Am. J. Hum. Genet. 73:612–626, 2003

Novel Case-Control Test in a Founder Population Identifies P-Selectin as an Atopy-Susceptibility Locus

Catherine Bourgain,[1,*] Sabine Hoffjan,[1] Raluca Nicolae,[1] Dina Newman,[1] Lori Steiner,[3] Karen Walker,[3] Rebecca Reynolds,[3] Carole Ober,[1] and Mary Sara McPeek[1,2]

Departments of [1]Human Genetics and [2]Statistics, University of Chicago, Chicago; and [3]Roche Molecular Systems, Alameda, CA

To avoid problems related to unknown population substructure, association studies may be conducted in founder populations. In such populations, however, the relatedness among individuals may be considerable. Neglecting such correlations among individuals can lead to seriously spurious associations. Here, we propose a method for case-control association studies of binary traits that is suitable for any set of related individuals, provided that their genealogy is known. Although we focus here on large inbred pedigrees, this method may also be used in outbred populations for case-control studies in which some individuals are relatives. We base inference on a quasi-likelihood score (QLS) function and construct a QLS test for allelic association. This approach can be used even when the pedigree structure is far too complex to use an exact-likelihood calculation. We also present an alternative approach to this test, in which we use the known genealogy to derive a correction factor for the case-control association χ^2 test. We perform analytical power calculations for each of the two tests by deriving their respective noncentrality parameters. The QLS test is more powerful than the corrected χ^2 test in every situation considered. Indeed, under certain regularity conditions, the QLS test is asymptotically the locally most powerful test in a general class of linear tests that includes the corrected χ^2 test. The two methods are used to test for associations between three asthma-associated phenotypes and 48 SNPs in 35 candidate genes in the Hutterites. We report a highly significant novel association ($P = 2.10^{-6}$) between atopy and an amino acid polymorphism in the P-selectin gene, detected with the QLS test and also, but less significantly ($P = .0014$), with the transmission/disequilibrium test.

Introduction

Association studies are an essential step in the genetic dissection of complex traits. Whereas linkage studies yield relatively broad locations for susceptibility loci, association studies can be used to test the role of particular candidate genes. However, classical case-control tests might detect differences between cases and controls owing to ignored population substructure or improperly accounted relatedness among individuals and not necessarily owing to a true association between a locus and a trait. We focus here on the problem of performing valid association studies for binary traits in samples with related individuals in which the relationships are known. Such a situation may be encountered in outbred populations (Risch and Teng 1998; Slager and Schaid 2001). In isolated populations, the relatedness among individuals may be considerable, with many or all individuals related through multiple lines of descent. Neglecting such correlations among individuals can lead to seriously spurious associations. This has been illustrated by Newman et al. (2001) in a study conducted in a sample of Hutterites from South Dakota. The ~750 members of their sample are descendants of just 64 founders and are related to each other by a 13-generation, 1,623-member pedigree. Using an association test for quantitative traits developed by Abney et al. (2002) that takes the pedigree structure into account, Newman et al. (2001) illustrated the dramatic effect on false-positive rates of neglecting interindividual correlations.

To overcome this problem, family-based association tests have been widely, if not systematically, used, even though they have some disadvantages. In particular, the need for genotype information on family members, such as parents or sibs, can drastically reduce the number of cases available for a study, a concern that may be particularly relevant for late-onset diseases. Devlin and Roeder (1999) have proposed to use genomic controls to correct association tests for unknown relatedness among individuals. However, when the genealogy of the sample is entirely known, it is preferable to use this information. Slager and Schaid (2001) have derived a correction factor for the Armitage trend test to account for the presence of close relatives (e.g., siblings and cous-

Received May 5, 2003; accepted for publication June 26, 2003; electronically published August 15, 2003.

Address for correspondence and reprints: Dr. Catherine Bourgain, INSERM U535, Hôpital Paul Brousse, Batiment Leriche, BP 100, 94817 Villejuif Cedex, France. E-mail: bourgain@vjf.inserm.fr

* Present affiliation: INSERM U535, Villejuif, France.

0002-9297/2003/7303-0014$15.00

ins) from outbred populations, but this method cannot handle complex inbred pedigrees. Furthermore, little is known about the relative power of these different approaches.

Here, we propose a method for case-control association studies of binary traits, a method suitable for any set of related individuals, provided that their genealogy is known. In particular, it can be used in large inbred pedigrees. The method takes into account interindividual correlations, as well as intraindividual correlations due to inbreeding, by conditioning on the pedigree structure. Because we want the method to be suitable even when the pedigree structure is far too complex to use an exact-likelihood calculation, we base inference on the quasi-likelihood score (QLS) function (also known as the "quasi-score function" [e.g., see McCullagh and Nelder 1989; Heyde 1997]). This quasi-score function has been proposed by Wu et al. (2000) and by M.S.M., X. Wu, and C.O. (unpublished data) for estimation of allele frequencies in large inbred pedigrees. In that case, it results in the best linear unbiased estimator. We extend their approach to construct a QLS test for allelic association. The exact computation of the correlations among alleles requires inbreeding coefficients for all individuals and kinship coefficients for all pairs of individuals included in the analysis.

We also present an alternative approach to this test, in which we derive a correction factor for the case-control association χ^2 test based on the variance that appropriately accounts for inter- and intraindividual correlations. As for the QLS test, this correction factor is performed conditional on the pedigree structure. The latter strategy is similar to the one used by Slager and Schaid (2001), although they used a variance computed conditional on the identity-by-descent (IBD) information obtained using the marker genotype data rather than an unconditional variance. We note that exact computation of the conditional variance is generally infeasible in complex pedigrees. By deriving the noncentrality parameters of both the QLS and corrected χ^2 statistics, we obtain analytical results that allow us to compare the power of the two approaches. Furthermore, we show that both tests belong to a general class of linear tests, with the QLS test being asymptotically the locally most powerful test of this class under certain regularity conditions.

In what follows, we shall first describe the QLS test statistic for allelic association and derive the variance correction of the case-control χ^2 test. We will then study the null distributions of the tests and compare their power in Hutterite samples. Finally we will use these statistics to test, in Hutterite samples, the association between three different asthma-related phenotypes (asthma [MIM 600807], bronchial hyperresponsiveness [BHR], and atopy) and a set of 48 SNPs located in 35 candidate genes that were selected because of their known or suspected role in the inflammatory process.

Methods

The QLS Test for Allelic Association

We develop a test for association between a single marker and a binary trait, based on case-control data from a founder population but also useful in outbred populations with sampled relatives. We first focus on the case of a biallelic marker and then extend the method to the multiallelic case.

Biallelic Case

Consider a group of N subjects sampled from a population of known genealogy. Consider a biallelic marker with alleles labeled "0" and "1." We start by considering the situation in which the marker has no association with the trait (the null model), and we briefly review the results of Wu et al. (2000) and M.S.M., X. Wu, and C.O. (unpublished data) on allele-frequency estimation for this case. We then present the alternative model and derive the QLS statistic. Let $Y = (Y_1, \ldots Y_i \ldots, Y_N)^{\mathrm{T}}$ be a vector with element Y_i equal to $\frac{1}{2} \times$ (the number of alleles of type 1 in individual i). $Y_i = 0$, $\frac{1}{2}$, or 1. Let p be the frequency of allele 1, $0 < p < 1$, and let $\mathbf{\Sigma}$ be the covariance matrix of Y. It can be shown that $\mathbf{\Sigma} = \frac{1}{2}p(1-p)\mathbf{K}$, where

$$\mathbf{K} = \begin{pmatrix} 1+h_1 & 2\phi_{12} & \ldots & 2\phi_{1n} \\ 2\phi_{12} & 1+h_2 & \ldots & 2\phi_{2n} \\ \vdots & \ldots & \ldots & \vdots \\ 2\phi_{1n} & 2\phi_{2n} & \ldots & 1+h_n \end{pmatrix} \qquad (1)$$

with h_i being the inbreeding coefficient of individual i and ϕ_{ij} the kinship coefficient between individuals i and j. We note that $\mathbf{K}$ will be invertible, provided that each MZ twin pair (if any) is entered as a single individual in the matrix.

Define $m = E(Y)$. By construction, m is a column vector of length N (N-vector) with $m = p\mathbf{1}$, where $\mathbf{1}$ is an N-vector of 1s. Let $D_p = \frac{\partial m}{\partial p} = \mathbf{1}$. Wu et al. (2000) and M.S.M., X. Wu, and C.O. (unpublished data) have proposed to use the quasi-score function $U = D_p^{\mathrm{T}}\mathbf{\Sigma}^{-1}(Y - m)$ to estimate p by setting $U(\hat{p}) = 0$ (e.g., see McCullagh and Nelder 1989). The solution to this equation is

$$\hat{p} = (D_p^{\mathrm{T}}\mathbf{K}^{-1}D_p)^{-1}D_p^{\mathrm{T}}\mathbf{K}^{-1}Y\ , \qquad (2)$$

which M.S.M. and X. Wu (unpublished data) have shown to be the best linear unbiased estimator of p.

We now consider the case in which the marker is associated with the trait (the alternative model). Suppose

that N_c subjects among the N are cases and that N_t are controls so that $N_c + N_t = N$. To test for an allelic association between the marker and the disease, we propose to consider the following model: $E(Y) = \mu = (\mu_1, \dots \mu_i, \dots \mu_N)^T$ with

$$\mu_i = \begin{cases} p + r & \text{if } i \text{ is } a \text{ case, with } 0 < p + r < 1 \\ p & \text{if } i \text{ is } a \text{ control, with } 0 < p < 1 \end{cases} . \quad (3)$$

Under the null hypothesis of no association, $r = 0$, whereas, under the alternative hypothesis, $r \neq 0$. p is now considered a nuisance parameter. We allow the covariance matrix of Y, $\mathbf{\Omega}$, to depend on both p and r. However, it turns out that we do not need to specify the exact form of $\mathbf{\Omega}$. We simply require that $\mathbf{\Omega} = \mathbf{\Sigma}$ when $r = 0$ and that $\mathbf{\Omega}$ be differentiable and invertible. The quasi-score function corresponding to our model is

$$U(r,p) = \begin{pmatrix} U_r(r,p) \\ U_p(r,p) \end{pmatrix} = \begin{pmatrix} D_r^T \mathbf{\Omega}^{-1}(Y - \mu) \\ D_p^T \mathbf{\Omega}^{-1}(Y - \mu) \end{pmatrix} ,$$

where

$$D_r = \frac{\partial \mu}{\partial r} \text{ and } D_p = \frac{\partial \mu}{\partial p} .$$

D_p and D_r are N-vectors with D_p as previously described and $D_r = (d_1 \dots d_N)^T$ where $d_i = 1$ if i is a case and $d_i = 0$ if i is a control. We propose to use this quasi-score function to build a QLS statistic. The classical score statistic when the null hypothesis is composite ($r = r_0$ and p is a nuisance parameter), as described by Cox and Hinkley (1974), has the following form

$$W_u = U_r^T(r_0,\hat{p}_0) i^{rr}(r_0,\hat{p}_0) U_r(r_0,\hat{p}_0) ,$$

where $\hat{p}_0$ is the maximum likelihood estimate of p when $r = r_0$. $U_r(r_0,\hat{p}_0)$ and $i^{rr}(r_0,\hat{p}_0)$ are, respectively, the derivative over r of the log-likelihood and the (r,r)th entry of the inverse of the information matrix, both computed with $p = \hat{p}_0$ and $r = r_0$. This statistic does not involve the existence of a likelihood function from which the score function is derived by differentiation. It can therefore be generalized to the case of quasi likelihood (Heyde 1997), where the derivative of the log-likelihood is replaced by the quasi-score function U and the information matrix is replaced by $E(UU^T)$. In our case, this substitution results in the statistic

$$\begin{aligned} W_{QLS} = {} & (Y - \hat{\mu}_0)^T \mathbf{\Sigma}_0^{-1} D_r [D_r^T \mathbf{\Sigma}_0^{-1} D_r \\ & - D_r^T \mathbf{\Sigma}_0^{-1} D_p (D_p^T \mathbf{\Sigma}_0^{-1} D_p)^{-1} D_p^T \mathbf{\Sigma}_0^{-1} D_r]^{-1} \\ & D_r^T \mathbf{\Sigma}_0^{-1} (Y - \hat{\mu}_0) , \end{aligned} \quad (4)$$

where $\hat{\mu}_0$ and $\mathbf{\Sigma}_0$ are, respectively, the expectation and the covariance matrix of Y evaluated at $r = 0$ and $p = \hat{p}_0$, where $\hat{p}_0$ is the maximum quasi-likelihood estimate of p when $r = 0$. Thus, $\hat{\mu}_0 = \hat{p}_0 D_p$ and $\mathbf{\Sigma}_0 = \frac{1}{2}\hat{p}_0(1 - \hat{p}_0)\mathbf{K}$ with $\hat{p}_0$ calculated using eq. (2). As demonstrated by Heyde (1997), W_{QLS} should follow a χ^2 distribution with 1 df under the null hypothesis, provided that $\mathrm{Var}_0(U)^{-1/2} U \sim \mathcal{MVN}(0,I)$ under the null. Simulations testing the accuracy of this null distribution will be presented in the "Results" section.

Suppose, now, that the N subjects belong to F independent families sampled in an outbred population. For each family f of size N_f, we define $\mathbf{\Sigma}_f$ and $\mathbf{K}_f$, its covariance and correlation matrices. $\mathbf{\Sigma}_f$ and $\mathbf{K}_f$ are $N_f \times N_f$ matrices with entries given in eq. (1) for $\mathbf{K}_f$ and $\mathbf{\Sigma}_f = \frac{1}{2}p(1 - p)\mathbf{K}_f$. When all the individuals in the sample are outbred, $\mathbf{K}_f$ has all its diagonal elements equal to 1. It comes from eq. (2) that

$$\hat{p} = \Big(\sum_{f=1}^{F} D_{pf}^T \mathbf{K}_f^{-1} D_{pf}\Big)^{-1} \Big(\sum_{f=1}^{F} D_{pf}^T \mathbf{K}_f^{-1} Y_f\Big) , \quad (5)$$

where Y_f is the N_f- vector of allele indicators for the N_f members of the fth family and D_{pf} is an N_f-vector of 1s. Similarly, W_{QLS} is computed as in eq. (4), with all the terms of the form $X^T \mathbf{\Sigma}_0^{-1} B$ in this formula (where X and B are N-vectors: either D_p D_r or $[Y - \hat{\mu}_0]$), computed as $X^T \mathbf{\Sigma}_0^{-1} B = \sum_{f=1}^{F} X_f^T \mathbf{\Sigma}_{0f}^{-1} B_f$, where X_f and B_f are the N_f-subvectors of X and B corresponding to the N_f members of the fth family.

Multiallelic Case

Consider a locus with a different alleles. Let $Y = (Y_1, \dots, Y_{a-1})^T$ be an $[(a - 1)N]$-vector with $Y_k = (Y_{k1}, \dots, Y_{kN})^T$ an N-vector and Y_{ki} equal to $\frac{1}{2}$(the number of alleles of type k in individual i). If a particular allele k is suspected to be associated with the disease, the locus might be treated as biallelic k/non-k and the test performed just as described in the previous case. Arguably, when there is no prior idea of which allele might be associated with the disease, a more general alternative model should be considered. As a generalization of the biallelic case, we consider $E(Y) = \mu = (\mu_1, \dots, \mu_{a-1})^T$, where $\mu_k = (p_k + s_1 r_k, \dots, p_k + s_N r_k)^T$ with $s_i = 1$ if i is a case and $s_i = 0$ if i is a control. Here, we write $p = (p_1, \dots, p_{a-1})^T$ and $r = (r_1, \dots, r_{a-1})^T$, and we assume $0 < p_k + r_k < 1$ and $0 < p_k < 1$ for all k. As in the biallelic case, we allow for a very general form for the covariance matrix of Y. When $r = 0$, this matrix is $\mathbf{\Sigma} = \mathbf{F} \otimes \mathbf{K}$

($\otimes$ is defined in appendix A), where $\mathbf{F}$ is the $(a-1)\times(a-1)$ matrix of the form

$$\mathbf{F}=\frac{1}{2}\begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_{a-1} \\ -p_1p_2 & p_2(1-p_2) & \dots & -p_2p_{a-1} \\ \vdots & \dots & \dots & \vdots \\ -p_1p_{a-1} & -p_2p_{a-1} & \dots & p_{a-1}(1-p_{a-1}) \end{pmatrix}. \tag{6}$$

Note that, in the biallelic case, $\mathbf{F}$ reduces to $\frac{1}{2}p(1-p)$ and $\mathbf{\Sigma}=\frac{1}{2}p(1-p)\mathbf{K}$, as given in the "Biallelic Case" subsection above. We show in appendix A that, in this case,

$$W_{\mathrm{QLS}}=\sum_{k=1}^{a-1}\sum_{i=1}^{a-1}(\hat{\mathbf{F}}^{-1})_{ik}(Y_k-\hat{\mu}_{0k})^{\mathrm{T}}\mathbf{K}^{-1}D_r[D_r^{\mathrm{T}}\mathbf{K}^{-1}D_r -D_r^{\mathrm{T}}\mathbf{K}^{-1}D_p(D_p^{\mathrm{T}}\mathbf{K}^{-1}D_p)^{-1}D_p^{\mathrm{T}}\mathbf{K}^{-1}D_r]^{-1}D_r^{\mathrm{T}}\mathbf{K}^{-1}(Y_i-\hat{\mu}_{0i}), \tag{7}$$

where, for all k, $\hat{\mu}_{0k}=\hat{p}_{0k}D_p$, and $(\hat{\mathbf{F}}^{-1})_{ik}$ is the (i,k) entry of $\mathbf{F}$ evaluated at $p=\hat{p}_0=(\hat{p}_{01},\dots,\hat{p}_{0k},\dots,\hat{p}_{0a-1})^{\mathrm{T}}$. Here, $\hat{p}_0$ is the quasi-likelihood estimator of p when $r=\mathbf{0}$. For each k, $\hat{p}_{0k}$ is calculated using eq. (2) with Y replaced by Y_k (Wu et al. [2000] and M.S.M., X. Wu, and C.O., unpublished data). In this case, we also expect the null distribution of W_{QLS} to be $\sim\chi^2$ with $(a-1)$ df, and we examine the accuracy of this approximation in the "Results" section. In a manner similar to the biallelic case, when the N subjects belong to F independent families, W_{QLS} is computed as in eq. (7) with all the terms of the form $X^{\mathrm{T}}\mathbf{K}^{-1}B$ in this formula—where X and B are N-vectors: either D_p, D_r or $(Y_k-\hat{\mu}_{0k})$—computed as $X^{\mathrm{T}}\mathbf{K}^{-1}B=\sum_{f=1}^{F}X_f^{\mathrm{T}}\mathbf{K}_f^{-1}B_f$, where X_f and B_f are the N_f-subvectors of X and B corresponding to the N_f members of the f family.

General Framework for the QLS and χ^2 Tests

The W_{QLS} statistic can be seen as a particular case of a more general class of linear statistics of the form $W=S^{\mathrm{T}}[\mathrm{Var}_0(S)]^{-1}S$, where $S=V^{\mathrm{T}}Y$ with $V\neq\mathbf{0}$ a known $[N(a-1)\times(a-1)]$ matrix and $E_0(S)=\mathbf{0}$. $E_0(S)$ and $\mathrm{Var}_0(S)$ are, respectively, the expectation and the variance of S when $r=r_0$. $\mathrm{Var}_0(S)$ depends on p and, in practice, is computed at $p=\hat{p}_0$. In what follows, we refer to this class of linear statistics as the "W class." In the biallelic case, we can define $V_{\mathrm{QLS}}=V_1$, where $V_1=\mathbf{K}^{-1}D_r-(D_r^{\mathrm{T}}\mathbf{K}^{-1}D_p)(D_p^{\mathrm{T}}\mathbf{K}^{-1}D_p)^{-1}\mathbf{K}^{-1}D_p$. Then, $W_{\mathrm{QLS}}=S_{\mathrm{QLS}}^{\mathrm{T}}\mathrm{Var}_0(S_{\mathrm{QLS}})^{-1}S_{\mathrm{QLS}}$, where $S_{\mathrm{QLS}}=V_{\mathrm{QLS}}^{\mathrm{T}}Y$ (this result can be obtained by inserting eq. [2] into eq. [4] and moving all the factors of $\frac{1}{2}\hat{p}_0[1-\hat{p}_0]$ into the variance term). More generally, we show, in appendix B, that, in the multiallelic case, we have a similar result with $V_{\mathrm{QLS}}=\mathbf{I}_{a-1}\otimes V_1$, where $\mathbf{I}_{a-1}$ is the identity matrix of dimension $(a-1)$.

In the special case when the correlations among all the individuals, as well as between the two alleles of an individual, are zero, the classical case-control χ^2 test for association also fits in the W class of statistics. For a biallelic locus, $V_{\chi^2}=V_2$, where $V_2=D_r-(D_r^{\mathrm{T}}D_p)(D_p^{\mathrm{T}}D_p)^{-1}D_p$ (note that V_2 is just V_1 with $\mathbf{K}$ replaced by $\mathbf{I}$). Indeed, $W_{\chi^2}=S_{\chi^2}^{\mathrm{T}}\mathrm{Var}_0(S_{\chi^2})^{-1}S_{\chi^2}$ with $S_{\chi^2}=V_{\chi^2}^{\mathrm{T}}Y$ is equal to the classical case-control χ^2 test statistic, as shown in appendix B. The multiallelic case is similar, with $V_{\chi^2}=\mathbf{I}_{a-1}\otimes V_2$ (see appendix B).

Correction Factor for the Classical χ^2 Test

One way to extend the classical χ^2 test so that it is valid when the above correlations are not zero, is to use the same $S_{\chi^2}=V_{\chi^2}^{\mathrm{T}}Y$ and recompute $\mathrm{Var}_0(S_{\chi^2})$ to take into account the correlations. In the biallelic case, this is done by making use of the fact that $\mathrm{Var}_0(Y)=\frac{1}{2}p(1-p)\mathbf{K}$, rather than $\frac{1}{2}p(1-p)\mathbf{I}$. We call the resulting statistic "$W_{\chi^2_{\mathrm{corr}}}$," and we have

$$W_{\chi^2_{\mathrm{corr}}}=\frac{\left[\sum_{i\in\mathrm{cases}}(Y_i-\bar{Y})\right]^2}{\frac{1}{2}\bar{Y}(1-\bar{Y})(D_r^{\mathrm{T}}\mathbf{K}D_r-2\frac{N_c}{N}D_p^{\mathrm{T}}\mathbf{K}D_r+(\frac{N_c}{N})^2D_p^{\mathrm{T}}\mathbf{K}D_p)}, \tag{8}$$

where $\hat{p}=\bar{Y}=\frac{1}{N}\sum_{i=1}^{N}Y_i$ has been substituted for p.

Let $\rho_{\chi^2_{\mathrm{corr}}}$ be the correction factor to be applied to the W_{χ^2} to have a valid test. $W_{\chi^2_{\mathrm{corr}}}=\rho_{\chi^2_{\mathrm{corr}}}W_{\chi^2}$ with

$$\rho_{\chi^2_{\mathrm{corr}}}=\frac{(N_c-\frac{N_c^2}{N})}{D_r^{\mathrm{T}}\mathbf{K}D_r-2\frac{N_c^2}{N}D_p^{\mathrm{T}}\mathbf{K}D_r+(\frac{N_c^2}{N})^2D_p^{\mathrm{T}}\mathbf{K}D_p}.$$

Note that $\rho_{\chi^2_{\mathrm{corr}}}$ depends only on the sample composition (i.e., who are the cases and who are the controls), not on allele-frequency estimates. It can be shown that the same correction applies when there are a alleles at a locus, so that $\rho_{\chi^2_{\mathrm{corr}}}W_{\chi^2}$ approximately follows a χ^2 distribution with $(a-1)$ df under the null hypothesis. Here again, if the N individuals belong to F independent families, $W_{\chi^2_{\mathrm{corr}}}$ and $\rho_{\chi^2_{\mathrm{corr}}}$ can be computed as above with all the terms of the form $X^{\mathrm{T}}\mathbf{K}B$ (where X and B are N-vectors) in these formulas, computed as $X^{\mathrm{T}}\mathbf{K}B=\sum_{f=1}^{F}X_f^{\mathrm{T}}\mathbf{K}_fB_f$.

Null Distribution of the W_{QLS} and $W_{\chi^2_{\mathrm{corr}}}$ Statistics

To determine whether the χ^2 approximation provides the correct type I error for the tests based on W_{QLS} and $W_{\chi^2_{\mathrm{corr}}}$, we performed simulations based on three different real case-control samples of Hutterites from South Da-

kota. The Hutterites are a North American religious isolate that originated in eastern Europe and whose entire population can be traced back to 90 ancestors in the 1700s/1800s. Sample 1 consisted of 701 Hutterites who were phenotyped for atopy, defined as a positive skin-prick testing (+SPT) to at least 1 of 14 airborne allergens. This sample included 310 individuals with atopy and 391 controls (no +SPT to any of the 14 allergens tested). Sample 2 consisted of 156 individuals with BHR and 434 controls without any asthma symptoms or BHR. Sample 3 consisted of the same 434 controls and 76 individuals out of the 156 cases of sample 2 with self-reported asthma symptoms and a doctor's diagnosis of asthma and BHR. The latter phenotype will be referred to as "asthma" in what follows. The details of the phenotype have been described elsewhere (Ober et al. 2000). The complete genealogy of these 719 different individuals was constructed from a Hutterite pedigree of $\geqslant$12,000 individuals. This yielded a 1,623-person pedigree that included all known ancestors of the individuals in the three samples. The inbreeding coefficients for all 719 individuals, as well as the kinship coefficients between any pair of individuals in each sample, were computed on the basis of the 1,623-person pedigree, using the algorithm of Boyce (1983). Mean values of the inbreeding and kinship coefficients for the three samples are presented in table 1. Two smaller subsamples were also considered, to evaluate the effect of sample size on the type I error. Sample 4 consisted of the 76 cases from sample 3 and 76 controls randomly drawn from the 434 corresponding controls. Sample 5 consisted of 30 cases randomly drawn from the 76 cases in sample 3 and 30 controls randomly drawn from the corresponding 434 controls.

Genotype information for markers unlinked to the phenotypes under study was simulated by randomly drawing alleles for the founders of the 1,623-person pedigree with fixed allele frequencies and then simulating the Mendelian transmission of these alleles throughout the pedigree. The validity of the χ^2 null distribution was assessed by comparing the proportion of simulations showing a statistic whose value is greater than the χ^2 threshold for a nominal type I error and the value of this nominal type I error.

Power of the W_{QLS} and $W_{\chi^2_{corr}}$ Statistics

We show, in appendix C, that, under certain regularity conditions, W_{QLS} is asymptotically the locally most powerful test of the W class of linear tests described earlier. We provide here analytical power calculations for W_{QLS} and $W_{\chi^2_{\mathrm{corr}}}$ to quantify the difference between the two tests. The basic assumption underlying these calculations is that, under the alternative hypothesis H_1, both W_{QLS} and $W_{\chi^2_{\mathrm{corr}}}$ have a noncentral χ^2 distribution

Table 1

Mean Inbreeding and Kinship Coefficients in Cases and in Controls from Three Hutterite Samples

Sample	Size	Kinship Coefficient	Inbreeding Coefficient
+SPT:			
Cases	310	.0436	.0363
Controls	391	.0419	.0321
BHR:			
Cases	156	.0423	.0337
Controls	434	.0426	.0348
Asthma:			
Cases	76	.0445	.0336
Controls	434	.0426	.0348

with the respective noncentrality parameters λ_{QLS} and $\lambda_{\chi^2_{\mathrm{corr}}}$. (For instance, this would hold asymptotically for local alternatives [i.e., alternatives that are close to the null] under certain conditions on $\mathbf{K}$.) λ_{QLS} is obtained by calculating W_{QLS} with Y replaced by $E_{H_1}(Y)$, where $E_{H_1}(Y)$ is the expectation of Y under the alternative hypothesis H_1. Similarly, $\lambda_{\chi^2_{\mathrm{corr}}}$ is equal to $W_{\chi^2_{\mathrm{corr}}}$ computed with Y replaced by $E_{H_1}(Y)$. We focus, in what follows, on the biallelic case. Using the expression of W_{QLS} as a function of S_{QLS} given earlier, we have

$$\sqrt{\lambda_{\mathrm{QLS}}} = \left| \frac{[D_r^{\mathrm{T}}\mathbf{K}^{-1} - (D_r^{\mathrm{T}}\mathbf{K}^{-1}D_p)(D_p^{\mathrm{T}}\mathbf{K}^{-1}D_p)^{-1}D_p^{\mathrm{T}}\mathbf{K}^{-1}]}{\sqrt{\frac{1}{2}p(1-p)[D_r^{\mathrm{T}}\mathbf{K}^{-1}D_r - (D_r^{\mathrm{T}}\mathbf{K}^{-1}D_p)^2(D_p^{\mathrm{T}}\mathbf{K}^{-1}D_p)^{-1}]}} E_{H_1}(Y) \right| . \tag{9}$$

The alternative presented in (3) may also be written as $E_{H_1}(Y) = pD_p + rD_r$. Thus,

$$\lambda_{\mathrm{QLS}} = \frac{r^2}{\frac{1}{2}p(1-p)}[(D_r^{\mathrm{T}}\mathbf{K}^{-1}D_r) - (D_r^{\mathrm{T}}\mathbf{K}^{-1}D_p)^2(D_p^{\mathrm{T}}\mathbf{K}^{-1}D_p)^{-1}] . \tag{10}$$

Similarly, it comes from eq. (8) that

$$\lambda_{\chi^2_{\mathrm{corr}}} = \frac{r^2(N_c - \frac{N_c^2}{N})^2}{\frac{1}{2}p(1-p)\left[D_r^{\mathrm{T}}\mathbf{K}D_r - 2\frac{N_c^2}{N}D_p^{\mathrm{T}}\mathbf{K}D_r + (\frac{N_c^2}{N})^2 D_p^{\mathrm{T}}\mathbf{K}D_p\right]} . \tag{11}$$

For fixed values of (p,r) defining an alternative, the power of W_{QLS} and $W_{\chi^2_{\mathrm{corr}}}$ for a nominal type I error α is $\beta_{\mathrm{QLS}} = 1 - \mathcal{R}_{\lambda_{\mathrm{QLS}},1}(K_{\alpha,1})$ and $\beta_{\chi^2_{\mathrm{corr}}} = 1 - \mathcal{R}_{\lambda_{\chi^2_{\mathrm{corr}}},1}(K_{\alpha,1})$, respectively, where $K_{\alpha,1}$ is the upper αth quantile of a χ^2_1 distribution and where $\mathcal{R}_{\lambda,1}$ is the distribution function of a noncentral χ^2_1 with noncentrality parameter λ.

Testing Candidate Genes for Asthma in the Hutterites

The W_{QLS} and the $W_{\chi^2_{corr}}$ statistics were used to test the association between asthma, BHR, and atopy and 48 biallelic markers located in 35 different genes that were selected because of their known or suspected role in the inflammatory process. Typing was done by multiplex PCR and an immobilized-probe linear-array system (LAS) (Mirel et al. 2002). These genes (n = number of polymorphisms per gene) included interleukin 4 (*IL4* [MIM 147780], $n = 1$), interleukin 4 receptor α chain (*IL4RA* [MIM 147730], $n = 3$), interleukin 13 (*IL13* [MIM 147683], $n = 1$), β_2-adrenergic receptor (*ADRB2* [MIM 109690], $n = 3$), intercellular adhesion molecule 1 (*ICAM1* [MIM 147840], $n = 2$), vascular cell adhesion molecule 1 (*VCAM1* [MIM 192225], $n = 1$), E-selectin (*SELE* [MIM 131210], $n = 1$), P-selectin (*SELP* [MIM 173610], $n = 2$), Fcϵ receptor β chain (*FCERB1* [MIM 147138], $n = 1$), monocyte differentiation antigen CD14 (*CD14* [MIM 158120], $n = 1$), uteroglobin (*UGB* [MIM 192020], $n = 1$), transforming growth factor β1 (*TGFB1* [MIM 190180], $n = 1$), Eotaxin (*SCYA11* [MIM 601156], $n = 2$), chemokine receptor 2 (*CCR2* [MIM 601267], $n = 1$), chemokine receptor 3 (*CCR3* [MIM 601268], $n = 1$), chemokine receptor 5 (*CCR5* [MIM 601373], $n = 2$), T cell–specific transcription factor 7 (*TCF7* [MIM 189908], $n = 1$), interleukin 9 (*IL9* [MIM 146931], $n = 1$), interleukin 1 α chain (*IL1A* [MIM 147760], $n = 1$), interleukin 1 β chain (*IL1B* [MIM 147720], $n = 2$), interleukin 5 receptor α chain (*IL5RA* [MIM 147851], $n = 1$), interleukin 6 (*IL6* [MIM 147620], $n = 2$), interleukin 10 (*IL10* [MIM 124092], $n = 1$), complement 3 (*C3* [MIM 120700], $n = 1$), complement 5 (*C5* [MIM 120900], $n = 1$), colony-stimulating factor 2 (*CSF2* [MIM 138960], $n = 1$), cytotoxic T lymphocyte–associated protein (*CTLA4* [MIM 123890], $n = 2$), leukotriene C4 synthase (*LTC4S* [MIM 246530], $n = 1$), nitrous oxide synthetase 3 (*NOS3* [MIM 163729], $n = 2$), nitrous oxide synthetase 2A (*NOS2A* [MIM 163730], $n = 1$), stromal cell–derived factor 1 (*SDF1* [MIM 600835], $n = 1$), lymphotoxin α (*LTA* [MIM 153440], $n = 1$), tumor necrosis factor (*TNF* [MIM 191160], $n = 2$), vitamin D receptor (*VDR* [MIM 601769], $n = 1$), and group-specific component (*GC* [MIM 139200], $n = 1$). Full descriptions of the polymorphisms included in this study are available at the authors' Web site (Association Studies in Hutterites). The samples 1, 2, and 3 (described in the "Null Distribution of the W_{QLS} and $W_{\chi^2_{corr}}$ Statistics" subsection above) corresponding to the three phenotypes—atopy, BHR, and asthma, respectively—were used to conduct the analysis. The association between these 48 SNPs and the three phenotypes was also tested in the Hutterites, using the transmission/disequilibrium test (TDT), as implemented in ASPEX (Hinds and Risch 1999), considering all the cases for which parental genotypes were available.

Results

Null Distribution of the W_{QLS} *and* $W_{\chi^2_{corr}}$ *Statistics*

Table 2 presents the results of simulation studies in samples 1 (+SPT), 2 (BHR), and 3 (asthma), assessing the empirical type I error of both the QLS and the corrected χ^2 statistics when the χ^2 distribution is used as the null distribution of the statistics. Four different choices of allele-frequency distribution are considered. The results of the noncorrected χ^2 test are also displayed, to highlight the increase in type I error when interindividual correlations are neglected. For a given sample, this increase shows nonnegligible variation from one allele case to another, even though the correction factor $\rho_{\chi^2_{corr}}$ does not depend on allele frequency. This presumably reflects the fact that $W_{\chi^2_{corr}}$ is only χ^2 distributed and that, in finite samples, the accuracy of the χ^2 approximation varies slightly, depending on the allele-frequency distribution. In our largest sample, sample 1, the real type I error for the noncorrected χ^2 with a nominal P value of 5% may be as large as 18% in the three-allele case. For both the QLS and the corrected χ^2 tests, the nominal type I error lies within the 95% CI of the real type I errors in all three samples for two of the three biallelic situations considered and for the triallelic situation. Similar results are obtained for the two smaller Hutterite subsamples, as can be seen in table 3. When the allele frequency becomes low (0.05), the χ^2 approximation seems to be slightly conservative or anticonservative for both tests in all five samples, a deviation possibly due to small numbers of observations of the minor allele when its frequency is low. Even though not exhaustive, these results tend to confirm that the χ^2 distribution is a reasonably good approximation of the null distribution of these two tests, as long as neither the allele frequency nor the sample size is too small. The sample size and allele frequency required for the χ^2 distribution to hold depend on the relationship among the individuals of the sample and are likely to differ from one population sample to another. We would recommend that exact simulations be performed for confirmation in case of a significant result associated with a small number of alleles in the case and/or the control sample.

Power of the W_{QLS} *and* $W_{\chi^2_{corr}}$ *Statistics*

Figure 1 displays the power of both the W_{QLS} and $W_{\chi^2_{corr}}$ statistics based on the analytical power calculations, for various alternatives defined by (p,r) pairs (allele frequency in the controls and difference in allele frequency between cases and controls), using the three dif-

Table 2

Empirical Type I Error of the QLS Test, the Corrected Case-Control χ^2 Test, and the Noncorrected χ^2 Test Estimated with 5,000 Simulations in Three Hutterite Samples

	Empirical Type I Error with Nominal Type I Error of					
	.05			.01		
Number of Alleles (Frequency) and Test	+SPT	BHR	Asthma	+SPT	BHR	Asthma
2 (.5/.5):						
QLS	.052	.046	.055	.013	.011	.012
Corrected χ^2	.055	.050	.054	.009	.013	.011
Noncorrected χ^2	.145	.071	.071	.054	.021	.019
2 (.2/.8):						
QLS	.045	.050	.047	.010	.010	.010
Corrected χ^2	.055	.056	.049	.010	.010	.010
Noncorrected χ^2	.140	.072	.069	.054	.016	.017
2 (.05/.95):						
QLS	.041	.038	.045	.013	.013	.014
Corrected χ^2	.045	.050	.042	.008	.008	.012
Noncorrected χ^2	.123	.066	.060	.043	.043	.016
3 (.3/.3/.4):						
QLS	.055	.055	.056	.013	.011	.012
Corrected χ^2	.054	.055	.051	.010	.014	.012
Noncorrected χ^2	.180	.082	.079	.073	.020	.020

Note.—Values outside the 95% CI of the nominal type I error are underlined.

ferent Hutterite samples described above. Note that these power calculations are expected to be more accurate for smaller values of *r*, which represent alternative models close to the null.

The corresponding noncentrality parameter ratios ($\lambda_{QLS}/\lambda_{\chi^2_{corr}}$) are 4.01 for sample 1, 2.36 for sample 2, and 2.14 for sample 3. These ratios depend only on the sample composition (i.e., the actual choice of cases and controls) and not on the alternative model, as can be seen from eqs. (10) and (11). In every situation considered, the approximate power calculated for W_{QLS} is higher than that calculated for $W_{\chi^2_{corr}}$, and, in fact, we show in appendix C that $\lambda_{QLS} \geqslant \lambda_{\chi^2_{corr}}$. The difference in power between the two tests tends to become smaller for small values of *r*, but the gain in power when using W_{QLS} instead of $W_{\chi^2_{corr}}$ remains nonnegligible. This point is of particular interest, because small values of *r* are more likely to occur in real data sets. Indeed, the difference in allele frequencies between cases and controls would reach its maximum when the marker is the functional variant. However, the most common situation is that the marker has an allele that is in linkage disequilibrium with the functional variant, corresponding to smaller values of *r*.

Table 3

Empirical Type I Error of the QLS Test and the Corrected χ^2 Test Estimated with 5,000 Simulations, for a Nominal Type I Error of .01 in Smaller Hutterite Samples and for Biallelic Markers

	Empirical Type I Error in					
	76 Cases/76 Controls with Allele Frequency of			30 Cases/30 Controls with Allele Frequency of		
Test	.5	.2	.05	.5	.2	.05
QLS	.012	.012	.019	.012	.01	.01
Corrected χ^2	.012	.012	.004	.012	.009	.0024

Note.—Values outside the 95% CI of the nominal type I error are underlined.

Testing Candidate Genes for Asthma in the Hutterites

The results for all the 48 SNPs with a noncorrected *P* value < .05 for at least one of the three tests (QLS test, corrected χ^2 test, or TDT) are presented in table 4 for asthma, BHR, and atopy. As expected, the number of observations available for the two case-control tests is larger than that for the TDT. Although, for example, 269 atopy cases and 323 controls are genotyped for the Val640Leu amino acid polymorphism in *SELP* (SELP_640), only 136 heterozygous parents are available for the TDT. Two association signals reached the 5% significance threshold after adjustment for 105 tests (35 different genes tested for three phenotypes), using the Bonferroni correction ($P = .000476$): SELP_640 and atopy, using the QLS test, and SCYA11_−1328 and BHR, using the corrected χ^2 test. If we consider the uncorrected threshold of .05, association signals were detected when the QLS test was used, with 15 other polymorphisms when the corrected χ^2 test was used but with only 3 polymorphisms when the TDT was used and none of the polymorphisms reached the significance threshold. Interestingly, the smallest *P* values obtained

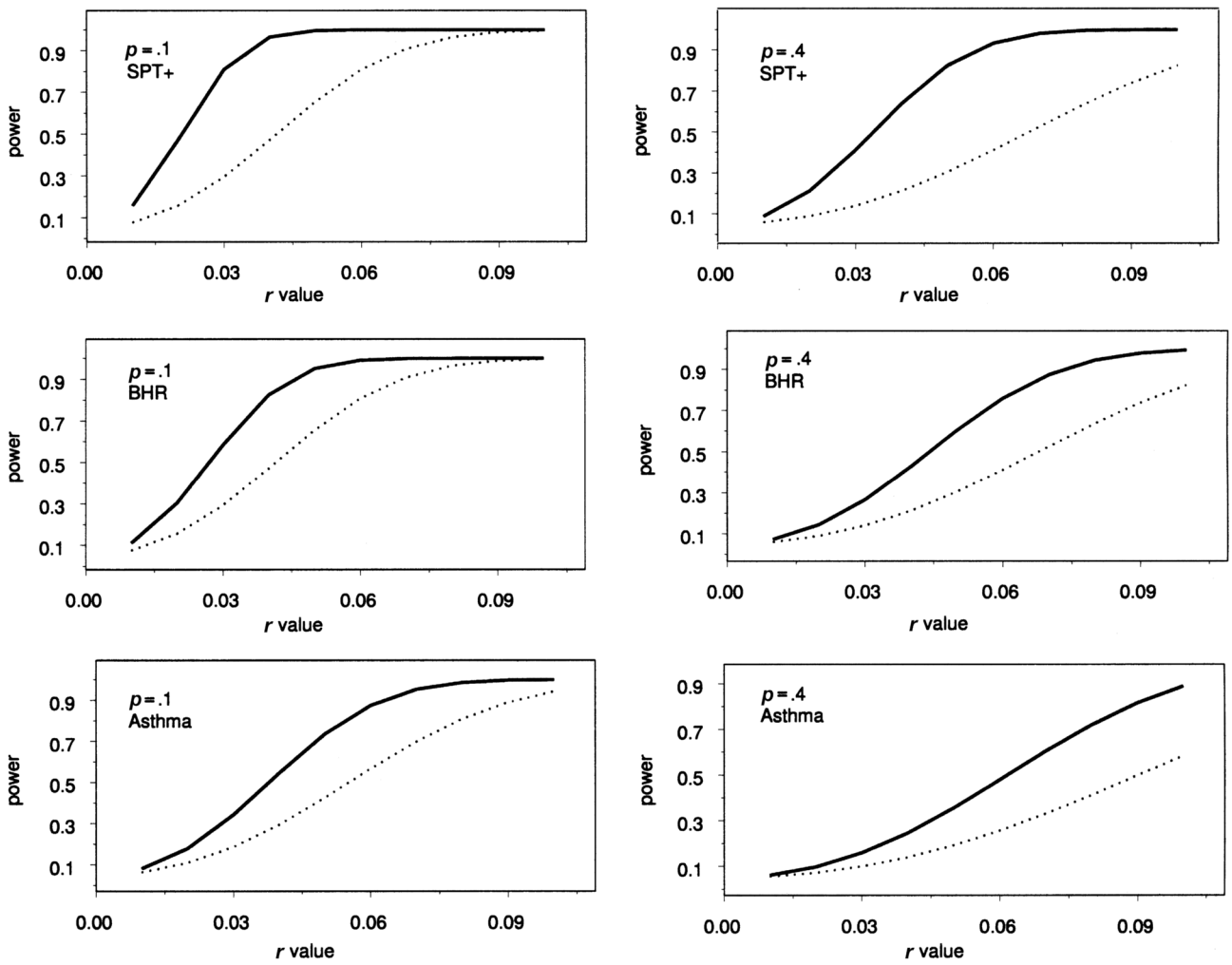

Figure 1 Power with a 5% nominal type I error of the QLS test (*solid lines*) and the corrected χ^2 test (*dotted lines*) for different alternative models defined by (p, r) pairs in three different Hutterite case-control samples: +SPT, BHR, and asthma. Power is presented as a function of r for two different values of p.

using the corrected χ^2 test and the QLS test were not systematically observed at the same loci. Except for the associations between CSF2_117 and atopy, which had a P value <.01 with both tests ($P = .0087$ with the corrected χ^2 test and $P = .0038$ with the QLS test) and between NOS3_298 and asthma, which had a P value close to .01 ($P = .0084$ with the corrected χ^2 test and $P = .011$ with the QLS test), all other association signals with a P value <.01 were observed with only one of the two tests (SCYA11_−1328 and asthma, SCYA11_−1328 and BHR, and SDF1_3UTR and asthma, using the corrected χ^2 test; LTC4S_444 and BHR, C3_102 and atopy, and SELP_640 and atopy, using the QLS test). Even though the QLS test is locally more powerful than the corrected χ^2 test under certain regularity conditions, the probability that the corrected χ^2 test provides a smaller P value than the QLS test in any particular case is not negligible. Furthermore, we have not studied the power of both tests when the alternative is not local (e.g., in the event of strong difference in allele frequency between cases and controls; note that analytic power calculations are not feasible for that case). In particular, the results shown on figure 1 were valid for values of r close to 0. We cannot rule out the possibility that the corrected χ^2 test might perform better for some alternatives, as suggested by the results of our data analysis.

Discussion

Recent progress in unraveling the genetic complexity of common diseases suggests that susceptibility is due to numerous genetic factors with modest effects. In this context, the study of isolated populations with negligible

Table 4

Results of Association Studies with Asthma, BHR, and Atopy, Using the QLS Test, the Corrected χ^2 Test, and the TDT

Phenotype and RS Number	Marker	N_c	N_t	p	P			
					χ^2_{corr}	QLS	TDT	TR:NT
Asthma:								
1801157	SDF1_3UTR[a]	65	357	.64	.00087	.037	.032	22:9
4795895	SCYA11_−1328[b]	65	362	.79	.0096	.14	.38	14:20
1799983	NOS3_298[c]	72	404	.52	.0084	.011	.89	26:24
25882	CSF2_117[d]	66	372	.84	.037	.04	.83	11:9
17611	C5_802[e]	66	362	.53	.034	.04	1.0	21:20
1800872	IL10_−571[f]	65	366	.55	.033	.077	.65	18:22
1042713	ADRB2_16[g]	71	400	.64	.025	.97	1.0	27:27
BHR:								
361525	TNF_−238[h]	147	409	.87	.043	.46	.74	23:20
4795895	SCYA11_−1328[b]	135	262	.79	.00038	.099	.33	29:38
1799983	NOS3_298[c]	144	404	.52	.013	.018	.26	55:43
1137933	NOS2A_346[i]	136	370	.76	.046	.12	.61	28:33
730012	LTC4S_−444[j]	133	354	.8	.06	.002	.47	22:28
25882	CSF2_117[d]	136	372	.83	.21	.04	.4	29:22
1800872	IL10_−571[f]	134	366	.56	.035	.19	.19	42:45
6133	SELP_640[k]	137	372	.82	.11	.012	.027	43:24
Atopy:								
2569190	CD14_−260[l]	259	319	.61	.27	.03	.2	87:70
1800779	NOS3_−922[m]	285	361	.56	.36	.034	1.0	89:89
5742909	CTLA4_−318[n]	261	320	.77	.013	.013	.51	59:67
25882	CSF2_117[d]	265	330	.81	.0087	.0038	.05	63:42
2230199	C3_102[o]	238	297	.72	.64	.0049	.23	46:34
2290608	IL5R_−80[p]	258	326	.93	.028	.074	.45	34:27
6133	SELP_640[k]	269	323	.82	.068	.000002	.0014	87:49
1041163	VCAM1_−1594[q]	271	330	.91	.49	.017	.48	40:33

Note.—SNPs with an associated P value <.05 for at least one of the three tests are reported, and P values <.05 are underlined. Descriptions of the SNPs can be found on the dbSNP Home Page, using their reference SNP (RS) number. The associated alleles are underlined in the footnotes below. The number of genotypes available in cases (N_c) and controls (N_t), the major allele frequency in the case-control sample as a whole (p) and the number of transmitted:nontransmitted major alleles in the TDT sample (TR:NT) are displayed.

[a] G→A in 3′ UTR in *SDF1*.
[b] G→A in promoter region position −1328 in *SCYA11* (eotaxin).
[c] Glu298Asp in *NOS3*.
[d] Ile 117Thr in *CSF2*.
[e] Val 802Ile in *C5*.
[f] C→A in promoter region position −571 in *IL10*.
[g] Gly16 Arg in *ADRB2*.
[h] G→A in promoter region position −238 in *TNF*.
[i] C→T synonymous change (Asp346) in *NOS2A*.
[j] A→C in promoter region position −444 in *LTC4S*.
[k] Val640Leu in *SELP*.
[l] C→T in promoter region position −260 in *CD14*.
[m] A→G in promoter region position −922 in *NOS3*.
[n] T→C in promoter region position −318 in *CTLA4*.
[o] Arg102Gly in *C3*.
[p] G→A in promoter region position −80 in *IL5R*.
[q] T→C in promoter region position −1594 in *VCAM1*.

migration will continue to be important, because the relative genetic (and often environmental) homogeneity may result in less complex underlying models of susceptibility. In such studies, however, limited sample sizes can be a serious problem. Making use of all affected individuals, not only those for whom parents are available (as in the TDT), will increase the power to detect susceptibility loci. However, to use case-control association tests, rather than family-based association tests, one needs to correct for the relatedness among individuals. In populations with known genealogy, it is preferable to use this information. This is the rationale for

developing the QLS test described in the present article. Indeed, making use of the quasi-likelihood framework, we were able to derive a valid test for allelic association in the presence of strong but known correlations among alleles. We showed that this approach may be more powerful than simply correcting the variance of the χ^2 test under certain conditions and is asymptotically the locally most powerful test in a general class of linear tests. Furthermore, we detected a highly significant association by use of this test.

Recent attempts to correct association tests either for unknown population stratification and cryptic relatedness (Devlin and Roeder 1999) or for the sampling of related subjects in outbred populations (Slager and Schaid 2001) used the Armitage trend test (Armitage 1955), which is genotype based rather than allele based. Indeed, as shown by Sasieni (1997) and further explored by Devlin and Roeder (1999), both the Armitage trend test and the allele-based test contrast allelic frequencies between cases and controls, while considering an additive effect for alleles. In addition, the Armitage trend test corrects for possible departure from Hardy-Weinberg equilibrium in the sample, whereas the allelic test does not. Apart from genotyping errors, departures from Hardy-Weinberg equilibrium in isolated populations such as the Hutterites are mainly due to nonnegligible inbreeding (Bourgain et al. 2002). The QLS test presented here, which is performed conditional on the pedigree structure and explicitly models inbreeding, is thus likely to be a correct test for allelic association even though it is performed at the level of alleles rather than genotypes. We showed how the QLS test may also be used in outbred populations when relatives of any kind are sampled. We should stress that, for this approach to be correct in outbred pedigrees, no departure from Hardy-Weinberg equilibrium should be observed at the loci under study. We did not compare the power of the Slager and Schaid (2001) approach with ours, because we focused on pedigrees that are too complex to be handled by their method. We believe that in simpler pedigrees the Slager and Schaid (2001) approach might perform better than the corrected χ^2 test presented here, because the former method uses a corrected variance, computed conditional on the IBD information obtained from the pedigree data, whereas our method uses an unconditional corrected variance (exact computation of the conditional variance is infeasible in complex pedigrees). With regard to the comparison between the corrected χ^2 test and the QLS test in small pedigrees, it is not obvious which test would be more powerful. The outcome might depend on the kind of correlations among the individuals in the sample, as well as on the informativeness of the markers used in the analysis. We note that, if desired, a QLS-type version of the Slager and Schaid (2001) approach could be performed, in small pedigrees, that should be more powerful than both our QLS approach and their approach.

The controversy as to whether case-control tests are preferable to family-based association tests has been ongoing for several years (Morton and Collins 1998; Risch and Teng 1998; McGinnis 2000). In the present article, we did not formally compare the power of the QLS test and the TDT. We note that the TDT can be expressed as a conditional-likelihood–score test (Clayton 1999), whereas our approach is an unconditional QLS test, which can be viewed as an approximation to the unconditional-likelihood–score test. The unconditional approach would be expected to perform better in the absence of confounding population substructure. However, the formal comparison of these two tests is not straightforward, because the QLS test is only an approximation to the unconditional-likelihood–score test. Furthermore, the correction for relatedness among individuals may be interpreted as a reduction of the effective sample size. Indeed, the weight associated with an allele in the W_{QLS} statistic decreases as the amount of relatedness of this allele with other alleles of the sample increases. The difference in effective sample size between the QLS test and the TDT is thus not as large as it might first appear. No general statement on power comparison can be made from our analysis of real data, because none of the genes investigated have been definitively established as risk factors for asthma, BHR, or atopy. Nonetheless, no associations were detected by the TDT that were not detected by the QLS test, and, in each case, the association signal was stronger when the QLS test was used. A number of quite significant associations would have been missed if we used only the TDT.

The most significant association among the 48 markers examined in the present study was between a polymorphism at amino acid 640 (Val→Leu) in *SELP* and atopy, detected by the QLS test ($P = .000002$) and the TDT ($P = .0014$). Although associations between polymorphisms in *SELP* and asthma-related phenotypes have not been reported previously, P-selectin is an outstanding functional candidate. Indeed, P-selectin is an adhesion molecule expressed on the surface of activated platelets and endothelial cells. It contributes to both bronchoconstriction and inflammation in murine models of allergic airway reactivity (Lukacs et al. 2002). The results of this study indicate that the common Val630 allele is a risk allele for atopy. The other significant association observed in this study was between a promoter polymorphism in *SCYA11* (−1328G→A) and BHR, detected by the corrected χ^2 test ($P = .000383$). The −1328A allele was significantly associated with BHR and with asthma ($P = .00956$), although the latter association likely reflects the fact that our definition of asthma included BHR. The *SCYA11* gene encodes

eotaxin, the predominant eosinophil chemoattractant involved in allergic inflammation. Another variant in the promoter region of this gene was associated with IgE level in patients with atopic dermatitis (AD) but not with AD itself or asthma (Tsunemi et al. 2002). Thus, variation in this gene may influence a variety of atopic phenotypes. Other novel associations identified in the present study are between asthma and the 880G allele in the 3′ UTR of *SDF1* and between atopy and the 102Gly allele in *C3*. These genes are both good functional candidates for asthma-related phenotypes. *SDF1* encodes a small chemokine (C-K-C motif) that is a highly potent lymphocyte chemoattractant and is the principal ligand for CXCR4, which is also a coreceptor for CD4. Furthermore, the 3′ UTR polymorphism investigated in this study has been associated with delayed progression to AIDS (Winkler et al. 1998), suggesting that its role as a viral receptor might influence asthma susceptibility. C3 deficiency in an allergen-induced model of airway allergy was associated with diminished airway responsiveness and lung eosinophilia (Drouin et al. 2001). Thus, variation in this gene may influence allergic responses, as indicated by our study. Most of the remaining associations detected in the present study, including many of the more modest associations, have been reported elsewhere for the same or related phenotypes (Genetic Association Database.

Finally, we believe that the two tests presented here are not only of general interest for studies involving related individuals but may also be particularly interesting tools to take full advantage of founder populations for gene mapping of complex traits.

Acknowledgments

The authors acknowledge Dr. Mark Abney for discussion and critical comments. This study was supported by National Institutes of Health grants HL56399, HL63533, HG001645, and DK55889 and an Institut National de la Santé et de la Recherche Medicale postdoctoral fellowship (to C.B.).

Appendix A

Expression of W_{QLS} in the Multiallelic Case

Notation: Given the $n \times m$ matrix $\mathbf{A}$ with (i,j)th element a_{ij} and the $p \times q$ matrix $\mathbf{B}$ with (i,j)th element b_{ij}, their Kronecker product, denoted $\mathbf{A} \otimes \mathbf{B}$, is the $np \times mq$ matrix with block structure

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{1,1}\mathbf{B} & \dots & a_{1,m}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{n,1}\mathbf{B} & \dots & a_{n,m}\mathbf{B} \end{pmatrix}_{np \times mq} .$$

Consider a locus with a distinct alleles. Define $p = (p_1, \dots p_{a-1})^{\mathrm{T}}$ as the $(a-1)$-vector of control allele frequencies, and define $r = (r_1, \dots r_{a-1})^{\mathrm{T}}$ as the $(a-1)$-vector of differences in allele frequencies between cases and controls. Our model stipulates that under the null hypothesis, $r = r_0 = \mathbf{0}$, in which case we obtain $E_0(Y) = \mu_0 = p \otimes D_p$ and $\mathrm{Var}_0(Y) = \mathbf{\Sigma} = \mathbf{F} \otimes \mathbf{K}$ with $\mathbf{K}$ defined in eq. (1) and $\mathbf{F}$ defined in eq. (6). Our model for the alternative hypothesis specifies $E(Y) = \mu = p \otimes D_p + r \otimes D_r$, and, as in the biallelic case, we allow $\mathrm{Var}(Y) = \mathbf{\Omega}$ to depend on both r and p, provided that, when $r = \mathbf{0}$, $\mathbf{\Omega} = \mathbf{\Sigma}$ and that $\mathbf{\Omega}$ be differentiable and invertible. Let $\mathbf{D}_\pi = \partial\mu/\partial p$ and $\mathbf{D}_\rho = \partial\mu/\partial r$. $\mathbf{D}_\pi$ and $\mathbf{D}_\rho$ are $N(a-1) \times (a-1)$ matrices with the kth column equal to $\partial\mu/\partial p_k$ and $\partial\mu/\partial r_k$, respectively. Then $\mathbf{D}_\pi = \mathbf{I}_{a-1} \otimes D_p$ and $\mathbf{D}_\rho = \mathbf{I}_{a-1} \otimes D_r$, where $\mathbf{I}_{a-1}$ is the identity matrix of dimension $(a-1)$. $U_r(r_0,p) = \mathbf{D}_\rho^{\mathrm{T}}\mathbf{\Sigma}^{-1}(Y - \mu_0)$ becomes $U_r(r_0,p) = (\mathbf{I}_{a-1} \otimes D_r)^{\mathrm{T}}(\mathbf{F} \otimes \mathbf{K})^{-1}(Y - p \otimes D_p)$. From the properties of the Kronecker product (e.g., see Schott 1996), it follows that $U_r(r_0,p) = (\mathbf{I}_{a-1} \otimes D_r^{\mathrm{T}})(\mathbf{F}^{-1} \otimes \mathbf{K}^{-1})(Y - p \otimes D_p) = [\mathbf{F}^{-1} \otimes (D_r^{\mathrm{T}}\mathbf{K}^{-1})](Y - p \otimes D_p)$ and $U_r(r_0,p)^{\mathrm{T}} = (Y - p \otimes D_p)^{\mathrm{T}}[\mathbf{F}^{-1} \otimes (\mathbf{K}^{-1}D_r)]$.

Similarly, $\mathrm{Var}_0[U_r(r_0,p)] = \mathbf{D}_\rho^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{D}_\rho - \mathbf{D}_\rho^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{D}_\pi(\mathbf{D}_\pi^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{D}_\pi)^{-1}\mathbf{D}_\pi^{\mathrm{T}}\mathbf{\Sigma}^{-1}\mathbf{D}_\rho$ becomes $\mathrm{Var}_0[U_r(r_0,p)] = \mathbf{F}^{-1} \otimes [D_r^{\mathrm{T}}\mathbf{K}^{-1}D_r - D_r^{\mathrm{T}}\mathbf{K}^{-1}D_p(D_p^{\mathrm{T}}\mathbf{K}^{-1}D_p)^{-1}D_p^{\mathrm{T}}\mathbf{K}^{-1}D_r]$. Then, evaluating $W_{\mathrm{QLS}} = U_r(r_0,p)^{\mathrm{T}}\{\mathrm{Var}_0[U_r(r_0,p)]\}^{-1}U_r(r_0,p)$ at $p = \hat{p}_0$ (the quasi-likelihood estimator of p when $r = r_0$), we get

$$W_{\mathrm{QLS}} = (Y - \hat{p}_0 \otimes D_p)^{\mathrm{T}}\{\hat{\mathbf{F}}^{-1} \otimes \{\mathbf{K}^{-1}D_r[D_r^{\mathrm{T}}\mathbf{K}^{-1}D_r - D_r^{\mathrm{T}}\mathbf{K}^{-1}D_p(D_p^{\mathrm{T}}\mathbf{K}^{-1}D_p)^{-1}D_p^{\mathrm{T}}\mathbf{K}^{-1}D_r]^{-1}D_r^{\mathrm{T}}\mathbf{K}^{-1}\}\}(Y - \hat{p}_0 \otimes D_p) ,$$

where $\hat{\mathbf{F}}^{-1}$ corresponds to matrix $\mathbf{F}^{-1}$ evaluated at $p = \hat{p}_0$. Eq. (7) is easily derived from the latter expression of W_{QLS}.

Appendix B

General Framework for the QLS and the χ^2 Test in the Multiallelic Case

In the multiallelic case, using the framework presented in appendix A, we note that $\hat{p}_0 = \{\mathbf{I}_{a-1} \otimes [(D_p^T \mathbf{K}^{-1} D_p)^{-1} D_p^T \mathbf{K}^{-1}]\} Y$. Define $V_{QLS} = \mathbf{I}_{a-1} \otimes V_1$ with V_1 given in the text and let $S_{QLS} = V_{QLS}^T Y$. We note that $U_r(r_0,\hat{p}_0) = \hat{F}^{-1} S_{QLS}$. We have $\mathrm{Var}_0(S_{QLS}) = V_{QLS}^T(\mathbf{F} \otimes \mathbf{K})V_{QLS}$. Because p is unknown, we evaluate $\mathrm{Var}_0(S_{QLS})$ at $p = \hat{p}_0$ to obtain $V_{QLS}^T(\hat{\mathbf{F}} \otimes \mathbf{K})V_{QLS}$. Then $S_{QLS}^T \mathrm{Var}_0^{-1}(S_{QLS}) S_{QLS} = S_{QLS}^T \hat{\mathbf{F}}^{-1} [\hat{\mathbf{F}}^{-1} V_{QLS}^T (\hat{\mathbf{F}} \otimes \mathbf{K}) V_{QLS} \hat{\mathbf{F}}^{-1}]^{-1} \hat{\mathbf{F}}^{-1} S_{QLS} = U_r^T(r_0,\hat{p}_0)\,[(\hat{\mathbf{F}}^{-1} \otimes V_1^T)(\hat{\mathbf{F}} \otimes \mathbf{K})(\hat{\mathbf{F}}^{-1} \otimes V_1)]^{-1} U_r(r_0,\hat{p}_0)$. Note that $(\hat{\mathbf{F}}^{-1} \otimes V_1^T)(\hat{\mathbf{F}} \otimes \mathbf{K})(\hat{\mathbf{F}}^{-1} \otimes V_1) = \hat{\mathbf{F}}^{-1} \otimes (V_1^T \mathbf{K} V_1) = \hat{\mathbf{F}}^{-1} \otimes [D_r^T \mathbf{K}^{-1} D_r - D_r^T \mathbf{K}^{-1} D_p (D_p^T \mathbf{K}^{-1} D_p)^{-1} D_p^T \mathbf{K}^{-1} D_r]$, which matches the formula for $\mathrm{Var}_0[U_r(r_0,p)]$ given in appendix A once $\hat{p}_0$ is substituted in for p. Thus, $S_{QLS}^T \mathrm{Var}_0^{-1}(S_{QLS}) S_{QLS} = W_{QLS}$.

A possible expression for the classical χ^2 test for case-control association in the biallelic case is

$$T_{\chi^2} = \frac{\left[\sum_{i \in \text{cases}} (Y_i - \bar{Y})\right]^2}{\frac{1}{2}\bar{Y}(1-\bar{Y})(N_c - \frac{N_c^2}{N})}$$

with $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$. We note that T_{χ^2} may also be written

$$T_{\chi^2} = \frac{(S_{\chi^2})^2}{\frac{1}{2}\bar{Y}(1-\bar{Y})[D_r^T D_r - (D_r^T D_p)^2 (D_p^T D_p)^{-1}]}$$

with $S_{\chi^2} = [D_r^T - (D_r^T D_p)(D_p^T D_p)^{-1} D_p^T] Y = V_2^T Y$. When the correlations among all the individuals as well as between the two alleles of an individual are zero then the variance of Y when $r = r_0$ is $\mathrm{Var}_0(Y) = \frac{1}{2}p(1-p)\mathbf{I}_N$ where $\mathbf{I}_N$ is the $N \times N$ identity matrix. Thus, $\mathrm{Var}_0(S_{\chi^2}) = \frac{1}{2}p(1-p)(D_r^T D_r - (D_r^T D_p)^2 (D_p^T D_p)^{-1})$. If $\bar{Y}$ is used as an estimator of p to compute $\mathrm{Var}_0(S_{\chi^2})$, then

$$T_{\chi^2} = \frac{[S_{\chi^2}]^2}{\mathrm{Var}_0(S_{\chi^2})} = W_{\chi^2}\ .$$

Similarly, in the multiallelic case, $\mathrm{Var}_0(Y) = \mathbf{F} \otimes \mathbf{I}_N$. Thus, $S_{\chi^2} = V_{\chi^2}^T Y$ with $V_{\chi^2} = \mathbf{I}_{a-1} \otimes V_2$ and $\mathrm{Var}_0(S_{\chi^2}) = \mathbf{F} \otimes V_2^T V_2$. Finally,

$$W_{\chi^2} = \sum_{k=1}^{a-1} \sum_{i=1}^{a-1} (\hat{\mathbf{F}}^{-1})_{ik} Y_k^T V_2 (V_2^T V_2)^{-1} V_2^T Y_i = T_{\chi^2}$$

Appendix C

Proof That the QLS Test Is Asymptotically the Locally Most Powerful Test of the *W* Class

We focus here on the biallelic situation. Consider the statistics of the W class described in the text, that is, $W = S^T(\mathrm{Var}_0(S))^{-1} S$ with $S = V^T Y$, $V \neq \mathbf{0}$ and known, $E_0(S) = 0$. Under the alternative hypothesis, $E_{H_1}(Y) = pD_p + rD_r$. Under the null hypothesis, $r = 0$. $E_0(S) = 0$ implies that $pV^T D_p = 0$, which must hold for any p, so that $V^T D_p = 0$. As described in the text, $\mathrm{Var}_0(Y) = \frac{1}{2}p(1-p)\mathbf{K}$, where $\mathbf{K}$ is the correlation matrix described in eq. (1). Thus, $\mathrm{Var}_0(S) = \frac{1}{2}p(1-p)V^T \mathbf{K} V$ and

$$W = \frac{(V^T Y)^2}{\frac{1}{2}p(1-p)V^T \mathbf{K} V}\ .$$

If we assume that W has a χ^2 distribution under the null hypothesis and a noncentral χ^2 distribution with noncentrality parameter λ under local alternatives, the locally most powerful statistic of this class is the one maximizing

the value of λ over all $V \neq \mathbf{0}$, such that $V^{\mathrm{T}}D_p = 0$. By definition $\sqrt{\lambda}$ is the expectation of $\sqrt{W}$ under the alternative hypothesis; thus,

$$\sqrt{\lambda} = \left| \frac{V^{\mathrm{T}}}{\sqrt{\frac{1}{2}p(1-p)V^{\mathrm{T}}\mathbf{K}V}} E_{\mathrm{H}_1}(Y) \right| .$$

In practice, p is a nuisance parameter, and we use its estimator $\hat{p}$. Under certain regularity conditions, this is asymptotically equivalent to the case in which the true p is used. In what follows, we use the true value of p and will thus derive the asymptotic value of λ:

$$\sqrt{\lambda} = \left| \frac{V^{\mathrm{T}}}{\sqrt{\frac{1}{2}p(1-p)V^{\mathrm{T}}\mathbf{K}V}} (pD_p + rD_r) \right| .$$

Maximizing λ is thus equivalent to maximizing

$$m = \left| \frac{V^{\mathrm{T}}D_r}{\sqrt{V^{\mathrm{T}}\mathbf{K}V}} \right|$$

over all $V \neq \mathbf{0}$, such that $V^{\mathrm{T}}D_p = 0$.

We first consider a modified version of the problem by maximizing

$$m' = \left| \frac{\omega^{\mathrm{T}}Z}{\sqrt{\omega^{\mathrm{T}}\omega}} \right|$$

over all $\omega \neq 0$ such that $\omega^{\mathrm{T}}R = 0$ with Z and R being N-vectors and $R \neq \mathbf{0}$. By definition, $|\omega^{\mathrm{T}}Z| = |\omega||Z|| \cos(\theta_{(\omega,Z)})|$, where $\theta_{(\omega,Z)}$ is the angle between ω and Z. Thus, $m' = |Z|| \cos(\theta_{(\omega,Z)})|$, and we need only to maximize $\cos(\theta_{(\omega,Z)})$ over all $\omega \neq 0$ subject to $\omega^{\mathrm{T}}R = 0$. By geometry, the maximizing ω, $\omega_{\max}$, is any scalar multiple of the projection of Z onto the subspace orthogonal to R. Thus, $\omega_{\max} \propto (\mathbf{I}_N - R(R^{\mathrm{T}}R)^{-1}R^{\mathrm{T}})Z$, where $\mathbf{I}_N$ is the $N \times N$ identity matrix (e.g., see Schott 1996).

We consider now the initial problem. $\mathbf{K}$ being symmetric positive definite, we can derive its Cholesky decomposition $\mathbf{K} = \mathbf{C}^{\mathrm{T}}\mathbf{C}$, where $\mathbf{C}$ is an invertible upper triangular matrix. Define ω by $\omega = \mathbf{C}V$. Because $\mathbf{C}$ is invertible, $V \neq 0 \Leftrightarrow \omega \neq 0$. Furthermore, $V^{\mathrm{T}}D_p = 0 \Leftrightarrow \omega^{\mathrm{T}}\mathbf{C}^{-\mathrm{T}}D_p = 0$, and we have

$$m = \frac{\omega^{\mathrm{T}}\mathbf{C}^{-\mathrm{T}}D_r}{\sqrt{\omega^{\mathrm{T}}\omega}} .$$

Define $R = \mathbf{C}^{-\mathrm{T}}D_p$ and $Z = \mathbf{C}^{-\mathrm{T}}D_r$. We get from the previous modified version of the problem that $\omega_{\max} \propto [\mathbf{I_N} - \mathbf{C}^{-\mathrm{T}}\mathbf{D_p}(\mathbf{D_p^T}\mathbf{C}^{-1}\mathbf{C}^{-\mathrm{T}}\mathbf{D_p})^{-1}\mathbf{D_p^T}\mathbf{C}^{-1}]\mathbf{C}^{-\mathrm{T}}D_r$ and the corresponding V is $V_{\max} \propto (\mathbf{K}^{-1}D_r - \mathbf{K}^{-1}D_p(D_p^{\mathrm{T}}\mathbf{K}^{-1}D_p)^{-1}D_p^{\mathrm{T}}\mathbf{K}^{-1}D_r)$. Note that $V_{\max} \propto V_{\mathrm{QLS}}$ so that

$$W_{\max} = \frac{(V_{\max}^{\mathrm{T}}Y)^2}{\frac{1}{2}p(1-p)V_{\max}^{\mathrm{T}}\mathbf{K}V_{\max}} = \frac{(V_{\mathrm{QLS}}^{\mathrm{T}}Y)^2}{V_{\mathrm{QLS}}^{\mathrm{T}}\mathbf{\Sigma}V_{\mathrm{QLS}}} = W_{\mathrm{QLS}} .$$

It follows that, when p is known, the noncentrality parameter for W_{QLS} is larger than that for any other statistic in the W class. When p is unknown, we argue that replacement of p by

$$\hat{p}_0 = p + O_p\left(\frac{1}{\sqrt{n}}\right)$$

makes W_{QLS} asymptotically the locally most powerful test in the W class. Note that all our data are correlated with

an $n \times n$ correlation matrix **K**. Thus, whether or not we have (i) W asymptotically χ^2 under the null and (ii) noncentral χ^2 for local alternatives and (iii)

$$\hat{p}_0 = p + O_p\left(\frac{1}{\sqrt{n}}\right)$$

for local alternatives, will depend on assumptions about how **K** behaves as $n \to \infty$. Subject to regularity conditions on **K** ensuring that statements (i), (ii), and (iii) hold, then W_{QLS} will be asymptotically the locally most powerful test in the W class.

Electronic-Database Information

URLs for data presented herein are as follows:

Association Studies in Hutterites, http://www.genes.uchicago.edu/hutterite/inflasnps/asthma (for full descriptions of the polymorphisms included in this study)

dbSNP Home Page, http://www.ncbi.nlm.nih.gov/SNP/index.html

Genetic Association Database, http://geneticassociationdb.nih.gov/

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for *IL4, IL4RA, IL13, ADRB2, ICAM1, VCAM1, SELE, SELP, FCERB1, CD14, UGB, TGFB1, SCYA11, CCR2, CCR3, CCR5, TCF7, IL9, IL1A, IL1B, IL5RA, IL6, IL10, C3, C5, CSF2, CTLA4, LTC4S, NOS3, NOS2A, SDF1, LTA, TNF, VDR,* and *GC*)

References

Abney M, Ober C, McPeek MS (2002) 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 70:920–934

Armitage P (1955) Tests for linear trends in proportions and frequencies. Biometrics 11:375–386

Bourgain C, Newmann D, Ober C, McPeek MS (2002) Performing classical statistical genetic tests in founder populations. Am J Hum Genet Suppl 71:177

Boyce AJ (1983) Computation of inbreeding and kinship coefficients on extended pedigrees. J Hered 74:400–404

Clayton D (1999) A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. Am J Hum Genet 65:1170–1177

Cox DR, Hinkley DV (1974) Theoretical statistics. Chapman and Hall, London

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004

Drouin SM, Kildsgaard J, Haviland J, Zabner J, Jia HP, McCray PB, Tack BF, Wetsel RA (2001) Expression of the complement anaphylatoxin C3a and C5a receptors on bronchial epithelial and smooth muscle cells in models of sepsis and asthma. J Immunol 166:2025–2032

Heyde C (1997) Quasi-likelihood and its application: a general approach to optimal parameter estimation. Springer, New York

Hinds D, Risch N (1996) The ASPEX package: affected sib pair exclusion mapping. ftp://lahmed.stanford.edu/pub/aspex/

Lukacs NW, John A, Berlin A, Bullard DC, Knibbs R, Stoolman LM (2002) E- and P-selectins are essential for the development of cockroach allergen–induced airway responses. J Immunol 169:2120–2125

McCullagh P, Nelder JA (1989) Generalized linear models, 2nd ed. Chapman and Hall, London

McGinnis R (2000) General equations for P_t, P_s, and the power of the TDT and the affected–sib-pair test. Am J Hum Genet 67:1340–1347

Mirel DB, Valdes AM, Lazzeroni LC, Reynolds RL, Erlich HA, Noble JA (2002) Association of IL4R haplotypes with type 1 diabetes. Diabetes 51:3336–3341

Morton NE, Collins A (1998) Tests and estimates of allelic association in complex inheritance. Proc Natl Acad Sci USA 95:11389–11393

Newman DL, Abney M, McPeek MS, Ober C, Cox NJ (2001) The importance of genealogy in determining genetic associations with complex traits. Am J Hum Genet 69:1146–1148

Ober C, Leavitt SA, Tsalenko A, Howard TD, Hoki DM, Daniel R, Newman DL, Wu X, Parry R, Lester LA, Solway J, Blumenthal M, King RA, Xu J, Meyers DA, Bleecker ER, Cox NJ (2000) Variation in the interleukin 4–receptor α gene confers susceptibility to asthma and atopy in ethnically diverse populations. Am J Hum Genet 66:517–526

Risch N, Teng J (1998) The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. I. DNA pooling. Genome Res 8:1273–1288

Sasieni PD (1997) From genotypes to genes: doubling the sample size. Biometrics 53:1253–1261

Schott JR (1996) Matrix analysis for statistics. John Wiley, New York

Slager SL, Schaid DJ (2001) Evaluation of candidate genes in case-control studies: a statistical method to account for related subjects. Am J Hum Genet 68:1457–1462

Tsunemi Y, Saeki H, Nakamura K, Sekiya T, Hirai K, Fujita H, Asano N, Tanida Y, Kakinuma T, Wakugawa M, Torii H, Tamaki K (2002) Eotaxin gene single nucleotide polymorphisms in the promoter and exon regions are not associated with susceptibility to atopic dermatitis, but two of

them in the promoter region are associated with serum IgE levels in patients with atopic dermatitis. J Dermatol Sci 29: 222–228

Winkler C, Modi W, Smith MW, Nelson GW, Wu X, Carrington M, Dean M, Honjo T, Tashiro K, Yabe D, Buchbinder S, Vittinghoff E, Goedert JJ, O'Brien TR, Jacobson LP, Detels R, Donfield S, Willoughby A, Gomperts E, Vlahov D, Phair J, O'Brien SJ (1998) Genetic restriction of AIDS pathogenesis by an SDF-1 chemokine gene variant: ALIVE Study, Hemophilia Growth and Development Study (HGDS), Multicenter AIDS Cohort Study (MACS), Multicenter Hemophilia Cohort Study (MHCS), San Francisco City Cohort (SFCC). Science 279:389–393

Wu X, Ober C, McPeek MS (2000) A quasi-likelihood method for allele frequency estimation. Am J Hum Genet Suppl 67: A1810