Variants Associated with Common Disease Are Not Unusually Differentiated in Frequency across Populations

Kirk E. Lohmueller,[1,3] Matthew M. Mauney,[2] David Reich,[4,5] and John M. Braverman[1]

Departments of [1]Biology and [2]Computer Science, Georgetown University, and [3]Institute for Molecular and Human Genetics, Georgetown University Medical Center, Washington, DC; [4]Department of Genetics, Harvard Medical School, Boston; and [5]Broad Institute of Harvard and MIT, Cambridge, MA

Genetic variants that contribute to risk of common disease may differ in frequency across populations more than random variants in the genome do, perhaps because they have been exposed to population-specific natural selection. To assess this hypothesis empirically, we analyzed data from two groups of single-nucleotide polymorphisms (SNPs) that have shown reproducible ($n = 9$) or reported ($n = 39$) associations with common diseases. We compared the frequency differentiation (between Europeans and Africans) of the disease-associated SNPs with that of random SNPs in the genome. These common-disease–associated SNPs are not significantly more differentiated across populations than random SNPs. Thus, for the data examined here, ethnicity will not be a good predictor of genotype at many common-disease–associated SNPs, just as it is rarely a good predictor of genotype at random SNPs in the genome.

An open question in medical and population genetics is how much information a person's self-identified ancestry (ethnicity) conveys about his or her risk of common disease (Risch et al. 2002; Burchard et al. 2003; Cooper et al. 2003). One way in which ethnicity could be informative about common-disease risk is if risk alleles vary in frequency among populations, which would allow ethnicity to be a predictor of whether a person has a risk allele. This correlation between ethnicity and genotype would be strongest if the disease-associated variants were differentiated in frequency. Although it is known that random variants in the genome are not particularly differentiated across populations on average (Lewontin 1972; Bowcock et al. 1991; Rosenberg et al. 2002), it has been hypothesized that, because of population-specific natural selection, functional SNPs associated with common disease may be more differentiated (Akey et al. 2002; Bamshad et al. 2004). There has been no empirical attempt to address this question, largely because so few disease-associated SNPs have been identified to date.

We set out to test the hypothesis that common-disease–associated SNPs are more differentiated than random SNPs by conducting an empirical evaluation of population differentiation in 48 SNPs associated with common disease. We wanted to study SNPs that were associated with common, complex traits, so we explicitly excluded variants associated with Mendelian diseases. The SNPs were all identified in a way that would not create a bias toward unusually high or low levels of frequency differentiation across populations, since each of them was initially identified in studies of single populations.

We first studied nine SNPs reproducibly associated with common disease (table 1). These SNPs satisfied two criteria: (1) >75% of replication studies showed a statistically significant association (Hirschhorn et al. 2002) or the association was significant after meta-analysis of replication studies (Lohmueller et al. 2003) and (2) allele-frequency information was publicly available for the SNPs in both West African and European-derived populations.

Second, we studied 39 SNPs that have been reported to be associated with common disease (table 2) but for which association has not necessarily been replicated. These were identified by checking the genes sequenced by the Seattle SNPs project (Seattle SNPs Web site) for overlap with the SNPs reported to be associated with common disease in the OMIM and PubMed databases or in table 1 of Hirschhorn et al. (2002).

Received June 8, 2005; accepted for publication October 13, 2005; electronically published November 16, 2005.

Address for correspondence and reprints: Dr. John M. Braverman, Department of Biology, Georgetown University, 3700 O Street NW, Washington, DC 20057-1229. E-mail: jmb24@georgetown.edu

Am. J. Hum. Genet. 2006;78:130–136. 0002-9297/2006/7801-0014$15.00

Table 1

Allele-Frequency Data for Nine Reproducible Associations

Gene	Disease[a]	SNP	Associated Allele[b]	Frequency		δ[f]	F_{ST}	Reference(s)[c]
				European[d]	African[e]			
CTLA4	T1DM	Thr17Ala	Ala	.38 (1,670)	.209 (402)	.171	.06	Osei-Hyiaman et al. 2001; Lohmueller et al. 2003
DRD3	Schizophrenia	Ser9Gly	Ser/Ser	.67 (202)	.116 (112)	.554	.458	Crocq et al. 1996; Lohmueller et al. 2003
AGT	Hypertension	Thr235Met	Thr	.42 (3,034)	.91 (658)	.49	.358	Rotimi et al. 1996; Nakajima et al. 2002
PRNP	CJD	Met129Val	Met	.72 (138)	.556 (72)	.164	.049	Hirschhorn et al. 2002; Soldevila et al. 2003
F5	DVT	Arg506Gln	Gln	.044 (1,236)	.00 (251)	.044	.03	Rees et al. 1995; Hirschhorn et al. 2002
HFE	HFE	Cys382Tyr	Tyr	.038 (2,900)	.00 (806)	.038	.024	Feder et al. 1996; Merryweather-Clarke et al. 1997
MTHFR	DVT	C677T	T	.3 (188)	.066 (468)	.234	.205	Schneider et al. 1998; Ray et al. 2002
PPARG	T2DM	Pro12Ala	Pro	.925 (120)	1.0 (120)	.075	.067	Altshuler et al. 2000; HapMap Project
KCNJ11	T2DM	Asp23Lys	Lys	.36 (96)	.09 (98)	.27	.182	Florez et al. 2004

[a] CJD = Creutzfeldt-Jacob disease; DVT = deep venous thrombosis; HFE = hemochromatosis; T1DM = type I diabetes; T2DM = type II diabetes.

[b] The associated allele is the SNP associated with disease, regardless of whether it is the derived or the ancestral allele. The frequencies for this allele are given.

[c] The reference that claims this to be a reproducible association, as well as the reference from which the allele frequencies were taken. For allele frequencies obtained from a meta-analysis, only the reference claiming reproducible association is given.

[d] Allele frequency obtained from the literature involving a European population. Either the general population frequency or the frequency in control groups in an association study was used. To reduce bias, when a control frequency was used for Europeans, a control frequency was also used for Africans. The total number of chromosomes surveyed is given in parentheses after each frequency.

[e] Allele frequency obtained from the literature involving a West African population. The total number of chromosomes surveyed is given in parentheses after each frequency.

[f] δ = The difference in the allele frequency between Europeans and Africans.

To assess whether the disease-associated SNPs are more differentiated across populations than random SNPs in the genome, we compared the two groups of disease-associated SNPs with SNPs from two public databases (table 3). The first database ("WICGR") was generated by the Whitehead/MIT Center for Genome Research and includes frequency data for SNPs genotyped in European American and Nigerian populations (see The SNP Consortium Allele Frequency/Genotype Project Web site). Since this data set includes a West African population, it was compared to the reproducible-disease-association group. The second database ("Perlegen") consists of SNPs for which frequency information is available in both European and African American populations (Hinds et al. 2005); it also has the virtue of including genotypes of the same samples that were studied for the Seattle SNPs project. For both databases, the physical map position, gene name, and SNP type were downloaded from dbSNP by a batch query of "rs" numbers (National Center for Biotechnology Information, dbSNP build 120, March 2004). SNPs were excluded from analysis if they were (*a*) not polymorphic, (*b*) mapped to more than one chromosomal location, or (*c*) within 20 kb of each other. The final WICGR data set consisted of 2,377 SNPs, and the final Perlegen data set consisted of 103,536 SNPs. To measure differentiation between European- and African-derived populations for the SNPs in all four data sets, we calculated F_{ST} (Weir and Cockerham 1984; Weir 1996), a classic measure of the frequency differentiation of a polymorphism.

To determine whether the average F_{ST} of 0.159 in the group of nine SNPs that were reproducibly associated with common disease was significantly larger than the average for random SNPs in the WICGR data set, we subsampled the WICGR data 10,000 times, counting the proportion of times that nine SNPs randomly chosen from WICGR had an average $F_{ST} \geq 0.159$. We did not find a significant increase in average F_{ST} in the reproducible-association set relative to the random group ($P = .12$). The same subsampling method also did not detect an excess in the percentage of SNPs with $F_{ST} > 0.3$ ($P = .26$). To obtain an upper bound on the level of differentiation at common-disease–associated SNPs, we performed bootstrap resamplings of the data from the nine reproducibly associated SNPs. Of 10,000 bootstrap replicates, 95% had average F_{ST} values in the range 0.074–0.274, which, as expected, is consistent with the average F_{ST} of the random SNPs.

A potential concern with this analysis is that different numbers of samples were used to calculate F_{ST} for the disease-associated and WICGR data sets. We therefore repeated our analysis after randomly dropping samples from the WICGR data set and the reproducible-disease-association data set until only 72 African and 72 European alleles for each SNP remained. More specifically, for the reproducible-disease-association group, we performed the random sample-dropping procedure for the

Table 2

Allele-Frequency Data for 39 Reported Associations

Gene	Disease/Phenotype[a]	SNP	Associated Allele[b]	Frequency			F_{ST}	Reference[c]
				European[d]	African[e]	δ[f]		
ADRB1	MI	Arg389Gly	Arg	.717 (46)	.467 (30)	.251	.1	Iwai et al. 2003
ALOX5AP	MI, stroke	*rs10507391*	T	.682 (44)	.159 (44)	.523	.425	Helgadottir et al. 2004
CAT	Hypertension	−844 (C/T)	T[g]	.714 (42)	.659 (44)	.055	0	Jiang et al. 2001
CCR2	AIDS susceptibility	Ile64Val	Val	.87 (46)	.813 (48)	.057	0	Smith et al. 1997
CD36	Malaria	Y to stop	Stop	0 (46)	.083 (48)	.083	.062	Aitman et al. 2000
F13	MI	Val34Leu	Val	.762 (42)	.795 (44)	.033	0	Kohler et al. 1999
FGA	Pulmonary embolism	Thr312Ala	Ala	.2 (40)	.5 (42)	.3	.159	Carter et al. 2000
GP1BA	CAD	Thr145Met	Met	.022 (46)	.167 (48)	.145	.095	Gonzalez-Conejero et al. 1998
ICAM1	MS	Lys469Glu	Lys	.643 (42)	.875 (48)	.232	.12	Nejentsev et al. 2003
ICAM1	Malaria	Lys29Met	Met	0 (46)	.354 (48)	.354	.335	Fernandez-Reyes et al. 1997
IFNGR1	Hp infection	−56 (C/T)	T	.455 (44)	.604 (48)	.15	.023	Thye et al. 2003
IL13	Asthma	−1055 (C/T)	T	.196 (46)	.25 (44)	.054	0	van der Pouw Kraan et al. 1999
IL13	Bronchial asthma	Arg110Gln	Gln	.273 (44)	.119 (42)	.154	.05	Heinzmann et al. 2003
IL1A	AD	−889 (C/T)	T	.295 (44)	.391 (46)	.096	0	Nicoll et al. 2000
IL1B	Gastric cancer	−31 (C/T)	T	.826 (46)	.375 (48)	.451	.335	El-Omar et al. 2000
IL3	RA	−16 (C/T)	C	.739 (46)	.875 (48)	.136	.037	Yamada et al. 2001
IL4	Asthma	−590 (T/C)	T	.174 (46)	.708 (48)	.534	.436	Noguchi et al. 1998
IL4R	Asthma	Gln576Arg	Arg	.295 (44)	.565 (46)	.27	.118	Hershey et al. 1997
IL6	Juvenile arthritis	−174 (C/G)	G	.5 (44)	1 (46)	.5	.494	Fishman et al. 1998
IL8	RSV bronchiolitis	−251 (T/A)	T[h]	.659 (44)	.229 (48)	.43	.301	Hull et al. 2000
ITGA2	MI	807 (C/T)	T	.316 (38)	.25 (48)	.066	0	Moshfegh et al. 1999
LTA	MI	Thr26Asn	Asn	.357 (42)	.5 (44)	.143	.018	Ozaki et al. 2002
MC1R	Fair skin	Val92Met	Met	.068 (44)	0 (44)	.068	.047	Valverde et al. 1995
NOS3	MI	Glu298Asp	Asp	.5 (44)	.136 (44)	.364	.247	Shimasaki et al. 1998
PLAU	AD	Pro141Leu	Pro	.659 (44)	.979 (48)	.32	.287	Finckh et al. 2003
PON1	CAD	Arg192Gln	Arg	.174 (46)	.727 (44)	.553	.461	Serrato and Marian 1995
PON2	CAD	Cys311Ser	Ser	.826 (46)	.762 (42)	.064	0	Sanghera et al. 1998
PTGS2	Colon cancer	−765 (G/C)	C	.238 (42)	.292 (48)	.054	0	Koh et al. 2004
PTPN22[i]	RA	Arg620Trp	Trp	.084 (1,120)	.024 (818)	.059	.03	Begovich et al. 2004
SELE	CAD	Ser128Arg	Arg	.091 (44)	.021 (48)	.07	.025	Wenzel et al. 1994
SELL	IgA nephropathy	Pro238Ser	Ser	.065 (46)	.333 (48)	.268	.183	Takei et al. 2002
SELP	MI	Thr715Pro	Thr	.864 (44)	.977 (44)	.114	.063	Herrmann et al. 1998
SFTPB	ARDS	Ile131Thr	Thr	.5 (44)	.348 (46)	.152	.025	Lin et al. 2000
SPD	RSV infection	Met11Thr	Met	.568 (44)	.478 (46)	.09	0	Lahti et al. 2002
TF	AD	Pro570Ser	Pro	.957 (46)	.935 (46)	.022	0	Zhang et al. 2003
THBD	MI	Ala455Val	Ala	.87 (46)	.848 (46)	.022	0	Norlund et al. 1997
THBS4	MI	Ala387Pro	Pro	.341 (44)	.083 (48)	.258	.166	Topol et al. 2001
TNFA	Infectious disease	−308 (A/G)	A	.182 (44)	.205 (44)	.023	0	Bayley et al. 2004
VCAM1	Stroke in SCD	Gly413Ala	Gly	1 (46)	.938 (48)	.063	.041	Taylor et al. 2002

[a] AD = Alzheimer disease; AIDS = acquired immunodeficiency syndrome; ARDS = acute respiratory distress syndrome; CAD = coronary artery disease; Hp = *Helicobacter pylori*; MI = myocardial infarction; MS = multiple sclerosis; RA = rheumatoid arthritis; RSV = respiratory syncytial virus; SCD = sickle cell disease.

[b] The associated allele is the SNP associated with disease, regardless of whether it is the derived or the ancestral allele. The frequencies for this allele are given.

[c] The reference that reported association with the listed disease/phenotype.

[d] Frequency obtained from the Seattle SNPs database for the European sample. The total number of chromosomes surveyed is given in parentheses after each frequency.

[e] Frequency obtained from the Seattle SNPs database for the African American sample. The total number of chromosomes surveyed is given in parentheses after each frequency.

[f] δ = The difference in the allele frequency between African Americans and Europeans.

[g] Associated allele in database is A.

[h] Associated allele in reference is A.

[i] This SNP was not from the Seattle SNPs database; instead, allele frequencies from Begovich et al. (2004) were used.

Table 3

Summary of F_{ST} Values for Comparison of Disease-Association and Genomewide Data Sets

Data Set	Populations Studied	No. of SNPs	Average F_{ST}[a]	No. (%) with $F_{ST} > .3$
WICGR:				
All SNPs	European and Nigerian	2,377	.119	237 (9.97)
72 alleles[b]	European and Nigerian	2,348	.113	233 (9.92)
Perlegen:				
All SNPs	European and African American	1,465,325	.083	88,138 (6.01)
⩾20 kb apart[c]	European and African American	103,536	.085	6,717 (6.49)
Reproducible associations	European and West African	9	.159 (.074–.274)[d]	2 (22.22)
Reported associations	European and African American	39	.120 (.077–.171)[d]	7 (17.95)

[a] Average F_{ST} values for African-derived and European-derived populations.

[b] WICGR data set after the sample size was decreased to 72 chromosomes at each SNP for both populations (see main text).

[c] Only SNPs that are at least 20 kb apart were used; this should decrease correlations among SNPs that are due to linkage disequilibrium.

[d] Average F_{ST} and, in parentheses, 95% CIs obtained by bootstrapping.

nine SNPs 1,000 times and recalculated F_{ST} for each replicate. The average F_{ST} and the percentage of F_{ST} values >0.3 in both data sets were extremely similar and were nearly identical to those observed in our original data sets, and we again could not reject the null hypothesis of no difference between the disease-associated and random SNPs (table 3).

We followed an identical protocol to compare F_{ST} in the group of SNPs with reported disease association to F_{ST} of SNPs in the Perlegen database. Here, there was no problem of sample size or sample mismatch, since the Perlegen set was genotyped in the same European American and African American individuals who were assayed for the disease-associated SNPs by the Seattle SNPs project. We did not find a significant increase in average F_{ST} ($P = .13$) or in the percentage of SNPs with $F_{ST} > 0.3$ ($P = .29$) in the reported-disease-association group relative to the Perlegen data set. To obtain a 95% CI for the differentiation, we performed 10,000 bootstrap resamplings of the 39 SNPs. The 95% CI is 0.074–0.171 and includes the genomewide average F_{ST} of the Perlegen data set (0.083) (table 3), which explains why the null hypothesis of no excess differentiation in the disease-associated SNPs relative to the random SNPs cannot be rejected. Because more SNPs are available for the reported-association group than for the reproducible-association group, we were able to put a more stringent upper bound on F_{ST} for the reported-association group.

Another question in medical genetics is whether disease-associated SNPs in the genome are more differentiated than random nonsynonymous SNPs (Freedman et al. 2004). To test this, we performed an analysis of the reported-association group in comparison with 6,763 nonsynonymous SNPs from the Perlegen data set (which may, of course, include some disease-associated SNPs). We did not find a significant increase in average F_{ST} ($P = .06$) or in the percentage of SNPs with $F_{ST} > 0.3$ ($P = .13$) in the reported-disease-association group relative to the nonsynonymous SNPs from the Perlegen data set.

The SNPs associated with common disease that we investigated do not show much higher levels of differentiation than those of random SNPs. Thus, in these cases, ethnicity is a poor predictor of an individual's genotype, which is also the pattern for random variants in the genome. This lends support to the hypothesis that many population differences in disease risk are environmental, rather than genetic, in origin. However, some exceptional SNPs associated with common disease are highly differentiated in frequency across populations, because of either a history of random drift or natural selection. The exceptional SNPs given in tables 1 and 2 are located in *AGT, DRD3, ALOX5AP, ICAM1, IL1B, IL4, IL6, IL8,* and *PON1*. Of note, evidence of selection has been observed for *AGT* (Nakajima et al. 2004), *IL4* (Rockman et al. 2003), *IL8* (Hull et al. 2001), and *PON1* (Allebrandt et al. 2002). Yet, for the vast majority of the common-disease–associated polymorphisms we examined, ethnicity is likely to be a poor predictor of an individual's genotype.

Acknowledgments

We thank Joel Hirschhorn for helpful discussions and for providing formatted versions of The SNP Consortium data. We also thank Jason Lohmueller and two anonymous reviewers for critical readings of the manuscript. K.E.L. is supported by a Barry Goldwater Scholarship and is a Georgetown-Hughes Undergraduate Research Scholar. D.R. is the recipient of a Burroughs Wellcome Career Development Award in the Biomedical Sciences.

Web Resources

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/SNP/

HapMap Project, http://www.hapmap.org/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/
PubMed, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed
Seattle SNPs, http://pga.gs.washington.edu/ (for National Heart, Lung, and Blood Institute Program for Genomic Application, Seattle SNPs, Seattle, WA [July 2004])
The SNP Consortium Allele Frequency/Genotype Project, http://snp.cshl.org/allele_frequency_project/ (for the WICGR data set)

References

Aitman TJ, Cooper LD, Norsworthy PJ, Wahid FN, Gray JK, Curtis BR, McKeigue PM, Kwiatkowski D, Greenwood BM, Snow RW, Hill AV, Scott J (2000) Malaria susceptibility and *CD36* mutation. Nature 405:1015–1016

Akey JM, Zhang G, Zhang K, Jin L, Shriver MD (2002) Interrogating a high-density SNP map for signatures of natural selection. Genome Res 12:1805–1814

Allebrandt KV, Souza RL, Chautard-Freire-Maia EA (2002) Variability of the paraoxonase gene (*PON1*) in Euro- and Afro-Brazilians. Toxicol Appl Pharmacol 180:151–156

Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, Tuomi T, Gaudet D, Hudson TJ, Daly M, Groop L, Lander ES (2000) The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. Nat Genet 26:76–80

Bamshad M, Wooding S, Salisbury BA, Stephens JC (2004) Deconstructing the relationship between genetics and race. Nat Rev Genet 5:598–609

Bayley JP, Ottenhoff TH, Verweij CL (2004) Is there a future for TNF promoter polymorphisms? Genes Immun 5:315–329

Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, et al (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (*PTPN22*) is associated with rheumatoid arthritis. Am J Hum Genet 75:330–337

Bowcock AM, Kidd JR, Mountain JL, Hebert JM, Carotenuto L, Kidd KK, Cavalli-Sforza LL (1991) Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. Proc Natl Acad Sci USA 88:839–843

Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, Mountain JL, Perez-Stable EJ, Sheppard D, Risch N (2003) The importance of race and ethnic background in biomedical research and clinical practice. N Engl J Med 348:1170–1175

Carter AM, Catto AJ, Kohler HP, Ariens RA, Stickland MH, Grant PJ (2000) α-Fibrinogen Thr312Ala polymorphism and venous thromboembolism. Blood 96:1177–1179

Cooper RS, Kaufman JS, Ward R (2003) Race and genomics. N Engl J Med 348:1166–1170

Crocq MA, Buguet A, Bisser S, Burgert E, Stanghellini A, Uyanik G, Dumas M, Macher JP, Mayerova A (1996) *Bal*I and *Msp*I polymorphisms of the dopamine D3 receptor gene in African blacks and Caucasians. Hum Hered 46:58–60

El-Omar EM, Carrington M, Chow WH, McColl KE, Bream JH, Young HA, Herrera J, Lissowska J, Yuan CC, Rothman N, Lanyon G, Martin M, Fraumeni JF Jr, Rabkin CS (2000) Interleukin-1 polymorphisms associated with increased risk of gastric cancer. Nature 404:398–402

Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, Dormishian F, et al (1996) A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 13:399–408

Fernandez-Reyes D, Craig AG, Kyes SA, Peshu N, Snow RW, Berendt AR, Marsh K, Newbold CI (1997) A high frequency African coding polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. Hum Mol Genet 6:1357–1360

Finckh U, van Hadeln K, Muller-Thomsen T, Alberici A, Binetti G, Hock C, Nitsch RM, Stoppe G, Reiss J, Gal A (2003) Association of late-onset Alzheimer disease with a genotype of *PLAU,* the gene encoding urokinase-type plasminogen activator on chromosome 10q22.2. Neurogenetics 4:213–217

Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin JS, Humphries S, Woo P (1998) The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. J Clin Invest 102:1369–1376

Florez JC, Burtt N, de Bakker PI, Almgren P, Tuomi T, Holmkvist J, Gaudet D, Hudson TJ, Schaffner SF, Daly MJ, Hirschhorn JN, Groop L, Altshuler D (2004) Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. Diabetes 53:1360–1368

Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D (2004) Assessing the impact of population stratification on genetic association studies. Nat Genet 36:388–393

Gonzalez-Conejero R, Lozano ML, Rivera J, Corral J, Iniesta JA, Moraleda JM, Vicente V (1998) Polymorphisms of platelet membrane glycoprotein Ib associated with arterial thrombotic disease. Blood 92:2771–2776

Heinzmann A, Jerkic SP, Ganter K, Kurz T, Blattmann S, Schuchmann L, Gerhold K, Berner R, Deichmann KA (2003) Association study of the *IL13* variant Arg110Gln in atopic diseases and juvenile idiopathic arthritis. J Allergy Clin Immunol 112:735–739

Helgadottir A, Manolescu A, Thorleifsson G, Gretarsdottir S, Jonsdottir H, Thorsteinsdottir U, Samani NJ, et al (2004) The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke. Nat Genet 36:233–239

Herrmann SM, Ricard S, Nicaud V, Mallet C, Evans A, Ruidavets JB, Arveiler D, Luc G, Cambien F (1998) The P-selectin gene is highly polymorphic: reduced frequency of the Pro715 allele carriers in patients with myocardial infarction. Hum Mol Genet 7:1277–1284

Hershey GK, Friedrich MF, Esswein LA, Thomas ML, Chatila TA (1997) The association of atopy with a gain-of-function mutation in the α subunit of the interleukin-4 receptor. N Engl J Med 337:1720–1725

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR (2005) Whole-genome patterns of common DNA variation in three human populations. Science 307:1072–1079

Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. Genet Med 4:45–61

Hull J, Ackerman H, Isles K, Usen S, Pinder M, Thomson A, Kwiatkowski D (2001) Unusual haplotypic structure of *IL8,* a susceptibility locus for a common respiratory virus. Am J Hum Genet 69:413–419

Hull J, Thomson A, Kwiatkowski D (2000) Association of respiratory syncytial virus bronchiolitis with the interleukin 8 gene region in UK families. Thorax 55:1023–1027

Iwai C, Akita H, Kanazawa K, Shiga N, Terashima M, Matsuda Y, Takai E, Miyamoto Y, Shimizu M, Kajiya T, Hayashi T, Yokoyama M (2003) Arg389Gly polymorphism of the human β1-adrenergic receptor in patients with nonfatal acute myocardial infarction. Am Heart J 146:106–109

Jiang Z, Akey JM, Shi J, Xiong M, Wang Y, Shen Y, Xu X, Chen H, Wu H, Xiao J, Lu D, Huang W, Jin L (2001) A polymorphism in the promoter region of catalase is associated with blood pressure levels. Hum Genet 109:95–98

Koh WP, Yuan JM, van den Berg D, Lee HP, Yu MC (2004) Interaction

between *cyclooxygenase-2* gene polymorphism and dietary n-6 polyunsaturated fatty acids on colon cancer risk: the Singapore Chinese Health Study. Br J Cancer 90:1760–1764

Kohler HP, Futers TS, Grant PJ (1999) Prevalence of three common polymorphisms in the A-subunit gene of factor XIII in patients with coronary artery disease. Thromb Haemost 81:511–515

Lahti M, Lofgren J, Marttila R, Renko M, Klaavuniemi T, Haataja R, Ramet M, Hallman M (2002) Surfactant protein D gene polymorphism associated with severe respiratory syncytial virus infection. Pediatr Res 51:696–699

Lewontin RC (1972) The apportionment of human diversity. Evol Biol 6:381–398

Lin Z, Pearson C, Chinchilli V, Pietschmann SM, Luo J, Pison U, Floros J (2000) Polymorphisms of human *SP-A, SP-B,* and *SP-D* genes: association of *SP-B* Thr131Ile with ARDS. Clin Genet 58:181–191

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33:177–182

Merryweather-Clarke AT, Pointon JJ, Shearman JD, Robson KJ (1997) Global prevalence of putative haemochromatosis mutations. J Med Genet 34:275–278

Moshfegh K, Wuillemin WA, Redondo M, Lammle B, Beer JH, Liechti-Gallati S, Meyer BJ (1999) Association of two silent polymorphisms of platelet glycoprotein Ia/IIa receptor with risk of myocardial infarction: a case-control study. Lancet 353:351–354

Nakajima T, Jorde LB, Ishigami T, Umemura S, Emi M, Lalouel JM, Inoue I (2002) Nucleotide diversity and haplotype structure of the human angiotensinogen gene in two populations. Am J Hum Genet 70:108–123

Nakajima T, Wooding S, Sakagami T, Emi M, Tokunaga K, Tamiya G, Ishigami T, Umemura S, Munkhbat B, Jin F, Guan-Jun J, Hayasaka I, Ishida T, Saitou N, Pavelka K, Lalouel JM, Jorde LB, Inoue I (2004) Natural selection and population history in the human angiotensinogen gene (*AGT*): 736 complete *AGT* sequences in chromosomes from around the world. Am J Hum Genet 74:898–916

Nejentsev S, Laaksonen M, Tienari PJ, Fernandez O, Cordell H, Ruutiainen J, Wikstrom J, Pastinen T, Kuokkanen S, Hillert J, Ilonen J (2003) Intercellular adhesion molecule-1 K469E polymorphism: study of association with multiple sclerosis. Hum Immunol 64:345–349

Nicoll JA, Mrak RE, Graham DI, Stewart J, Wilcock G, MacGowan S, Esiri MM, Murray LS, Dewar D, Love S, Moss T, Griffin WS (2000) Association of interleukin-1 gene polymorphisms with Alzheimer's disease. Ann Neurol 47:365–368

Noguchi E, Shibasaki M, Arinami T, Takeda K, Yokouchi Y, Kawashima T, Yanagi H, Matsui A, Hamaguchi H (1998) Association of asthma and the interleukin-4 promoter gene in Japanese. Clin Exp Allergy 28:449–453

Norlund L, Holm J, Zoller B, Ohlin AK (1997) A common thrombomodulin amino acid dimorphism is associated with myocardial infarction. Thromb Haemost 77:248–251

Osei-Hyiaman D, Hou L, Zhiyin R, Zhiming Z, Yu H, Amankwah AA, Harada S (2001) Association of a novel point mutation (C159G) of the *CTLA4* gene with type 1 diabetes in West Africans but not in Chinese. Diabetes 50:2169–2171

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Hori M, Nakamura Y, Tanaka T (2002) Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. Nat Genet 32:650–654

Ray JG, Shmorgun D, Chan WS (2002) Common C677T polymorphism of the methylenetetrahydrofolate reductase gene and the risk of venous thromboembolism: meta-analysis of 31 studies. Pathophysiol Haemost Thromb 32:51–58

Rees DC, Cox M, Clegg JB (1995) World distribution of factor V Leiden. Lancet 346:1133–1134

Risch N, Burchard E, Ziv E, Tang H (2002) Categorization of humans in biomedical research: genes, race, and disease. Genome Biol 3: 2007.1–2007.12

Rockman MV, Hahn MW, Soranzo N, Goldstein DB, Wray GA (2003) Positive selection on a human-specific transcription factor binding site regulating IL4 expression. Curr Biol 13:2118–2123

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385

Rotimi C, Puras A, Cooper R, McFarlane-Anderson N, Forrester T, Ogunbiyi O, Morrison L, Ward R (1996) Polymorphisms of renin-angiotensin genes among Nigerians, Jamaicans, and African Americans. Hypertension 27:558–563

Sanghera DK, Aston CE, Saha N, Kamboh MI (1998) DNA polymorphisms in two paraoxonase genes (PON1 and PON2) are associated with the risk of coronary heart disease. Am J Hum Genet 62:36–44

Schneider JA, Rees DC, Liu YT, Clegg JB (1998) Worldwide distribution of a common methylenetetrahydrofolate reductase mutation. Am J Hum Genet 62:1258–1260

Serrato M, Marian AJ (1995) A variant of human paraoxonase/arylesterase (HUMPONA) gene is a risk factor for coronary artery disease. J Clin Invest 96:3005–3008

Shimasaki Y, Yasue H, Yoshimura M, Nakayama M, Kugiyama K, Ogawa H, Harada E, Masuda T, Koyama W, Saito Y, Miyamoto Y, Ogawa Y, Nakao K (1998) Association of the missense Glu298Asp variant of the endothelial nitric oxide synthase gene with myocardial infarction. J Am Coll Cardiol 31:1506–1510

Smith MW, Dean M, Carrington M, Winkler C, Huttley GA, Lomb DA, Goedert JJ, O'Brien TR, Jacobson LP, Kaslow R, Buchbinder S, Vittinghoff E, Vlahov D, Hoots K, Hilgartner MW, Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC), ALIVE Study, O'Brien SJ (1997) Contrasting genetic influence of *CCR2* and *CCR5* variants on HIV-1 infection and disease progression. Science 277:959–965

Soldevila M, Calafell F, Andres AM, Yague J, Helgason A, Stefansson K, Bertranpetit J (2003) Prion susceptibility and protective alleles exhibit marked geographic differences. Hum Mutat 22:104–105

Takei T, Iida A, Nitta K, Tanaka T, Ohnishi Y, Yamada R, Maeda S, Tsunoda T, Takeoka S, Ito K, Honda K, Uchida K, Tsuchiya K, Suzuki Y, Fujioka T, Ujiie T, Nagane Y, Miyano S, Narita I, Gejyo F, Nihei H, Nakamura Y (2002) Association between single-nucleotide polymorphisms in selectin genes and immunoglobulin A nephropathy. Am J Hum Genet 70:781–786

Taylor JG VI, Tang DC, Savage SA, Leitman SF, Heller SI, Serjeant GR, Rodgers GP, Chanock SJ (2002) Variants in the *VCAM1* gene and risk for symptomatic stroke in sickle cell disease. Blood 100: 4303–4309

Thye T, Burchard GD, Nilius M, Muller-Myhsok B, Horstmann RD (2003) Genomewide linkage analysis identifies polymorphism in the human interferon-γ receptor affecting *Helicobacter pylori* infection. Am J Hum Genet 72:448–453

Topol EJ, McCarthy J, Gabriel S, Moliterno DJ, Rogers WJ, Newby LK, Freedman M, Metivier J, Cannata R, O'Donnell CJ, Kottke-Marchant K, Murugesan G, Plow EF, Stenina O, Daley GQ (2001) Single nucleotide polymorphisms in multiple novel thrombospondin genes may be associated with familial premature myocardial infarction. Circulation 104:2641–2644

Valverde P, Healy E, Jackson I, Rees JL, Thody AJ (1995) Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. Nat Genet 11:328–330

van der Pouw Kraan TC, van Veen A, Boeije LC, van Tuyl SA, de Groot ER, Stapel SO, Bakker A, Verweij CL, Aarden LA, van der Zee JS (1999) An IL-13 promoter polymorphism associated with increased risk of allergic asthma. Genes Immun 1:61–65

Weir BS (1996) Genetic data analysis II. Sinauer, Sunderland, MA
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370
Wenzel K, Felix S, Kleber FX, Brachold R, Menke T, Schattke S, Schulte KL, Glaser C, Rohde K, Baumann G (1994) E-selectin polymorphism and atherosclerosis: an association study. Hum Mol Genet 3: 1935–1937
Yamada R, Tanaka T, Unoki M, Nagai T, Sawada T, Ohnishi Y, Tsunoda T, Yukioka M, Maeda A, Suzuki K, Tateishi H, Ochi T, Nakamura Y, Yamamoto K (2001) Association between a single-nucleotide polymorphism in the promoter of the human interleukin-3 gene and rheumatoid arthritis in Japanese patients, and maximum-likelihood estimation of combinatorial effect that two genetic loci have on susceptibility to the disease. Am J Hum Genet 68:674–685
Zhang P, Yang Z, Zhang C, Lu Z, Shi X, Zheng W, Wan C, Zhang D, Zheng C, Li S, Jin F, Wang L (2003) Association study between late-onset Alzheimer's disease and the transferrin gene polymorphisms in Chinese. Neurosci Lett 349:209–211