

Promises and Challenges of Using Clinical Data for Discovery: Clinical Research IT Priorities in AHCs

Daniel E Ford, MD, MPH
Vice Dean for Clinical Investigation
Johns Hopkins School of Medicine

Sources of Phenotype Data

- Clinical Medical Records
 - One hospital or larger network
- Research Systematic Collection
 - Sample of those presenting for care
 - Population-based sample
 - Combining research samples

eMERGE NETWORK

- National consortium to develop, disseminate and apply approaches that combine DNA biorepositories with electronic medical records (EMR)
- Members
 - Group Health/University of Washington
 - Marshfield Clinic
 - Mayo Clinic
 - Northwestern University
 - Vanderbilt University

Cardiovascular Research Consortium

- Framingham Study (5-15,000)
- ARIC Study (15,000)
- Cardiovascular Health Study (8,000)
- CARDIA Study (5,000)

- All with at least 10 years of followup

Genetic Variations in Nitric Oxide Synthase 1 Adaptor Protein Are Associated With Sudden Cardiac Death in US White Community-Based Populations

Kao, W H. Linda PhD, MHS*; Arking, Dan E. PhD*; Post, Wendy MD, MS*; Rea, Thomas D. MD, MPH; Sotoodehnia, Nona MD, MPH; Prineas, Ronald J. MD, PhD; Bishe, Bryan BS; Doan, Betty Q. PhD, MPH; Boerwinkle, Eric PhD; Psaty, Bruce M. MD, PhD; Tomaselli, Gordon F. MD; Coresh, Josef MD, PhD, MHS; Siscovick, David S. MD; Marbán, Eduardo MD, PhD; Spooner, Peter M. PhD; Burke, Gregory L. MD, MSc; Chakravarti, Aravinda PhD

Background— The ECG QT interval is associated with risk of sudden cardiac death (SCD). A previous genome-wide association study demonstrated that allelic variants (rs10494366 and rs4657139) in the nitric oxide synthase 1 adaptor protein (NOS1AP), which encodes a carboxy-terminal PDZ ligand of neuronal nitric oxide synthase, are associated with the QT interval in white adults. The present analysis was conducted to validate the association between NOS1AP variants and the QT interval and to examine the association with SCD in a combined population of 19 295 black and white adults from the Atherosclerosis Risk In Communities Study and the Cardiovascular Health Study.

Methods and Results— We examined 19 tagging single-nucleotide polymorphisms in the genomic blocks containing rs10494366 and rs4657139 in NOS1AP. SCD was defined as a sudden pulseless condition of cardiac origin in a previously stable individual. General linear models and Cox proportional hazards regression models were used. Multiple single-nucleotide polymorphisms in NOS1AP, including rs10494366, rs4657139, and rs16847548, were significantly associated with adjusted QT interval in whites ($P < 0.0001$). In whites, after adjustment for age, sex, and study, the relative hazard of SCD associated with each C allele at rs16847548 was 1.31 (95% confidence interval 1.10 to 1.56, $P = 0.002$), assuming an additive model. In addition, a downstream neighboring single-nucleotide polymorphism, rs12567209, which was not correlated with rs16847548 or QT interval, was also independently associated with SCD in whites (relative hazard 0.57, 95% confidence interval 0.39 to 0.83, $P = 0.003$). Adjustment for QT interval and coronary heart disease risk factors attenuated but did not eliminate the association between rs16847548 and SCD, and such adjustment had no effect on the association between rs12567209 and SCD. No significant associations between tagging single-nucleotide polymorphisms in NOS1AP and either QT interval or SCD were observed in blacks.

Conclusions— In a combined analysis of 2 population-based prospective cohort studies, sequence variations in NOS1AP were associated with baseline QT interval and the risk of SCD in white US adults.

THE LANCET

Volume 372, Issue 9654, 6 December 2008-12 December 2008, Pages 1953-1961

Association of three genetic loci with uric acid concentration and risk of gout: a genome-wide association study

[Dehghan A](#), [Köttgen A](#), [Yang Q](#), [Hwang SJ](#), [Kao WL](#), [Rivadeneira F](#), [Boerwinkle E](#), [Levy D](#), [Hofman A](#), [Astor BC](#), [Benjamin EJ](#), [van Duijn CM](#), [Witteman JC](#), [Coresh J](#), [Fox CS](#).

Background

Hyperuricaemia, a highly heritable trait, is a key risk factor for gout. We aimed to identify novel genes associated with serum uric acid concentration and gout.

Methods

Genome-wide association studies were done for serum uric acid in 7699 participants in the Framingham cohort and in 4148 participants in the Rotterdam cohort. Genome-wide significant single nucleotide polymorphisms (SNPs) were replicated in white (n=11 024) and black (n=3843) individuals who took part in the study of Atherosclerosis Risk in Communities (ARIC). The SNPs that reached genome-wide significant association with uric acid in either the Framingham cohort ($p < 5.0 \times 10^{-8}$) or the Rotterdam cohort ($p < 1.0 \times 10^{-7}$) were evaluated with gout. The results obtained in white participants were combined using meta-analysis.

Findings

Three loci in the Framingham cohort and two in the Rotterdam cohort showed genome-wide association with uric acid. Top SNPs in each locus were: missense rs16890979 in *SLC2A9* ($p = 7.0 \times 10^{-168}$ and 2.9×10^{-18} for white and black participants, respectively); missense rs2231142 in *ABCG2* ($p = 2.5 \times 10^{-60}$ and 9.8×10^{-4}), and rs1165205 in *SLC17A3* ($p = 3.3 \times 10^{-26}$ and 0.33). All SNPs were direction-consistent with gout in white participants: rs16890979 (OR 0.59 per T allele, 95% CI 0.52–0.68, $p = 7.0 \times 10^{-14}$), rs2231142 (1.74, 1.51–1.99, $p = 3.3 \times 10^{-15}$), and rs1165205 (0.85, 0.77–0.94, $p = 0.002$). In black participants of the ARIC study, rs2231142 was direction-consistent with gout (1.71, 1.06–2.77, $p = 0.028$). An additive genetic risk score of high-risk alleles at the three loci showed graded associations with uric acid (272–351 $\mu\text{mol/L}$ in the Framingham cohort, 269–386 $\mu\text{mol/L}$ in the Rotterdam cohort, and 303–426 $\mu\text{mol/L}$ in white participants of the ARIC study) and gout (frequency 2–13% in the Framingham cohort, 2–8% in the Rotterdam cohort, and 1–18% in white participants in the ARIC study).

Interpretation

We identified three genetic loci associated with uric acid concentration and gout. A score based on genes with a putative role in renal urate handling showed a substantial risk for gout.

Common variants at ten loci modulate the QT interval duration in the QTSCD Study

[Pfeufer A](#), [Sanna S](#), [Arking DE](#), [Müller M](#), [Gateva V](#), [Fuchsberger C](#), [Ehret GB](#), [Orrú M](#), [Pattaro C](#), [Köttgen A](#), [Perz S](#), [Usala G](#), [Barbalic M](#), [Li M](#), [Pütz B](#), [Scuteri A](#), [Prineas RJ](#), [Sinner MF](#), [Gieger C](#), [Najjar SS](#), [Kao WH](#), [Mühleisen TW](#), [Dei M](#), [Happle C](#), [Möhlenkamp S](#), [Crisponi L](#), [Erbel R](#), [Jöckel KH](#), [Naitza S](#), [Steinbeck G](#), [Marroni F](#), [Hicks AA](#), [Lakatta E](#), [Müller-Myhsok B](#), [Pramstaller PP](#), [Wichmann HE](#), [Schlessinger D](#), [Boerwinkle E](#), [Meitinger T](#), [Uda M](#), [Coresh J](#), [Kääb S](#), [Abecasis GR](#), [Chakravarti A](#).

The QT interval, a measure of cardiac repolarization, predisposes to ventricular arrhythmias and sudden cardiac death (SCD) when prolonged or shortened. A common variant in NOS1AP is known to influence repolarization. We analyze genome-wide data from five population-based cohorts (ARIC, KORA, SardiNIA, GenNOVA and HNR) with a total of 15,842 individuals of European ancestry, to confirm the NOS1AP association and identify nine additional loci at $P < 5 \times 10^{-8}$. Four loci map near the monogenic long-QT syndrome genes KCNQ1, KCNH2, SCN5A and KCNJ2. Two other loci include ATP1B1 and PLN, genes with established electrophysiological function, whereas three map to RNF207, near LITAF and within NDRG4-GINS3-SETD6-CNOT1, respectively, all of which have not previously been implicated in cardiac electrophysiology. These results, together with an accompanying paper from the QTGEN consortium, identify new candidate genes for ventricular arrhythmias and SCD.

Factors to consider in measuring phenotypes

- Representativeness of the sample
 - Who does not seek care?
 - Who is lost to followup?
- Validity of the data
 - Smoking status
 - Clinical measurements – BP, mental disorders, etc
- Costs of data collection
 - Cost of EMR and cost of research data collection