

## Extremes of Unexplained Variation as a Phenotype An Efficient Approach for Genome-Wide Association Studies of Cardiovascular Disease

Matthew B. Lanktree, BSc; Robert A. Hegele, MD, FRCPC;  
Nicholas J. Schork, PhD; J. David Spence, MD, FRCPC

Because cardiovascular disease (CVD) is the leading cause of morbidity and mortality in the Western world,<sup>1</sup> reduction of CVD is the focus of a large body of research. Ever since the initial Framingham study, researchers have sought to identify and validate CVD risk factors, now often referred to as “traditional” risk factors, including age, sex, blood pressure, lipid parameters, diabetes, and smoking.<sup>2</sup> More recently, extraordinary efforts have been undertaken to construct large study samples to carry out population-based genome-wide association studies (GWAS) to further deconstruct the heritable component of CVD risk, with the successful identification of molecular determinants, such as the robustly associated chromosome 9p21 locus.<sup>3–6</sup> However, a substantial component of heritable CVD risk still exists,<sup>7</sup> and new study designs and statistical methods will be necessary for uncovering additional associated variants.

Mounting evidence suggests that common complex disorders represent the extremes of quantitative traits.<sup>8</sup> The evaluation of individuals with extreme values of a continuously distributed trait in studies seeking to identify quantitative trait loci has a long history in genetics. “Selective genotyping” was first described by Lander and Botstein<sup>9</sup> in the context of animal models, and later Carey and Williamson<sup>10</sup> and then Risch and Zhang<sup>11</sup> described the approach in family-based linkage studies. Schork et al<sup>12</sup> demonstrated formally that selective genotyping, or genotyping of the extremes, showed substantial efficiency gains in human genetic association studies. Since the publication by Schork et al, a number of authors have extended the extreme sampling strategy to the use of more general selection strategy designs,<sup>13</sup> considerations when loci are multiallelic and exhibit varying strengths of linkage disequilibrium,<sup>14</sup> and for use in DNA pooling strategies.<sup>15</sup>

The various ultrasound techniques used to monitor carotid atherosclerosis, including intima-media thickness, plaque area, and plaque volume, are correlated with each other<sup>16</sup> but appear to represent different components of the atherosclerotic process.<sup>17,18</sup> Among the carotid imaging phenotypes,

plaque area as measured by 2-dimensional B-mode ultrasound is so far best correlated with traditional risk factors and best able to predict coronary events, stroke, and death.<sup>18–21</sup> Here we update a previously described approach<sup>22</sup> in the context of GWAS, namely, genotyping individuals with extreme values of the multiple-regression residual of plaque area involving correction for traditional risk factors, to efficiently detect genetic variants with small effects on CVD.

Studying the extremes of plaque area unexplained by traditional risk factors permits the omission of those individuals who are not of interest, specifically those whose extreme plaque area is predicted by traditional risk factors that are neither genetically based nor modifiable or whose extreme plaque area is essentially predicted by treatable traditional risk factors, such as tobacco smoking, blood pressure, and cholesterol. The extremes of unexplained atherosclerosis design permits a focus on patients who are of great interest with respect to identifying new disease pathways and therapeutic targets. We find that to obtain similar statistical power, the number needed to genotype is 4 times greater when using a population-based approach compared with an approach based on targeted genotyping of individuals at the extremes of unexplained atherosclerosis and unexplained protection. The approach described here could be easily adapted to other complex phenotypes with known risk factors.

### Methods

#### Subjects

Patients were ascertained at the Vascular Disease Prevention Clinics at University Hospital in London, Ontario, Canada. Patients were referred to the Premature Atherosclerosis Clinic because of either clinical evidence of excessive premature atherosclerosis given Framingham risk factors or a strong family history for early CVD. They were referred to the Stroke Prevention Clinic because of history of stroke or transient ischemic attack. The study was approved by the ethics review board of the University of Western Ontario. Medical history questionnaires and physical examination were performed on all patients. Plasma lipid and lipoprotein concentrations were determined from blood, after a 12-hour fast, with standard methodolo-

Received December 23, 2009; accepted February 22, 2010.

From the Departments of Medicine (M.B.L., R.A.H., J.D.S.) and Biochemistry (M.B.L., R.A.H.) and the Stroke Prevention and Atherosclerosis Research Centre (J.D.S.), Robarts Research Institute and Schulich School of Medicine and Dentistry, University of Western Ontario, London, Ontario, Canada; and the Scripps Translational Sciences Institute (N.J.S.), La Jolla, Calif.

Correspondence to Robert A. Hegele, MD, FRCPC, FACP, Blackburn Cardiovascular Genetics Laboratory, Robarts Research Institute, University of Western Ontario, London, Ontario, Canada, N6A 5K8. E-mail hegele@robarts.ca

(*Circ Cardiovasc Genet.* 2010;3:215-221.)

© 2010 American Heart Association, Inc.

*Circ Cardiovasc Genet* is available at <http://circgenetics.ahajournals.org>

DOI: 10.1161/CIRCGENETICS.109.934505

gies.<sup>22</sup> Two-dimensional plaque area was determined by using a previously reported, validated approach on an Advanced Technologies Laboratory Mark 9 ultrasound machine before the year 2000 and subsequently on an Advanced Technologies Laboratory (Phillips) HDI 5000 instrument.<sup>18,22,23</sup> Plaque area was calculated as the sum of the cross-sectional areas of all plaques seen in the longitudinal views of the common, external, and internal carotid arteries. For plaque area observations, within-observer reliability was 0.94 (intraclass correlation;  $\kappa$ ), and between-observer reliability was 0.86.<sup>23</sup>

All patients provided informed consent for DNA analysis and participation in research. The authors had full access to and take full responsibility for the integrity of the data. All authors have read and agreed to the manuscript as written.

### Statistical Analysis

To obtain a normal distribution, plaque area was cube root transformed, as previously reported.<sup>18,22,23</sup> To determine the extent of variation explained by different genetic factors, we first performed a multiple regression with a forward stepwise modeling approach to obtain residual plaque area values that remove variation induced by nongenetic factors. The variables included age, sex, smoking (pack-years), systolic blood pressure, diastolic blood pressure, HDL cholesterol, diabetes diagnosis, blood pressure prescription therapy, and lipid lowering prescription therapy. Blood pressure and lipid parameters included a mix of subjects on and off treatment, a factor mitigated by including prescription therapy for each as an additional variable. Analyses were performed in SAS version 9.1 (SAS Institute, Cary, NC). Each term in the regression model was required to meet a statistical significance of 0.01. A "predicted atherosclerosis" value was calculated by summing the product of each individual's independent variables and the unstandardized parameter coefficients from the multiple linear regression. Subtracting an individual's predicted plaque area from his or her actual plaque area yielded a residual plaque area value.

Power calculations were performed with "PS Power and Sample Size calculator"<sup>24</sup> and the "genetic power calculator."<sup>25</sup> Correction for multiple-hypothesis testing is difficult in the context of high-throughput genotyping technologies. Although some alternative methods have been proposed, such as false discovery rate or permutation-based methods, Bonferroni correction remains the standard methodology for GWAS. Power calculations were performed with a type I error rate ( $\alpha$ ) of  $5 \times 10^{-7}$ , a threshold argued to maintain the false-positive rate at an acceptable level given an adequately powered study.<sup>6</sup> The standard statistical technique for assessing genetic association between a biallelic marker and a quantitative trait

is either a genotype-based ANOVA with 2 *df* or a 1 *df* linear regression, assuming an additive model.<sup>26</sup> Alternatively, if a recessive or dominant inheritance model is hypothesized, either a 1 *df* ANOVA between genotype groups or a 2 *df* linear regression with an additional term for dominance deviation can be performed. However, after targeted genotyping of individuals from the extreme tails of the distribution, the normal distribution of the quantitative trait is lost, and  $\chi^2$  analysis, testing the number of risk alleles in each tail defined as a discrete trait, would be the standard analysis method.<sup>27</sup> However, additional nonparametric tests including a score test,<sup>27</sup> likelihood-based,<sup>28</sup> and permutation methodologies<sup>29</sup> are currently being developed for testing association with quantitative traits at the extremes of a distribution and are theoretically more powerful.<sup>30</sup>

For power analysis of the extremes of unexplained atherosclerosis design, necessary parameters included the number of cases (*n*); the ratio of individuals in 1 extreme to the other (*m*); the minor allele frequency (*f*); the percentage of trait variance explained by the genetic factor ( $\varphi$ ); and the number of SDs from the mean for threshold-defined extremes ( $\delta$ ). The following parameters were used:  $\alpha = 5 \times 10^{-7}$ ; *m* = 1;  $\varphi = 0.01$  to 0.02; *f* = 0.10 to 0.40; and  $\delta = \pm 1.65$  (denoting 5th and 95th percentiles). For power analysis of the population-based sample, necessary parameters for power calculations included the sample size (*n*); SD of the independent variable ( $\sigma_x$ ), a function of the minor allele frequency; the SD of the dependent variable ( $\sigma_y$ ), which we calculated from observed plaque area; the predicted effect size ( $\lambda$ ); and the type 1 error rate ( $\alpha$ ). The following parameters were used:  $\alpha = 5 \times 10^{-7}$ ;  $\sigma_x = 0.43$  to 0.70 (0.65 given a minor allele frequency of 30%);  $\sigma_y = 1.98$  (observed SD of plaque area in the whole sample, Table 1); and  $\lambda = 0.297$  (15% of 1 SD) or  $\lambda = 0.436$  (22% of 1 SD). If the assayed variant is not causally related to the trait in question but is merely in linkage disequilibrium with the causal genetic variant, then the effect size of the causal variant will be attenuated to a degree dictated by the linkage disequilibrium between the causal and tested variants.<sup>12</sup> Owing to the density of modern arrays, in both study designs we have made the assumption that we are testing a marker with high linkage disequilibrium with the causative variant.

## Results

### Forward Stepwise Multiple-Regression Analysis of Plaque Area

Baseline characteristics for study participants are shown in Table 1. Forward stepwise multiple regression validated the inclusion of the 9 independent variables ( $P \leq 0.005$ ; Table 2),

**Table 1. Baseline Patient Characteristics**

Characteristic	Top 5% "Unexplained Protection"	Whole Sample	Bottom 5% "Unexplained Atherosclerosis"	Top vs Bottom <i>P</i> *
n/N	175	3146	175	
Female sex	79 (45)	1504 (48)	73 (41)	NS
Age, y	59.9 ± 13.1	59.9 ± 14.8	58.9 ± 11.1	NS
Smoking, pack-years	8 (0 to 30)	4 (0 to 25)	15 (1.5 to 30)	NS
HDL cholesterol, mmol/L	1.30 ± 0.40	1.32 ± 0.44	1.27 ± 0.41	NS
Systolic BP, mm Hg	139 (129 to 159)	140 (128 to 155)	140 (125 to 155)	NS
Diastolic BP, mm Hg	80 (74 to 89)	81 (73 to 90)	81 (74 to 90)	NS
Diabetes	19 (11)	353 (11)	21 (12)	NS
BP Rx	84 (48)	1664 (53)	78 (45)	NS
Lipid Rx	73 (42)	1346 (43)	71 (41)	NS
Predicted plaque area <sup>1/3</sup>	5.78 ± 1.49	5.65 ± 1.58	5.68 ± 1.11	NS
Observed plaque area <sup>1/3</sup>	1.05 ± 1.47	4.05 ± 1.98	7.16 ± 1.15	$3 \times 10^{-142}$

Values are presented as n (%), mean ± SD, or median (interquartile range). BP indicates blood pressure; Rx, prescription therapy; NS, not significant.

\*Two-sided significance of ANOVA for continuous traits or  $\chi^2$  for discrete traits.

**Table 2. Multiple Regression of Ultrasonographic 2D Plaque Area**

Factor	Variable Type	Parameter Estimate	Standard Error	$\beta$	Partial $R^2$	$P$
Age	Continuous	0.076	0.002	0.52	0.396	<0.001
Smoking, pack-years	Continuous	0.022	0.001	0.21	0.063	<0.001
Male sex	Binary	-0.684	0.056	-0.16	0.027	<0.001
Systolic BP	Continuous	0.014	0.002	0.14	0.013	<0.001
Lipid Rx	Binary	0.410	0.054	0.09	0.011	<0.001
Diastolic BP	Continuous	-0.011	0.003	-0.07	0.003	<0.001
BP Rx	Binary	0.231	0.053	0.05	0.003	<0.001
Diabetes	Binary	0.324	0.082	0.05	0.002	<0.001
HDL cholesterol	Continuous	-0.217	0.062	0.04	0.002	0.005

$R^2=0.522$ ;  $P<0.0001$ .  $\beta$  indicates standardized parameter estimate or parameter as a percentage of the SD of the risk factor;  $R^2$ , coefficient of determination; BP, blood pressure; Rx, prescription therapy;  $P$ , 2-tailed statistical significance that the parameter is different from zero.

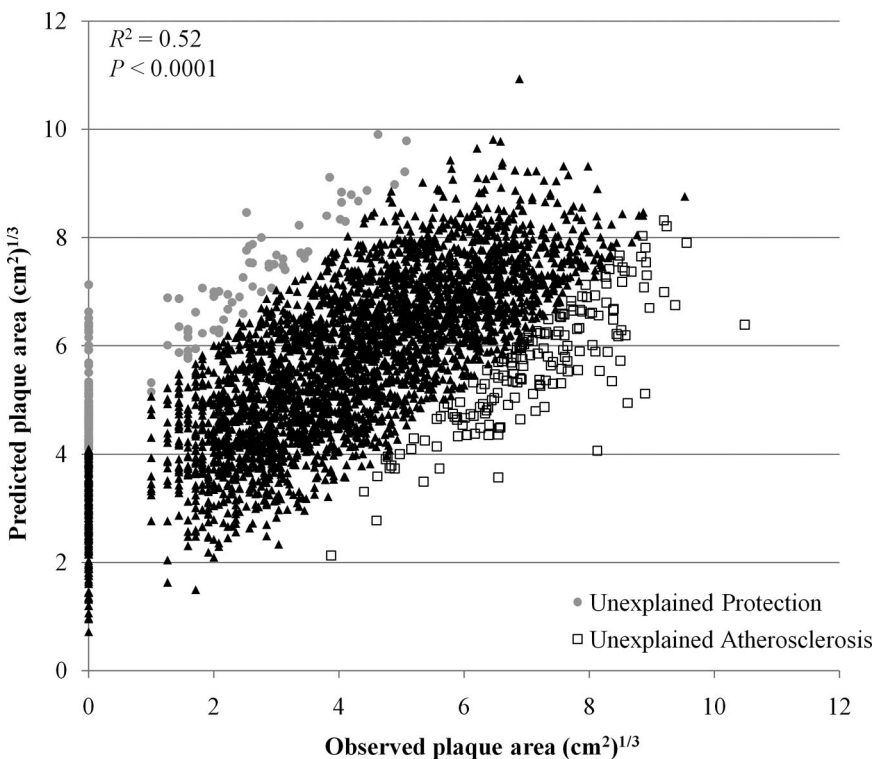
and the overall model explained 52% of plaque area variation ( $R^2=0.522$ ;  $F_9=422$ ;  $P<0.0001$ ; Table 2). The relation between the predicted and observed plaque area is shown in Figure 1. The bottom 5% of regression residuals, or the patients furthest below the regression line, were defined as those with “unexplained atherosclerosis,” and the top 5% of regression residuals, or the patients furthest above the regression line, were defined as those with “unexplained protection.” There was no significant difference in the characteristics of the individuals falling into these 2 groups, with the exception of observed plaque area (Table 1).

### Power Calculations

For a comparison of targeted genotyping of the extremes of unexplained variation versus genotyping the entire population, power calculations of both approaches were performed.

In the context of a genetic factor with a minor allele frequency of 30%, genotyping 250 individuals from the extremes of unexplained plaque area (500 total) yields 93% power if the genetic factor explains 2% of trait variation. If individuals are selected from across the population to be genotyped, 4 times the number of individuals is required to obtain the same level of power (Table 3;  $n=2000$  for population versus 500 for extremes).

Power to detect association with unexplained plaque area using either the extreme selection or the population-based approach decreased with reductions in the expected effect size and with reductions in the minor allele frequency of the responsible variant (Figure 2). The relative performance of the extreme selection approach improves with a reduction in minor allele frequency to 20%, because it maintains 80% power, whereas the power of 2000 population-based samples



**Figure 1.** Distribution of ultrasound-measured 2-dimensional plaque area compared with plaque area predicted by multiple regression of age, sex, systolic and diastolic blood pressure, diabetes status, smoking (pack-years), HDL cholesterol, and treatment of blood pressure and cholesterol as independent variables. The residual score from the multiple regression, or the difference between the observed plaque area and the plaque area predicted by the independent variables, is computed as a quantitative trait. The black triangles represent the 90% of individuals whose total plaque area is best explained by traditional risk factors. Individuals represented by white squares are those with “unexplained atherosclerosis,” forming the bottom 5% of regression residual scores. Individuals represented by gray circles are those with “unexplained protection,” forming the top 5% of regression residual scores.

**Table 3. Comparison of Power to Detect Genome-Wide Associations in a Population Sample and the Extremes of Unexplained Atherosclerosis**

Parameter	Top and Bottom 5% of "Unexplained Atherosclerosis"		Population Sample	
Trait type	Discrete		Continuous	
Analysis type	$\chi^2$		Regression	
N	500 (250 "unexplained" + 250 "protected")		2000	
$\alpha$	$5 \times 10^{-7}$		$5 \times 10^{-7}$	
Minor allele frequency	0.30		0.30	
Percent of variation explained	1	2	1	2
Effect size	Odds ratio of 1.9	Odds ratio of 2.5	15% of 1 SD/allele	22% of 1 SD/allele
Power	33%	93%	25%	92%

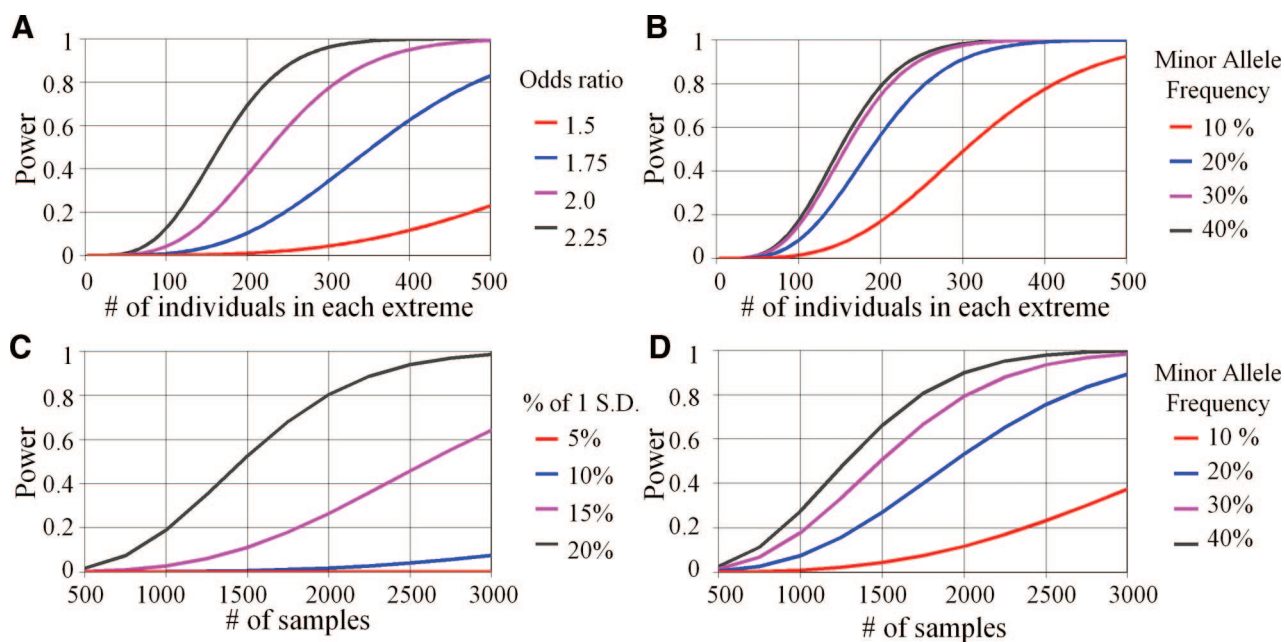
$\alpha$  indicates type I error rate.

falls to 55%. If the effect of the genetic variant is reduced to the lower level of most currently identified genetic variants, explaining only 1% of trait variation, genotyping 250 individuals from the extremes of unexplained plaque area (500 total) yields 34% power, whereas the population-based approach has 25% power with 2000 genotyped individuals. To obtain 90% power in this context, the extreme selection approach requires genotyping of 940 individuals (470 in each extreme), whereas 4169 genotypes are required with the population-based approach.

## Discussion

To obtain an equally powered study, the number needed to genotype is 4 times greater when using a population-based approach compared with targeted genotyping of individuals at the extremes of quantitative trait variation. Using previous knowledge of CVD risk factors, we are able to identify individuals whose degree of atherosclerotic plaque area is significantly greater or less than predicted, representing "unexplained atherosclerosis" and "unexplained protection" respectively, to efficiently identify novel genetic associations with the atherosclerotic process.

By targeting genotyping to individuals with extremes of unexplained atherosclerosis, the monetary cost of the study can be markedly reduced without loss of power. The primary objection would be that the reduction in genotyping cost needs to be balanced with the increased cost of screening to obtain individuals in the extremes of the distribution, especially in the context of falling genotyping costs. The number of individuals screened is an important consideration because simply sampling further into the distribution, for instance, taking the top and bottom 10 percentiles, increases the sample size but reduces the expected difference in allele frequency between the 2 groups, thus reducing power. However, as collection of traditional CVD risk factors and determination of plaque area are standard care at our Vascular Prevention Clinics, selection of individuals at the extremes of the distribution incurs no additional subject ascertainment cost. Clinicians continue to play an increasingly important role in the study of complex disease genetics, and the proliferation of electronic medical records in patient care should further assist in similar ascertainment schemes to efficiently obtain maxi-



**Figure 2.** A, Sample power examining the extremes of unexplained plaque area, as a function of expected allelic odds ratio. Minor allele frequency is set at 30%. B, Sample power examining the extremes of unexplained plaque area as a function of minor allele frequency. Expected allelic odds ratio between extremes is set at 2.3, corresponding to a variant explaining 1.69% of trait variation. C, Sample power examining unexplained plaque area across the population as a function of expected effect size. Minor allele frequency is set at 30%. D, Sample power examining unexplained plaque area across the population as a function of minor allele frequency. Expected effect is set at 20% of 1 SD, corresponding to a variant explaining 1.69% of trait variation.

mally powered studies of a multitude of complex quantitative phenotypes.

For the purpose of studying atherosclerosis per se, measurement of carotid plaque may afford an advantage over studies of clinical end points, such as traditional case-control studies contrasting those with premature myocardial infarction versus controls. Within the control group, there may be individuals with extensive atherosclerosis who have not yet experienced an event. Moreover, significant findings will likely point to factors affecting mechanisms of atherogenesis that are not captured by assessment of traditional epidemiological risk factors, such as plaque rupture and thrombosis. By measuring plaque area and adjusting for traditional risk factors, the approach described here offers great potential to uncover therapeutic targets related to the burden of atherosclerosis per se.

Two contrasting hypotheses have been presented to explain the distribution of quantitative traits in the population as a whole—the common disease, common variant hypothesis<sup>31</sup> and the common disease, rare variant hypothesis.<sup>32,33</sup> In common disease, common variant, a large number of variants with appreciable frequencies in the population but with relatively small effects and low penetrance cumulatively affect the quantitative trait. Conversely, the common disease, rare variant hypothesis suggests that many different rare variants, each with variable effect size and penetrance, are responsible for the distribution of the quantitative trait. GWAS have identified hundreds of single-nucleotide polymorphisms (SNPs) associated with variation in quantitative traits, and deep resequencing has successfully identified an excess (or an accumulation) of individually rare missense mutations in individuals within the extreme of multiple quantitative traits. For example, genes that contain common SNPs associated with normolipidemic variation in plasma triglycerides<sup>34–36</sup> also contain common SNPs<sup>37</sup> and rare mutations<sup>38</sup> associated with hypertriglyceridemia.<sup>39</sup> Similarly, SNPs associated with common variation in body mass index are also overrepresented in individuals with extreme obesity.<sup>40</sup> Investigations continue on the basis of both the common disease, common variant and common disease, rare variant hypotheses, and it is becoming increasingly apparent that both common and rare alleles will play a role in explaining the distribution of a quantitative trait.<sup>8,33</sup>

A recent deep resequencing study of blood pressure questioned the validity of extreme sampling to detect associations by comparing the significance obtained when testing the number of rare variants in the top and bottom 10% of the population (bottom 10%,  $n=323$ ) with the significance obtained when testing the number of rare variants in the top and bottom halves of the population (bottom 50%,  $n=1672$ ).<sup>41</sup> However, additional blood pressure screening to increase the number of participants for sequencing in the extreme tails of blood pressure should provide a greater increment in power than moving into the middle of the distribution to add additional samples. Indeed, if the allele frequencies in the top and bottom 10% of the population remained with additional sequencing in the extremes, similar levels of significance as observed in the whole population would be expected with less than half the number of individuals sequenced. Another study

**Table 4. Advantages and Disadvantages of Population-Based and Targeted Genotyping Designs**

Sampling Strategy	Advantages	Disadvantages
Population based	Uses all data points Can study a range of phenotypes	Cost and time of genotyping Majority of data points do not contribute signal
Targeting extremes	Less genotyping required Increased study efficiency	Upfront phenotype screening requirement Limited to selected phenotype

that examined blood pressure used the increase in power afforded through an extreme sampling approach to study gene-gene and gene-sex interactions.<sup>42</sup>

Investigators should be aware of several potential limitations of the extreme sampling approach (Table 4). Owing to the density of modern genotyping arrays and advances in imputation methodologies, it is assumed that either the allele causative for the quantitative trait is assayed or imputed directly or the causative allele is in tight linkage disequilibrium with 1 of the assayed or imputed alleles. Power will substantially decrease as the degree of linkage disequilibrium between the assayed and causative variant decreases,<sup>12</sup> but this is a limitation of both study designs. A spectrum of minor allele frequencies should be queried because maximal power is attained when the frequency of the assayed allele is close to the frequency of the causative allele.<sup>12</sup> A noted limitation of sampling from the extremes of a quantitative trait distribution is that the sample is only valid for the trait for which the sample was selected. However, the selected sample can be interrogated for variants of different frequencies and sizes, including both rare or common variants and single-nucleotide changes or copy number variation. Using the extremes of unexplained variation as the quantitative trait would be suitable only if a solid knowledge regarding nongenetic risk factors is present. Perhaps the largest limitation of the extreme selection approach is the assumption of a homogeneous effect for the tested genetic variant throughout the distribution of the trait. This assumption could be rendered false by gene-gene or gene-environment interactions, causing the effect of the tested genetic variant to change on the basis of additional factors.

In conclusion, we present a methodology, adaptable to GWAS of other complex phenotypes, for identifying novel loci involved in CVD by targeting genotyping toward individuals with extreme atherosclerosis unexplained by traditional risk factors. Thus, a study can be designed and executed with maximal efficiency for identifying genetic variants working through novel pathways.

### Sources of Funding

This work was supported by the Canadian Institutes of Health Research (MOP-13430, MOP-79523, CTP-79853), the Heart and Stroke Foundation of Ontario (NA-4990, NA-6059, NA-6018, NA-5912, T-2956, T-5017, T-5704, T-6018, and PRG-4854), Genome Canada through Ontario Genomics Institute, and the Pfizer Jean Davignon Distinguished Cardiovascular and Metabolic Research Award. Dr Schork was supported, in part, by the National Institutes

of Health grants U19 AG023122-01, MH078151-01A1, N01 MH22005, U01 DA024417-01, and P50 MH081755-01; the Scripps Translational Sciences Institute Clinical Translational Science Award (U54 RR0252204-01); the Price Foundation; and the Scripps Genomic Medicine. M.B. Lanktree was supported by the Canadian Institutes of Health Research MD/PhD Studentship Award and the University of Western Ontario MD/PhD Program. M.B. Lanktree is a Canadian Institutes of Health Research Fellow in Vascular Research. Dr Hegele is a Career Investigator of the Heart and Stroke Foundation of Ontario and holds the Edith Schulich Vinet Canada Research Chair (Tier I) in Human Genetics, the Martha G. Blackburn Chair in Cardiovascular Research, and the Jacob J. Wolfe Distinguished Medical Research Chair at the University of Western Ontario.

## Disclosures

None.

## References

1. Yusuf S, Reddy S, Ounpuu S, Anand S. Global burden of cardiovascular diseases, I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization. *Circulation*. 2001;104:2746–2753.
2. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health Nations Health*. 1951;41:279–281.
3. Helgadottir A, Thorleifsson G, Manolescu A, Gretarsdottir S, Blondal T, Jonasdottir A, Sigurdsson A, Baker A, Palsson A, Masson G, Gudbjartsson DF, Magnusson KP, Andersen K, Levey AI, Backman VM, Matthiasdottir S, Jonsdottir T, Palsson S, Einarsdottir H, Gunnarsdottir S, Gylfason A, Vaccarino V, Hooper WC, Reilly MP, Granger CB, Austin H, Rader DJ, Shah SH, Quyyumi AA, Gulcher JR, Thorgerisson G, Thorsteinsdottir U, Kong A, Stefansson K. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491–1493.
4. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316:1488–1491.
5. Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, König IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357:443–453.
6. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661–678.
7. Lanktree M, Oh J, Hegele RA. Genetic testing for atherosclerosis risk: inevitability or pipe dream? *Can J Cardiol*. 2008;24:851–854.
8. Plomin R, Haworth CM, Davis OS. Common disorders are quantitative traits. *Nat Rev Genet*. 2009;10:872–878.
9. Lander ES, Botstein D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*. 1989;121:185–199.
10. Carey G, Williamson J. Linkage analysis of quantitative traits: increased power by using selected samples. *Am J Hum Genet*. 1991;49:786–796.
11. Risch N, Zhang H. Extreme discordant sib pairs for mapping quantitative trait loci in humans. *Science*. 1995;268:1584–1589.
12. Schork NJ, Nath SK, Fallin D, Chakravarti A. Linkage disequilibrium analysis of biallelic DNA markers, human quantitative trait loci, and threshold-defined case and control subjects. *Am J Hum Genet*. 2000;67:1208–1218.
13. Abecasis GR, Cookson WO, Cardon LR. The power to detect linkage disequilibrium with quantitative traits in selected samples. *Am J Hum Genet*. 2001;68:1463–1474.
14. Tenesa A, Knott SA, Carothers AD, Visscher PM. Power of linkage disequilibrium mapping to detect a quantitative trait locus (QTL) in selected samples of unrelated individuals. *Ann Hum Genet*. 2003;67:557–566.
15. Jawaid A, Bader JS, Purcell S, Cherny SS, Sham P. Optimal selection strategies for QTL mapping using pooled DNA samples. *Eur J Hum Genet*. 2002;10:125–132.
16. Pollex RL, Spence JD, House AA, Fenster A, Hanley AJ, Zinman B, Harris SB, Hegele RA. A comparison of ultrasound measurements to assess carotid atherosclerosis development in subjects with and without type 2 diabetes. *Cardiovasc Ultrasound*. 2005;3:15.
17. de Labriolle A, Mohty D, Pacouret G, Giraudeau B, Fichet J, Fremont B, Fauchier L, Charbonnier B, Arbeille P. Comparison of degree of stenosis and plaque volume for the assessment of carotid atherosclerosis using 2-D ultrasound. *Ultrasound Med Biol*. 2009;35:1436–1442.
18. Spence JD, Hegele RA. Noninvasive phenotypes of atherosclerosis: similar windows but different views. *Stroke*. 2004;35:649–653.
19. Al-Shali K, House AA, Hanley AJ, Khan HM, Harris SB, Mamakesick M, Zinman B, Fenster A, Spence JD, Hegele RA. Differences between carotid wall morphological phenotypes measured by ultrasound in one, two and three dimensions. *Atherosclerosis*. 2005;178:319–325.
20. Johnsen SH, Mathiesen EB, Joakimsen O, Stensland E, Wilsgaard T, Lochen ML, Njolstad I, Arnesen E. Carotid atherosclerosis is a stronger predictor of myocardial infarction in women than in men: a 6-year follow-up study of 6226 persons: the Tromso Study. *Stroke*. 2007;38:2873–2880.
21. Ebrahim S, Papacosta O, Whincup P, Wannamethee G, Walker M, Nicolaides AN, Dhanjal S, Griffin M, Belcaro G, Rumley A, Lowe GD. Carotid plaque, intima media thickness, cardiovascular risk factors, and prevalent cardiovascular disease in men and women: the British Regional Heart Study. *Stroke*. 1999;30:841–850.
22. Spence JD, Barnett PA, Bulman DE, Hegele RA. An approach to ascertain probands with a non-traditional risk factor for carotid atherosclerosis. *Atherosclerosis*. 1999;144:429–434.
23. Spence JD, Eliasziw M, DiCicco M, Hackam DG, Galil R, Lohmann T. Carotid plaque area: a tool for targeting and evaluating vascular preventive therapy. *Stroke*. 2002;33:2916–2922.
24. Dupont WD, Plummer WD Jr. Power and sample size calculations for studies involving linear regression. *Control Clin Trials*. 1998;19:589–601.
25. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003;19:149–150.
26. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet*. 2006;7:781–791.
27. Wallace C, Chapman JM, Clayton DG. Improved power offered by a score test for linkage disequilibrium mapping of quantitative-trait loci by selective genotyping. *Am J Hum Genet*. 2006;78:498–504.
28. Huang BE, Lin DY. Efficient association mapping of quantitative trait loci with selective genotyping. *Am J Hum Genet*. 2007;80:567–576.
29. Manichaikul A, Palmer AA, Sen S, Broman KW. Significance thresholds for quantitative trait locus mapping under selective genotyping. *Genetics*. 2007;177:1963–1966.
30. Tenesa A, Visscher PM, Carothers AD, Knott SA. Mapping quantitative trait loci using linkage disequilibrium: marker- versus trait-based methods. *Behav Genet*. 2005;35:219–228.
31. Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet*. 2001;17:502–510.
32. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*. 2001;69:124–137.
33. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev*. 2009;19:212–219.
34. Aulchenko YS, Ripatti S, Lindqvist I, Boomsma D, Heid IM, Pramstaller PP, Penninx BW, Janssens AC, Wilson JF, Spector T, Martin NG, Pedersen NL, Kyvik KO, Kaprio J, Hofman A, Freimer NB, Jarvelin MR, Gyllenstein U, Campbell H, Rudan I, Johansson A, Marroni F, Hayward C, Vitart V, Jonasson I, Pattaro C, Wright A, Hastie N, Pichler I, Hicks AA, Falchi M, Willemsen G, Hottenga JJ, de Geus EJ, Montgomery GW, Whitfield J, Magnusson P, Saharinen J, Perola M, Silander K, Isaacs A, Sijbrands EJ, Uitterlinden AG, Witteman JC, Oostra BA, Elliott P, Ruokonen A, Sabatti C, Gieger C, Meitinger T, Kronenberg F, Doring A, Wichmann HE, Smit JH, McCarthy MI, van Duijn CM, Peltonen L. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet*. 2009;41:47–55.
35. Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanaka T, Voight BF, Bonnycastle LL, Jackson AU, Crawford G, Surti A, Guiducci C, Burt NP, Parish S, Clarke R, Zelenika D, Kubalanza KA, Morken MA, Scott LJ, Stringham HM, Galan P, Swift AJ, Kuusisto J, Bergman RN, Sundvall J, Laakso M, Ferrucci L, Scheet P, Sanna S, Uda M, Yang Q, Lunetta KL, Dupuis J, de Bakker PI, O'Donnell CJ, Chambers JC, Kooner JS, Hercberg S, Meneton

- P, Lakatta EG, Scuteri A, Schlessinger D, Tuomilehto J, Collins FS, Groop L, Altshuler D, Collins R, Lathrop GM, Melander O, Salomaa V, Peltonen L, Orho-Melander M, Ordovas JM, Boehnke M, Abecasis GR, Mohlke KL, Cupples LA. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet.* 2009;41:56–65.
36. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, Brodsky J, Jones CG, Zaitlen NA, Varilo T, Kaakinen M, Sovio U, Ruokonen A, Laitinen J, Jakkula E, Coin L, Hoggart C, Collins A, Turunen H, Gabriel S, Elliot P, McCarthy MI, Daly MJ, Jarvelin MR, Freimer NB, Peltonen L. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet.* 2009;41:35–46.
  37. Hegele RA, Ban MR, Hsueh N, Kennedy BA, Cao H, Zou GY, Anand S, Yusuf S, Huff MW, Wang J. A polygenic basis for four classical Fredrickson hyperlipoproteinemia phenotypes that are characterized by hypertriglyceridemia. *Hum Mol Genet.* 2009;18:4189–4194.
  38. Wang J, Cao H, Ban MR, Kennedy BA, Zhu S, Anand S, Yusuf S, Pollex RL, Hegele RA. Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). *Arterioscler Thromb Vasc Biol.* 2007;27:2450–2455.
  39. Hegele RA. Plasma lipoproteins: genetic influences and clinical implications. *Nat Rev Genet.* 2009;10:109–121.
  40. Cotsapas C, Speliotes EK, Hatoum IJ, Greenawalt DM, Dobrin R, Lum PY, Suver C, Chudin E, Kemp D, Reitman M, Voight BF, Neale BM, Schadt EE, Hirschhorn JN, Kaplan LM, Daly MJ. Common body mass index-associated variants confer risk of extreme obesity. *Hum Mol Genet.* 2009;18:3502–3507.
  41. Ji W, Foo JN, O’Roak BJ, Zhao H, Larson MG, Simon DB, Newton-Cheh C, State MW, Levy D, Lifton RP. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet.* 2008;40:592–599.
  42. Rana BK, Insel PA, Payne SH, Abel K, Beutler E, Ziegler MG, Schork NJ, O’Connor DT. Population-based sample reveals gene-gender interactions in blood pressure in White Americans. *Hypertension.* 2007;49:96–106.

---

KEY WORDS: genetics ■ cardiovascular disease ■ atherosclerosis ■ risk factors ■ stroke ■ genome-wide association studies ■ ultrasound

**Extremes of Unexplained Variation as a Phenotype: An Efficient Approach for  
Genome-Wide Association Studies of Cardiovascular Disease**

Matthew B. Lanktree, Robert A. Hegele, Nicholas J. Schork and J. David Spence

*Circ Cardiovasc Genet.* 2010;3:215-221

doi: 10.1161/CIRCGENETICS.109.934505

*Circulation: Cardiovascular Genetics* is published by the American Heart Association, 7272 Greenville Avenue,  
Dallas, TX 75231

Copyright © 2010 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/3/2/215>

**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/>