Genome- and Phenome-Wide Analyses of Cardiac Conduction Identifies Markers of Arrhythmia Risk

Marylyn D. Ritchie, PhD*; Joshua C. Denny, MD, MS*; Rebecca L. Zuvich, PhD*; Dana C. Crawford, PhD; Jonathan S. Schildcrout, PhD; Lisa Bastarache, MS; Andrea H. Ramirez, MD; Jonathan D. Mosley, MD, PhD; Jill M. Pulley, MBA; Melissa A. Basford, MBA; Yuki Bradford, MS; Luke V. Rasmussen; Jyotishman Pathak, PhD; Christopher G. Chute, MD, DrPH; Iftikhar J. Kullo, MD; Catherine A. McCarty, PhD; Rex L. Chisholm, PhD; Abel N. Kho, MD, MS; Christopher S. Carlson, PhD; Eric B. Larson, MD, MPH; Gail P. Jarvik, MD, PhD; Nona Sotoodehnia, MD, MPH; on Behalf of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) QRS Group, Teri A. Manolio, MD, PhD; Rongling Li, PhD; Daniel R. Masys, MD; Jonathan L. Haines, PhD; Dan M. Roden, MD

Background—ECG QRS duration, a measure of cardiac intraventricular conduction, varies ∼2-fold in individuals without cardiac disease. Slow conduction may promote re-entrant arrhythmias.

Methods and Results—We performed a genome-wide association study to identify genomic markers of QRS duration in 5272 individuals without cardiac disease selected from electronic medical record algorithms at 5 sites in the Electronic Medical Records and Genomics (eMERGE) network. The most significant loci were evaluated within the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium QRS genome-wide association study meta-analysis. Twenty-three single-nucleotide polymorphisms in 5 loci, previously described by CHARGE, were replicated in the eMERGE samples; 18 single-nucleotide polymorphisms were in the chromosome 3 SCN5A and SCN10A loci, where the most significant single-nucleotide polymorphisms were rs1805126 in SCN5A with \( P = 1.2 \times 10^{-8} \) (eMERGE) and \( P = 2.5 \times 10^{-20} \) (CHARGE) and rs6795970 in SCN10A with \( P = 6 \times 10^{-6} \) (eMERGE) and \( P = 5 \times 10^{-7} \) (CHARGE). The other loci were in NFIA, near CDKN1A, and near C6orf204. We then performed phenotype-wide association studies on variants in these 5 loci in 13 859 European Americans to search for diagnoses associated with these markers. Phenome-wide association study identified atrial fibrillation and cardiac arrhythmias as the most common associated diagnoses with SCN10A and SCN5A variants. SCN10A variants were also associated with subsequent development of atrial fibrillation and arrhythmia in the original 5272 “heart-healthy” study population.

Conclusions—We conclude that DNA biobanks coupled to electronic medical records not only provide a platform for genome-wide association study but also may allow broad interrogation of the longitudinal incidence of disease associated with genetic variants. The phenotype-wide association study approach implicated sodium channel variants modulating QRS duration in subjects without cardiac disease as predictors of subsequent arrhythmias. (Circulation. 2013;127:1377-1385.)

Key Words: atrial fibrillation ■ electronic health records ■ genetics ■ genome-wide association study

Electrocardiographic parameters of cardiac conduction and repolarization, PR, QRS, and QT intervals, are widely used in clinical medicine and display substantial variability when measured across large populations. QRS duration represents activation time in the cardiac ventricle, and prolongation of this interval, representing global or regional slow...
conduction, has been associated with adverse outcomes such as sudden cardiac death. This variability reflects modulators such as abnormal electrolytes, underlying heart disease, or concomitant drug therapy; as well as heritable components; published estimates suggest that up to 40% of variability in the QRS interval is heritable. Using automated methods described further below, we have shown that 99% of QRS durations in >30,000 healthy subjects not receiving confounding medications fall between 65 and 108 milliseconds.

### Methods

#### Normal QRS Algorithm

We developed and deployed an algorithm to identify individuals with normal ECGs and without any cardiac disease, abnormal electrolyte values, or QRS-active medication across the 5 eMERGE-I sites and identified 5272 white patients (2488 male and 2784 female patients; Table 1). The algorithm was developed and validated in the Synthetic Derivative, a deidentified image of the Vanderbilt EMR that currently contains >120 million documents on ~2 million patients. The Synthetic Derivative is refreshed regularly to add new clinical information from the EMR as it is accrued.

The study population consisted of subjects with a normal ECG without evidence of cardiac disease any time before or within 1 month after the ECG, who were without concurrent use of medications that interfere with ventricular conduction, and who did not have abnormal potassium, calcium, or magnesium laboratory values at the time of the ECG. The algorithm has been described in detail previously. The algorithm used natural language processing to analyze narrative text, billing code queries, and laboratory queries to exclude any subjects with evidence of arrhythmia, heart failure, cardiomyopathy, myocardial ischemia/infarct, or cardiac conduction defect. The algorithm considered all physician-generated clinical documentation, including clinical notes and cardiologist-generated ECG impressions. Patients with family histories of cardiac disease were allowed by the natural language processing algorithm. In addition, ECGs had normal Bazett-corrected QT intervals (<450 milliseconds), heart rates (between 50 and 100 bpm), and QRS (60–120 milliseconds). The algorithm was reviewed by 2 physicians not involved in algorithm development and achieved a positive predictive value of 97% to identify patients with normal ECGs who did not have normal QRSs and who did not have normal QRSs and who did not have normal QRSs and who did not have normal QRSs and who did not have normal QRSs and who did not have normal QRSs. Analysis of clinical covariates in ~30,000 records with algorithm-defined normal ECGs identified sex and ancestry as modulators of QRS duration. Complete details of the algorithm are available from the Phenome-Wide Association Study (PheWAS) database (http://phekb.org). The algorithm was then deployed across the DNA repositories at the 4 other eMERGE-I sites (Marshall Clinic, Northwestern University, Mayo Clinic, and Group Health Research Institute) to identify subjects with extant eMERGE-based genotyping data (based on other phenotypes; Table 1) who met algorithm-defined criteria for normal QRS. The eMERGE cohorts are described in more detail by McCarthy et al and at the Phenotype Knowledge Base (http://phekb.org). Thus, all eMERGE individuals used in the analysis underwent the same algorithm to select those with normal ECGs and without prior heart disease, interfering medications, and abnormal electrolytes. To assess the performance of the algorithm when applied within external EMR systems, trained chart abstractors at Northwestern and Marshallfield performed random database selections of 100 subjects at Marshallfield and 45 subjects at Northwestern to determine the accuracy of the algorithm at external sites. The evaluation at Northwestern also included an independent review by a board-certified internal medicine physician, with discrepancies resolved by consensus. This study included only subjects designated as non-Hispanic white European Americans in the EMR from each site. We have previously shown that the EMR ancestry performs similar to self-report.

This study was approved by the Institutional Review Board at each site. Because BioVU is deidentified and accurses individuals through leftover blood remaining after routine clinical testing, it operates as nonhuman subjects research according to the provisions of Code of Federal Regulations Title 45 Part 46.

### Table 1. eMERGE Sites Contributing Samples

<table>
<thead>
<tr>
<th>Site</th>
<th>Primary Site Phenotype</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Health</td>
<td>Dementia</td>
<td>187</td>
<td>351</td>
<td>538</td>
</tr>
<tr>
<td>Marshallfield</td>
<td>Cataract</td>
<td>69</td>
<td>149</td>
<td>218</td>
</tr>
<tr>
<td>Mayo Clinic</td>
<td>Peripheral artery disease</td>
<td>1056</td>
<td>730</td>
<td>1786</td>
</tr>
<tr>
<td>Northwestern</td>
<td>Type 2 diabetes mellitus</td>
<td>99</td>
<td>118</td>
<td>217</td>
</tr>
<tr>
<td>Vanderbilt</td>
<td>Normal QRS</td>
<td>1077</td>
<td>1436</td>
<td>2513</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2488</td>
<td>2784</td>
<td>5272</td>
</tr>
</tbody>
</table>

All patients were of European ancestry.
Genotyping and Data Analysis

Genotyping was performed at the Center for Genotyping and Analysis at the Broad Institute and the Center for Inherited Disease Research at Johns Hopkins University. Samples of European ancestry or unknown ancestry were analyzed with the Illumina Human660W-Quadv1-A genotyping platform, consisting of 561 490 single-nucleotide polymorphisms (SNPs) and 95 876 intensity-only probes. Data were cleaned using the quality control (QC) pipeline developed by the eMERGE Genomics Working Group. This process includes evaluation of sample and marker call rate, sex mismatch and anomalies, duplicate and HapMap concordance, batch effects, Hardy-Weinberg equilibrium, sample relatedness, and population stratification. After QC, 528 508 SNPs were used for analysis on the basis of the following QC criteria: SNP call rate >99%, sample call rate >99%, minor allele frequency >0.001, unrelated samples only (removing all parent-offspring, full and half-siblings), and individuals of European descent only (based on STRUCTURE22 analysis of >90% probability of being in the Centre d’Etude du Polymporphism Humain Utah residents with ancestry from northern and western Europe [CEU] cluster). Each eMERGE site used the QC pipeline to clean its initial data set before all the samples were merged. QC procedures were then performed on the merged eMERGE data set in which data from all 5 sites were combined, and no significant differences across sites or genotyping center were identified. In addition, all sites had comparable QC results, including similar SNP and sample call rates, Hardy-Weinberg equilibrium P values overall, and minor allele frequencies. The detailed QC report on the merged data set will be deposited in the database of Genotypes and Phenotypes (dbGaP), along with the merged data set.

Single-locus tests of association were performed with linear regression assuming an additive genetic model for all 528 508 SNPs with a normal QRS duration. Our studies of ECG intervals in 32 949 healthy individuals identified sex as a major modulator of normal QRS duration, with minor effects of age and ancestry.1 All analyses were performed unadjusted and then adjusted for age, sex, body mass index, and the first principal component from Eigenstrat23 to adjust for potential population stratification, without significantly changing the key results. Because only sex is significantly associated with QRS duration in the literature, we report here that sex, sex, body mass index, and the first principal component from Eigenstrat were included in the analysis (n=502 905 SNP). The genetic relationship matrix was computed for all 5272 subjects and all SNPs using GCTA. To eliminate possible cryptic relationships, subjects with genetic relationship matrix >0.25 were pruned from the analysis, which removed 310 subjects. The proportion of variance explained by either all SNPs or all SNPs excluding the subset of 23 SNPs significant in the CHARGE GWAS was computed on the remaining subjects for QRS duration. We compared this with a linear regression analysis using the 5 SNPs in Table 2 to estimate the proportion of variance explained by these loci. All analyses were adjusted for age, sex, and the first principal component (previously computed).

PheWAS of QRS-Associated SNPs

We selected the most significant SNP associations for analysis by PheWAS.16,17 For this analysis, we combined the entire eMERGE cohort of European American individuals (n=13 859) identified across the 5 eMERGE sites. These individuals represent a superset of the 5272 individuals with normal ECGs and without heart disease used for the GWAS. To define diseases, we queried all International Classification of Disease, ninth edition, codes from the respective EMRs of the 5 eMERGE sites. The PheWAS software uses occurrences of International Classification of Disease codes to classify each person as having 1 or more of 778 possible clinical phenotypes (typically diseases). For each disease, the PheWAS algorithm constructs a control population by selecting all patients who do not have the case disease or closely related diseases (eg, a patient with a bundle-branch block cannot serve as a control for complete heart block). The PheWAS methodology has previously been validated through rediscovery of known associations.16,17 Analysis of each phenotype then proceeds using a pairwise analysis of all case and control groups for each tested SNP (n=23). We have observed that the positive predictive values increase when individual codes are present more than once in the EMR, and here we required each case to have at least 4 instances of the same International Classification of Disease code in a PheWAS case group. In addition, we did not analyze phenotypes occurring in <50 patients (a prevalence of 0.36% in the data set). Association analyses were performed with PLINK using logistic regression adjusted for age, sex, and the first principal component analyses as calculated by Eigenstrat because, on this larger population, the third principal component was statistically significant.23 Analysis adjusted with and without principal components did not substantively change the results. After identification of PheWAS case and control groups using the PheWAS software, the association analyses were performed using PLINK.26

*This SNP was coded on the opposite strand from that in the CHARGE study. Therefore, although the allele is not identical, the direction of effect is the same.

Table 2. Replicated SNPs in the QRS GWAS analysis

<table>
<thead>
<tr>
<th>CHR</th>
<th>SNP</th>
<th>Location</th>
<th>Coded Allele</th>
<th>Coded Allele Frequency</th>
<th>β, ms</th>
<th>P Value</th>
<th>Coded Allele</th>
<th>Coded Allele Frequency</th>
<th>β, ms</th>
<th>P Value</th>
<th>Nearest Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>rs1805126</td>
<td>38567410</td>
<td>T*</td>
<td>0.665</td>
<td>−1.002</td>
<td>1.45E−8</td>
<td>A</td>
<td>0.655</td>
<td>−0.6568</td>
<td>2.52E−20</td>
<td>SCN5A</td>
</tr>
<tr>
<td>3</td>
<td>rs6795970</td>
<td>38741679</td>
<td>A</td>
<td>0.401</td>
<td>0.765</td>
<td>6.00E−6</td>
<td>C</td>
<td>0.396</td>
<td>0.7476</td>
<td>5.08E−27</td>
<td>SCN10A</td>
</tr>
<tr>
<td>6</td>
<td>rs6906287</td>
<td>119069433</td>
<td>C</td>
<td>0.545</td>
<td>0.717</td>
<td>2.26E−5</td>
<td>C</td>
<td>0.451</td>
<td>0.5383</td>
<td>5.56E−16</td>
<td>C6orf204</td>
</tr>
<tr>
<td>6</td>
<td>rs1321313</td>
<td>36726799</td>
<td>G*</td>
<td>0.760</td>
<td>−0.793</td>
<td>6.13E−5</td>
<td>C</td>
<td>0.742</td>
<td>−0.8129</td>
<td>4.60E−25</td>
<td>CDKN1A</td>
</tr>
<tr>
<td>1</td>
<td>rs2207790</td>
<td>61670555</td>
<td>T*</td>
<td>0.482</td>
<td>−0.622</td>
<td>2.07E−4</td>
<td>A</td>
<td>0.461</td>
<td>−0.5956</td>
<td>6.31E−18</td>
<td>NFIA</td>
</tr>
</tbody>
</table>

CHARGE indicates Cohorts for Heart and Aging Research in Genomic Epidemiology; CHR, chromosome; eMERGE, Electronic Medical Records and Genomics; GWAS, genome-wide association study; and SNP, single-nucleotide polymorphism.

Reported β and P values for the eMERGE analysis are adjusted for sex, and all β values for CHARGE and eMERGE analyses are for the coded allele. Each SNP represents the strongest association in the region and was the target of a phenotype-wide association study (as shown in Table 3).

1379

Ritchie et al  QRS GWAS and PheWAS in Electronic Records
Survival Analysis of the QRS Population

After the PheWAS analysis, we analyzed the original set of 5272 patients who met our algorithm definition for normal cardiac conduction/normal heart for subsequent development of atrial fibrillation and cardiac arrhythmias with the SCN5A rs1805126 and SCN10A rs6795970 SNPs. Phenotype definitions were drawn from the PheWAS analysis using billing codes. Kaplan-Meier analysis and Cox proportional hazard models were calculated using the starting time as the initial normal ECG with a time-to-event analysis. Cox proportional hazard models were adjusted for age, sex, principal components as calculated above, and QRS duration.

Results

Population Identification

We identified 5272 white patients (2488 male and 2784 female patients; Table 1) across the 5 eMERGE-I sites. The positive predictive value of the automated phenotype algorithm to find cases with normal ECGs and without exclusions at the development site, Vanderbilt, to identify study subjects was 97% (95% confidence interval, 91–99). The positive predictive values at Northwestern University and Marshfield Clinic were 97% (95% confidence interval, 83–100) and 100% (95% confidence interval, 96–100), respectively. Combining all reviewed samples across the 3 sites gives a positive predictive value of 98% (95% confidence interval, 96–100). The mean QRS duration was 87.9 milliseconds (SD, 9.5 milliseconds; median, 88.0 milliseconds; Figure 1A).

GWAS Results

A total of 528 508 SNPs passed QC of eMERGE-supported Illumina 660Quad genotyping data in these subjects. Figure 1B shows the GWAS analysis for QRS duration adjusted for sex; the findings were nearly identical for the unadjusted analysis. There was a single association between QRS duration and an SNP (rs1805126) in SCN5A, encoding the cardiac sodium channel gene, that survived Bonferroni correction (β=1.002 milliseconds per copy of the T allele; P=1.45×10−8). The points in green are those single-nucleotide polymorphisms that were also identified at genome-wide significance (P<5×10−7). The red line indicates genome-wide significance (P=5×10−8). The points in green are those single-nucleotide polymorphisms that were also identified at genome-wide significance in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium QRS meta-analysis as described in the text.

Figure 1. A, Distribution of QRS durations in 5272 normal ECGs. B, Genome-wide association analysis of QRS duration using sex-adjusted linear regression. Red line indicates genome-wide significance (P=5×10−8). The points in green are those single-nucleotide polymorphisms that were also identified at genome-wide significance in the CHARGE consortium QRS meta-analysis as described in the text.

The set taken forward to the CHARGE QRS meta-analysis consortium included 108 SNPs with values of P<10−4. The retrieved P values for this set divided into 2 distinct groups: 23 SNPs with P values in the CHARGE set from 10−4 to 10−2 and 85 with P values >0.003. These 23 associations (Table I in the online-only Data Supplement) are located in the 5 loci with the lowest P values reported by the CHARGE consortium: 18 of 23 are in the chromosome 3 locus that includes SCN5A and SCN10A, as well as other genes (eg, EXOG and XYL1B1), and the other 3 loci are near SLCE5F1 and C6orf204 (chromosome 6), near CDRN1A (chromosome 6), and in NFI A (chromosome 1). The most significant SNP for each locus is presented in Table 2. The locus zoom plot (Figure 1 in the online-only Data Supplement) shows little linkage disequilibrium in the chromosome 3 region in HapMap phase III (CEU), consistent with the suggestion that the SCN5A-10A finding may actually indicate multiple independent associations. Specifically, the most significant variants in SCN5A (rs1805126) and SCN10A (rs6795970) are not in linkage disequilibrium (r²<0.20).

With the GTCA approach, we estimated heritability for QRS at 31.1% (SE, 6.9%; P=5.7×10−7) using all SNPs in the data set. Conducting the analysis without the 23 SNPs significant in CHARGE decreased the estimated heritability to 30.3%, a decrease of 0.8%. This was somewhat conservative compared with a linear regression model, which estimated an adjusted r² value of 1.6% for the 5 loci in Table 2.

PheWAS Analysis

The PheWAS data set consisted of 13 859 European American subjects in the entire genotyped eMERGE cohort. The analysis focused on the most significant SNPs in each of the 5 loci associated with QRS (Table 3). Although no associations survived a strict Bonferroni correction for significance (P=0.05/778/5=1.3×10−5), the most significant associations were particularly relevant to cardiac disease and demonstrated significantly different patterns of associations for the 5 QRS-associated loci. The strongest associations for the SNPs in both SCN5A and SCN10A were with the diagnoses of cardiac arrhythmias (P=7.21×10−4 for SCN10A and P=1.1×10−3 for SCN5A) and, for SCN10A, atrial fibrillation (P=8.5×10−4; Figure 2). Table 3 lists associations for the most significant SCN5A (rs1805126) and SCN10A (rs6795970) SNPs, as well as those at the other QRS-associated loci, chromosome 1 (rs2207790) and the 2 chromosome 6 loci (rs6906287 and rs1321313), also graphed in Figures II through IV in the online-only Data Supplement. The CDRN1A and C6orf204 loci were not associated with cardiac arrhythmias (P>0.3, with 80% power to detect an odds ratio >1.12 at P=0.05), and NFI A was weakly associated with cardiac arrhythmias (odds ratio,
Table 3. Most Significant Associations Between SNPs Predicting Normal QRS Duration and Diagnostic Codes

<table>
<thead>
<tr>
<th>Associated Phenotype</th>
<th>Case Count, n</th>
<th>Odds Ratio</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCN5A (Chr 3, rs1805126)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac arrhythmias</td>
<td>3075</td>
<td>0.877</td>
<td>7.21E-04</td>
</tr>
<tr>
<td>Other diseases of upper respiratory tract</td>
<td>509</td>
<td>0.8131</td>
<td>3.83E-03</td>
</tr>
<tr>
<td>Benign prostatic hypertrophy</td>
<td>1615</td>
<td>0.8527</td>
<td>5.54E-03</td>
</tr>
<tr>
<td>Chronic kidney disease</td>
<td>1212</td>
<td>0.8772</td>
<td>6.02E-03</td>
</tr>
<tr>
<td>Melanoma</td>
<td>151</td>
<td>0.7033</td>
<td>8.18E-03</td>
</tr>
<tr>
<td>Dermatophytosis</td>
<td>1229</td>
<td>0.8791</td>
<td>8.77E-03</td>
</tr>
<tr>
<td>SCN10A (Chr 3, rs6795970)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac arrhythmias</td>
<td>3075</td>
<td>0.8781</td>
<td>7.21E-04</td>
</tr>
<tr>
<td>Atrial fibrillation and flutter</td>
<td>1758</td>
<td>0.8519</td>
<td>8.45E-04</td>
</tr>
<tr>
<td>Arterial embolism and thrombosis</td>
<td>150</td>
<td>0.6928</td>
<td>3.20E-03</td>
</tr>
<tr>
<td>Cholecystitis</td>
<td>121</td>
<td>0.6627</td>
<td>3.44E-03</td>
</tr>
<tr>
<td>Convulsions</td>
<td>311</td>
<td>1.249</td>
<td>7.01E-03</td>
</tr>
<tr>
<td>Aphasia</td>
<td>69</td>
<td>0.5999</td>
<td>7.25E-03</td>
</tr>
<tr>
<td>C6orf204 (Chr 6, rs9006287)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>316</td>
<td>0.7592</td>
<td>9.90E-04</td>
</tr>
<tr>
<td>Benign prostatic hypertrophy</td>
<td>1615</td>
<td>0.8576</td>
<td>4.19E-03</td>
</tr>
<tr>
<td>Penicillin allergy</td>
<td>65</td>
<td>1.68</td>
<td>4.01E-03</td>
</tr>
<tr>
<td>Melanoma</td>
<td>151</td>
<td>1.383</td>
<td>5.68E-03</td>
</tr>
<tr>
<td>Disorders of pancreatic secretion</td>
<td>84</td>
<td>0.6421</td>
<td>6.22E-03</td>
</tr>
<tr>
<td>Nasal polyps</td>
<td>127</td>
<td>1.413</td>
<td>7.15E-03</td>
</tr>
<tr>
<td>Prostatitis</td>
<td>150</td>
<td>0.7166</td>
<td>8.07E-03</td>
</tr>
<tr>
<td>NFIA (Chr 1, rs2207790)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postinflammatory pulmonary fibrosis</td>
<td>173</td>
<td>0.7048</td>
<td>1.76E-03</td>
</tr>
<tr>
<td>Multiple myeloma</td>
<td>54</td>
<td>0.5425</td>
<td>2.80E-03</td>
</tr>
<tr>
<td>Neutropenia and leukopenia</td>
<td>238</td>
<td>0.7545</td>
<td>2.81E-03</td>
</tr>
<tr>
<td>Facial nerve disorders</td>
<td>67</td>
<td>1.677</td>
<td>3.70E-03</td>
</tr>
<tr>
<td>Cardiomyopathy</td>
<td>420</td>
<td>0.8033</td>
<td>4.40E-03</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>80</td>
<td>0.6391</td>
<td>6.59E-03</td>
</tr>
<tr>
<td>Disorders of function of stomach</td>
<td>355</td>
<td>1.23</td>
<td>7.73E-03</td>
</tr>
<tr>
<td>Prostatitis</td>
<td>150</td>
<td>0.7239</td>
<td>8.85E-03</td>
</tr>
<tr>
<td>Toxic diffuse goiter</td>
<td>76</td>
<td>1.514</td>
<td>9.47E-03</td>
</tr>
<tr>
<td>CDKN1A (Chr 6, rs1321313)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anorexia</td>
<td>53</td>
<td>0.3649</td>
<td>1.57E-03</td>
</tr>
<tr>
<td>Bacteremia</td>
<td>118</td>
<td>0.5684</td>
<td>2.01E-03</td>
</tr>
<tr>
<td>Anal fissure and fistula</td>
<td>108</td>
<td>1.552</td>
<td>3.04E-03</td>
</tr>
<tr>
<td>Abnormal loss of weight and underweight</td>
<td>398</td>
<td>0.7642</td>
<td>3.39E-03</td>
</tr>
<tr>
<td>Dementias</td>
<td>841</td>
<td>0.833</td>
<td>5.69E-03</td>
</tr>
<tr>
<td>Overweight, obesity, and other</td>
<td>2501</td>
<td>1.105</td>
<td>7.39E-03</td>
</tr>
<tr>
<td>hyperalimentation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spondylosis and allied disorders</td>
<td>1244</td>
<td>1.156</td>
<td>9.39E-03</td>
</tr>
</tbody>
</table>

Chr indicates chromosome; and SNP single-nucleotide polymorphism. For each SNP displayed, all phenotype-wide association study associations P<0.01 are displayed.

0.91; P=0.02). Table II in the online-only Data Supplement presents PheWAS association data for all 23 SNPs significant in CHARGE. Although most SNPs in a given gene displayed similar patterns of PheWAS associations, rs11129801, the strongest SCN10A SNP in our adjusted analysis but a lesser association in CHARGE, had a very different PheWAS pattern, with the strongest associations being epilepsy, uterine cancer, and migraines; atrial fibrillation was not associated (odds ratio, 1.07; P=0.18). In agreement with these data, rs11129801 was only in weak linkage disequilibrium to the other SCN10A SNPs such as rs6800541 (r²=0.20). Likewise, the EXOG locus had a very different PheWAS pattern of associations (prostatic hyperplasia, sexual and gender-identity disorders, liver disease, kidney disease, cerebral degenerations, diarrhea) despite being near the SCN5A-10A region.

**Analysis of QRS Population for Arrhythmias**

After PheWAS analysis, we analyzed the original set of 5272 patients who met our algorithm definition for normal cardiac conduction/normal heart for subsequent development of atrial fibrillation and cardiac arrhythmias with the SCN5A rs1805126 and SCN10A rs6795970. In this population, 173 (3%) developed atrial fibrillation or atrial flutter at some point at least 1 month after the normal ECG, and 605 (11%) were coded as having any arrhythmia. QRS duration itself was associated with the future development of atrial fibrillation (P=0.015) by logistic regression. As with the eMERGE PheWAS, SCN10A rs6795970 was associated with both arrhythmias (hazard ratio [HR], 0.81 per copy of the A allele; P=0.002) and atrial fibrillation/flutter (HR, 0.67 per copy of the A allele; P=0.001). Moreover, this association was essentially unchanged when QRS was also included in the model (HR, 0.68), indicating that the association between rs6795970 and atrial fibrillation is independent of the association between QRS and rs6795970. Similarly, the association between rs6795970 and cardiac arrhythmias was independent of QRS (HR, 0.80 without QRS). Our analysis did not demonstrate an association between SCN5A rs1805126 and either atrial fibrillation (HR, 1.2; P=0.14) or cardiac arrhythmias (HR, 1.03; P=0.66) in the normal QRS population. Figure 3 presents a Kaplan-Meier plot for the relationship between rs6795970 and the development of atrial fibrillation.

**Discussion**

The present study demonstrates that common variants in the SCN5A-SCN10A locus are associated with QRS duration in subjects without clinical evidence of prior heart disease. These patients were derived from clinical practice settings, adding to the growing body of evidence of supporting the utility of EMR-based genomic analysis.12–16,27 The data replicate earlier studies. The major new finding here is that using the PheWAS study paradigm, we were able to examine the longitudinal associations of these genomic variants on disease in a hypothesis-free manner. This analysis revealed that SNPs in SCN5A-10A specifically is associated with atrial fibrillation and cardiac arrhythmias. SCN10A was not as precisely controlled or excluded from all included studies. The major new finding here is that using the PheWAS study paradigm, we were able to examine the longitudinal associations of these genomic variants on disease in a hypothesis-free manner. This analysis revealed that SNPs in SCN5A-10A specifically is associated with atrial fibrillation and cardiac arrhythmias. SCN10A was not as precisely controlled or excluded from all included studies. The major new finding here is that using the PheWAS study paradigm, we were able to examine the longitudinal associations of these genomic variants on disease in a hypothesis-free manner.
in the original heart-healthy study population and were independent of the association of SNP with QRS duration. The latter finding suggests that although variants at the SCN5A-SCN10A locus determine QRS and subsequent arrhythmia susceptibility, they may do so by divergent (pleiotropic) pathways or that conduction slowing occurs not only in the ventricle but also in the atrium, where it contributes to susceptibility to atrial fibrillation. Importantly, the selection logic for the case selection algorithm in our GWAS required the absence of cardiovascular disease at the time of the ECG. Therefore, these associations represent the subsequent development of cardiac arrhythmias in subjects with these variants. This result highlights the potential of the EMR, with multiple diagnoses and longitudinal follow-up, to identify not only variants associated with disease susceptibility or trait variability but also subsequent outcomes associated with these variants.

Drugs that block SCN5A-encoded sodium channels slowed ventricular conduction, prolonged QRS duration,28 and increased mortality after myocardial infarction in the Cardiac Arrhythmia Suppression Trial (CAST).29,30 Available evidence supports the view that slow conduction, particularly in the setting of scarred or ischemic myocardium, promotes re-entrant excitation that leads to fatal arrhythmias,31–33 and in CAST, longer QRS durations also predicted increased mortality among patients treated with placebo.1 In addition, a genetic disease caused by SCN5A loss-of-function mutations (Brugada syndrome) is characterized by slowed ventricular conduction and an increased risk for fatal arrhythmias.34 Interestingly, previous analyses of variable PR duration have also identified strong associations with variants in SCN10A,13,35–37 and multiple mechanisms are currently being examined to explain this effect: expression of SCN10A-encoded channels in cardiomyocytes and/or cardiac neurons or regulation of SCN5A-10A expression.38–40

Our PheWAS analysis suggests that SCN10A, SCN5A, and EXOG variants are associated with cardiac arrhythmia billing codes (entered by either physicians or professional coders); this includes atrial fibrillation and flutter, supraventricular and ventricular tachycardia, cardiac arrest, and other unspecified arrhythmias. SCN10A rs6799570 was specifically associated with atrial fibrillation and flutter, which was noted by Pfeufer

Figure 2. Phenome-wide association study plots of the most significant single-nucleotide polymorphisms associated with QRS duration. A, SCN5A (rs1805126). B, SCN10A (rs6799570). Blue lines indicate P<0.05.
et al.\textsuperscript{36} but not Holm et al.\textsuperscript{35} Chambers et al.\textsuperscript{37} previously demonstrated associations between \textit{SCN10A} rs6795970 and heart block and ventricular fibrillation; we also noted an association with first-degree atrioventricular block (\(P=0.009\); Table II in the online-only Data Supplement). Mouse studies have demonstrated expression of Nav1.8 in vagal and spinal afferents in gastrointestinal mucosa and myenteric plexes,\textsuperscript{41} and interestingly, \textit{SCN10A} was also associated with cholecystitis in the PheWAS. Variants in \textit{CDKN1A} and \textit{C6orf204}, however, were not associated with cardiac arrhythmias, although we cannot exclude that weak associations may exist. In contrast, the \textit{C6orf204} locus seems most associated with neoplastic disorders (colorectal cancer, prostatic hypertrophy, and melanoma); its strongest cardiovascular disease association was atherosclerosis (odds ratio, 1.094; \(P=0.03\)). Thus, PheWAS suggests that although all these regions may be associated with QRS interval, only those SNPs in the \textit{SCN5A-10A} region seem significantly associated with subsequent development of arrhythmias.

Recent growth in large GWAS meta-analyses has shown the power of large numbers to find genomic influences of given traits and disease. In this study, the development of the phenotype algorithm at 1 site was followed by its use at 4 other sites with different EMR systems. Algorithm performance was similar in finding patients meeting inclusion and exclusion criteria at the 3 sites that evaluated the algorithm, providing further validation of the transportability of EMR phenotype algorithms.\textsuperscript{36,42} Estimates of heritability of QRS using all SNPs are consistent with other reports; the heritability associated with the very limited subset of 23 SNPs implicated here appears to explain a surprisingly large proportion of this heritability estimate. This analysis also indicates that as with other traits, there is extensive “missing heritability” when QRS duration is analyzed as a function of common genomic variation.

This report highlights both limitations and real and potential advantages of EMR-based genomic research. The present study involved analysis of subjects accrued in the initial stages of the eMERGE network; thus, a major limitation was the relatively small size of the study set. Large consortia such as CHARGE have used meta-analysis to aggregate individual data sets across many sites and have demonstrated the power of the large numbers to generate highly significant results by this approach. One of the key lessons in the eMERGE experience to date has been that algorithms to identify cases and controls for genomic or other study can be successfully deployed across multiple EMR systems.\textsuperscript{16,42,43} Thus, as the number of subjects with dense genomic information across multiple EMR systems grows, this and other EMR-based studies highlight the potential for accrual of increasingly large sample sets to identify genomic predictors of variability in phenotypes such as physiological traits or disease susceptibility.

Furthermore, EMRs hold the promise, as suggested here, of examining longitudinal healthcare outcomes such as disease complications or response to drug therapies. Identifying cases for such studies requires especially large resources because subsets of subsets (eg, drug response X in disease Y) are required. Current efforts demonstrate the feasibility of accrual of DNA collections coupled to EMRs large enough to support statistically valid analyses of rare variants and/or rare clinical events. In the current eMERGE network, there are 9 sites with EMR-based biobanks that include \(>250,000\) subjects, and \(>50,000\) have been genotyped on a genome-wide platform. Other resources that should expand the reach of EMR-based genomic research include the Kaiser Northern California biobank (>100,000 individuals),\textsuperscript{44} the Million Veteran’s Project (currently >100,000 individuals),\textsuperscript{45} and biobanks that will be coupled to national healthcare systems (eg, the UK Biobank).\textsuperscript{46}

Studies in the EMR environment enable PheWAS analyses; a PheWAS experiment cannot be executed in the absence of diverse diagnoses across many diagnostic classes in study subjects. The PheWAS-based associations reported here did not achieve significance using a strict Bonferroni correction; however, \textit{SCN10A} rs6795970 also was strongly associated with both atrial fibrillation and cardiac arrhythmias in the normal QRS population. The PheWAS approach is still in development, and it is likely that the strategy of using only standard diagnostic codes is a limitation. The refinement of disease classifications and abstraction methods will enable more granular phenotypic subsetting, particularly when applied to increasingly large data sets described above.

A fundamental limitation of the PheWAS technique is that every diagnosis is not explicitly included or excluded in each record. Another potential limitation of EMR-based phenotyping is the accuracy of the information contained in the record. Extracting phenotypes from the EMR can result in errors if the data in the source EMR are incorrect (eg, the EMR specifies a disease the patient does not have). Sex and ancestry testing suggests that these demographic features are rarely incorrect in an EMR.\textsuperscript{7,20} Similarly, the electronic phenotyping experience in eMERGE indicates that recurrent mentions of specific diagnoses or combinations of diverse data types (such as medications plus free text plus diagnostic codes) greatly improve diagnostic accuracy of electronic phenotyping algorithms when assessed as positive predictive value using hand curation as the gold standard.\textsuperscript{47} EMRs contain data on subjects exposed to a healthcare system; therefore, EMR-based studies may not be
generalizable to a broad population. The conduct of EMR-based studies across a variety of geographical locations and practice settings (as in eMERGE) potentially mitigates this issue.

Conclusions

A GWAS conducted across multiple EMR systems replicated known associations for a readily available index of cardiac conduction, the QRS duration. The algorithm deployed allowed us to identify subjects with normal ECGs and no evidence of heart disease, confounding drugs, or electrolyte abnormalities, and the PheWS established genomic variants predicting slower conduction in this population also associated with subsequent development of arrhythmias. Thus, the present findings are consonant with a view that individual susceptibility to serious arrhythmias is determined in part by genetically determined variability in cardiac electrophysiological behaviors. Furthermore, this study highlights the advantages of a genotyped EMR population to explore the subsequent emergence of clinically important phenotypes not ascertained in the original study design.

Sources of Funding

This work was supported by the Electronic Medical Records and Genomics (eMERGE) Network, initiated and funded by the National Human Genome Research Institute, with additional funding from the National Institute of General Medical Sciences, through the following grants: U01-HG-004610 (Group Health Cooperative), U01-HG-004608 (Marshfield Clinic), U01-HG-004599 (Mayo Clinic), U01-HG-004609 (Northwestern University), and U01-HG-004603 (Vanderbilt University, also serving as the Administrative Coordinating Center). BioVU also receives support through the National Center for Research Resources UL1 RR024975, which is now at the National Center for Advancing Translational Sciences, 2 UL1 TR000445. PheWAS method was also supported by RO1-LM-010685 from the National Library of Medicine. Dr. Sotoodehnia was supported by RO1-HL088456. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosures

None.

References

QRS duration, a measure of cardiac intraventricular conduction, varies 2- to 4-fold in individuals without cardiac disease. Slow conduction may promote re-entrant arrhythmias. We identified 5272 individuals with normal ECGs and no evidence of cardiac conduction abnormalities, that were replicated in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) cohort. QRS duration, a measure of cardiac intraventricular conduction, varies 2- to 4-fold in individuals without cardiac disease. Slow conduction may promote re-entrant arrhythmias. We identified 5272 individuals with normal ECGs and no evidence of cardiac conduction abnormalities, that were replicated in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) cohort. QRS duration, a measure of cardiac intraventricular conduction, varies 2- to 4-fold in individuals without cardiac disease. Slow conduction may promote re-entrant arrhythmias. We identified 5272 individuals with normal ECGs and no evidence of cardiac conduction abnormalities, that were replicated in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) cohort. QRS duration, a measure of cardiac intraventricular conduction, varies 2- to 4-fold in individuals without cardiac disease. Slow conduction may promote re-entrant arrhythmias. We identified 5272 individuals with normal ECGs and no evidence of cardiac conduction abnormalities, that were replicated in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) cohort. QRS duration, a measure of cardiac intraventricular conduction, varies 2- to 4-fold in individuals without cardiac disease. Slow conduction may promote re-entrant arrhythmias. We identified 5272 individuals with normal ECGs and no evidence of cardiac conduction abnormalities, that were replicated in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) cohort. QRS duration, a measure of cardiac intraventricular conduction, varies 2- to 4-fold in individuals without cardiac disease. Slow conduction may promote re-entrant arrhythmias. We identified 5272 individuals with normal ECGs and no evidence of cardiac conduction abnormalities, that were replicated in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) cohort. QRS duration, a measure of cardiac intraventricular conduction, varies 2- to 4-fold in individuals without cardiac disease. Slow conduction may promote re-entrant arrhythmias. We identified 5272 individuals with normal ECGs and no evidence of cardiac conduction abnormalities, that were replicated in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) cohort. QRS duration, a measure of cardiac intraventricular conduction, varies 2- to 4-fold in individuals without cardiac disease. Slow conduction may promote re-entrant arrhythmias. We identified 5272 individuals with normal ECGs and no evidence of cardiac conduction abnormalities, that were replicated in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) cohort. QRS duration, a measure of cardiac intraventricular conduction, varies 2- to 4-fold in individuals without cardiac disease. Slow conduction may promote re-entrant arrhythmias. We identified 5272 individuals with normal ECGs and no evidence of cardiac conduction abnormalities, that were replicated in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) cohort.
Genome- and Phenome-Wide Analyses of Cardiac Conduction Identifies Markers of Arrhythmia Risk


on Behalf of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) QRS Group

Circulation. 2013;127:1377-1385; originally published online March 5, 2013;
doi: 10.1161/CIRCULATIONAHA.112.000604

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2013 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/127/13/1377

Data Supplement (unedited) at:
http://circ.ahajournals.org/content/suppl/2013/03/05/CIRCULATIONAHA.112.000604.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation 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 is online at:
http://circ.ahajournals.org/subscriptions/
Supplemental Material

Supplemental Table 1: Replicated SNPs in the QRS GWAS analysis. Reported betas and p-values for eMERGE analysis are adjusted for sex, and all betas for CHARGE and eMERGE analyses are for the coded allele.

<table>
<thead>
<tr>
<th>CHR</th>
<th>SNP</th>
<th>Location</th>
<th>Coded Allele</th>
<th>Coded Allele Frequency</th>
<th>BETA (msec)</th>
<th>p-value</th>
<th>Coded Allele</th>
<th>Coded Allele Frequency</th>
<th>BETA (msec)</th>
<th>p-value</th>
<th>Nearest Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>rs1805126*</td>
<td>38567410</td>
<td>T**</td>
<td>0.665</td>
<td>-1.002</td>
<td>1.45E-08</td>
<td>A</td>
<td>0.655</td>
<td>-0.6568</td>
<td>2.52E-20</td>
<td>SCN5A*</td>
</tr>
<tr>
<td>3</td>
<td>rs9832895</td>
<td>38636337</td>
<td>C</td>
<td>0.518</td>
<td>0.895</td>
<td>1.18E-07</td>
<td>C</td>
<td>0.497</td>
<td>0.6576</td>
<td>8.68E-18</td>
<td>SCN5A</td>
</tr>
<tr>
<td>3</td>
<td>rs6771881</td>
<td>38553268</td>
<td>A</td>
<td>0.756</td>
<td>-1.003</td>
<td>2.23E-07</td>
<td>A</td>
<td>0.745</td>
<td>-0.7659</td>
<td>2.54E-22</td>
<td>EXOG</td>
</tr>
<tr>
<td>3</td>
<td>rs12053903</td>
<td>38568397</td>
<td>C</td>
<td>0.329</td>
<td>0.881</td>
<td>7.54E-07</td>
<td>C</td>
<td>0.330</td>
<td>0.646</td>
<td>1.87E-19</td>
<td>SCN5A</td>
</tr>
<tr>
<td>3</td>
<td>rs7374138</td>
<td>38580736</td>
<td>C</td>
<td>0.737</td>
<td>0.912</td>
<td>1.49E-06</td>
<td>C</td>
<td>0.718</td>
<td>0.5041</td>
<td>5.59E-10</td>
<td>SCN5A</td>
</tr>
<tr>
<td>3</td>
<td>rs11129801</td>
<td>38725379</td>
<td>A</td>
<td>0.294</td>
<td>-0.877</td>
<td>1.82E-06</td>
<td>A</td>
<td>0.301</td>
<td>-0.6296</td>
<td>2.60E-17</td>
<td>SCN10A</td>
</tr>
<tr>
<td>3</td>
<td>rs6800541</td>
<td>38749836</td>
<td>C</td>
<td>0.409</td>
<td>0.803</td>
<td>1.97E-06</td>
<td>C</td>
<td>0.398</td>
<td>0.7394</td>
<td>1.88E-27</td>
<td>SCN10A</td>
</tr>
<tr>
<td>3</td>
<td>rs12636358</td>
<td>38430905</td>
<td>A</td>
<td>0.047</td>
<td>1.867</td>
<td>2.30E-06</td>
<td>A</td>
<td>0.070</td>
<td>1.185</td>
<td>2.46E-12</td>
<td>XYLBI</td>
</tr>
<tr>
<td>3</td>
<td>rs7430439</td>
<td>38787643</td>
<td>A</td>
<td>0.597</td>
<td>-0.794</td>
<td>4.03E-06</td>
<td>A</td>
<td>0.588</td>
<td>-0.5516</td>
<td>8.32E-16</td>
<td>SCN10A</td>
</tr>
<tr>
<td>3</td>
<td>rs6798015</td>
<td>38773840</td>
<td>C</td>
<td>0.374</td>
<td>0.793</td>
<td>4.25E-06</td>
<td>C</td>
<td>0.370</td>
<td>0.7187</td>
<td>5.88E-24</td>
<td>SCN10A</td>
</tr>
<tr>
<td>3</td>
<td>rs4130467</td>
<td>38586708</td>
<td>G**</td>
<td>0.279</td>
<td>0.846</td>
<td>5.88E-06</td>
<td>C</td>
<td>0.289</td>
<td>0.5274</td>
<td>5.98E-12</td>
<td>SCN5A</td>
</tr>
<tr>
<td>3</td>
<td>rs6795970*</td>
<td>38741679</td>
<td>A</td>
<td>0.401</td>
<td>0.765</td>
<td>6.60E-06</td>
<td>A</td>
<td>0.396</td>
<td>0.7476</td>
<td>5.08E-27</td>
<td>SCN10A*</td>
</tr>
<tr>
<td>3</td>
<td>rs7374540</td>
<td>38609146</td>
<td>A</td>
<td>0.609</td>
<td>-0.758</td>
<td>1.04E-05</td>
<td>A</td>
<td>0.598</td>
<td>-0.5246</td>
<td>7.67E-14</td>
<td>SCN5A</td>
</tr>
<tr>
<td>3</td>
<td>rs6797133</td>
<td>38631037</td>
<td>A</td>
<td>0.394</td>
<td>-0.750</td>
<td>1.14E-05</td>
<td>A</td>
<td>0.395</td>
<td>-0.667</td>
<td>2.61E-21</td>
<td>SCN5A</td>
</tr>
<tr>
<td>3</td>
<td>rs9861242</td>
<td>38584338</td>
<td>A</td>
<td>0.250</td>
<td>0.848</td>
<td>1.20E-05</td>
<td>A</td>
<td>0.262</td>
<td>0.5273</td>
<td>7.43E-11</td>
<td>SCN5A</td>
</tr>
<tr>
<td>3</td>
<td>rs7373102</td>
<td>38655632</td>
<td>C</td>
<td>0.496</td>
<td>0.680</td>
<td>1.65E-05</td>
<td>C</td>
<td>0.494</td>
<td>0.4128</td>
<td>1.10E-08</td>
<td>SCN5A</td>
</tr>
<tr>
<td>6</td>
<td>rs6906287*</td>
<td>119069433</td>
<td>C</td>
<td>0.454</td>
<td>0.717</td>
<td>2.26E-05</td>
<td>C</td>
<td>0.451</td>
<td>0.5383</td>
<td>5.56E-16</td>
<td>C6orf204*</td>
</tr>
<tr>
<td>6</td>
<td>rs13970286</td>
<td>118787067</td>
<td>C</td>
<td>0.549</td>
<td>-0.715</td>
<td>2.33E-05</td>
<td>C</td>
<td>0.534</td>
<td>-0.5502</td>
<td>4.89E-16</td>
<td>SLC3F1</td>
</tr>
<tr>
<td>6</td>
<td>rs4307206</td>
<td>118920013</td>
<td>A</td>
<td>0.451</td>
<td>0.712</td>
<td>2.59E-05</td>
<td>A</td>
<td>0.450</td>
<td>0.5399</td>
<td>4.56E-16</td>
<td>C6orf204</td>
</tr>
<tr>
<td>6</td>
<td>rs1321313*</td>
<td>36726799</td>
<td>G**</td>
<td>0.760</td>
<td>-0.793</td>
<td>6.13E-05</td>
<td>C</td>
<td>0.742</td>
<td>-0.8129</td>
<td>4.60E-25</td>
<td>CDKN1A*</td>
</tr>
<tr>
<td>3</td>
<td>rs7430477</td>
<td>38740494</td>
<td>C</td>
<td>0.530</td>
<td>0.661</td>
<td>8.18E-05</td>
<td>C</td>
<td>0.524</td>
<td>0.6061</td>
<td>2.72E-19</td>
<td>SCN10A</td>
</tr>
<tr>
<td>3</td>
<td>rs6599257</td>
<td>38779592</td>
<td>C</td>
<td>0.313</td>
<td>0.705</td>
<td>9.58E-05</td>
<td>C</td>
<td>0.306</td>
<td>0.5508</td>
<td>2.23E-14</td>
<td>SCN10A</td>
</tr>
<tr>
<td>1</td>
<td>rs2207790*</td>
<td>61670555</td>
<td>T**</td>
<td>0.482</td>
<td>-0.622</td>
<td>2.07E-04</td>
<td>A</td>
<td>0.461</td>
<td>-0.5956</td>
<td>6.31E-18</td>
<td>NF1A*</td>
</tr>
</tbody>
</table>

*The phenotype-wide association analysis results for these variants (best or near-best association at each locus/gene in replication analysis) are shown in Table 3.

**This SNP was coded on the opposite strand from that in the CHARGE study. Therefore, while the allele is not identical, the direction of effect is the same.
**Supplemental Figure 1:** LocusZoom plot of the region in chromosome 3 in which most of the replicated associations (18/23) were identified. All variants with $p<10^{-4}$ in the eMERGE data set were replicated at $p<5\times10^{-8}$ in the CHARGE data set.
Supplemental Table 2: PheWAS associations for replicated SNPs (N=23). Data for all associations with P<0.01 are shown. ICD9 codes for given associations can be downloaded from http://knowledgemap.mc.vanderbilt.edu/research/content/phewas.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>Nearest gene</th>
<th>Disease association</th>
<th>cases</th>
<th>odds ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs11129801</td>
<td>3</td>
<td>SCN10A</td>
<td>Epilepsy and recurrent seizures</td>
<td>181</td>
<td>0.71</td>
<td>6.23E-03</td>
</tr>
<tr>
<td>rs11129801</td>
<td>3</td>
<td>SCN10A</td>
<td>Malignant neoplasm of uterus</td>
<td>103</td>
<td>1.49</td>
<td>6.31E-03</td>
</tr>
<tr>
<td>rs11129801</td>
<td>3</td>
<td>SCN10A</td>
<td>Migraine</td>
<td>420</td>
<td>0.81</td>
<td>9.81E-03</td>
</tr>
<tr>
<td>rs6599257</td>
<td>3</td>
<td>SCN10A</td>
<td>Atrial fibrillation and flutter</td>
<td>1758</td>
<td>0.84</td>
<td>6.53E-04</td>
</tr>
<tr>
<td>rs6599257</td>
<td>3</td>
<td>SCN10A</td>
<td>Drug dependence</td>
<td>71</td>
<td>1.68</td>
<td>2.88E-03</td>
</tr>
<tr>
<td>rs6599257</td>
<td>3</td>
<td>SCN10A</td>
<td>Cardiac arrhythmias</td>
<td>3075</td>
<td>0.89</td>
<td>3.03E-03</td>
</tr>
<tr>
<td>rs6599257</td>
<td>3</td>
<td>SCN10A</td>
<td>Lupus erythematosus</td>
<td>93</td>
<td>0.59</td>
<td>4.31E-03</td>
</tr>
<tr>
<td>rs6599257</td>
<td>3</td>
<td>SCN10A</td>
<td>Cholecystitis</td>
<td>121</td>
<td>0.65</td>
<td>4.77E-03</td>
</tr>
<tr>
<td>rs6795970</td>
<td>3</td>
<td>SCN10A</td>
<td>Migraine</td>
<td>420</td>
<td>1.23</td>
<td>5.78E-03</td>
</tr>
<tr>
<td>rs6795970</td>
<td>3</td>
<td>SCN10A</td>
<td>Cardiac arrhythmias</td>
<td>3075</td>
<td>0.88</td>
<td>7.21E-04</td>
</tr>
<tr>
<td>rs6795970</td>
<td>3</td>
<td>SCN10A</td>
<td>Atrial fibrillation and flutter</td>
<td>1758</td>
<td>0.85</td>
<td>8.45E-04</td>
</tr>
<tr>
<td>rs6795970</td>
<td>3</td>
<td>SCN10A</td>
<td>Arterial embolism and thrombosis</td>
<td>150</td>
<td>0.69</td>
<td>3.20E-03</td>
</tr>
<tr>
<td>rs6795970</td>
<td>3</td>
<td>SCN10A</td>
<td>Cholecystitis</td>
<td>121</td>
<td>0.66</td>
<td>3.44E-03</td>
</tr>
<tr>
<td>rs6795970</td>
<td>3</td>
<td>SCN10A</td>
<td>Convulsions</td>
<td>311</td>
<td>1.25</td>
<td>7.01E-03</td>
</tr>
<tr>
<td>rs6795970</td>
<td>3</td>
<td>SCN10A</td>
<td>Aphasia</td>
<td>69</td>
<td>0.60</td>
<td>7.25E-03</td>
</tr>
<tr>
<td>rs6798015</td>
<td>3</td>
<td>SCN10A</td>
<td>Cholecystitis</td>
<td>121</td>
<td>0.56</td>
<td>1.32E-04</td>
</tr>
<tr>
<td>rs6798015</td>
<td>3</td>
<td>SCN10A</td>
<td>Arterial embolism and thrombosis</td>
<td>150</td>
<td>0.67</td>
<td>1.73E-03</td>
</tr>
<tr>
<td>rs6798015</td>
<td>3</td>
<td>SCN10A</td>
<td>Atrial fibrillation and flutter</td>
<td>1758</td>
<td>0.86</td>
<td>2.39E-03</td>
</tr>
<tr>
<td>rs6798015</td>
<td>3</td>
<td>SCN10A</td>
<td>Genital prolapse</td>
<td>511</td>
<td>0.82</td>
<td>8.87E-03</td>
</tr>
<tr>
<td>rs6800541</td>
<td>3</td>
<td>SCN10A</td>
<td>Cardiac arrhythmias</td>
<td>3075</td>
<td>0.89</td>
<td>1.90E-03</td>
</tr>
<tr>
<td>rs6800541</td>
<td>3</td>
<td>SCN10A</td>
<td>Atrial fibrillation and flutter</td>
<td>1758</td>
<td>0.86</td>
<td>2.15E-03</td>
</tr>
<tr>
<td>rs6800541</td>
<td>3</td>
<td>SCN10A</td>
<td>Cholecystitis</td>
<td>121</td>
<td>0.66</td>
<td>2.64E-03</td>
</tr>
<tr>
<td>rs6800541</td>
<td>3</td>
<td>SCN10A</td>
<td>Arterial embolism and thrombosis</td>
<td>150</td>
<td>0.69</td>
<td>3.05E-03</td>
</tr>
<tr>
<td>rs6800541</td>
<td>3</td>
<td>SCN10A</td>
<td>Aphasia</td>
<td>69</td>
<td>0.61</td>
<td>8.15E-03</td>
</tr>
<tr>
<td>rs6800541</td>
<td>3</td>
<td>SCN10A</td>
<td>Gingival and periodontal diseases</td>
<td>229</td>
<td>0.77</td>
<td>9.66E-03</td>
</tr>
<tr>
<td>rs7430439</td>
<td>3</td>
<td>SCN10A</td>
<td>Intestinal malabsorption</td>
<td>156</td>
<td>1.42</td>
<td>2.34E-03</td>
</tr>
<tr>
<td>rs7430439</td>
<td>3</td>
<td>SCN10A</td>
<td>Allergic rinitis</td>
<td>865</td>
<td>1.16</td>
<td>4.83E-03</td>
</tr>
<tr>
<td>rs7430439</td>
<td>3</td>
<td>SCN10A</td>
<td>Disturbance of skin sensation</td>
<td>664</td>
<td>0.85</td>
<td>7.85E-03</td>
</tr>
<tr>
<td>rs7430439</td>
<td>3</td>
<td>SCN10A</td>
<td>First degree AV block</td>
<td>181</td>
<td>1.34</td>
<td>9.09E-03</td>
</tr>
<tr>
<td>rs7430439</td>
<td>3</td>
<td>SCN10A</td>
<td>Cholecystitis</td>
<td>121</td>
<td>0.70</td>
<td>9.44E-03</td>
</tr>
<tr>
<td>rs7430477</td>
<td>3</td>
<td>SCN10A</td>
<td>Cholecystitis</td>
<td>121</td>
<td>1.71</td>
<td>5.85E-05</td>
</tr>
<tr>
<td>rs7430477</td>
<td>3</td>
<td>SCN10A</td>
<td>Epilepsy and recurrent seizures</td>
<td>181</td>
<td>0.72</td>
<td>2.90E-03</td>
</tr>
<tr>
<td>rs7430477</td>
<td>3</td>
<td>SCN10A</td>
<td>Tinnitus</td>
<td>159</td>
<td>1.39</td>
<td>4.46E-03</td>
</tr>
<tr>
<td>rs7430477</td>
<td>3</td>
<td>SCN10A</td>
<td>Gingival and periodontal diseases</td>
<td>229</td>
<td>1.30</td>
<td>6.12E-03</td>
</tr>
<tr>
<td>rs7430477</td>
<td>3</td>
<td>SCN10A</td>
<td>Convulsions</td>
<td>311</td>
<td>0.80</td>
<td>7.36E-03</td>
</tr>
<tr>
<td>rs7430477</td>
<td>3</td>
<td>SCN10A</td>
<td>Migraine</td>
<td>420</td>
<td>0.83</td>
<td>7.87E-03</td>
</tr>
<tr>
<td>rs7430477</td>
<td>3</td>
<td>SCN10A</td>
<td>Inflammatory disease of cervix, vagina, and vulva</td>
<td>266</td>
<td>0.79</td>
<td>8.21E-03</td>
</tr>
<tr>
<td>rs11170286</td>
<td>6</td>
<td>SLC35F1</td>
<td>Colorectal cancer</td>
<td>316</td>
<td>0.77</td>
<td>1.37E-03</td>
</tr>
<tr>
<td>rs11170286</td>
<td>6</td>
<td>SLC35F1</td>
<td>Nasal polyps</td>
<td>127</td>
<td>1.48</td>
<td>2.32E-03</td>
</tr>
<tr>
<td>rs11170286</td>
<td>6</td>
<td>SLC35F1</td>
<td>Malignant neoplasm of colon</td>
<td>226</td>
<td>0.75</td>
<td>3.23E-03</td>
</tr>
<tr>
<td>rs11170286</td>
<td>6</td>
<td>SLC35F1</td>
<td>Personal history of malignant</td>
<td>130</td>
<td>0.69</td>
<td>3.80E-03</td>
</tr>
<tr>
<td>rs</td>
<td>SNP</td>
<td>Gene</td>
<td>Condition</td>
<td>Odds Ratio</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-----------</td>
<td>---------</td>
<td>------------------------------------------------</td>
<td>------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>rs11970286</td>
<td>6</td>
<td>SLC35F1</td>
<td>Personal history of malignant melanoma of skin</td>
<td>151</td>
<td>1.37</td>
<td>7.37E-03</td>
</tr>
<tr>
<td>rs11970286</td>
<td>6</td>
<td>SLC35F1</td>
<td>Hyperplasia of prostate</td>
<td>1615</td>
<td>0.87</td>
<td>9.66E-03</td>
</tr>
<tr>
<td>rs12053903</td>
<td>3</td>
<td>SCN5A</td>
<td>Chronic kidney disease</td>
<td>1212</td>
<td>0.85</td>
<td>9.16E-04</td>
</tr>
<tr>
<td>rs12053903</td>
<td>3</td>
<td>SCN5A</td>
<td>Atherosclerosis</td>
<td>1511</td>
<td>0.87</td>
<td>1.60E-03</td>
</tr>
<tr>
<td>rs12053903</td>
<td>3</td>
<td>SCN5A</td>
<td>Cardiac arrhythmias</td>
<td>3075</td>
<td>0.89</td>
<td>2.80E-03</td>
</tr>
<tr>
<td>rs12053903</td>
<td>3</td>
<td>SCN5A</td>
<td>Other peripheral vascular disease</td>
<td>1228</td>
<td>0.88</td>
<td>5.84E-03</td>
</tr>
<tr>
<td>rs12053903</td>
<td>3</td>
<td>SCN5A</td>
<td>Disorders resulting from impaired renal function</td>
<td>173</td>
<td>0.71</td>
<td>5.92E-03</td>
</tr>
<tr>
<td>rs12053903</td>
<td>3</td>
<td>SCN5A</td>
<td>Osteomyelitis, periostitis, and other infections involving bone</td>
<td>192</td>
<td>0.73</td>
<td>7.56E-03</td>
</tr>
<tr>
<td>rs12053903</td>
<td>3</td>
<td>SCN5A</td>
<td>Other paralytic syndromes</td>
<td>65</td>
<td>1.60</td>
<td>8.78E-03</td>
</tr>
<tr>
<td>rs12053903</td>
<td>3</td>
<td>SCN5A</td>
<td>Thyrotoxicosis with or without goiter</td>
<td>189</td>
<td>0.74</td>
<td>9.64E-03</td>
</tr>
<tr>
<td>rs12053903</td>
<td>3</td>
<td>SCN5A</td>
<td>Fracture of other and unspecified parts of femur</td>
<td>85</td>
<td>0.62</td>
<td>9.96E-03</td>
</tr>
<tr>
<td>rs1805126</td>
<td>3</td>
<td>SCN5A</td>
<td>Cardiac arrhythmias</td>
<td>3075</td>
<td>0.88</td>
<td>1.10E-03</td>
</tr>
<tr>
<td>rs1805126</td>
<td>3</td>
<td>SCN5A</td>
<td>Diseases of upper respiratory tract</td>
<td>509</td>
<td>0.81</td>
<td>3.83E-03</td>
</tr>
<tr>
<td>rs1805126</td>
<td>3</td>
<td>SCN5A</td>
<td>Hyperplasia of prostate</td>
<td>1615</td>
<td>0.85</td>
<td>5.54E-03</td>
</tr>
<tr>
<td>rs1805126</td>
<td>3</td>
<td>SCN5A</td>
<td>Chronic kidney disease</td>
<td>1212</td>
<td>0.88</td>
<td>6.02E-03</td>
</tr>
<tr>
<td>rs1805126</td>
<td>3</td>
<td>SCN5A</td>
<td>Personal history of malignant melanoma of skin</td>
<td>151</td>
<td>0.70</td>
<td>8.18E-03</td>
</tr>
<tr>
<td>rs1805126</td>
<td>3</td>
<td>SCN5A</td>
<td>Dermatophytosis</td>
<td>1229</td>
<td>0.88</td>
<td>8.77E-03</td>
</tr>
<tr>
<td>rs4130467</td>
<td>3</td>
<td>SCN5A</td>
<td>Late effects of musculoskeletal and connective tissue injuries</td>
<td>81</td>
<td>1.67</td>
<td>1.55E-03</td>
</tr>
<tr>
<td>rs4130467</td>
<td>3</td>
<td>SCN5A</td>
<td>Hyperplasia of prostate</td>
<td>1615</td>
<td>0.85</td>
<td>6.97E-03</td>
</tr>
<tr>
<td>rs4130467</td>
<td>3</td>
<td>SCN5A</td>
<td>Systemic lupus erythematosus</td>
<td>81</td>
<td>1.56</td>
<td>7.07E-03</td>
</tr>
<tr>
<td>rs7373102</td>
<td>3</td>
<td>SCN5A</td>
<td>Postinflammatory pulmonary fibrosis</td>
<td>173</td>
<td>1.50</td>
<td>1.31E-04</td>
</tr>
<tr>
<td>rs7373102</td>
<td>3</td>
<td>SCN5A</td>
<td>Irritable Bowel Syndrome</td>
<td>498</td>
<td>0.81</td>
<td>1.54E-03</td>
</tr>
<tr>
<td>rs7373102</td>
<td>3</td>
<td>SCN5A</td>
<td>Diseases of white blood cells</td>
<td>219</td>
<td>1.31</td>
<td>3.98E-03</td>
</tr>
<tr>
<td>rs7373102</td>
<td>3</td>
<td>SCN5A</td>
<td>Abnormal glucose</td>
<td>837</td>
<td>0.87</td>
<td>4.46E-03</td>
</tr>
<tr>
<td>rs7373102</td>
<td>3</td>
<td>SCN5A</td>
<td>Functional digestive disorders</td>
<td>553</td>
<td>0.84</td>
<td>4.95E-03</td>
</tr>
<tr>
<td>rs7373102</td>
<td>3</td>
<td>SCN5A</td>
<td>Elevated blood pressure</td>
<td>829</td>
<td>0.88</td>
<td>7.67E-03</td>
</tr>
<tr>
<td>rs7373102</td>
<td>3</td>
<td>SCN5A</td>
<td>Abnormal cardiovascular function test</td>
<td>838</td>
<td>0.88</td>
<td>7.87E-03</td>
</tr>
<tr>
<td>rs7373102</td>
<td>3</td>
<td>SCN5A</td>
<td>Anxiety, phobic and dissociative disorders</td>
<td>412</td>
<td>0.84</td>
<td>9.09E-03</td>
</tr>
<tr>
<td>rs7374138</td>
<td>3</td>
<td>SCN5A</td>
<td>Malignant neoplasm of female breast</td>
<td>645</td>
<td>1.22</td>
<td>2.10E-03</td>
</tr>
<tr>
<td>rs7374138</td>
<td>3</td>
<td>SCN5A</td>
<td>Chronic pharyngitis and nasopharyngitis</td>
<td>462</td>
<td>1.24</td>
<td>3.14E-03</td>
</tr>
<tr>
<td>rs7374540</td>
<td>3</td>
<td>SCN5A</td>
<td>Dermatophytosis</td>
<td>1229</td>
<td>0.87</td>
<td>2.05E-03</td>
</tr>
<tr>
<td>rs7374540</td>
<td>3</td>
<td>SCN5A</td>
<td>Spondylosis and allied disorders</td>
<td>1244</td>
<td>0.86</td>
<td>2.54E-03</td>
</tr>
<tr>
<td>rs7374540</td>
<td>3</td>
<td>SCN5A</td>
<td>Fracture of patella</td>
<td>65</td>
<td>0.56</td>
<td>3.58E-03</td>
</tr>
<tr>
<td>rs7374540</td>
<td>3</td>
<td>SCN5A</td>
<td>Deviated nasal septum</td>
<td>275</td>
<td>0.77</td>
<td>3.86E-03</td>
</tr>
<tr>
<td>rs7374540</td>
<td>3</td>
<td>SCN5A</td>
<td>Emphysema</td>
<td>138</td>
<td>0.69</td>
<td>5.01E-03</td>
</tr>
<tr>
<td>Variant</td>
<td>Gene</td>
<td>Condition</td>
<td>P-value</td>
<td>OR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------</td>
<td>-----------------------------------------------</td>
<td>---------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7374540</td>
<td>3</td>
<td>SCN5A</td>
<td>Osteopenia</td>
<td>2161</td>
<td>0.89</td>
<td>5.77E-03</td>
</tr>
<tr>
<td>rs7374540</td>
<td>3</td>
<td>SCN5A</td>
<td>Candidiasis</td>
<td>201</td>
<td>0.75</td>
<td>6.29E-03</td>
</tr>
<tr>
<td>rs7374540</td>
<td>3</td>
<td>SCN5A</td>
<td>Sprains and strains</td>
<td>2928</td>
<td>0.91</td>
<td>6.34E-03</td>
</tr>
<tr>
<td>rs7374540</td>
<td>3</td>
<td>SCN5A</td>
<td>Late effects of musculoskeletal and connective tissue injuries</td>
<td>81</td>
<td>1.54</td>
<td>6.47E-03</td>
</tr>
<tr>
<td>rs7374540</td>
<td>3</td>
<td>SCN5A</td>
<td>Urinary tract infection</td>
<td>2094</td>
<td>0.89</td>
<td>8.31E-03</td>
</tr>
<tr>
<td>rs9832895</td>
<td>3</td>
<td>SCN5A</td>
<td>Candidiasis</td>
<td>201</td>
<td>0.75</td>
<td>6.29E-03</td>
</tr>
<tr>
<td>rs9832895</td>
<td>3</td>
<td>SCN5A</td>
<td>Mitral valve disorders</td>
<td>551</td>
<td>1.20</td>
<td>5.79E-03</td>
</tr>
<tr>
<td>rs9832895</td>
<td>3</td>
<td>SCN5A</td>
<td>Debility</td>
<td>309</td>
<td>1.25</td>
<td>6.76E-03</td>
</tr>
<tr>
<td>rs9832895</td>
<td>3</td>
<td>SCN5A</td>
<td>Disorders of tooth development and eruption</td>
<td>55</td>
<td>1.69</td>
<td>7.48E-03</td>
</tr>
<tr>
<td>rs9861242</td>
<td>3</td>
<td>SCN5A</td>
<td>Peptic ulcers</td>
<td>427</td>
<td>1.20</td>
<td>9.29E-03</td>
</tr>
<tr>
<td>rs9861242</td>
<td>3</td>
<td>SCN5A</td>
<td>Late effects of musculoskeletal and connective tissue injuries</td>
<td>81</td>
<td>1.66</td>
<td>2.18E-03</td>
</tr>
<tr>
<td>rs6797133</td>
<td>3</td>
<td>SCN5A</td>
<td>Sleep disorders</td>
<td>52</td>
<td>0.50</td>
<td>2.48E-03</td>
</tr>
<tr>
<td>rs6797133</td>
<td>3</td>
<td>SCN5A</td>
<td>Other diseases of appendix</td>
<td>59</td>
<td>0.59</td>
<td>9.88E-03</td>
</tr>
<tr>
<td>rs12636358</td>
<td>3</td>
<td>XLYB</td>
<td>Sleep disorders</td>
<td>52</td>
<td>2.46</td>
<td>5.26E-03</td>
</tr>
<tr>
<td>rs12636358</td>
<td>3</td>
<td>XLYB</td>
<td>Diseases of hard tissues of teeth</td>
<td>209</td>
<td>1.65</td>
<td>9.73E-03</td>
</tr>
<tr>
<td>rs1321313</td>
<td>6</td>
<td>CDKN1A</td>
<td>Anorexia</td>
<td>53</td>
<td>0.36</td>
<td>1.57E-03</td>
</tr>
<tr>
<td>rs1321313</td>
<td>6</td>
<td>CDKN1A</td>
<td>Bacteremia</td>
<td>118</td>
<td>0.57</td>
<td>2.01E-03</td>
</tr>
<tr>
<td>rs1321313</td>
<td>6</td>
<td>CDKN1A</td>
<td>Anal fissure and fistula</td>
<td>108</td>
<td>1.55</td>
<td>3.04E-03</td>
</tr>
<tr>
<td>rs1321313</td>
<td>6</td>
<td>CDKN1A</td>
<td>Abnormal loss of weight</td>
<td>398</td>
<td>0.76</td>
<td>3.39E-03</td>
</tr>
<tr>
<td>rs1321313</td>
<td>6</td>
<td>CDKN1A</td>
<td>Dementias</td>
<td>841</td>
<td>0.83</td>
<td>5.69E-03</td>
</tr>
<tr>
<td>rs1321313</td>
<td>6</td>
<td>CDKN1A</td>
<td>Overweight, obesity and other hyperalimentation</td>
<td>2501</td>
<td>1.11</td>
<td>7.39E-03</td>
</tr>
<tr>
<td>rs1321313</td>
<td>6</td>
<td>CDKN1A</td>
<td>Spondylosis and allied disorders</td>
<td>1244</td>
<td>1.16</td>
<td>9.39E-03</td>
</tr>
<tr>
<td>rs2207790</td>
<td>1</td>
<td>NFIA</td>
<td>Postinflammatory pulmonary fibrosis</td>
<td>173</td>
<td>0.70</td>
<td>1.76E-03</td>
</tr>
<tr>
<td>rs2207790</td>
<td>1</td>
<td>NFIA</td>
<td>Abnormal serum enzyme levels (acid phosphatase, alkaline phosphatase, amylase, lipase)</td>
<td>205</td>
<td>1.37</td>
<td>1.97E-03</td>
</tr>
<tr>
<td>rs2207790</td>
<td>1</td>
<td>NFIA</td>
<td>Multiple myeloma and immunoproliferative neoplasms</td>
<td>54</td>
<td>0.54</td>
<td>2.80E-03</td>
</tr>
<tr>
<td>rs2207790</td>
<td>1</td>
<td>NFIA</td>
<td>Neutropenia and leukopenia</td>
<td>238</td>
<td>0.75</td>
<td>2.81E-03</td>
</tr>
<tr>
<td>rs2207790</td>
<td>1</td>
<td>NFIA</td>
<td>Facial nerve disorders</td>
<td>67</td>
<td>1.68</td>
<td>3.70E-03</td>
</tr>
<tr>
<td>rs2207790</td>
<td>1</td>
<td>NFIA</td>
<td>Cardiomyopathy</td>
<td>420</td>
<td>0.80</td>
<td>4.40E-03</td>
</tr>
<tr>
<td>rs2207790</td>
<td>1</td>
<td>NFIA</td>
<td>Pneumothorax</td>
<td>80</td>
<td>0.64</td>
<td>6.59E-03</td>
</tr>
<tr>
<td>rs2207790</td>
<td>1</td>
<td>NFIA</td>
<td>Disorders of function of stomach</td>
<td>355</td>
<td>1.23</td>
<td>7.73E-03</td>
</tr>
<tr>
<td>rs2207790</td>
<td>1</td>
<td>NFIA</td>
<td>Prostatitis</td>
<td>150</td>
<td>0.72</td>
<td>8.85E-03</td>
</tr>
<tr>
<td>rs2207790</td>
<td>1</td>
<td>NFIA</td>
<td>Toxic diffuse goiter</td>
<td>76</td>
<td>1.54</td>
<td>9.47E-03</td>
</tr>
<tr>
<td>rs4307206</td>
<td>6</td>
<td>C6orf204</td>
<td>Colorectal cancer</td>
<td>316</td>
<td>0.76</td>
<td>1.01E-03</td>
</tr>
<tr>
<td>rs4307206</td>
<td>6</td>
<td>C6orf204</td>
<td>Malignant neoplasm of colon</td>
<td>226</td>
<td>0.74</td>
<td>2.76E-03</td>
</tr>
<tr>
<td>rs4307206</td>
<td>6</td>
<td>C6orf204</td>
<td>Hyperplasia of prostate</td>
<td>1615</td>
<td>0.86</td>
<td>3.92E-03</td>
</tr>
<tr>
<td>rs4307206</td>
<td>6</td>
<td>C6orf204</td>
<td>Personal history of malignant neoplasm of colon</td>
<td>130</td>
<td>0.69</td>
<td>4.14E-03</td>
</tr>
<tr>
<td>rs4307206</td>
<td>6</td>
<td>C6orf204</td>
<td>Penicillin allergy</td>
<td>65</td>
<td>1.64</td>
<td>6.07E-03</td>
</tr>
<tr>
<td>rs4307206</td>
<td>6</td>
<td>C6orf204</td>
<td>Disorders of pancreatic internal secretion</td>
<td>84</td>
<td>0.65</td>
<td>7.40E-03</td>
</tr>
<tr>
<td>rs4307206</td>
<td>6</td>
<td>C6orf204</td>
<td>Nasal polyps</td>
<td>127</td>
<td>1.40</td>
<td>8.87E-03</td>
</tr>
<tr>
<td>rs4307206</td>
<td>6</td>
<td>C6orf204</td>
<td>Personal history of malignant melanoma of skin</td>
<td>151</td>
<td>1.36</td>
<td>9.30E-03</td>
</tr>
<tr>
<td>rs6906287</td>
<td>6</td>
<td>C6orf204</td>
<td>Colorectal cancer</td>
<td>316</td>
<td>0.76</td>
<td>9.90E-04</td>
</tr>
<tr>
<td>rs6906287</td>
<td>6</td>
<td>C6orf204</td>
<td>Malignant neoplasm of colon</td>
<td>226</td>
<td>0.75</td>
<td>3.03E-03</td>
</tr>
<tr>
<td>rs6906287</td>
<td>6</td>
<td>C6orf204</td>
<td>Penicillin allergy</td>
<td>65</td>
<td>1.68</td>
<td>4.01E-03</td>
</tr>
<tr>
<td>rs6906287</td>
<td>6</td>
<td>C6orf204</td>
<td>Hyperplasia of prostate</td>
<td>1615</td>
<td>0.86</td>
<td>4.19E-03</td>
</tr>
<tr>
<td>rs6906287</td>
<td>6</td>
<td>C6orf204</td>
<td>Personal history of malignant neoplasm of colon</td>
<td>130</td>
<td>0.70</td>
<td>5.04E-03</td>
</tr>
<tr>
<td>rs6906287</td>
<td>6</td>
<td>C6orf204</td>
<td>Personal history of malignant melanoma of skin</td>
<td>151</td>
<td>1.38</td>
<td>5.68E-03</td>
</tr>
<tr>
<td>rs6906287</td>
<td>6</td>
<td>C6orf204</td>
<td>Disorders of pancreatic internal secretion</td>
<td>84</td>
<td>0.64</td>
<td>6.22E-03</td>
</tr>
<tr>
<td>rs6906287</td>
<td>6</td>
<td>C6orf204</td>
<td>Nasal polyps</td>
<td>127</td>
<td>1.41</td>
<td>7.15E-03</td>
</tr>
<tr>
<td>rs6906287</td>
<td>6</td>
<td>C6orf204</td>
<td>Prostatitis</td>
<td>150</td>
<td>0.72</td>
<td>8.07E-03</td>
</tr>
<tr>
<td>rs6771881</td>
<td>3</td>
<td>EXOG</td>
<td>Hyperplasia of prostate</td>
<td>1615</td>
<td>0.81</td>
<td>8.10E-04</td>
</tr>
<tr>
<td>rs6771881</td>
<td>3</td>
<td>EXOG</td>
<td>Chronic liver disease</td>
<td>254</td>
<td>1.33</td>
<td>3.55E-03</td>
</tr>
<tr>
<td>rs6771881</td>
<td>3</td>
<td>EXOG</td>
<td>Cardiac arrhythmias</td>
<td>3075</td>
<td>0.88</td>
<td>3.86E-03</td>
</tr>
<tr>
<td>rs6771881</td>
<td>3</td>
<td>EXOG</td>
<td>Sexual and gender identity disorders</td>
<td>76</td>
<td>0.51</td>
<td>4.20E-03</td>
</tr>
<tr>
<td>rs6771881</td>
<td>3</td>
<td>EXOG</td>
<td>Chronic kidney disease</td>
<td>1212</td>
<td>0.86</td>
<td>5.07E-03</td>
</tr>
<tr>
<td>rs6771881</td>
<td>3</td>
<td>EXOG</td>
<td>Cerebral degenerations</td>
<td>137</td>
<td>0.63</td>
<td>5.29E-03</td>
</tr>
<tr>
<td>rs6771881</td>
<td>3</td>
<td>EXOG</td>
<td>Diarrhea</td>
<td>880</td>
<td>0.85</td>
<td>9.06E-03</td>
</tr>
</tbody>
</table>
Supplemental Figure 2: Phenome-wide association study plot for CDKN1A (rs1321313).
Supplemental Figure 3: Phenome-wide association study plot for *C6orf204* (rs6906287).
Supplemental Figure 4: Phenome-wide association study plot for NFIA (rs2207790).