ARTICLE

Are common disease susceptibility alleles the same in outbred and founder populations?

Dina L Newman*[,1,5], Sabine Hoffjan[1], Catherine Bourgain[1,6], Mark Abney[1], Raluca I Nicolae[1], Elle T Profits[1], Michael A Grow[2], Karen Walker[2], Lori Steiner[2], Rodney Parry[3], Rebecca Reynolds[2,7], Mary Sara McPeek[1,4], Suzanne Cheng[2] and Carole Ober[1]

[1]Department of Human Genetics, University of Chicago, Chicago, IL, USA; [2]Roche Molecular Systems, Alameda, CA, USA; [3]Department of Medicine, University of South Dakota, Sioux Falls, SD, USA; [4]Department of Statistics, University of Chicago, Chicago, IL, USA; [5]Department of Biological Sciences, Rochester Institute of Technology, Rochester, NY; [6]INSERM U535, Villejuif, France; [7]Bayer Corporation, E Walpole, MA, USA

Founder populations have been the subjects of complex disease studies because of their decreased genetic heterogeneity, increased linkage disequilibrium and more homogeneous environmental exposures. However, it is possible that disease alleles identified in founder populations may not contribute significantly to susceptibility in outbred populations. In this study we examine the Hutterites, a founder population of European descent, for 103 polymorphisms in 66 genes that are candidates for cardiovascular or inflammatory diseases. We compare the frequencies of alleles at these loci in the Hutterites to their frequencies in outbred European-American populations and test for associations with cardiovascular disease-associated phenotypes in the Hutterites. We show that alleles at these loci are found at similar frequencies in the Hutterites and in outbred populations. In addition, we report associations between 39 alleles or haplotypes and cardiovascular disease phenotypes ($P < 0.05$), with five loci remaining significant after adjusting for multiple comparisons. These data indicate that this founder population is informative and offers considerable advantages for genetic studies of common complex diseases.

Introduction

Many founder populations are currently the subjects of genetic studies of common diseases. The advantages of these populations are well known.[1–4] As many of the diseases under investigation in founder populations are major public health concerns, the primary rationale for these studies is to identify genetic factors that are also important in the general (outbred) population. However, it is possible that susceptibility alleles present in founder populations may not contribute to susceptibility in outbred individuals, as is the case for many Mendelian disorders.[5–8] Although identifying mutations associated with common diseases only in founder populations could identify pathways important for disease and possibly new therapeutic approaches, there would be considerably less enthusiasm for gene discovery in founder populations if their susceptibility alleles were population-specific.

In order to assess the utility of founder populations for identifying alleles that are associated with susceptibility to common diseases in outbred populations, we conducted a survey of polymorphisms in 66 candidate genes for cardiovascular disease-related phenotypes in the Hutterites

*Correspondence: Dr DL Newman, Department of Biological Sciences, Rochester Institute of Technology, 85 Lomb Memorial Dr., Rochester, NY 14623, USA. Tel: +1 585 475 4482; Fax: +1 585 475 5766; E-mail: dnewman@mail.rit.edu

and compared the results to published studies of the same variants in outbred populations. The Hutterites are particularly well suited for this study because the more than 35 000 extant Hutterites are derived from fewer than 90 ancestors who were born in Europe in the early 17th to early 18th century.[9] The Hutterites of South Dakota, the subjects of our study, are descendants of only 64 of these 90 ancestors.[10] The small number of Hutterite founders increases the likelihood that rare alleles could have risen to modest frequencies due to drift, making it possible that alleles that are very rare in outbred populations contribute significantly to disease risk in the Hutterites. Conversely, if rare variants contribute significantly to common diseases in outbred populations,[11] these variants may not be present in the Hutterites because either they were not carried by one of the founders or were lost due to drift.

In this investigation, we asked three questions: (1) Are variants thought to be associated with cardiovascular disease-related phenotypes in outbred populations present in the Hutterites? (2) Are allele frequencies at these loci similar in the Hutterites and outbred populations? (3) Are the alleles with evidence for association with disease in the Hutterites the same as in outbred populations? Our results indicate that founder populations are indeed informative for genetic studies of common diseases.

Materials and methods
Founder and outbred population samples
Measurements of blood pressure and serum lipids were determined as part of a larger study of complex trait genetics in the Hutterites; the details of these studies were described previously.[4] Briefly, phenotype information and blood samples for DNA extraction were collected from all individuals age 6 years or older during field trips to nine Hutterite colonies in 1993–1994 and 1996–1997. The 813 South Dakota Hutterites in our studies are related to each other in a 13-generation, 1623-person pedigree[12] and have been the subjects of genome-screens for asthma and atopy,[10,13,14] fasting serum insulin,[15] and serum triglyceride[16] susceptibility loci.

We used two published data sets to assess allele frequencies in outbred populations: a cohort of 142 Caucasian individuals at high risk for cardiovascular disease recruited from the San Francisco Bay area (UCSF)[17] and a cohort of 207 Caucasian children at high risk for developing asthma or atopy (COAST).[18]

Genotyping
Genotyping was performed with two linear array panels developed by Roche Molecular Systems, Inc (Alameda, CA, USA). Genotyping accuracy by this assay is estimated at >99% (S. Cheng *et al.*, personal communication). The first panel, hereafter referred to as the 'CVD' set, consisted of 65 biallelic polymorphisms in 36 genes. These polymorphisms

were selected as candidate markers for cardiovascular disease based on published reports in outbred populations. An earlier version of this panel with 35 polymorphisms in 15 genes, 26 of them overlapping with our panel, was previously described.[17] Four single-nucleotide polymorphisms (SNPs) in two genes (*CETP* intron 14 +1, +3 and asp442gly,[19] and *TNF* -244[20]) were excluded because they were originally identified in non-European populations and were very rare (<0.5%) in European populations (S. Cheng *et al.*, unpublished). The second panel, hereafter referred to as the 'INF' set, included 50 biallelic polymorphisms in 35 genes. These polymorphisms were selected as candidates for inflammatory diseases. Although not all of the SNPs included in the INF panel had been studied in patients with cardiovascular disease or associated phenotypes, we considered them to be good candidates because of the important role of inflammation in cardiovascular disease (reviewed by Libby[21]). Eight SNPs in five genes were present in both sets. The final study set included 103 polymorphisms in 66 genes. In total, 721 Hutterites were genotyped for the CVD panel and 794 for the INF panel; 692 individuals were genotyped for both. Error checking was performed by comparing blind duplicates, utilizing the known pedigree structure to detect Mendelian errors and testing for Hardy–Weinberg equilibrium.

Statistical analyses
We compared the allele frequencies in the populations using a case–control (CC) association test that takes into account both the relatedness between individuals and inbreeding.[22] Instead of contrasting the frequency of an allele between a group of cases and a group of controls, we contrasted the frequency of the allele between a sample of Hutterites and a sample of outbred individuals. Because we did not have access to the raw data of the outbred samples, but only to their size and allele frequency, we created dummy samples of genotypes corresponding to these sizes and frequencies, assuming Hardy–Weinberg equilibrium. Thus, the data did not allow us to contrast genotype frequencies between the Hutterite sample and the outbred samples. We note however that there was no departure from Hardy–Weinberg equilibrium in the Hutterite sample once the pedigree structure was taken into account. Differences in the genotype frequencies between the two samples are thus due to the differences in allele frequencies and to inbreeding, which is the same for every locus.

We also compared allele frequencies in the populations by calculating F_{ST} with the program GENDIST in the PHYLIP software package.[23]

We tested for associations in the Hutterites using 89 polymorphic markers (multiple SNPs in the *APOE*, *GC* and *LTA* genes and the *CETP* promoter were analyzed as haplotypes) and six quantitative phenotypes associated with cardiovascular disease: systolic blood pressure (SBP),

diastolic blood pressure (DBP), low-density lipoprotein–cholesterol (LDL), high-density lipoprotein–cholesterol (HDL), triglycerides (TG), and lipoprotein (a) [Lp(a)]. The descriptions of these phenotypes in the Hutterites are shown in supplementary Table 1. To test for association, we used two approaches that take into account the relatedness between all pairs of individuals in our sample.[15,22] The first test, called the general two-allele model (GTAM), allows for a quantitative trait to follow any two-allele model (including dominant, recessive, and additive).[15] In these analyses, age and gender were included as covariates. In order to eliminate excess skewness and kurtosis, a transformation was applied to each trait, resulting in a phenotype that closely matched a normal distribution. Cube root was used for LDL and HDL, $\log_{10}$ was used for TG, Lp(a) and SBP, and square root was used for DBP. The second approach utilized a case–control design whereby the cases were individuals in our sample in the highest quartile for LDL, Lp(a), and TG and the lowest quartile for HDL, while the controls were the individuals in the lowest quartile for LDL, Lp(a), and TG and the highest quartile for HDL, after adjusting for age and gender (113, 93, 121, and 121 cases and an equivalent number of controls for each phenotype, respectively). For the HTN phenotype, 110 cases (adults taking antihypertensive medication or stage 1 or higher hypertension as defined by the World Health Organization-International Society of Hypertension;[24] children with high blood pressure as defined by Council on Cardiovascular Disease in the Young[25]) were compared with 84 controls (age and gender matched normotensive controls for all individuals under age 40 years, plus all normotensive adults over age 40 years). To test for association with these categorical phenotypes, we used two CC association tests (the quasi-likelihood score test and the corrected χ^2-test).[22]

Results

Frequencies of the minor alleles in the Hutterites are shown for all three populations in Supplementary Table II (Hutterite and COAST data are also on our websites). Ten polymorphisms were monomorphic in the Hutterites (supplementary Table 2; shown in bold). All of these SNPs had low minor allele frequencies (<0.07) or were monomorphic in the UCSF or COAST samples or in other published studies.[17,26–28] Thus, 93 of the 103 (90%) SNPs that were associated with disease or identified as candidates in outbred European and European–American populations were present in the Hutterites. Among the more common SNPs (minor allele frequencies >0.10) in the outbred population samples, all (58 of 58) were present in the Hutterites, whereas among SNPs with minor allele frequencies <0.10 in the outbred population samples, only 35 of 45 (78%) were present in the Hutterites. Thus,

approximately 25% of the less common alleles (<0.10) were either not present in the Hutterite founders or were lost due to drift.

The correlation between minor allele frequencies in the Hutterites and the COAST and UCSF samples are shown in Figure 1a and b, respectively. Among the 50 polymorphisms in the INF panel, allele frequencies in the Hutterite and COAST samples were strongly correlated ($R = 0.854$), as were the 26 polymorphisms from the CVD panel that could be compared between the Hutterite and UCSF samples ($R = 0.939$). The average difference between minor allele frequencies in the Hutterites and outbred samples was 0.066 (range 0–0.210). Overall, however, among alleles present in the Hutterites, the frequencies are similar to those observed in outbred Caucasian populations. Only three of 75 comparisons (4%) were significantly different ($P < 0.05$ for *ACE* intron 16 ins/del, *CTLA4* −318 C/T and *IL5RA* −80 G/A), as would be expected by chance. Additionally, in comparisons of Hutterite *vs* UCSF allele frequencies, $F_{ST} = 0.0191$, and of
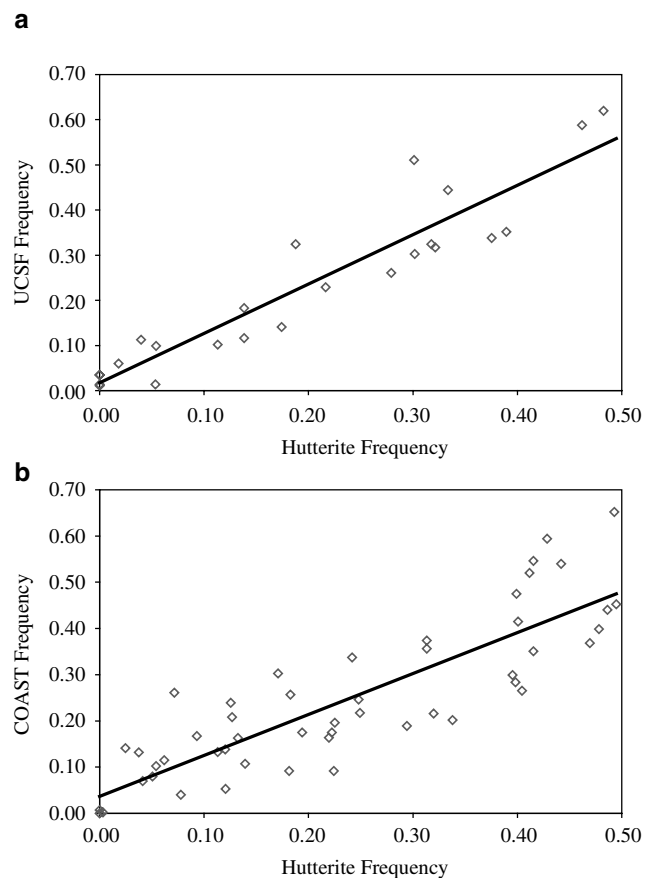
Figure 1 Comparison of allele frequencies in Hutterites and outbred populations. (**a**) Hutterites compared to UCSF for 26 biallelic polymorphisms from the CVD set. $R = 0.939$. (**b**) Hutterites compared to COAST for 50 biallelic polymorphisms from the INF set. $R = 0.854$.

Hutterite *vs* COAST, $F_{ST} = 0.0230$. These F_{ST} values are similar to those obtained in comparisons between European populations.[29,30] Thus, the Hutterites do not appear to have significantly diverged from other European-American population samples.

We found 39 significant associations ($P < 0.05$) of lipid levels or blood pressure with 29 different polymorphisms using the GTAM (supplementary Table 3). Four of these associations were highly significant ($P < 0.001$) and remained significant even after adjusting for 89 loci by a permutation test.[15] We found 56 significant associations ($P < 0.05$) of abnormal lipid levels or hypertension with 42 different polymorphisms using the CC tests (supplementary Table 4). The same four alleles that were highly significant by the GTAM plus one more were highly significantly associated by CC tests and remained significant after adjusting for testing of polymorphisms in 63 genes and two CC tests using a Bonferroni correction (adjusted critical P-value $= 0.0004$). These include the *APOE* ε2 allele with low LDL levels (GTAM $P = 1.0 \times 10^{-3}$; CC $P = 6.0 \times 10^{-5}$), the *LDLR* NcoI+ allele with high LDL levels (CC $P = 2.2 \times 10^{-4}$), the *CETP* −631C/−629A haplotype with high HDL levels (GTAM $P = 1.3 \times 10^{-5}$; CC $P = 3.9 \times 10^{-5}$), the *APOC3* 3175G allele with high TG levels (GTAM $P = 9.3 \times 10^{-5}$; CC $P = 1.6 \times 10^{-5}$), and the *LPA* +93T allele with high Lp(a) levels (GTAM $P = 2.9 \times 10^{-8}$; CC $P = 2.9 \times 10^{-4}$).

These five associated variants had frequencies ranging from 0.02 (*APOE*) to 0.56 (*CETP*) in the Hutterites. The average minor allele frequency for all 49 variants that showed association with one or more phenotype in the Hutterites was 0.24 (range 0.02–0.49), which was the same as the average allele frequency for the 41 loci that were not associated with these phenotypes in the Hutterites (average minor allele frequency 0.25, range 0.03–0.49).

Discussion

Although classical population genetic theory predicts that human founder populations will have allele frequencies on average similar to those in the ancestral population, to our knowledge this is the first study to provide empiric evidence of this prediction for both biallelic markers and loci selected because of their potential role as susceptibility alleles for common diseases. In this investigation, all of the variants identified in outbred populations, except for approximately one fourth of low frequency (<0.10) variants, were present in the Hutterites. These results are consistent with results of a previous analysis of microsatellite alleles in the Hutterites and CEPH families.[13] In that study of 164 microsatellite alleles, all alleles that were present in the CEPH families but absent in the Hutterites had frequencies <0.10 in the CEPH sample. Among those alleles that were present in the Hutterites, the frequencies were similar to those found in

the CEPH families.[13] Likewise, in this study SNP allele frequencies were similar in Hutterite and outbred populations. Thus, common alleles (>0.10) that are identified and associated with diseases in outbred populations should be present in the Hutterites and will often show similar patterns of association.

Owing to the complex etiology of cardiovascular disease and associated phenotypes, we would not expect to replicate all associations in the Hutterites, particularly those reported only in non-Caucasian populations or in only a single sample. However, we would expect to find some associations with related phenotypes, particularly with those that have been replicated in multiple samples. All five of the highly significant associations in the Hutterites have been reported in other populations. The *APOC3* 3175 polymorphism (a.k.a. 'the SstI polymorphism') is located in exon 4 in the 3′ untranslated region (3′UTR) of the gene encoding apolipoprotein C-III (apo C-III). Apo C-III is a component of triglyceride-rich lipoproteins, and numerous studies have demonstrated an association between high triglycerides levels and the rare (G) allele of this SNP (reviewed by Groenendijk *et al*.[31]), as reported in this study. Apolipoprotein E is the primary ligand for the LDL receptor, and the ε2 variant binds with significantly lower affinity to the receptor than either the ε3 or ε4 isoforms. It has been well established that the ε2 variant is associated with lower levels of cholesterol and LDL, and that carriers of the ε2 variant have a lower incidence of coronary heart disease (reviewed by Eichner *et al*.[32]). In the Hutterites, this variant was associated with significantly lower levels of LDL by the GTAM, whereas the ε4 variant was significantly associated with 'high LDL' in the CC study. More than 770 mutations have been discovered in the LDL receptor to date.[33,34] Mutations in LDLR have been associated with familial hypercholesterolemia, a disorder characterized by high levels of LDL, and variation in the gene also contributes significantly to LDL levels in the general population.[35–37] The NcoI site is a polymorphism in exon 18, where the '+' allele is on a haplotype that is associated with increased total and LDL-cholesterol,[38,39] consistent with our finding of an association with 'high LDL' in the Hutterites. Cholesteryl ester transfer protein (CETP) transfers cholesteryl esters from HDL to LDL and very low-density lipoprotein (VLDL), thus promoting the redistribution of lipids from anti-atherogenic lipoproteins to proatherogenic lipoproteins. Several polymorphisms in the coding region of CETP have been associated with CETP activity, HDL concentration, CHD and/or atherosclerosis (reviewed by Barter *et al*.[40]). In addition, the C allele at −629 in the promoter has been associated with high-CETP and low-HDL concentrations.[41,42] Here, we report significant associations between the −631C/−629A haplotype and high HDL by the GTAM and between the −631C/−629C haplotype and low HDL in the CC test. *LPA* encodes apolipoprotein(a), a protein

component of Lp(a), and variation in this gene is known to affect the concentration of Lp(a) in the blood. The +93 promoter polymorphism influences expression of *LPA*[43] and has been associated with serum Lp(a) levels in Japanese and Africans.[44,45] In the Hutterites, the +93T allele was very significantly associated with high Lp(a) levels by the GTAM and CC tests. Many other loci that have been associated with cardiovascular phenotypes in outbred populations were also associated with similar phenotypes in the Hutterites (eg, *AGTR1*, *AGT*, *LIPC*, *ACE*).

A number of novel associations with SNPs in the INF panel are also reported here, although none of these remained significant after adjusting for multiple comparisons and some may represent false positives. Nevertheless, alleles at some of these loci were associated with multiple phenotypes with $P < 0.01$ (eg, *TGFB1*, *IL13*, *IL6*, *CTLA4*) making them particularly intriguing candidates that may contribute to the inflammatory and immune processes that underlie atherosclerosis. Overall, among the polymorphic markers tested, 30 of 51 (59%) in the CVD panel and 24 of 47 (51%) in the INF panel showed some association with blood pressure or lipid levels in the Hutterites, indicating that many common disease-associated variants that are present in outbred populations also show evidence for association with related phenotypes in the Hutterites.

One question that cannot be addressed directly with these data is whether disease-associated variants first detected in the Hutterites are also present and associated with disease in outbred populations. However, data from our laboratory suggest that this is likely to be the case. For example, all SNPs that we have identified in the Hutterites have also been detected in outbred populations,[46–48] although our strategy for SNP detection in the Hutterites would not detect very low-frequency SNPs (less than approximately 0.02). However, because we would have less power to detect susceptibility alleles in this very low-frequency range, the disease susceptibility alleles that we identify in the Hutterites will likely be higher frequency, and therefore should be present in the general population. On the other hand, we do not expect all disease associations detected in the Hutterites to be replicated in all outbred populations. In particular, susceptibility to common diseases involves gene–gene and gene–environment interactions. The frequencies of interacting alleles differ among populations, and the Hutterites' unique lifestyle (eg, prohibition of cigarette smoking and uniformly high-fat, high-salt diet) may not be replicated in other populations. In addition, linkage disequilibrium (LD) is increased in the Hutterites and other founder populations. This should result in longer ancestral haplotypes on which disease susceptibility alleles reside, and a greater likelihood that an associated allele detected in the Hutterites is merely a marker for the true susceptibility allele on the extended haplotype. Because the ancestral haplotype is not likely to

extend as far in the outbred population, the association may not be replicated. It should be noted that characterizing LD among the SNPs was not the purpose of this study and indeed, we do not have the correct data or methods to assess this important question.

It is not possible at this time to predict whether most susceptibility alleles for common diseases can be identified in founder populations, because the proportional contributions of common vs rare variants to common diseases are still being debated.[11,49] Indeed, given the size of the Hutterite sample it is unlikely that we would have sufficient power to detect associations with alleles in the very low-frequency range, particularly if they have modest effects on susceptibility, as is expected for genes underlying complex phenotypes. However, our results indicate that populations like the Hutterites will be useful for identifying common variants that contribute to common diseases in the general (outbred) population and that interpretation of our results will not be limited to this unique population. On the other hand, the Hutterite's large and well-characterized genealogy, along with a naturally high fertility rate resulting in large family sizes, provide significant advantages for mapping complex trait genes compared with outbred populations. Further, the fact that association studies in young founder populations such as the Hutterites provide equivalent power with fewer markers and smaller sample sizes, and therefore lower costs, compared with outbred populations,[50] makes them well suited for identifying risk alleles for common diseases.

Electronic-database information

The URLs for data in this article are: www.genes.uchicago.edu/hutterite/inflasnps (INF panel on Hutterites), www.genes.uchicago.edu/hutterite/cvdsnps (CVD panel on Hutterites), and www.genes.uchicago.edu/coast/inflasnps (INF panel on COAST). The rs numbers in Supplementary Table II are from dbSNP: www.ncbi.nlm.nih.gov/SNP/.

References

1 Kruglyak L: Genetic isolates: separate but equal? *Proc Natl Acad Sci USA* 1999; **96**: 1170–1172.
2 Lander ES, Schork NJ: Genetic dissection of complex traits. *Science* 1994; **265**: 2037–2048.

3 Peltonen L: Positional cloning of disease genes: advantages of genetic isolates. *Hum Hered* 2000; **50**: 66–75.

4 Ober C, Abney M, McPeek MS: The genetic dissection of complex traits in a founder population. *Am J Hum Genet* 2001; **69**: 1068–1079.

5 McKusick VA, Hostetler JA, Egeland JA, Eldridge R: The distribution of certain genes in the Old Order Amish. *Cold Spring Harbor Symp Quant Biol* 1964; **XXIX**: 99–114.

6 Bowen P: Workshop on genetic disorders in the Hutterites – Edmonton, Canada, October 12-13, 1983. *Am J Med Genet* 1985; **22**: 449–451.

7 Kalaydjieva L, Gresham D, Calafell F: Genetic studies of the Roma (Gypsies): a review. *BMC Med Genet* 2001; **2**: 5.

8 Sipila K, Aula P: Database for the mutations of the Finnish disease heritage. *Hum Mutat* 2002; **19**: 16–22.

9 Martin AO: The founder effect in a human isolate: evolutionary implications. *Am J Phys Anthropol* 1970; **32**: 351–368.

10 Ober C, Tsalenko A, Parry R, Cox NJ: A second-generation genomewide screen for asthma-susceptibility alleles in a founder population. *Am J Hum Genet* 2000; **67**: 1154–1162.

11 Pritchard JK: Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001; **69**: 124–137.

12 Abney M, McPeek MS, Ober C: Estimation of variance components of quantitative traits in inbred populations. *Am J Hum Genet* 2000; **66**: 629–650.

13 Ober C, Cox NJ, Abney M *et al*: Genome-wide search for asthma susceptibility loci in a founder population. The Collaborative Study on the Genetics of Asthma. *Hum Mol Genet* 1998; **7**: 1393–1398.

14 Ober C, Tsalenko A, Willadsen S *et al*: Genome-wide screen for atopy susceptibility alleles in the Hutterites. *Clin Exp Allergy* 1999; **29** (Suppl 4): 11–15.

15 Abney M, Ober C, McPeek MS: Quantitative-trait homozygosity and association mapping and empirical genomewide significance in large, complex pedigrees: fasting serum-insulin level in the hutterites. *Am J Hum Genet* 2002; **70**: 920–934.

16 Newman DL, Abney M, Dytch H *et al*: Major loci influencing serum triglyceride levels on 2q14 and 9p21 localized by homozygosity-by-descent mapping in a large Hutterite pedigree. *Hum Mol Genet* 2003; **12**: 137–144.

17 Cheng S, Grow MA, Pallaud C *et al*: A multilocus genotyping assay for candidate markers of cardiovascular disease risk. *Genome Res* 1999; **9**: 936–949.

18 Hoffjan S, Nicolae D, Ostrovnaja I *et al*: Genetic variation in immunoregulatory pathways and atopic phenotypes in infancy. *J Allergy Clin Immunol* 2004; **113**: 511–518.

19 Inazu A, Jiang XC, Haraki T *et al*: Genetic cholesteryl ester transfer protein deficiency caused by two prevalent mutations as a major determinant of increased levels of high density lipoprotein cholesterol. *J Clin Invest* 1994; **94**: 1872–1882.

20 Zimmerman PA, Guderian RH, Nutman TB: A new TNFA promoter allele identified in South American Blacks. *Immunogenetics* 1996; **44**: 485–486.

21 Libby P: Inflammation in atherosclerosis. *Nature* 2002; **420**: 868–874.

22 Bourgain C, Hoffjan S, Nicolae R *et al*: Novel case–control test in a founder population identifies P-selectin as an atopy susceptibility locus. *Am J Hum Genet* 2003; **73**: 612–626.

23 Felsenstein J PHYLIP (Phylogeny Inference Package). In: version 3.5c. ed. *Department of Genetics*. University of Washington: Seattle, Distributed by the author: 1993.

24 World Health Organization-International Society of Hypertension: 1999 World Health Organization-International Society of Hypertension Guidelines for the Management of Hypertension. *Guidelines Subcommittee. J Hypertens* 1999; **17**: 151–183.

25 Williams CL, Hayman LL, Daniels SR *et al*: Cardiovascular health in childhood: a statement for health professionals from the Committee on Atherosclerosis, Hypertension, and Obesity in the Young (AHOY) of the Council on Cardiovascular Disease in the Young, American Heart Association. *Circulation* 2002; **106**: 143–160.

26 von Eckardstein A, Funke H, Schulte M *et al*: Nonsynonymous polymorphic sites in the apolipoprotein (apo) A-IV gene are associated with changes in the concentration of apo B- and apo A-I- containing lipoproteins in a normal population. *Am J Hum Genet* 1992; **50**: 1115–1128.

27 Rosendaal FR, Siscovick DS, Schwartz SM *et al*: A common prothrombin variant (20210 G to A) increases the risk of myocardial infarction in young women. *Blood* 1997; **90**: 1747–1750.

28 Rubattu S, Ridker P, Stampfer MJ *et al*: The gene encoding atrial natriuretic peptide and the risk of human stroke. *Circulation* 1999; **100**: 1722–1726.

29 Cavalli-Sforza LL, Menozzi P, Piazza A: Analysis of classical markers in forty-two selected populations; In: *The history and geography of human genes*, Abridged paperback edition ed. Princeton, NJ: Princeton University Press,; 1993, pp. 73–77.

30 Nicholson G, Donnelly P: Assessing population differentiation and isolation from single-nucleotide polymorphism data. *J R Statist Soc B* 2002; **64**: 1–21.

31 Groenendijk M, Cantor RM, de Bruin TW, Dallinga-Thie GM: The apoAI-CIII-AIV gene cluster. *Atherosclerosis* 2001; **157**: 1–11.

32 Eichner JE, Dunn ST, Perveen G *et al*: Apolipoprotein E polymorphism and cardiovascular disease: a HuGE review. *Am J Epidemiol* 2002; **155**: 487–495.

33 Varret M, Rabes JP, Thiart R *et al*: LDLR Database (second edition): new additions to the database andvthe software, and results of the first molecular analysis. *Nucleic Acids Res* 1998; **26**: 248–252.

34 www.ucl.ac.uk/fh.

35 Schuster H, Humphries S, Rauh G *et al*: Association of DNA-haplotypes in the human LDL-receptor gene with normal serum cholesterol levels. *Clin Genet* 1990; **38**: 401–409.

36 Haviland MB, Ferrell RE, Sing CF: Association between common alleles of the low-density lipoprotein receptor gene region and interindividual variation in plasma lipid and apolipoprotein levels in a population-based sample from Rochester, Minnesota. *Hum Genet* 1997; **99**: 108–114.

37 Knoblauch H, Bauerfeind A, Krahenbuhl C *et al*: Common haplotypes in five genes influence genetic variance of LDL and HDL cholesterol in the general population. *Hum Mol Genet* 2002; **11**: 1477–1485.

38 Hansen PS, Defesche JC, Kastelein JJ *et al*: Phenotypic variation in patients heterozygous for familial defective apolipoprotein B (FDB) in three European countries. *Arterioscler Thromb Vasc Biol* 1997; **17**: 741–747.

39 Ahn YI, Kamboh MI, Aston CE *et al*.: Role of common genetic polymorphisms in the LDL receptor gene in affecting plasma cholesterol levels in the general population. *Arterioscler Thromb Vasc Biol* 1994; **14**: 663–670.

40 Barter PJ, Brewer HB Jr, Chapman MJ *et al*: Cholesteryl ester transfer protein: a novel target for raising HDL and inhibiting atherosclerosis. *Arterioscler Thromb Vasc Biol* 2003; **23**: 160–167.

41 Dachet C, Poirier O, Cambien F *et al*: New functional promoter polymorphism, CETP/−629, in cholesteryl ester transfer protein (CETP) gene related to CETP mass and high density lipoprotein cholesterol levels: role of Sp1/Sp3 in transcriptional regulation. *Arterioscler Thromb Vasc Biol* 2000; **20**: 507–515.

42 Klerkx AH, Tanck MW, Kastelein JJ *et al*.: Haplotype analysis of the CETP gene: not TaqIB, but the closely linked −629C→A polymorphism and a novel promoter variant are independently associated with CETP concentration. *Hum Mol Genet* 2003; **12**: 111–123.

43 Suzuki K, Kuriyama M, Saito T, Ichinose A: Plasma lipoprotein(a) levels and expression of the apolipoprotein(a) gene are dependent on the nucleotide polymorphisms in its 5′-flanking region. *J Clin Invest* 1997; **99**: 1361–1366.

44 Ichinose A, Kuriyama M: Detection of polymorphisms in the 5′-flanking region of the gene for apolipoprotein(a). *Biochem Biophys Res Commun* 1995; **209**: 372–378.

45 Kraft HG, Windegger M, Menzel HJ, Utermann G: Significant impact of the +93 C/T polymorphism in the apolipoprotein(a)

gene on Lp(a) concentrations in Africans but not in Caucasians: confounding effect of linkage disequilibrium. *Hum Mol Genet* 1998; **7**: 257–264.

46 Ober C, Leavitt SA, Tsalenko A *et al*: Variation in the interleukin 4-receptor alpha gene confers susceptibility to asthma and atopy in ethnically diverse populations. *Am J Hum Genet* 2000; **66**: 517–526.

47 Donfack J, Kogut P, Forsythe S *et al*: Sequence variation in the promoter region of the cholinergic receptor muscarinic 3 gene and asthma and atopy. *J Allergy Clin Immunol* 2003; **111**: 527–532.

48 Ober C, Aldrich CL, Chervoneva I *et al*: Variation in the HLA-G promoter region influences miscarriage rates. *Am J Hum Genet* 2003; **72**: 1425–1435.

49 Lohmueller KE, Pearce CL, Pike M *et al*: Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003; **33**: 177–182.

50 Service SK, Sandkuijl LA, Freimer NB: Cost-effective designs for linkage disequilibrium mapping of complex traits. *Am J Hum Genet* 2003; **72**: 1213–1220.

51 McPeek MS, Wu X, Ober C: Bestlinear unbiased allele-frequency estimation in complex pedigrees. *Biometrics* 2004; **60**: in press.

Supplementary information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)