

Correcting for Cryptic Relatedness in Population-Based Association Studies of Continuous Traits

Feng Zhang^a Hong-Wen Deng^{a–c}

^aThe Key Laboratory of Biomedical Information Engineering of Ministry of Education, and Institute of Molecular Genetics, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, ^bCollege of Life Sciences and Engineering, Beijing Jiao Tong University, Beijing, PR China; ^cDepartments of Orthopedic Surgery and Basic Medical Science, School of Medicine, University of Missouri-Kansas City, Kansas City, Mo., USA

Key Words

Cryptic relatedness · Population-based association studies · Likelihood ratio tests

Abstract

Cryptic relatedness was suggested to be an important source of confounding in population-based association studies (PBAS). The magnitude and manner of cryptic relatedness affecting the performance of PBAS of continuous traits remain to be investigated. We simulated a set of related samples through biased sampling and inbreeding, and evaluated the power and type I error rates of simple association tests (SAT) without correcting for cryptic relatedness. We also used extended likelihood ratio tests (ELRT) to conduct PBAS accounting for cryptic relatedness, and compared it with genomic control (GC). Cryptic relatedness decreased the power as well as increased the type I error rates of SAT in both biased sampling and inbreeding models. The impact of cryptic relatedness on the performance of SAT appeared to be limited in the biased sampling model. However, cryptic relatedness in inbred populations may result in excessive false positive results of SAT. Compared with SAT and GC, ELRT obtained improved power and type I error rates under various scenarios. Ignoring cryptic relatedness may increase

spurious association results in PBAS. Our ELRT provides a novel approach to control cryptic relatedness in PBAS of human continuous traits.

Copyright © 2009 S. Karger AG, Basel

Introduction

Population-based association studies (PBAS) are a powerful strategy for susceptible gene mapping of human complex diseases [1–3]. With the rapid development of high-throughput genotyping technologies, genome-wide PBAS are widely used to identify causal alleles of human complex diseases, such as osteoporosis, diabetes and obesity [4–7]. Nonetheless, an outstanding issue complicating PBAS is population structure, which can cause spurious association results and limit the robustness and efficiencies of PBAS [8–10].

Population structure mainly refers to population stratification and cryptic relatedness. In contrast to population stratification, which has been extensively studied and was well addressed by Pritchard et al. [11, 12], Zhu et al. [13, 14], Zhang et al. [15], Chen et al. [16], Price et al. [17] and Devlin and Roeder [18], information about the impact of cryptic relatedness on PBAS is limited. Cryptic

relatedness, which means some or all subjects of study samples are related, was suggested to be an important source of confounding in PBAS [18, 19]. Because PBAS assume individual independence of study samples, cryptic relatedness may make these statistical tests invalid and reduce the robustness and efficiencies of PBAS. Biased sampling and inbreeding are two major reasons of cryptic relatedness.

By now, few studies have been conducted to assess the impact of cryptic relatedness on case-control studies [19], but none for quantitative traits. The magnitude and way of cryptic relatedness affecting the performance of PBAS of human continuous traits remain to be investigated. There are several PBAS methods that can take individual relationships into account [18, 20–22]. However, most of these methods require known individual relationships, which are usually not certain or available in practice. When individual relationships are not known in advance, genomic control (GC) was suggested to control cryptic relatedness in PBAS [18, 19]. GC was originally developed to correct for population stratification and had been found to be conservative in stratified populations [23–25]. Information about the performance of GC correcting for cryptic relatedness in PBAS is limited.

Variance component models were first introduced to genetic studies in the 20th century [26, 27]. Fisher divided the total phenotypic variance of a quantitative trait into environment variance and genetic variance due to additive, dominance, and epistasis genetic effects [26]. Through including the variance components of genes linked to particular loci, variance component models have been widely used for genetic linkage and association mapping of human complex diseases [28–31].

PLINK is a popular genome-wide PBAS software package, which can estimate individual genome-wide identity by descent (IBD) sharing coefficients using genotypic data in seeming unrelated individuals [32]. The estimated genome-wide IBD sharing coefficients can be used to infer individual relationships, which can then be included into a variance component model to control the impact of cryptic relatedness on PBAS. By now, to the best of our knowledge, no work about the performance of PLINK in IBD estimation has been reported.

In this study, we simulated a set of related samples through biased sampling and inbreeding, and evaluated the power and type I error rates of simple association tests (SAT) without correcting for cryptic relatedness. Based on a variance component model, we also used extended likelihood ratio tests (ELRT) to conduct PBAS accounting for cryptic relatedness, and compared it with GC in

both biased sampling and inbreeding models. Our study aims to assess how serious confounding from cryptic relatedness is in PBAS of continuous traits, and to develop an efficient PBAS approach to control cryptic relatedness.

Materials and Methods

Likelihood Ratio Tests

PLINK is first applied to genotypic data to estimate genome-wide IBD sharing coefficients for each pair of individuals [32]. The estimated IBD sharing coefficients can then be converted to kinship coefficients: $K_{ij} = 0.5P_{ij1} + P_{ij2}$, where K_{ij} represents the kinship coefficient between individuals i and j ; P_{ij1} and P_{ij2} are estimated by PLINK, and denote the general possibilities of sharing one and two IBD allele(s) between individuals i and j on a genome-wide scale, respectively. Based on the inferred kinship coefficients, classical likelihood ratio tests are extended to conduct PBAS accounting for individual relationships. Supposed genotypic and phenotypic data of n individuals were collected. The log-likelihood functions under null hypothesis (H_0) and alternate hypothesis (H_1) can be expressed as

$$H_0 : a = 0$$

$$L_0(\beta, \sigma_{poly}, \sigma_e / y, \Omega) = -\frac{N}{2} \ln 2\pi - \frac{1}{2} \ln |\Omega \sigma_{poly} + I \sigma_e| - \frac{1}{2} (y - \beta)^T (\Omega \sigma_{poly} + I \sigma_e)^{-1} (y - \beta)$$

$$H_1 : a \neq 0$$

$$L_1(\beta, a, \sigma_{poly}, \sigma_e / y, \Omega, Z) = -\frac{N}{2} \ln 2\pi - \frac{1}{2} \ln |\Omega \sigma_{poly} + I \sigma_e| - \frac{1}{2} (y - \beta - Za)^T (\Omega \sigma_{poly} + I \sigma_e)^{-1} (y - \beta - Za)$$

where a is phenotypic effect of candidate locus; β is a $n \times 1$ vector of fixed effect; σ_{poly} and σ_e are two $n \times 1$ vectors representing polygenic and environmental effects, respectively; y is a $n \times 1$ vector of observed phenotypic values; Ω is a $n \times n$ kinship coefficient matrix with element K_{ij} ($i, j = 1, 2, 3, \dots, n$); Z is a $n \times 1$ vector of individual genotype at candidate locus; I is a $n \times n$ identity matrix. Log-likelihood ratio test statistic U can be written as

$$U = -2 \log \frac{L_1}{L_0},$$

where L_0 and L_1 are the maximized log-likelihood values estimated under H_0 and H_1 , respectively.

Simulations

We considered two common cryptic relatedness models: biased sampling and inbreeding. Genotype data of 1,000 bi-allelic loci were simulated for each individual. Allele frequencies of the 1,000 loci were randomly generated from beta distribution in the first generation. Recombination rates were assigned 1.0×10^{-8} for all pairs of adjacent loci. Mutation rates were set to be 1.0×10^{-5} for each locus. All loci were assumed to be under Hardy-

Weinberg equilibrium and randomly recombined and mutated during genotype simulations.

For the biased sampling, we simulated 400 nuclear families with two parents and four children in each family. Two parents were first simulated based on the randomly generated allele frequencies, and then randomly mated and recombined to generate four children in each family. We randomly selected one parent and two children from each of the families as related individuals in the total sample (400 individuals). The remaining unrelated individuals in the total sample were obtained through randomly selecting one individual from each of the remaining families. For the inbreeding model, our simulation procedure includes two stages. In stage 1, based on the randomly generated allele frequencies, a small unrelated population was first simulated as the founder population. The founder population was then randomly mated and recombined for some generations to generate a population of size of 3,200, with non-overlapping and random pairing of parents in each generation. Four children were simulated in each family and population size was assumed to increase two times per generation in stage 1. In stage 2, the simulated population was forward-randomly mated and recombined for five generations to obtain an inbred population with non-overlapping and random pairing of parents in each generation. Population size was kept constant in stage 2. 400 subjects were finally randomly selected from the simulated inbred population (3,200 individuals) as study sample.

A bi-allelic quantitative trait locus (QTL) was assumed to be associated with an individual quantitative phenotype. The QTL was randomly selected from the simulated 1,000 loci with $0.18 \leq$ minor allele frequency ≤ 0.22 . An additive genetic model was implemented here for quantitative phenotype simulation. Let y_j be the phenotypic value of individual j , the linear model is expressed as

$$y_i = \beta + z_j a + p_j + e_j,$$

where β is a fixed effect; z_j is the genotype of individual j at QTL ($z_j = 0, 1$ or 2); a is the additive genetic effect of QTL; and p_j is the residual polygenic effect of individual j attributed to other potential susceptible loci. During the simulation, p_j was randomly generated from a normal distribution with mean 0 and variance σ_{poly} in the first generation. In the second or more generation (for the inbreeding model), p_j equaled the average value of two parents' p_j , which ensured the phenotypic relatedness among family members due to polygenic effect. e_j is the residual environmental effect of individual j , following zero-mean normal distribution with variance σ_e .

Proportions of sib pairs and polygenic variance in the biased sampling model and the founder population sizes in the inbreeding model were controlled to model various relatedness levels. The simulated QTL was assumed to explain 2% of phenotypic variation in both biased sampling and inbreeding models. Detailed parameter designs are presented in table 1.

Data Analyses

Individual kinship coefficients of the study sample (400 individuals) were first inferred by PLINK. To assess the possible bias caused by kinship coefficient inference, we also recorded the real kinship coefficients for each pair of individuals in a simulation for the biased sampling model. The simulated genotypic and phe-

Table 1. Parameter configurations and corresponding inflation factors estimated by GC in the studies

| Models ^a | Mean inflation factors ^b | | |
|------------------------|-------------------------------------|--------------------|------|
| | power | type I error rates | |
| <i>Biased sampling</i> | | | |
| PRS | PV | | |
| 0.0 | 0.4 | 1.10 | 1.10 |
| 0.1 | 0.4 | 1.13 | 1.11 |
| 0.2 | 0.4 | 1.14 | 1.14 |
| 0.3 | 0.4 | 1.16 | 1.16 |
| 0.3 | 0.2 | 1.12 | 1.12 |
| 0.3 | 0.3 | 1.14 | 1.13 |
| <i>Inbreeding</i> | | | |
| FPS | PV | | |
| 100 | 0.4 | 1.20 | 1.21 |
| 200 | 0.4 | 1.12 | 1.11 |
| 400 | 0.4 | 1.12 | 1.10 |

^a PRS = Proportions of related subjects; PV = polygenic variances; FPS = founder population sizes.

^b Average inflation factors were calculated by genomic control based on 1,000 replicates for each parameter setting.

notypic data were simultaneously analyzed by SAT, GC and ELRT using PLINK inferred kinship coefficients (ELRTP) and real kinship coefficients (ELRTR, only for the biased sampling model), respectively. 1,000 simulations were conducted for each parameter setting. Power and type I error rates were calculated, respectively, as the proportions of positive results (P values ≤ 0.05) obtained at the simulated QTL with and without phenotypic effect in 1,000 simulations. All the simulations and ELRT analyses were implemented in R [33].

Results

The mean inflation factors estimated by GC under various scenarios are presented in table 1. The performances of SAT, ELRTP, ELRTR and GC in the biased sampling and inbreeding models are detailed in the following:

Biased Sampling

Proportions of related subjects and polygenic variances were varied to investigate the potential effect of biased sampling on PBAS. As shown in table 2, with proportions of related subjects increasing from 0.0 to 0.3, we observed consistent decreasing trends in power (from 88.8 to 83.6%) as well as increasing type I error rates

Table 2. Performance of the 4 analytical methods in related samples with various proportions of related individuals

| | Proportions of related individuals | | | |
|-----------------------|------------------------------------|------|------|------|
| | 0.0 | 0.1 | 0.2 | 0.3 |
| Power, % | | | | |
| SAT | 88.8 | 88.4 | 86.0 | 83.6 |
| ELRTP | 88.6 | 88.6 | 85.0 | 83.8 |
| ELRTR | 88.6 | 88.8 | 85.8 | 84.6 |
| GC | 88.0 | 86.2 | 82.4 | 81.8 |
| Type I error rates, % | | | | |
| SAT | 5.0 | 6.2 | 6.4 | 7.0 |
| ELRTP | 3.6 | 3.4 | 4.0 | 3.5 |
| ELRTR | 3.4 | 3.2 | 3.2 | 3.0 |
| GC | 4.4 | 5.1 | 5.4 | 5.5 |

SAT = Simple association test without correcting for cryptic relatedness; ELRTP = extended likelihood ratio test using PLINK inferred kinship coefficients; ELRTR = extended likelihood ratio test using real kinship coefficients; GC = genomic control.

(from 5.0 to 7.0%) for SAT. Compared with GC, both ELRTP and ELRTR obtained higher power and lower type I error rates under the same proportions of related subjects that we investigated. The performance of ELRTR was slightly better than that of ELRTP. In addition, the performance of GC showed similar varying trends with SAT and obtained the lowest power (from 88.0 to 81.8%) under various studied proportions of related subjects.

Table 3 provides an overview of comparison results with respect to polygenic variances. With polygenic variances increasing from 0.2 to 0.4, SAT presented consistently decreasing power (from 85.6 to 83.6%) as well as increasing type I error rates (from 4.8 to 7.0%). ELRTP and ELRTR had similar performances and performed better in power and type I error rates than GC under the same polygenic variances investigated.

Inbred Populations

Table 4 summarized the association test results of the 4 methods in inbred populations. We observed a high type I error rate of 7.3% for SAT at the founder population size = 100. When founder population sizes increased to 200 or 400, SAT obtained normal type I error rates ($\leq 5\%$). Compared with GC, ELRTP generally showed higher power (from 95.0 to 96.0%) and lower type I error rates (from 4.2 to 4.9%) within the range of founder population size we investigated.

Table 3. Performance of the 4 analytical methods in related samples with various polygenic variances

| | Polygenic variances | | |
|-----------------------|---------------------|------|------|
| | 0.2 | 0.3 | 0.4 |
| Power, % | | | |
| SAT | 85.6 | 85.2 | 83.6 |
| ELRTP | 84.9 | 85.1 | 83.8 |
| ELRTR | 84.7 | 84.5 | 84.6 |
| GC | 82.7 | 81.9 | 81.8 |
| Type I error rates, % | | | |
| SAT | 4.8 | 5.5 | 7.0 |
| ELRTP | 3.0 | 3.4 | 3.5 |
| ELRTR | 3.0 | 3.0 | 3.0 |
| GC | 4.1 | 5.0 | 5.5 |

SAT = Simple association test without correcting for cryptic relatedness; ELRTP = extended likelihood ratio test using PLINK inferred kinship coefficients; ELRTR = extended likelihood ratio test using real kinship coefficients; GC = genomic control.

Table 4. Performance of the 3 analytical methods in inbred samples with various founder population sizes

| | Founder population sizes | | |
|-----------------------|--------------------------|------|------|
| | 100 | 200 | 400 |
| Power, % | | | |
| SAT | 96.0 | 96.2 | 94.6 |
| ELRTP | 96.0 | 95.8 | 95.0 |
| GC | 93.4 | 93.8 | 94.6 |
| Type I error rates, % | | | |
| SAT | 7.3 | 4.4 | 4.9 |
| ELRTP | 4.9 | 4.2 | 4.7 |
| GC | 5.8 | 5.2 | 4.8 |

SAT = Simple association test without correcting for cryptic relatedness; ELRTP = extended likelihood ratio test using PLINK inferred kinship coefficients; GC = genomic control.

Discussion

To answer how important it is to consider cryptic relatedness in PBAS of human continuous traits, we simulated a set of related samples through biased sampling and inbreeding, and investigated the power and type I error rates of SAT. We found that biased sampling decreased the power as well as increased the type I error

rates of SAT. However, the confounding from biased sampling was limited in our study. For instance, even if 30% of the samples were closely related sib pairs, the type I error rates of SAT just increased to 7.0%. The effects of biased sampling on the power of SAT also appeared to be limited in our study. To investigate the impact of biased sampling on the performance of SAT, we simulated extremely related samples, which are usually not available in practice. Our simulation results are consistent with Voight and Pritchard's study, which assessed the effect of cryptic relatedness on case-control studies through theoretical derivation [19]. Based on our simulation results and on the aforementioned study [19], we suggest that the impact of biased sampling on PBAS might be limited and could generally be ignored in practice.

Due to inherent advantages, some inbred populations, such as founder and island populations, were recommended for PBAS [9, 34–36]. Some or all individuals from these inbred populations are usually related because of their common ancestries [9]. Information about the potential impact of inbreeding on PBAS of continuous traits is limited. In our study, we observed a high type I error rate of 7.3% for SAT, when founder population size was 100. With founder population sizes increasing to 200 or 400, type I error rates of SAT decreased to normal levels ($\leq 5\%$). Our simulation results suggest that cryptic relatedness in inbred populations might increase spurious results in PBAS of continuous traits. For PBAS conducted in small and closely related inbred populations, it may be better to carefully address cryptic relatedness.

Because cryptic relatedness may be a serious problem in some situations [18, 19], we extended classical likelihood ratio tests to conduct PBAS accounting for cryptic relatedness (ELRT), and compared it with GC under various scenarios. ELRT presented improved power and type I error rates compared to GC in both biased sampling and inbreeding models. It should be emphasized that ELRT uses genome-wide IBD sharing coefficients estimated by PLINK to infer individual kinship coefficients, and does not require known individual relationships [32]. On the other hand, the performance of ELRT may be affected by the accuracy of genome-wide IBD sharing coefficients estimation. To assess the possible effect of genome-wide IBD sharing coefficients estimation on the performance of ELRT, in the biased sampling model, we compared the performance of ELRT using the kinship coefficients inferred by PLINK (ELRTP) and the real kinship coefficients obtained from simulations (ELRTR), respectively. The performance of ELRTP was close to that of ELRTR under various scenarios, which may demonstrate the

good performance of PLINK in genome-wide IBD sharing coefficients estimation, and suggests no significant effect of kinship coefficients inference on the performance of ELRT in our study. Additionally, we observed that the computational cost of ELRT significantly increased with increasing sample sizes, due to the large kinship coefficient matrix used by ELRT. For example, execution of ELRT on a data set with 2,000 samples and 1,000 markers requires about 26 hours of computation time (Intel Xeon dual quad-core CPUs with 4 GB memories), which is usually acceptable for real studies.

It should be noted that using PLINK to identify related individuals and excluding them in following studies may also help to decrease the impact of cryptic relatedness on PBAS. However, it may be difficult to define a suitable excluding criterion in practice. A too strict excluding criterion may significantly decrease sample sizes and power of PBAS, while a too loose one may not eliminate the spurious associations caused by cryptic relatedness. GC is a popular PBAS method correcting for population stratification and cryptic relatedness [18]. In our studies, GC generally showed moderate decreasing trends in power and moderate increasing trends in type I error rates with increasing relatedness levels in both biased sampling and inbreeding models. The performance of GC appeared to be slightly affected by relatedness levels.

In summary, our study results show that cryptic relatedness may decrease the power as well as increase the type I error rates of PBAS of continuous traits. The impact of cryptic relatedness caused by biased sampling on PBAS is limited. In contrast, cryptic relatedness in inbred populations may be serious and should be carefully addressed. Our ELRT provides a novel approach to control spurious results caused by cryptic relatedness in PBAS of human continuous traits.

Acknowledgements

Investigators of this work were partially supported by grants from NIH (R01 AR050496, R21 AG027110, R01 AG026564, P50 AR055081 and R21 AA015973). The study also benefited from grants from the National Science Foundation of China, the Huo Ying Dong Education Foundation, HuNan Province, Xi'an Jiaotong University, and the Ministry of Education of China.

References

- 1 Risch N, Teng J: The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. *Genome Res* 1998; 8:1273–1288.
- 2 Morton NE, Collins A: Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci USA* 1998;95:11389–11393.
- 3 Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996;273:1516–1517.
- 4 Liu YJ, Liu XG, Wang L, Dina C, Yan H, Liu JF, Levy S, Papiasian CJ, Drees BM, Hamilton JJ, Meyre D, Delplanque J, Pei YF, Zhang L, Recker RR, Froguel P, Deng HW: Genome-wide association scans identified CTNBL1 as a novel gene for obesity. *Hum Mol Genet* 2008;17:1803–1813.
- 5 Richards JB, Rivadeneira F, Inouye M, Pastinen TM, Soranzo N, Wilson SG, Andrew T, Falchi M, Gwilliam R, Ahmadi KR, Valdes AM, Arp P, Whittaker P, Verlaan DJ, Jhama M, Kumanduri V, Moorhouse M, van Meurs JB, Hofman A, Pols HA, Hart D, Zhai G, Kato BS, Mullin BH, Zhang F, Deloukas P, Uitterlinden AG, Spector TD: Bone mineral density, osteoporosis, and osteoporotic fractures: a genome-wide association study. *Lancet* 2008;371:1505–1512.
- 6 Liu YZ, Wilson SG, Wang L, Liu XG, Guo YF, Li J, Yan H, Deloukas P, Soranzo N, Chinapen-Horsley U, Cervino A, Williams FM, Xiong DH, Zhang YP, Jin TB, Levy S, Papiasian CJ, Drees BM, Hamilton JJ, Recker RR, Spector TD, Deng HW: Identification of PLCL1 gene for hip bone size variation in females in a genome-wide association study. *PLoS ONE* 2008;3:e3160.
- 7 Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Lawson ML, Robinson LJ, Skraban R, Lu Y, Chiavacci RM, Stanley CA, Kirsch SE, Rappaport EF, Orange JS, Monos DS, Devoto M, Qu HQ, Polychronakos C: A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 2007;448:591–594.
- 8 Deng HW: Population admixture may appear to mask, change or reverse genetic effects of genes underlying complex traits. *Genetics* 2001;159:1319–1323.
- 9 Newman DL, Abney M, McPeck MS, Ober C, Cox NJ: The importance of genealogy in determining genetic associations with complex traits. *Am J Hum Genet* 2001;69:1146–1148.
- 10 Cooper RS, Tayo B, Zhu X: Genome-wide association studies: implications for multiethnic samples. *Hum Mol Genet* 2008;17:R151–R155.
- 11 Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: Association mapping in structured populations. *Am J Hum Genet* 2000; 67:170–181.
- 12 Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945–959.
- 13 Zhu X, Zhang S, Zhao H, Cooper RS: Association mapping, using a mixture model for complex traits. *Genet Epidemiol* 2002;23: 181–196.
- 14 Zhu X, Li S, Cooper RS, Elston RC: A unified association analysis approach for family and unrelated samples correcting for stratification. *Am J Hum Genet* 2008;82:352–365.
- 15 Zhang S, Zhu X, Zhao H: On a semiparametric test to detect associations between quantitative traits and candidate genes using unrelated individuals. *Genet Epidemiol* 2003; 24:44–56.
- 16 Chen HS, Zhu X, Zhao H, Zhang S: Qualitative semi-parametric test for genetic associations in case-control designs under structured populations. *Ann Hum Genet* 2003;67: 250–264.
- 17 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;38:904–909.
- 18 Devlin B, Roeder K: Genomic control for association studies. *Biometrics* 1999;55:997–1004.
- 19 Voight BF, Pritchard JK: Confounding from cryptic relatedness in case-control association studies. *PLoS Genet* 2005;1:e32.
- 20 Slager SL, Schaid DJ: Evaluation of candidate genes in case-control studies: a statistical method to account for related subjects. *Am J Hum Genet* 2001;68:1457–1462.
- 21 Bourgain C, Hoffjan S, Nicolae R, Newman D, Steiner L, Walker K, Reynolds R, Ober C, McPeck MS: Novel case-control test in a founder population identifies P-selectin as an atopy-susceptibility locus. *Am J Hum Genet* 2003;73:612–626.
- 22 Abney M, Ober C, McPeck MS: Quantitative-trait homozygosity and association mapping and empirical genome-wide significance in large, complex pedigrees: fasting serum-insulin level in the Hutterites. *Am J Hum Genet* 2002;70:920–934.
- 23 Epstein MP, Allen AS, Satten GA: A simple and improved correction for population stratification in case-control studies. *Am J Hum Genet* 2007;80:921–930.
- 24 Wawro N, Bammann K, Pigeot I: Testing for association in the presence of population stratification: a simulation study comparing the S-TDT, STRAT and the GC. *Biom J* 2006; 48:420–434.
- 25 Zhang F, Wang Y, Deng HW: Comparison of population-based association study methods correcting for population stratification. *PLoS ONE* 2008;3:e3392.
- 26 Fisher R: The correlation between relatives on the supposition of Mendelian inheritance. *Trans R Soc Edinb* 1918;52:399–433.
- 27 Weinberg W: Über Vererbungsgesetze beim Menschen. *Induktive Abstammungs-Vererbungslehre* 1909;1:377–392.
- 28 Amos CI: Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* 1994;54:535–543.
- 29 Goldgar DE: Multipoint analysis of human quantitative genetic variation. *Am J Hum Genet* 1990;47:957–967.
- 30 Almasy L, Blangero J: Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998;62:1198–1211.
- 31 Scuteri A, Sanna S, Chen WM, Uda M, Albai G, Strait J, Najjar S, Nagaraja R, Orru M, Usala G, Dei M, Lai S, Maschio A, Busonero F, Mulas A, Ehret GB, Fink AA, Weder AB, Cooper RS, Galan P, Chakravarti A, Schlessinger D, Cao A, Lakatta E, Abecasis GR: Genome-wide association scan shows genetic variants in the FTO gene are associated with obesity-related traits. *PLoS Genet* 2007;3: e115.
- 32 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–575.
- 33 R-Development-Core-Team: R: A language and environment for statistical computing. Vienna, Austria, 2007.
- 34 Shifman S, Darvasi A: The value of isolated populations. *Nat Genet* 2001;28:309–310.
- 35 Lander ES, Schork NJ: Genetic dissection of complex traits. *Science* 1994;265:2037–2048.
- 36 Wright AF, Carothers AD, Pirastu M: Population choice in mapping genes for complex diseases. *Nat Genet* 1999;23:397–404.