td>
<td>α-Fetuin, resistin, ghrelin</td>
</tr>
<tr>
<td></td>
<td>Retinol binding protein-4</td>
</tr>
<tr>
<td></td>
<td>Fatty acid binding protein-4</td>
</tr>
<tr>
<td></td>
<td>Adiponectin</td>
</tr>
<tr>
<td>Growth factors</td>
<td>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</td>
</tr>
<tr>
<td>Molecules interacting with vessel wall and platelets</td>
<td>Matrix remodeling markers: matrix metalloproteinase-9, matrix metalloproteinase-3, tissue inhibitor of matrix metalloproteinase-1, N-terminal propeptide of type III procollagen</td>
</tr>
<tr>
<td></td>
<td>Plasma homocysteine, asymmetrical dimethyl arginine</td>
</tr>
<tr>
<td>Markers of brain injury</td>
<td>β-Amyloid, clusterin</td>
</tr>
<tr>
<td>Lipid subfractions</td>
<td>Lipoprotein(a), apolipoproteins A1, B48, B100, C1, CII, E, H, J</td>
</tr>
<tr>
<td>Hormones and vitamins</td>
<td>Renin-angiotensin-aldosterone pathway</td>
</tr>
<tr>
<td></td>
<td>Thyroid function (eg, thyroid-stimulating hormone)</td>
</tr>
<tr>
<td></td>
<td>Sex steroid hormones</td>
</tr>
<tr>
<td></td>
<td>Vitamin D, folate, B₁₂, B₆, vitamin E</td>
</tr>
</tbody>
</table>

2. Available FHS Genetic Data

Framingham Participants With Consent for Genetic Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>ALL Cohorts</th>
<th>Original Cohort</th>
<th>Offspring Cohort</th>
<th>Omni Gr 1</th>
<th>Third Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>14,271</td>
<td>5079</td>
<td>5013</td>
<td>101</td>
<td>4078</td>
</tr>
<tr>
<td>CARe</td>
<td>7,547</td>
<td>647</td>
<td>3023</td>
<td>0</td>
<td>3877</td>
</tr>
<tr>
<td>CHARGES_TARGET</td>
<td>10,95</td>
<td>36</td>
<td>957</td>
<td>0</td>
<td>102</td>
</tr>
<tr>
<td>CHARGES_WES_FREEZE1</td>
<td>621</td>
<td>13</td>
<td>487</td>
<td>0</td>
<td>121</td>
</tr>
<tr>
<td>CHARGES_WES_FREEZE3</td>
<td>850</td>
<td>13</td>
<td>715</td>
<td>0</td>
<td>122</td>
</tr>
<tr>
<td>CHARGES_WES_FREEZE4</td>
<td>1,271</td>
<td>45</td>
<td>1,107</td>
<td>0</td>
<td>119</td>
</tr>
<tr>
<td>CHARGES_WGS_FREEZE1</td>
<td>320</td>
<td>8</td>
<td>282</td>
<td>0</td>
<td>30</td>
</tr>
</tbody>
</table>

SNPs from ~2100 candidate genes across a range of cardiovascular, metabolic, and inflammatory syndromes, produced as part of the NHLBI CARe project (phs000282.v12.p8).

targeted sequencing produced as part of the CHARGE sequencing consortium (phs000651.v3,p8)

First release of whole-exome sequencing produced as part of the CHARGE sequencing consortium (phs000651.v3,p8)

Third release of whole-exome sequencing produced as part of the CHARGE sequencing consortium (phs000651.v3,p8)

Fourth release of whole-exome sequencing produced as part of the CHARGE sequencing consortium (phs000651.v3,p8)

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>ALL Cohorts</th>
<th>Original Cohort</th>
<th>Offspring Cohort</th>
<th>Omni Gr 1</th>
<th>Third Generation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESP</td>
<td>464</td>
<td>0</td>
<td>291</td>
<td>0</td>
<td>173</td>
</tr>
<tr>
<td>MARSH</td>
<td>4112</td>
<td>454</td>
<td>1400</td>
<td>0</td>
<td>2258</td>
</tr>
<tr>
<td>METHYLATION</td>
<td>2202</td>
<td>0</td>
<td>2202</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NHGRI_MEDSEQ</td>
<td>1703</td>
<td>0</td>
<td>1703</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>OMNI5</td>
<td>2472</td>
<td>0</td>
<td>2472</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ONEK</td>
<td>1342</td>
<td>258</td>
<td>1084</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PERLEGEN</td>
<td>1649</td>
<td>0</td>
<td>1649</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>PGA</td>
<td>1749</td>
<td>0</td>
<td>1749</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RTPCR</td>
<td>1943</td>
<td>0</td>
<td>1943</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SABRE_EXPRESS</td>
<td>5622</td>
<td>0</td>
<td>2442</td>
<td>0</td>
<td>3180</td>
</tr>
<tr>
<td>SABRE_IMMUNO</td>
<td>7315</td>
<td>0</td>
<td>3264</td>
<td>0</td>
<td>4051</td>
</tr>
<tr>
<td>SABRE_ITRAQ</td>
<td>269</td>
<td>0</td>
<td>269</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SABRE_miRNA</td>
<td>5718</td>
<td>0</td>
<td>2484</td>
<td>0</td>
<td>3234</td>
</tr>
<tr>
<td>SABRE_MRM</td>
<td>665</td>
<td>0</td>
<td>665</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SHARe</td>
<td>9261</td>
<td>1529</td>
<td>3747</td>
<td>97</td>
<td>3888</td>
</tr>
<tr>
<td>SHARe_FOLLOWUP</td>
<td>8070</td>
<td>659</td>
<td>3295</td>
<td>96</td>
<td>4020</td>
</tr>
<tr>
<td>SHARe_IMP</td>
<td>8468</td>
<td>954</td>
<td>3559</td>
<td>95</td>
<td>3860</td>
</tr>
</tbody>
</table>

Whole-exome sequence data produced as part of the NHLBI GO-ESP project (phs000401.v6.p8)

Microsatellite markers used in linkage analyses (phs000342.v11.p8)

Methylation data; Offspring Cohort examination 8 (phs000724.v1.p8)

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)

SNPs from the Illumina HumanOmni5M-4v1 array designed to target variation down to 1% minor allele frequency (phs000342.v11.p8)

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)

SNPs in inflammatory candidate genes (phs000342.v11.p8)

SNP data of candidate genes involved in human congenital heart disease and cardiomyopathy (phs000342.v11.p8)

Reverse transcriptase—polymerase chain reaction gene expression, Offspring Cohort examination 8 (phs000325.v1.p5)

Gene expression profiling of white blood cell—derived RNA to characterize the genomic signatures of atherosclerosis and metabolic syndrome (phs000363.v10.p8)

Immunomassays of 180 circulating protein biomarkers of atherosclerosis and metabolic syndrome (phs000363.v10.p8)

iTRAQ Px data set 135 case/control pairs; iTRAQ is used in proteomics to study quantitative changes in the proteome (phs000363.v10.p8)

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)

Targeted multiple reaction monitoring proteomics of 33 targets measured in the CVD study; method for quantitative measurement of target proteins (phs000363.v10.p8)

SNPs from Affymetrix 500 000 mapping array plus Affymetrix 50 000 supplemental array (phs000342.v11.p8)

SNPs performed as follow-up to the SHARe GWAS (phs000342.v11.p8)

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

<table>
<thead>
<tr>
<th>Available phenotype data</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical events (validated and adjudicated)</td>
<td>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</td>
</tr>
<tr>
<td>Subclinical disease</td>
<td>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</td>
</tr>
<tr>
<td>Pulmonary disease and sleep traits</td>
<td>Spirometry, Sleep questionnaire</td>
</tr>
</tbody>
</table>

(Continued)
1. **Continued**

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

<table>
<thead>
<tr>
<th>Available phenotype data</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye disease</td>
<td>Retinopathy grades</td>
</tr>
<tr>
<td>Traditional risk factors (directly measured)</td>
<td>Systolic and diastolic blood pressures, 24-Hour ambulatory blood pressure, Fasting blood glucose, hemoglobin A1c, fasting insulin, Fasting lipids</td>
</tr>
<tr>
<td>Anthropometry (directly measured)</td>
<td>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</td>
</tr>
<tr>
<td>Lifestyle</td>
<td>Diet: food frequency questionnaire (calories, supplements), Smoking, Exercise (self-report, objective measurement with accelerometry), Alcohol intake</td>
</tr>
<tr>
<td>Measures of function</td>
<td>Physical function and mobility, Cognitive function (global and multiple domains), Depression (Center for Epidemiologic Studies Depression Scale), Social network</td>
</tr>
<tr>
<td>Medications</td>
<td>All 3 examinations</td>
</tr>
<tr>
<td>Medicare data</td>
<td>Centers for Medicare &amp; Medicaid Services data since 1991, including International Classification of Diseases codes, charges, medications, procedures on individuals ≥65 years of age enrolled in fee for service</td>
</tr>
</tbody>
</table>
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 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 data, and to indicate whether for-profit entities could access their data.

### Genetic Data

<table>
<thead>
<tr>
<th>Genetic Data</th>
<th>Number of Participants</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>CARe (Illumina ITMAT-Broad-CARe (IBC Array))</td>
<td>2948</td>
<td>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).</td>
</tr>
<tr>
<td>CARe GWAS</td>
<td>3029</td>
<td>SNPs from Affymetrix 6.0 array (phs000499.v2.p1).</td>
</tr>
<tr>
<td>CARe IMP</td>
<td>3029</td>
<td>Imputed SNPs using CARe GWAS genotypes (dbGaP upload to be initiated)</td>
</tr>
<tr>
<td>ESP</td>
<td>401</td>
<td>Whole-exome sequence data produced as part of the NHLBI GO-ESP project (phs000402.v2.p1)</td>
</tr>
<tr>
<td>ESP follow-up*</td>
<td>2790</td>
<td>Approximately 240,000 uncommon and rare nonsexonsonymous coding variants and other selected content of the Illumina V1.0 Exome Chip, with support from NHLBI R01HL1107816 (Kathiresan)</td>
</tr>
<tr>
<td>Minority Health Genomics and Translational Research Bio-Repository Database</td>
<td>312</td>
<td>Whole-exome sequence data produced as part of the NHLBI project Minority Health Genomics and Translational Research Bio-Repository Database (dbGaP upload pending)</td>
</tr>
<tr>
<td>Type 2 Diabetes Genetic Exploration by Next-Generation Sequencing in Multi-Ethnic Samples Project</td>
<td>1036</td>
<td>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)</td>
</tr>
<tr>
<td>Jointly called exomes*</td>
<td>3237</td>
<td>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 MiCIP. Joint calling of all exomes was performed by the Broad/Kathiresan Laboratory.</td>
</tr>
<tr>
<td>JHS_AllelicSpectrum_Seq</td>
<td>1979</td>
<td>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).</td>
</tr>
<tr>
<td>MARSHFIELD Marker Set 16</td>
<td>1486</td>
<td>Microsatellite markers used in linkage analyses (dbGaP upload to be initiated)</td>
</tr>
<tr>
<td>Consortium on Asthma Among African-Ancestry Populations in the Americas Project*</td>
<td>50</td>
<td>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)</td>
</tr>
</tbody>
</table>

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

References


**KEY WORDS:** American Heart Association ■ cardiovascular diseases ■ genome