Original Article

Multiancestry Study of Gene–Lifestyle Interactions for Cardiovascular Traits in 610 475 Individuals From 124 Cohorts

Design and Rationale

D. C. Rao, PhD; Yun J. Sung, PhD; Thomas W. Winkler, PhD; Karen Schwander, MS; Ingrid Borecki, PhD; L. Adrienne Cupples, PhD; W. James Gauderman, PhD; Kenneth Rice, PhD; Patricia B. Munroe, PhD; Bruce M. Psaty, MD, PhD; on behalf of the CHARGE Gene-Lifestyle Interactions Working Group*

Background—Several consortia have pursued genome-wide association studies for identifying novel genetic loci for blood pressure, lipids, hypertension, etc. They demonstrated the power of collaborative research through meta-analysis of study-specific results.

Methods and Results—The Gene-Lifestyle Interactions Working Group was formed to facilitate the first large, concerted, multiancestry study to systematically evaluate gene–lifestyle interactions. In stage 1, genome-wide interaction analysis is performed in 53 cohorts with a total of 149 684 individuals from multiple ancestries. In stage 2 involving an additional 71 cohorts with 460 791 individuals from multiple ancestries, focused analysis is performed for a subset of the most promising variants from stage 1. In all, the study involves up to 610 475 individuals. Current focus is on cardiovascular traits including blood pressure and lipids, and lifestyle factors including smoking, alcohol, education (as a surrogate for socioeconomic status), physical activity, psychosocial variables, and sleep. The total sample sizes vary among projects because of missing data. Large-scale gene–lifestyle or more generally gene–environment interaction (G×E) meta-analysis studies can be cumbersome and challenging. This article describes the design and some of the approaches pursued in the interaction projects.

Conclusions—The Gene-Lifestyle Interactions Working Group provides an excellent framework for understanding the lifestyle context of genetic effects and to identify novel trait loci through analysis of interactions. An important and novel feature of our study is that the gene–lifestyle interaction (G×E) results may improve our knowledge about the underlying mechanisms for novel and already known trait loci. **(*Circ Cardiovasc Genet*. 2017;10:e001649. DOI: 10.1161/CIRCGENETICS.116.001649.)**

Key Words: blood pressure ■ genome-wide association study ■ life style ■ meta-analysis ■ molecular epidemiology

Remarkable advances in genomics, including the Human Genome Project and 1000 Genomes (1000G) Project, have revolutionized methods for genetic dissection of common complex diseases and disease traits. Using genome-wide association studies (GWAS), large consortia, such as CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology),[1] ICBP (International Consortium of Blood Pressure), AGEN (Asian Genetic Epidemiology Network), GLGC (Global Lipids Genetics Consortium), and DIAGRAM (Diabetes Genetics Replication and Meta-Analysis), have identified hundreds of common genetic variants associated with many common complex disease traits (https://www.genome.gov/26525384/catalog-of-published-genomewide-association-studies/). However, most of the identified genetic variants explain small proportions of the trait heritability, mostly through small main effects of common variants. It has been recognized that this focus on main effects may have become a barrier to further progress.[2,3]

See Editorial by Kirk
See Clinical Perspective

Hypertension and dyslipidemia are common complex disorders that contribute to 2 of the leading causes of death (cardiovascular and cerebrovascular diseases) and exhibit

Received October 18, 2016; accepted February 14, 2017.
*A list of all CHARGE Gene-Lifestyle Interactions Working Group participants is given in the Data Supplement.
Guest Editor for this article was Christopher Semsarian, MBBS, PhD, MPH.
The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health.
The Data Supplement is available at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.116.001649/-/DC1.
Correspondence to Dabeeru C. Rao, PhD, Division of Biostatistics, Washington University in St. Louis, 660 S Euclid Ave, Campus Box 8067, St. Louis, MO 63110. E-mail rao@wustl.edu

significant patterns of health disparity among racial/ancestral groups in the United States.[4,5] Although lifestyle factors have long been recognized as risk factors, modulation of the effects of genetic variants by lifestyle factors and the underlying candidate pathobiological mechanisms have not received much attention. Understanding these genetic modifiers is important because it may provide valuable clues for lifestyle-based interventions that may result in a more successful management of these health conditions through personalized therapies and may explain part of the missing heritability.[2,6]

The Gene-Lifestyle Interactions Working Group (hereafter referred to as this study) investigates gene–lifestyle interactions for uncovering more of the unexplained genetic variance in blood pressure (BP) and lipids and for gaining insights into the biological mechanisms influencing these important morbid conditions. We will do this by leveraging the CHARGE infrastructure and the extensive resources of existing studies in multiple ancestries that have data on phenotypes, lifestyle factors, and dense genotype data.

Research involving gene–environment (G×E) interactions is now being reported.[7–9] These demonstrated the promise of G×E interactions for identifying genetic variants with large effects.[10–13] For example, mean triglyceride levels are 23 mg/dL lower in physically active than in sedentary individuals (88 versus 111 md/dL) who carry a C-allele at rs2070744 in *NOS*3, but there is little difference by physical activity status in TT homozygotes.[11] This shows the utility of G×E interactions for using genetic information to identify subpopulations in whom modifying the environmental factors is beneficial,[14–16] and that the main effect (of the genetic variant) alone is inadequate to inform lifestyle interventions that need to be personalized based on genotype.[17,18] In addition, G×E interactions may provide additional insight into biological mechanisms and pathways.

This is the first large, concerted, multiancestry study to evaluate gene–lifestyle interactions systematically using data from 610 475 individuals. Large-scale G×E meta-analysis studies can be cumbersome and challenging. This article describes the design and some of the approaches pursued in our ongoing Gene-Lifestyle Interaction projects.

Study Design

The CHARGE Consortium

This study leverages the infrastructure created by the CHARGE consortium,[1] which created several Working Groups (WGs), an internal wiki site, guidelines for collaboration and authorship, and periodic CHARGE meetings where WGs meet in person.

The Gene-Lifestyle Interactions WG

A new WG has been established for pursuing the major goals of this study. The WG includes investigators and analysts from the large group of studies participating in stage 1 (genome-wide discovery) as discussed later. Another large group of studies participates in stage 2 (focused discovery/replication). The WG is assisted by a Coordinating Center at Washington University in St. Louis.

This study operates through the WG, an Analysis Committee, a Harmonization Committee, and multiple Project Teams. The WG meets twice a year at CHARGE meetings and meets by conference call twice a month. Research direction and priorities are set by the WG. The analysis and harmonization committees meet together once a year and by conference calls twice a month. All harmonization and analytic issues are resolved by these committees. There are multiple Project Teams, each leading interaction analyses for a combination of the phenotypes (BP or lipids) and lifestyle domains (smoking, alcohol, education, physical activity, psychosocial, sleep). Finally, institutional review board approval has been obtained for the study.

Mission and Aim

The overall mission of the WG is to promote and facilitate large collaborative analysis of gene–lifestyle interactions on disease traits across a large number of cohorts from multiple ancestries. Primarily, the WG aims to better understand the lifestyle context of genetic effects and to discover new trait loci through analysis of interactions, thereby explaining part of the missing heritability[2] in the disease traits. An important and novel feature of our study is that the gene–lifestyle interaction (G×E) results may improve our knowledge about the underlying mechanisms for novel as well as already known trait loci.

Primary Hypothesis

We hypothesize that lifestyle (environment) variables modulate some of the genetic effects on cardiovascular traits and that accounting for lifestyle factors and gene–lifestyle interactions in genome-wide scans will identify multiple novel genetic variants.

Phenotypes and Lifestyle Variables

The primary phenotypes include BP and lipids. An analysis plan in the Data Supplement discusses data definitions and adjustments. Future initiatives may consider other cardiometabolic traits in collaboration with other WGs.

The primary BP phenotypes are resting/sitting systolic blood pressure (SBP; mm Hg) and diastolic blood pressure (DBP; mm Hg). For individuals taking any antihypertensive (BP lowering) medications, their SBP and DBP values are first adjusted by adding 15 mm Hg to SBP and adding 10 mm Hg to DBP. Mean arterial pressure and pulse pressure are also derived, using the adjusted SBP and DBP values:

a. Mean arterial pressure=DBP+(SBP−DBP)/3, and
b. Pulse pressure=SBP−DBP

The primary lipids phenotypes are high-density lipoprotein cholesterol (mg/dL), triglycerides (mg/dL), and low-density lipoprotein cholesterol (LDL, mg/dL), either directly assayed (LDL_{da}) or derived using the Friedewald equation (LDL_F). For individuals with triglycerides >400 mg/dL, only directly assayed LDL (LDL_{da}) is used. When using nonfasting samples or fasting <8 hours, only LDL_{da} and high-density lipoprotein are used (not LDL_F or triglycerides). Log transformations are used for high-density lipoprotein and triglycerides, and LDL is adjusted for statin use (see the analysis plan in the Data Supplement).

The initial set of dichotomized lifestyle are smoking (current smoking and ever smoking), alcohol consumption (current drinking, current regular drinking, and quantity of drinks [>7 drinks per week]), education (as a measure of socioeconomic status; some college, and graduated college), physical activity (physically inactive), psychosocial attributes (depression, trait anxiety, and social support), and sleep duration (short sleep and long sleep). Future initiatives may consider other domains, such as diet.

GWAS Data

Dosages derived from 1000G imputation are the primary resource for GWAS analysis. The 1000G imputations are based on the all ancestry panel from 1000G Phase I Integrated Release Version 3 Haplotypes (2010-11 data freeze, 2012-03-14 haplotypes) that contains haplotypes of 1092 individuals of all ancestral backgrounds. Dosages based on HapMap Phase II/III reference panel are used if 1000G imputations are not available for a specific study. In general, rare variants (mean allele frequency <1%) and poorly imputed variants (Rsq <0.1) are excluded. Variants mapping to sex chromosomes or mitochondria have also been excluded. Although we refer to single-nucleotide polymorphism (SNP) variants, the imputed data also include indels (insertions and deletions).

Participating Studies and Ancestry Groups

Five ancestry groups are represented: European (EA), African (AA), Hispanic (HA), Asian (AS), and Brazilian admixed (BR). Men and women between the ages of 18 and 80 years are included in the analyses. Although the participating studies are based on different study designs and populations, most of them have data on BP and lipid traits, a range of lifestyle variables, and genotypes across the genome. In total, this study comprises up to 610 475 individuals.

Stage 1 (Genome-Wide Discovery)

A total of 32 studies with data on 53 cohorts (Table I in the Data Supplement) participate in the discovery phase (stage 1), which involves genome-wide interaction analyses. In total, this stage includes up to 95 911 EA, 27 116 AA, 8805 HA, 13 438 AS, and 4414 BR individuals, to an overall total of 149 684 individuals in stage 1.

Stage 2 (Focused Discovery/Replication)

A total of 46 studies with data on 71 cohorts (Table II in the Data Supplement) participate in stage 2, which involves analyses of small sets of variants that were identified in stage 1 as either genome-wide significant (with $P<10^{-8}$) or suggestive (with $P<10^{-6}$). In total, this stage includes up to 290 552 EA, 7785 AA, 13 522 HA, and 148 932 AS individuals to a total of 460 791 individuals in stage 2. There are no BR cohorts in stage 2.

Analysis Models

The participating studies have considerable prior experience contributing to GWAS-based consortia studying the main effects of common variants (without interactions). For G×E work, existing analysis pipelines had to be modified. Based on extensive discussions with the Analysis Committee and the Working Group, an Analysis Plan was developed, addressing critical issues, including data preparation, analysis models, analysis methods, and software packages. Individual project teams made appropriate modifications to the analysis plan as needed. The most critical elements are summarized below. An example of a full analysis plan (education-lipids) is provided in the Data Supplement.

We consider 3 different analysis models, each with slightly different purposes.

Joint Model (Model 1)

This is our primary model that features joint analysis of the effects of the SNP, lifestyle, and their interaction. For each combination of phenotype (Y) and lifestyle exposure variable (E), each study fits the following linear model, separately by ancestry:

$$Y \sim E + SNP + E * SNP + C, \text{ or more formally,}$$
$$E(Y) = \beta_0 + \beta_E E + \beta_G SNP + \beta_{GE} E * SNP + \beta_C C$$

where SNP is the dosage of the genetic variant and C is the set of covariates (age, sex, principal components for controlling population stratification effects as needed, and other study-specific covariates), and therefore β_C is a vector; body mass index was specifically excluded as a covariate so that lifestyle interactions with related pathway genes (such as inflammation genes) can be identified. Participating studies provide estimates of β_G and β_{GE} along with their covariance matrix. If E is dichotomous (E=0 or 1), the SNP effect (β_G) represents the SNP effect in those who are unexposed (environmental variable E=0), and thus needs to be interpreted with caution. If E is continuous, it is often desirable to center it on its sample mean, so that β_G approximates the overall effect of the SNP on Y (as is estimated by model 2). In either case, the SNP effect is context dependent and therefore should not be interpreted as the main effect.

Model 1 was used by all studies in both stages. In addition to model 1, each study in stage 1 (only) uses at least 1 of 2 additional models presented below, depending on the specific needs of the respective project.

Main Effects Model (Model 2)

Analysis of the main effect only: For each phenotype (Y), each study fits the following linear model, separately by ancestry:

$$Y \sim SNP + C, \text{ or more formally,}$$
$$E(Y) = \lambda_0 + \lambda_G SNP + \lambda_C C.$$

Model 2 is used as a benchmark to identify which of our discoveries from the joint model would be found using analysis of main effects alone. Some projects also fit this model separately in the exposed and unexposed groups (ie, they performed stratified analysis) and provide a 1 *df* test of the interaction term as well as a 2 *df* joint test of the SNP and interaction effects.[19,20] For each analysis, participating studies provide estimates of λ_G and its SE. Stratified analysis and the joint analysis using model 1 in stage 1 cohorts have been shown to yield largely similar results.[21] Stratified analysis can help reduce inflation of type I error rates by fitting separate covariate effects and error variances by strata.[22–24]

Refined Main Effects Model (Model 3)

Analysis of the SNP and lifestyle effects, without interaction: For each phenotype (Y) and lifestyle exposure variable

(E), each study fits the following linear model, separately by ancestry:

$$Y \sim E + SNP + C, \text{ or more formally,}$$

$$E(Y) = \gamma_0 + \gamma_E E + \gamma_G SNP + \gamma_C C.$$

Model 3 is used to identify which of our discoveries from the joint model would be missed when the interaction term is not used. For each analysis, participating studies provide estimates of γ_G and its SE.

Analysis Methods

Analysis Methods for Low Frequency and Common Variants

We identify novel loci through SNP×E interaction effects alone or jointly with the SNP effects (or only through SNP effects in models 2 and 3). For continuous traits, the joint test of the SNP and SNP×E interaction effects is known to be powerful for this aim.[20,25,26] Because our interaction projects involve many studies, we rely on existing methods and software, such as ProbABEL, Sandwich, and MMAP (see the analysis plan in the Data Supplement), or those that are straightforward to implement using these tools.

Testing the Significance of the SNP and the SNP×E Interaction Effects

In model 1, the focus is on the test of the interaction effect and the joint effects of the SNP and the interaction. The interaction effect (β_{GE}) is evaluated using a 1-*df* Wald test. The effects of both SNP (β_G) and interaction (β_{GE}) are tested jointly, using a 2-*df* Wald test.[25] In model 2, which does not include E or SNP×E terms, λ_G is the familiar main effect of the SNP which is tested using a 1-*df* Wald test. A 1-*df* Wald test is also used in model 3 for evaluating the SNP effect (γ_G) in the presence of E, which may be referred to as the refined SNP effect or context-dependent SNP effect. In all cases, we will use the robust Wald tests by using robust estimates of the SEs and covariances to protect against misspecification of the mean model.[27,28] When the SNP effect is weak and the SNP×E interaction effect is moderate, the joint 2 *df* test has been shown to be more powerful than either the 1-*df* test of the SNP effect or the 1-*df* test of the interaction effect alone.[25] The increase in power for the 2-*df* over 1-*df* test can be particularly dramatic, especially when the type I error rate is controlled at low levels (eg, 5×10^{-8}) as in this project.[29]

Analyses Needed From Each Cohort

Each cohort performs a genome-wide analysis of the SNP and SNP×E interaction effects, separately within each ancestry by accounting for possible population stratification (see the Analysis Plan in the Data Supplement), and provides estimates of regressions coefficients and robust estimates of the corresponding SEs and covariance. Because the model is based on a standard regression framework, software to compute the relevant statistics is widely available. For studies of unrelated individuals, standard commands and the R sandwich package[30] implement bivariate robust covariance estimates for SNP-specific analyses. To implement the analyses for all SNPs, the R interface in PLINK[31] may be used; ProbABEL[32] also provides appropriate utilities. For family studies in which relatedness must be taken into account, programs such as GenABEL/MixABEL[33] and MMAP (J.O., PhD, unpublished data, 2017, personal communication) implement mixed models that allow for relatedness. All cohorts analyze their data using these methods/software following a detailed Analysis Plan and upload results to a secure server.

Meta-Analysis for Combining Results Across Studies

To combine estimates of the regression coefficients and their corresponding 2×2 covariance matrix provided by each cohort, we use the joint meta-analysis method developed by Manning et al[26] who modified METAL[34] to handle this joint 2 *df* meta-analysis. The joint meta-analysis provides inference on the SNP and SNP×E interaction effects pooled across all cohorts. Manning et al[7] used this approach and demonstrated power enhancement for detecting interactions. We use the modified METAL for the joint meta-analysis and use METAL for carrying out meta-analysis of the 1-*df* analyses (interaction effect in model 1, main effect in model 2, and refined SNP effect in model 3). We use a genome-wide significance threshold of 5×10^{-8} for identifying significant results and use 10^{-6} for identifying suggestive results. Heterogeneity χ^2 test is used to test for differences in any of the regression coefficients among the contributing studies. Early results indicate minimal differences even in the interaction coefficient.

Quality Control

Quality assurance is emphasized by preparing detailed analysis plans with step by step instructions for preparing and analyzing data, and formatting results for uploading (see the Education-Lipids analysis plan in the Data Supplement). Extensive quality control (QC) measures are used for processing all study-specific results centrally by each project team, at 2 levels. Study-level QC involves reviewing and harmonization of each individual result file separately. Meta-level QC involves reviewing and harmonizing results files across all available cohorts for a single analysis (eg, comparing summary statistics across all SBP-Current Smoking-Model1 discovery cohorts). QC was performed using customized EasyQC scripts that provide a wide variety of QC checks for GWAS results.[35]

Analysis of Interactions Involving Rare Variants

Power of the joint test of SNP and SNP×E for testing individual rare variants is limited primarily because of their low frequency. Burden tests[36–39] collapse all rare variants in a genomic region (typically a gene) into a single burden variable (essentially a mega variant, giving each subjects' total dosage across a gene) and regress the phenotype on the burden variable to test for the combined effects of all rare variants in the region/gene. We apply the 2 *df* test directly to each burden variable. Each cohort creates burden variables by collapsing variants within the genomic regions using a mean allele frequency (pooled across studies) threshold (eg, mean allele frequency <0.01). We then perform meta-analysis of these results, similar to the meta-analysis described earlier but now with as

many burden variables as the number of genomic regions. To assess the significance for the analysis of rare variants, we will use a Bonferroni-corrected significance threshold ($\alpha=0.05/N_b$, where N_b=number of burden variables). The CHARGE consortium has provided detailed analysis guidelines for exome chip data and the Coordinating Center has used some of these rare variant methods.[40–43]

Analysis of Stage 1 and Stage 2 Results

Primary publications resulting from the various analyses in stages 1 and 2 are pursuing 2 approaches as shown in Figure 1: combined analysis of stages 1 and 2 and traditional discovery/replication.

Combined Analysis of Stages 1 and 2

This approach can be more powerful than other approaches.[44] For a given combination of phenotype and lifestyle, all significant and suggestive results (with $\alpha=10^{-6}$) from stage 1 cohorts and the corresponding results from stage 2 cohorts are pooled through meta-analysis (first within each stage and then meta-analyzing the 2-stage–specific meta-analyses) separately by ancestry. A significance threshold of $\alpha=5\times10^{-8}$ is used to identify significant results from the combined meta-analysis results. Finally, all ancestry-specific meta-analyses are meta-analyzed as an approximate transancestry analysis for identifying additional associations (if any) that are missed by ancestry-specific analyses.

Traditional Discovery/Replication Analysis

In this approach, all genome-wide significant results are identified from stage 1 results only, separately by ancestry, using a significance threshold of $\alpha=5\times10^{-8}$. Stage 2 results are then used to formally replicate the stage 1 findings, using appropriate Bonferroni correction such as 0.05 divided by the number of independent novel loci discovered in stage 1. Variants that are suggestive but not significant in stage 1 are only considered in the combined analysis approach.

The combined approach is more powerful than the traditional approach. However, the traditional approach can identify additional novel validated loci missed by the combined approach (as shown most recently using a slight variation of this approach[45]). This justifies using both approaches. If only 1 approach were to be used, the combined one is the method of choice.

Statistical Power for Detecting Associations

With the overall sample size used, this study is well powered for identifying novel discoveries even with moderately small effect sizes. To demonstrate this, we illustrate the sample sizes required to achieve at least 80% power to identify the genetic

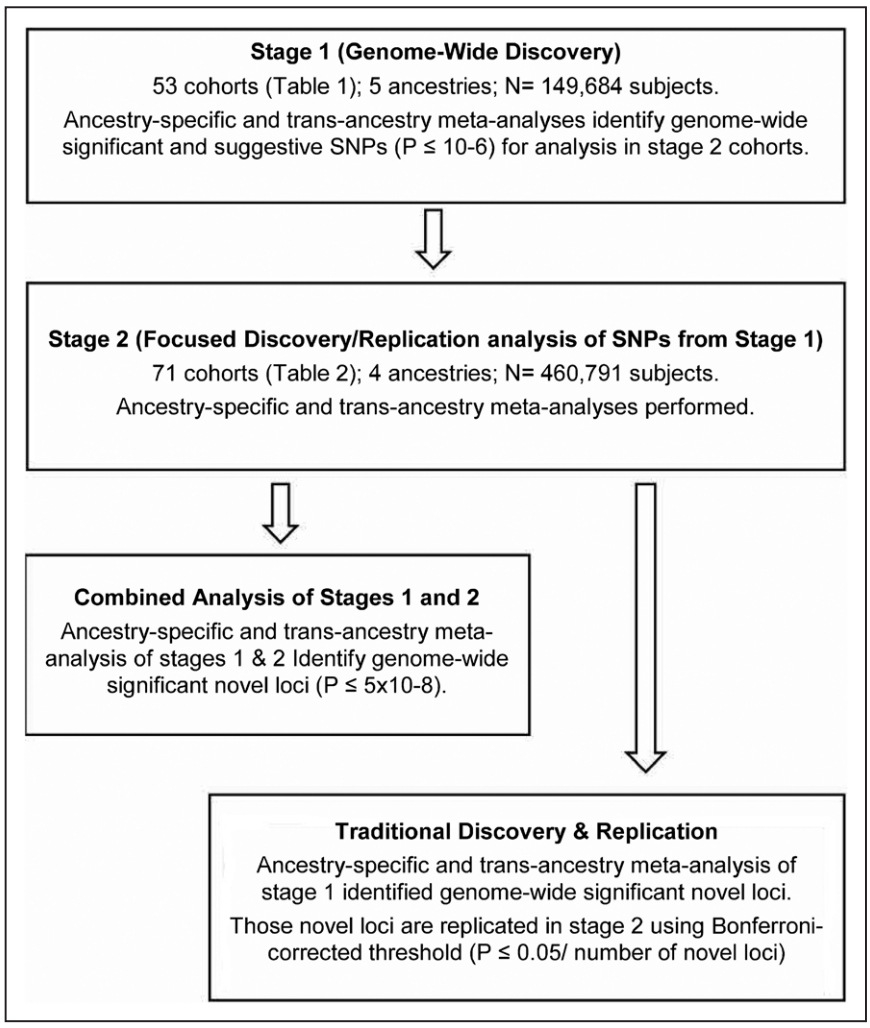


Figure 1. Overall flow of analyses. Combined analysis leverages the full power of stages 1 and 2. The traditional discovery and replication approach identifies additional loci missed by the combined approach. Both approaches can be used for maximizing discovery. SNP indicates single nucleotide polymorphism.

(G) effect and the G×E interaction effect using the 2 *df* joint test for a range of model parameters. We used QUANTO,[46] which computes power and sample size for both disease and quantitative trait studies of genes (G), environment factors (E), and G×E interactions. For our study of quantitative traits, the required sample sizes depend on the proportions of variance explained by the G (R2G), the lifestyle factor (R2E), and their interaction effect (R2GE). A wide range of R2E values yielded similar results, and therefore we fixed R2E=0.1% and examined the effect of varying the other 2 parameters. Although low-frequency variants explain large proportions of variance in some cases,[47] we limited this investigation to lower R2G values of 0.01%, 0.02%, 0.05%, and 0.1% because most variants identified through GWAS have much smaller effect sizes. Figure 2 shows the sample sizes required for a range of R2GE values corresponding to each of the 4 values for R2G using a significance threshold of 5×10^{-8}. These values are smaller than what we found in our preliminary studies (not reported), suggesting that our power estimates may be conservative.

The sample sizes should be more than adequate for 80% power in EA and AA using stage 1 samples alone, so long as the SNP effect is not small (eg, R2G>0.05%). In fact, for R2G=0.05%, significance level of 5×10^{-8}, and the stage 1 sample sizes shown in Table I in the Data Supplement, the minimum detectable R2GE at 80% power are <0.01%, 0.11%, 0.44%, and 0.27% for EA, AA, HA, and AS, respectively. When stages 1 and 2 are combined, even smaller effect sizes are detectable (although the exact calculations are complex because stage 2 studies did not carry out genome-wide interaction analyses). In any case, the combined sample size of stages 1 and 2 seems well poised for powerful discoveries even with smaller effect sizes than assumed in these estimates.

Discussion

Current Status and Anticipated Benefits

Our study has made considerable progress to date. Four projects have completed all analyses in stages 1 and 2 and are processing the final results for publications (smoking-BP, smoking-lipids, alcohol-BP, and alcohol-lipids). In addition, 3 other projects (education-BP, education-lipids, and PA-lipids) have completed stage 1 analyses for which stage 2 analyses are in progress, and 2 projects (psychosocial-BP and psychosocial-lipids) have completed stage 1 analyses and are preparing for stage 2 analysis. More projects are getting underway. We think that these projects will make major contributions to the genetic dissection of cardiovascular traits and that the G×E analysis can help improve understanding of the mechanisms underlying the novel as well as known loci that have been identified previously through main effects.

What Are the Unique Benefits of Our Approach? How Critical Is the Consideration of Lifestyle and Interactions (Models 1 and 3)?

Emerging results indicate that a large proportion of novel findings originate from models 1 and 3 (ie, results that would be missed by limiting analyses to main effects; model 2). This suggests that inclusion of the lifestyle context or gene–lifestyle interaction is important for identifying novel signals.

Collaboration Levels Are Unprecedented

In an area where direct competition among study groups was the norm until about a decade ago, collaborative GWAS-based consortia such as CHARGE represent an innovative model for research. Through working together,

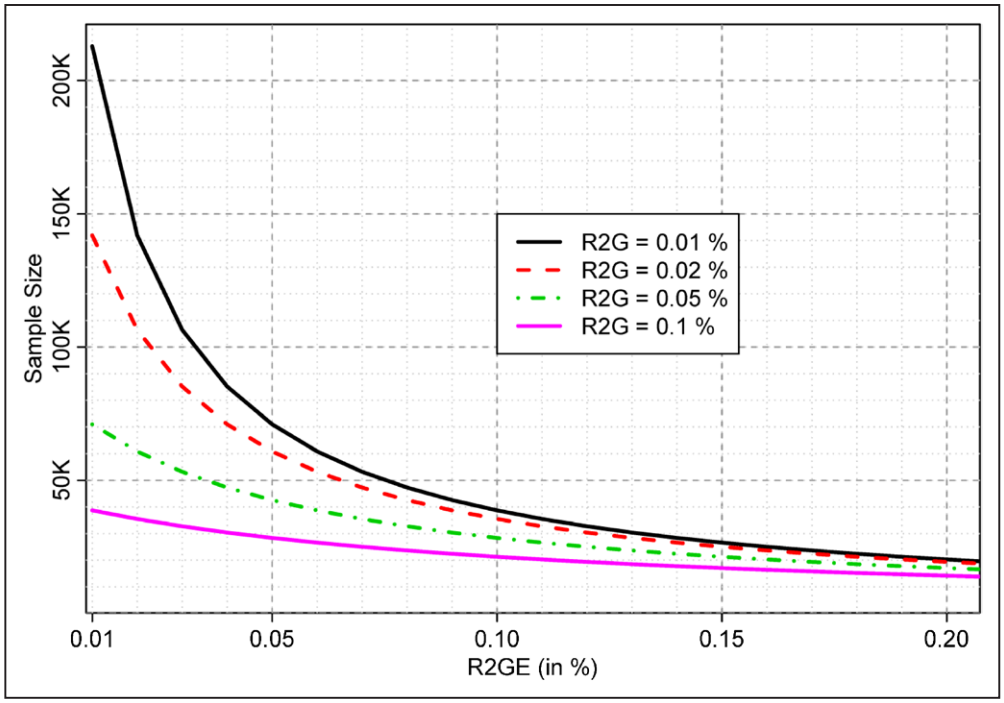


Figure 2. Sample sizes needed for 80% power using the 2 *df* joint test. Sample size (Y-axis) is plotted as a function of the percent variance explained by the interaction (R2GE; X-axis), for each of 4 different values of the percent variance explained by the genetic effect (R2G); that due to the lifestyle factor (R2E) is fixed at 0.1% (see the text).

the contributing studies have achieved much more than they could have working alone. The Gene-Lifestyle Interactions Working Group takes this model further, assembling 610 475 subjects in 124 cohorts. Other studies with appropriate data are welcome to join. Although the collaborative nature of the work requires some compromises (eg, using standard software and meta-analysis of relatively simple analyses), the results will hopefully deepen what has already been learned from GWAS.

Acknowledgments

We thank several members of the Working Group (WG) for their overall contributions to the WG, notably Hugues Aschard, Sharon Kardia, Ruth Loos, Alisa Manning, Jeff O'Connell, Michael Province, Patricia Peyser, Jerome Rotter, Xiaofeng Zhu, among others. The full list of the WG members can be found at http://depts.washington.edu/chargeco/wiki/Gene-Lifestyle_Interactions. We also thank Matthew Brown for his critical help in preparing some of the materials for this publication. Study descriptions and study-specific acknowledgments are included in the Data Supplement along with an example analysis plan.

Sources of Funding

This multiancestry study of gene–lifestyle interactions is sponsored by R01HL118305 from the National Heart, Lung, and Blood Institute (NHLBI), National Institute of Health. The CHARGE (Cohorts for Heart and Aging Research in Genomic Epidemiology) infrastructure on which this study is based is also sponsored by another NHLBI grant HL105756.

Disclosures

Dr Psaty serves on the DSMB of a clinical trial funded by Zoll LifeCor and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. The other authors report no conflicts.

Appendix

From the Division of Biostatistics (D.C.R., Y.J.S., K.S.) and Department of Genetics (I.B.), Washington University in St. Louis, School of Medicine, St. Louis, MO; Department of Genetic Epidemiology, University of Regensburg, Rogensburg, Germany (T.W.W.); Department of Biostatistics, Boston University School of Public Health, Boston & NHLBI Framingham Heart Study, MA (L.A.C.); Division of Biostatistics, Department of Preventive Medicine, University of Southern California, Los Angeles (W.J.G.); Department of Biostatistics (K.R.) and Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services (B.M.P.), University of Washington, and Group Health Research Institute, Group Health Cooperative, Seattle, WA; Clinical Pharmacology, William Harvey Research Institute & NIHR Barts Cardiovascular Biomedical Research Unit, Queen Mary, University of London, London, United Kingdom (P.B.M.).

References

1. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, Rotter JI, et al; CHARGE Consortium. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ Cardiovasc Genet*. 2009;2:73–80. doi: 10.1161/CIRCGENETICS.108.829747.
2. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747–753. doi: 10.1038/nature08494.
3. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90:7–24. doi: 10.1016/j.ajhg.2011.11.029.
4. Murphy SL, Xu J, Kochanek KD. *Deaths: Preliminary Data for 2010, in National Vital Statistics Reports*. Hyattsville, MD: National Center for Health Statistics; 2012.
5. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, et al; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics–2012 update: a report from the American Heart Association. *Circulation*. 2012;125:e2–e220. doi: 10.1161/CIR.0b013e31823ac046.
6. Zheng JS, Arnett DK, Lee YC, Shen J, Parnell LD, Smith CE, et al. Genome-wide contribution of genotype by environment interaction to variation of diabetes-related traits. *PLoS One*. 2013;8:e77442. doi: 10.1371/journal.pone.0077442.
7. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al; DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium; Multiple Tissue Human Expression Resource (MUTHER) Consortium. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet*. 2012;44:659–669. doi: 10.1038/ng.2274.
8. Hutter CM, Mechanic LE, Chatterjee N, Kraft P, Gillanders EM; NCI Gene-Environment Think Tank. Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. *Genet Epidemiol*. 2013;37:643–657. doi: 10.1002/gepi.21756.
9. Sung YJ, de Las Fuentes L, Schwander KL, Simino J, Rao DC. Gene-smoking interactions identify several novel blood pressure loci in the Framingham Heart Study. *Am J Hypertens*. 2015;28:343–354. doi: 10.1093/ajh/hpu149.
10. Montasser ME, Shimmin LC, Hanis CL, Boerwinkle E, Hixson JE. Gene by smoking interaction in hypertension: identification of a major quantitative trait locus on chromosome 15q for systolic blood pressure in Mexican-Americans. *J Hypertens*. 2009;27:491–501.
11. Higashibata T, Hamajima N, Naito M, Kawai S, Yin G, Suzuki S, et al. eNOS genotype modifies the effect of leisure-time physical activity on serum triglyceride levels in a Japanese population. *Lipids Health Dis*. 2012;11:150. doi: 10.1186/1476-511X-11-150.
12. Grarup N, Andreasen CH, Andersen MK, Albrechtsen A, Sandbaek A, Lauritzen T, et al. The -250G>A promoter variant in hepatic lipase associates with elevated fasting serum high-density lipoprotein cholesterol modulated by interaction with physical activity in a study of 16,156 Danish subjects. *J Clin Endocrinol Metab*. 2008;93:2294–2299. doi: 10.1210/jc.2007-2815.
13. Parnell LD, Blokker BA, Dashti HS, Nesbeth PD, Cooper BE, Ma Y, et al. CardioGxE, a catalog of gene-environment interactions for cardiometabolic traits. *BioData Min*. 2014;7:21. doi: 10.1186/1756-0381-7-21.
14. Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet*. 2005;6:287–298. doi: 10.1038/nrg1578.
15. Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. *Am J Epidemiol*. 2009;169:219–226. doi: 10.1093/aje/kwn353.
16. Thomas D. Gene–environment-wide association studies: emerging approaches. *Nat Rev Genet*. 2010;11:259–272. doi: 10.1038/nrg2764.
17. Taylor JY, Maddox R, Wu CY. Genetic and environmental risks for high blood pressure among African American mothers and daughters. *Biol Res Nurs*. 2009;11:53–65. doi: 10.1177/1099800409334817.
18. Green ED, Guyer MS. Charting a course for genomic medicine from base pairs to bedside. *Nature*. 2011;470:204–213. doi: 10.1038/nature09764.
19. Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, et al; DIAGRAM Consortium; MAGIC Investigators. Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet*. 2013;9:e1003500. doi: 10.1371/journal.pgen.1003500.
20. Aschard H, Hancock DB, London SJ, Kraft P. Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. *Hum Hered*. 2010;70:292–300. doi: 10.1159/000323318.
21. Sung YJ, Winkler TW, Manning AK, Aschard H, Gudnason V, Harris TB, et al. An empirical comparison of joint and stratified frameworks for studying G × E interactions: systolic blood pressure and smoking in the CHARGE Gene-Lifestyle Interactions Working Group. *Genet Epidemiol*. 2016;40:404–415. doi: 10.1002/gepi.21978.
22. Vanderweele TJ, Ko YA, Mukherjee B. Environmental confounding in gene-environment interaction studies. *Am J Epidemiol*. 2013;178:144–152. doi: 10.1093/aje/kws439.
23. Dudbridge F, Fletcher O. Gene-environment dependence creates spurious gene-environment interaction. *Am J Hum Genet*. 2014;95:301–307. doi: 10.1016/j.ajhg.2014.07.014.
24. Keller MC. Gene × environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol Psychiatry*. 2014;75:18–24. doi: 10.1016/j.biopsych.2013.09.006.

25. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered*. 2007;63:111–119. doi: 10.1159/000099183.
26. Manning AK, LaValley M, Liu CT, Rice K, An P, Liu Y, et al. Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP × environment regression coefficients. *Genet Epidemiol*. 2011;35:11–18. doi: 10.1002/gepi.20546.
27. Tchetgen Tchetgen EJ, Kraft P. On the robustness of tests of genetic associations incorporating gene-environment interaction when the environmental exposure is misspecified. *Epidemiology*. 2011;22:257–261. doi: 10.1097/EDE.0b013e31820877c5.
28. Voorman A, Lumley T, McKnight B, Rice K. Behavior of QQ-plots and genomic control in studies of gene-environment interaction. *PLoS One*. 2011;6:e19416. doi: 10.1371/journal.pone.0019416.
29. Morris N, Elston R. A note on comparing the power of test statistics at low significance levels. *Am Stat*. 2011;65:164–166. doi: 10.1198/tast.2011.10117.
30. Zeileis A. Object-oriented computation of sandwich estimators. *J Stat Softw*. 2006;1-16.
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575. doi: 10.1086/519795.
32. Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics*. 2010;11:134. doi: 10.1186/1471-2105-11-134.
33. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*. 2007;23:1294–1296. doi: 10.1093/bioinformatics/btm108.
34. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–2191. doi: 10.1093/bioinformatics/btq340.
35. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R, et al; Genetic Investigation of Anthropometric Traits (GIANT) Consortium. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc*. 2014;9:1192–1212. doi: 10.1038/nprot.2014.071.
36. Morgenthaler S, Thilly WG. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat Res*. 2007;615:28–56. doi: 10.1016/j.mrfmmm.2006.09.003.
37. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet*. 2008;83:311–321. doi: 10.1016/j.ajhg.2008.06.024.
38. Madsen BE, Browning SR. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet*. 2009;5:e1000384. doi: 10.1371/journal.pgen.1000384.
39. Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet Epidemiol*. 2010;34:188–193. doi: 10.1002/gepi.20450.
40. Sung YJ, Rice TK, Rao DC. Application of collapsing methods for continuous traits to the Genetic Analysis Workshop 17 exome sequence data. *BMC Proc*. 2011;5(suppl 9):S121. doi: 10.1186/1753-6561-5-S9-S121.
41. Sun YV, Sung YJ, Tintle N, Ziegler A. Identification of genetic association of multiple rare variants using collapsing methods. *Genet Epidemiol*. 2011;35(suppl 1):S101–S106. doi: 10.1002/gepi.20658.
42. Mallaney C, Sung YJ. Rare variant analysis of blood pressure phenotypes in the Genetic Analysis Workshop 18 whole genome sequencing data using sequence kernel association test. *BMC Proc*. 2014;8(suppl 1 Genetic Analysis Workshop 18Vanessa Olmo):S10. doi: 10.1186/1753-6561-8-S1-S10.
43. Sung YJ, Basson J, Rao DC. Whole genome sequence analysis of the simulated systolic blood pressure in Genetic Analysis Workshop 18 family data: long-term average and collapsing methods. *BMC Proc*. 2014;8(suppl 1 Genetic Analysis Workshop 18Vanessa Olmo):S12. doi: 10.1186/1753-6561-8-S1-S12.
44. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*. 2006;38:209–213. doi: 10.1038/ng1706.
45. Surendran P, Drenos F, Young R, Warren H, Cook JP, Manning AK, et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet*. 2016;48:1151–1161. doi: 10.1038/ng.3654.
46. Gauderman W, Morrison J. QUANTO 1:2: a computer program for power and sample size calculations for genetic-epidemiology studies. 2009; http://biostats.usc.edu/software. Last accessed 04/12/2017.
47. Bowden DW, An SS, Palmer ND, Brown WM, Norris JM, Haffner SM, et al. Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. *Hum Mol Genet*. 2010;19:4112–4120. doi: 10.1093/hmg/ddq327.

CLINICAL PERSPECTIVE

Gene–lifestyle interaction studies provide an excellent framework for understanding the lifestyle context of genetic effects on disease traits. Understanding these genetic modifiers is important because it may provide valuable clues for lifestyle-based interventions which may result in a more successful management of these health conditions through personalized therapies. An important feature of our study is that such interaction results may improve our knowledge about the underlying mechanisms for novel as well as already known trait loci.

Circulation
Cardiovascular Genetics

Multiancestry Study of Gene–Lifestyle Interactions for Cardiovascular Traits in 610 475 Individuals From 124 Cohorts: Design and Rationale

D. C. Rao, Yun J. Sung, Thomas W. Winkler, Karen Schwander, Ingrid Borecki, L. Adrienne Cupples, W. James Gauderman, Kenneth Rice, Patricia B. Munroe and Bruce M. Psaty on behalf of the CHARGE Gene-Lifestyle Interactions Working Group*

Circ Cardiovasc Genet. 2017;10:
doi: 10.1161/CIRCGENETICS.116.001649
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1942-325X. Online ISSN: 1942-3268

SUPPLEMENTAL MATERIAL

Table 1. Studies and ancestry groups participating in Stage 1 (Genome-wide discovery)

No	Study/ Cohort	Type of Study	European Ancestry	African Ancestry	Hispanic Ancestry	Asians	Brazilian Admixed
1	AGES	Population study of GxE in elderly	2,410	-	-	-	-
2	ARIC	Population-based study of Atherosclerosis	9,465	2,862	-	-	-
3	Baependi	Family-based study of CVD traits	-	-	-	-	873
4	CARDIA	Population-based study of CVD traits	1,649	945	-	-	-
5	CHS	Population-based study of CVD traits	2,975	734	-	-	-
6	CROATIA	Population-based study of Croatians: Vis	483	-	-	-	-
		Population-based study of Croatians: Korcula	456	-	-	-	-
7	Fam HS	Family study of CVD related traits	3,683	617	-	-	-
8	FHS	Longitudinal family study of CVD traits	8,195	-	-	-	-
9	GENOA	Sibling study of Atherosclerosis and HT	1,064	941	-	-	-
10	GenSalt	Family study of salt sensitivity	-	-	-	1,835	-
11	GENSCOT	Population-based study in Scotland	6,439	-	-	-	-
12	GOLDN	Family-based study of HT & CVD traits	820	-	-	-	-
13	HANDLS	Diversity study of aging and CVD traits	-	903	-	-	-
14	Health ABC	Study of health, aging and body comp	1,663	1,136	-	-	-
15	HERITAGE	Fam study of responses to exercise	499	-	-	-	-
16	HUFS	Family study of hypertension in AA	-	1,686	-	-	-
17	HyperGEN	Family-based study of HT & CVD traits	1,251	1,240	-	-	-
18	JHS	Population-based study of CVD traits	-	2,134	-	-	-
19	Maywood-L	Population study of CVD traits in AA	-	75	-	-	-
20	Maywood-N	Study of CVD traits in Nigerians	-	1,229	-	-	-
21	MESA	Family-based study of Atherosclerosis	2,591	1,594	1,455	748	-
22	Mt. Sinai IPM	Hospital-based / Biobank patients	1,480	3,101	3,973	-	-
23	NEO	Population-based study of obesity related traits	5,735	-	-	-	-
24	Pelotas	Population-based birth cohort in Brazil	-	-	-	-	3,541

25	RS	Rotterdam study of CVD traits: RS1	4,990	-	-	-	-
		Rotterdam study of CVD traits: RS2	1,998	-	-	-	-
		Rotterdam study of CVD traits: RS3	2,966	-	-	-	-
		Rotterdam family study of CVD traits: RS-ERF	2,491	-	-	-	-
26	SCES	Singapore Chinese eye study	-	-	-	1,848	-
27	SCHS	Singapore Chinese Health Study: Cases	-	-	-	674	-
		Singapore Chinese Health Study: Controls	-	-	-	1,218	-
28	SiMES	Singapore Malay eye study	-	-	-	2,531	-
29	SINDI	Singapore Indian eye study	-	-	-	2,491	-
30	SP2	Singapore 2: 1M	-	-	-	949	-
		Singapore 2: 610	-	-	-	1,144	-
31	WGHS	Popn-based; genomics; women's health	22,983	-	-	-	-
32	WHI	Popn-based study of women's health	-	7,919	3,377	-	-
		Popn-based study of women's health: GARNET	4,423	-	-	-	-
		Popn-based study of women's health: WHIMS	5,202	-	-	-	-
TOTALS			**95,911**	**27,116**	**8,805**	**13,438**	**4,414**

Note: Sample sizes may vary across phenotype-exposure combinations due to missing data.

Table 2. Studies and ancestry groups participating in Stage 2 (Focused Discovery/Replication)

No	Study/ Cohort	Type of Study	European Ancestry	African Ancestry	Hispanic Ancestry	Asians	Brazilian Admixed
1	AADHS	Case-Control study of diabetes in AAs	-	584	-	-	-
2	ASCOT-SC	Population-based study of cardiac outcomes	2,389	-	-	-	-
3	BBJ	Population-based biobank in Japan	-	-	-	126,413	-
4	BES	Population-based study of eye disease:610	-	-	-	601	-
		Popn-based study of eye disease:OmniExpress	-	-	-	545	-
5	BRIGHT	Population-based study of hypertension	1,823	-	-	-	-
6	CAGE	Popn-based study of CVD traits: Amagasaki	-	-	-	952	-
7	CARL	Family-based study of auditory traits in Italy	462	-	-	-	-
8	CFS	Family-based study of sleep apnea in AA	-	561	-	-	-
9	DESIR1	Epidemiological study on insulin resistance	697	-	-	-	-
10	DFTJ	Popn-based study of health and retirement	-	-	-	1,406	-
11	DHS	Family-based study of diabetes	1,173	-	-	-	-
12	DR's EXTRA	Unrelated study of exercise training	1,230	-	-	-	-
13	EGCUT	Popn-based biobank in Estonia:OmniExpress	5,937	-	-	-	-
		Popn-based biobank in Estonia:CoreExome	4,911	-	-	-	-
		Popn-based biobank in Estonia:Human370CNV	1,870	-	-	-	-
14	EPIC	Popn-based study of cancer/nutrition in Europe	20,458	-	-	-	-
15	Fenland	Popn-based study of metabolic traits: GWAS	1,345	-	-	-	-
		Popn-based study of metabolic traits: OMICS	8,471	-	-	-	-
16	FUSION	Case-Control Study of NIDDM:CASES	674	-	-	-	-
		Case-Control Study of NIDDM:CONTROLS	277	-	-	-	-
17	FVG	Family-based study of auditory traits in Italy	951	-	-	-	-
18	GeneSTAR	Family study of atherosclerosis risk	1,699	1,107	-	-	-
19	GLACIER	Population-based study of lobular cardinoma	5,909	-	-	-	-
20	GRAPHIC	Population-based study of arterial pressure	1,010	-	-	-	-
21	HRS	Population-based study of health & retirement	8,367	1,993	-	-	-

22	HyperGEN	Family-based study of HT & CVD traits:AXIOM	-	418	-	-	-
23	InterAct	Case-contrl study of T2DM:CoreExome:CASES	3,996	-	-	-	-
		CC study of T2DM:CoreExome:SUBCOHORT	6,405	-	-	-	-
		Case-control study of T2DM:GWAS:CASES	2,793	-	-	-	-
		CC study of T2DM:GWAS:SUBCOHORT	3,188	-	-	-	-
24	IRAS	Popn-based study of atherosclerosis:IRASC	-	-	185	-	-
		Family-based study of atherosclerosis:IRASFS	-	-	957	-	-
25	JUPITER	Population-based study of lipids and statin use	8,400	1,606	-	-	-
26	KORA	Population-based German research cohort:S3	3,095	-	-	-	-
		Population-based German research cohort:S4	3,770	-	-	-	-
27	LBC	Lothian Birth Cohort study:1921	511	-	-	-	-
		Lothian Birth Cohort study:1936	996	-	-	-	-
28	Lifelines	Biobank cohort in the Netherlands	12,323	-	-	-	-
29	LLFS	Family-based study on aging	3,133	-	-	-	-
30	LOLIPOP	London Population study of CVD traits: EW610	927	-	-	-	-
		London Population study of CVD traits: EWA	582	-	-	-	-
		London Population study of CVD traits: EWP	644	-	-	-	-
		London Population study of CVD traits: IA317	-	-	-	2,059	-
		London CC study of CVD traits: IA610-case	-	-	-	2,791	-
		London CC study of CVD traits: IA610-ctrl	-	-	-	3,757	-
		London Population study of CVD traits: IAP	-	-	-	501	-
		London Popn study of CVD traits: OmniEE	-	-	-	899	-
31	LOYOLA	Population-based Jamaican cohort of BP:GXE	-	612	-	-	-
		Population-based Jamaican health cohort:SPT	-	904	-	-	-
32	METSIM	Men-only unrelated study; metabolic syndrome	8,353	-	-	-	-
33	OBA	Unrelated French obese cases	669	-	-	-	-
34	PROCARDIS	Case-control study of CAD:Cases	5,651	-	-	-	-
		Case-control study of CAD:Controls	1,668	-	-	-	-
35	RHS	Popn-based cohort of metabolic syndrome	-	-	-	2,468	-
36	SHEEP	Case-control study of CVD traits:Cases	1,165	-	-	-	-

		Case-control study of CVD traits:Controls	1,528	-	-	-	-
37	SHIP	Population-based health study:0 Cohort	4,046	-	-	-	-
		Population-based health study:Trend Cohort	982	-	-	-	-
38	SMWHS	Population-based men/women health study	-	-	-	3,862	-
39	SOL	Hispanic community health study	-	-	12,380	-	-
40	TAICHI	Popn-based study of atherosclerosis:Zhonghua	-	-	-	1,505	-
41	THRV	Population-based Taiwan study of hypertension	-	-	-	287	-
42	TRAILS	Population-based study of adolescents	1,266	-	-	-	-
43	TUDR	Population-based study of diabetes	-	-	-	886	-
44	TWINGENE	Family-based study of twins in Sweden	5,358	-	-	-	-
45	UK Biobank	Population-based Biobank in the UK	137,426	-	-	-	-
46	YFS	Population-based CV study of young adults	2,024	-	-	-	-
TOTALS			**290,552**	**7,785**	**13,522**	**148,932**	**0**

Note: Sample sizes may vary across phenotype-exposure combinations due to missing data.

Gene-Lifestyle Interactions

Analysis Plan for Education and Lipids

Karen Schwander (karen2@wubios.wustl.edu); Yun Ju Sung (yunju@wubios.wustl.edu);
DC Rao (rao@wubios.wustl.edu)

DRAFT: 09/30/15

ABRIDGED VERSION FOR THIS PUBLICATION

AIM:

The primary goals of this investigation are:

1. To identify novel genetic loci for lipid traits that may not have been detected previously when interaction was not considered,
2. To characterize gene-education interactions in known and novel lipid loci, and
3. To integrate results across race-ethnic groups by carrying out meta-analysis for possible identification of additional novel loci.

TYPES OF STUDIES:

1. Cross-sectional (single visit) studies of unrelated subjects.
2. Cross-sectional (single visit) family studies** of related subjects.
3. Longitudinal* (multiple visits) studies of unrelated subjects.
4. Longitudinal* (multiple visits) family studies** of related subjects.
5. Case-control studies ***

PLEASE NOTE:

* For longitudinal studies, **please choose a single visit for each race/ethnic group that maximizes the sample size.** Once chosen, these studies can then follow the instructions as for a cross-sectional study.

** Requires adequate statistical correction for dependencies among family members while avoiding potential deflation.

*** **For case-control studies, run all analyses within case and control samples separately**.

SUBJECTS & RACE/ETHNIC GROUPS:

1. Men and women between 18 and 80 years of age with existing data.
2. Five race/ethnic/population groups are considered for analysis: European ancestry (EA), African ancestry (AA), Hispanic ancestry (HA), Asian descent (AS), and Brazilian admixed (BR).
3. Please analyze and report each race/ethnic group separately.

STRATIFICATION BY SEX:

No stratification by sex is proposed for these analyses.

PHENOTYPES AND DATA ADJUSTMENTS:

1. Phenotypes to be analysed: HDL, TG, LDL (ALL UPPER CASE LETTERS)
    a. High-density lipoprotein cholesterol (**HDL**, mg/dL)
    b. Triglycerides (**TG**, mg/dL)
    c. Low-density lipoprotein cholesterol (**LDL**, mg/dL), either directly assayed (LDL_{da}) or derived using the Friedewald equation (LDL_F). Otherwise, set LDL to missing.
        i. Note: For TG >400 mg/dL, only LDL_{da} should be used (if you only have LDL_F, set LDL values to missing in those subjects).
2. Fasting Status:
    a. If have fasting lipids (fasting ≥ 8 hours), use HDL, TG, and either LDL_{da} or LDL_F.
    b. If have non-fasting lipids only (fasting < 8 hours):
        i. Use LDL_{da} and HDL
        ii. Do not use LDL_F or TG
3. Transformations:
        i. Use natural log for TG and HDL
        ii. No transformation on LDL
4. Adjustment for statin use:
    a. Only LDL will be adjusted (not TG or HDL)
    b. Modelled after CHARGE Lipids ExomeChip Analysis Plan
    c. Assumptions: Before 1994, commonly used lipid therapy drugs (which were mostly non-statin) were relatively ineffective at lowering lipids. The benchmark “4S” study published in 1994 led to more wide-spread statin use as they were very effective. Therefore, if somebody was on an unspecified lipid lowering drug after 1994, it can be safely assumed to be statin. The following lipid adjustments reflect these assumptions (as implemented in the CHARGE Pharmacogenetics WG).
    d. Perform LDL adjustment as shown in the Table:

Lipid Lowering Drug used	Lipids measured before 1994	Lipids measured during or after 1994
Statin	**Adjust**	**Adjust**
Unspecified	No adjustment	**Adjust**
Non-statin (eg, fibrate and fibric acid derivatives, niacin, binding agents)	No adjustment	No adjustment

d.

    e. When adjustment is indicated per the table above, perform LDL adjustment as follows:
        i. If LDL was derived from Friedewald and TG <400 mg/dL,
            a) First adjust total cholesterol (TC) for statin use: adjTC = TC / 0.8
            b) Then adjust LDL using adjusted TC: $adjLDL_F$ = adjTC – HDL – (TG/5)
        ii. If LDL is directly assayed (not derived as above),
            a) Adjust LDL directly: $adjLDL_{da}$ = LDL_{da} / 0.7
    f. No adjustments for use of any other lipid lowering medications

LIFESTYLE VARIABLES: Education ("E"): "SOMECOL" AND "GRADCOL" (ALL UPPER CASE LETTERS)

As done for the education-BP project, we focus on 2 dichotomous education variables derived from existing education data. If your study has information on only one (say, SOMECOL) but not both variables, you may contribute analyses using that variable only. Studies lacking any data on education will not be able to participate.

1. **Some College (SOMECOL):**
   SOMECOL = 1 if subject attended any education beyond high school (i.e., E=1)
   SOMECOL = 0 if subject has no education beyond high school/GED (E=0)
2. **Graduated College (GRADCOL):**
   GRADCOL = 1 if subject completed at least a 4 year college degree (BA/BS) (E=1)
   GRADCOL = 0 if subject has not completed a 4 year college degree (E=0)

COVARIATES:

Include the following covariates in each analysis:

1. **AGE**, **SEX** (code male=0, female=1)
2. **Field Center (FC1 … FCn-1,** for multi-center studies); create n-1 dichotomous indicators where n= number of field centers
3. **Principal components (PC1 … PC10)** denoting stratification derived using genotyped SNPs. Please conduct some initial exploratory analyses to decide which PCs should be included in every model for each race / ethnic group in your study. We suggest the following approach: Include the first PC, and then in a stepwise manner, determine whether / which additional PCs (up to the first 10) should be included. *Note that in previous GWAS efforts for African-American subjects, the analysis plan typically calls for inclusion of all 10 PCs.* The needed PCs may differ by race / ethnic group. Once it is decided for each race / ethnic group which PCs to use to control for stratification, use those for all analyses in that race / ethnic group consistently.
4. **Additional cohort-specific covariates**, if any, to control for additional confounding.

Note: For the initial analyses, no adjustment for BMI will be made, though BMI may be added as a covariate in later analyses.

GENOTYPES:

1. Use dosage of imputed SNPs from data of the 1000 Genomes Project (1000G).
2. Imputation should be based on the ALL ancestry panel from 1000G Phase I Integrated Release Version 3 Haplotypes (2010-11 data freeze, 2012-03-14 haplotypes) that contains haplotypes of 1,092 individuals of all ethnic backgrounds. For MACH, it is ALLGIANT.phase1_release_v3.20101123.snps_indels_svs.genotypes.refpanel.ALL.vcf.gz.tgz available at http://www.sph.umich.edu/csg/abecasis/MACH/download/1000G.2012-03-14.html. For IMPUTE2, it is "ALL_1000G_phase1integrated_v3_impute_macGT1.tgz" available at http://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html

3. Use dosage of imputed SNPs based on HapMap Phase II / III reference panel if 1000G imputations are not available.
4. **SNP EXCLUSIONS**: BEFORE ANALYSIS *please exclude* the following SNPs to reduce the overall analysis burden (and file sizes):
    a. SNPs with *very low imputation quality* (r^2<0.1 if using MACH OR information metric < 0.1 if using IMPUTE2; and
    b. *SNPs with MAF < 1%* (the allele frequency of an imputed SNP can be computed as the average of dosage values for all subjects in the sample divided by 2; if this value is > 0.5, subtract it from 1 to get the MAF).
    c. Any SNPs mapping to *sex chromosomes or mitochondria*.

    Include all other SNPs. Results submitted for meta-analyses will undergo additional QC procedures centrally, such as additional filters based on minor allele counts, imputation quality measures (e.g., filtering at higher cutoffs for r^2 and information), and genomic control values. Results should not be filtered within individual cohorts except as noted above; however, if cohorts have previously filtered out some of the SNPs as part of their cohort-specific QC, include a brief description of such filters in the README file, which is described below.

CHROMOSOMES:

All autosomes (chromosomes 1-22) will be included. Sex chromosomes will be considered later as a separate project.

SUBJECT EXCLUSIONS:

1. Men and women below 18 or over 80 years of age.
2. Subjects without any GWAS data.
3. Subjects with missing data for any of the common covariates: age, sex, and PCs denoting stratification.

MODELS & ANALYSIS FOR EACH LIPID TRAIT WITH EACH EDUCATION VARIABLE:

Review of the alcohol-lipids results on M3 calls suggested that Model 2 (main effects only) resulted in novel discoveries distinct from Models 1 and 3. Accordingly, we are adding Model 2 back (only in the total sample; analysis within strata will not be considered). Thus, for each education-lipid combination, we will run 3 models (Models 1, 2, and 3) using the total sample within each race-ethnic group.

***NOTE**: To avoid confusion across the analysis plans, we continue to use "**β**" to designate the regression coefficients across all models although their interpretation varies from model to model as discussed on several calls recently. Different symbols will be used in presentations and publications to underscore the differences.*

MODEL 1 (Joint analysis of main and interaction effects):

$$Y = \beta_0 + \beta_E E + \beta_G SNP + \beta_{GE} E * SNP + \beta_C C$$

MODEL 2 (Analysis of main effect only):

$$Y = \beta_0 + \beta_G SNP + \beta_C C$$

MODEL 3 (Analysis of main effect in the presence of education):

$$Y = \beta_0 + \beta_E E + \beta_G SNP + \beta_C C$$

Where:

a) Y is the lipid phenotype value
b) β_0 is the intercept
c) E is the education variable
d) SNP is the dosage of the genetic variant, coded additively
e) C is the vector of covariates: including age, sex, study-specific confounders, PCs

What to report: For all models, provide results as shown in the "Results Format" section below (pages 7-8). All models are standard linear regression models (or linear mixed effect model for family data). For MODEL 1, please report the betas for the SNP main effect, the SNP*E interaction term, their **robust SEs** as well as the **robust covariance** between the betas. Likewise, for Models 2 and 3, report the SNP main effect, and its **robust SE** (see Tables on pages 7-8).

METHODS/SOFTWARE FOR ANALYSIS OF EACH TYPE OF STUDY:

a) **For longitudinal studies** (i.e. if more than one visit/measurement per participant is available): Use data from the one visit with the maximum information and follow the same methods as proposed for cross-sectional studies (below).
b) **For cross-sectional studies** (i.e. if only one visit/measurement per participant is available): For **unrelated studies** (with unrelated subjects), use linear regression with robust estimates of standard errors: you may use ProbABEL or MMAP. We will also accept continued use of the R/sandwich currently used by some studies. For **family studies** (with related subjects), use linear mixed model using kinship matrix to account for family relationships: we recommend using only the 1-step method of analysis, also providing robust SEs and robust covariance. At the current time, only MMAP appears to be suitable for analysis of family studies.
    i. **ProbABEL**: ProbABEL has been widely used for analysis of unrelated studies in the first three interaction projects (smoking-BP, smoking-lipids, and alcohol-lipids) and recently also used for analysis of family studies. We recommend its continued use with unrelated studies only, but not with family studies. **Code is provided in the Appendix.**
    ii. **MMAP** (http://edn.som.umaryland.edu/mmap/): MMAP (Mixed Model Analysis in Pedigrees and Populations) by Jeff O'Connell implements the 1-step method for family studies (simultaneous analysis of the association models while adjusting for family relationships) and provides robust SEs and robust covariance. **Currently, we**

recommend using only MMAP for analysis of family studies. Several studies have already used MMAP which has recently been shown to be computationally 3-4 times faster than alternatives even for unrelated studies. Therefore, MMAP can also be used for analysis of unrelated studies. Please contact Jeff O'Connell (joconnel@medicine.umaryland.edu) for the Code.

iii. Based on the experience to date and as discussed at the analysis committee meeting on January 8, 2015, use of generalized estimating equations (GEE) (e.g. geepack) for analysis of family data is strongly discouraged.

Type of study	ProbABEL	MMAP	R packages
Unrelated	Use this with the code provided (Appendix1)	Contact Jeff for Code	Sandwich
Family	*Do **not** use ProbABEL*	Contact Jeff for Code	*Do not use GEE*

N-EXPOSED (N_E):

In order to implement a t-distribution based approach developed recently by Ken Rice and colleagues, we need the value of "N-exposed" from each cohort (for computing approximate degrees of freedom). For each education variable, define **N-exposed** as the number of subjects with E=1 (i.e., the number of exposed [higher educated] individuals). For family studies, compute N-exposed as (n+N)/2 where n= the number of sibships with at least one exposed member, and N= the total number of exposed subjects across all the sibships. This is a compromise between using "n" alone (which underestimates N_E as multiple exposed subjects in a family are counted only once) or "N" alone (which overestimates N_E as this ignores sibling correlation).

Note: Typically, N for E=1 (exposed) is smaller than that for E=0 (unexposed). However, if the N for E=0 (unexposed) is smaller than that for E=1 (exposed), define N_E as the N for E=0 (unexposed). That is, N_E should be the smaller of the two N's for "exposed" and "unexposed" groups (even though it is called N-exposed). Note that this applies only for the calculation of the "N-exposed", which will be used for calculating the "approximate degrees of freedom".

RESULTS TO BE PROVIDED FOR META-ANALYSIS

For Model 1, provide results with all columns listed below**:**

Column header	Description	Recommended format	Examples
rsID	The rs-number of the variant analyzed	rs-number	rs3845291
CHR	Chromosome Number	Numeric, integer	1
POS	Position of the variant	Numeric, integer	132146
STRAND	Orientation of the site to the human genome strand used	A single character, - or +. Strong preference for + strand	+

IMPUTATION	A value (range 0-1) corresponding to the imputation quality measure (Rsq from MACH/Minimac or info from IMPUTE2)	Numeric fraction 0 to 1	0.954565
EFFECT_ALLELE	Allele for which the effect has been estimated	A single upper-case character (A, C, G, or T)	A
NON_EFFECT_ALLELE	The alternative to the effect allele	A single upper-case character (A, C, G, or T)	T
EAF	Analysis-specific allele frequency of the EFFECT_ALLELE	At least 5 digits to the right of the decimal. Scientific E notation is acceptable.	0.354125
BETA_SNP	Beta-coefficient for the association of SNP with DEPENDENT VARIABLE	At least 6 digits to the right of the decimal.	0.045228
SE_SNP	Standard error for the association of SNP with the DEPENDENT VARIABLE	At least 6 digits to the right of the decimal.	0.018343
P_SNP	P value for the association of SNP with the DEPENDENT VARIABLE	At least 6 digits and use scientific E notation	6.219424E-10
BETA_INT	Beta-coefficient for the SNPxE interaction	At least 6 digits to the right of the decimal.	0.045228
SE_INT	Robust standard error for the SNPxE interaction	At least 6 digits to the right of the decimal.	0.018343
P_INT	P value for the SNPxE interaction	At least 6 digits and use scientific E notation	6.212423E-10
COV_SNP_INT	Robust covariance between BETA_SNP and BETA_INT	At least 6 digits to the right of the decimal.	0.002343

For Models 2 and 3, provide results with all columns listed below (separately for each of the models)**:**

Column header	**Description**	**Recommended format**	**Examples**
rsID	The rs-number of the variant analyzed	rs-number	rs3845291
CHR	Chromosome Number	Numeric, integer	1
POS	Position of the variant	Numeric, integer	132146
STRAND	Orientation of the site to the human genome strand used	A single character, - or +. Strong preference for + strand	+
IMPUTATION	A value (range 0-1) corresponding to the imputation quality measure (Rsq from MACH/Minimac or info from IMPUTE2)	Numeric fraction 0 to 1	0.954565
EFFECT_ALLELE	Allele for which the effect has been estimated	A single upper-case character (A, C, G, or T)	A
NON_EFFECT_ALLELE	The alternative to the effect allele	A single upper-case character (A, C, G, or T)	T
EAF	Analysis-specific allele frequency of the EFFECT_ALLELE	At least 5 digits to the right of the decimal. Scientific E notation is acceptable.	0.354125
BETA_SNP	Beta-coefficient for the association of SNP with DEPENDENT VARIABLE	At least 6 digits to the right of the decimal.	0.045228
SE_SNP	Standard error for the association of SNP with the DEPENDENT VARIABLE	At least 6 digits to the right of the decimal.	0.018343
P_SNP	P value for the association of SNP with the DEPENDENT VARIABLE	At least 6 digits and use scientific E notation	6.219424E-10

README FILE: EXCEL TEMPLATE DISTRIBUTED WITH ANALYSIS PLAN

When uploading the result files to the CHARGE Google Drive (instructions will be provided in the future), please fill in, rename, and upload the accompanying Excel file "**STUDY.RACE.LIPIDS.EDUC.README.DATE.xls**". This excel file asks for information about the lipid phenotypes and the education variables used in the analyses. In particular, the following information is required:

1. Contacts. List the contact information for the study: name of the Principal Investigator (PI) and the Contact Analyst, and their email and telephone number.

2. Study characteristics. Provide information about the characteristics of your study and genotype data, such as whether the study is unrelated (UN), family-based (FB), or case-control (CC), and how many (if any) principal components (PCs) were used in the models.

3. N-exposed. Report the $N_{exposed}$ for each analysis. Note that this corresponds to **the smaller of the two sample sizes for the exposed (E=1) and the unexposed (E=0) groups** for each race-phenotype-exposure combination. Please use the estimation formula described on page 6 if the study is family-based (again referring to the group with the lower N).

4. Descriptive Statistics. Summary statistics for HDL, TG and LDL in the total sample and within the stratified groups for each education variable.

Please report the required information in each sheet for each race/ethnic group separately, as indicated. Please remember to rename the file. Until instructions are provided for uploading the files to the CHARGE Google Drive, please store all results files and the excel files locally.

META-ANALYSIS:

1. Meta analyses will be conducted separately by race; we will consider pooling across races using meta-analysis and/or trans-ethnic meta-analysis using MANTRA.
2. For MODEL 1, we will perform joint fixed-effects meta-analysis of the SNP main effect term and the SNP*E interaction term while considering the covariance between them in METAL following Manning et al. [Genetic Epidemiology 2011].
3. For MODEL 2 and 3, we will perform fixed-effects meta-analysis of the SNP main effect term in METAL.

CONTACT INFORMATION:

Lead Investigators: Karen Schwander (karen2@wubios.wustl.edu); Yun Ju Sung (yunju@wubios.wustl.edu); and DC Rao (rao@wubios.wustl.edu)

Code for running 1-step MMAP is available from Jeff O'Connell (joconnel@medicine.umaryland.edu).

Appendix: Example Code for ProbABEL
To be used with unrelated cohorts only

Date: 09/30/15

This example code was originally developed for the Education-BP analyses and has been modified for the Education-Lipid analysis. The code can be used as a guide for your analyses. Please direct any questions to Karen Schwander (karen2@wubios.wustl.edu) or Yun Ju Sung (yunju@wubios.wustl.edu).

This appendix includes 4 steps. They are:

1. Prepare genotype (dose and info) input files for ProbABEL – If you have used R (such as sandwich and GEE) for other projects, then your genotype data must have been already filtered and saved as RData objects. This step converts these filtered imputed GWAS data in RData objects to two text files (dose and info files). Note that if these genotype files (one per chromosome) were already created for previous projects, then this step can be skipped.
2. Prepare phenotype file for ProbABEL – This step creates an input file for ProbABEL that includes the trait and covariates (e.g., SOMECOL, AGE, SEX).
3. Run ProbABEL – It describes how to run Models 1, 2, and 3 in ProbABEL.
4. Post-process ProbABEL result files – This step is recommended for all studies that are using ProbABEL. It uses ProbABEL output, selects/renames required variables, and calculates p-values.

The table below shows an overview for those who have already run our previous ProbABEL code for any previous projects.

Model	1. prepare genotype file	2. prepare phenotype file	3. run ProbABEL	4. post-processing*
1	Same as previous projects	One per each educ-lipid combination	Use interaction option	Need to run using the updated function
2	Same as previous projects	Exclude educ column from model 2 phenotype file	Use default option	Need to run using the updated function
3	Same as previous projects	Same as model 1 phenotype file	Use default option	Need to run using the updated function

*Note that prep.upload() function is modified to include position for output file

How to Use the Code Below: All R codes below are in courier new font. All functions created and used below are available in a separate text file called *functions.R*. Any code commented in #red below is part of the function. It is provided here for your reference, but should not need to be edited. Any code commented in #green inside a box is an example of using a function; this is the code that you will need

to copy and edit for your analysis. To use our provided R function, type source("functions.R"), and then use the function, as shown below.

Step 1: Prepare genotype input files

If you have used GEE or R/sandwich for previous analyses, then your genotype data must have been already filtered and are available as RData. If so, use this code to create dose and info files for use in ProbABEL.

```
# This function turns filtered RData objects into info and dose files
prep.geno=function(RDatafile,infofile,dosefile,change.id=F) {
# set change.id=T if your imputations were done with MACH to fix the id

# Load in previously filtered RData file
  load(RDatafile)

# Get 7 columns from info and write to info file
  info=info[,1:7]
  write.table(info,file=infofile, quote=F, row.names=F)

# Add dose column, fix ID if necessary, and write to dose file
 if (change.id)
   rownames(geno)= sapply(strsplit(rownames(geno), "->"),function(x) x[2])
  write.table(cbind("DOSE",geno),file=dosefile,quote=F,row.names=T,col.names=F)
}
```

```
# Example using the prep.geno function
source("functions.R")  # read the provided R code with functions
prep.geno("ch22-filtered.RData", "ch22-filtered.info", "ch22 filtered.dose",
change.id=T)  # note here we used change.id=T for our MACH imputed data
```

Step 2: Prepare phenotype input files

This step create phenotype input file that includes the pedigree-adjusted trait residuals and needed covariates (e.g., SOMECOL, AGE, SEX).

Note: In order for these R scripts to work, the subject ID must be unique in the dataset, not just within families. For example, there can only be one subject with ID=1. If you need help dealing with this issue, please contact Karen Schwander at karen2@wubios.wustl.edu.

The phenotype file must have at least two columns (first for id, second for the trait, in that order), then the remaining columns are considered as covariates (see Step 3 for examples of input files). Please note that for ProbABEL, the phenotype and genotype data have to be in the same order of subjects (one row per subject) and therefore should have the same number of rows. If a subject has partly missing phenotype data, they should still be included in the phenotype file; ProbABEL will automatically exclude them from the analysis as appropriate. In addition, the order of the SNPs in the info files should be the same as the order of the SNPs in the dose files. The function prep.pheno will take care of this.

```
# This function creates phenotype input file that includes trait,
# the environmental covariate and other covariates to the file.
```

```
prep.pheno = function(phen.outfile, phen, trait, env.cov, add.cov, geno.id){

### minimal checks for pheno file
  if (!any(colnames(phen)=='ID'))
    stop('pheno file does not have the column for ID')
  if (!any(colnames(phen)==trait))
    stop(paste('pheno file does not have the column for',trait))
  if (!any(colnames(phen)==env.cov))
    stop(paste('pheno file does not have the column for',env.cov))

  col.out=c('ID',trait,env.cov,add.cov)
  index=match(col.out,colnames(phen))
  if (any(is.na(index)))
    stop('pheno file dose not have the column for additional covariates')

  out=phen[,index]

### reorder rows of out to match with geno.id
### geno.id is a text file containing a list of IDs (with no header), that are ###
in the dosage files, in the same order that they appear in the dosage files.
### This code reorders the phenotype data, if needed, to match the dosage files.
  index=match(geno.id,phen$ID)
  out=out[index,]
  out$ID=geno.id
  colnames(out)[1]='id'

  write.table(out, file=phen.outfile, row.names=F, quote=F)
  print(paste('created',phen.outfile))
}
```

```
# Example code to use prep.pheno
phen = read.csv("pheno-white.csv", head=T)
geno.id = scan('geno.id')

# Prepare pheno file for ProbABEL for Model 1 (also Model 3)
add.cov = c('AGE','SEX','FC1','FC2')
prep.pheno('pheno-HDL-SOMECOL-M1.txt', phen, 'HDL', 'SOMECOL', add.cov,
  geno.id)

# Prepare pheno file for ProbABEL for Model 2
phen = read.table("pheno-HDL-SOMECOL-M1.txt", head=T)
phen = phen[,-3] # exclude the 3rd column: SOMECOL
write.table(phen, file='pheno-HDL-SOMECOL-M2.txt', row.names=F, quote=F)
```

Step 3: Run ProbABEL for Models 1, 2, and 3

Here are examples of what your input files for ProbABEL should look like:

Input files:

pheno-HDL-SOMECOL-M1.txt:

```
id HDL SOMECOL AGE SEX
1 145.1 1 68 2
3 150.2 0 42 1
4 132.4 1 40 2
5 141.0 1 21 2
```

pheno-HDL-SOMECOL-M2.txt:

```
id HDL AGE SEX
1 145.1 68 2
3 150.2 42 1
4 132.4 40 2
5 141.0 21 2
```

ch22.info:

```
SNP Al1 Al2 Freq1 MAF AvgCall Rsq
rs131526 G A 0.92785 0.07215 0.93295 0.23028
rs131527 C T 0.9301 0.0699 0.93304 0.22309
rs131531 A G 0.92464 0.07536 0.93179 0.2331
rs131538 G A 0.92667 0.07333 0.93431 0.24482
```

ch22.dose:

```
1 DOSE      1.999 1.989 1.999 1.999 1.478
3 DOSE      1.999 1.986 1.999 1.999 1.731
4 DOSE      1.997 1.983 1.999 1.996 1.593
5 DOSE      1.999 1.990 1.999 1.999 1.718
```

The phenotype file must have at least two columns (first for id, second for the trait, in that order), then the remaining columns are considered as covariates (see above examples of input files). Therefore, if the phenotype file includes a lifestyle covariate, such as SOMECOL column, it will run Model 3 by default (no interactions). See the ProbABEL_manual.pdf (available from http://www.genabel.org/sites/default/files/pdfs/ProbABEL_manual.pdf) for more details.

Below is the code that shows how to run ProbABEL for Models 1, 2, and 3, while getting robust standard errors. To calculate robust standard errors in ProbABEL, use the --robust option.

Commands to run ProbABEL for Models 1, 2, and 3:

```
# Example code for running Model 1 in ProbABEL
palinear --pheno pheno-HDL-SOMECOL-M1.txt --info ch22.info --dose ch22.dose --
robust --out --interaction=1 out-ch22-HDL-SOMECOL-M1.txt

# Example code for running Model 2 in ProbABEL
palinear --pheno pheno-HDL-SOMECOL-M2.txt --info ch22.info --dose ch22.dose --
robust --out out-ch22-HDL-SOMECOL-M2.txt

# Example code for running Model 3 in ProbABEL
palinear --pheno pheno-HDL-SOMECOL-M1.txt --info ch22.info --dose ch22.dose --
robust --out out-ch22-HDL-SOMECOL-M3.txt
```

In the above code for Model 1, --interaction=1 tells ProbABEL to use the 1st listed covariate as the interaction covariate, because in the phenotype file (see above input file for example), SOMECOL is the 1st listed covariate, i.e., the 1st column after the subject ID and trait. It is crucial that the column containing your lifestyle covariate in your phenotype file is correctly identified in the interaction statement above. For example, if the covariates were in the order AGE SEX SOMECOL, then you would use --interaction=3. An example output file for Model 1 is shown below; the bold-faced column names correspond to the beta estimates and their robust standard errors. For Models 2 and 3, it is similar, except there are no interaction or covariance terms.

Output file: out-ch22_add.out

```
name A1 A2 Freq1 MAF Quality Rsq n Mean_predictor_allele beta_SNP_add
sebeta_SNP_add beta_SNP_SOMECOL sebeta_SNP_SOMECOL cov_SNP_int_SNP_SOMECOL
loglik
rs149201999 T C 0.93864 0.06136 0.93864 2045 0.0129584 3.65995 4.12266 -
11.6668 4.73242 -16.999 -7230.14
rs146752890 C G 0.91575 0.08425 0.91575 2045 0.0867971 1.27578 1.57563 -
4.95806 2.14644 -2.47974 -7229.94
rs139377059 C T 0.94826 0.05174 0.94826 2045 0.452078 -0.561018 0.953585
0.491564 1.29279 -0.907429 -7232.15
```

Step 4: Post-process ProbABEL result files

Because the ProbABEL output doesn't include the p-values, one can use the following function in R to obtain them. This code also renames and reorders the variables to match the analysis plan. Note prep.upload assumes that you have one ProbABEL result file for each chromosome. If you have chopped chromosomes, then you need to edit this function. *NOTE: This code now selects the appropriate variable for EAF, so no additional work is required to get the analysis-specific EAFs.*

```
# This function reads ProbABEL output, renames variables, and computes p-values
p.probabel = function(probabel.outfile,include.int=TRUE) {

# read ProbABEL output file
  out=read.table(probabel.outfile,head=T)

# match column headers to those in the table on page 7 of the analysis plan
  new.name=c("rsID","EFFECT_ALLELE","NON_EFFECT_ALLELE","EAF","IMPUTATION",
    "BETA_SNP","SE_SNP")
  old.name=c("name","A1","A2","Mean_predictor_allele","Rsq","beta_SNP_add",
    "sebeta_SNP_add")
  index=match(old.name,colnames(out))
  new.out=out[,index]
  colnames(new.out)=new.name
  out=out[,-index]
  if (include.int) {
    index=grep("SNP",names(out))
    new.names=c("BETA_INT","SE_INT","COVAR_SNP_INT")
    out=out[,index]
    names(out)=new.names
    out=cbind(new.out,out)
  } else out=new.out

# Compute p-value and create output
  out$P_SNP=pchisq((out$BETA_SNP/out$SE_SNP)^2,df=1,lower.tail=F)
  if (include.int)
    out$P_INT=pchisq((out$BETA_INT/out$SE_INT)^2,df=1,lower.tail=F)

  return(out)
}
### This function will create the output file ready to upload
## This function is modified to include position column in the output file
prep.upload =
function(uploadfile,probabel.files,position.files,strand='+',include.int=T) {
  chr=1:22
```

```
  if (length(probabel.files)!=22)
    stop('probabel files should be 22, one for each chromosome')
  if (length(position.files)!=22)
    stop('position files should be 22, one for each chromosome')

  for (i in chr) {
    out=p.probabel(probabel.files[i],include.int=include.int)
    out$CHR=i
    out$STRAND=strand
    if (setdiff(colnames(position.files[i]), c('name','position')))
      stop(sprintf('position file for chr %d does not have name and position
column',i))

    out = merge(out, position.files[i], by='name')
    if (i==1) write.table(out,file=uploadfile,quote=F,row.names=F,na='.')
    else write.table(out,file=uploadfile,quote=F,row.names=F,
                     col.names=F,append=T,na='.')
    print(paste('processed',probabel.files[i]))
  }
  system(paste('gzip',uploadfile))
  print(paste(uploadfile,'is ready to upload'))
}
```

```
# Example using the prep.upload function

# Separately, you need to prepare 22 position files.
# Each file (position-ch1.txt,…) should include only two columns: name and position
position.files=sprint("positions-ch%d.txt",1:22)

# Call this function for ProbABEL Model 1 output with interaction effects
probabel.files=sprintf("EA-HDL-SOMECOL-M1-ch%d_add.out.txt",1:22)
prep.upload(uploadfile="EA.HDL.SOMECOL.M1.txt",probabel.files,position.files)

# Call this function for ProbABEL Model 2 output without interaction effects
probabel.files=sprintf("EA-HDL-SOMECOL-M2-ch%d_add.out.txt",1:22)
prep.upload(uploadfile="EA.HDL.SOMECOL.M2.txt",probabel.files,position.files,includ
e.int=FALSE)

# Call this function for ProbABEL Model 3 output without interaction effects
probabel.files=sprintf("EA-HDL-SOMECOL-M3-ch%d_add.out.txt",1:22)
prep.upload(uploadfile="EA.HDL.SOMECOL.M3.txt",probabel.files,position.files,includ
e.int=FALSE)
```

Supplement: Quality Control Using EasyQC[1]

Each discovery cohort completed a preliminary filter on their 1000G imputed data files, before undertaking any analysis. They removed variants with minor allele frequency (MAF) < 1% or imputation quality measure (Rsq) < 0.1. Depending on the ancestry group, this reduced the number of variants analyzed from over 30 million to approximately 8-15 million high quality variants.

After completing analysis using this reduced set of variants, all discovery cohorts submitted result files containing effect sizes and SEs (as well as a covariance term in model 1), chromosome and position, strand, effect and non-effect alleles, effect allele frequency (EAF), and Rsq.

For each project, the number of result files sent per-cohort was the number of phenotypes x number of covariates x number of models run. For example, in the Education-Lipids project, each cohort sent 18 files (3 Lipids x 2 Education covariates x 3 models).

In addition to result files, the discovery cohorts sent "readme" files containing other important information, such as Ns (total, unexposed, exposed) and summary statistics of the phenotypes. Further details were provided about the genotyping platform, imputation software used, analysis software, and which additional study-specific covariates (such as PCs), were used in the association analysis.

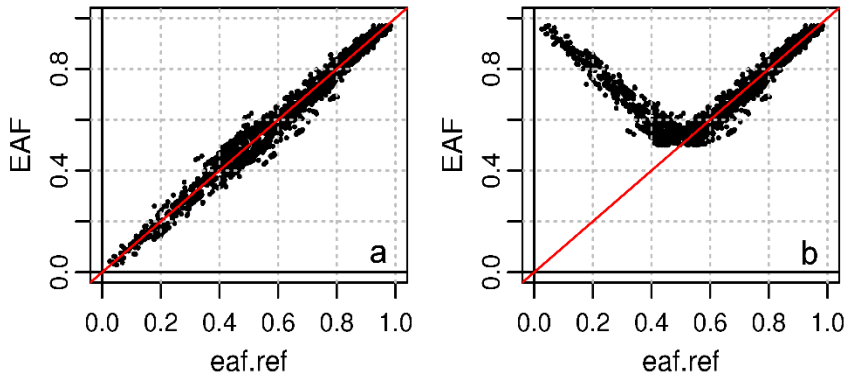
Two "levels" of quality control (QC) were performed consecutively: "Study-level" QC involved multiple steps to review result files from each study individually and "Meta-level" QC involved reviewing result files of a specific analysis (e.g., HDL-Some College-Model1) across all available discovery cohorts. All QC was performed using the EasyQC R package that provides functions to perform a wide variety of QC checks for genome-wide association studies (www.genepi-regensburg.de/easyQC).[1]

During Study-level QC, the result files were first checked to see if cohorts met the preliminary filter requirements of excluding low quality variants with MAF<1% or Rsq< 0.1, or of excluding sex chromosomal variants (as specified in the analysis plan). If variants were included that did not pass these filters, they were removed at this point. We also checked for and removed any variants with missing or invalid values as well as duplicates.

In order to prepare for meta-analysis, we harmonized study results with regards to column names, file formats, allele codes and variant names. We applied two novel EasyQC functions 'HARMONIZEALLELES' and 'CREATECPTID' that automatically reformat varying versions of allele codes and marker names given across studies. The functions compile unique allele codes (A/C/G/T for SNPs, and I/D for insertions and deletions, INDELs) and variant names (CPTIDs that follow the format of CHR:POSITION:TYPE, where TYPE id 'ID' FOR INDELs). For example, for INDELs, I/D allele codes were derived from the MACH reference formats of R/D and R/I as well as from IMPUTE reference

format that involve the actual sequence (e.g., C/CTGT for the INDEL at base position 46402 on chromosome 1). Harmonizing both the variant names and alleles in this way ensures consistency across cohorts during the meta-analysis stage.

Next, we compared the allele frequencies provided in the result files against the ancestry-specific 1000G reference panel [2]; Supplementary Figure 1 shows a normal cohort (a), and a problematic cohort (b). Any major differences from the ancestry-specific reference panel were discussed with the relevant cohorts, and corrections were made as needed; in the example below, the cohort had provided "major" allele frequency instead of EAF; they corrected and resent the results. Finally, strand information, along with effect and non-effect alleles, were used to adjust allele directions according to reference allele directions, if necessary. A small number of variants may be removed in this step, where the appropriate adjustment was not clear (e.g., A/T in discovery cohort, A/G in reference).



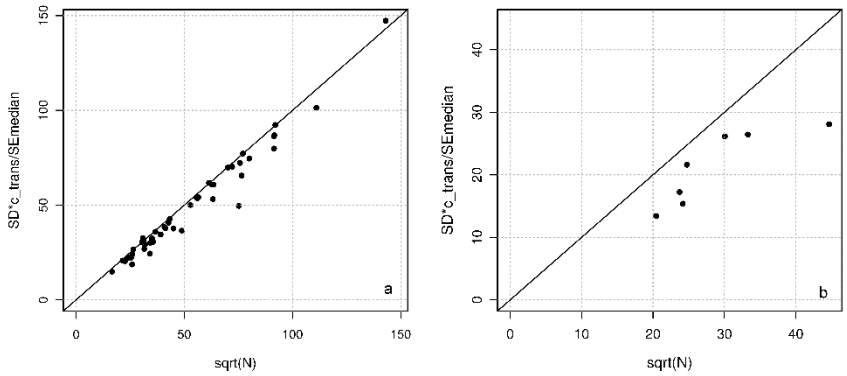
Supplementary Figure 1: Comparison of Cohort EAFs with 1000G Reference Panel

Once all the discovery cohort files were harmonized and QC'd, a "cleaned" version of each result file was created. At this point, the "Meta-level" QC was performed, by comparing result files of a specific analysis across all contributing cohorts.

This first involved calculating the 2df joint test [3] from the 1df SNP and interaction terms, and then visually comparing summary statistics (mean, median, standard deviation, inter quartile range, minimum, maximum) on all effect estimates, standard errors (SEs), and p-values, to check for consistency across cohorts, or errors that occurred in the original analyses. Additionally, we checked to see the number of variants with unusually large effect sizes.

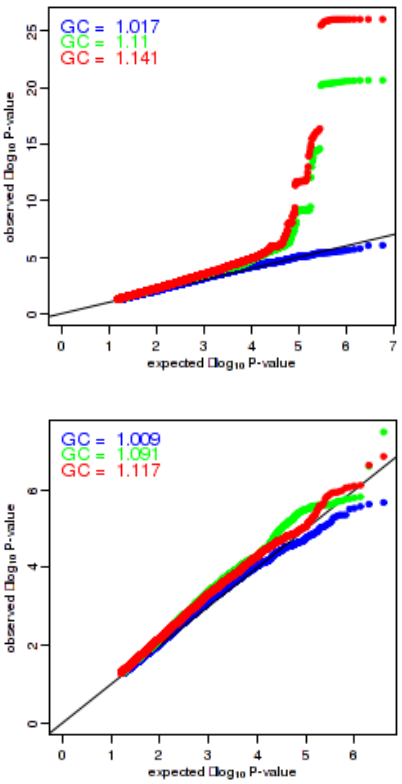
Next, so-called SEN plots were created with EasyQC that can be used to check for issues with trait transformation or other analytical problems. The SEN plot displays the square root of the total sample size (N), against c*SD/medianSE, where c=median(1/sqrt(2*MAF*(1-MAF))), and SD is the standard deviation of the phenotype. Supplementary Figure 2 shows an example of a normal SEN plot (a), and a problematic SEN plot (b); we expect to see each cohort along the diagonal, since the variance of the

beta estimate of a specified SNP depends on the variance of the phenotype, the variance of the genotype, and the sample size, i.e., X and Y axis should increase proportionally.



Supplementary Figure 2: Example of SEN Plots.

Using both the review of the summary statistics and the SEN plot, we were able to pinpoint a number of problems that were then corrected, including (for projects involving blood pressure traits): incorrect calculation of the MAP phenotype; switched MAP/PP result files; 'extra' covariates in the model that were not expected; or incorrectly reporting the sample size. Any problems found during this step were resolved with the individual cohorts.



The last step in the meta-QC process was to review QQ plots. During this step, we looked at the 'raw' QQ plots, but also explored various filters to use in meta-analysis to remove unstable and unreliable results. To this end, we created derived variables MAC0 (minor allele count in the unexposed group), MAC1 (minor allele count in the exposed group), and approximate degrees of freedom (to approximate DF from a t-distribution).[4] QQ plots were created using both the 'raw' result files as well as various filters involving MAC0/MAC1/DF and Rsq, to see which filter did the most efficient job of reducing noise while minimizing the loss of variants for meta-analysis.

Each gene-lifestyle analysis project decides the most appropriate filter to use for meta-analysis; the majority of projects have chosen to use the DF<20 filter. Supplementary Figure 3 shows an example of a raw cohort result file (a), and the same cohort filtered at DF<20 (b).

***Supplementary Figure 3**: Example of QQ Plots on Unfiltered and DF20 Filtered Variants.*

Gene-Lifestyle Interactions WG: STAGE 1 STUDY DESCRIPTIONS:

Brief descriptions are provided below for each of the discovery studies some of which are based outside the United States:

AGES (Age Gene/Environment Susceptibility Reykjavik Study): The AGES Reykjavik study originally comprised a random sample of 30,795 men and women born in 1907-1935 and living in Reykjavik in 1967. A total of 19,381 people attended, resulting in a 71% recruitment rate. The study sample was divided into six groups by birth year and birth date within month. One group was designated for longitudinal follow up and was examined in all stages; another was designated as a control group and was not included in examinations until 1991. Other groups were invited to participate in specific stages of the study. Between 2002 and 2006, the AGESReykjavik study re-examined 5,764 survivors of the original cohort who had participated before in the Reykjavik Study. The midlife data blood pressure measurement was taken from stage 3 of the Reykjavik Study (1974-1979), if available. Half of the cohort attended during this period. Otherwise an observation was selected closest in time to the stage 3 visit. The supine blood pressure was measured twice by a nurse using a mercury sphygmomanometer after 5 minutes rest following World Health Organization recommendations.

ARIC (Atherosclerosis Risk in Communities): The ARIC study is a population-based prospective cohort study of cardiovascular disease sponsored by the National Heart, Lung, and Blood Institute (NHLBI). ARIC included 15,792 individuals, predominantly European American and African American, aged 45-64 years at baseline (1987-89), chosen by probability sampling from four US communities. Cohort members completed three additional triennial follow-up examinations and a fifth exam in 2011-2013. The ARIC study has been described in detail previously.[5] Blood pressure was measured using a standardized Hawksley random-zero mercury column sphygmomanometer with participants in a sitting position after a resting period of 5 minutes. The size of the cuff was chosen according to the arm circumference. Three sequential recordings for systolic and diastolic blood pressure were obtained; the mean of the last two measurements was used in this analysis, discarding the first reading. Blood pressure lowering medication use was recorded from the medication history.

Baependi Heart Study (Brazil): The Baependi Heart Study, is an ongoing family-based cohort conducted in a rural town of the state of Minas Gerais. The study has enrolled approximate 2,200 individuals (over 10% of the town's adult population) and 10-year follow up period of longitudinal data. Briefly, probands were selected at random across 11 out of the 12 census districts in Baependi. After enrolment, the proband's first-degree (parents, siblings, and offspring), second-degree (half-siblings, grandparents/grandchildren, uncles/aunts, nephews/nieces, and double cousins), and third-degree (first cousins, great uncles/aunts, and great nephews/nieces) relatives, and his/her respective spouse's relatives resident both within Baependi (municipal and rural area) and surrounding towns were invited to participate. Only individuals age 18 and older were eligible to participate in the study. The study is conducted from a clinic/office in an easily accessible sector of the town, where the questionnaires were completed. A broad range of phenotypes ranging from cardiovascular, neurocognitive, psychiatric, imaging, physiologic and several layers of endophenotypes like metabolomics and lipidomics have been collected throughout the years Details about follow-up visits and available data can be found in the cohort profile paper.[6] DNA samples were genotyped using the Affymetrix 6.0 genechip. After quality control, the data were prephased using SHAPEIT and imputed using IMPUTE2 based on 1000 Genomes haplotypes.

Bio*Me* Biobank (Bio*Me* Biobank of Institute for Personalized Medicine at Mount Sinai): The Bio*Me* Biobank, founded in September 2007, is an ongoing, consented electronic medical record (EMR)-linked bio- and data repository that enrolls participants non-selectively from the Mount Sinai Medical Center patient population. The Bio*Me* Biobank currently (Winter 2015) comprises over 31,000 participants from diverse ancestries characterized by a broad spectrum of (longitudinal) biomedical traits. On average 400

new participants are consented each month. Bio*Me* participants represent the broad ancestral, ethnic and socioeconomic diversity with a distinct and population-specific disease burden, characteristic of Northern Manhattan communities served by Mount Sinai Hospital. Enrolled participants consent to be followed throughout their clinical care (past, present, and future) at Mount Sinai in real-time, integrating their genomic information with their electronic health record for discovery research and clinical care implementation. Bio*Me* participants are predominantly of African, Hispanic/Latino, and European ancestry. Participants who self-identify as Hispanic/Latino further report to be of Puerto Rican (39%), Dominican (23%), Central/South American (17%), Mexican (5%) or other Hispanic (16%) ancestry. More than 40% of European ancestry participants are genetically determined to be of Ashkenazi Jewish ancestry.

The IRB-approved Bio*Me* Biobank consent permits use of samples and de-identified linkable past, present and future clinical information from EMRs; re-contacting participants for enrollment in future research; unlimited duration of storage, and access to clinical information from the entire medical records, as well as local and external sharing of specimens and data.

The Bio*Me* Biobank has a longitudinal design as participants consent to make any EMR data from past (dating back as far as 2003), present and future inpatient or outpatient encounters available for research. The median number of clinical encounters per participant is 21, reflecting predominant enrollment of participants with common chronic conditions from primary care facilities. Mount Sinai's system-wide Epic EMR implementation captures a full spectrum of biomedical phenotypes, including clinical outcomes, covariate and exposure data. This clinical information is complemented by detailed information on ancestry, residence history, familial medical history, education, socio-economic status, physical activity, smoking, alcohol use, and weight history being collected in a systematic manner by interview-based questionnaire at time of enrollment. Phenotype harmonization and validation is critical to facilitate consortium-wide analyses. By applying advanced medical informatics and data mining tools, high-quality and validated phenotype data can be culled from Mount Sinai's Epic EMR. Fully-implemented phenotype algorithms include; T2D, CKD, CAD, lipid disorders, peripheral artery disease, resistant hypertension, blood cell traits, abdominal aortic aneurism, venous thromboembolism among others (see also Phenotype KnowledgeBase (PheKB) of the eMERGE Network (http://emerge.mc.vanderbilt.edu/emerge-network).

A total of 14,017 participants have been genotyped for both GWAS (11,150 Illumina OmniExpress BeadChip, 2,867 Affymetrix Human SNP Array 6.0) *and* ExomeChip (Illumina HumanExome v1.0 BeadChip) arrays funded by institutional sources. An additional 16,000 Bio*Me* participants are scheduled for genotyping using the Illumina MEGA Chip (by April 2015), funded by NHGRI through our PAGEII grant (U01HG007417) (n=12,500) and through institutional funds (n=3,500).

CARDIA (Coronary Artery Risk Development in Young Adults): CARDIA is a prospective multicenter study with 5,115 adults Caucasian and African American participants of the age group 18-30 years, recruited from four centers at the baseline examination in 1985-1986. The recruitment was done from the total community in Birmingham, AL, from selected census tracts in Chicago, IL and Minneapolis, MN; and from the Kaiser Permanente health plan membership in Oakland, CA. The details of the study design for the CARDIA study have been previously published. Eight examinations have been completed since initiation of the study, respectively in the years 0, 2, 5, 7, 10, 15, 20 and 25. Written informed consent was obtained from participants at each examination and all study protocols were approved by the institutional review boards of the participating institutions. Systolic and diastolic blood pressure was measured in triplicate on the right arm using a random-zero sphygmomanometer with the participant seated and following a 5-min. rest. The average of the second and third measurements was taken as the blood pressure value. Blood pressure medication use was obtained by questionnaire.

CHS (Cardiovascular Health Study): CHS is a population-based cohort study of risk factors for cardiovascular disease in adults 65 years of age or older conducted across four field centers.[7] The original predominantly European ancestry cohort of 5,201 persons was recruited in 1989-1990 from random samples of the Medicare eligibility lists and an additional predominately African-American cohort of 687 persons was enrolled in 1992-93 for a total sample of 5,888. Blood samples were drawn from all participants at their baseline examination and DNA was subsequently extracted from available samples. European ancestry participants were excluded from the GWAS study sample due to prevalent coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke, or transient ischemic attack at baseline. After QC, genotyping was successful for 3271 European ancestry and 823 African-American participants. CHS was approved by institutional review committees at each site and individuals in the present analysis gave informed consent including consent to use of genetic information for the study of cardiovascular disease.

Participants with missing BMI (N=10) or BP measurements (N=8) were also excluded. Research staff with central training in blood pressure measurement assessed repeated right-arm seated systolic and diastolic blood pressure levels at baseline with a Hawksley random-zero sphygmomanometer.

ERF (Erasmus Rucphen Family study): Erasmus Rucphen Family is a family based study that includes inhabitants of a genetically isolated community in the South-West of the Netherlands, studied as part of the Genetic Research in Isolated Population (GRIP) program. The goal of the study is to identify the risk factors in the development of complex disorders. Study population includes approximately 3,000 individuals who are living descendants of 22 couples who lived in the isolate between 1850 and 1900 and had at least six children baptized in the community church. All data were collected between 2002 and 2005. All participants gave informed consent, and the Medical Ethics Committee of the Erasmus University Medical Centre approved the study.

Fam HS (Family Heart Study): The NHLBI FamHS study design, collection of phenotypes and covariates as well as clinical examination have been previously described (https://dsgweb.wustl.edu/fhscc/).[8] In brief, the FamHS recruited 1,200 families (approximately 6,000 individuals), half randomly sampled, and half selected because of an excess of coronary heart disease (CHD) or risk factor abnormalities as compared with age- and sex-specific population rates. The participants were sampled from four population-based parent studies: the Framingham Heart Study, the Utah Family Tree Study, and two centers for the Atherosclerosis Risk in Communities study (ARIC: Minneapolis, and Forsyth County, NC). These individuals attended a clinic exam (1994-1996) and a broad range of phenotypes were assessed in the general domains of CHD, atherosclerosis, cardiac and vascular function, inflammation and hemostasis, lipids and lipoproteins, blood pressure, diabetes and insulin resistance, pulmonary function, diet, education, socioeconomic status, habitual behavior, physical activity, anthropometry, medical history and medication use. Approximately 8 years later, study participants belonging to the largest pedigrees were invited for a second clinical exam (2002-04). The most important CHD risk factors were measured again, including lipids, parameters of glucose metabolism, blood pressure, anthropometry, and several biochemical and hematologic markers. In addition, a computed tomography examination provided measures of coronary and aortic calcification, and abdominal and liver fat burden. Medical history and medication use was updated. A total of 2,756 European ancestry subjects in 510 extended random and high CHD risk families were studied. Also, 633 African ancestry subjects were recruited at ARIC field center at the University of Alabama in Birmingham. Informed consent was obtained from all participants.

FHS (Framingham Heart Study): FHS began in 1948 with the recruitment of an original cohort of 5,209 men and women (mean age 44 years; 55 percent women). In 1971 a second generation of study participants was enrolled; this cohort (mean age 37 years; 52% women) consisted of 5,124 children and spouses of children of the original cohort. A third generation cohort of 4,095 children of offspring cohort participants (mean age 40 years; 53 percent women) was enrolled in 2002-2005 and are seen every 4 to

8 years. Details of study designs for the three cohorts are summarized elsewhere. At each clinic visit, a medical history was obtained with a focus on cardiovascular content, and participants underwent a physical examination including measurement of height and weight from which BMI was calculated. Systolic and diastolic blood pressures were measured twice by a physician on the left arm of the resting and seated participant using a mercury column sphygmomanometer. Blood pressures were recorded to the nearest even number. The means of two separate systolic and diastolic blood pressure readings at each clinic examination were used for statistical analyses.

GENOA (Genetic Epidemiology Network of Arteriopathy): GENOA is one of four networks in the NHLBI Family-Blood Pressure Program (FBPP).[9-10] GENOA's long-term objective is to elucidate the genetics of target organ complications of hypertension, including both atherosclerotic and arteriolosclerotic complications involving the heart, brain, kidneys, and peripheral arteries. The longitudinal GENOA Study recruited European-American and African-American sibships with at least 2 individuals with clinically diagnosed essential hypertension before age 60 years. All other members of the sibship were invited to participate regardless of their hypertension status. Participants were diagnosed with hypertension if they had either 1) a previous clinical diagnosis of hypertension by a physician with current anti-hypertensive treatment, or 2) an average systolic blood pressure ≥ 140 mm Hg or diastolic blood pressure ≥ 90 mm Hg based on the second and third readings at the time of their clinic visit. Exclusion criteria were secondary hypertension, alcoholism or drug abuse, pregnancy, insulin-dependent diabetes mellitus, or active malignancy. During the first exam (1995-2000), 1,583 European Americans from Rochester, MN and 1,854 African Americans from Jackson, MS were examined. Between 2000 and 2005, 1,241 of the European Americans and 1,482 of the African Americans returned for a second examination. Because African-American probands for GENOA were recruited through the Atherosclerosis Risk in Communities (ARIC) Jackson field center participants, we excluded ARIC participants from analyses.

GenSalt (Genetic Epidemiology Network of Salt Sensitivity): GenSalt is a multi-center, family based study designed to identify, through dietary sodium and potassium intervention, salt-sensitivitivity susceptibility genes which may underlie essential hypertension in rural Han Chinese families. Approximately 629 families with at least one ‘proband’ with high blood pressure were recruited and tested for a wide variety of physiological, metabolic and biochemical measures at baseline and at multiple times during the 3-week intervention. The intervention consisted of one week on a low sodium diet, followed by one week on a high sodium diet, and finally one week on a high sodium diet with a potassium supplement.

GOLDN (Genetics of Diet and Lipid Lowering Network): GOLDN is a multi-center family pharmacogenetic study that is investigating gene- environment interactions on lipid profiles. 1,200 subjects in extended pedigrees were measured before and after two environmental exposures: 1) a dietary fat challenge to assess genetic regulators of fat uptake and clearance and 2) a 3 week clinical trial of fenofibrate to assess pharmacogenetic influences on response to treatment. The goals of the study are to identify and characterize genetic loci that predict the lipid profile treatment responses. https://dsgweb.wustl.edu/PROJECTS/MP5.html

HANDLS (Healthy Aging in Neighborhoods of Diversity across the Life Span): HANDLS is a community-based, longitudinal epidemiologic study examining the influences of race and socioeconomic status (SES) on the development of age-related health disparities among a sample of socioeconomically diverse African Americans and whites. This unique study will assess over a 20-year period physical parameters and also evaluate genetic, biologic, demographic, and psychosocial, parameters of African American and white participants in higher and lower SES to understand the driving factors behind persistent black-white health disparities in overall longevity, cardiovascular disease, and cognitive decline. The study recruited 3,722 participants from Baltimore, MD with a mean age of 47.7 years, 2,200 African Americans and 1,522 whites, with 41% reporting household incomes below the 125% poverty delimiter.

Genotyping was done on a subset of self-reporting African American participants by the Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health (NIH). A larger genotyping effort included a small subset of self-reporting European ancestry samples. This research was supported by the Intramural Research Program of the NIH, NIA and the National Center on Minority Health and Health Disparities.

Health ABC (Health, Aging, and Body Composition): Cohort description: The Health ABC study is a prospective cohort study investigating the associations between body composition, weight-related health conditions, and incident functional limitation in older adults. Health ABC enrolled well-functioning, community-dwelling black (n=1281) and white (n=1794) men and women aged 70-79 years between April 1997 and June 1998. Participants were recruited from a random sample of white and all black Medicare eligible residents in the Pittsburgh, PA, and Memphis, TN, metropolitan areas. Participants have undergone annual exams and semi-annual phone interviews. The current study sample consists of 1559 white participants who attended the second exam in 1998-1999 with available genotyping data.

Genotyping: Genotyping was performed by the Center for Inherited Disease Research (CIDR) using the Illumina Human1M-Duo BeadChip system. Samples were excluded from the dataset for the reasons of sample failure, genotypic sex mismatch, and first-degree relative of an included individual based on genotype data. Genotyping was successful in 1663 Caucasians. Analysis was restricted to SNPs with minor allele frequency ≥ 1%, call rate ≥97% and HWE p≥10-6. Genotypes were available on 914,263 high quality SNPs for imputation based on the HapMap CEU (release 22, build 36) using the MACH software (version 1.0.16). A total of 2,543,888 imputed SNPs were analyzed for association with vitamin D levels.

Association analysis: Linear regression models were used to generate cohort-specific residuals of naturally log transformed vitamin D levels adjusted for age, sex, BMI and season defined as summer (June-August), fall (September-November), winter (December to February) and spring (March to May) standardized to have mean 0 and variance of 1. Association between the additively coded SNP genotypes and the vitamin D residuals standardized was assessed using linear regression models. For imputed SNPs, expected number of minor alleles (i.e. dosage) was used in assessing association with the vitamin D residuals.

HERITAGE (Health, Risk Factors, Exercise Training and Genetics): The HERITAGE is the only known family-based study of exercise intervention to evaluate the role of genes and sequence variants involved in the response to a physically active lifestyle. The current study is based on the data collected at baseline of the study from 99 White families (244 males, 255 females). All subjects were required to be sedentary and free of chronic diseases at baseline. There are over 18 trait domains (e.g. dietary, lipids and lipoproteins, glucose and insulin metabolism [fasting and IVGTT], steroids, body composition and body fat distribution, cardiorespiratory fitness), for a grand total of over one thousand variables. Moreover, most of the outcome traits were measured twice on two separate days both at baseline and after exercise training was completed. Marker data include a genome-wide linkage scan and GWAS, in addition to a large number of candidate genes.

HUFS (Howard University Family Study): HUFS followed a population-based selection strategy designed to be representative of African American families living in the Washington, DC metropolitan area. The major objectives of the HUFS were to study the genetic and environmental basis of common complex diseases including hypertension, obesity and associated phenotypes. Participants were sought through door-to-door canvassing, advertisements in local print media and at health fairs and other community gatherings. In order to maximize the utility of this cohort for the study of multiple common traits, families were not ascertained based on any phenotype. During a clinical examination, demographic information was collected by interview.

HyperGEN (Hypertension Genetic Epidemiology Network): HyperGEN is a family-based study that looks at the genetic causes of hypertension and related conditions in EA and AA subjects.[11] HyperGEN recruited hypertensive sibships, along with their normotensive adult offspring, and an age-matched random sample. HyperGEN has collected data on 2,471 Caucasian-American subjects and 2,300 African-American subjects, from five field centers in Alabama, Massachusetts, Minnesota, North Carolina, and Utah.

IGMM (Institute of Genetics and Molecular Medicine): IGMM oversees three participating studies: CROATIA-Korcula:; CROATIA-Vis; GS:SFHS (Generation Scotland: Scottish Family Health Study. **CROATIA-Korcula:** The CROATIA-Korcula study is a family-based, cross-sectional study in the isolated island of Korcula that included 965 examinees aged 18-95. Blood samples were collected in 2007 along with many clinical and biochemical measures and lifestyle and health questionnaires. **CROATIA-Vis:** The CROATIA-Vis study is a family-based, cross-sectional study in the isolated island of Vis that included 1,056 examinees aged 8-93. Blood samples were collected in 2003 and 2004 along with many clinical and biochemical measures and lifestyle and health questionnaires. **GS:SFHS:** The Generation Scotland (www.generationscotland.org) Scottish Family Health Study (GS:SFHS) is a family-based genetic epidemiology cohort with DNA, other biological samples (serum, urine and cryopreserved whole blood) and socio-demographic and clinical data from approximately 24,000 volunteers, aged 18-98 years, in ~7,000 family groups. An important feature of GS:SFHS is the breadth of phenotype information, including detailed data on cognitive function, personality traits and mental health. Although data collection was cross-sectional, GS:SFHS becomes a longitudinal cohort as a result of the ability to link to routine NHS data, using the community health index (CHI) number.

JHS (Jackson Heart Study): The Jackson Heart Study is a longitudinal, community-based observational cohort study investigating the role of environmental and genetic factors in the development of cardiovascular disease in African Americans. Between 2000 and 2004, a total of 5301 participants were recruited from a tri-county area (Hinds, Madison, and Rankin Counties) that encompasses Jackson, MS. Details of the design and recruitment for the Jackson Heart Study cohort has been previously published.[12-14] Briefly, approximately 30% of participants were former members of the Atherosclerosis Risk in Communities (ARIC) study. The remainder were recruited by either 1) random selection from the Accudata list, 2) commercial listing, 3) a constrained volunteer sample, in which recruitment was distributed among defined demographic cells in proportions designed to mirror those in the overall population, or through the Jackson Heart Study Family Study.

Maywood-Loyola Study: Participants were self-identified African Americans from a working class suburb of Chicago, Illinois, USA who were enrolled in studies of BP at the Loyola University Medical Center in Maywood, Illinois, USA as part of the International Collaborative Study on Hypertension in Blacks (ICSHIB) which is described in detail elsewhere.[15] Briefly, nuclear families were identified through middle-aged probands who were not ascertained based on any phenotype. Thereafter all available first-degree relatives 18 years old and above were enrolled into the study cohort of families. A screening exam was completed by trained and certified research staff using a standardized protocol.[15-16] Information was obtained on medical history, age, body weight and height. Protocols were reviewed and approved by the IRB at the Loyola University Chicago Stritch School of Medicine prior to recruitment activities. This present study included unrelated adults sampled and for whom information on anthropometrics, BP and use of antihypertensive medication was available. BP measurements were obtained using an oscillometric device, previously evaluated in our field settings.[16] Three measurements were taken three minutes apart and the average of the final two was used in the analysis. Individuals with SBP ≥140 mmHg, DBP ≥90 mmHg or on anti-hypertensive medication at time of exam were defined as hypertensive. Participants with hypertension were offered treatment after detection at the screening exam.

Maywood-Nigeria Study: The sampling frame for the Nigeria cohort was also provided by the International Collaborative Study on Hypertension in Blacks (ICSHIB) as described in detail elsewhere.[15]

Study participants were recruited from Igbo-Ora and Ibadan in southwest Nigeria as part of a long-term study on the environmental and genetic factors underlying hypertension. The base cohort consists of over 15,000 participants with information available on anthropometrics, BP and use of antihypertensive medication. BP measurements followed the same protocol described in the Loyola-Maywood study. This present study included unrelated adults samples from the cohort and some hypertensive participants who were recruited as controls in the Africa-America Diabetes Mellitus (AADM) Study recruited from Ibadan in similar neighborhoods.[17] Both projects were reviewed and approved by the sponsoring US institutions (Loyola University Chicago and Howard University) and the University of Ibadan. All participants signed informed consent administered in either English or Yoruba. BP measurements were obtained using an oscillometric device, previously evaluated in our field settings.[16] Three measurements were taken three minutes apart and the average of the final two was used in the analysis. Individuals with SBP ≥140 mmHg, DBP ≥90 mmHg or on anti-hypertensive medication at time of exam were defined as hypertensive. Participants with hypertension were offered treatment after detection at the screening exam.

MESA (Multi-Ethnic Study of Atherosclerosis): The Multi-Ethnic Study of Atherosclerosis (MESA) is a study of the characteristics of subclinical cardiovascular disease and the risk factors that predict progression to clinically overt cardiovascular disease or progression of the subclinical disease.[18] MESA consisted of a diverse, population-based sample of an initial 6,814 asymptomatic men and women aged 45-84. 38 percent of the recruited participants were white, 28 percent African American, 22 percent Hispanic, and 12 percent Asian, predominantly of Chinese descent. Participants were recruited from six field centers across the United States: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University and University of California - Los Angeles. Each participant received an extensive physical exam and determination of coronary calcification, ventricular mass and function, flow-mediated endothelial vasodilation, carotid intimal-medial wall thickness and presence of echogenic lucencies in the carotid artery, lower extremity vascular insufficiency, arterial wave forms, electrocardiographic (ECG) measures, standard coronary risk factors, sociodemographic factors, lifestyle factors, and psychosocial factors. Selected repetition of subclinical disease measures and risk factors at follow-up visits allowed study of the progression of disease. Participants are being followed for identification and characterization of cardiovascular disease events, including acute myocardial infarction and other forms of coronary heart disease (CHD), stroke, and congestive heart failure; for cardiovascular disease interventions; and for mortality. The first examination took place over two years, from July 2000 - July 2002. It was followed by four examination periods that were 17-20 months in length. Participants have been contacted every 9 to 12 months throughout the study to assess clinical morbidity and mortality.

MESA Family (Family data were not used in the analyses.)
In the MESA Family Study, the goal is to locate and identify genes contributing to the genetic risk for cardiovascular disease (CVD), by looking at the early changes of atherosclerosis within families (mainly siblings). 2128 individuals from 594 families, yielding 3,026 sibpairs divided between African Americans and Hispanic-Americans, were recruited by utilizing the existing framework of MESA. MESA Family studied siblings of index subjects from the MESA study and from new sibpair families (with the same demographic characteristics) and is determining the extent of genetic contribution to the variation in coronary calcium (obtained via CT Scan) and carotid artery wall thickness (B-mode ultrasound) in the two largest non-majority U.S. populations. The MESA Family cohort was recruited from the six MESA Field Centers. MESA Family participants underwent the same examination as MESA participants during May 2004 - May 2007.

NEO (The Netherlands Epidemiology of Obesity study): The NEO was designed for extensive phenotyping to investigate pathways that lead to obesity-related diseases. The NEO study is a population-based, prospective cohort study that includes 6,671 individuals aged 45–65 years, with an oversampling of individuals with overweight or obesity. At baseline, information on demography, lifestyle, and medical history have been collected by questionnaires. In addition, samples of 24-h urine, fasting

and postprandial blood plasma and serum, and DNA were collected. Genotyping was performed using the Illumina HumanCoreExome chip, which was subsequently imputed to the 1000 genome reference panal. Participants underwent an extensive physical examination, including anthropometry, electrocardiography, spirometry, and measurement of the carotid artery intima-media thickness by ultrasonography. In random subsamples of participants, magnetic resonance imaging of abdominal fat, pulse wave velocity of the aorta, heart, and brain, magnetic resonance spectroscopy of the liver, indirect calorimetry, dual energy X-ray absorptiometry, or accelerometry measurements were performed. The collection of data started in September 2008 and completed at the end of September 2012. Participants are currently being followed for the incidence of obesity-related diseases and mortality.

Pelotas Birth Cohort Study (The 1982 Pelotas Birth Cohort Study, Brazil): The maternity hospitals in Pelotas, a southern Brazilian city (current population ~330,000), were visited daily in the year of 1982. The 5,914 liveborns whose families lived in the urban area were examined and their mothers interviewed. Information was obtained for more than 99% of the livebirths. These subjects have been followed-up at the following mean ages: 11.3 months (all children born from January to Abril 1982; n=1457), 19.4 months (entire cohort; n=4934), 43.1 months (entire cohort; n=4742), 13.1 years (random subsample; n=715), 14.7 years (systematic subsample; n=1076); 18.2 (male cohorts attending to compulsory Army recruitment examination; n=2250), 18.9 (systematic subsample; n=1031), 22.8 years (entire cohort; n=4297) and 30.2 years (entire cohort; n=3701). Details about follow-up visits and available data can be found in the two Cohort Profile papers.[19-20] DNA samples (collected at the mean age of 22.8 years) were genotyped for ~2.5 million of SNPs using the Illumina HumanOmni2.5-8v1 array (which includes autosomal, X and Y chromosomes, and mitochondrial variants). After quality control, the data were prephased using SHAPEIT and imputed using IMPUTE2 based on 1000 Genomes haplotypes.

RS (Rotterdam Study): The Rotterdam Study is a prospective, population-based cohort study among individuals living in the well-defined Ommoord district in the city of Rotterdam in The Netherlands. The aim of the study is to determine the occurrence of cardiovascular, neurological, ophthalmic, endocrine, hepatic, respiratory, and psychiatric diseases in elderly people. The cohort was initially defined in 1990 among approximately 7,900 persons, aged 55 years and older, who underwent a home interview and extensive physical examination at the baseline and during follow-up rounds every 3-4 years (RS-I). Cohort was extended in 2000/2001 (RS-II, 3,011 individuals aged 55 years and older) and 2006/2008 (RS-III, 3,932 subjects, aged 45 and older). Written informed consent was obtained from all participants and the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam, approved the study.

SCHS-CHD (Singapore Chinese Health Study - Coronary Heart Disease): SCHS-CHD is a case-control study of coronary heart disease that was nested within the Singapore Chinese Health Study (SCHS), a prospective cohort study of 63,257 Singaporean Chinese men and women aged 45-74 years living in Singapore. We selected cases and controls from participants that provided blood samples and were free of coronary heart disease and stroke at the time of blood collection (N=24,454). Cases (N=760) had acute myocardial infarction (AMI) or died of coronary heart disease. AMI was identified through the Singapore Myocardial Infarction Registry or through the nationwide hospital discharge database followed by confirmation of AMI by cardiologists' review of medical records using the Multi-Ethnic Study of Atherosclerosis criteria (available at: http://www.mesa-nhlbi.org/manuals.aspx). Coronary heart disease deaths were identified through the Singapore Registry of Births and Deaths (ICD9 410-414 as first stated cause of death). Matched controls (N=1,491) were selected using a risk-set sampling strategy. Controls were participants who were alive and free of coronary heart disease at the time of the diagnosis or death of the index cases and were matched for age, sex, dialect group, year of recruitment and date of blood collection. In-person interviews and phlebotomy were conducted before the onset of disease and non-fasting venous blood was stored at -80^{0}C for extraction of DNA and blood biochemistry.

Singapore: SCES (Singapore Chinese Eye Study): SCES is a population-based, cross-sectional study of Chinese adults aged 40–80+ years residing in the South-Western part of Singapore, which is part of

the Singapore Epidemiology of Eye Disease (SEED). Age stratified random sampling was used to select 6,350 eligible participants, of which 3,300 participated in the study (73% response rate). Detailed methodology has been published. Two readings of blood pressure were taken from participants after 5 minutes of rest, seated, using an automated blood pressure monitor (Dinamap Pro100V2; Criticon, Norderstedt, Germany) by trained observers. One of two cuff sizes (regular, large) was chosen on the basis of the circumference of the participant's arm. A third reading was performed if the difference between two readings of either the systolic blood pressure was greater than 10mmHg or the diastolic blood pressure was greater than 5mmHg. The mean values of the closest two readings were calculated. **SiMES (Singapore Malay Eye Study):** SiMES is a population-based cross-sectional epidemiological study of 3,280 individuals from one of the three major ethnic groups residing in Singapore. SiMES is part of the Singapore Epidemiology of Eye Disease (SEED) study. In summary, 5,600 individuals have been selected by an age-stratified sampling strategy. Among these 4,168 individuals are eligible for this study. 3,280 individuals finally participated in the study. All subjects were Malay and aged 40-79 years. Two readings of blood pressure were taken from participants after 5 minutes of rest, seated, using an automated blood pressure monitor (Dinamap Pro100V2; Criticon, Norderstedt, Germany) by trained observers. One of two cuff sizes (regular, large) was chosen on the basis of the circumference of the participant's arm. A third reading was performed if the difference between two readings of either the systolic blood pressure was greater than 10mmHg or the diastolic blood pressure was greater than 5mmHg. The mean values of the closest two readings were calculated. **SINDI (Singapore Indian Eye Study):** is a population-based, cross-sectional study of Asian Indian adults aged 40–80+ years residing in the South-Western part of Singapore, which is part of the Singapore Epidemiology of Eye Disease (SEED). Age stratified random sampling was used to select 6,350 eligible participants, of which 3,400 participated in the study (75.6% response rate). Detailed methodology has been published. Two readings of blood pressure were taken from participants after 5 minutes of rest, seated, using an automated blood pressure monitor (Dinamap Pro100V2; Criticon, Norderstedt, Germany) by trained observers. One of two cuff sizes (regular, large) was chosen on the basis of the circumference of the participant's arm. A third reading was performed if the difference between two readings of either the systolic blood pressure was greater than 10mmHg or the diastolic blood pressure was greater than 5mmHg. The mean values of the closest two readings were calculated. **SP2 (Singapore 2):** The SP2 is a population-based study of diabetes and cardiovascular disease in Singapore. It first surveyed subjects (Chinese, Malay and Indian) from four cross-sectional studies that were conducted in Singapore between 1982 and 1998. Subjects were between the ages of 24-95 years and represented a random sample of the Singapore population. Subjects were re-visited between 2003 and 2007. Among the 10,747 individuals who were eligible, 5,157 subjects completed a questionnaire and the subsequent clinical examinations. Data from this re-visit were utilized for this study. Two readings of blood pressure were taken from participants after 5 min of rest, seated, using an automated blood pressure monitor (Dinamap Pro100V2; Criticon, Norderstedt, Germany) by trained observers. One of two cuff sizes (regular, large) was chosen on the basis of the circumference of the participant's arm. A third reading was performed if the difference between two readings of either the systolic blood pressure was greater than 10mmHg or the diastolic blood pressure was greater than 5mmHg. The mean values of the closest two readings were calculated.

WGHS (Women's Genome Health Study): WGHS is a prospective cohort of female North American health care professionals representing participants in the Women's Health Study (WHS) trial who provided a blood sample at baseline and consent for blood-based analyses. Participants in the WHS were 45 years or older at enrollment and free of cardiovascular disease, cancer or other major chronic illness. The current data are derived from 23,294 WGHS participants for whom whole genome genotype information was available at the time of analysis and for whom self-reported European ancestry could be confirmed by multidimensional scaling analysis of 1,443 ancestry informative markers in PLINK v. 1.06. At baseline, BP and lifestyle habits related to smoking, consumption of alcohol, and physical activity as well as other general clinical information were ascertained by a self-reported questionnaire, an approach which has been validated in the WGHS demographic, namely female health care professionals. Questionnaires recorded systolic BPe in 9 categories (<110, 110-119, 120-129, 130-139, 140-149, 150-

159, 160-169, 170-179, ≥180 mmHg), and diastolic BP in 7 categories (<65, 65-74, 75-84, 85-89, 90-94, 95-104, ≥105 mmHg). All analyses treated these BP responses as quantitative variables representing each category with its midpoint value. Hypertension was defined as one or more of reported physician diagnosis, systolic BP ≥140 mmHg, or diastolic BP ≥90 mmHg.

WHI (Women's Health Initiative): WHI is a long-term national health study that focuses on strategies for preventing common diseases such as heart disease, cancer and fracture in postmenopausal women. A total of 161,838 women aged 50–79 years old were recruited from 40 clinical centers in the US between 1993 and 1998. WHI consists of an observational study, two clinical trials of postmenopausal hormone therapy (HT, estrogen alone or estrogen plus progestin), a calcium and vitamin D supplement trial, and a dietary modification trial. Study recruitment and exclusion criteria have been described previously.[21] Recruitment was done through mass mailing to age-eligible women obtained from voter registration, driver's license and Health Care Financing Administration or other insurance list, with emphasis on recruitment of minorities and older women.[22] Exclusions included participation in other randomized trials, predicted survival < 3 years, alcoholism, drug dependency, mental illness and dementia. For the CT, women were ineligible if they had a systolic BP > 200 mm Hg or diastolic BP > 105 mm Hg, a history of hypertriglyceridemia or breast cancer. Study protocols and consent forms were approved by the IRB at all participating institutions. Socio-demographic characteristics, lifestyle, medical history and self-reported medications were collected using standardized questionnaires at the screening visit. Physical measures of height, weight and blood pressure were measured at a baseline clinical visit.[22] BP was measured by certified staff using standardized procedures and instruments.[23] Two BP measures were recorded after 5 minutes rest using a mercury sphygmomanometer. Appropriate cuff bladder size was determined at each visit based on arm circumference. Diastolic BP was taken from the phase V Korotkoff measures. The average of the two measurements, obtained 30 seconds apart, was used in analyses. The genome wide association study (GWAS) non-overlapping samples are composed of a case-control study (WHI Genomics and Randomized Trials Network – GARNET, which included all coronary heart disease, stroke, venous thromboembolic events and selected diabetes cases that happened during the active intervention phase in the WHI HT clinical trials and aged matched controls), women selected to be "representative" of the HT trial (mostly younger white HT subjects that were also enrolled in the WHI memory study - WHIMS) and the WHI SNP Health Association Resource (WHI SHARe), a randomly selected sample of 8,515 African American and 3,642 Hispanic women from WHI. GWAS was performed using Affymetrix 6.0 (WHI-SHARe), HumanOmniExpressExome-8v1_B (WHIMS), Illumina HumanOmni1-Quad v1-0 B (GARNET). Extensive quality control (QC) of the GWAS data included alignment ("flipping") to the same reference panel, imputation to the 1000G data (using the recent reference panel - v3.20101123), identification of genetically related individuals, and computations of principal components (PCs) using methods developed by Price et al. (using EIGENSOFT software 53), and finally the comparison with self-reported ethnicity. After QC and exclusions from analysis protocol, the number of women included in analysis is 4,423 whites for GARNET, 5,202 white for WHIMS, 7,919 for SHARe African American and 3,377 for SHARe Hispanics.

Gene-Lifestyle Interactions WG: STAGE 2 STUDY DESCRIPTIONS:

Brief descriptions are provided below for each of the replication studies/cohorts:

AA-DHS (African American Diabetes Heart Study): AA-DHS objectives are to improve understanding of ethnic differences in CAC and CP in populations of African and European ancestry. The AA-DHS consists of self-reported African Americans with T2D recruited from two Wake Forest School of Medicine (WFSM) studies: the family-based Diabetes Heart Study (DHS) and unrelated individuals in the AA-DHS. DHS is a cross-sectional study of European American and African American families with siblings concordant for T2D. AA-DHS started after DHS and enrolled unrelated African Americans. The AA-DHS GWAS utilized the Illumina 5M chip with imputation to 1,000 Genomes.

ASCOT (Anglo-Scandinavian Cardiac Outcomes Trial): ASCOT is a randomised control clinical trial investigating the cardiac outcomes of blood pressure lowering and lipid lowering treatments. Of 19,342 hypertensive patients (40–79 years of age with at least three other cardiovascular risk factors) who were randomized to one of two antihypertensive regimens in ASCOT (atenolol, Beta-Blocker vs amlodipine, Calcium-Channel-Blocker), 10,305 patients with non-fasting total cholesterol concentrations of 6.5 mmol/l or less (measured at the non-fasting screening visit) had been randomly assigned additional atorvastatin 10 mg or placebo. Only a proportion of United Kingdom, Irish, Sweden, Norway, Finland and Denmark consented to contribute DNA and participate in genetic studies.[24]

BBJ (Biobank Japan Project): The Biobank Japan (BBJ) Project was established in 2003 with the aim of the implementation of personalized medicine as a leading project of Ministry of Education, Culture, Sports, Science and Technology (MEXT). In collaboration with twelve cooperating institutes, the BBJ has recruited a total of 200,000 people, suffering from at least one of the 47 target common diseases, in the first phase (5-year period). BBJ has collected biospecimens including DNA and serum as well as various clinical and lifestyle information through interview or medical records by using standardized questionnaire. All participants gave written informed consent to this project and this study was approved by ethical committees of RIKEN and participating institutes.

BES (Beijing Eye Study): Beijing Eye Study is a population-based study that assess the associated and risk factors of ocular and general diseases in China population. The study was initialized in 2001, collected data from 4439 subjects aged ≥ 40 years from seven communities in Beijing area, where three of the communities were located in rural districts and four were located in urban districts. BES was followed-up in 2006, with 3251 of the original subjects participated, and in 2011, with 2695 subjects returned for the follow-up examination. At the examinations in 2006 and 2011, trained research staffs asked the subjects questions from a standard questionnaire providing information on family status, level of education, income, quality of life, psychic depression, physical activity, and known major systemic diseases. Fasting blood samples were taken for measurement of blood lipids, glucose, and glycosylated hemoglobin. Individuals were classified as self-reported non-smokers or self-reported current smokers. Alcohol consumption habits based on number of drinks per day were collected. All variables used in analyses were taken from examinations in 2006 or in 2011. The BES subjects were genotyped on two arrays, Illumina Human610-Quad (N = 832) and Illumina OmniExpress (N = 814).

BRIGHT (British Genetics of Hypertension): Participants of the BRIGHT Study are recruited from the Medical Research Council General Practice Framework and other primary care practices in the UK. Each case had a history of hypertension diagnosed prior to 60 years of age with confirmed blood pressure recordings corresponding to seated levels >150/100mmHg (1 reading) or mean of 3 readings >145/95 mmHg. BRIGHT is focused on recruitment of hypertensive individuals with BMI<30. Sample selection for GWAS was based on DNA availability and quantity.[25]

CAGE-Amagasaki (Cardio-metabolic Genome Epidemiology Network, Amagasaki Study): The Amagasaki Study (CAGE-Amagasaki) is an ongoing population-based cohort study of 5,743 individuals (3,435 males and 2,310 females), aged >18 years and recruited for a baseline examination between September 2002 to August 2003. Participants were interviewed by trained personnel to obtain information on medical and lifestyle variables, and consented to provide DNA for genotyping of molecular variants to investigate genetic susceptibility for so-called lifestyle-related diseases such as hypertension and cardiovascular disorder.

CFS (Cleveland Family Study): The Cleveland Family Study (CFS) is a family-based, longitudinal study designed to characterize the genetic and non-genetic risk factors for sleep apnea. In total, 2534 individuals (46% African American) from 352 families were studied on up to 4 occasions over a period of 16 years (1990-2006). The initial aim of the study was to quantify the familial aggregation of sleep apnea. 632 African Americans were genotyped on the Affymetrix array 6.0 platform through the CARe Consortium with suitable genotying quality control. A further 122 African-Americans had genotyping based on the Illumina OmniExpress + Exome platform. Genomes were imputed separately for each chip based on a 1000 Genomes Project Phase 3 Version 5 cosmopolitan template using SHAPEIT and IMPUTE2. Participants had three supine BP measurements each performed after lying quietly for 10 minutes, before bed (10:00 P.M.) and upon awakening (7:00 A.M.), and another three sitting at 11 am, following standardized guidelines using a calibrated sphygmomanometer. Cuff size was determined by the circumference of the upper arm and the appropriate bladder size from a standard chart. BP phenotypes were determined from the average of the nine measurements.

Colaus (Cohorte Lausannoise): The cohort is a random population sample of the city of Lausanne aged 35-75 years. Recruitment began in June 2003 and ended in May 2006, and the first follow-up was conducted between April 2009 and September 2012. The CoLaus study was approved by the Institutional Ethics Committee of the University of Lausanne and informed consent was appropriately obtained by all participants. Both at baseline and follow-up, all participants attended the outpatient clinic of the University Hospital of Lausanne in the morning after an overnight fast. Data were collected by trained field interviewers in a single visit lasting about 60 min.

DESIR (Data from an Epidemiological Study on the Insulin Resistance): The DESIR cohort study aims to: describe and understand the relations between the abnormalities of the syndrome, their evolution, according to age and sex; search for risk factors of insulin resistance, in particular factors associated with the environment, lifestyle and genetic markers; quantify the links between the syndrome and both cardiovascular disease and diabetes; evaluate the frequency of the syndrome in terms of its consequences on public health.

DFTJ (Dongfeng-Tongji Cohort Study): The DFTJ-cohort study includes 27,009 retired employees from a state-owned automobile enterprise in China. This study was launched in 2008 and will be followed up every 5 years. In 2013 we conducted the first follow-up. By using semi-structural questionnaire and health examination, those having cancer or severe diseases were excluded. Fasting blood samples and detailed epidemiology data were collected. The main goal of the cohort was to identify the environmental and genetic risk factors and the gene-environment interactions on chronic diseases, and to find novel biomarkers for chronic disease and mortality prediction. Finally, 1,461 included in the present study with GWAS data. All of the participants wrote informed consent and the ethical committees in the Tongji Medical College approved this research project. Detailed information has been described in elsewhere.[26]

QC criteria and imputation methods:

We did the GWAS scan on the DFTJ-cohort with Affymetrix Genome-Wide Human SNP Array 6.0 chips. In total, we genotyped 906,703 SNPs among 1,461 subjects. After stringent QC filtering, SNPs with MAF < 0.01, Hardy-Weinberg Equilibrium (HWE) < 0.0001, and SNP call rate < 95% were excluded. Individuals with call rates < 95% were also not included for further analysis. In total, we retained 1,452 subjects with

658,288 autosomal SNPs for statistical analyses, with an overall call rate of 99.68%. We used MACH 1.0 software to impute untyped SNPs using the LD information from the HapMap phase II database (CHB+JPT as a reference set (2007-08_rel22, released 2007-03-02). Imputed SNPs with high genotype information content (Rsq > 0.3 for MACH) were kept for the further association analysis.

DHS (Diabetes Heart Study): The Diabetes Heart Study (DHS) is an ongoing family-based cohort study investigating the epidemiology and genetics of cardiovascular disease (CVD) in a population-based sample. The DHS recruited T2D-affected siblings without advanced renal insufficiency from 1998 through 2005 in western North Carolina. DHS has collected genetic data on 1,220 self-described European American (EA) individuals from 475 families. Genotyping was completed using an Affymetrix Genome-Wide Human SNP Array 5.0 with imputation of 1,000 Genomes project SNPs from this array using IMPUTE2 and the Phase I v2, cosmopolitan (integrated) reference panel, build 37.

Dr's EXTRA (Dose Responses to Exercise Training): The Dose-Responses to Exercise Training (DR's EXTRA) Study is a 4-year RCT on the effects of regular physical exercise and healthy diet on endothelial function, atherosclerosis and cognition in a randomly selected population sample (n=3000) of Eastern Finnish men and women, identified from the national population register, aged 55-74 years. Of the eligible sample, 1410 individuals were randomized into one of the 6 groups: aerobic exercise, resistance exercise, diet, combined aerobic exercise and diet, combined resistance exercise and diet, or reference group following baseline assessments. During the four year intervention the drop-out rate was 15%.

EGCUT (Estonian Genome Center - University of Tartu (Estonian Biobank)): The Estonian Biobank is the population-based biobank of the Estonian Genome Center at the University of Tartu (www.biobank.ee; EGCUT). The entire project is conducted according to the Estonian Gene Research Act and all of the participants have signed the broad informed consent. The cohort size is up to 51535 individuals from 18 years of age and up, which closely reflects the age, sex and geographical distribution of the Estonian population. All of the subjects are recruited randomly by general practitioners and physicians in hospitals. A Computer Assisted Personal interview is filled within 1-2 hours at a doctor's office, which includes personal, genealogical, educational, occupational history and lifestyle data. Anthropometric measurements, blood pressure and resting heart rate are measured and venous blood taken during the visit. Medical history and current health status is recorded according to ICD-10 codes.

EPIC (European Prospective Investigation into Cancer and Nutrition): The European Prospective Investigation of Cancer (EPIC) began as a large multi-centre cohort study primarily looking at the connection between diet, lifestyle factors and cancer, although the study was broadened from the outset to include other conditions. The EPIC-Norfolk participants are men and women who were aged between 40 and 79 when they joined the study and who lived in Norwich and the surrounding towns and rural areas. They have been contributing information about their diet, lifestyle and health through questionnaires and health checks over two decades. The Norwich Local Research Ethics Committee granted ethical approval for the study. All participants gave written informed consent.

FENLAND (The Fenland Study): The Fenland study is a population-based cohort study that uses objective measures of disease exposure to investigate the influence of diet, lifestyle and genetic factors on the development of diabetes and obesity. The volunteers are recruited from general practice lists in and around Cambridgeshire (Cambridge, Ely, and Wisbech) in the United Kingdom from birth cohorts from 1950–1975.

FUSION (Finland-United States Investigation of NIDDM Genetics): The Finland-United States Investigation of NIDDM Genetics (FUSION) study is a long-term effort to identify genetic variants that predispose to type 2 diabetes (T2D) or that impact the variability of T2D-related quantitative traits. The FUSION GWAS sample consists of 1,161 Finnish T2D cases and 1,174 Finnish normal glucose-tolerant (NGT) controls.[27] Cases are defined by fasting plasma glucose ≥ 7.0 mmol/l or 2-h plasma glucose ≥

11.1 mmol/l, by report of diabetes medication use, or based on medical record review. 789 FUSION cases each reported at least one T2D sibling; 372 Finrisk 2002 T2D cases came from a Finnish population-based risk factor survey. NGT controls are defined by fasting glucose < 6.1 mmol/l and 2-h glucose < 7.8 mmol/l. FUSION controls include 119 subjects from Vantaa, Finland who were NGT at ages 65 and 70 years, 304 NGT spouses from FUSION families, and 651 Finrisk 2002 subjects. The controls were approximately frequency matched to the cases by age, sex, and birth province. Smoking and alcohol data are only available in the FUSION subset of our GWAS samples.

GeneSTAR (Genetic Studies of Atherosclerosis Risk): GeneSTAR is a family-based prospective study of more than 4000 participants begun in 1983 to determine phenotypic and genetic causes of premature cardiovascular disease. Families were identified from 1983-2006 from probands with a premature coronary disease event prior to 60 years of age who were identified at the time of hospitalization in any of 10 hospitals in the Baltimore, Maryland area. Their apparently healthy 30-59 year old siblings without known coronary disease were recruited and screened between 1983 and 2006. From 2003-2006, adult offspring over 21 years of age of all participating siblings and probands, as well as the coparents of the offspring were recruited and screened. Genotyping was performed in 3,232 participants on the Illumina 1Mv1_c platform.

GLACIER (Gene x Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk): The Gene-Lifestyle interactions And Complex traits Involved in Elevated disease Risk (GLACIER) Study is nested within the Västerbotten Health Survey, which is part of the Northern Sweden Health and Disease Study, a population-based prospective cohort study from northern Sweden. Participants were genotyped with Illumina CardioMetaboChip array. This array contains ~200,000 variants, the majority being common variants. Systolic and diastolic blood pressures were measured once following a period of five minutes rest with the participant in the supine position using a mercury-gauge sphygmomanometer. Analysis of serum lipids (HDL-C, triglycerides and total cholesterol) were undertaken at the Department of Clinical Chemistry at Umeå University Hospital using routine methods. LDL-C was determined using the Friedewald formula. All participants completed a detailed, optically readable, health and lifestyle questionnaire including questions about smoking status and alcohol intake (FFQ).[28]

GRAPHIC (Genetic Regulation of Arterial Pressure of Humans in the Community): The GRAPHIC Study comprises 2024 individuals from 520 nuclear families recruited from the general population in Leicestershire, UK between 2003-2005 for the purpose of investigating the genetic determinants of blood pressure and related cardiovascular traits. A detailed medical history was obtained from study subjects by standardized questionnaires and clinical examination was performed by research nurses following standard procedures. Measurements obtained included height, weight, waist-hip ratio, clinic and ambulatory blood pressure and a 12-lead ECG.

HCHS/SOL (Hispanic Community Health Study/ Study of Latinos): The HCHS/SOL is a community-based cohort study of 16,415 self-identified Hispanic/Latino persons aged 18–74 years and selected from households in predefined census-block groups across four US field centers (in Chicago, Miami, the Bronx, and San Diego). The census-block groups were chosen to provide diversity among cohort participants with regard to socioeconomic status and national origin or background. The HCHS/SOL cohort includes participants who self-identified as having a Hispanic/Latino background; the largest groups are Central American (n = 1,730), Cuban (n = 2,348), Dominican (n = 1,460), Mexican (n = 6,471), Puerto Rican (n = 2,728), and South American (n = 1,068). The HCHS/SOL baseline clinical examination occurred between 2008 and 2011 and included comprehensive biological, behavioral, and sociodemographic assessments. Consenting HCHS/SOL subjects were genotyped at Illumina on the HCHS/SOL custom 15041502 B3 array. The custom array comprised the Illumina Omni 2.5M array (HumanOmni2.5-8v.1-1) ancestry-informative markers, known GWAS hits and drug absorption, distribution, metabolism, and excretion (ADME) markers, and additional custom content including

~150,000 SNPs selected from the CLM (Colombian in Medellin, Colombia), MXL (Mexican Ancestry in Los Angeles, California), and PUR (Puerto Rican in Puerto Rico) samples in the 1000Genomes phase 1 data to capture a greater amount of Amerindian genetic variation. QA/QC procedures yielded a total of 12,803 unique study participants for imputation and downstream association analyses.

HRS (Health & Retirement Study): The Health and Retirement Study (HRS) is a longitudinal survey of a representative sample of Americans over the age of 50.[29-31] he current sample is over 26,000 persons in 17,000 households. Respondents are interviewed every two years about income and wealth, health and use of health services, work and retirement, and family connections. DNA was extracted from saliva collected during a face-to-face interview in the respondents' homes. These data represent respondents who provided DNA samples and signed consent forms in 2006, 2008, and 2010. Respondents were removed if they had missing genotype or phenotype data.

HyperGEN-AXIOM (Hypertension Genetic Epidemiology Network): HyperGEN is a family-based study that investigates the genetic causes of hypertension and related conditions in EA and AA subjects. HyperGEN recruited hypertensive sibships, along with their normotensive adult offspring, and an age-matched random sample. HyperGEN has collected data on 2,471 Caucasian-American subjects and 2,300 African-American subjects, from five field centers in Alabama, Massachusetts, Minnesota, North Carolina, and Utah. HyperGEN participates as a discovery study using GWAS available in a large subset of the samples. The remaining AA subjects without GWAS data were genotyped on the Affymetrix Axiom chip as part of a HyperGEN admixture mapping ancillary study. After excluding subjects already included in the original HyperGEN (or with family members included), this subset of approximately 450 AA subjects are included in the HyperGEN-AXIOM study which participates in replications.

INGI-CARL (Italian Network Genetic Isolates): The Carlantino cohort (INGI-CARL) is a population-based study including approximately 1000 samples from an isolated village of Southern Italy.

INGI-FVG (Italian Network Genetic Isolates): INGI-FVG is a population-based study including approximately 1700 samples from six isolated villages of Northern Italy.

InterAct (The EPIC-InterAct Case-Cohort Study): The large prospective InterAct type 2 diabetes case-cohort study is coordinated by the MRC Epidemiology Unit in Cambridge and nested within the European Prospective Investigation into Cancer and Nutrition (EPIC). EPIC was initiated in the late 1980s and involves collaboration between 23 research institutions across Europe in 10 countries (Denmark, France, Germany, Greece, Italy, the Netherlands, Norway, Spain, Sweden and the United Kingdom). The majority of EPIC cohorts were recruited from the general population, with some exceptions. French cohorts included women who were members of a health insurance scheme for school and university employees; Turin and Ragusa (Italy) and the Spanish centres included some blood donors. Participants from Utrecht (Netherlands) and Florence (Italy) were recruited via a breast cancer screening program. The majority of participants recruited by the EPIC Oxford (UK) centre consisted of vegetarian and “health conscious” volunteers from England, Wales, Scotland, and Northern Ireland.

IRAS (Insulin Resistance Atherosclerosis Study): The Insulin Resistance Atherosclerosis Study (IRAS) was an epidemiologic cohort study designed to examine the relationship between insulin resistance and carotid atherosclerosis across a range of glucose tolerance. Individuals of self-reported Mexican-American ethnicity were recruited in San Antonio, TX and San Luis Valley, CO. Recruitment was balanced across age and glucose tolerance status. Inclusion of IRAS data is limited to 194 normoglycemic individuals with genotype data from the Illumina OmniExpress and Omni 1S arrays and imputation to the 1000 Genome Integrated Reference Panel (phase I).

IRAS Family Study (Insulin Resistance Atherosclerosis Study): The IRASFS was a family study designed to examine the genetic and epidemiologic basis of glucose homeostasis traits and abdominal

adiposity. Briefly, self-reported Mexican pedigrees were recruited in San Antonio, TX and San Luis Valley, CO. Probands with large families were recruited from the initial non-family-based IRAS, which was modestly enriched for impaired glucose tolerance and T2D. Inclusion of IRASFS data is limited to 1040 normoglycemic individuals in 88 pedigrees with genotype data from the Illumina OmniExpress and Omni 1S arrays and imputation to the 1000 Genome Integrated Reference Panel (phase I).

JUPITER (Justification for the Use of Statins in Primary Prevention: An Intervention Trial Evaluating Rosuvastatin): Genetic analysis was performed in a sub-population from JUPITER (Justification for the Use of statins in Prevention: an Intervention Trial Evaluating Rosuvastatin), an international, randomized, placebo-controlled trial of rosuvastatin (20mg/day) in the primary prevention of cardiovascular disease conducted among apparently healthy men and women with LDL-C < 130 mg/dL and hsCRP ≥ 2 mg/L.[32-33] Individuals with diabetes or triglyceride concentration >500mg/dL were excluded. The present analysis includes only individuals who provided consent for genetic analysis, had successfully collected genotype information, and who had either verified European or verified South African black ancestry.

KORA (Cooperative Health Research in the Augsburg Region): The KORA study is a series of independent population-based epidemiological surveys of participants living in the region of Augsburg, Southern Germany. All survey participants are residents of German nationality identified through the registration office and were examined in 1994/95 (KORA S3) and 1999/2001 (KORA S4). In the KORA S3 and S4 studies 4,856 and 4,261 subjects have been examined implying response rates of 75% and 67%, respectively. 3,006 subjects participated in a 10-year follow-up examination of S3 in 2004/05 (KORA F3), and 3080 of S4 in 2006/2008 (KORA F4). The age range of the participants was 25 to 74 years at recruitment. Informed consent has been given by all participants. The study has been approved by the local ethics committee. Individuals for genotyping in KORA F3 and KORA F4 were randomly selected and these genotypes are taken for the analysis of the phenotypes in KORA S3 and KORA S4.

LBC1921 (Lothian Birth Cohort 1921): LBC1921 consists of 550 (234 male) relatively healthy individuals, assessed on cognitive and medical traits at a mean age of 79.1 years (SD = 0.6). They were born in 1921, most took part in the Scottish Mental Survey of 1932, and almost all lived independently in the Lothian region (Edinburgh City and surrounding area) of Scotland.[1]

LBC1936 (Lothian Birth Cohort 1936): LBC1936 consists of 1091 (548 male) relatively healthy individuals who underwent cognitive and medical testing at a mean age of 69.6 years (SD = 0.8). They were born in 1936, most took part in the Scottish Mental Survey of 1947, and almost all lived independently in the Lothian region of Scotland.[34]

LifeLines (Netherlands Biobank): Lifelines (https://lifelines.nl/) is a multi-disciplinary prospective population-based cohort study using a unique three-generation design to examine the health and health-related behaviors of 165,000 persons living in the North East region of The Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multimorbidity. In addition, the LifeLines project comprises a number of cross-sectional sub-studies which investigate specific age-related conditions. These include investigations into metabolic and hormonal diseases, including obesity, cardiovascular and renal diseases, pulmonary diseases and allergy, cognitive function and depression, and musculoskeletal conditions. All survey participants are between 18 and 90 years old at the time of enrollment. Recruitment has been going on since the end of 2006, and over 130,000 participants had been included by April 2013. At the baseline examination, the participants in the study were asked to fill in a questionnaire (on paper or online) before the first visit. During the first and second visit, the first or second part of the questionnaire, respectively, are checked for completeness, a number of investigations are conducted, and blood and urine samples are taken.

LLFS (The Long Life Family Study): LLFS is a family-based cohort study, including four clinical centers: Boston University Medical Center in Boston, MA, USA, Columbia College of Physicians and Surgeons in New York City, NY, USA, the University of Pittsburgh in Pittsburgh PA, USA, and University of Southern Denmark, Denmark. The study characteristics, recruitment, eligibility and enrollment have been previously described.[35-37] In brief, the LLFS was designed to determine genetic, behavioral, and environmental factors related to families of exceptionally healthy, elderly individuals. Phase 1 was conducted between 2006 and 2009 recruiting 4,953 individuals from 539 families. The probands were at least 79 years old in the USA centers, and 90 years old or above in Denmark. The families were selected to participate in the study based on The Family Longevity Selection Score (FLoSS),[36] a score generated according to birth-year cohort survival probabilities of the proband and siblings; probands and their families with FLoSS score of 7 or higher, at least one living sibling, and at least one living offspring (minimum family size of 3), who were able to give informed consent and willing to participate were recruited. The individuals were genotyped using ~2.3 million SNPs from the Illumina Omni chip, and then imputed on phased 1000 Genomes with Cosmopolitan data as a reference using MACH and MINIMAC. After excluding participants with 80 years and older, ~3,200 individuals have been included in the analyses for replication.

LOLIPOP (London Life Sciences Prospective Population Study): LOLIPOP is a population based prospective study of about 28K Indian Asian and European men and women, recruited from the lists of 58 General Practitioners in West London, United Kingdom between 2003 and 2008 [1]. Indian Asians had all four grandparents born on the Indian subcontinent. Europeans were of self-reported white ancestry. At enrolment all participants completed an interviewer-administered questionnaire for demographic data, medical history, and smoking and alcohol drinking habits. Anthropometric data were collected and blood pressure measured using an Omron 705CP with the mean of three measurements recorded. Blood samples were collected for the measurement of lipid profile after an overnight fasting of at least 8 hours. Aliquots of whole blood were stored at -80C for extraction of genomic DNA. The LOLIPOP study is approved by the local Research Ethics Committees and all participants provided written informed consent.

Loyola GxE (Kingston Gene-by-environment; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): The Kingston GxE cohort was obtained from a survey conducted in Kingston, Jamaica as part of a larger project to examine gene by environment interactions in the determination of blood pressure among adults 25-74 years.[38] The principal criterion for eligibility was a body mass index in either the top or bottom third of BMI for the Jamaican population. Participants were identified principally from the records of the Heart Foundation of Jamaica, a non-governmental organization based in Kingston, which provides low-cost screening services (height and weight, blood pressure, glucose, cholesterol) to the general public. Other participants were identified from among participants in family studies of blood pressure at the Tropical Metabolism Research Unit (TMRU) and from among staff members at the University of the West Indies, Mona.

Loyola SPT (Spanish Town; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): Participants were recruited from Spanish Town, a stable, residential urban area neighboring the capital city of Kingston, Jamaica as part of the ICSHIB.[38] A stratified random sampling scheme was used to recruit adult males and females aged 25–74 years from the general population. Spanish Town was chosen because its demographic make-up was broadly representative of Jamaica as a whole.

METSIM (Metabolic Syndrome In Men): The METSIM Study includes 10,197 men, aged from 45 to 73 years at recruitment, randomly selected from the population register of the Kuopio town, Eastern Finland, and examined in 2005-2010.[39] The aim of the study is to investigate genetic and non-genetic factors associated with type 2 diabetes and cardiovascular disease and its risk factors.

NESDA (Netherlands Study of Depression and Anxiety): NESDA is a multi-center study designed to examine the long-term course and consequences of depressive and anxiety disorders (http://www.nesda.nl). NESDA included both individuals with depressive and/or anxiety disorders and controls without psychiatric conditions. Inclusion criteria were age 18-65 years and self-reported western European ancestry while exclusion criteria were not being fluent in Dutch and having a primary diagnosis of another psychiatric condition (psychotic disorder, obsessive compulsive disorder, bipolar disorder, or severe substance use disorder).

OBA (French obese cases): Study of the genetic of obesity in adults.

PROCARDIS (Precocious Coronary Artery Disease): The PROCARDIS (European collaborative study of the genetics of precocious coronary artery disease) study is a multi-centre case-control study in which CAD cases and controls were recruited from the United Kingdom, Italy, Sweden and Germany. Cases were defined as symptomatic CAD before age 66 years and 80% of cases also had a sibling in whom CAD had been diagnosed before age 66 years. CAD was defined as clinically documented evidence of myocardial infarction (MI) (80%), coronary artery bypass graft (CABG) (10%), acute coronary syndrome (ACS) (6%), coronary angioplasty (CA) (1%) or stable angina (hospitalization for angina or documented obstructive coronary disease) (3%). The cases included 2,136 cases who were half or full siblings. PROCARDIS controls had no personal or sibling history of CAD before age 66 years.

RHS (Ragama Health Study): The Ragama Health Study (RHS) is a population-based study of South Asian men and women aged 35-64yrs living in the Ragama Medical Officer of Health (MOH) area, near Colombo, Sri Lanka.[40] Consenting adults attended a clinic after a 12-h fast with available health records, and were interviewed by trained personnel to obtain information on medical, sociodemographic, and lifestyle variables. A 10-mL sample of venous blood was obtained from each subject. The concurrent study was performed in two tea plantation estates in the Lindula MOH area, near Nuwara Eliya (180 km from Colombo), to investigate the gene-environment interaction in a community with differing lifestyles (e.g., physical activity and diet). BP was measured using the Omron 750CP (Omron Co., Japan) in the seated position. The average of two readings was used for the analysis. The RHS is a collaborative effort between the Faculty of Medicine, University of Kelaniya and the National Center for Global Health and Medicine, Japan.

SHEEP (Stockholm Heart Epidemiology Project): The SHEEP is a population based case-control study of risk factors for first episode of acute myocardial infarction. The study base comprised all Swedish citizens resident in the Stockholm county 1992-1994 who were 45-70 years of age and were free of previous clinically diagnosed myocardial infarction.

Cases were identified using three different sources: 1) coronary units and internal medicine wards for acute care in all Stockholm hospitals; 2) the National Patient Register; and 3) death certificates. For the present study, only cases who survived at least 28 days were considered (n=1213).

First time incident myocardial infarction cases (n=1213) were identified during a 2-year period (1992-1993) for men and during a 3-year period (1992-1994) for women. Controls (n=1561) were randomly recruited from the study population continuously over time within 2 days of the case occurrence and matched to cases on age (5-years interval), sex and hospital catchment area using computerized registers of the population of Stockholm. Five control candidates were sampled simultaneously to be able to replace potential non-respondent controls. Occasionally, because of late response of the initial control, both the first and alternative controls were considered resulting in the inclusion of more controls than cases. Postal questionnaires covering a wide range of exposure areas including occupational exposures, life style factors, social factors and health related factors were distributed to the participants. Clinical investigations were performed at least three months after myocardial infarction of cases and their matched controls. The investigations included blood samplings under fasting conditions with collection of

whole blood for DNA extraction, serum and plasma. A biobank was established containing DNA, serum and plasma.

Exposure information based on both the questionnaire and biological data from the health examination was available for 78% of the male and 67% of the female non-fatal cases; the corresponding figures for their controls were 68% and 64%.

SHIP (Study of Health in Pomerania): The Study of Health In Pomerania (SHIP) is a prospective longitudinal population-based cohort study in Mecklenburg-West Pomerania assessing the prevalence and incidence of common diseases and their risk factors.[41] SHIP encompasses the two independent cohorts SHIP and SHIP-TREND. Participants aged 20 to 79 with German citizenship and principal residency in the study area were recruited from a random sample of residents living in the three local cities, 12 towns as well as 17 randomly selected smaller towns. Individuals were randomly selected stratified by age and sex in proportion to population size of the city, town or small towns, respectively. A total of 4,308 participants were recruited between 1997 and 2001 in the SHIP cohort. Between 2008 and 2012 a total of 4,420 participants were recruited in the SHIP-TREND cohort. Individuals were invited to the SHIP study centre for a computer-assisted personal interviews and extensive physical examinations. The study protocol was approved by the medical ethics committee of the University of Greifswald. Oral and written informed consents was obtained from each of the study participants

Genome-wide SNP-typing was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 or the Illumina Human Omni 2.5 array (SHIP-TREND samples). Array processing was carried out in accordance with the manufacturer's standard recommendations. Genotypes were determined using GenomeStudio Genotyping Module v1.0 (GenCall) for SHIP-TREND and the Birdseed2 clustering algorithm for SHIP. Imputation of genotypes in SHIP and SHIP-TREND was performed with the software IMPUTE v2.2.2 based on 1000 Genomes release March 2012.

SWHS/SMHS (Shanghai Women's Health Study/ Shanghai Men's Health Study): The Shanghai Women's Health Study (SWHS) is an ongoing population-based cohort study of approximately 75,000 women who were aged 40-70 years at study enrollment and resided in in urban Shanghai, China; 56,832 (75.8%) provided a blood samples. Recruitment for the SWHS was initiated in 1997 and completed in 2000. The self-administered questionnaire includes information on demographic characteristics, disease and surgery histories, personal habits (such as cigarette smoking, alcohol consumption, tea drinking, and ginseng use), menstrual history, residential history, occupational history, and family history of cancer.

The blood pressure were measured by trained interviewers (retired nurses) with a conventional mercury sphygmomanometer according to a standard protocol, after the participants sat quietly for 5 min at the study recruitment. Included in the current project were 2970 women who had GWAS data and blood pressure measurements at the baseline interview or 892 women who had GWAS data and lipids data.

The Shanghai Men's Health Study (SMHS) is an ongoing population-based cohort study of 61,480 Chinese men who were aged between 40 and 74 years, were free of cancer at enrollment, and lived in urban Shanghai, China; 45,766 (74.4%) provided a blood samples. Recruitment for the SMHS was initiated in 2002 and completed in 2006. The self-administered questionnaire includes information on demographic characteristics, disease and surgery histories, personal habits (such as cigarette smoking, alcohol consumption, tea drinking, and ginseng use), residential history, occupational history, and family history of cancer. The blood pressure were measured by trained interviewers (retired nurses) with a conventional mercury sphygmomanometer according to a standard protocol, after the participants sat quietly for 5 min at the study recruitment. Included in the current project were 892 men who had GWAS data and blood pressure measurements at the baseline interview or 298 men who had GWAS data and lipids data.

Genotyping and imputation: Genomic DNA was extracted from buffy coats by using a Qiagen DNA purification kit (Valencia, CA) or Puregene DNA purification kit (Minneapolis, MN) according to the manufacturers' instructions and then used for genotyping assays. The GWAS genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 (Affy6.0) platform or Illumina 660, following manufacturers' protocols. After sample quality control, we exclude SNPs with 1) MAF <0.01; 2) call rate <95%; 2) bad genotyping cluster; and 3) concordance rate <95% among duplicated QC samples. Genotypes were imputed using the program MACH (http://www.sph.umich.edu/csg/abecasis/MACH/download/), which determines the probable distribution of missing genotypes conditional on a set of known haplotypes, while simultaneously estimating the fine-scale recombination map. Phased autosome SNP data from HapMap Phase II Asians (release 22) were used as the reference. To test for associations between the imputed SNP data with BMI, linear regression (additive model) was used, in which SNPs were represented by the expected allele count, an approach that takes into account the degree of uncertainty of genotype imputation (http://www.sph.umich.edu/csg/abecasis/MACH/download/).

The lipid profiles were measured at Vanderbilt Lipid Laboratory. Total cholesterol, high-density lipoprotein (HDL) cholesterol, and triglycerides (TG) were measured using an ACE Clinical Chemistry System (Alfa Wassermann, Inc, West Caldwell, NJ). Low-density lipoprotein (LDL) cholesterol levels were calculated by using the Friedwald equation. The levels of LDL cholesterol were directly measured using an ACE Clinical Chemistry System for subjects with TG levels ≥ 400 mg/dL. Fasting status was defined as an interval between the last meal and blood draw of 8 hours or longer.

TAICHI-G: The TaiChi consortium consists of 7 studies that collaborated initially in a large scale metabochip study, and became an ongoing consortium for studies of cardiometabolic disease in the Chinese population in Taiwan. The seven studies included the following: 1) HALST (Healthy Aging Longitudinal Study in Taiwan), a population based epidemiologic study of older adults living in all major geographic regions of Taiwan established by the Taiwan National Health Research Institutes (NHRI); 2) SAPPHIRe (Stanford-Asian Pacific Program in Hypertension and Insulin Resistance), a family based study established in 1995 with an initial goal of identifying major genetic loci underlying hypertension and insulin resistance in East Asian populations, with Taiwan subjects participating in the TaiChi consortium; 3) TCAGEN (Taiwan Coronary Artery Disease GENetic), a cohort study that that enrolled patients undergoing coronary angiography or percutaneous intervention at the National Taiwan University Hospital (NTUH) in the setting of either stable angina pectoris or prior myocardial infarction; 4) TACT (TAiwan Coronary and Transcatheter intervention), a cohort study enrolled patients with angina pectoris and objective documentation of myocardial ischemia who underwent diagnostic coronary angiography and/or revascularization any time after October 2000 at the National Taiwan University Hospital (NTUH) (similar to TCAGEN but recruitment was independent of TCAGEN); 5) Taiwan DRAGON (Taiwan Diabetes and RelAted Genetic COmplicatioN), acohort study of Type 2 diabetes at Taichung Veterans General Hospital (Taichung VGH) in Taiwan, with participants including individuals with either newly diagnosed or established diabetes (subjects with hyperglycemia who did not meet diagnostic criteria for Type 2 DM were not included); 6) TCAD (Taichung CAD study), includes patients with a variety of cardiovascular diseases who received care at the Taichung Veterans General Hospital (Taichung VGH), i.e. specifically individuals who were hospitalized for diagnostic and interventional coronary angiography examinations and treatment; 7) TUDR (Taiwan US Diabetic Retinopathy) enrolled subjects with Type 2 diabetes who received care at Taichung Veteran General Hospital (Taichung VGH), and a small number of subjects from Taipei Tri-Service General Hospital (TSGH); TUDR subjects underwent a complete ophthalmic and fundus examination to carefully document the presence and extent of retinopathy. From these 7 studies, samples for over 1,800 subjects were selected based on completeness of standard metabolic phenotyping and knowledge of cardiac disease status, to undergo GWAS genotyping with an Illumina human-omni 'chip' specific for Asian population (Illumina, San Diego, CA; cat. No. 20004337), hence TAICHI-G.

THRV (Taiwan study of Hypertensives Rare Variants): THRV proposed to identify rare and low frequency genetic variants for blood pressure and hypertension through whole exome sequencing of a subset of highly enriched Taiwan Chinese hypertensive families and as many matched controls. The Taiwan Chinese families (approximately N=1,200 subjects) were previously recruited as part of the NHLBI-sponsored SAPPHIRe Network which is part of the Family Blood Pressure Program (FBPP). The SAPPHIRe families were recruited to have multiple hypertensive sibs and some of them also included one normotensive/hypotensive sib. The matched controls (N=1,200) were selected from the large population-based HALST Study and a Hospital-based population, both in Taipei, Taiwan.

TRAILS (Tracking Adolescents' Individual Lives Survey): TRAILS is a prospective cohort study of Dutch adolescents with bi- or triennial measurements from age 11 to at least age 25 and consists of a general population and a clinical cohort (https://www.trails.nl/en/home). In the population cohort, four assessment waves have been completed to date, which ran from March 2001 to July 2002 (T1), September 2003 to December 2004 (T2), September 2005 to August 2007 (T3), and October 2008 to September 2010 (T4). Data for the present study were collected in the population cohort only, during the third assessment wave.

TUDR (Taiwan-US Diabetic Retinopathy): 2009 to present, is a cohort that enrolled subjects with Type 2 diabetes receiving care at Taichung Veteran General Hospital (Taichung VGH), and a small number of subjects from Taipei Tri-Service General Hospital. All TUDR subjects underwent a complete ophthalmic and fundus examination to carefully document the presence and extent of retinopathy.

TWINGENE (TwinGene of the Swedish Twin Registry): The aim of the TwinGene project has been to systematically transform the oldest cohorts of the Swedish Twin Registry (STR) into a molecular-genetic resource. Beginning in 2004, about 200 twins were contacted each month until the data collection was completed in 2008. A total of 21 500 twins were contacted where of 12 600 participated. Invitations to the study contained information of the study and its purpose. Along with the invitations consent forms and health questionnaire were sent to the subjects. When the signed consent forms where returned, the subjects were sent blood sampling equipment and asked to contact a local health facility for blood sampling. The study population was recruited among twins participating in the Screening Across the Lifespan Twin Study (SALT) which was a telephone interview study conducted in 1998-2002. Other inclusion criteria was that both twins in the pair had to be alive and living in Sweden. Subjects were excluded from the study if they preciously declined participation in future studies or if they had been enrolled in other STR DNA sampling projects. The subjects were asked to make an appointment for a health check-up at their local health-care facility on the morning Monday to Thursday and not the day before a national holiday, this to ensure that the sample would reach the KI biobank the following morning by over nights mail. The subjects were instructed to fast from 20.00 the previous night. By venipuncture a total of 50 ml of blood was drawn from each subject. Tubes with serum and blood for biobanking as well as for clinical chemistry tests were sent to KI by over night mail. One 7ml EDTA tube of whole blood is stored in -80°C while a second 7ml EDTA tube of blood is used for DNA extraction using Puregene extraction kit (Gentra systems, Minneapolis, USA). After excluding subjects in which the DNA concentration in the stock-solution was below 20ng/µl as well as subset of 302 female monozygous twin pairs participating in a previous genome wide effort DNA from 9896 individual subjects was sent to SNP&SEQ Technology Platform Uppsala, Sweden for genome wide genotyping with Illumina OmniExpress bead chip (all available dizygous twins + one twin from each available MZ twin pair).

UKB (United Kingdom Biobank, www.ukbiobank.ac.uk): UK Biobank is a major national health resource with the aim of improving the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses. UK Biobank includes data from 502,682 individuals (94% of self-reported European ancestry), with extensive health and lifestyle questionnaire data, physical measures and genetic data. A total of 152,249 participants had genetic and phenotypic (blood pressure) data. Central

genotyping quality control (QC) had been performed by UK Biobank [The UK Biobank. UK Biobank Genotyping QC documentation. (2015)]. Further QC was also performed locally.

UKHLS (Understanding Society / The UK Household Longitudinal Study): The United Kingdom Household Longitudinal Study, also known as Understanding Society (https://www.understandingsociety.ac.uk) is a longitudinal panel survey of 40.000 UK households (England, Scotland, Wales and Northern Ireland) representative of the UK population. Participants are surveyed annually since 2009 and contribute information relating to their socioeconomic circumstances, attitudes, and behaviors via a computer assisted interview. The study includes phenotypical data for a representative sample of participants for a wide range of social and economic indicators as well as a biological sample collection encompassing biometric, physiological, biochemical, and haematological measurements and self-reported medical history and medication use. The United Kingdom Household Longitudinal Study has been approved by the University of Essex Ethics Committee and informed consent was obtained from every participant.

YFS (The Cardiovascular Risk in Young Finns Study): The YFS is a population-based follow up-study started in 1980. The main aim of the YFS is to determine the contribution made by childhood lifestyle, biological and psychological measures to the risk of cardiovascular diseases in adulthood. In 1980, over 3,500 children and adolescents all around Finland participated in the baseline study. The follow-up studies have been conducted mainly with 3-year intervals. The latest 30-year follow-up study was conducted in 2010-11 (ages 33-49 years) with 2,063 participants. The study was approved by the local ethics committees (University Hospitals of Helsinki, Turku, Tampere, Kuopio and Oulu) and was conducted following the guidelines of the Declaration of Helsinki. All participants gave their written informed consent.

Gene-Lifestyle Interactions WG: STAGE 1 STUDY ACKNOWLEDGMENTS:

Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. Infrastructure for the Gene-Lifestyle Working Group is supported by the National Heart, Lung, and Blood Institute grant R01HL118305.

AGES (Age Gene/Environment Susceptibility Reykjavik Study): This study has been funded by NIH contract N01-AG012100, the NIA Intramural Research Program, an Intramural Research Program Award (ZIAEY000401) from the National Eye Institute, an award from the National Institute on Deafness and Other Communication Disorders (NIDCD) Division of Scientific Programs (IAA Y2-DC_1004-02), Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament). The study is approved by the Icelandic National Bioethics Committee, VSN: 00-063. The researchers are indebted to the participants for their willingness to participate in the study.

ARIC (Atherosclerosis Risk in Communities) Study: The ARIC study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

Baependi Heart Study (Brazil): The Baependi Heart Study was supported by Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) (Grant 2013/17368-0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Hospital Samaritano Society (Grant 25000.180.664/2011-35), through Ministry of Health to Support Program Institutional Development of the Unified Health System (SUS-PROADI).

Bio*Me* Biobank (Bio*Me* Biobank of Institute for Personalized Medicine at Mount Sinai): The Mount Sinai IPM Biobank Program is supported by The Andrea and Charles Bronfman Philanthropies.

CARDIA (Coronary Artery Risk Development in Young Adults): The CARDIA Study is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging. Genotyping was funded as part of the NHLBI Candidate-gene Association Resource (N01-HC-65226) and the NHGRI Gene Environment Association Studies (GENEVA) (U01-HG004729, U01-HG04424, and U01-HG004446). This manuscript has been reviewed and approved by CARDIA for scientific content.

CHS (Cardiovascular Health Study): This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, and R01HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National

Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

ERF (Erasmus Rucphen Family study): The ERF study as a part of EUROSPAN (European Special Populations Research Network) was supported by European Commission FP6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme "Quality of Life and Management of the Living Resources" of 5th Framework Programme (no. QLG2-CT-2002-01254). The ERF study was further supported by ENGAGE consortium and CMSB. High-throughput analysis of the ERF data was supported by joint grant from Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). ERF was further supported by the ZonMw grant (project 91111025). We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work, P. Snijders for his help in data collection and E.M. van Leeuwen for genetic imputation.

Fam HS (Family Heart Study): The FamHS is funded by R01HL118305 and R01HL117078 NHLBI grants, and 5R01DK07568102 and 5R01DK089256 NIDDK grant.

FHS (Framingham Heart Study): This research was conducted in part using data and resources from the Framingham Heart Study of the National Heart Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. This work was partially supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract Nos. N01-HC-25195 and HHSN2682015000001I) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278). A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston Medical Center. This research was partially supported by grant R01-DK089256 from the National Institute of Diabetes and Digestive and Kidney Diseases (MPIs: Ingrid B. Borecki, L. Adrienne Cupples, Kari North).

GENOA (Genetic Epidemiology Network of Arteriopathy): Support for GENOA was provided by the National Heart, Lung and Blood Institute (HL119443, HL118305, HL054464, HL054457, HL054481, HL071917 and HL87660) of the National Institutes of Health. Genotyping was performed at the Mayo Clinic (Stephan T. Turner, MD, Mariza de Andrade PhD, Julie Cunningham, PhD). We thank Eric Boerwinkle, PhD and Megan L. Grove from the Human Genetics Center and Institute of Molecular Medicine and Division of Epidemiology, University of Texas Health Science Center, Houston, Texas, USA for their help with genotyping. We would also like to thank the families that participated in the GENOA study.

GenSalt (Genetic Epidemiology Network of Salt Sensitivity): The Genetic Epidemiology Network of Salt Sensitivity is supported by research grants (U01HL072507, R01HL087263, and R01HL090682) from the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD.

GOLDN (Genetics of Diet and Lipid Lowering Network): Support for the genome-wide association studies in GOLDN was provided by the National Heart, Lung, and Blood Institute grant U01HL072524-04.

HANDLS (Healthy Aging in Neighborhoods of Diversity across the Life Span): The Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study was supported by the Intramural Research Program of the NIH, National Institute on Aging and the National Center on Minority Health and Health Disparities (project # Z01-AG000513 and human subjects protocol number 09-AG-N248).

Data analyses for the HANDLS study utilized the high-performance computational resources of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. (http://biowulf.nih.gov; http://hpc.nih.gov)).

Health ABC (Health, Aging, and Body Composition): Health ABC was funded by the National Institutes of Aging. This research was supported by NIA contracts N01AG62101, N01AG62103, and N01AG62106. The GWAS was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HHSN268200782096C. This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

HERITAGE (Health, Risk Factors, Exercise Training and Genetics): The HERITAGE Family Study was supported by National Heart, Lung, and Blood Institute grant HL-45670.

HUFS (Howard University Family Study): The Howard University Family Study was supported by National Institutes of Health grants S06GM008016-320107 to Charles Rotimi and S06GM008016-380111 to Adebowale Adeyemo. We thank the participants of the study, for which enrollment was carried out at the Howard University General Clinical Research Center, supported by National Institutes of Health grant 2M01RR010284. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official view of the National Institutes of Health. This research was supported in part by the Intramural Research Program of the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute, the National Institute of Diabetes and Digestive and Kidney Diseases, the Center for Information Technology, and the Office of the Director at the National Institutes of Health (Z01HG200362). Genotyping support was provided by the Coriell Institute for Medical Research.

HyperGEN (Hypertension Genetic Epidemiology Network): The hypertension network is funded by cooperative agreements (U10) with NHLBI: HL54471, HL54472, HL54473, HL54495, HL54496, HL54497, HL54509, HL54515, and 2 R01 HL55673-12. The study involves: University of Utah: (Network Coordinating Center, Field Center, and Molecular Genetics Lab); Univ. of Alabama at Birmingham: (Field Center and Echo Coordinating and Analysis Center); Medical College of Wisconsin: (Echo Genotyping Lab); Boston University: (Field Center); University of Minnesota: (Field Center and Biochemistry Lab); University of North Carolina: (Field Center); Washington University: (Data Coordinating Center); Weil Cornell Medical College: (Echo Reading Center); National Heart, Lung, & Blood Institute. For a complete list of HyperGEN Investigators: http://www.biostat.wustl.edu/hypergen/Acknowledge.html

IGMM (Institute of Genetics and Molecular Medicine): CROATIA-Korcula: We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools and the Croatian Institute for Public Health. We would like to acknowledge the invaluable contributions of the recruitment team in Korcula, the administrative teams in Croatia and Edinburgh and the participants. The SNP genotyping for the CROATIA-Korcula cohort was performed in Helmholtz Zentrum München, Neuherberg, Germany. CROATIA-Korcula (CR-Korcula) was funded by the Medical Research Council UK, The Croatian Ministry of Science, Education and Sports (grant 216-1080315-0302), the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947) and the Croatian Science Foundation (grant 8875). **CROATIA-Vis:** We would like to acknowledge the staff of several institutions in Croatia that supported the field work, including but not limited to The University of Split and Zagreb Medical Schools, the Institute for Anthropological Research in Zagreb and Croatian Institute for Public Health. The SNP genotyping for the CROATIA-Vis cohort was performed in the core genotyping laboratory of the Wellcome Trust Clinical Research Facility at the Western General Hospital, Edinburgh, Scotland. CROATIA-Vis (CR-Vis) was funded by the Medical Research Council UK, The Croatian Ministry of Science, Education

and Sports (grant 216-1080315-0302), the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947) and the Croatian Science Foundation (grant 8875). **GS:SFHS:** Generation Scotland received core funding from the Chief Scientist Office of the Scottish Government Health Directorate CZD/16/6 and the Scottish Funding Council HR03006. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the UK's Medical Research Council. Ethics approval for the study was given by the NHS Tayside committee on research ethics (reference 05/S1401/89). We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses.

JHS (Jackson Heart Study): The Jackson Heart Study is supported by contracts HSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung, and Blood Institute on Minority Health and Health Disparities. The authors acknowledge the Jackson Heart Study team institutions (University of Mississippi Medical Center, Jackson State University and Tougaloo College) and participants for their long-term commitment that continues to improve our understanding of the genetic epidemiology of cardiovascular and other chronic diseases among African Americans.

Maywood-Loyola Study: Maywood African-American study is supported in part by the National Institutes of Health grant numbers HL074166, R01HL074166, R01HG003054, R37HL45508 and R01HL53353.

Maywood-Nigeria Study: The Loyola-Nigeria study was supported by National Institutes of Health grant number R01HL053353 and the Intramural Research Program of the Center for Research on Genomics and Global Health, National Human Genome Research Institute (Z01HG200362). The authors acknowledge the assistance of the research staff and participants in Ibadan and Igbo-Ora, Oyo State, Nigeria.

MESA (Multi-Ethnic Study of Atherosclerosis): This research was supported by the Multi-Ethnic Study of Atherosclerosis (MESA) contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, by grant HL071205 and by UL1-DR-001079 from NCRR . Funding for MESA SHARe genotyping was provided by NHLBI Contract N02-HL-6-4278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The authors thank the participants of the MESA study, the Coordinating Center, MESA investigators, and study staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

NEO (The Netherlands Epidemiology of Obesity study): The authors of the NEO study thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Petra Noordijk, Pat van Beelen and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. Dennis Mook-Kanamori is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023).

Pelotas Birth Cohort Study (The 1982 Pelotas Birth Cohort Study, Brazil): The 1982 Pelotas Birth Cohort Study is conducted by the Postgraduate Program in Epidemiology at Universidade Federal de Pelotas with the collaboration of the Brazilian Public Health Association (ABRASCO). From 2004 to 2013, the Wellcome Trust supported the study. The International Development Research Center, World Health Organization, Overseas Development Administration, European Union, National Support Program for Centers of Excellence (PRONEX), the Brazilian National Research Council (CNPq), and the Brazilian Ministry of Health supported previous phases of the study.

Genotyping of 1982 Pelotas Birth Cohort Study participants was supported by the Department of Science and Technology (DECIT, Ministry of Health) and National Fund for Scientific and Technological Development (FNDCT, Ministry of Science and Technology), Funding of Studies and Projects (FINEP, Ministry of Science and Technology, Brazil), Coordination of Improvement of Higher Education Personnel (CAPES, Ministry of Education, Brazil).

RS (Rotterdam Study): The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

The generation and management of GWAS genotype data for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera, Marjolein Peters and Carolina Medina-Gomez for their help in creating the GWAS database, and Karol Estrada, Yurii Aulchenko and Carolina Medina-Gomez for the creation and analysis of imputed data.

SCHS-CHD (Singapore Chinese Health Study - Coronary Heart Disease): The Singapore Chinese Health Study is supported by the National Institutes of Health, USA (RO1 CA144034 and UM1 CA182876), the nested case-control study of myocardial infarction by the Singapore National Medical Research Council (NMRC 1270/2010) and genotyping by the HUJ-CREATE Programme of the National Research Foundation, Singapore (Project Number 370062002).

SCES (Singapore Chinese Eye Study), SiMES (Singapore Malay Eye Study), (SINDI) Singapore Indian Eye Study: The Singapore Malay Eye Study (SiMES), the Singapore Indian Eye Study (SINDI), and the Singapore Chinese Eye Study (SCES) are supported by the National Medical Research Council (NMRC), Singapore (grants 0796/2003, 1176/2008, 1149/2008, STaR/0003/2008, 1249/2010, CG/SERI/2010, CIRG/1371/2013, and CIRG/1417/2015), and Biomedical Research Council (BMRC), Singapore (08/1/35/19/550 and 09/1/35/19/616). Ching-Yu Cheng is supported by an award from NMRC (CSA/033/2012). The Singapore Tissue Network and the Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore provided services for tissue archival and genotyping, respectively. **SP2 (Singapore Prospective Study Program):** SP2 is supported by the individual research grant and clinician scientist award schemes from the National Medical Research Council and the Biomedical Research Councils of Singapore.

WGHS (Women's Genome Health Study): The WGHS is supported by HL043851 and HL080467 from the National Heart, Lung, and Blood Institute and CA047988 from the National Cancer Institute with collaborative scientific support and funding for genotyping provided by Amgen.

WHI (Women's Health Initiative): The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C. The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at:
http://www.whi.org/researchers/Documents%20%20Write%20a%20Paper/WHI%20Investigator%20Short%20List.pdf

Gene-Lifestyle Interactions WG: STAGE 2 STUDY ACKNOWLEDGMENTS:

Infrastructure for the CHARGE Consortium is supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. Infrastructure for the Gene-Lifestyle Working Group is supported by the National Heart, Lung, and Blood Institute grant R01HL118305.

AA-DHS (African American Diabetes Heart Study): The investigators acknowledge the cooperation of our Diabetes Heart Study (DHS) and AA-DHS participants. This work was supported by NIH R01 DK071891, R01 HL092301 and the General Clinical Research Center of Wake Forest School of Medicine M01-RR-07122.

ASCOT (Anglo-Scandinavian Cardiac Outcomes Trial): The ASCOT study was supported by Pfizer, New York, NY, USA for the ASCOT study and the collection of the ASCOT DNA repository; by Servier Research Group, Paris, France; and by Leo Laboratories, Copenhagen, Denmark. We thank all ASCOT trial participants, physicians, nurses, and practices in the participating countries for their important contribution to the study. In particular we thank Clare Muckian and David Toomey for their help in DNA extraction, storage, and handling. Genotyping was funded by the CNG, and the National Institutes of Health Research (NIHR). We would also like to acknowledge the Barts and The London Genome Centre staff for genotyping the Exome chip array. This work forms part of the research programme of the NIHR Cardiovascular Biomedical Research Unit at Barts and The London, QMUL. P.B.M. wishes to acknowledge the NIHR Cardiovascular Biomedical Research Unit at Barts and The London, Queen Mary University of London, UK for support.

BBJ (Biobank Japan Project): BioBank Japan project is supported by the Japan Agency for Medical Research and Development and by the Ministry of Education, Culture, Sports, Sciences and Technology of the Japanese government.

BES (Beijing Eye Study): BES was supported by the National Key Laboratory Fund, Beijing, China.

BRIGHT (British Genetics of Hypertension): This work was supported by the Medical Research Council of Great Britain (grant number G9521010D) and the British Heart Foundation (grant number PG/02/128). The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. This work forms part of the research program of the National Institutes of Health Research (NIHR Cardiovascular Biomedical Research) Cardiovascular Biomedical Unit at Barts and The London, QMUL. P.B.M. wishes to acknowledge the NIHR Cardiovascular Biomedical Research Unit at Barts and The London, Queen Mary University of London, UK for support.

CAGE-Amagasaki (Cardio-metabolic Genome Epidemiology Network, Amagasaki Study): The CAGE Network studies were supported by grants for the Core Research for Evolutional Science and Technology (CREST) from the Japan Science Technology Agency; the Program for Promotion of Fundamental Studies in Health Sciences, National Institute of Biomedical Innovation Organization (NIBIO); and the Grant of National Center for Global Health and Medicine (NCGM).

CFS (Cleveland Family Study): The CFS was supported by the National Institutes of Health, the National Heart, Lung, Blood Institute grant HL113338, R01HL098433, HL46380.

CoLaus (Cohorte Lausannoise): The CoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 33CSCO-122661, 33CS30-139468 and 33CS30-148401).

DESIR (Data from an Epidemiological Study on the Insulin Resistance): The DESIR Study Group is composed of Inserm-U1018 (Paris: B. Balkau, P. Ducimetière, E. Eschwège), Inserm-U367 (Paris: F.

Alhenc-Gelas), CHU d'Angers (A. Girault), Bichat Hospital (Paris: F. Fumeron, M. Marre, R. Roussel), CHU de Rennes (F. Bonnet), CNRS UMR-8199 (Lille: A. Bonnefond, P. Froguel), Medical Examination Services (Alençon, Angers, Blois, Caen, Chartres, Chateauroux, Cholet, LeMans, Orléans and Tours), Research Institute for General Medicine (J. Cogneau), the general practitioners of the region and the Cross- Regional Institute for Health (C. Born, E. Caces, M. Cailleau, N. Copin, J.G. Moreau, F. Rakotozafy, J. Tichet, S. Vol).

The DESIR study was supported by Inserm contracts with CNAMTS, Lilly, Novartis Pharma and Sanofi-aventis, and by Inserm (Réseaux en Santé Publique, Interactions entre les déterminants de la santé, Cohortes Santé TGIR 2008), the Association Diabète Risque Vasculaire, the Fédération Française de Cardiologie, La Fondation de France, ALFEDIAM, ONIVINS, Société Francophone du Diabète, Ardix Medical, Bayer Diagnostics, Becton Dickinson, Cardionics, Merck Santé, Novo Nordisk, Pierre Fabre, Roche and Topcon.

DFTJ (Dongfeng-Tongji Cohort Study): This work was supported by grants from the National Basic Research Program grant (2011CB503800), the Programme of Introducing Talents of Discipline, the grants from the National Natural Science Foundation (grant NSFC-81473051, 81522040 and 81230069), and the Program for the New Century Excellent Talents in University (NCET-11-0169).

DHS (Diabetes Heart Study): The authors thank the investigators, staff, and participants of the DHS for their valuable contributions. This study was supported by the National Institutes of Health through HL67348 and HL092301.

Dr's EXTRA (Dose Responses to Exercise Training): The study was supported by grants from Ministry of Education and Culture of Finland (722 and 627; 2004-2010); Academy of Finland (102318, 104943, 123885, 211119); European Commission FP6 Integrated Project (EXGENESIS), LSHM-CT-2004-005272; City of Kuopio; Juho Vainio Foundation; Finnish Diabetes Association; Finnish Foundation for Cardiovascular Research; Kuopio University Hospital; Päivikki and Sakari Sohlberg Foundation; Social Insurance Institution of Finland 4/26/2010.

EGCUT (Estonian Genome Center - University of Tartu (Estonian Biobank)): This study was supported by EU H2020 grants 692145, 676550, 654248, Estonian Research Council Grant IUT20-60, NIASC, EIT – Health and NIH-BMI Grant No: 2R01DK075787-06A1 and EU through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012 GENTRANSMED.

EPIC (European Prospective Investigation into Cancer and Nutrition): The EPIC Norfolk Study is funded by Cancer Research, United Kingdom, British Heart Foundation, the Medical Research Council, the Ministry of Agriculture, Fisheries and Food, and the Europe against Cancer Programme of the Commission of the European Communities. We thank all EPIC participants and staff for their contribution to the study. We thank staff from the Technical, Field Epidemiology and Data Functional Group Teams of the Medical Research Council Epidemiology Unit in Cambridge, UK, for carrying out sample preparation, DNA provision and quality control, genotyping and data handling work. We specifically thank Sarah Dawson for coordinating the sample provision for biomarker measurements, Abigail Britten for coordinating DNA sample provision and genotyping of candidate markers, Nicola Kerrison, Chris Gillson and Abigail Britten for data provision and genotyping quality control, Matt Sims for writing the technical laboratory specification for the intermediate pathway biomarker measurements and for overseeing the laboratory work.

FENLAND (The Fenland Study): The Fenland Study is funded by the Wellcome Trust and the Medical Research Council (MC_U106179471). We are grateful to all the volunteers for their time and help, and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland

Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams. We further acknowledge support from the Medical research council (MC_UU_12015/1).

FUSION (Finland-United States Investigation of NIDDM Genetics): The FUSION study was supported by DK093757, DK072193, DK062370, and ZIA-HG000024.

Genotyping was conducted at the Genetic Resources Core Facility (GRCF) at the Johns Hopkins Institute of Genetic Medicine.

GeneSTAR (Genetic Studies of Atherosclerosis Risk): [for the smoking/lipids and smoking/BP analyses] GeneSTAR was supported by National Institutes of Health grants from the National Heart, Lung, and Blood Institute (HL49762, HL59684, HL58625, HL071025, U01 HL72518, and HL087698), National Institute of Nursing Research (NR0224103), and by a grant from the National Center for Research Resources to the Johns Hopkins General Clinical Research Center (M01-RR000052).

[for the alcohol/lipids and alcohol/BP analyses] GeneSTAR was supported by National Institutes of Health grants from the National Heart, Lung, and Blood Institute (HL49762, HL59684, HL58625, HL071025, U01 HL72518, HL087698, HL092165, HL099747, and K23HL105897), National Institute of Nursing Research (NR0224103), National Institute of Neurological Disorders and Stroke (NS062059), and by grants from the National Center for Research Resources to the Johns Hopkins General Clinical Research Center (M01-RR000052) and the Johns Hopkins Institute for Clinical & Translational Research (UL1 RR 025005).

GLACIER (Gene x Lifestyle Interactions and Complex Traits Involved in Elevated Disease Risk): We thank the participants, health professionals and data managers involved in the Västerbottens Intervention Project. We are also grateful to the staff of the Northern Sweden Biobank for preparing materials and to K Enqvist and T Johansson (Västerbottens County Council, Umeå, Sweden) for DNA preparation. The current study was supported by Novo Nordisk (PWF), the Swedish Research Council (PWF), the Swedish Heart Lung Foundation (PWF), the European Research Council (PWF), and the Skåne Health Authority (PWF).

GRAPHIC (Genetic Regulation of Arterial Pressure of Humans in the Community): The GRAPHIC Study was funded by the British Heart Foundation (BHF/RG/2000004). This work falls under the portfolio of research supported by the NIHR Leicester Cardiovascular Biomedical Research Unit. CPN and NJS are funded by the BHF and NJS is a NIHR Senior Investigator.

HCHS/SOL (Hispanic Community Health Study/ Study of Latinos): The baseline examination of HCHS/SOL was supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research (NIDCR), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Neurological Disorders and Stroke, and NIH Office of Dietary Supplements additionally contributed funding to HCHS/SOL. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03). Additional analysis support was provided by 1R01DK101855-01 and 13GRNT16490017. Genotyping was also supported by National Center for Advancing Translational Sciences UL1TR000124 and NIDDK DK063491 to the Southern California Diabetes Endocrinology Research Center. This research was also supported in part by the Intramural Research Program of the NIDDK, contract no. HHSB268201200054C, and Illumina.

HRS (Health & Retirement Study): HRS is supported by the National Institute on Aging (NIA U01AG009740). Genotyping was funded separately by NIA (RC2 AG036495, RC4 AG039029). Our genotyping was conducted by the NIH Center for Inherited Disease Research (CIDR) at Johns Hopkins University. Genotyping quality control and final preparation of the data were performed by the Genetics Coordinating Center at the University of Washington.

HyperGEN-AXIOM (Hypertension Genetic Epidemiology Network): The study was support by the National Institutes of Health, the National Heart, Lung, Blood Institute grant HL086718.

INGI-CARL (Italian Network Genetic Isolates): This study was partially supported by Regione FVG (L.26.2008) and Italian Ministry of Health (GR-2011-02349604).

INGI-FVG (Italian Network Genetic Isolates): This study was partially supported by Regione FVG (L.26.2008) and Italian Ministry of Health (GR-2011-02349604).

InterAct (The EPIC-InterAct Case-Cohort Study): We thank all EPIC participants and staff for their contribution to the study. The InterAct study received funding from the European Union (Integrated Project LSHM-CT-2006-037197 in the Framework Programme 6 of the European Community).

IRAS (Insulin Resistance Atherosclerosis Study): The IRAS is supported by the National Heart Lung Institute (HL047887, HL047889, HL047890, and HL47902). Genotyping for this study was supported by the GUARDIAN Consortium with grant support from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; DK085175) and in part by UL1TR000124 (CTSI) and DK063491 (DRC). The authors thank study investigators, staff, and participants for their valuable contributions.

IRAS Family Study (Insulin Resistance Atherosclerosis Study): The IRASFS is supported by the National Heart Lung and Blood Institute (HL060944, HL061019, and HL060919). Genotyping for this study was supported by the GUARDIAN Consortium with grant support from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; DK085175) and in part by UL1TR000124 (CTSI) and DK063491 (DRC). The authors thank study investigators, staff, and participants for their valuable contributions.

JUPITER (Justification for the Use of Statins in Primary Prevention: An Intervention Trial Evaluating Rosuvastatin): Support for genotype data collection and collaborative genetic analysis in JUPITER was provided by Astra-Zeneca.

KORA (Cooperative Health Research in the Augsburg Region): The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria.

LBC1921 (Lothian Birth Cohort 1921): LBC1921 consists of 550 (234 male) relatively healthy individuals, assessed on cognitive and medical traits at a mean age of 79.1 years (SD = 0.6). They were born in 1921, most took part in the Scottish Mental Survey of 1932, and almost all lived independently in the Lothian region (Edinburgh City and surrounding area) of Scotland.[1]

LBC1936 (Lothian Birth Cohort 1936): LBC1936 consists of 1091 (548 male) relatively healthy individuals who underwent cognitive and medical testing at a mean age of 69.6 years (SD = 0.8). They were born in 1936, most took part in the Scottish Mental Survey of 1947, and almost all lived independently in the Lothian region of Scotland.[1]

(1) Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. Int J Epidemiol 2012;41:1576-1584.

LifeLines (Netherlands Biobank): The Lifelines Cohort Study, and generation and management of GWAS genotype data for the Lifelines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation.

The authors wish to acknowledge the services of the Lifelines Cohort Study, the contributing research centers delivering data to Lifelines, and all the study participants.

LLFS (The Long Life Family Study): The study is supported by the National Institute on Aging (NIA) grant U01AG023746.

LOLIPOP (London Life Sciences Prospective Population Study): The LOLIPOP study is supported by the National Institute for Health Research (NIHR) Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0601966, G0700931), the Wellcome Trust (084723/Z/08/Z), the NIHR (RP-PG-0407-10371), European Union FP7 (EpiMigrant, 279143), and Action on Hearing Loss (G51). We thank the participants and research staff who made the study possible.

Loyola GxE (Kingston Gene-by-environment; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): The Loyola GxE project was supported by NIH Grant R01HL53353.

Loyola SPT (Spanish Town; subset of International Collaborative Study of Hypertension in Blacks (ICSHIB)): The Loyola SPT project was supported by NIH Grant R01HL53353.

METSIM (Metabolic Syndrome In Men): The METSIM study was supported by the Academy of Finland (contract 124243), the Finnish Heart Foundation, the Finnish Diabetes Foundation, Tekes (contract 1510/31/06), and the Commission of the European Community (HEALTH-F2-2007 201681), and the US National Institutes of Health grants DK093757, DK072193, DK062370, and ZIA- HG000024. Genotyping was conducted at the Genetic Resources Core Facility (GRCF) at the Johns Hopkins Institute of Genetic Medicine.

NESDA (Netherlands Study of Depression and Anxiety): The infrastructure for the NESDA study is funded through the Geestkracht programme of the Dutch Scientific Organization (ZON-MW, grant number 10-000-1002) and matching funds from participating universities and mental health care organizations. Genotyping in NESDA was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health. Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation.

OBA (French obese cases): The obese French adults were recruited by the laboratory "Integrated Genomics and Metabolic Diseases Modeling" (UMR 8199 CNRS / Université de Lille 2 / Institut Pasteur de Lille) of Pr. Philippe Froguel.

PROCARDIS (Precocious Coronary Artery Disease): PROCARDIS was supported by the European Community Sixth Framework Program (LSHM-CT- 2007-037273), AstraZeneca, the British Heart

Foundation, the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Swedish Heart-Lung Foundation, the Torsten and Ragnar Söderberg Foundation, the Strategic Cardiovascular Program of Karolinska Institutet and Stockholm County Council, the Foundation for Strategic Research and the Stockholm County Council (560283). M.F and H.W acknowledge the support of the Wellcome Trust core award (090532/Z/09/Z) and the BHF Centre of Research Excellence. A.G and H.W acknowledge European Union Seventh Framework Programme FP7/2007-2013 under grant agreement no. HEALTH-F2-2013-601456 (CVGenes@Target) & and A.G, the Wellcome Trust Institutional strategic support fund.

RHS (Ragama Health Study): The RHS was supported by the Grant of National Center for Global Health and Medicine (NCGM).

SHEEP (Stockholm Heart Epidemiology Project): This study was supported by grants from the Swedish Research Council for Health, Working Life and Welfare (http://www.forte.se/en/), the Stockholm County Council (http://www.sll.se/om-landstinget/Information-in-English1/), the Swedish Research Council (http://www.vr.se/inenglish.4.12fff4451215cbd83e4800015152.html), the Swedish Heart and Lung Foundation (https://www.hjart-lungfonden.se/HLF/Om-Hjart-lungfonden/About-HLF/), and the Cardiovascular Programme at Karolinska Institutet (http://ki.se/en/mmk/cardiovascular-research-networks).

SHIP (Study of Health in Pomerania): SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania, and the network 'Greifswald Approach to Individualized Medicine (GANI_MED)' funded by the Federal Ministry of Education and Research (grant 03IS2061A). Genome-wide data were supported by the Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg- West Pomerania. The University of Greifswald is a member of the 'Center of Knowledge Interchange' program of the Siemens AG.

SWHS/SMHS (Shanghai Women's Health Study/ Shanghai Men's Health Study): We thank all the individuals who took part in these studies and all the researchers who have enabled this work to be carried out. The Shanghai Women's Health Study and the Shanghai Men's Health Study are supported by research grants UM1CA182910 and UM1CA173640 from the U.S. National Cancer Institute, respectively.

TAICHI_G: This study was supported by the National Health Research Institutes, Taiwan (PH-100-SP-01, BS-094-PP-01, PH-100-PP-03), the National Science Council, Taiwan (Grant Nos NSC 98-2314-B-075A-002-MY3, NSC 96-2314-B-002-151, NSC 96-2314-B-002-152, NSC 98-2314-B-002-122-MY2, NSC 100-2314-B-002-115, NSC 101-2325-002-078, 101-2314-B-075A-006-MY3), the National Taiwan University Hospital, Taiwan (NTUH 98-N1266, NTUH 100-N1775, NTUH 101-N2010, NTUH 101-N, VN101-04, NTUH 101-S1784).

THRV (Taiwan study of Hypertensives Rare Variants): The THRV study is supported by National Heart, Lung, and Blood Institute grant R01HL111249.

TRAILS (Tracking Adolescents' Individual Lives Survey): TRAILS (TRacking Adolescents' Individual Lives Survey) is a collaborative project involving various departments of the University Medical Center and University of Groningen, the Erasmus University Medical Center Rotterdam, the University of Utrecht, the Radboud Medical Center Nijmegen, and the Parnassia Bavo group, all in the Netherlands. TRAILS has been financially supported by grants from the Netherlands Organization for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004;

ZonMw Risk Behavior and Dependence grant 60-60600-97-118; ZonMw Culture and Health grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MaGW 480-01-006 and GB-MaGW 480-07-001; Social Sciences Council project grants GB-MaGW 452-04-314 and GB-MaGW 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013); the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), the participating universities, and Accare Center for Child and Adolescent Psychiatry. Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation.

We are grateful to all adolescents who participated in this research and to everyone who worked on this project and made it possible.

TUDR (Taiwan-US Diabetic Retinopathy): This study was supported by the National Eye Institute of the National Institutes of Health (EY014684 to J.I.R. and Y.-D.I.C.) and ARRA Supplement (EY014684-03S1, -04S1), the Eye Birth Defects Foundation Inc., the National Science Council, Taiwan (NSC 98-2314-B-075A-002-MY3 to W.H.S.) and the Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1003001C to W.H.S .). DNA handling and genotyping were supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124 and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

TWINGENE (TwinGene of the Swedish Twin Registry): The Swedish Twin Registry is financially supported by Karolinska Institutet. TwinGene project received funding from the Swedish Research Council (M-2005-1112), GenomEUtwin (EU/QLRT-2001-01254; QLG2-CT-2002-01254), NIH DK U01-066134, The Swedish Foundation for Strategic Research (SSF) and the Heart and Lung foundation no. 20070481

UKB (United Kingdom Biobank, www.ukbiobank.ac.uk): This research has been conducted using the UK Biobank Resource. The UK Biobank data was analysed from the data set corresponding to UK Biobank access application no. 236, application title “Genome-wide association study of blood pressure”, with Paul Elliott as the PI/applicant. Central analysts from Queen Mary University of London (QMUL) and Imperial College London are funded by the NIHR Cardiovascular Biomedical Research Unit at Barts and The London School of Medicine, QMUL, and the NIHR Imperial College Health Care NHS Trust and Imperial College London Biomedical Research Centre respectively. This work was supported by the UK Biobank CardioMetabolic Consortium (UKB-CMC) and the BP working group.

UKHLS (Understanding Society / The UK Household Longitudinal Study): These data are from Understanding Society: The UK Household Longitudinal Study, which is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The data were collected by NatCen and the genome wide scan data were analysed by the Wellcome Trust Sanger Institute. The Understanding Society DAC have an application system for genetics data and all use of the data should be approved by them. The application form is at:

https://www.understandingsociety.ac.uk/about/health/data.

YFS (The Cardiovascular Risk in Young Finns Study): The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Kuopio, Tampere and Turku University Hospital Medical Funds (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; Tampere

Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association.

The expert technical assistance in the statistical analyses by Leo-Pekka Lyytikäinen and Irina Lisinen is gratefully acknowledged.

CHARGE Gene-Lifestyle Interactions Working Group Roster

Gudnason, Vilmundur	AGES
Harris, Tamara	AGES
Launer, Lenore	AGES
Smith, Albert V.	AGES
Boerwinkle, Eric	ARIC/GENOA
Brown, Michael R	ARIC
De Vries, Paul	ARIC
Grove, Megan	ARIC
Morrison, Alanna	ARIC
Wang, Zhe (JJ)	ARIC
Pereira, Alexandre	Baependi Heart Study (Brazil)
Lu, Kevin	Bio*Me* Biobank
Loos, Ruth	Bio*Me* Biobank
Fornage, Myriam	CARDIA
Richard, Melissa A	CARDIA
Shikany, James	CARDIA
Bartz, Traci	CHS
Psaty, Bruce	CHS
Rice, Kenneth	CHS
Barata, Llilda	Fam HS
Feitosa, Mary	Fam HS
Kraja, Aldi	Fam HS
Province, Mike	Fam HS
Wojczynski, Mary	Fam HS
Wang, Judy	Fam HS
Chen, Brian	FHS
Cupples, Adrienne	FHS
Deng, Xuan	FHS
Levy, Dan	FHS
Liu, Ching-Ti	FHS
Bielak, Lawrence	GENOA
Kardia, Sharon	GENOA
Peyser, Patricia	GENOA
Ware, Erin	GENOA
Zhao, Wei	GENOA
Kelly, Tanika	GenSalt
Nierenberg, Amelia L	GenSalt
Arnett, Donna	GOLDN/HyperGEN
Aslibekyan, Stella	GOLDN
Baixeras, Sergi Sayols	GOLDN
Evans, Michele	HANDLS
Zonderman, Alan	HANDLS
Tajuddin, Salman	HANDLS
Liu, Youngmei	HABC
Lohman, Kurt	HABC
Bouchard, Claude	HERITAGE
Rankinen, Tuomo	HERITAGE
Baker, Jenna	HUFS
Bentley, Amy	HUFS
Rotimi, Charles	HUFS
de las Fuentes, Lisa	HyperGEN
Gu, Charles	HyperGEN
Li, Yize	HyperGEN
Rao, D.C.	HyperGEN
Schwander, Karen	HyperGEN
Sung, Yun Ju	HyperGEN
Waken, Robert J.	HyperGEN
Caroline Hayward	IGMM: GS, CROATIA (Korcula & Vis)
Jonathan Marten	IGMM: GS, CROATIA (Korcula & Vis)
Fox, Ervin	JHS
Grant, Abril	JHS
Musani, Solomon	JHS
Sims, Mario	JHS (Diet)
Cooper, Richard	Maywood-Nigeria
Tayo, Bamidele	Maywood-Nigeria
Burke, Gregory	MESA
Rotter, Jerome	MESA
Guo, Xiuqing	MESA
Yao, Jie	MESA
Lu, Yang	MESA
de Haan, Hugoline	NEO
Li, Ruifang	NEO
Mook-Kanamori, Dennis	NEO
Noordam, Raymond	NEO
Rosendaal, Frits	NEO
Hartwig, Fernando	Pelotas Birth Cohort (Brazil)
Amin, Najaf	RS
van Duijn, Cornelia	RS
Vojinovic, Dina	RS
Chai, Jin Fang	Singapore (SCES, SiMES, SINDI, SP2)
Sim, Xueling	Singapore (SCES, SiMES, SINDI, SP2)
Tai, E Shyong	Singapore (SCES, SiMES, SINDI, SP2)
Dorajoo, Rajkumar	Singapore (SCHS)
Van Dam, Rob Martinus	Singapore (Diet) (All cohorts)
Chasman, Daniel	WGHS
Chu, Audrey	WGHS
Ridker, Paul	WGHS
Franceschini, Nora	WHI
Isaacs, Steve	WHI
Reiner, Alex	WHI
Aschard, Hugo	Collaborator
Gauderman, Jim	Collaborator
Kilpelainen, Tuomas	Collaborator
Manning, Alisa	Collaborator
Munroe, Patricia	Collaborator
Ntalla, Ioanna	Collaborator
O'Connell, Jeff	Collaborator
Winkler, Thomas	Collaborator
Zhu, Xiaofeng	Collaborator

REFERENCES (for the Supplementary Material)

1. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Magi R, et al. Genetic Investigation of Anthropometric Traits (GIANT) Consortium. Quality control and conduct of genome-wide association meta-analyses. *Nature Protocols*. 2014;9:1192-1212.
2. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68-74.
3. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered.* 2007;63:111-119.
4. Sitlani CM, Rice KM, Lumley T, McKnight B, Cupples LA, Avery CL, et al. Generalized estimating equations for genome-wide association studies using longitudinal phenotype data. *Stat Med.* 2015;34:118-130.
5. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) study: Design and objectives. *Am J Epidemiol.* 1989;129:687-702.
6. de Oliveira CM, Pereira AC, de Andrade M, Soler JM, Krieger JE. Heritability of cardiovascular risk factors in a Brazilian population: Baependi Heart Study. *BMC Med Genet.* 2008;9:32.
7. Fried LP, Borhani NO, Enright P, Furberg CD, Gardin JM, Kronmal RA, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol.* 1991;1:263-276.
8. Higgins M, Province M, Heiss G, Eckfeldt J, Ellison RC, Folsom AR, et al. NHLBI Family Heart Study: objectives and design. *Am J Epidemiol.* 1996;143:1219-1228.
9. The FBPP Investigators. Multi-center genetic study of hypertension: The Family Blood Pressure Program (FBPP). *Hypertension*. 2002;39:3-9.
10. Daniels PR, Kardia SL, Hanis CL, Brown CA, Hutchinson R, Boerwinkle E, et al. Genetic Epidemiology Network of Arteriopathy study. Familial aggregation of hypertension treatment and control in the Genetic Epidemiology Network of Arteriopathy (GENOA) study. *Am J Med.* 2004;116:676-681.
11. Williams RR, Rao DC, Ellison RC, Arnett DK, Heiss G, Oberman A, et al. NHLBI family blood pressure program: methodology and recruitment in the HyperGEN network. Hypertension genetic epidemiology network. *Ann Epidemiol.* 2000;10:389-400.
12. Wyatt SB, Diekelmann N, Henderson F, Andrew ME, Billingsley G, Felder SH, et al. A community-driven model of research participation: the Jackson Heart Study Participant Recruitment and Retention Study. *Ethn Dis*. 2003;13:438-455.
13. Taylor HA Jr, Wilson JG, Jones DW, Sarpong DF, Srinivasan A, Garrison RJ, et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis*. 2005;15:S6-17.
14. Fuqua SR, Wyatt SB, Andrew ME, Sarpong DF, Henderson FR, et al. Recruiting African-American research participation in the Jackson Heart Study: methods, response rates, and sample description. *Ethn Dis*. 2005;15:S6-29.
15. Cooper R, Rotimi C, Ataman S, McGee D, Osotimehin B, Kadiri S, et al. The prevalence of hypertension in seven populations of west African origin. *Am J Public Health.* 1997;87:160-168.
16. Cooper R, Puras A, Tracy J, Kaufman J, Asuzu M, Ordunez P, et al. Evaluation of an electronic blood pressure device for epidemiological studies. *Blood Press Monit.* 1997;2:35-40.

17. Rotimi CN, Dunston GM, Berg K, Akinsete O, Amoah A, Owusu S, et al. In search of susceptibility genes for type 2 diabetes in West Africa: the design and results of the first phase of the AADM study. *Ann Epidemiol.* 2001;11:51-58.

18. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol.* 2002;156:871-881.

19. Victora CG, Barros FC. Cohort profile: the 1982 Pelotas (Brazil) birth cohort study. *Int J Epidemiol.* 2006;35:237-242.

20. Horta BL, Gigante DP, Gonçalves H, dos Santos Motta J, Loret de Mola C, Oliveira IO, et al. Cohort Profile Update: The 1982 Pelotas (Brazil) Birth Cohort Study. *Int J Epidemiol.* 2015;44:441, 441a-441e.

21. Hays J, Hunt JR, Hubbell FA, Anderson GL, Limacher M, Allen C, et al. The women's health initiative recruitment methods and results. *Ann Epidemiol.* 2003;13:S18-77.

22. Design of the women's health initiative clinical trial and observational study. The women's health initiative study group. *Control Clin Trials.* 1998;19:61-109.

23. Hsia J, Margolis KL, Eaton CB, Wenger NK, Allison M, Wu L, et al. Prehypertension and cardiovascular disease risk in the women's health initiative. *Circulation.* 2007;115:855-860.

24. Sever PS, Dahlöf B, Poulter NR, Wedel H, Beevers G, Caulfield M, et al. Anglo-Scandinavian Cardiac Outcomes Trial: a brief history, rationale and outline protocol. *J Hum Hypertens.* 2001;15: Suppl 1:S11-12.

25. Caulfield M, Munroe P, Pembroke J, Samani N, Dominiczak A, Brown M, et al. Genome-wide mapping of human loci for essential hypertension. *Lancet.* 2003;361:2118-2123.

26. Wang F, Zhu J, Yao P, Li X, He M, Liu Y, et al. Cohort Profile: the Dongfeng-Tongji cohort study of retired workers. *Int J Epidemiol.* 2013;42:731-740.

27. Scott, LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science.* 2007;316:1341-1345.

28. Kurbasic A, Poveda A, Chen Y, Agren A, Engberg E, Hu FB, et al. Gene-Lifestyle Interactions in Complex Diseases: Design and Description of the GLACIER and VIKING Studies. *Curr Nutr Rep.* 2014;3:400-411.

29. Juster FT, Suzman R. An Overview of the Health and Retirement Study. *Journal of Human Resources*. 1995;30:Suppl: S7-S56.

30. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JWR, Weir DR. Cohort Profile: the Health and Retirement Study (HRS). *Int J Epidemiol.* 2014;43:576-585.

31. Crimmins EM, Guyer H, Langa KM, Ofstedal MB, Wallace RB, Weir DR. Documentation of Physical Measures, Anthropometrics and Blood Pressure in the Health and Retirement Study. HRS Documentation Report DR-011. 2008;http://hrsonline.isr.umich.edu/sitedocs/userg/dr-011.pdf

32. Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto AM Jr, Kastelein JJ et al., JUPITER Study Group. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med*. 2008;359:2195-2207.

33. Chasman DI, Giulianini F, MacFadyen J, Barratt BJ, Nyberg F, Ridker PM. Genetic determinants of statin-induced low-density lipoprotein cholesterol reduction: the Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) trial. *Circ Cardiovasc Genet.* 2012;5:257-264. Erratum in: Circ Cardiovasc Genet. 2012;5:e27.

34. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *Int J Epidemiol*. 2012;41:1576-1584.

35. Pedersen CB, Gøtzsche H, Møller JO, Mortensen PB. The Danish Civil Registration System. A cohort of eight million persons. *Med Bull.* 2006;53:441-449.

36. Sebastiani P, Hadley EC, Province M, Christensen K, Rossi W, Perls TT, et al. A family longevity selection score: ranking sibships by their longevity, size, and availability for study. *Am J Epidemiol.* 2009;170:1555-1562.

37. Newman AB, Glynn NW, Taylor CA, Sebastiani P, Perls TT, Mayeux R et al. Health and function of participants in the Long Life Family Study: A comparison with other cohorts. *Aging (Albany NY).* 2011;3:63-76.

38. Cooper R, Rotimi C, Ataman S, McGee D, Osotimehin B, Kadiri S et al. The prevalence of hypertension in seven populations of west African origin. *Am J Public Health.* 1997;87:160-168.

39. Stancakova A, Javorsky M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M. Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,416 Finnish men. *Diabetes*. 2009;58:1212-1221.

40. Dassanayake, AS, Kasturiratne A, Rajindrajith S, Kalubowila U, Chakrawarthi S, De Silva AP, et al. Prevalence and risk factors for non-alcoholic fatty liver disease among adults in an urban Sri Lankan population. *J Gastroenterol Hepatol*. 2009;24:1284-1288.

41. Völzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N, et al. Cohort profile: the study of health in Pomerania. *Int J Epidemiol.* 2011;40:294-307.

Editorial

Genes, Environment, and the Heart
Putting the Pieces Together

Edwin P. Kirk, MB BS, PhD

The study of the interaction between genes and environment dates back through the work of R.A. Fisher and Lancelot Hogben in the early 20th century, to that of Charles Darwin and Alfred Russel Wallace in the mid-19th. Darwin and Wallace's focus was, of course, evolution by natural selection rather than human disease. However, their fundamental insight was that different members of the same species, faced with a change in environment, respond in different ways and that these differences in response are heritable. In the context of human disease, we think of gene–environment interactions in terms of people with different genotypes at a particular locus responding to an environmental stimulus, such as exposure to tobacco smoke, in different ways.

See Article by Rao et al

More broadly, there is abundant evidence that practically all human disease can be seen as being due to the effect of multiple genetic variants, and in the interaction of those genetic variants with each other, and with the environment. For example, although we think of injuries as purely environmental in nature, in fact trauma is also strongly genetically determined. The Y chromosome is a major genetic risk factor for trauma at all ages after 12 months, and beyond that, genetic variants contributing to personality traits, such as impulsivity, also serve as risk factors.[1] Similarly, infectious disease represents an interaction between environmental exposure (contact with a pathogen) and genetic predisposition. Host genetic factors are important determinants of whether exposure to a pathogen passes unnoticed by the individual or causes severe morbidity or even mortality.[2]

Conversely, it has become clear that apparently single gene disorders are nothing of the sort. Readers of *Circulation: Cardiovascular Genetics* will be familiar with the extreme variability of Mendelian disorders, such as hypertrophic cardiomyopathy, even between individuals with the same causative variant and even within the same family. This is, at least in part, because of the ameliorating or exacerbating effects of variants in genes other than the causative gene responsible for the condition (known as modifier genes), as well as environmental factors. Environmental influences on monogenic disease are mostly poorly understood, but can be striking. The impact of dietary modification on phenylketonuria and the effects of exposure to *Burkholderia cepacia* complex organisms in cystic fibrosis are just 2 examples among many.

So, we know that genetic variation is important in causing human disease. We know that environmental factors are also important. And we know that the two interact. This means that a full understanding of the basis of any disorder will never be possible, until we understand the genetic (and epigenetic) and the environmental factors, which give rise to that disorder, as well as the way that they interact with one another. Currently, however, we are a long way from this goal, even in a field as important and well-studied as cardiovascular disease. A great deal is known about cardiac environmental risk factors. Considerable effort has been expended, particularly in the past decade, on identifying genetic loci which modify disease risk. But little is known about the intersection between the 2.

Part of the reason for this is the formidable difficulty of studying such interactions. A major tool for the study of the genetic basis of common disease is the genome-wide association study (GWAS). A GWAS involves searching for associations between single-nucleotide polymorphisms and a phenotype of interest. A large number of single-nucleotide polymorphisms, typically hundreds of thousands, distributed across the genome, are tested in a cohort of thousands of individuals. The National Human Genome Research Institute - European Bioinformatics Institute Catalog of published GWAS Catalog lists ≈2500 such studies published since the first (in 2005), identifying ≈25 000 single-nucleotide polymorphism trait associations.[3] This approach has the benefit of being hypothesis-free, and unexpected associations between genetic variants and disease states discovered by GWAS have led to new biological insights.

However, reproducibility of GWAS results has been problematic, and it is demanding and expensive to conduct such studies; large numbers of well-phenotyped patients need to be recruited, consented, and genotyped. To adapt the technique to the study of gene–environment interactions adds a further layer of difficulty. The variants identified in GWAS studies are only sometimes located within genes or close enough to a particular gene that they might be surmised to be relevant to its function. This means that it is often difficult to determine the underlying biological mechanism behind the effect of a particular variant, even if the effect is relatively large and has been replicated in more than one study. Moreover, to date, most of the single-nucleotide polymorphisms which have been associated with a human trait to date, have only a modest impact on the phenotype being studied, whether that is a discrete outcome such as stroke or a continuous measure such as blood

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

From the Centre for Clinical Genetics, Sydney Children's Hospital; SEALS Genetics Laboratory, NSW Health Pathology; School of Women's and Children's Health, University of New South Wales, Randwick, Australia.

Correspondence to Edwin P. Kirk, MB BS, PhD, Centre for Clinical Genetics, Sydney Children's Hospital, High St, Randwick, NSW 2031, Australia. E-mail Edwin.kirk@health.nsw.gov.au

(*Circ Cardiovasc Genet*. 2017;10:e001764.
DOI: 10.1161/CIRCGENETICS.117.001764.)

***Circ Cardiovasc Genet* is available at**
http://circgenetics.ahajournals.org
DOI: 10.1161/CIRCGENETICS.117.001764

pressure. Studying the interaction between a variant of modest effect and an environmental exposure adds a layer of difficulty and increases the required sample sizes.

The CHARGE Consortium (Cohorts for Heart and Aging Research in Genomic Epidemiology) is a major international collaboration, the aim of which is to facilitate GWAS meta-analyses and the replication of GWAS results. In this issue of *Circulation: Cardiovascular Genetics*, Rao et al[4] report the formation, structure, and administration of, and methodology used by, the Gene-Lifestyle Interactions Working Group. This Working Group works with the resources of CHARGE to study the interactions of genetic variation and environmental factors. The phenotypes being studied in this first phase of the project are blood pressure and lipids. The environmental factors are smoking, alcohol consumption, education (as a surrogate for socioeconomic status), physical activity, a set of psychosocial attributes, and sleep duration.

The scale of this undertaking is enormous. In a field which was once highly competitive, no fewer than 124 cohorts, including 610 475 subjects from 5 ancestry groups, have been drawn together into a coordinated whole. The approach involves agreement on a standard study design, which is then implemented in each separate cohort. The resulting data are uploaded in a standard format to a central server, and meta-analysis is conducted across the cohorts. The investigators leverage the power of their huge sample size by conducting analyses using different models and with different structures. For example, the now standard approach of doing genome-wide analyses in a discovery cohort, followed by targeted analysis in a replication cohort is used, with 149 684 individuals in stage 1 and 490 791 individuals in stage 2. However, combined analyses using all 124 cohorts are also performed; using both approaches allows the maximum possible discovery to be made with the resources available.

Why have the investigators gone to the considerable lengths that they have? Why does it matter to the field that they have done so, and what kind of results can we expect and hope for from this study? The GWAS Catalog lists just 22 studies of gene–environment interaction, all published after 2010 and only 4 of which are directly relevant to cardiovascular disease. Already, 4 projects from the Working Group have completed all analyses and are being taken toward publication; effectively, they are already set to double the world literature in this field. Five other projects are well underway and more are being planned. We can expect to see a steady flow of papers, reporting new associations and, importantly, bridging the gene/environment gap, in a systematic way that has not been possible in the past. It is likely that novel biology will be revealed. Moreover, there is evident potential to expand the scope of the phenotypes and interactions that the group studies.

What about impacts for patient care? It seems unlikely we will ever be saying to patients "your particular set of genetic variants means that it's fine for you to smoke"…or to refrain from exercise, or subsist on fast food. However, it is not unrealistic to expect that better understanding of the genetic basis of cardiovascular phenotypes, especially the role that genotypes play in modifying response to environmental exposures, will allow improved stratification of patients, both by risk and by likely response to interventions. It is no stretch to say that this article represents an important milestone on the path toward a far more complete understanding of the origins of cardiovascular disease and a better understanding of how to manage it.

Disclosures

None.

References

1. Bevilacqua L, Goldman D. Genetics of impulsive behaviour. *Philos Trans R Soc Lond B Biol Sci*. 2013;368:20120380. doi: 10.1098/rstb.2012.0380.
2. Antonelli G, Roilides E. Host genetics: deciphering the variability in susceptibility to infections. *Clin Microbiol Infect*. 2014;20:1235–1236. doi: 10.1111/1469-0691.12789.
3. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res*. 2017;45(D1):D896–D901. doi: 10.1093/nar/gkw1133.
4. Rao DC, Sung YJ, Winkler TW, Schwander K, Borecki I, Cupples LA, et al; on behalf of the CHARGE Gene-Lifestyle Interactions Working Group. A Multiancestry study of gene–lifestyle interactions for cardiovascular traits in 610475 individuals from 124 cohorts: design and rationale. *Circ Cardiovasc Genet*. 2017;10:e001649. doi: 10.1161/CIRCGENETICS.116.001649.

KEY WORDS: Editorials ■ cardiovascular diseases ■ environment ■ genetics ■ gene-environment interaction ■ genotype