BIOINFORMATICS ORIGINAL PAPER

Vol. 28 no. 13 2012, pages 1729–1737
doi:10.1093/bioinformatics/bts259

Genetics and population analysis

Advance Access publication May 3, 2012

Analysis of case–control association studies with known risk variants

Noah Zaitlen[1,2,3,4,*], Bogdan Paşaniuc[1,2,3,4], Nick Patterson[3], Samuela Pollack[1], Benjamin Voight[3,5,6], Leif Groop[7], David Altshuler[3,5,6], Brian E. Henderson[8], Laurence N. Kolonel[9], Loic Le Marchand[9], Kevin Waters[8], Christopher A. Haiman[8], Barbara E. Stranger[3,6,10], Emmanouil T. Dermitzakis[11], Peter Kraft[1,2,3,4] and Alkes L. Price[1,2,3,4,*]

[1]Department of Epidemiology, [2]Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, [3]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, [4]Program in Molecular and Genetic Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA, [5]Center for Human Genetic Research, Department of Molecular Biology and Diabetes Unit, Massachusetts General Hospital, [6]Departments of Genetics and of Medicine, Harvard Medical School, Boston, MA 02115, [7]Department of Clinical Sciences, Diabetes and Endocrinology Research Unit, Scania University Hospital Lund University, Malm, Sweden SE-205, [8]Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA 90089, [9]Epidemiology Program, Cancer Research Center, University of Hawaii, Honolulu, HI 96813, [10]Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115 and [11]Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland CH-1211

Associate Editor: Jeffrey Barrett

ABSTRACT

Motivation: The question of how to best use information from known associated variants when conducting disease association studies has yet to be answered. Some studies compute a marginal P-value for each Several Nucleotide Polymorphisms independently, ignoring previously discovered variants. Other studies include known variants as covariates in logistic regression, but a weakness of this standard conditioning strategy is that it does not account for disease prevalence and non-random ascertainment, which can induce a correlation structure between candidate variants and known associated variants even if the variants lie on different chromosomes. Here, we propose a new conditioning approach, which is based in part on the classical technique of liability threshold modeling. Roughly, this method estimates model parameters for each known variant while accounting for the published disease prevalence from the epidemiological literature.

Results: We show via simulation and application to empirical datasets that our approach outperforms both the no conditioning strategy and the standard conditioning strategy, with a properly controlled false-positive rate. Furthermore, in multiple data sets involving diseases of low prevalence, standard conditioning produces a severe drop in test statistics whereas our approach generally performs as well or better than no conditioning. Our approach may substantially improve disease gene discovery for diseases with many known risk variants.

Availability: LTSOFT software is available online http://www.hsph.harvard.edu/faculty/alkes-price/software/

Contact: nzaitlen@hsph.harvard.edu; aprice@hsph.harvard.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on November 4, 2011; revised on March 29, 2012; accepted on April 25, 2012

1 INTRODUCTION

The NHGRI catalog of Published Genome Wide Association Studies (GWAS) (Hindorff *et al.*, 2009) lists thousands of single nucleotide polymorphisms (SNPs) associated with several hundred complex phenotypes. However, it is currently unknown how to optimally use these discovered SNPs when conducting additional GWAS. Typically, known variants are ignored and SNPs are tested independently for association via logistic regression for case–control phenotypes and linear regression for quantitative phenotypes (McCarthy *et al.*, 2008). Occasionally, known variants are used as covariates in the regression models to determine additional signals exist in the data beyond those already discovered, as in recent studies of Type 2 diabetes (Voight *et al.*, 2010). We show that for standard case–control studies neither one of these strategies, testing SNPs marginally or standard conditioning on associated variants, is optimally powered to discover new loci. Surprisingly, standard conditioning will often dramatically decrease power (Kuo and Feingold, 2010). For example, in the Welcome Trust Case Control Consortium (WTCCC), Type 1 diabetes (T1D) dataset (WTCCC, 2007b), conditioning on a known variant on Chromosome 6 decreases the one degree of freedom (df) χ^2 statistic from a logistic regression likelihood ratio test by an average of 27% at independent known associated variants on entirely different chromosomes relative to the same test without conditioning on the

[*]To whom correspondence should be addressed.

Chromosome 6 variant. However, we find that if used properly, known variants can substantially improve study power and therefore represent an important resource in conducting future GWAS.

In this work, we thoroughly examine the use of known associated variants in the analysis of GWAS and their effects on SNPs that are completely unlinked to the known variants (i.e. on different chromosomes or distant loci). Previously, (Neuhaus, 1998) showed that in the case of logistic regression, including a covariate increased power when it was uncorrelated to the random variable being tested, but decreased power when it was correlated. Our extensive simulations and analysis of real gene expression and case–control data indicate that for randomly ascertained individuals such as those in a cross-sectional study, the practice of standard conditioning on known variants does indeed improve the power to discover new variants (Ma *et al.*, 2010; Neuhaus, 1998), with larger gains in power as the fraction of variance explained by the conditioned SNPs increases. However, in a balanced case–control study in which an equal number of cases and controls are ascertained based on disease status, standard conditioning on known variants significantly decreases power when the disease prevalence is low. This power loss is due to an induced non-independence between associated variants in case–control datasets. That is, SNPs that were completely uncorrelated in the population become correlated when individuals are collected in a case–control study design, and as predicted by Neuhaus (1998), there is a corresponding loss in power due to this correlation. This is true regardless of whether the data are generated under a liability threshold model of disease or the logit model of disease assumed by logistic regression. We show that the effect of standard conditioning on known variants is a function of prevalence, sample ascertainment, and the total phenotypic variance explained by the known variants. We give full analytic derivations of the non-centrality parameter of the conditioned and unconditioned tests detailing the scenarios when each improves or diminishes power.

To address this power loss in the case–control setting, we develop a new statistic, called LTSCORE, based on the liability threshold model (Dempster and Lerner, 1950; Falconer, 1967). LTSCORE properly accounts for study design and disease prevalence while still leveraging the known associated SNPs. The basis for the improvement of our statistic is the incorporation of external prevalence information, which is readily available. The liability threshold model models individuals as having an unobserved continuous phenotype called the liability (Dempster and Lerner, 1950; Falconer, 1967). Cases are individuals whose liability exceeds some threshold while all other individuals are controls. We compute the posterior mean of the residual of the liability given an individual's disease status, the disease prevalence and the known associated variants. This posterior mean is then treated as a continuous phenotype and tested for association via linear regression while easily incorporating covariates such as principal components (Price *et al.*, 2006) (see Supplementary Material). The crucial distinction between our approach and previous applications of liability threshold modeling (Duggirala *et al.*, 1997; Falconer, 1967; Jewell, 2004; Yang *et al.*, 2010) is that we incorporate ascertainment strategy and disease prevalence, which is the source of the power loss for logistic regression with covariates when estimating the parameters of the model. We show that accounting for ascertainment can also be done in a relative risk framework, but the liability threshold approach is more versatile.

In practice, our disease model changes dichotomous phenotypes to continuous ones. Cases are assigned positive-valued phenotypes and controls negative-valued phenotypes. Individuals carrying a smaller number of risk alleles are given a larger phenotype. The size of these shifts are a function of SNP effect size and disease prevalence, which is not accounted for in standard logistic regression. Our approach, unlike standard logistic regression, does not suffer any loss of power when the disease prevalence is low. This is not an issue with the logit model, which may also be adapted to account for ascertainment, but with the commonly used approach of adding genetic covariates to standard logistic regression in ascertained data, without accounting for disease prevalence (see Section 4).

Results on empirical data, including a large Type 2 diabetes (T2D) case–control study and the (WTCCC, 2007b) T1D, Rheumatoid Arthritis (RA), and T2D GWAS, demonstrate the pitfalls of using logistic regression with covariates as well as the power gains of LTSCORE when compared with both logistic regression with and without covariates. The gain in power is a function of prevalence and total variance explained by the known SNPs. Our method matches or outperforms conditioned or unconditioned linear or logistic regression for nearly all values of prevalence or ascertainment examined. Its performance relative to these methods will continue to increase as more variance in disease risk is explained by risk variants that are identified. We release a software package implementing LTSCORE for use in future association studies.

2 METHODS

Given a normally distributed continuous phenotype Y or a case–control phenotype Z, we want to test candidate SNP s_0 for association with the phenotype. There are K independent SNPs $s_1,...,s_K$ with genotypes $g_1,...,g_K$ and minor allele frequencies $p_1,...,p_K$ known to be associated with the phenotype and in complete linkage equilibrium (e.g. on different chromosomes) with SNP s_0. SNP s_0 has genotypes g_0 and minor allele frequency p_0. In this work, we explore three classes of statistical tests of association: NOCOND, STDCOND and LTSCORE. NOCOND-log is logistic regression of the genotypes g_0 against the phenotypes without conditioning on any known genetic covariates. STDCOND-log is logistic regression where the genotypes $g_1,...,g_K$ are included as covariates. LTSCORE is linear regression applied to the posterior mean of the residual of the liability threshold model described below. NOCOND-lin and STDCOND-lin refer to linear instead of logistic regression. Each test generates a χ^2 one df test statistic by performing a likelihood ratio test. Under the alternate hypothesis the effect size of s_0 is a free parameter and under the null hypothesis the effect size of s_0 is fixed at 0. The details of logistic and linear regression models are described in (Wasserman, 2005). For reasons of simplicity, the derivations below all use linear instead of logistic regression. Linear regression is commonly used in place of logistic regression in association studies (Armitage, 1955; Price *et al.*, 2006; Wallace *et al.*, 2006). Furthermore, we perform simulations and experiments under both linear and logistic regression frameworks to demonstrate that the theory described below holds under both models in practice (see Section 3). The extension of these tests to recessive and dominant models is straightforward. LTSCORE is publicly available in the LTSOFT software package.

2.1 Randomly ascertained case–control phenotypes

We begin with the case of cross-sectional dichotomous phenotypes (see Supplementary Material for Continuous phenotypes). We create a dichotomous phenotype Z under a liability threshold model (Falconer, 1967) by labeling all N individual cases when $Y \geq t$, for a threshold t, and controls otherwise. We consider the simple case of conditioning on one SNP i.e. $K=1$. In this case, the non-centrality parameter of NOCOND-lin is

$$N*\mathrm{corr}(g_0,Z)^2. \tag{1}$$

The non-centrality parameter of STDCOND-lin is

$$N*\frac{\mathrm{corr}(g_0,Z)^2}{(1-\mathrm{corr}(g_1,Z)^2)}, \tag{2}$$

where $\mathrm{corr}(g_0,Z)$ is the correlation between the genotypes g_0 and the phenotypes Z. The full details of the derivation are given in Supplementary Material S1. The non-centrality parameter increases in proportion to the inverse of (1−fraction of variance explained by s_1). That is, as the known variants explain more of the phenotype, the greater our power to discover new variants by conditioning in randomly ascertained study designs (Robinson and Jewell, 1991). The non-centrality parameter for STDCOND-lin is also larger than that of NOCOND-lin in the case of randomly ascertained continuous phenotypes (see Supplementary Material).

2.2 Non-randomly ascertained case–control phenotypes

A key assumption in the derivations above is that candidates SNP s_0 and SNP s_1 are *independent*. However, in an ascertained case–control study, especially one for a disease of low-prevalence this assumption no longer holds. That is, candidate, SNP s_0 and SNP s_1 that are independent in the population will become correlated in the study cohort. Consider the extreme example of drawing cases from one tail of Y and controls from the other. Under both a logit and liability threshold model of disease, controls will have relatively fewer copies of g_0 and g_1, while the cases will have relatively more, and so in the study, g_0 and g_1 will be correlated. This exaggerated example provides the intuition for why we see correlation in the ascertained study as we are drawing all of our cases from an extreme tail of an underlying distribution. As shown in Section 3 below, this correlation and the corresponding effects of conditioning exist under both the liability threshold model of disease and the logit model assumed by logistic regression.

The covariance between g_0,g_1 in the case of haploid individuals (this is easily extended to the diploid case) is

$$\mathrm{cov}(g_0,g_1)=E[g_0*g_1]-p_0*p_1. \tag{3}$$

The expectation of the product of g_0*g_1 is

$$\begin{aligned} &P(g_0=1,g_1=1|Z=1)+P(g_0=1,g_1=1|Z=0) \\ &\quad=P(Z=1|g_0=1,g_1=1)p_0p_1F_S/F \\ &\quad+P(Z=0|g_0=1,g_1=1)p_0p_1(1-F_S)/(1-F), \end{aligned} \tag{4}$$

where F_S is the frequency of cases in the study and F is the frequency of cases in the population. In the case of random ascertainment, F_S and F are the same and so the covariance will be 0. However, in a disease of low prevalence, F_S and F will be different and so s_0,s_1 will be correlated in the study due to ascertainment-induced correlation.

When we test SNP s_0 marginally (NOCOND-lin), the non centrality parameter is

$$N*\frac{\alpha_0^2+2\alpha_0\alpha_1\mathrm{cov}(g_0,g_1)}{\mathrm{var}(Z)} \tag{5}$$

When we test s_0 conditioned on s_1 (STDCOND-lin), the non-centrality parameter is

$$N*\frac{\alpha_0^2}{\mathrm{var}(Z)-\alpha_1^2}, \tag{6}$$

where α_0,α_1 are the expected SNP effect sizes in the study (as opposed to β_0,β_1 the effect sizes in the population). The full details of the derivation are given in Supplementary Material S1. In the marginal case (NOCOND-lin), the shared signal of s_0 and s_1 is added to the non-centrality parameter. This implies that the power to detect s_0 in the marginal case is greater if there exists another SNP s_1 that explains a significant fraction of the variance. In the conditioned case (STDCOND-lin), the numerator is decreased because the shared signal of s_0 and s_1 is conditioned out. However, the denominator is also smaller since the variance of Z conditioned on s_1 is smaller than the unconditioned variance of Z. The power of STDCOND-lin relative to NOCOND-lin is therefore a function of effect size, prevalence and ascertainment. Yang and colleagues (Yang *et al.*, 2010) provide alternative derivations of the non-centrality parameter in the unconditioned case for both quantitative and case–control phenotypes based on the liability threshold model. In the case of non-randomly ascertained quantitative phenotypes, two associated variants s_0 and s_1 that are independent in the population will be correlated in the study for the reasons given above. We do not consider this case in detail in this work but note that in many cases, STDCOND-lin will reduce power significantly and we therefore caution against this statistic for non-randomly ascertained quantitative phenotypes.

2.3 LTSCORE statistic

We model a case–control phenotype as arising from an underlying normally distributed phenotype

$$\phi=-m+\epsilon;\epsilon\sim\mathcal{N}(0,1) \tag{7}$$

called the liability (Falconer, 1967). Cases are those individuals with $\phi\geq 0$ and controls are those individuals with $\phi<0$. There is a relationship between this liability scale and the relative risk model of disease described in detail previously in Wray *et al.* (2010) and Yang *et al.* (2010). If F is the prevalence of the disease in the population then $m=\Phi^{-1}(1-F)$, where $\Phi^{-1}(x)$ is the inverse of the cumulative normal distribution function with mean 0 and variance 1 evaluated at x, so that the expected proportion of individuals with $\phi\geq 0$ is F. A SNP s_1 associated with the disease and having mean adjusted genotypes $g_1\in\{0-2p,1-2p,2-2p\}$ is incorporated into the model as $\phi=-m+\beta_1 g_1+\epsilon$, where $\epsilon\sim\mathcal{N}(0,\sqrt{1-\mathrm{var}(\beta_1*g_1)})$ so that the total variance of ϕ is 1. Given a case–control study where SNP s_1 has frequency p_1^+ in the cases and frequency p_1^- in the controls, we estimate β_1 via a method (described below) that relies on published prevalence data for the disease. This prevalence represents a source of external data not available to STDCOND-lin.

The estimation procedure is repeated for independent known associated SNPs $s_2,\ldots,s_K$ giving a final model

$$\phi=-m+X\beta+\epsilon, \tag{8}$$

$$\epsilon\sim\mathcal{N}(0,\sigma_e=\sqrt{1-\mathrm{var}(X\beta)}), \tag{9}$$

where X are the genotypes of the K known SNP, and β is a vector of the effects size $\beta_1,\ldots\beta_K$.

To use both the prevalence information and the effects of the known associated variants $s_1,\ldots,s_K$ when testing a new candidate SNP s_0, we compute the posterior mean of the residual of the liability given the genotypes of the known variants X, their effect sizes β, the disease prevalence F and the case–control status Z $E(\epsilon|X,\beta,F,Z)$:

$$E(\epsilon|X,\beta,F,Z=\mathrm{Case})=\frac{\int_{m-X\beta}^{\infty}\epsilon\frac{1}{\sqrt{2\pi\sigma_e^2}}e^{(\frac{-\epsilon^2}{2\sigma_e^2})}d\epsilon}{\int_{m-X\beta}^{\infty}\frac{1}{\sqrt{2\pi\sigma_e^2}}e^{(\frac{-\epsilon^2}{2\sigma_e^2})}d\epsilon}, \tag{10}$$

$$E(\epsilon|X,\beta,F,Z=\mathrm{Control})=\frac{\int_{-\infty}^{m-X\beta}\epsilon\frac{1}{\sqrt{2\pi\sigma_e^2}}e^{(\frac{-\epsilon^2}{2\sigma_e^2})}d\epsilon}{\int_{-\infty}^{m-X\beta}\frac{1}{\sqrt{2\pi\sigma_e^2}}e^{(\frac{-\epsilon^2}{2\sigma_e^2})}d\epsilon}, \tag{11}$$

where σ_e^2 is $1-\mathrm{var}(X\beta)$ the residual variance of ϕ after subtracting the variance from the known SNPs. The prevalence-aware liability threshold based statistic is then computed by running standard linear regression between the genotypes of the new SNP s_0 and the posterior mean of the residual of the liability of each individual as calculated above. Although the posterior mean is not normally distributed, the use of linear regression in place of logistic regression is common practice in association studies (Armitage, 1955; Price *et al.*, 2006; Wallace *et al.*, 2006).

Intuitively, the above integrals have the following effect. Cases without risk alleles at other loci are assigned more extreme phenotypes than cases with risk alleles at other loci (and analogously for controls). Consider a case with no risk alleles at any of the known associated variants. To exceed the liability threshold, such an individual will require a large value ϵ relative to a case with many risk alleles at the known associated variants. Another implication of this model (as well as the relative risk model) is that the odds ratio at s_0 will be higher when computed with cases having no known risk alleles (Guey *et al.*, 2011).

For fixed effect sizes β, as the prevalence of the disease approaches 0, the computation of $E[\epsilon|X,\beta,F,Z]$ is dominated by the threshold m. All of the case individuals will have approximately the same value of $E[\epsilon|X,\beta,F,Z=1]$ (E_{case}), and all of the controls will have approximately the same value of $E[\epsilon|X,\beta,F,Z=0]$ (E_{control}). Since the LTSCORE statistic is linear regression applied to $E[\epsilon|X,\beta,F,Z]$, it is equivalent to the marginal test NOCOND-lin in this case of near 0 prevalence.

The liability threshold model is not the only model of disease and we also derive a prevalence aware statistic from the relative risk model of disease (RRCOND)(Jewell, 2004). The RRCOND model is presented in Supplementary Material S1, but we primarily focus on the LTSCORE because the relative risk model does not easily handle non-SNP covariates such as principal components.

2.4 Estimating β using published prevalence

We require an estimate of the disease prevalence F taken from the literature. In the liability threshold model, any estimates $\hat{\beta_1}$ of β_1 and $\hat{p_1}$of p_1 give an expected frequency of s_1 in the cases and controls. Our estimate of the population minor allele frequency is

$$\hat{p_1}=p_1^+F+p_1^-(1-F), \tag{12}$$

where p_1^+ and p_1^- are the observed frequencies of s_1 in the cases and controls.

Given an estimated effect size $\hat{\beta_1}$ of s_1

$$P(Z=1|g_1=0)=(1-\Phi(m,\hat{\beta_1}(-2\hat{p_1}),\sigma_e^2)) \tag{13}$$

$$P(Z=1|g_1=1)=(1-\Phi(m,\hat{\beta_1}(1-2\hat{p_1}),\sigma_e^2)) \tag{14}$$

$$P(Z=1|g_1=2)=(1-\Phi(m,\hat{\beta_1}(2-2\hat{p_1}),\sigma_e^2)), \tag{15}$$

where $\Phi(x,y,z)$ is the cumulative normal distribution evaluated at x, with mean y and variance z. Then

$$P(g_1=0|Z=1)=\frac{P(Z=1|g_1=0)(1-\hat{p_1})^2}{F} \tag{16}$$

and similarly for $g_1=1,2$. Finally, we compute the frequency of s_1 in the cases given $\hat{\beta_1}$ and $\hat{p_1}$ as

$$\hat{p_1}^+=P(g_1=1|Z=1)+2P(g_1=2|Z=1) \tag{17}$$

and similarly for controls. Using these frequencies, we can compute the squared error between the observed and expected frequencies in the cases and controls $S_e=(p_1^+-\hat{p_1}^+)^2+(p_1^--\hat{p_1}^-)^2$. We perform a binary search for 10 iterations to identify the $\hat{\beta_1}$ that minimizes S_e. For multiple known SNPs, the $\hat{\beta_i}$ are estimated independently and combined, and only one associated SNP from any locus can be used.

3 RESULTS

The theory presented in Section 2 above modeled case–control phenotypes under a liability threshold model and estimated the power of linear regression with no covariates (NOCOND-lin), linear regression conditioned on known variants (STDCOND-lin) and our liability threshold model-based LTSCORE, under various ascertainment scenarios. Here, we examine the relative benefits of the three classes of statistical tests NOCOND, STDCOND and LTSCORE over simulated and real data. For NOCOND and STDCOND, we conduct most of our analyses using the logistic regression versions NOCOND-log and STDCOND-log, but we have verified that NOCOND-lin and STDCOND-lin produce very similar results (see below). There are many equivalencies between the logit model, the liability threshold model and the multiplicative relative risk model (So *et al.*, 2011; Wray *et al.*, 2010). To be maximally conservative and to demonstrate that the results derived in Methods section hold for different disease models, we simulate our case–control phenotypes under a logit model. This prevents our method from having an unfair advantage due to testing the same model that generated the data. As shown below similar results were obtained when using linear instead of logistic regression and the liability threshold model instead of the logit model.

LTSCORE computes posterior mean of the residual of the liability, using liability threshold model parameters that account for disease prevalence and study design, and then uses posterior mean as input to linear regression (see Section 2). The LTSCORE parameters are estimated from published disease prevalence data. This external information, unavailable to either NOCOND-log or STDCOND-log, is the basis of the improvement of LTSCORE. We are interested in the effects of known associated SNPs on association tests for undiscovered SNPs that are in complete linkage equilibrium (e.g. those on completely different chromosomes) with the known associated SNPs in the population. In both the simulated and real datasets below, we never condition on SNPs that are in LD with the candidate SNP. The derivations above assumed a liability threshold model of disease. However, both the STDCOND-log and NOCOND-log tests assume a logit model of disease as they are applications of logistic regression. We compare the the performance of the methods by measuring the ratio of the average χ^2 test-statistics produced by each method. This has a natural interpretation of the increase in sample size needed to obtain the equivalent power (Pritchard and Przeworski, 2001). For example, if LTSCORE gives 10% increase in test-statistic over STDCOND-log, this corresponds to adding 10% more individuals to a study analyzed with STDCOND-log to achieve the power of the original study analyzed by LTSCORE.

3.1 Simulated datasets

3.1.1 Randomly ascertained case–control phenotypes To examine the effect of conditioning in randomly ascertained (cross-sectional) case–control phenotypes, we generated case–control data from a logit model $P(\text{Disease})=\frac{e^{g_0\alpha+X\beta+z}}{1+e^{g_0\alpha+X\beta+z}}$.

The affine term z determines the prevalence F of the disease in the population. To test the effects of conditioning we tested candidate SNP s_0 under NOCOND-log, STDCOND-log and LTSCORE. We ran 5000 simulations of 1000 cases and 1000 controls. In each simulation, there was one candidate SNP with effect size α and one known variant with effect size β. The fraction of variance explained with K SNPs of effect size $\beta/\sqrt(K)$ is the same as the fraction of variance explained by one SNP with effect size β. LTSCORE with K SNPs of effect size $\beta/\sqrt(K)$ produced equivalent results to using LTSCORE with one SNP of effect size β (see Supplementary Material) and so we chose to use one SNP for simplicity. The genotypes were generated as random draws from a binomial distribution for each simulation. We examined a range of known variant effect sizes β, a fixed candidate SNP effect size $\alpha = 0.35$, minor allele frequencies $p_0 = p_1 = 0.2$ and z (the affine term in the logit model) corresponding to a prevalence of $F = 50\%$. The results are presented in Figure 1a. STDCOND-log always improves on NOCOND-log and the improvement is a function of the total variance explained by the known variants. LTSCORE assumes the data were generated with a liability threshold model. Despite generating data under a logit model, LTSCORE and STDCOND-log perform similarly. Reducing the prevalence F (the fraction of case individuals in the population) decreases the number of cases and increases the number of controls, but both LTSCORE and STDCOND-log still outperform NOCOND-log. Results for the liability threshold-based simulation are presented in Supplementary Figure S2 (a) and are similar to those presented in Figure 1(a) Standard conditioning also improves power for randomly ascertained continuous phenotypes in simulations (see Supplementary Material and Fig. S1).

3.1.2 Non-randomly ascertained case–control phenotypes We have seen that STDCOND-log improves the power to detect new variants at independent loci relative to NOCOND-log. Surprisingly, in a balanced case–control study, this is not always the case and STDCOND-log often significantly decreases the power to detect new loci. The reason for this reduction in power is the non-random ascertainment of the samples which induces a correlation between all the causal variants. The strength of the correlation between associated variants is a function of disease prevalence. The STDCOND-log test on any set of associated variants will not only remove their signal but also some of the signal from the SNP being tested. We simulated a low-prevalence case–control phenotype under a logit model as in the randomly ascertained experiments described above with $F = 0.1\%$, $\alpha = 2.0$ and $\beta = 2.0$. We then sampled 1000 cases and 1000 controls and measured the correlation between candidate SNP s_0 and SNP s_1. The average correlation in 5000 simulations was $r^2 = 0.11$ ($\chi^2 = 220$ via Armitage trend test). We used an extreme β to demonstrate the effect with a small number of simulations.

To examine the relative behaviors of the three classes of tests in case–control data, we simulated a case–control phenotype under a logit model as in the randomly ascertained experiments described above with $F = 4.0\%$, $\alpha = 0.35$ and minor allele frequency MAF= 0.20 for both SNPs. We then sampled 1000 cases and 1000 controls. The results are presented in Figure 1(b). When β is small, LTSCORE is nearly identical to NOCOND-log losing 0.5% in the worst case. The improvement of LTSCORE relative to NOCOND-log increases as the known variant β explains more the population phenotypic variance. STDCOND-log decreases in performance relative to NOCOND-log until the known variant explains at least 35% of the population phenotypic variance, at which point STDCOND-log starts to improve. However, even after the known variant explains 50% of the in study phenotypic variation, STDCOND-log achieves only 96.8% of the NOCOND-log statistic. Note that the results presented in Figure 1(b) refer to the fraction of study variance not population variance explained. Because of the ascertainment strategy, there is a significant difference between the effect sizes of the SNPs in the study and their effect size in the population. In the population, only 4.0% of individuals are cases, while in the study, 50% of individuals are cases. This skew causes the variance explained by a SNP in the population to be much smaller than the variance explained in the study (Guey *et al.*, 2011; Yang *et al.*, 2011). Results for the liability threshold based simulation are similar and presented in Supplementary Figure S2b. We repeated the experiments for Figure1a and b replacing logistic regression with linear regression and found nearly identical results, Supplementary Figure S3a and b. We conclude that replacing linear with logistic regression makes little difference in this context and use only logistic regression for the remaining experiments (McCarthy *et al.*, 2008).

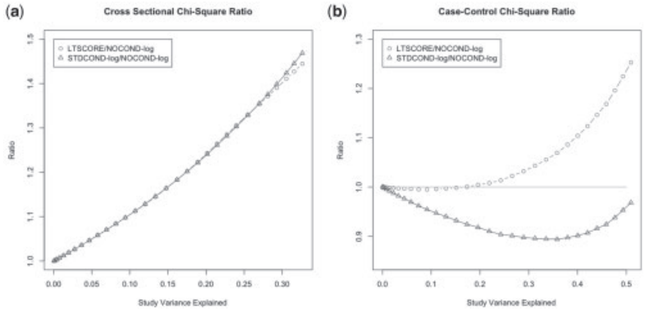

Fig. 1. NOCOND-log, STDCOND-log, and LTSCORE simulations on case–control phenotypes. Results of NOCOND-log (logistic regression), STDCOND-log (logistic regression with covariates) and LTSCORE tests for simulated case–control datasets from a logit model. Study variance explained is the proportion of phenotypic variance explained in the study by the known association variant. For randomly ascertained data (**a**) both the LTSCORE and STDCOND-log tests improve over the NOCOND-log tests and have similar performance. However, for non-randomly ascertained case–control data (**b**) with prevalence of 4.0% the STDCOND-log test performs significantly worse than the NOCOND-log test

To examine the effects of prevalence on the three tests, we fixed $\beta = 1.5$, $\alpha = 0.35$ MAF = 0.2 and varied the disease prevalence F under the same model as above. The results presented in Figure 2 show that the LTSCORE always outperforms STDCOND-log. STDCOND-log reduces power compared with the NOCOND-log test when the prevalence is low. However, as the prevalence increases, the study becomes more like a randomly ascertained study and the STDCOND-log test performance increases above the NOCOND-log test. LTSCORE is slightly (<2%) worse than NOCOND-log for very low-prevalence (0.1%) disease and improves as the prevalence increases. This modest loss in power is removed when the data are generated under a liability threshold model (see Supplementary Fig. S4). In this case LTSCORE always outperforms or matches NOCOND-log and STDCOND-log. It is unknown which model better represents the truth about disease.

We tested the sensitivity of our model to the misspecification of the prevalence F by generating data under the same model as above for a disease with true prevalence of 3%. We tested under our LTSCORE model for a range of 'estimated' prevalences. We repeated the simulation 5000 times, with 1000 cases and 1000 controls. The results are presented in Supplementary Figure S6. Changing the estimated prevalence between 1% and 5% had a minimal effect and the performance in this case was greater than either the NOCOND-log or STDCOND-log tests. The power was greater than NOCOND-log until the specified prevalence was greater than twice the true prevalence. The maximum power was not attained at the true prevalence and we believe this is because the disease model tested (liability threshold) is different than the disease model

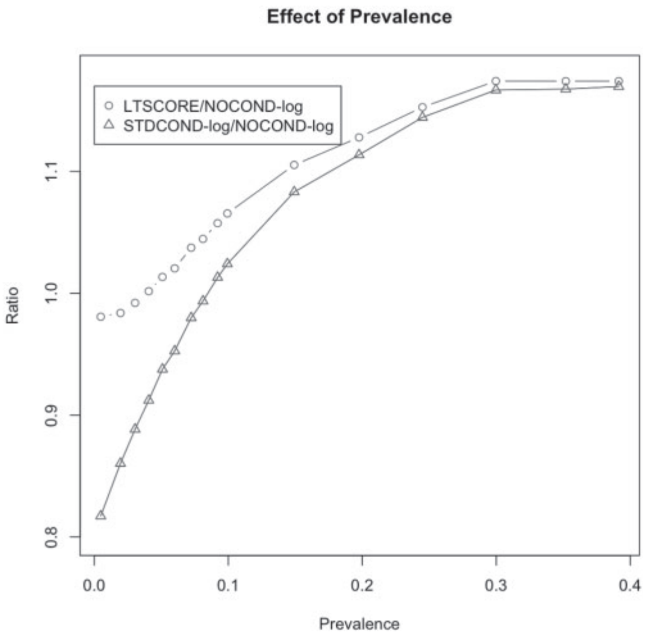

Fig. 2. Effects of prevalence on NOCOND-log, STDCOND-log and LTSCORE Results of STDCOND-log versus NOCOND-log and LTSCORE vs. NOCOND-log for simulated ascertained case–control phenotypes from a logit model, as a function of disease prevalence. Under low-disease prevalence, there is an induced correlation between associated variants causing a sever loss of power for the STDCOND-log test relative to the NOCOND-log test. As the prevalence increases, the design is more like a randomly ascertained study and the STDCOND-log test outperforms the NOCOND-log test. The LTSCORE always outperforms STDCOND-log. For low-prevalence disease, LTSCORE is slightly worse than NOCOND-log and improves as the prevalence increases

used to generate the data (logit). Results for the liability threshold-based simulation are presented in Supplementary Figure S5 and in this case, the maximum is attained at the true prevalence.

To examine the behavior of the tests under the null, we repeated the experiments for a range of prevalences $F = 0.01, 0.03, 0.05, 0.1$ and setting the effect size $\alpha = 0$ and keeping $\beta = 1.5$. For each prevalence, we generated 1000 cases and 1000 controls 1000000 times. All three tests were well behaved maintaining a false positive rate of 0.050 as desired.

To inform researchers about the potential gains available in their case–control datasets, we include the average χ^2 test statistics for all three tests for a range of realistic disease parameters in Supplementary Table S8.

3.2 Real data sets

3.2.1 Non-randomly ascertained datasets for low-prevalence disease (T1D, RA) We begin with an analysis of low-prevalence case–control phenotypes (see Supplementary Material for real continuous phenotypes). We examined the performance of the NOCOND-log, STDCOND-log and LTSCORE statistics on the WTCCC T1D and RA datasets (WTCCC, 2007b). There were 1924 and 1860 cases for RA and T1D respectively, and the same set of 2938 controls for the two datasets. For T1D, we used a prevalence of 0.125% (Cooper and Stroehla, 2003) and HLA SNP rs9273363 from Chromosome 6 as the known variant which explained 12.4% phenotypic variation (Nejentsev *et al.*, 2007) in the study. For RA, we used a prevalence of 1% (Cooper and Stroehla, 2003) and HLA SNP rs6457620 from Chromosome 6 as the known variants which explained 7.1% phenotypic variation in the study. We filtered out all SNPs with MAF <5% and applied the NOCOND-log, STDCOND-log and LTSCORE, tests to all SNPs not found on Chromosome 6.

Although the WTCCC studies identified a relatively small number of risk loci due to limited sample size, for T1D and RA this includes HLA, a locus of large effect. The prevalences of T1D and RA are low so the expected improvement of LTSCORE relative to STDCOND-log is not expected to be large (see Section 3.1). However, these datasets demonstrate the potential for a severe loss in power of using STDCOND-log and that LTSCORE is well behaved for low-prevalence diseases. Indeed, for T1D, there was a greater than 27% drop in test statistic using STDCOND-log relative to NOCOND-log and a 4% increase using LTSCORE relative to NOCOND-log as measured by the average change in test statistic at all published GWAS variants according to the NHGRI (Hindorff *et al.*, 2009) (see Supplementary Tables S1–S8).

The Q–Q plots of NOCOND-log, STDCOND-log and LTSCORE are shown in Figure 3a and b and serve as one means of assessing the relative performance of the methods. The significant SNPs lie at the tail of the distribution and methods with larger values at the tail are better powered. All of the test statistics had a similar λ_{GC} and all were genomic control (GC) corrected before analysis (Devlin and Roeder, 1999). On the RA dataset for example the λ_{GC} values were 1.046, 1.047 and 1.041, for the NOCOND-log, STDCOND-log and LTSCORE tests, respectively. It is clear that STDCOND-log reduces the χ^2 test statistic relative to NOCOND-log and LTSCORE in T1D (Fig. 3a) and RA (Fig. 3b). The reduction in T1D is the most dramatic because it has a very low-prevalence and the SNPs explain a larger fraction of the variance.

As another means of assessing the relative performance of the methods, we look at the test statistics of known associated variants published in the NHGRI catalog (Hindorff *et al.*, 2009). When the known associated variant was missing from the dataset, we used the best tag as measured by r^2, removing any SNP where the best tag had $r^2 < 0.5$. The results are presented in Supplementary Tables S1,S2 and are analogous to the Q–Q plot results. STDCOND-log performs poorly for T1D and RA with a reduction in the sum of test statistics of roughly 27% in T1D equivalent to removing 27% of the individuals from the study (Pritchard and Przeworski, 2001). On the other hand, LTSCORE has slightly larger sum χ^2 test statistics relative to NOCOND-log. We simulated 1000 case–control studies with effect sizes, prevalences and sample sizes matching the WTCCC studies. We generated the data under a liability threshold model and found expected gains for both studies close to 2% relative to NOCOND-log.

3.2.2 Non-randomly ascertained datasets for high-prevalence disease (T2D) We examined the performance of the NOCOND-log, STDCOND-log and LTSCORE statistics over of 6142 cases and 7403 controls genotyped at 19 known associated SNPs from the Multiethnic Cohort (MEC) (African Americans, Latinos, Japanese Americans, Native Hawaiians, and European Americans) (Waters *et al.*, 2010) and used a prevalence of 9% (Scott *et al.*, 2007). Unfortunately, the known associated variants together explain only

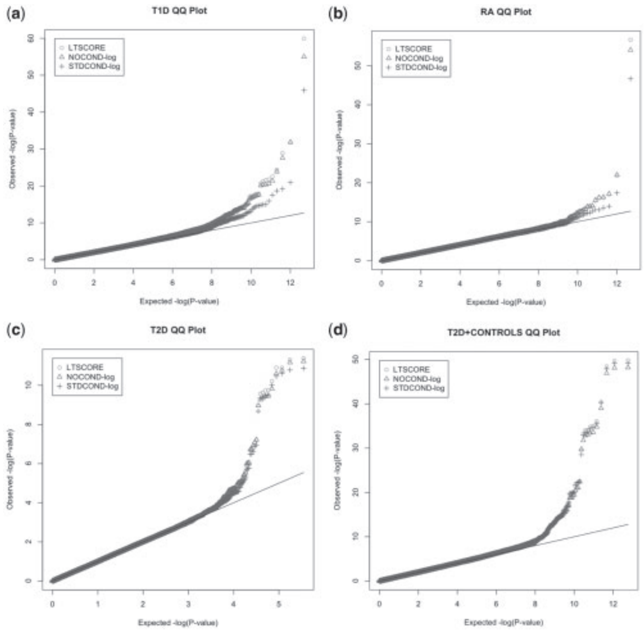

Fig. 3. Q–Q plots of NOCOND-log, STDCOND-log and LTSCORE on WTCCC datasets Q–Q plots for the NOCOND-log, STDCOND-log and LTSCORE tests applied to the WTCCC T1D, RA, T2D, and T2D+ datasets. The tail of the plots serves as an empirical measure of improvement. In T1D (**a**), the LTSCORE outperforms the NOCOND-log test and the STDCOND-log test suffers significant power loss. In RA (**b**) the LTSCORE matches the performance of the NOCOND-log test and again the STDCOND-log test suffers significant power loss. In T2D (**c**) and T2D+CONTROLS (**d**) LTSCORE and NOCOND-log perform similarly. STDCOND-log improves significantly with the addition of controls, which mimics a randomly ascertained design

4% phenotypic variation in the study. We simulated 1000 datasets with the same sample size, a disease prevalence of 9%, and a known associated variant that accounted for 4% of the phenotypic variation. For an SNP with a minor allele frequency of 20% and an effect size on the liability scale of 0.05 (corresponding to 1.6% of the variance on the liability scale), the average improvement of LTSCORE was 3% with a standard error of 10% in the simulations. Using many SNPs of small effect produced, the same result as one SNP of large effect. The results on the MEC data are shown in Supplementary Table S3 with LTSCORE slightly outperforming NOCOND-log, but not significantly different from STDCOND-log. The variance of the expected improvement is large and this improvement is within the expected range. As expected, the relative performance of STDCOND-log in this high-prevalence disease is much better than it was in T1D and RA.

We examined the relative performance of NOCOND-log, LTSCORE and STDCOND-log in the WTCCC T2D study with 1924 cases. We used the 2938 controls in the original study and we created a large control set (+CONTROLS) for T2D containing individuals in all other diseases with a sample size of 14255. We note that the use of cases from other diseases as shared controls is commonplace in WTCCC and other studies (WTCCC, 2007a, b). The known variants explained 1.42% and 0.64% in the original study and T2D+CONTROLS respectively. The results are shown in Supplementary Tables S5 and S6. The expected improvement is even smaller than in the MEC study above as a smaller fraction of the variance is explained and LTSCORE performed slightly worse than STDCOND-log but within the range predicted by simulations (1 ± 6%). The performance of STDCOND-log is affected by the addition of controls as this simulates the properties of random ascertainment where STDCOND-log is expected to perform better. In the original study NOCOND-log had an 8% higher sum of test statistics than STDCOND-log, while in the T2D+CONTROLS study, this was reduced to 2%.

The Q–Q plots of NOCOND-log, STDCOND-log and LTSCORE are shown in Figure 3c and d. STDCOND-log reduces the χ^2 test statistic relative to NOCOND-log and LTSCORE in T2D 3(c). In the case of T2D+CONTROLS, the large number of controls create a study that is more similar to random ascertainment. As expected, STDCOND-log improves over NOCOND-log in this case as shown in Figure 3d. The LTSCORE method performs well in all instances, matching or outperforming each of the other tests.

4 DISCUSSION

We have shown that the practice of standard conditioning on known associated variants does not account for study design and disease prevalence potentially leading to significant power loss. This power loss is due to the induced correlation between associated variants in case–control studies. The phenomenon of higher odds ratios in cases with fewer risk alleles at other loci than in cases with more risk alleles at other loci can be viewed as a gene–gene interaction (Cordell, 2009). This is a statistical, rather than biological, interaction. By properly modeling the ascertainment and prevalence while still leveraging known associated variants, our LTSCORE statistic improves study power relative to NOCOND-log and STDCOND-log tests in case–control studies of mid-to-low prevalence diseases. This increase in power is a function of the total phenotypic variance explained by known variants and disease prevalence. The datasets examined here had either a low-prevalence or a small fraction of the variance explained and therefore we did not expect a large improvement. However, as more associated variants are discovered, the performance of LTSCORE will increase giving rise to power gains as a function of covariate effect size and disease prevalence. This approach can also be applied to clinical covariates, and in this case, an average power gain of >17% was achieved (Zaitlen *et al.*, unpublished data). We have verified that results similar to Supplementary Table S3 are obtained when comparing genetic + clinical covariates to clinical covariates only (see Supplementary Table S4). However, conditioning on clinical covariates is a fundamentally different problem, both because a different parameter estimation method is needed and because with clinical covariates, it is often the case that samples are non-randomly ascertained for covariate value as well as case–control status.

A recent T2D meta-analysis (Voight *et al.*, 2010) uses the standard conditioning statistic and shows a significant gain in power. Their ratio of cases to controls is closer to a randomly ascertained study and in this case we expect STDCOND-log to outperform NOCOND-log and increase power. In addition to their beneficial study design, some of the conditioned variants are proximal to the new discoveries. Both of the elements serve to improve the power of standard conditioning. (Yang *et al.*, 2012) also examine the potential benefits of genome-wide conditioning in T2D. However, we believe the use of our LTSCORE statistic on these data could improve the power further by accounting for prevalence and ascertainment. In a recent meta-analysis of GWAS height data (Lango Allen *et al.*, 2010), a randomly ascertained continuous phenotype, standard conditioning revealed no new associated variants. This is due to the nature of their study design and not a contradiction of our results (see Supplementary Material S1). In their landmark T1D paper, Barrett *et al.* (2009) find a correlation between disease risk computed from HLA SNPs and disease risk computed from SNPs in the rest of the genome. They suggest that this is due to a departure from a multiplicative model of disease. However, this effect may also be explained from the non-independence of the genotypes that we described in case–control studies. That is, some or all of the effect that was described (correlation between MHC major histocompatibility complex risk score and non-MHC risk score) may be due to ascertainment-induced correlation. We caution that in tests for epistatic interaction (Moore and Williams, 2009), this induced correlation could give rise to a spurious signal of epistatic interaction at true (marginally) associated variants.

Adjustment for informative covariates is not unique to genetics and the problem of estimation from case–control data has received considerable attention in the epidemiological literature. It is well known that regressing or stratifying on a covariate which is related to disease but not exposure of interest causes a reduction in power unless one matches on the covariate when sampling controls (Hosmer and Lemeshow, 2000; Jewell, 2004; Moolgavkar *et al.*, 1985; Nam, 1992; Neuhaus, 1998). We derive this power loss in terms of the liability threshold model. (Neuhaus, 1998) shows the reduction in power under a logit model for any correlated covariate (i.e. not just due to ascertainment). Although we focus on adapting the liability threshold model to incorporate prevalence information, it may be possible to achieve the same result in a logistic framework. For example, if there is only one known variant, one could construct a $2\times2\times2$ table of case–control status, candidate SNP s_0 and known covariate s_1. Much larger tables would be required as the number of known variants increased.

We recently proposed (Monsees *et al.*, 2009) a weighted logistic regression method (IPW) in the case of conditioning on environmental variables in case–control studies. Rose and van der Laan (2008) also offer an efficient estimator for case–control studies to account for ascertainment-induced biases. However, the focus of these works is obtaining an unbiased estimate of effect size while our concern is power (and a valid test under the null). In the case of genetic association studies, the effect sizes are generally small and the emphasis of the community is on discovery as opposed to effect size estimation. In the case of IPW, unbiased effect sizes are indeed obtained, but it under-performed relative to STDCOND-log, NOCOND-log and LTSCORE in simulations so is not considered. If the objective is to obtain unbiased effect sizes, IPW is recommended over LTSCORE. Note that the basis for the improvement of LTSCORE is the published prevalence data and not published SNP effect sizes. It is not equivalent to using STDCOND-log with an offset, which will perform similarly to STDCOND-log in the presence of ascertainment. Including an explicit interaction term in the logistic model introduces an extra df reducing the overall power.

Although this paper focuses exclusively on the use of conditioning to discover new loci that are completely unlinked to the known variants, conditioning is also a widely used tool for SNPs in the same locus. In this case, the purpose is to perform fine-mapping and better understand the genetic architecture of the known associated locus (Chang *et al.*, 2009). Therefore, any drop in power due to induced correlation should not prevent researchers from using conditioning in this same-locus context. LTSCORE may improve fine-mapping efforts in some situations (see Supplementary Material). A discussion of usage and meta-analysis is given in the Supplementary Material.

ACKNOWLEDGEMENTS

The authors are grateful to D. Stram for helpful discussions.

Funding: NIH fellowship [5T32ES007142-27 N.Z.]; NHGRI [R01 HG006399 and N.Z., B.P., N.P., A.L.P.]; US National Institutes of Health; National Cancer Institute [cooperative agreements U01-CA98233-07 to David J. Hunter, U01-CA98710-06 to Susan Gapstur, U01-CA98216-06 to Elio Riboli and Rudolf Kaaks, and U01-CA98758-07 to Brian E. Henderson, and Intramural Research Program of NIH/National Cancer Institute, Division of Cancer Epidemiology and Genetics]. The T2D study in the MEC was supported by NIH CA63464, CA54281, 1U01HG004802 and Multiethnic Cohort (MEC) [R37-CA54281 to Laurence N. Kolonel].

Conflict of Interest: none declared.

REFERENCES

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

Barrett,J.C. *et al.* (2009) Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.*, **41**, 703–707.

Chang,B.L. *et al.* (2009) Fine mapping association study and functional analysis implicate a snp in msmb at 10q11 as a causal variant for prostate cancer risk. *Hum. Mol. Genet.*, **18**, 1368–1375.

Cooper,G.S. and Stroehla,B.C. (2003) The epidemiology of autoimmune diseases. *Autoimmun. Rev.*, **2**, 119–125.

Cordell,H.J. (2009) Detecting gene-gene interactions that underlie human diseases. *Nat. Rev. Genet.*, **10**, 392–404.

Dempster,E. and Lerner,M. (1950) Heritability of threshold characters. *Genetics*, **35**, 212, 236.

Devlin,B. and Roeder,K. (1999) Genomic control for association studies. *Biometrics.*, **55**.

Duggirala,R. *et al.* (1997) A variance component approach to dichotomous trait linkage analysis using a threshold model. *Genet. Epidemiol.*, **14**, 987–992.

Falconer,D.S. (1967) The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus. *Ann. Hum. Genet.*, **31**, 1–20.

Guey,L.T. *et al.* (2011) Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. *Genetic Epidemiology*, **In Revision**.

Hindorff,L. *et al.* (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci.*, **106**, 9362–9367.

Hosmer,D. and Lemeshow,S. (2000) *Applied Logistic Regression*. Wiley Series in Probability and Statistics. John Wiley and Sons Inc, New York, USA.

Jewell,N.P. (2004) *Statistics for epidemiology*. Texts in statistical science series. Chapman & Hall CRC, Boca Raton.

Kuo,C.L. and Feingold,E. (2010) What's the best statistic for a simple test of genetic association in a case–control study? *Genet. Epidemiol*, **34**, 246–253.

Lango Allen,H. *et al.* (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, **467**, 832–838.

Ma,L. *et al.* (2010) Multi-locus test conditional on confirmed effects leads to increased power in genome-wide association studies. *PLoS One*, **5**, e15006.

McCarthy,M.I. *et al.* (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.*, **9**, 356–369.

Monsees,G.M. *et al.* (2009) Genome-wide association scans for secondary traits using case–control samples. *Genet. Epidemiol*, **33**, 717–728.

Moolgavkar,S. *et al.* (1985) Assessing the adequacy of the logistic regression model for matched case–control studies. *Stat. Med.*, **4**.

Moore,J. and Williams,S. (2009) Epistasis and its implications for personal genetics. *Am. J. Hum. Genet.*, **85**, 309–320.

Nam,J. (1992) Improved power offered by a score test for linkage disequilibrium mapping of quantitative-trait loci by selective genotyping. *Biometrics.*, **48**.

Nejentsev,S. *et al.* (2007) Localization of type 1 diabetes susceptibility to the mhc class i genes hla-b and hla-a. *Nature*, **450**, 887–892.

Neuhaus,J. (1998) Estimation efficiency with omitted covariates in generalized linear models. *J. Am. Stat. Assoc.*, **4**.

Price,A.L. *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.

Pritchard,J. and Przeworski,M. (2001) Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.*, **69**.

Robinson,L. and Jewell,N. (1991) Some surprising results about covariate adjustment in logistic regression models. *Int. Stat. Rev.*, **59**, 227–240.

Rose,S. and van der Laan,M. (2008) Simple optimal weighting of cases and controls in case–control studies. *Int. J. Biostat.*, **4**.

Scott,L.J. *et al.* (2007) A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science*, **316**, 1341–1345.

So,H.C. *et al.* (2011) Risk prediction of complex diseases from family history and known susceptibility loci, with applications for cancer screening. *Am. J. Hum. Genet.*, **88**, 548–565.

Voight,B.F. *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.*, **42**, 579–589.

Wallace,C. *et al.* (2006) Improved power offered by a score test for linkage disequilibrium mapping of quantitative-trait loci by selective genotyping. *Am. J. Hum. Genet.*, **78**.

Wasserman,L. (2005) All of statistics. *Springer*.

Waters,K.M. *et al.* (2010) Consistent association of type 2 diabetes risk variants found in europeans in diverse racial and ethnic groups. *PLoS Genet.*, **6**.

Wray,N.R. *et al.* (2010) The genetic interpretation of area under the roc curve in genomic profiling. *PLoS Genet.*, **6**, e1000864.

WTCCC (2007a) Association scan of 14,500 nonsynonymous snps in four diseases identifies autoimmunity variants. *Nat. Genet.*, **39**, 1329–1337.

WTCCC (2007b) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, **447**, 661–678.

Yang,J. *et al.* (2010) Comparing apples and oranges: equating the power of case–control and quantitative trait association studies. *Genet. Epidemiol*, **34**, 254–257.

Yang,J. *et al.* (2011) Gcta: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.

Yang,J. *et al.* (2012) Conditional and joint multiple-snp analysis of gwas summary statistics identifies additional variants influencing complex traits. *Nat. Genet.*