Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Association of Exome Sequences With Cardiovascular Traits among Blacks in the Jackson Heart Study

Gina M. Peloso, PhD[1,2,3], **Leslie A. Lange, PhD**[4], **Tibor V. Varga, PhD**[5], **Deborah A. Nickerson, PhD**[6], **Joshua D. Smith, PhD**[6], **Michael E. Griswold, PhD**[7], **Solomon Musani, PhD**[8], **Linda M. Polfus, PhD**[9], **Hao Mei, PhD**[7], **Stacey Gabriel, PhD**[3], **Rakale Collins Quarells, PhD**[10], **David Altshuler, MD, PhD**[3,11], **Eric Boerwinkle, PhD**[9], **Mark J. Daly, PhD**[3,12], **Benjamin Neale, PhD**[3,12], **Adolfo Correa, MD, PhD**[13], **Alex P. Reiner, MD**[14], **James G. Wilson, MD**[15], and **Sekar Kathiresan, MD**[2,3,11]

[1]Department of Biostatistics, Boston University School of Public Health, Boston

[2]Center for Human Genetic Research, Massachusetts General Hospital, Boston

[3]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA

[4]Department of Genetics, University of North Carolina, Chapel Hill, NC

[5]Genetic and Molecular Epidemiology Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden

[6]Department of Genome Sciences, University of Washington, Seattle, WA

[7]Center of Biostatistics & Bioinformatics, University of Mississippi Medical Center, Jackson, MS

[8]Department of Medicine, University of Mississippi Medical Center, Jackson, MS

[9]Human Genetics Center, University of Texas Health Science Center, Houston, TX

[10]Cardiovascular Research Institute, Morehouse School of Medicine, Atlanta, GA

[11]Department of Medicine, Harvard Medical School, Boston, MA

[12]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston

[13]Departments of Pediatrics & Medicine, University of Mississippi Medical Center, Jackson, MS

[14]Department of Epidemiology, University of Washington School of Public Health, Seattle, WA

[15]Department of Physiology & Biophysics, University of Mississippi Medical Center, Jackson, MS

Abstract

Background—The correlation of null alleles with human phenotypes can provide insight into gene function in humans. In individuals of African ancestry, we set out to identify null and

Correspondence: Gina M Peloso, PhD, Department of Biostatistics, Boston University, 801 Massachusetts Ave, 3rd Floor, Boston, MA 02118, Tel: 617-638-5168, gpeloso@bu.edu. Sekar Kathiresan, MD, Center for Human Genetic Research, Massachusetts General Hospital, 185 Cambridge St., CPZN 5.248, Boston, MA 02114, Tel: 617-643-6120, Fax: 617-507-7766, skathiresan@partners.org.

Disclosures: None

damaging missense variants, and test these variants for association with a range of cardiovascular phenotypes.

Methods and Results—We performed whole exome sequencing in 3,223 African American individuals from the Jackson Heart Study and found a total of 729,666 variant sites with minor allele frequency (MAF) < 5%, including 17,263 null variants and 49,929 missense variants predicted to be damaging by *in silico* algorithms. We tested null and damaging missense variants within each gene for association with 36 cardiovascular traits. We found three associations that met our pre-specified level of significance ($\alpha=1.1\times10^{-7}$). Null and damaging missense variants in *PCSK9* were associated with 36 mg/dl lower low density lipoprotein cholesterol (LDL-C) (p-value=3×10^{-21}). Three individuals in their 50s with complete PCSK9 deficiency (each compound heterozygote for *PCSK9* p.Y142X and p.C679X) were identified, with one having a coronary artery calcification score in the 83rd-percentile despite a LDL-C of 32 mg/dl. A damaging missense variant in *HBQ1* (p.G52A) was associated with a 2 pg/cell lower mean corpuscular hemoglobin (p-value=9×10^{-13}) and rare damaging missense variants in *VPS13A* with higher red blood cell distribution width (p-value=9.9×10^{-8}).

Conclusions—A limited number of null/damaging alleles with a large effect on cardiovascular traits were detectable in ~3,000 African American individuals.

Journal Subject Terms

Lipids and Cholesterol; Genetics; Genetic; Association Studies

Keywords

low-density lipoprotein cholesterol; exome; genetic association; lipids; PCSK9; loss of function; rare variant; VPS13A

A compelling therapeutic target for lowering low-density lipoprotein cholesterol (LDL-C) emerged from human genetic studies - the proprotein convertase subtilisin/kexin type 9 gene (*PCSK9*)[1]. Null alleles (also termed loss-of-function [LoF] protein-coding sequence variants) in *PCSK9* were identified in African Americans[2] and shown to associate with lower plasma LDL-C levels[2–4] as well as reduced risk for CHD (up to 88% reduction)[5, 6]. Based on this human genetic evidence as well as corroborating functional studies, several pharmaceutical companies have established drug development programs targeting PCSK9[7] and two inhibitors have been approved for reducing LDL-C in individuals with heterozygous familial hypercholesterolemia and individuals with clinical atherosclerotic cardiovascular disease[8, 9]. Based on the *PCSK9* example, it has been suggested that low-frequency or rare mutations of large effect may be paradigmatic for therapeutic target discovery[10].

To address whether additional such examples can be readily identified, we sequenced the exomes of 3,223 individuals from the Jackson Heart Study (JHS), a prospective cohort of African Americans living in Jackson, Mississippi, and catalogued null as well as damaging missense mutations across 18,465 genes. Subsequently, we performed an association study of these variants with a range of quantitative and qualitative cardiovascular traits.

Methods

Study Participants

The JHS is a community-based longitudinal cohort study located in the Jackson, Mississippi metropolitan area designed to investigate the determinants of cardiovascular disease in African Americans[11]. JHS recruited 5,301 African Americans, aged between 35–84, between September 2000 and March 2008[11]. The Institutional Review Board of the University of Mississippi Medical Center approved the study protocol and all participants provided written informed consent.

Exome Sequencing

Exome sequencing was performed at three sequencing centers (the Broad Institute [n = 2,317], University of Washington [n = 481], and Baylor University [n = 475]) across 5 projects (The U.S. National Heart, Lung, and Blood Institute's [NHLBI] Exome Sequencing Project [ESP], Myocardial Infarction Genetics Consortium Exome Sequencing Project [MIGen ExS], CHARGE-S, Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples [T2D-GENES], and Minority Health Genomics and Translational Research Bio-Repository Database [MH-GRID]) (Supplemental Table 1). The sequencing reads (i.e. fastq files) from exomes were aligned to the human genome reference (hg19) using bwa on a per lane basis and bam files were obtained from the three sequencing centers. The Genome Analysis Toolkit (GATK) v3.1 HaplotypeCaller algorithm was used for joint variant discovery and genotyping on both exomes and flanking 50bp of intronic sequence (http://www.broadinstitute.org/gatk/guide/article?id=3893). Single-sample gVCFs were created using the GATK HaplotypeCaller with the options -emitRefConfidence GVCF, --variant_index_type LINEAR, and --variant_index_parameter 128000. Then batches of ~200 gVCFs were merged into a single gVCF using the CombineGVCF command in GATK. Finally, GenotypeGVCFs was run on the combined gVCFs to create the raw SNP and indel VCFs. As a majority of individuals were sequenced at the Broad Institute, we limited analysis to the sequence intervals captured by the Broad's exome sequencing platform.

Variant Quality Control

GATK Variant Quality Score Recalibration (VQSR) was used with the recommended resources to filter variants. The SNP VQSR model was trained using HapMap3.3 and 1KG Omni 2.5 SNP sites and a 99.5% sensitivity threshold was applied to filter variants, while the INDEL VQSR model was trained using the Mills 1000G gold standard and Axiom Exome Plus sites for insertions/deletions and a 99.0% sensitivity threshold was applied to filter INDEL sites. Variants were filtered to VQSR PASS and quality depth (QD) 2. (Supplemental Table 2). Individual genotypes were set to missing if depth < 5.

Sample Quality Control

We performed quality control on the jointly-called samples. Individuals were checked for total number of variants, observed number of singletons and doubletons, Ti/Tv ratio, Het/Hom ratio, missingness, contamination with VerifyBamID[12], and non-reference

concordance with available genotype data from the Illumina HumanExome BeadChip v1.0. Individuals that were outliers ($> \pm 3$*interquartile range) on at least one metric were excluded (Supplemental Table 1, Supplemental Figure 1). Population structure was assessed using the multi-dimensional scaling (MDS) algorithm in the PLINK software[13] and ten principal components of ancestry were obtained (Supplemental Figure 2).

Annotation

All variant sites were annotated with the Variant Effect Predictor algorithm (VEP; http://useast.ensembl.org/info/docs/tools/vep) and dbSFP[14] (https://sites.google.com/site/jpopgen/dbNSFP). Analysis was limited to variants predicted to be null (nonsense, splice, frameshift) plus missense variation predicted to be damaging in at least five of the following seven variation prediction tools[15]: LRT[16], Mutation Taster[17], PolyPhen2[18] (HumDiv), PolyPhen2 (HumVar), SIFT[19], MutationAssessor[20] and FATHMM[21].

Phenotypes

We analyzed 36 cardiovascular traits (Figure 1) available in the Jackson Heart Study Vanguard Center data package (https://www.jacksonheartstudy.org/jhsinfo/ForResearchers/VanguardCenters/tabid/171/Default.aspx). For participants who were taking antihypertensive medication, we added 10 mm Hg to observed systolic blood pressure (SBP) values and 5 mm Hg to diastolic blood pressure (DBP) values[22]. We adjusted the total cholesterol values for individuals on lipid lowering medication by replacing their total cholesterol values by total cholesterol divided by 0.8[23]. No adjustment was made on high-density lipoprotein cholesterol (HDL-C) or triglycerides. Only fasting lipid measures were used and LDL-C was calculated using the Friedewald equation for those with triglycerides < 400 mg/dl, using the lipid adjusted total cholesterol for those on treatment.

Individuals with diabetes were excluded in analyses of fasting plasma glucose, fasting insulin, HOMA-IR, HOMA-B, and HbA1c. Individuals with QRS > 120, atrial fibrillation, or coronary heart disease were excluded for analysis of QRS interval. Individuals with QRS 120, ECG heart rate < 40, ECG heart rate > 120, or with atrial fibrillation were excluded from the analysis of QT interval. Individuals with end stage renal disease (ESRD) defined as eGRF < 15 or reporting being on dialysis, hemoglobinopathy defined as being homozygous for rs334, or myelotoxic drug use were excluded from the blood cell trait analyses.

Non-normality of the following raw traits was resolved by a natural log transform before analysis: triglycerides, leptin, hsCRP, endothelin, renin, aldosterone, and adiponectin. Non-normality was resolved by the log transformation.

Association Analysis

We performed gene-based analyses of 36 cardiovascular phenotypes. We limited analysis to null mutations plus missense variants predicted to be damaging by at least 5 of 7 *in silico* prediction algorithms (LRT, Mutation Taster, PolyPhen2 (HumDiv), PolyPhen2 (HumVar), SIFT, MutationAssessor and FATHMM)[15]. We aggregated variants with minor allele frequency (MAF) < 5% within each gene using four sets of variants: (1) null mutations only, (2) null mutations plus missense variants predicted to be damaging by 7 of 7 *in silico*

prediction algorithms, (3) null mutations plus missense variants predicted to be damaging in at least 6 of 7 *in silico* prediction algorithms, and (4) null mutations plus missense variants predicted to be damaging in at least 5 of 7 *in silico* prediction algorithms. All associations were performed using the EPACTS (http://genome.sph.umich.edu/wiki/EPACTS) software. EPACTS (Efficient and Parallelizable Association Container Toolbox) is a software pipeline to perform statistical tests of association using sequence data. It implements the EMMAX[24] (Efficient Mixed Model Association eXpedited) model, a mixed model association approach that captures pedigree, cryptic relatedness, and population structure by using a covariance matrix estimated from genome-wide data. To apply the EMMAX model, we used the epacts-group command with the emmaxCMC test option to perform collapsing burden gene-based tests. The single command with the q.emmax test option in EPACTS was used to obtain the single variant results for each variant going into the gene-based test. We used an additive genetic model. A kinship matrix of all individuals was created with EPACTS and used in analyses. All analyses were adjusted for age, sex, and 4 principal components of ancestry. Analyses for QT interval and QRS additionally included adjustments for height and BMI.

We excluded results with ≥ 10 minor alleles contributing to the gene-based test to ensure robust association statistics. We set our significance threshold to 1.1×10^{-7} (0.05/[36 traits*~12,500 genes after minor allele count exclusion]).

A Wilcoxon rank sum test was performed to compare *PCSK9* null compound heterozygous carriers to heterozygous carriers using the R software (version 3.1). Coronary artery calcification (CAC) percentiles were calculated with the MESA CAC Score Reference Values web tool (http://www.mesa-nhlbi.org/Calcium/input.aspx)[25].

Power

We performed power calculations using the Genetic Power calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/) with the "QTL association for sib-ships and singletons" option.

Results

After quality control, 3,223 individuals from the Jackson Heart Study were available for analysis (Table 1, Supplemental Table 1). We observed 17,263 null variants with MAF < 5% and 49,929 missense variants predicted to be damaging in at least 5 of 7 *in silico* prediction algorithms with MAF < 5% (Supplemental Table 2). Of the 18,465 genes sequenced, 14,058 have a null or damaging missense variant with MAF < 5%. On average, we observe 5 null or damaging missense variants per gene and an average of 7 null or damaging missense alleles per gene. Each individual carries, on average, a total of 153 null or damaging missense variants with MAF < 5%.

We found three gene-based associations that met our pre-specified significance threshold of 1.1×10^{-7} (Table 2, Supplemental Table 3, Supplemental Table 4). The most significant association was between LDL-C and *PCSK9*. Participants who carried null or damaging missense mutations in *PCSK9* had 36 mg/dl lower LDL-C compared with non-carriers (p-value=2.9×10^{-21}). Of note, we identified three individuals with complete PCSK9

deficiency (each compound heterozygote for *PCSK9* p.Y142X and p.C679X) (Table 3). These individuals had a lower median LDL-C (64.2 mg/dl) compared to individuals that carry only one null mutation (85.7 mg/dl; n=77) (p-value=0.044; Supplemental Figure 3). The three *PCSK9* null compound heterozygotes did not differ from heterozygotes in any other cardiometabolic trait tested except QT interval (Supplemental Table 5). Compound heterozygotes had a lower QT interval (mean=369, range=362–380) compared to individuals that carried only one null *PCSK9* variant (mean=413) (p-value=0.006 using a Wilcoxon rank sum test). Individuals carrying one null *PCSK9* variant had similar QT intervals compared with non-carriers (mean=413), suggesting a recessive effect. Two individuals carrying both *PCSK9* p.Y142X and p.679X had a coronary artery calcification (CAC) greater than the 80th-percentile for their age and sex. A 52-year-old man had a CAC of 24.9, which is in the 83rd-percentile for age and sex, despite an LDL-C of 32 mg/dl (Table 3).

The second most significant gene association was between mean corpuscular hemoglobin (MCH) and hemoglobin subunit theta 1 (*HBQ1*). Individuals carrying a damaging missense variant (p.G52A)[26] in *HBQ1* had lower MCH compared with non-carriers (p-value=8.4×10^{-13}). One additional association passed our significance threshold. Rare damaging missense variants in Vacuolar Protein Sorting-Associated Protein 13A (*VPS13A*) were associated with an increase in red blood cell distribution width (p-value=7.1×10^{-8}). Of the nine variants that contributed to the association between *VPS13A* and red blood cell distribution width, six were singletons, one a doubleton, one with four carriers (p.S2673L) and one with 22 minor allele carriers (p.K2672N) (Supplemental Table 4). *VPS13A* showed evidence for association with other hematologic phenotypes, including lower hemoglobin levels (p-value=7.0×10^{-04}; Supplemental Table 6).

Li et al[27] recently reported ten gene-based associations aggregating null variants with a p-value $< 4.4 \times 10^{-6}$. Individuals of African Ancestry contributed to seven of these associations. We attempted to replicate these seven associations in our data (Supplemental Table 7). We replicated the association of total cholesterol with *PCSK9* (beta = -39 mg/dl; p-value = 6.6×10^{-12}), and of triglycerides with apolipoprotein C-III (*APOC3*; p-value = 1.0×10^{-5})[2, 28–30]. We found suggestive evidence for the association of fasting glucose with thioredoxin domain containing 5 (*TXNDC5*), consistent with the report by Li et al; carriers of null alleles in *TXNDC5* had higher fasting glucose compared with non-carriers (p=0.07).

For 3,223 individuals and a significance level of 1.1×10^{-7}, we had 99% statistical power to detect a 1-standard deviation unit effect with a 1% cumulative minor allele frequency, and 64% statistical power to detect a 1-standard deviation unit effect with a 0.5% cumulative minor allele frequency. Analysis of Mendelian lipid genes as a 'positive control' shows several genes where a burden of null/damaging mutations alters the expected plasma lipid fraction in the appropriate direction (e.g., *LDLR* and higher LDL-C [P=4.7×10^{-5}], *CETP* and higher HDL-C [P=0.0001]) (Supplemental Table 8). However, even an analysis of positive controls is limited by the number of carriers, with the majority of the Mendelian lipid genes having < 10 observed null alleles.

Discussion

We set out to discover null or damaging missense variants that lead to a large effect on any of a range of cardiovascular traits. In a study of 3,223 African Americans, we found three associations that met our pre-specific significance threshold.

We report one new observation, that of *VPS13A* associated with an increase in red blood cell distribution width (RDW). RDW is a measure of the range of variation in red blood cells and higher values can indicate certain disorders such as anemia. Mutations in *VPS13A* have been reported to cause chorea-acanthocytosis, an autosomal recessive neurodegenerative disorder that causes red blood cells to appear spiky[31]. Ten *VPS13A* variants are reported in ClinVar with chorea-acanthocytosis listed as the condition. We did not find any of the reported ClinVar variants in our data nor any carriers of rare damaging recessive variants in *VPS13A*. Here, in a sample of individuals unselected for disease state, we report a milder phenotype resulting from heterozygous mutations in *VPS13A*. Similar to *VPS13A*, Mendelian lipid genes having a large effect on plasma lipid levels have been shown to harbor common variants with smaller effects on phenotype[32–34].

We found three individuals who are compound heterozygous for null mutations in *PCSK9*. Previously, only two individuals with *PCSK9* deficiency have been reported[35, 36]. Both of the previously reported individuals were young (21 and 31 years old) and had very low circulating LDL-C (14–16 mg/dl). The three individuals we have identified here are older (50–52 years old) and have higher circulating LDL-C (32–72 mg/dl). One of the three individuals had a CAC score in the 83rd-percentile despite a LDL-C of 32 mg/dl. CAC values over the 75th-percentile are considered abnormal.

Some limitations deserve mention. The association between *VPS13A* and RDW needs to be confirmed in an independent study. Furthermore, sequencing will be required for replication; none of the variants driving the novel gene-based association were available on the widely-used exome genotyping array. The few results passing our pre-specified significance level could be explained by statistical power given our sample size and the limited number of observed null alleles per gene. We also note that we have used a stringent significance threshold given the multiple testing burden inherent in our study design.

In conclusion, a limited number of null/damaging alleles with a large effect on cardiovascular traits were detectable from the exome sequences of 3,000 African American individuals.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Abifadel M, Varret M, Rabes JP, Allard D, Ouguerram K, Devillers M, et al. Mutations in pcsk9 cause autosomal dominant hypercholesterolemia. Nat Genet. 2003; 34:154–156. [PubMed: 12730697]

2. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low ldl cholesterol in individuals of african descent resulting from frequent nonsense mutations in pcsk9. Nat Genet. 2005; 37:161–165. [PubMed: 15654334]

3. Shioji K, Mannami T, Kokubo Y, Inamoto N, Takagi S, Goto Y, et al. Genetic variants in pcsk9 affect the cholesterol level in japanese. J Hum Genet. 2004; 49:109–114. [PubMed: 14727156]

4. Chen SN, Ballantyne CM, Gotto AM Jr, Tan Y, Willerson JT, Marian AJ. A common pcsk9 haplotype, encompassing the e670g coding single nucleotide polymorphism, is a novel genetic marker for plasma low-density lipoprotein cholesterol levels and severity of coronary atherosclerosis. J Am Coll Cardiol. 2005; 45:1611–1619. [PubMed: 15893176]

5. Cohen JC, Boerwinkle E, Mosley TH Jr, Hobbs HH. Sequence variations in pcsk9, low ldl, and protection against coronary heart disease. N Engl J Med. 2006; 354:1264–1272. [PubMed: 16554528]

6. Kathiresan S. A pcsk9 missense variant associated with a reduced risk of early-onset myocardial infarction. N Engl J Med. 2008; 358:2299–2300. [PubMed: 18499582]

7. Stein EA, Mellis S, Yancopoulos GD, Stahl N, Logan D, Smith WB, et al. Effect of a monoclonal antibody to pcsk9 on ldl cholesterol. N Engl J Med. 2012; 366:1108–1118. [PubMed: 22435370]

8. Robinson JG, Kastelein JJ. Pcsk9 inhibitors and cardiovascular events. N Engl J Med. 2015; 373:774. [PubMed: 26287854]

9. Sabatine MS, Wasserman SM, Stein EA. Pcsk9 inhibitors and cardiovascular events. N Engl J Med. 2015; 373:774–775. [PubMed: 26295078]

10. Cohen JC, Hobbs HH. Genetics. Simple genetics for a complex disease. Science. 2013; 340:689–690. [PubMed: 23661745]

11. Fuqua SR, Wyatt SB, Andrew ME, Sarpong DF, Henderson FR, Cunningham MF, Taylor HA Jr. Recruiting african-american research participation in the jackson heart study: Methods, response rates, and sample description. Ethn Dis. 2005; 15:S6-18-29. [PubMed: 16317982]

12. Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet. 2012; 91:839–848. [PubMed: 23103226]

13. 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. [PubMed: 17701901]

14. Liu X, Jian X, Boerwinkle E. Dbnsfp v2.0: A database of human non-synonymous snvs and their functional predictions and annotations. Hum Mutat. 2013; 34:E2393–E2402. [PubMed: 23843252]

15. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature. 2014; 506:185–190. [PubMed: 24463508]

16. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res. 2009; 19:1553–1561. [PubMed: 19602639]

17. Schwarz JM, Cooper DN, Schuelke M, Seelow D. Mutationtaster2: Mutation prediction for the deep-sequencing age. Nat Methods. 2014; 11:361–362. [PubMed: 24681721]

18. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7:248–249. [PubMed: 20354512]

19. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the sift algorithm. Nat Protoc. 2009; 4:1073–1081. [PubMed: 19561590]

20. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: Application to cancer genomics. Nucleic Acids Res. 2011; 39:e118. [PubMed: 21727090]

21. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden markov models. Hum Mutat. 2013; 34:57–65. [PubMed: 23033316]

22. Cui JS, Hopper JL, Harrap SB. Antihypertensive treatments obscure familial contributions to blood pressure variation. Hypertension. 2003; 41:207–210. [PubMed: 12574083]

23. Peloso GM, Auer PL, Bis JC, Voorman A, Morrison AC, Stitziel NO, et al. Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. Am J Hum Genet. 2014; 94:223–232. [PubMed: 24507774]

24. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010; 42:348–354. [PubMed: 20208533]

25. McClelland RL, Chung H, Detrano R, Post W, Kronmal RA. Distribution of coronary artery calcium by race, gender, and age: Results from the multi-ethnic study of atherosclerosis (mesa). Circulation. 2006; 113:30–37. [PubMed: 16365194]

26. Auer PL, Johnsen JM, Johnson AD, Logsdon BA, Lange LA, Nalls MA, et al. Imputation of exome sequence variants into population- based samples and blood-cell-trait-associated loci in african americans: Nhlbi go exome sequencing project. Am J Hum Genet. 2012; 91:794–808. [PubMed: 23103231]

27. Li AH, Morrison AC, Kovar C, Cupples LA, Brody JA, Polfus LM, et al. Analysis of loss-of-function variants and 20 risk factor phenotypes in 8,554 individuals identifies loci influencing chronic disease. Nat Genet. 2015; 47:640–642. [PubMed: 25915599]

28. Pollin TI, Damcott CM, Shen H, Ott SH, Shelton J, Horenstein RB, et al. A null mutation in human apoc3 confers a favorable plasma lipid profile and apparent cardioprotection. Science. 2008; 322:1702–1705. [PubMed: 19074352]

29. Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitziel NO, et al. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute. Loss-of-function mutations in apoc3, triglycerides, and coronary disease. N Engl J Med. 2014; 371:22–31. [PubMed: 24941081]

30. Jorgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjaerg-Hansen A. Loss-of-function mutations in apoc3 and risk of ischemic vascular disease. N Engl J Med. 2014; 371:32–41. [PubMed: 24941082]

31. Dobson-Stone C, Danek A, Rampoldi L, Hardie RJ, Chalmers RM, Wood NW, et al. Mutational spectrum of the chac gene in patients with chorea-acanthocytosis. Eur J Hum Genet. 2002; 10:773–781. [PubMed: 12404112]

32. Johansen CT, Wang J, Lanktree MB, Cao H, McIntyre AD, Ban MR, et al. Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. Nat Genet. 2010; 42:684–687. [PubMed: 20657596]

33. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466:707–713. [PubMed: 20686565]

34. Gloyn AL, McCarthy MI. Variation across the allele frequency spectrum. Nat Genet. 2010; 42:648–650. [PubMed: 20664646]

35. Hooper AJ, Marais AD, Tanyanyiwa DM, Burnett JR. The c679x mutation in pcsk9 is present and lowers blood cholesterol in a southern african population. Atherosclerosis. 2007; 193:445–448. [PubMed: 16989838]

36. Zhao Z, Tuakli-Wosornu Y, Lagace TA, Kinch L, Grishin NV, Horton JD, et al. Molecular characterization of loss-of-function mutations in pcsk9 and identification of a compound heterozygote. Am J Hum Genet. 2006; 79:514–523. [PubMed: 16909389]

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Clinical Perspective

The correlation of null alleles with human phenotypes can provide insight into gene function in humans. Here, we performed whole exome sequencing in 3,223 African American individuals living in Jackson, Mississippi in order to identify null and damaging missense variants and test these variants for association with 36 cardiovascular traits. We replicated the association of null and damaging missense variants in *PCSK9* with LDL cholesterol and found three individuals in their 50s each compound heterozygous for *PCSK9*. Of note, one of these three individuals had a coronary artery calcification score in the 83rd-percentile despite a LDL-C of 32 mg/dl. We also found individuals with rare damaging missense variants in *VPS13A* had higher red blood cell distribution width compared with non-carriers. Mutations in *VPS13A* have been previously reported to cause chorea-acanthocytosis, an autosomal recessive neurodegenerative disorder that causes red blood cells to appear spiky. Only a limited number of null/damaging alleles with a large effect on cardiovascular traits were detectable in ~3,000 African American individuals.
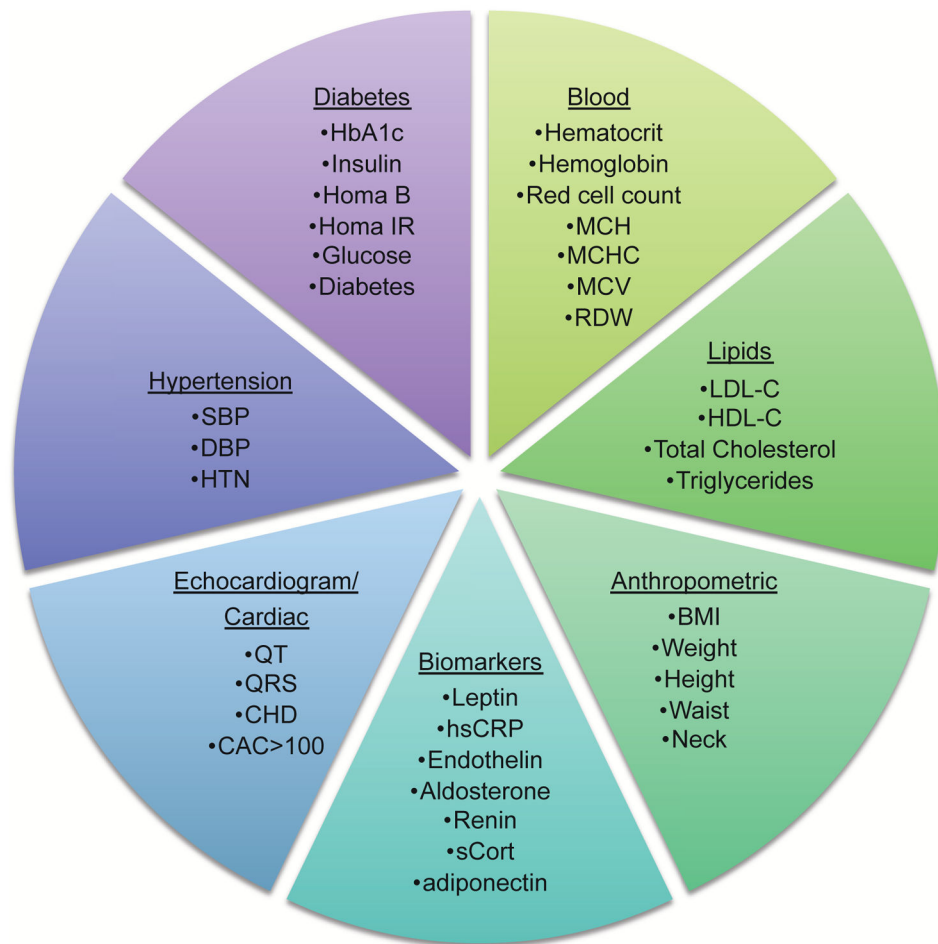
Figure 1.
Cardiovascular traits tested for association with null and damaging missense variants. 36 cardiovascular traits were tested for association with null and damaging missense variants in 3,223 African American individuals from the Jackson Heart Study.

Table 1

Descriptive statistics of Jackson Heart Study participants with exome sequences. Statistic provided as mean ±standard deviation for continuous variables and n (percent) for categorical variables.

Trait	N	Statistic
Demographic		
Female	3223	1211 (37.6%)
Age (years)	3223	55.59±12.82
Current Smoking status	3195	428 (13.4%)
Anthropometrics		
Body Mass Index [BMI] (kg/m^2)	3216	31.99±7.37
Weight (kg)	3218	91.37±21.71
Height (cm)	3218	169.06±9.25
Waist Circumference (cm)	3216	101.36±16.26
Neck Circumference (cm)	3219	38.72±3.76
Hypertension		
Hypertension, Yes [HTN]	3223	2012 (62.4%)
Systolic Blood Pressure [SBP] (mmHg) [*]	3217	132.11±19.87
Diastolic Blood Pressure [DBP] (mmHg) [*]	3217	81.50±10.80
Anti-hypertensive treatment	2619	1655 (63.2%)
Lipids		
LDL-C [†] (mg/dl)	2950	131.8±39.29
HDL-C (mg/dl)	2980	51.58±14.76
Triglycerides (mg/dl)	2979	107.61±82.77
Total Cholesterol [†] (mg/dl)	2395	206.08±43.44
Lipid-Lowering Treatment	2619	367 (14%)
Coronary Heart Disease		
Coronary Heart Disease Status [CHD]	3223	251 (7.8%)
Coronary Artery Calcium Score [CAC] (Agatston units)	1795	176.93±550.8
CAC>0	1795	882 (49.1%)
CAC>100	1795	439 (24.5%)
Diabetes		
Diabetic Status	3220	745 (23.1%)
Fasting Insulin (Plasma IU/mL) [‡]	2388	15.88±9.22
HOMA-B [‡] (mmol/l)	2357	215.75±107.75
HOMA-IR [‡] (mmol/l)	2386	3.59±2.29

Trait	N	Statistic
Fasting Plasma Glucose Level (mg/dL)[‡]	2390	90.53±8.97
Hemoglobin HbA1c (%)[‡]	2429	5.51±0.47
Biomarkers		
Leptin (Serum ng/mL)	3198	28.39±23.98
High Sensitivity C-Reactive Protein [hsCRP] (Serum mg/dL)	3214	0.53±1
Endothelin-1 (Serum pg/mL)	3214	1.34±0.6
Aldosterone (Serum ng/dL)	3213	5.81±4.92
Renin Activity RIA (Plasma ng/mL/hr)	1509	1.72±6.45
Cortisol Levels (Serum ug/dL)	3213	9.87±4.13
Adiponectin (Plasma ng/mL)	3166	5345.18±4236.78
Electrocardiogram		
QT Interval (msec)	3008	413.34±30.74
QRS Interval (msec)	2802	92.08±9.95
Blood		
Hematocrit level (%)	3110	39.27±4.2
Hemoglobin (g/dl)	3109	13.04±1.48
Mean corpuscular hemoglobin [MCH] (pg)	2781	28.88±2.51
Mean corpuscular hemoglobin concentration [MCHC] (%)	2781	33.16±0.91
Mean corpuscular volume [MCV] (fL)	2781	86.97±6.41
Red blood cell distribution width [RDW] (%)	2780	13.70±1.38
Red cell count (m/cmm)	2781	4.53±0.51

[*] Values were adjusted for individuals on blood pressure lowering medication.

[†] Values were adjusted for individuals on lipid lowering medication.

[‡] Values for non-diabetic individuals

Table 2

Genes with rare variant association signals meeting a significance threshold $< 1.1 \times 10^{-7}$ in the Jackson Heart Study

Outcome (units)	Gene	Chr	Best Test*		# variant sites	MAC	% carriers	Beta±SE	P-value
LDL (mg/dl)	PCSK9	1	Null+	6/7 damaging missense	7	119	3.9%	−35.8±3.8	2.9×10^{-21}
Mean corpuscular hemoglobin (pg/cell)	HBQ1	16	Null+	5/7 damaging missense	1	88	3.1%	−2.0±0.3	8.9×10^{-13}
Red blood cell distribution width (%)	VPS13A	9	Null+	5/7 damaging missense	9	34	1.2%	1.3±0.2	7.1×10^{-8}

*
Best Test indicates the group of variants that provided the most significant results for the gene. Null variants are defined as nonsense, splice-site, and frameshift variants. Damaging missense variants were classified according to the following 7 *in silico* prediction algorithms: LRT, Mutation Taster, PolyPhen2 (HumDiv), PolyPhen2 (HumVar), SIFT, MutationAssessor and FATHMM Chr, chromosome; # variant sites, number of sites going into the gene-based test; MAC (minor allele count), number of minor alleles across the variant sites; % Carriers, percent of individuals carrying a null or damaging missense variant tested; Beta, SE (standard error), p-value for association.

Table 3

Phenotypes of compound heterozygotes for *PCSK9* null mutations.

Trait	Individual 1	Individual 2	Individual 3
Demographic			
Sex	Female	Female	Male
Age (years)	50	50	52
Current Smoker	Yes	No	Yes
Anthropometrics			
Body Mass Index [BMI] (kg/m^2)	23.7	36.7	28.3
Weight (kg)	60.6	95.2	84.6
Height (cm)	160	161	173
Waist Circumference (cm)	82	118	96
Neck Circumference (cm)	34	36	40
Hypertension			
Hypertension, Yes [HTN]	No	Yes	Yes
Systolic Blood Pressure [SBP] (mmHg) [*]	96	140	166
Diastolic Blood Pressure [DBP] (mmHg) [*]	64	84	105
Anti-hypertensive treatment	No	Yes	No
Lipids			
LDL-C (mg/dl)	71.6	64.2	32
HDL-C (mg/dl)	63	98	41
Triglycerides (mg/dl)	97	34	142
Total Cholesterol (mg/dl)	154	169	101
Lipid-lowering Treatment	No	No	Not reported
Coronary Heart Disease			
Coronary Heart Disease Status [CHD]	No	No	No
Coronary Artery Calcium Score [CAC]	2.7	0	24.9
CAC percentile	87th	0th	83rd
Diabetes			
Diabetic Status	Yes	No	No
Fasting Insulin (Plasma IU/mL)	17	12	13
HOMA-B (mmol/l)	194.4	200.6	139.9
HOMA-IR (mmol/l)	4.0	2.5	3.1
Fasting Plasma Glucose Level (mg/dL)	95	85	97
Hemoglobin HbA1c (%)	6.3	5.4	5.4
Biomarkers			

Trait	Individual 1	Individual 2	Individual 3
Leptin (Serum ng/mL)	13.8	55.3	12.9
High Sensitivity C-Reactive Protein [hsCRP] (Serum mg/dL)	0.1	6.4	0.2
Endothelin-1 (Serum pg/mL)	1.4	1.5	3
Aldosterone (Serum ng/dL)	4.6	1.9	5.6
Renin Activity RIA (Plasma ng/mL/hr)	1.3	1.4	0.3
Cortisol Levels (Serum ug/dL)	6.2	12.2	8.8
Adiponectin (Plasma ng/mL)	2585.5	4847.6	3184.3
Electrocardiogram			
QT Interval (msec)	366	362	380
QRS Interval (msec)	84	102	114
Blood			
Hematocrit level (%)	41.1	38.7	47.8
Hemoglobin (g/dl)	13.8	13.2	15.3
Mean corpuscular hemoglobin [MCH] (pg)	30.6	30.3	27
Mean corpuscular hemoglobin concentration [MCHC] (%)	33.7	34.2	32
Mean corpuscular volume [MCV] (fL)	90.8	88.5	84.4
Red blood cell distribution width [RDW] (%)	13.2	14.3	13.4
Red cell count (m/cmm)	4.5	4.4	5.7

*
Values were adjusted for individuals on blood pressure lowering medication.