In the past few years, genome-wide association studies (GWAS) have successfully identified associations of common genetic variations with a variety of diseases and health-related quantitative traits. However, in most cases, neither the gene underlying disease susceptibility nor the spectrum of candidate functional variants has been identified. Within a genomic locus identified by GWAS, detailed examination of all genetic variants is required to discover causal variant(s), to estimate their impact on disease susceptibility, and to identify their functional roles. The large number of low-frequency and rare variants that exist within any given GWAS locus vastly outnumber common variants and may contribute significantly to the genetic architecture of disease. With the advent of genome sequencing using next-generation technologies, targeted sequencing can be performed at high throughput to identify lower frequency variants within regions identified by GWAS associations. Targeted sequencing of protein-coding genes identified by GWAS has been demonstrated to identify a large excess burden of rare functional alleles in people at extreme ends of...
quantitative traits, such as level of circulating triglycerides. However, many GWAS signals have been located in introns or flanking regions of protein-coding genes and are poorly correlated with functional variants in protein-coding genes, and ≥40% of GWAS signals are located in genomic regions uncorrelated with known missense variants, suggesting that most GWAS signals are regulatory in nature. Targeted sequencing of implicated genomic regions beyond exons may identify functional alleles involved in gene regulation. One emerging feature of GWAS is the existence of multiple apparently pleiotropic regions that underlie several different disease phenotypes, and targeted sequencing may aid in defining the genetic architecture of such regions.

The CHARGE Targeted Sequencing Study aims to follow up GWAS signals to comprehensively localize the functional variants and to evaluate the contribution of rare variants to a wide array of cardiovascular-related traits. A total of 77 genomic loci previously implicated by GWAS were selected and sequenced in participants from 3 CHARGE cohorts: the Atherosclerosis Risk in Communities Study (ARIC), the Cardiovascular Health Study (CHS), and the Framingham Heart Study (FHS). Here, we summarize the study design and the bioinformatic and statistical analysis strategies used in the CHARGE Targeted Sequencing Study.

Methods

Study Design

The CHARGE Targeted Sequencing Study used a case–cohort study design, in which a random sample was selected from all 3 cohorts at baseline. We planned for the cohort random sample to include ≥2000 individuals: 1000 participants from ARIC, 500 participants from CHS, and 500 participants from FHS, with proportions from each study reflecting relative cohort sizes with equal numbers of men and women. In addition to the cohort random sample, ≥200 participants (generally 100 from ARIC, 50 from CHS, and 50 from FHS) from each of 14 key phenotypes were selected for sequencing on the basis of either case status for discrete phenotypes or extreme values of quantitative traits.

The phenotypes studied (Table 1) were atrial fibrillation, blood pressure, body mass index, bone mineral density, C-reactive protein, carotid intima-media thickness, echocardiography, electrocardiographic PR and QRS interval, fasting insulin, hematocrit, pulmonary function, retinal venule diameter, and stroke. Because individuals initially selected for the cohort random sample or some phenotype groups could satisfy the criteria for the extreme sampling of ≥1 phenotype group, the achieved number with extreme values for each phenotype was often larger than the target number of 200. Detailed information on the criteria for the selection of study participants for each phenotype is provided in the Materials section in the Data Supplement.

Table 1. Phenotype Groups and Sample Selection Strategies

<table>
<thead>
<tr>
<th>Phenotype Groups</th>
<th>Strategy</th>
<th>FHS</th>
<th>CHS</th>
<th>ARIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECG PR interval</td>
<td>High residual†</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>ECG QRS interval</td>
<td>High</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Stroke</td>
<td>Ischemic stroke</td>
<td>50</td>
<td>70</td>
<td>80</td>
</tr>
<tr>
<td>Blood pressure</td>
<td>Low residual†</td>
<td>25</td>
<td>25</td>
<td>54</td>
</tr>
<tr>
<td>Body mass index</td>
<td>High residual†</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Fasting insulin</td>
<td>High</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Bone mineral density by DEXA</td>
<td>Low z-score</td>
<td>100</td>
<td>100</td>
<td>…</td>
</tr>
<tr>
<td>Left ventricular diastolic diameter</td>
<td>High residual†</td>
<td>100</td>
<td>100</td>
<td>…</td>
</tr>
<tr>
<td>C-reactive protein</td>
<td>High residual†</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Hematocrit</td>
<td>Low residual†</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Retinal venule diameter</td>
<td>High residual†</td>
<td>34</td>
<td>166</td>
<td></td>
</tr>
<tr>
<td>Carotid wall thickness</td>
<td>High</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Pulmonary: FEV1/FVC</td>
<td>Low</td>
<td>…</td>
<td>…</td>
<td>200</td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>Lone atrial fibrillation†</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

ARIC indicates Atherosclerosis Risk in Communities Study; CHS, Cardiovascular Health Study; DEXA, dual-energy X-ray absorptiometry; FEV1, forced expiratory volume in 1 second; FHS, Framingham Heart Study; and FVC, forced vital capacity.

*These numbers represent the number of participants with extreme phenotypes targeted for selection by each phenotype group but do not reflect the additional participants who may have met the criteria for an extreme phenotype but had already been selected by other phenotype groups or as part of the cohort random sample.

†Extreme samples were selected by taking either the extreme high or low distribution based on age, sex, and phenotype-specific variable-adjusted residuals.

‡200 cases with lone atrial fibrillation were selected from Massachusetts General Hospital.
region or locus containing one to many genes having displayed strong associations ($P<5\times10^{-8}$) with ≥2 traits in multiple GWAS (Table 3).

### Library Preparation, Sequencing, and Variant Calling

Detailed description of library preparation and sequencing is found in the Materials section in the Data Supplement. In brief, the targeted regions were captured by a specific SOLiD platform–based multiplexed capture sequencing protocol developed at the Baylor College of Medicine Human Genome Sequencing Center. The enriched libraries were then pooled to form an 8-sample pool for multiplexed sequencing. Each sequencing pool was subsequently sequenced on a quadrant of a SOLiD V4 slide using Life Technologies’ Barcode Fragment Sequencing Kits and methods.

### Quality Control

Because data from sequencing experiments can have errors at multiple levels, such as variant calls and read mapping, we implemented...
a multilevel approach to identify sites with true variation for use in downstream association analyses. All quality control (QC) procedures were performed in the statistical platform R or Java, in combination with SAMtools.11

Preliminary QC Procedures in Sequencing Laboratory

The first level of QC took place through laboratory procedures. After sequencing a sample to the target depth, we evaluated several QC metrics, including alignment rate and uniqueness, to validate that the sequencing performed as expected. Base and quality calling for the SOLiD data was performed on-instrument using standard vendor software and settings. To gauge the overall performance of the capture process, sample-level BAMs were also subjected to a capture analysis QC pipeline to obtain additional metrics, such as the proportion of the aligned reads that mapped to the targeted region and the proportion of targeted bases at various coverage levels. Samples that met a minimum of 65% of the targeted bases at 20× or greater coverage were submitted for subsequent analysis and QC.

For each successfully sequenced sample, we confirmed sample identity and checked purity by using the ERIS tool suite (https://github.com/dsextion2/ERIS) to compare sequence data with genotypes from available GWA SNP arrays. Using an e-GenoTyping approach, we screened all sequence reads for exact matches to probe sequences defined by the variant and position of interest, along with 11 bases of sequence flanking either side of the SNP site. In this process, we removed SNP array sites that were nonspecific and over- or under-covered before comparing the read data with the variants for all samples in the project. Based on our previous empirical experience, we used thresholds of 90% self-concordance and next-best matches <75% to identify samples that demonstrated minimal contamination and confirmed sample identity. We informatically unsampled any samples with clear evidence of mislabeling by attaching the appropriate sample names. Any samples that seemed to be either cryptically swapped or significantly contaminated were resequenced and rescreened for inclusion in the study.

Variant-Level QC

Each cohort individually implemented an extensive QC pipeline for all of their own samples that passed the laboratory QC procedures. Our QC pipeline consisted of a series of variant-level filtering steps followed by QC on individual samples (summarized in Table 4). Before applying these steps, we first prefiltered the raw data to remove any variants that mapped >100 bp from the requested target capture region. We further removed potentially low-quality reads by filtering variants with a Phred-scaled base quality score12 (−10 log10 P, where P is the probability of calling error) <30, with <2 reads of the alternate alleles, and variants with a depth of coverage of <10 total reads.

At the sample-SNP filtering stage, we assessed each variant within each sample in terms of allelic imbalance and strand bias. Heterozygote genotypes were removed if their alternate allele ratio was disproportionate, defined to be <0.2 or >0.8 for 1 allele. We did not take into account copy number variations (Materials section in the Data Supplement). For strand bias, we kept only variants with alternate allele reads obtained from both the positive and negative strands.

Finally, each variant was evaluated across all samples. We removed SNPs that had >20% missingness, >2 observed alleles, or were part of an overly dense SNP cluster (≥3 variants in a 10-bp window) because several variants within a short genomic interval can indicate regional sequencing errors. Then, using only samples from the cohort random sample, we filtered SNPs that deviated from the expectations of Hardy–Weinberg equilibrium (P<1×10−5) to identify excess heterozygosity that may have been induced by mismapped reads.

Table 3. Targets for Pleiotropic Loci*  

<table>
<thead>
<tr>
<th>Locus</th>
<th>Genes (5′→3′)†</th>
<th>Start Position‡</th>
<th>Stop Position‡</th>
<th>No. of SNPs</th>
<th>No. of Coding SNPs</th>
<th>No. of Novel SNPs§</th>
</tr>
</thead>
<tbody>
<tr>
<td>6q23.3</td>
<td>HB51L, MYB</td>
<td>135 322 113</td>
<td>135 582 124</td>
<td>1894</td>
<td>282</td>
<td>1486</td>
</tr>
<tr>
<td>7q22.3</td>
<td>PKC3G</td>
<td>105 917 292</td>
<td>106 379 327</td>
<td>766</td>
<td>159</td>
<td>539</td>
</tr>
<tr>
<td>7q36.1</td>
<td>PKRAG2</td>
<td>150 856 537</td>
<td>151 267 715</td>
<td>1022</td>
<td>83</td>
<td>664</td>
</tr>
<tr>
<td>8p21.1</td>
<td>SCARA5</td>
<td>27 783 553</td>
<td>27 906 232</td>
<td>631</td>
<td>108</td>
<td>436</td>
</tr>
<tr>
<td>11p11.2</td>
<td>DGKZ, AMBRA1, ATG13, ARHGAP1, ZNF408, F2, CKAP5, LR4P, C11ORF49, DD1, ACP2, NRI1H3, MADD, MYBPC3, SP11, SLC39A13, PSMC3, RAPSN, CELF1, NDUFS3, MITCH2, AGBL2, FNBP4, NUP160</td>
<td>46 308 577</td>
<td>47 851 121</td>
<td>3059</td>
<td>583</td>
<td>2309</td>
</tr>
<tr>
<td>12q24.12–13</td>
<td>ATXN2, BRAP, ACAD10, ALDH2, MAPKAPK5, ADAM1, TMEM116, NAA25, TRAFD1, C12ORF51, RPL6, PTPN11</td>
<td>110 374 301</td>
<td>111 436 622</td>
<td>3818</td>
<td>989</td>
<td>3130</td>
</tr>
<tr>
<td>13q34</td>
<td>COL4A1, COL4A2</td>
<td>109 599 195</td>
<td>109 967 539</td>
<td>1043</td>
<td>330</td>
<td>712</td>
</tr>
</tbody>
</table>

Table 4. SNP Quality Filters

<table>
<thead>
<tr>
<th>Stages</th>
<th>SNPs Should Satisfy All the Following Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefilter</td>
<td>Off-target distance</td>
</tr>
<tr>
<td></td>
<td>Phred score</td>
</tr>
<tr>
<td></td>
<td>Depth of coverage</td>
</tr>
<tr>
<td></td>
<td>Depth of alternate allele coverage</td>
</tr>
<tr>
<td>Sample SNP</td>
<td>Allelic imbalance</td>
</tr>
<tr>
<td></td>
<td>Strand bias</td>
</tr>
<tr>
<td>Whole SNP</td>
<td>% Missing</td>
</tr>
<tr>
<td></td>
<td>HWE exact test*</td>
</tr>
<tr>
<td></td>
<td>No. of Alleles</td>
</tr>
<tr>
<td></td>
<td>SNP cluster</td>
</tr>
</tbody>
</table>

HWE indicates Hardy–Weinberg equilibrium; and SNP, single-nucleotide polymorphisms.

*HWE was calculated based on cohort random sample only.
Sample-Level QC

After variant-level QC was completed, each cohort performed a quality assessment of the final sequence data based on several measures. Within each cohort, a sample was flagged as potentially poor quality if it fell below the lower or upper 2.5th percentile of any of 8 selected measures: mean mapping quality score across all variants; mean fold coverage; mean transition to transversion ratios; mean heterozygote to homozygote ratio; mean nonsynonymous to synonymous ratio; number of singletons; number of doubletons; and percentage of sites with coverage $\geq 20x$. However, none of samples showed systematically low quality. We therefore kept all the sequenced samples but recorded these quality metrics in a joint sample information file. Phenotype groups, however, could further examine these samples and decide whether to remove some of them in their respective association analyses.

SNP Information and Functional Annotation

An SNP information file combining information across the 3 cohorts and all sequence data was produced after QC, including summaries and functional annotations for the SNPs. The summaries included the SNP position, reference and alternative alleles, sample size, genotype counts, allele counts, allele frequencies, average mapping quality, average SNP calling quality, lower 2.5 and upper 97.5 percentiles of read depths, genotype missing rate, and minimum $P$ value of the Hardy–Weinberg equilibrium test within the cohort random sample. Functional annotations were produced using a combination of ANNOVAR,$^{23}$ dbNSFP,$^{14}$ and custom internal tools. SNP positions referring to the RefSeq15 gene and custom internal tools. SNP positions referring to the RefSeq15 gene and the ORegAnno database 23 and the TRANSFAC database 24 accessed through the UCSC Genome Browser.25 We recommended that phenotype groups take into account various types of supporting evidence in the interpretation of association results.

Statistical Analysis

Common Variants

The CHARGE Analysis and Bioinformatics Committee recommended performing single marker analyses for each common variant within a target. Although individual phenotype groups implemented this threshold differently, common variants were loosely defined as those with allelic frequency $\leq 2\%$, which corresponded to variants where there were $\geq 20$ individuals with 1 or 2 minor alleles across the entire study.

We performed 2 regression analyses: an unweighted analysis to obtain $P$ values for association and a weighted analysis to obtain effect estimates and estimated standard errors. The weighted analysis accounted for the sampling design by assigning different weights to extreme samples and to individuals from the cohort random sample. Extreme samples were weighted by 1, whereas individuals of cohort random sample were weighted by the inverse of their probability of inclusion in each cohort. More details of the sampling weight are described by T. Lumley, et al, http://stattech.wordpress.fos.auckland.ac.nz/files/2012/05/design-paper.pdf. For both analyses, we used data from all subjects. To produce $P$ values for association between each variant and the phenotype of interest, we used standard regression methods: linear regression or linear mixed-effects models (FHS) for dichotomous outcomes, and Cox proportional hazards regression with robust variance or Cox proportional hazards regression with robust variance or Cox proportional hazards regression (with clustering on pedigrees with robust variance in FHS)$^{26,27}$ for survival outcomes. The different models used in FHS aimed to address relatedness in FHS subjects. Because these analyses were intended to follow up on GWAS loci, working groups typically used the same phenotype definition, adjustment variables, and additive genetic models (0/1/2 copies) as in the discovery GWAS analyses. Results from each study (estimated regression coefficient [$\hat{\beta}$ and estimated standard error] were then shared and combined, applying inverse-variance–weighted fixed-effects meta-analysis. $P$ values from this meta-analysis were reported. Because of our sampling scheme, we reported the corresponding meta-analytic estimate of effect ($\hat{\beta}$-hat) from the weighted analysis and $P$ values from the unweighted analysis. Each working group made their own decisions toward control of type I error. Some groups used an $\alpha$ cutoff according to their previous hypotheses and others used $\alpha > 1$ cutoff, depending on the focus of their investigation. All the analyses were performed using R software (www.r-project.org/).

Rare Variants

Single-marker–based association analysis generally has low power for rare variants. Therefore, several methods for rare variant tests have recently been developed. Basu and Pan$^{28}$ performed an extensive comparison of many of the currently available methods under different circumstances. For the CHARGE Targeted Sequencing Study, we recommended that working groups use analyses that either collapse variants in each genomic region using a burden test or jointly analyze associations with variants in each genomic region by using the sequence kernel association test (SKAT).

Collapsing Tests

The primary recommendation for analyses that collapse variants in a genomic region into a single summary measure was to use the T1 count, defined as the number of variants with $\geq 1$ rare allele among variants in the region with a study-wide minor allele frequency (MAF) $< 1\%$. A secondary recommendation was a Madsen–Browning type test, which aggregates all variants with MAF $< 1\%$ in a genomic region, weighting each variant by a function of its MAF.29 Although all variants in a region can be considered in the Madsen–Browning test statistic, we recommended restricting to rare variants with MAF $< 1\%$. For these methods that collapse variants, the same regression analyses described above for common variants were used, with the aggregate collapsed regional burden replacing the usual genotype dosage.

Joint Analysis of Variants

The recommendation for jointly analyzing variants in a genomic region was a specific version of a general score test available as the SKAT.30 The SKAT score can be written as a weighted sum of squares of $z$-statistics from score tests in single-variant regression models. These single-variant tests were computed in each study and meta-analyzed using standard methods to give the SKAT statistic using weights based on combined allele frequencies across all studies. The reference distribution for the SKAT requires the covariance matrix of the genetic variants, which was computed as a simple weighted average of the covariance matrices in the 3 cohorts. Each study implemented the SKAT analyses by using custom R scripts that included an SKAT extension to account for familial relatedness.31 The scripts are provided in the CHARGE wiki Web site (http://depts.washington.edu/chargeco/wiki/CHARGE-S). Simulations confirmed that this approach agrees closely with the SKAT performed on individual data, and that the power is higher than when the meta-analysis is performed on $P$ values (T. Lumley, et al, http://stattech.wordpress.fos.auckland.ac.nz/files/2012/11/skat-meta-paper.pdf).

Results

A total of 4646 samples were target captured and sequenced for the project. After applying initial sequencing QC for sample identity, contamination, and target coverage described above, 4440 samples qualified for additional analysis, providing a 95.5% capture sequencing and QC success rate. Data produced from all these samples are summarized in Figure 1. Individual samples from the 3 cohorts (ARIC, CHS, and FHS) plus 1 additional sample set (200 lone atrial fibrillation cases from Massachusetts General Hospital) are shown with the percentage coverage of the target bases at $20\times$ coverage in relation to the actual megabases generated. Across the 3 studied cohorts, $\approx 70\%$ to $80\%$ of short reads were successfully aligned to the
reference genome (hg18). We found that 40% to 45% of short reads were mapped to the target regions. After removal of duplicate and low-quality reads, ~21% of total aligned reads were kept for downstream analyses. On average, 82% of the targeted bases were covered at ≥20x, and the average coverage for each sample was ≥45x. Nearly all the targeted probe sets were successfully captured, and 95% to 96% of the targeted bases had ≥1 read for coverage. The number of targeted bases with a given depth of coverage closely followed a Poisson distribution, indicating uniform capture and sequencing of the targeted regions. After removing duplicate samples, a total of 4231 unique individuals from the 3 cohorts were used for downstream analysis, including 2003 from ARIC, 1132 from CHS, and 1096 from FHS. The cohort random sample included 1917 individuals, including 2003 from ARIC, 1132 from CHS, and 1096 from FHS. The cohort random sample was used for downstream analysis, indicating uniform capture and sequencing of the targeted regions. After removing duplicate samples, a total of 4231 unique individuals from the 3 cohorts were used for downstream analysis, including 2003 from ARIC, 1132 from CHS, and 1096 from FHS. The cohort random sample included 1917 individuals, and the remaining 2314 individuals were distributed across the 14 phenotype groups. Demographic characteristics of the investigated participants are presented in Table 5.

A total of 52736 variants were identified that passed QC among the 3 cohorts. This number included 30912 variants in ARIC, 21150 in CHS, and 21267 in FHS. Across all samples, the average mean transition to transversion ratio after SNP filtering was 2.44, in accordance with what would be expected given that the CHARGE targeted sequencing regions were a mixture of exonic, intronic, and intergenic regions. A cross-validation with previous genotype data showed a concordance rate of 98.0% (Materials section in the Data Supplement). The summary statistics of SNPs found in each individual are shown in Table I in the Data Supplement.

Figure 2 displays the distribution of functional classes and MAF combining filtered variants from all cohorts. The majority of variants were located within the intergenic (31.0%) or intronic regions (50.7%), and only 11.7% of variants were within known protein-coding regions. A total of 4800 (9.1%) were common variants (MAF ≥1%), and the remaining 47936 were rare variants. Overall, most (93%) common variants were observed in multiple cohorts, whereas rare variants were more likely to be unique to a single cohort. Of the common variants, 98% have already been reported in phase 1 of the 1000 Genomes Project,21 whereas only 15% of rare variants have been reported. Among the 4800 common variants identified in this project, only 2501 (52.1%) of them were available in the HapMap CEU panel, which was used for genotype imputation and thus GWAS. In particular, we identified 70 damaging variants (missense, nonsense, or splicing variants), of which only half were available in the HapMap CEU panel. As an example, 4 gene regions were selected for sequencing because of previous associations with circulating C-reactive protein levels.32 We found that 13 SNPs remained significant after adjusting for multiple testing, including a missense SNP rs2228145 within the \( \text{IL6R} \) locus (Materials section in the Data Supplement). The SNP was not studied in GWAS, but it was in linkage disequilibrium with the GWAS lead SNP (rs4129267) at the \( \text{IL6R} \) locus. Previous studies have found that rs2228145 was strongly associated with circulating concentrations of interleukin-6 soluble receptor.33,34 which is a proinflammatory cytokine regulating a variety of inflammatory responses.35,36 Our results suggest that rs2228145 might be the functional SNP, explaining the association of the \( \text{IL6R} \) locus with C-reactive protein levels.

**Discussion**

The objective of the CHARGE Targeted Sequencing Study was to localize the GWA signals and to evaluate the contribution of rare variants to 14 phenotypes. We implemented a case–cohort design.
study design, in which both a random sample of participants and participants with extreme trait values were selected from each of 3 participating cohorts. We also developed and implemented robust analysis strategies to analyze sequence data in relation to each individual phenotype. In addition, our sequencing project was able to accommodate different hypotheses proposed by phenotype groups relating to the target selection. For some targets (e.g., ZFHX3 and SCN5A), only exonic regions were sequenced, and for some other targets (e.g., PLN and SCN10A), the entire gene region was sequenced. Some targeted regions were even outside of any known gene regions (e.g., 2q36.3 and MEF2C), demonstrating the flexibility of our target selection. The full data set has been registered with dbGaP and will be deposited soon.

Our study design provides a cost-effective way to evaluate genetic associations for multiple phenotypes. The same cohort random sample was included in the analyses of all phenotypes, and thus sample sizes were larger than would be achieved with phenotype-specific analysis populations. In addition, analyses were typically performed across all available samples from the phenotype groups. That is, extreme samples chosen by one phenotype working group were used by others, significantly increasing the overall sample size and allowing more rare variants to be observed in each analysis. Because the phenotype group sampling was based on trait values, we applied a weighting approach so that the distributions of all variables would be the same as in the full cohort. Although testing can, in our circumstances, be performed without the sampling weights, they are needed for unbiased estimation of effects (T. Lumley, et al., http://stattech.wordpress.fos.auckland.ac.nz/files/2012/05/design-paper.pdf). Under plausible scenarios, for a single phenotype, the use of our design of the cohort random sample is less powerful than sampling extreme values from both tails, but for studying multiple phenotypes the repeated use of the cohort random sample provides greater power. An alternative sampling strategy that selected control subjects only from those participants without extreme values for any phenotype of interest might offer larger power if a small number of phenotypes were studied. Given that a small proportion of samples in this study had familiar relatedness, we have limited power to perform family cosegregation analysis of rare variants.

In summary, we sequenced and analyzed 77 genomic loci associated with various phenotypes as implicated in previous GWAS. A cost-effective case–cohort study design and robust analysis strategies were implemented to analyze sequence data.

**Sources of Funding**

Support for Building on genome-wide association studies (GWAS) for the National Heart, Lung, and Blood Institute (NHLBI) diseases: the US Cohorts for Heart and Aging Research in Genomic Epidemiology Targeted Sequencing Study (CHARGE) consortium was provided by the National Institutes of Health through the American Recovery and Reinvestment Act of 2009 (5RC2HL102419). Data for Building on GWAS for NHLBI-diseases: the US CHARGE consortium was provided by Eric Boerwinkle on behalf of the Atherosclerosis Risk in Communities (ARIC) Study, L.A. Cupples, principal investigator for the Framingham Heart Study, and Bruce Psaty, principal investigator for the Cardiovascular Health Study (CHS). Sequencing was carried out at the Baylor Genome Center (U54 HG003273). The ARIC Study is performed as a collaborative study supported by the NHLBI contracts (HHSN268201100005C, HHSN268201100006C,
Disclosures

B.M. Psaty serves on the Data and Safety Monitoring Board of a clinical trial of a device funded by Zoll LifeCor and on the Steering Committee of the Yale Open Data Access Project funded by Medtronic. The other authors report no conflicts.

Appendix

From the Department of Medicine, Boston University School of Medicine, MA (H.L., E.J.B.); The NHLBI’s Framingham Heart Study, MA (H.L., J.D., A.D.J., C.J.O.D., A.L.D.S., K.L.L., F., et al.); Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX (M.W., J.G.R., C.L.K., H.D., Y.W., I.N., D.M.M.); Cardiovascular Health Research Unit, Department of Medicine (J.A.B., J.C.B., C.M.S., S.R.H., B.M.P.), Department of Biostatistics (B.M.K., K.M.R.), Department of Epidemiology (S.R.H., B.M.P.), and Department of Health Services (B.M.P.), University of Washington, Seattle; Department of General and Interventional Cardiology, University Heart Center, Hamburg, Hamburg, Germany (R.B.S.); and Group Health Research Institute, Group Health Cooperative, Seattle, WA (S.R.H., B.M.P.).

References

Strategies to Design and Analyze Targeted Sequencing Data: Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium Targeted Sequencing Study
Honghuang Lin, Min Wang, Jennifer A. Brody, Joshua C. Bis, Josée Dupuis, Thomas Lumley, Barbara McKnight, Kenneth M. Rice, Colleen M. Sitlani, Jeffrey G. Reid, Jan Bressler, Xiaoming Liu, Brian C. Davis, Andrew D. Johnson, Christopher J. O’Donnell, Christie L. Kovar, Huyen Dinh, Yuanqing Wu, Irene Newsham, Han Chen, Andi Broka, Anita L. DeStefano, Mayetri Gupta, Kathryn L. Lunetta, Ching-Ti Liu, Charles C. White, Chuanhua Xing, Yanhua Zhou, Emelia J. Benjamin, Renate B. Schnabel, Susan R. Heckbert, Bruce M. Psaty, Donna M. Muzny, L. Adrienne Cupples, Alanna C. Morrison and Eric Boerwinkle

Circ Cardiovasc Genet. 2014;7:335-343
doi: 10.1161/CIRCGENETICS.113.000350
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/7/3/335

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Genetics can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Genetics is online at:
http://circgenetics.ahajournals.org/subscriptions/