Rare Variant Association Analysis Methods for Complex Traits

Jennifer Asimit and Eleftheria Zeggini

Wellcome Trust Sanger Institute, Hinxton CB10 1SA, United Kingdom;
email: Eleftheria@sanger.ac.uk

Annu. Rev. Genet. 2010. 44:293–308

First published online as a Review in Advance on August 18, 2010

The *Annual Review of Genetics* is online at genet.annualreviews.org

This article's doi:
10.1146/annurev-genet-102209-163421

0066-4197/10/1201-0293$20.00

Key Words

1,000 Genomes Project, collapsing method, imputation, sequencing

Abstract

There has been increasing interest in rare variants and their association with disease, and several rare variant–disease associations have already been detected. The usual association tests for common variants are underpowered for detecting variants of lower frequency, so alternative approaches are required. In addition to reviewing the association analysis methods for rare variants, we discuss the limitations of genome-wide association studies in identifying rare variants and the problems that arise in the imputation of rare variants.

INTRODUCTION

There is growing interest in the role of rare variants in multifactorial disease etiology and increasing evidence that rare variants are associated with complex traits. The frequency of any single rare or low-frequency variant is low (<5%), but collectively the number of rare variants makes them quite common. According to the multiple rare variant (MRV) hypothesis, there are many large effect rare variants in the population and each case of a common inherited disease is due to the summation of the effects of a few of these moderate to high penetrance MRVs (2). On the contrary, the common disease common variant (CDCV) hypothesis asserts that the genetic risk of a common complex disease is mostly because of a small number of high-frequency variants with moderately small effects (25), so there must be many to explain the observed genetic variance.

Hundreds of genome-wide association (GWA) studies have been carried out with the focus of identifying common disease variants that are associated with complex diseases. Typically, only variants with minor allele frequency (MAF) greater than 1%–5% are followed up in such studies. Despite the extensive GWA studies resulting in the identification of many genetic variations that have strong evidence of disease association, these variants explain at most 5%–10% of the heritable component of disease. This suggests limitations in GWA studies to identify common variants associated with complex traits and leads in the direction of searching for associations with MRVs (28). The most likely scenario is that a combination of both common and rare variants contributes to disease risk.

One of the reasons why most studies (typically up to a few thousand subjects) identify common causal single-nucleotide polymorphisms (SNPs) is that there is an inverse relationship between sample size and the MAF that maximizes the power to detect a true association (7). Furthermore, SNP genotyping panels are typically designed with a focus on common SNPs, therefore containing a relatively small number of rare variants. Thus, an issue that often arises when performing a rare variant analysis is that most SNP typing platforms are not designed to detect many rare variants (**Table 1**).

The effects of rare variants tend to be larger than those of higher frequency SNPs. Based on published results, there is a clear difference in the distributions of the odds ratios (ORs) for common and rare variants (2), at least within the power constraints of the published studies. Only a few common disease-associated variants have ORs greater than two, and the majority currently fall between 1.1 and 1.4. On the other hand, most identified rare variants to date have an OR greater than two, and the mean OR is 3.74. Furthermore, the identification of rare variants may facilitate pinpointing causality. It can be more difficult to ascribe causality to the majority of loci identified through GWA studies, as high linkage disequilibrium (LD) makes it difficult to use association mapping to determine exactly which variant is functionally relevant. In addition, when (common or rare) SNPs map to genomic regions that do not have a clear role, elucidating their effects can become especially challenging. The problem may be simplified by searching for disease-associated rare variants in known functional genomic regions, defined as genes. In addition, it might be easier to at least infer causality at a locus that contains both common and rare disease-associated variants.

Table 1 Approximate-low frequency/rare variant GWAS platform content

Platform	Affymetrix 500k	Affymetrix 6.0M	Illumina 370k	Illumina 550k	Illumina 610k	Illumina 1.2M
MAF < 0.05	55k	106k	9k	32k	35k	62k
MAF < 0.01	17k	35k	1k	7k	8k	22k

Below, we discuss several known rare variant–disease associations and the approach that was used to find them. Various tests of association, from multiple single-marker to multimarker to collapsing methods are discussed. Finally, we review the issues that arise when studying rare variants, including those encountered when imputing genotypes.

KNOWN RARE VARIANT–DISEASE ASSOCIATIONS

The literature documents a growing body of evidence supporting the role of rare variants in complex trait associations. For example, in the search for causal variants of type 1 diabetes (T1D), Nejentsev et al. (23) identified four disease-associated rare variants in the *IFIH1* gene, which are protective of T1D. They resequenced exons and splice sites of 10 candidate genes for 480 patients and 480 controls, and tested for association with T1D at the 212 identified SNPs. They confirmed several previously identified common SNPs and detected associations with several rare variants occurring only in the *IFIH1* gene, of which rs35667974 and rs35337543 had the strongest associations (Fisher's-exact p-values of 4.4×10^{-5} and 0.0049). Next, association with T1D was tested for in case-control (8,379 cases, 10,575 controls) and family (3,165 families where at least one child has T1D) collections. Among the rare variants detected in *IFIH1* by sequencing, a total of 4 (rs35667974, rs35337543, rs35744605, and rs35732034) were found to be associated in the larger sample as well. The respective MAFs in cases were 0.011, 0.010, 0.0046, and 0.0069, whereas controls had respective MAFs of 0.022, 0.015, 0.0067, and 0.0093. The case-control odds ratios were 0.51 ($P = 1.3 \times 10^{-14}$), 0.68 ($P = 1.1 \times 10^{-4}$), 0.69 ($P = 9.0 \times 10^{-3}$), and 0.74 ($P = 0.012$), whereas the family study relative risks were 0.60 ($P = 5.9 \times 10^{-4}$), 0.85 ($P = 0.20$), 0.55 ($P = 0.028$), and 0.63 ($P = 0.021$). The combined p-values for the case-control and family studies were 2.1×10^{-16}, 1.4×10^{-4}, 1.3×10^{-3}, and 1.1×10^{-3}. Each of these rare variants protects from T1D, and they all have stronger protective effects than the common nonsynonymous SNP (nsSNP) rs1990760/T946A (OR = 0.86), identified in previous GWA studies. Moreover, the LD between any pair from these four rare variants is low ($r^2 < 0.04$), and each is found to be associated with T1D independently of each other and of the common nsSNP. This indicates that there are four rare polymorphisms and one common nsSNP in the *IFIH1* gene that show independent association with T1D.

Convincing evidence for the involvement of rare variants in hypertension has also been produced. In the offspring cohort of the Framingham Heart Study (FHS), Ji et al. (10) examined all codons and flanking intronic sequences of three genes (*SLC12A1*, *SLC12A3*, and *KCNJ1*) that are known to cause rare recessive diseases that are characterized by very low blood pressure. Using their validated criteria of phylogenetic conservation and rare allele frequency (MAF < 0.001), they reduced the 138 identified coding sequence variants to a set of 30 functional variants, of which almost all were predicted to be damaging by bioinformatics tools and have MAF < 0.0005. The mean long-term systolic and diastolic blood pressures among mutation carriers are both lower than the cohort mean by 6.3 mm Hg ($P = 0.0009$) and 3.4 mm Hg ($P = 0.003$), respectively, and there are similar patterns from measurements taken at various ages. This result was also confirmed by a within-family test comparing the blood pressures between siblings in FHS who were discordant for mutations. In each age group (25–40, 41–50, 51–60), the prevalence of hypertension was found to be lower for mutation carriers, and compared with noncarriers, the carriers also have a 59% reduction in the risk of developing hypertension by the age of 60, as revealed by a Kaplan-Meier analysis (log-rank $P < 0.003$).

Perhaps the earliest evidence in support of the contributions of rare variants to complex traits came from the examination of low plasma levels of high-density lipoprotein cholesterol (HDL-C) by Cohen et al. (5). In their study,

the coding regions and consensus splice sites of three candidate genes (*ABCA1*, *APOA1*, and *LCAT*) were sequenced in 256 individuals who formed the upper and lower 5% of HDL-C levels from the Dallas Heart Study population. Among the 128 individuals from the low HDL-C group, 21 had rare sequence variants (20 in *ABCA1*) that were not found in the high HDL-C group, compared with only three individuals from the high HDL-C group who had sequence variants not present in the low HDL-C group (Fisher's exact test $P < 0.0001$). A similar pattern was also found in their analysis on samples of Canadians with HDL-C at the extremes of the distribution (Fisher's exact test $P < 0.001$). In another study, Cohen et al. (6) identified multiple rare alleles that collectively contribute to a significant proportion of genetic variance in low-density lipoprotein cholesterol (LDL-C). First, the coding regions of a candidate gene, *NCP1L1*, were sequenced in individuals from the Dallas Heart Study who had the highest and lowest cholesterol absorption, as indicated by the Ca:L ratio (128 subjects in each group), such that each group consisted of 32 individuals from the groups of black males, white males, black females, and white females. They found 13 nsSNPs unique to the low level group and 3 nsSNPs present only in the high-level group (Fisher's exact test $P < 0.01$). A similar result was obtained when they repeated the analysis on the next set of 128 subjects from each of the two extremes (Fisher's exact test $P < 0.025$). Next, the variants were analyzed in the complete Dallas Heart Study (3,553 individuals) to examine the frequencies and phenotypic effects of the identified nsSNPs. The majority of nsSNPs identified were found in African-Americans. Because of the extremely low allele frequencies (0.03–0.6%), a particular rare variant exists in too few individuals for a variant-by-variant statistical analysis. However, the aggregate of these variants was associated with a significant reduction in the Ca:L ratio and in the plasma LDL-C concentrations in African-Americans (Wilcoxon's two-sample test $P < 0.01$).

TESTS OF ASSOCIATION

Although rare variants have a proven role in some complex traits, they have not yet been studied as extensively as common SNPs. There are several reasons for this, including the lack of a rare variant catalog with reference genotypes (poised to change with the 1,000 Genomes Project), current cost limitations in next generation sequencing technologies, and a lack of an appropriate analytical toolbox to enable powerful rare variant association analysis given currently available sample sizes. In a typical association analysis with a focus on identifying common disease-associated variants, the number of SNPs to test is reduced by taking into account LD because there is no gain in information by including highly correlated SNPs. Within a group of high LD SNPs, a tagSNP can be chosen such that the set of tagSNPs is of minimal size to explain the majority of the genome sequence variation. Sets of tagSNPs are usually chosen from publicly available genotype data based on unrelated individuals from various reference populations, such as HapMap. The International HapMap Consortium was designed with identifying tagSNP sets as a prime objective. However, it has been found that although common variation is captured quite well by creating tagSNP sets using HapMap, the sample sizes available are not sufficient for tagging variants with lower MAFs (<0.05) (37). Thus, common variants have only a limited capacity to tag rare variants.

Many indirect LD methods have been developed to identify disease-associated common variants based on the idea of analyzing tagSNPs, but such methods lose power when applied to rare variants. Rare variants are weakly correlated with common tag SNPs because MAFs must be similar in order for two variants to be highly correlated (34). This low correlation, together with the low frequency of rare variants, results in low power to detect associations via indirect LD mapping, so direct mapping through exhaustive genotyping or sequencing is necessary in the search for rare variants. Sequencing the entire genome is now an option to

sequencing only candidate genes, as the cost of generating massive amounts of sequencing data is dropping with the introduction of new sequencing technologies (29). This has led to the 1,000 Genomes Project, which will sequence 2,000 genomes in order to develop a high resolution human genome map that will include almost all variants with allele frequencies as low as 1%.

In the analysis of sequencing data to identify disease-associated rare variants, the tests for association fall into three main types: multiple univariate single-marker tests, multiple-marker tests, and collapsing methods. Within a functional unit of interest, all rare variants that are genotyped may be included in the analysis at the cost of a slight loss in power from inclusion of nonfunctional variants (see below). Alternatively, only those classified as functional by bioinformatics tools may be included in the analysis. In the latter case, this can be done by determining the confidence in the predictions of functionality for the variants [e.g., PolyPhen (26) and SIFT (24)] or by classifying the variants as potentially functional or neutral [e.g., Evolutionary Trace (14)].

Single-Marker Tests

The simplest approach to testing for association is to apply a univariate test at each rare variant and then assess significance, taking into account multiple testing by using an appropriately scaled p-value threshold for declaring significance. For case-control data, possible methods include the χ^2 test, Fisher's exact test, Cochran-Armitage test for trend, and logistic regression, whereas linear regression may be applied in the case of quantitative traits. Given that each variant is tested independently for an association, a correction for the multiple comparisons within the family of tests is needed so that the family-wise error rate (FWER) is controlled. This results in a loss of power. Several methods have been developed to account for the multiplicity that arises. The most common approaches are random permutation tests to obtain empirical p-values, or methods to control the false discovery rate (FDR), which is the expected proportion of incorrect rejections of the null hypothesis (1). The FDR is smaller than the FWER and equal only when all of the hypotheses are true. This indicates that by controlling the FDR, rather than FWER, there may be a gain in power.

To explore the properties of these tests, assume that there are n subjects and m rare variants in the locus. Regression (logistic or linear) can be used to test for association with a trait (binary or quantitative) by fitting a regression model at each of the m variants, possibly including covariates. Denote the phenotype for subject i by y_i, let x_{ij} be the minor allele count at variant j for subject i. In the case of linear regression, the relationship at variant j may be modeled by $y_i = \alpha_j + \beta_j x_{ij} + \boldsymbol{\eta_j z_{j_i}} + \varepsilon_i$, where $\boldsymbol{z_j}$ is a matrix of covariates, which may be included, and there is the usual assumption of ε_i being independent normal random variables with mean 0. The null hypothesis of no association at variant j is equivalent to testing $\beta_j = 0$ in the regression. For logistic regression y_i is replaced by $\log(\frac{p_i}{1-p_i})$, where p_i is the probability of the presence of the binary trait, e.g., disease presence.

For the χ^2 test, Fisher's exact test and the Cochran-Armitage test for trend, a 2×3 contingency table may be constructed to compare genotype frequencies between cases and controls at a specific variant; the rows are disease status and columns correspond to the three possible genotypes. Assume that a is the high risk and rare minor allele. In the Cochran-Armitage test for trend, it is assumed that the genotypes can be viewed as ordered categories: *AA*, *Aa*, *aa*, i.e., the number of rare alleles. The test for association measures a linear trend in proportions weighted by the category effects, which are typically taken to be the number of a alleles (31).

Both the χ^2 and Fisher exact tests consider the null hypothesis of equal genotype frequencies in cases and controls. However, the χ^2 test approximates the significance with an accuracy that increases with sample size, provided that cell counts are large enough (e.g.,

>5), whereas Fisher's exact test yields exact results, so it is recommended when there are small counts. The expected cell counts for *aa* will be extremely low because of the rare variants. An allele-based rather than genotype-based test can be carried out so that only a 2×2 table is required, but the expected combined counts will still be low (11). Thus, Fisher's exact test should be used rather than the χ^2 test to control the type I error. This comes at the cost of reduced power because Fisher's exact test is more conservative.

A quick method to correct for the multiple hypothesis tests is to use a Bonferroni correction, but because it is too conservative it is not used in practice. For completeness, a brief description of it is given below. Assuming that there are m variants being tested within the locus, and the significance level for the m independent hypothesis tests is α, the Bonferroni correction is to use α/m as the significance criterion for each of the m individual tests. A nonparametric approach is to obtain empirical p-values by using random permutation testing, which does not make any assumptions about the joint distribution of the m test statistics, but is computationally intensive. The idea behind random permutation testing is to approximate the reference distributions of the test statistics under the null assumption that there is no difference between the cases and controls. Denote the p-value from the hypothesis test for variant j, based on the original dataset, by p_j. A sample k is generated by permuting the case and control labels. Such samples are generated B times, and for each sample k, the same multiple tests are performed as in the original dataset to obtain B sets of m p-values p^*_{jk}, $j = 1, \ldots, m$. The empirical p-value p^*_j for variant j is the proportion of the p^*_{jk} that are at least as significant as the original p-value p_{jk}, i.e., $p^*_j = \#\{k : p^*_{jk} \leq p_{jk}\}/B$.

To control the FDR for independent test statistics, Benjamini & Hochberg (1) developed a sequential Bonferroni-type procedure. The m p-values from the single SNP tests are first ordered: $P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(m)}$. At FDR level q, let k be the largest i such that $P_{(i)} \leq \frac{i}{m}q$. Then, the null hypothesis is rejected for SNPs with p-values smaller than $P_{(k)}$.

As a result of the multiple testing penalty, single-marker tests lack power to detect low to moderate effects. Even without adjusting for multiple testing, the power of a single-marker test at a single low-frequency variant has been demonstrated to be very sensitive to the effect size. Approximate sample sizes (cases+controls, with equal sized groups) required to attain a power of 0.8 to detect an allelic odds ratio of 2 at an α level of 5×10^{-8} increase from 2,500 to 12,000 to 117,000 samples, as the MAF decreases, respectively, from 0.05 to 0.01 to 0.001. Likewise, given that joint associations are not accounted for with single-marker tests, multiple SNPs with moderate effect sizes will also have low power. When there is allelic heterogeneity within a locus, there is a further reduction in power because various individuals will contribute to a signal at different variants in the locus. In addition, the power is greatly affected by low allele frequency at a variant. In a simulation study by Li & Leal (11), exceptionally low powers were observed in the application of single-marker association tests to rare variants in a locus. It is common practice for common SNP GWA studies to impose a strict p-value threshold for declaring significance ($p < 5 \times 10^{-8}$). This is based on the approximate number of estimated independent common variants across the genome. It is not yet clear what the equivalent threshold for rare variants would be.

Multiple-Marker Tests

An alternative to testing each variant separately is to use multivariate methods to combine information across the variants and simultaneously test the multiple variant sites. Therefore, in the case of multiple moderate SNP effects, a multiple-marker test will have higher power than single-marker tests. Possible multimarker approaches include Fisher's method, Hotelling's T^2 test, and multiple regression (logistic or linear). All of these methods require multiple degrees of freedom, which lowers the

power of the test, especially in the case that there is only a single strong signal in the region.

Fisher's method is a way of combining the results from the m single-marker tests, but it is anticonservative when there is dependency among the single tests. Denoting the m p-values obtained from any of the single-marker tests by p_i, the test statistic is $X^2 = -2\sum_{i=1}^{m}\log(p_i)$. When all of the p_i are independent and all of the null hypotheses are true, X^2 follows a χ^2_{2m} distribution, from which the p-value can be obtained.

A multiple regression model may be used to jointly test for association between the variants and phenotype rather than fitting m separate regression models at each of the rare variants. In the simplest case of no covariates the regression model for a binary trait is

$$y_i = \alpha + \boldsymbol{X}\boldsymbol{\beta}_i + \varepsilon_i,$$

where $\boldsymbol{X}$ is the $n \times m$ matrix of minor allele counts for the n subjects at each of the m rare variants, and $\boldsymbol{\beta}$ is the m vector of regression coefficients. By jointly estimating the associations at each variant, m degrees of freedom are required for the fit so that the corresponding test statistics for each null hypothesis of $\beta_j = 0$ have $n - m$ degrees of freedom rather than $n - 1$ as in the single-marker case.

For case-control studies, Hotelling's two-sample T^2 test may be used, which is a multivariate generalization of Student's t-test (33). Assume that there are N_A affected and $N_{\bar{A}}$ unaffected individuals. In order to calculate the test statistic, indicator variables X_{ij} and Y_{ij} are defined for the genotype of the jth marker for the ith individual from the case and control populations, respectively. More explicitly, for the N_A cases we have

$$X_{ij} = \begin{cases} 1, & \text{if } aa \\ 0, & \text{if } Aa, \\ -1, & \text{if } AA \end{cases}$$

whereas Y_{ij} is defined in an identical fashion for the $N_{\bar{A}}$ controls. Let $X_i = (X_{i1}, \ldots, X_{im})^T$; $i = 1, \ldots, N_A$ for cases and $Y_i = (Y_{i1}, \ldots, Y_{ik})^T$; $i = 1, \ldots, N_{\bar{A}}$ for controls. Upon determining the pooled-sample covariance matrix S of the X_i and Y_i, Hotelling's two-sample T^2 test statistic is

$$T^2 = \frac{N_A N_{\bar{A}}}{N_A + N_{\bar{A}}}(\bar{X} - \bar{Y})^T S(\bar{X} - \bar{Y}), \quad (1)$$

and under H_0, $\frac{N_A+N_{\bar{A}}-m-1}{m(N_A+N_{\bar{A}}-2)}T^2$ follows an $F_{m,N_A+N_{\bar{A}}-m-1}$ distribution.

Several issues arise with the use of multi-marker tests in addition to the problem of multiple degrees of freedom. Multiple marker tests are also sensitive to allele frequencies. Li & Leal (11) demonstrated in a simulation study on rare variants that Hotelling's T^2 test is greatly affected by the MAF and has a reduction in power when the number of rare causal variants increases. Multi-marker approaches that have a reduction in the degrees of freedom include the ZGlobal statistic of Schaid et al. (27) and a variant of it developed by Wang & Elston (35). However, with these methods, as well as many others, the risk allele must first be identified at each variant, and the direction of the genotype scores at each SNP affects the test power. A powerful approach for testing markers within a small genomic region is multivariate distance matrix regression (MDMR) (36), which uses a matrix of genetic similarity among individuals, and via this matrix the marker genotypes are used as dependent variables. The similarity scores in the matrix for MDMR are calculated without knowledge of the risk allele, and the method can be applied to continuous or discrete traits, but it is not very powerful for a set of independent SNPs across the genome. The kernel-based association test (KBAT) also jointly tests multiple SNPs without making any assumptions on the direction of individual SNP effects, but it is also able to handle correlated and/or independent SNPs (22). In simulation studies done by the authors, KBAT was found to generally have more power than Zglobal (27) and MDMR (36), especially in the presence of rare causal SNPs.

KBAT is based on genotype similarity scores, measured by a kernel function, between individuals within the same group (e.g., cases or controls). First, similarity scores $y_{l(ij)}$ between

individuals i and j in group l (1 = cases, 2 = controls) are determined by using a kernel, such as the allele match (AM) kernel, which is the count of common alleles between the genotypes of two individuals. By defining the kernel in this way, there is no need to have knowledge of the risk allele at each SNP. The similarity scores $y^k_{l(ij)}$ between individuals i and j in group l at SNP k are modeled using a one-way ANOVA model at each SNP,

$$y^k_{l(ij)} = \mu^k + \alpha^k_l + \varepsilon^k_{l(ij)}, \quad i < j = 1, \ldots, n_l; \quad l = 1, 2,$$

where at SNP k, μ^k is the general effect for pairs of individuals, α^k_l is the group specific treatment effect, and to test for disease association the null hypothesis is $H_0 : \alpha^k_1 = \alpha^k_2$. Let n_l be the number of individuals in group l, and let m_l be the number of possible distinct pairs of individuals (i.e., $m_l = n_l(n_l - 1)/2$), which corresponds to the number of distinct similarity scores. Let $\bar{y}^k_l = \sum_{i<j} y^k_{l(ij)}/m_l$ be the mean for group l, and denote the grand mean by $\bar{y}^k = \sum_{l=1}^{2} \sum_{i<j} y^k_{l(ij)} / \sum_{l=1}^{2} m_l$. The between-group sum of squares at marker k is defined by $SSB_k = \sum_{l=1}^{2} m_l(\bar{y}^k_l - \bar{y}^k)^2$, whereas the corresponding within-group sum of squares is given by $SSW_k = \sum_{l=1}^{2} \sum_{i<j} (y^k_{l(ij)} - \bar{y}^k)^2$. The single SNP test statistic at marker k is then given by $\frac{SSB_k}{SSW_k}$, and the K-marker KBAT test statistic to test for disease association with the set of K markers is

$$\frac{\sum_{k=1}^{K} SSB_k}{\sum_{k=1}^{K} SSW_k}.$$

Permutation is required to obtain the p-value of the KBAT statistic (or any of the single SNP statistics) since the $y_{l(ij)}$ are not independent normals.

Collapsing Methods

Individually, low-frequency variants are rare, but in aggregate they may be common enough to account for variation in common traits, which is the basic idea behind collapsing methods. The goal of collapsing methods is to test for an association of an accumulation of rare minor alleles with some trait by combining information across multiple variant sites. For each individual, in some manner, genotypes are collapsed across variants within the same region (group) so that each individual has a single quantity for that region. For example, the genotypes of the variants in a group are collapsed for each individual to an indicator variable for the presence of at least one rare allele at any of the variant sites, using the rationale that there should be a low probability of an individual carrying more than one rare allele (11, 17). Groups may correspond to genes or may be defined by allele frequencies or functionality. When grouping by function or genes, the focus changes from the sole identification of highly associated genomic regions to the causal relations between genes and diseases (16). This leads to a complex relation between sample size, minor allele freuqency, and power. Gorlov et al. (7) illustrated that the power to detect a true association (joint probability of a SNP being identified as both functional and disease-associated) is influenced by both the MAF and sample size. Moreover, the MAF at which this power is maximal is inversely related with sample size, and power to detect an association at an identified functional variant is quite low.

The collapsing approach applies a single univariate test to the collapsed data within a group, resulting in enriched signals and fewer degrees of freedom, rather than facing multiple correction factors for many single-marker tests or high degrees of freedom in a multiple-marker test (11). The weighted sum statistic (16) and combined multivariate and collapsing (CMC) method (11) are specific to case-control data, whereas the regression approaches of Morris & Zeggini (17) can be applied to quantitative data as well. In describing these methods below, we assume that there are N individuals and that the genotype score at a variant is the number of minor alleles present for the individual.

The motivation behind the CMC method (11) is the development of a test that combines the high power for rare variant analyses (collapsing methods) with robustness to inclusion

of noncausal variants (multivariate test). The CMC method is used for multiple predefined groups and expands on the collapsing method for a single group of markers. In both methods, only variants classified as functional are included in the analysis. Choices for the univariate test to be applied to the collapsed data that are considered are the Cochran-Armitage test for trend (12), χ^2 test, and logistic regression (11).

In the application of the Cochran-Armitage test for trend, the rare variants within a locus are collapsed for each individual to the number of minor alleles that are carried by the subject. It is assumed that the probability of being diseased increases with the number of rare minor alleles, so the counts of minor alleles can be treated as ordered categories. This assumption is expected to hold for the allelic heterogeneity model, for which there is independence between multiple causal rare variants. A contingency table is then constructed to compare the frequencies of the minor allele counts between cases and controls (12).

For the χ^2 test, each individual is coded by the indicator variable of at least one rare allele present at any of the variants within the locus. The proportions of individuals with rare variants in cases and in controls, ϕ_A and $\phi_{\bar{A}}$, respectively, are then tested for a difference via a χ^2 statistic with 1 degree of freedom,

$$N\left[\frac{(\hat{\phi}_A-\hat{\phi}_{\bar{A}})^2}{\hat{\phi}_A+\hat{\phi}_{\bar{A}}}+\frac{(\hat{\phi}_A-\hat{\phi}_{\bar{A}})^2}{2-\hat{\phi}_A-\hat{\phi}_{\bar{A}}}\right],$$

where $\hat{\phi}_A$ and $\hat{\phi}_{\bar{A}}$ are the corresponding observed proportions. The power of the χ^2 test is determined by the noncentral χ^2_1 distribution with noncentrality parameter given by

$$\nu_c = N\left[\frac{(\phi_A-\phi_{\bar{A}})^2}{\phi_A+\phi_{\bar{A}}}+\frac{(\phi_A-\phi_{\bar{A}})^2}{2-\phi_A-\phi_{\bar{A}}}\right].$$

Expressions for ϕ_A and $\phi_{\bar{A}}$ and their derivations are provided in Li & Leal (11). The power of the test is given by $\eta_c = \Pr(\chi^2_1(\nu_c) \geq \chi^2_{1,1-\alpha})$.

Li & Leal (11) use allele frequencies to determine the partition of the variants into groups, with 0.1 as a criterion; they only collapse variants with allele frequencies below 0.1. They suggest multiple groups defined by several cut-offs when there is a wide spectrum of allele frequencies. As discovered in their simulation study, this avoids a massive loss of power from misclassification when variants with very different allele frequencies are collapsed into the same group. These authors also remark that in addition to ensuring that only variants with similar allele frequencies are collapsed together, care must also be taken so that protective and high-risk variants are collapsed separately (although we note that this is difficult to implement in practice). If all functional variants have the same affect on disease risk, then collapsing will enrich the signal. On the other hand, there will be a weakened signal if variants that increase disease risk are collapsed with those that reduce disease risk.

After the rare variants are partitioned into k groups, $\{g_j, j = 1, \ldots, k\}$, Li & Leal (11) collapse the variants within each group to an indicator of any rare allele presence and then apply a multivariate test, such as Hotelling's T^2 test or logistic regression, to the collection of these quantities to test the null hypothesis that none of the groups are associated with disease susceptibility. Details of the implementation of Hotelling's T^2 follow (11).

In the application of Hotelling's T^2, Li & Leal (11) define n_j to be the number of markers in group g_j. Within group j, the indicator of the presence of any rare allele for member i of the case population is denoted by X_{ij}, and Y_{ij} is defined similarly for subject i of the controls. In this manner, for each of the k groups, the n_j variants in group g_j are collapsed for each individual. Each subject is then represented by a k-vector of indicators: $X_i = (X_{i1}, \ldots, X_{ik})^T$; $i = 1, \ldots, N_A$ for cases and $Y_i = (Y_{i1ts}, \ldots, Y_{ik})^T$; $i = 1, \ldots, N_{\bar{A}}$ for controls. The proportions of cases and controls with at least one rare allele in group g_j are $\bar{X}_j$ and $\bar{Y}_j$, respectively, and letting S denote the pooled covariance matrix, the test statistic is as given in equation 1. Under the null hypothesis, $\frac{N_A+N_{\bar{A}}-k-1}{k(N_A+N_{\bar{A}}-2)}T^2$ follows an $F_{k,N_A+N_{\bar{A}}-k-1}$ distribution.

Li & Leal (11) analytically compare the power of single-marker tests, multiple-marker tests, the collapsing method, and the CMC method, taking into account a number of functional variants and functional misclassification of variants. In a simulation study, LD effects on test power are examined, as well as type I error rates at the 0.05 level. The CMC method (with Hotelling's T^2 or logistic regression) controls the type I error quite well, as does Hotelling's T^2 and the collapsing method based on the χ^2 test, but the latter two are slightly conservative. Logistic regression (uncollapsed data) displays poor performance, with an inflated type I error. When LD is present, among the collapsing method (χ^2 test), Hotelling's T^2 test, and the single-marker tests, the single-marker tests have the lowest power, whereas the collapsing method has the highest.

The power of the collapsing method increases with the number of functional variants, whereas the powers of the multiple- and single-marker tests have the inverse relationship with the counts. When nonfunctional variants are included in the analysis or functional variants are excluded from the analysis, all three tests experience a drop in power, with the single-marker test consistently having the lowest power. The collapsing method manages to have the highest power in all scenarios, despite being generally less robust to functional misclassification than Hotelling's T^2. As the proportion of variants excluded from the analysis increases, there is a larger drop in the power of the collapsing method than in Hotelling's T^2. However, the exclusion of high-frequency (e.g., frequency 0.02 or 0.05) functional variants results in a more dramatic loss of power for Hotelling's T^2 than the collapsing method. Hotelling's T^2 is found to be quite robust to the inclusion of nonfunctional variants, irrespective of their allele frequencies, and decreases slightly in power as the number of nonfunctional variants increases. On the other hand, the power of the collapsing method noticeably decreases as the allele frequency of a single nonfunctional variant increases, and there is a dramatic loss of power with the inclusion of two high-frequency (e.g., frequency 0.02 or 0.05) nonfunctional variants.

The CMC method is demonstrated to have high power and is robust against functional variant misclassification, combining the strengths of collapsing and multivariate methods. In the presence of functional misclassification, the power of the CMC method is much higher than that of the collapsing method, especially when a high-frequency noncausal variant is included in the analysis. The power of the CMC method decreases upon increasing the number of noncausal variants in the analysis, but it is not affected by the noncausal allele frequency. In the analysis of data with truly functional high-frequency variants (allele frequency of 0.02 or 0.05), the CMC method has only a slightly lower power than the collapsing method.

Collapsing methods in a regression framework have been developed by Morris & Zeggini (17). They focus on quantitative trait associations, assuming that a normally distributed trait is phenotyped for a sample of unrelated individuals who are typed for rare variants in a gene or small genomic region. However, the method is easily extended to binary traits by considering a logistic regression-modeling framework. They model the phenotype as a function of a collapsed summary of the variants in one of two ways, referred to as rare variant tests (RVT):

- RVT1: for each individual, the proportion of rare variants that carry at least one copy of the minor allele;
- RVT2: for each individual, the presence or absence of at least one minor allele at any rare variant.

The phenotype for individual i is denoted y_i, whereas n_i denotes the number of successfully genotyped rare variants for individual i, r_i denotes the number of rare variants that carry at least one copy of the minor allele, and $I(r_i) = 1\{r_i > 0\}$ is the indicator variable for the presence of at least one minor allele at any rare variant for subject i. Incorporating a vector of covariates for individual i, $\mathbf{x}_i$, the regression

models are given by

$$y_i = \alpha + \lambda \frac{r_i}{n_i} + \boldsymbol{\beta x_i} + \varepsilon_i, \quad (2)$$

$$y_i = \alpha + \lambda I(r_i) + \boldsymbol{\beta x_i} + \varepsilon_i, \quad (3)$$

where in equation 2, λ is the expected increase in the phenotype for an individual with a minor allele at all rare variants compared with one with none, and in equation 3, λ is the expected increase when an individual carries at least one minor allele at any rare variant compared with one that has a complete absence of rare minor alleles. Analysis of deviance is used to compare the maximized likelihoods of the null ($\lambda = 0$) and unconstrained λ models in the construction of likelihood ratio tests of disease association for an accumulation of rare variants.

Within a functional unit of interest, the power of the association tests based on equations 2 and 3, as well as independent Bonferroni-corrected trend tests of quantitative trait association at all SNPs and haplotype trend tests of association, are compared by Morris & Zeggini (17) in a simulation study in which haplotype data are generated in a 50-kb genomic region. Each of the four methods is applied to simulated data in a manner that is equivalent to testing at low frequency variants (MAF 1%–5%) present on the genome-wide SNP chip. When testing on the SNP chip, it is necessary to allow the MAF to include SNPs with MAFs as high as 5% so that the mean number of variants tested is 1.6 rather than only 0.2 when the analysis is restricted to variants with MAF < 1%. The regression methods are also applied to simulated data, testing in a cohort those rare variants that were identified through deep resequencing. They examine various simulation models for association of the trait with multiple causal variants in the same region, varying the maximum MAF of all causal variants, the total MAF of all causal variants, and their joint contribution to the phenotypic variance. Phenotypes are either generated under the assumption that the trait is determined by the proportion of causal variants at which a minor allele is present (equation 2) or by the presence or absence of a minor allele at any casual variant (equation 3).

Morris & Zeggini (17) find that rare variants discovered through resequencing tests based on equation 2 tend to be more powerful than those of equation 3, even when the simulated traits are determined by equation 3. This indicates that tests based on the proportion of rare variants that carry at least one minor allele are more robust than those based on the presence/absence of any minor allele. A drawback of the proportions-based method is that it can be adversely affected by the presence of LD. Tests based on low-frequency variants on genome-wide SNP arrays display a distinct loss in power in comparison to those based on rare variants identified via resequencing, with the largest differences in power resulting when there is substantial allelic heterogeneity (e.g., maximum MAF of 0.5% for causal rare variants in simulated data). In addition, the regression methods, which are based on accumulations of minor alleles, have low power to identify rare variant associations at low-frequency variants on genome-wide SNP chips because of their scarcity on the chips. They also confirm the result that when there is substantial allelic heterogeneity, rare variant associations are detected with greater power by haplotype-based tests than by single-locus tests (21).

Both the cohort allelic sums test (CAST) (5, 19) and the weighted sum statistic (16) test for a difference in mutation counts in a group of variants between cases and controls by coding individuals according to the number of mutations. Both rare and common variants within the functional unit are included in calculations for these two methods, but in the weighted sum statistic the variants are weighted according to their frequency in the controls. By differing the weight contributions of the variants, the impact of the common mutations is not as high as in the CAST method, where many individuals may be grouped as having at least one mutation.

In the construction of the weighted sum statistic, each variant within the functional unit is assigned a weight $\hat{w}_i$, which is the estimated standard deviation of the mutation count in the

sample, under the null assumption of equal frequencies in affected and unaffected subjects

$$\hat{w}_i = \sqrt{n_i q_i (1 - q_i)}, \quad \text{where} \quad q_i = \frac{m_i^U + 1}{2n_i^U + 1},$$

m_i^U is the controls count of mutant alleles for variant i, n_i^U is the number of genotyped controls for variant i, and n_i is the total number of subjects genotyped for variant i. These weights are used to downweight the mutation counts I_{ij}, for variant i, subject j, when calculating the genetic score of individual j

$$\gamma_j = \sum_{j=1}^{L} \frac{I_{ij}}{\hat{w}_i},$$

where in the generic genetic model $I_{ij} \in \{0, 1, 2\}$. These weighted sums of mutation counts are then used to rank all individuals, regardless of affected status, and the test statistic x is the sum of the ranks for the cases, which is equivalent to the Wilcoxon rank statistic.

Madsen & Browning (16) proceed in two ways to determine the p-value, both depending on random permutation methods: a normal approximation and the standard method. Here we will focus on the standard method, which is the preferred, quicker method because it incorporates a stopping rule so that fewer permutations may be required. In either method, a sample under the null hypothesis is obtained by permuting the affected/unaffected status, and the test statistic is calculated for the sample. In this manner, the LD structure of the genetic data is preserved so that regardless of whether or not the variants are in LD, the test has a correct false-positive rate. This sampling is repeated k times to obtain $x_1^*, \ldots, x_k^*$. Let $k_0 = \#\{j : x_j^* \geq x\}$, so that the p-value is found by $\hat{p} = \frac{k_0+1}{k+1}$, under the standard approach. Sampling may be suppressed if $\hat{p}$ and its precision are satisfactory, such as when $\hat{p} - 3\hat{\sigma}_{\hat{p}}$ is above the significance threshold, where $\hat{\sigma}_{\hat{p}}$ is the estimated standard deviation of $\hat{p}$.

A discussion on the power of the weighted sum statistic is deferred to follow the description of the data-adaptive sum test (8), which tests for disease association with multiple rare and/or common variants in a region. In the usual sum test, a common-effect logistic regression model is fit for a region of k SNPs, assuming that all SNPs have a common odds ratio for disease association

$$\text{logit } Pr(Y_i = 1) = \beta_{c0} + \sum_{j=1}^{k} X_{ij}\beta_c,$$

where β_c is the common association strength between the disease and each SNP (8). This avoids multiple degrees of freedom and the use of multiple test adjustments given that the single test of interest is $H_0{:}\beta_c = 0$, which can be tested by a score statistic (or Wald statistic). However, the least-squares estimate of β_c is a function of the k single SNP regression coefficients that would result from fitting a marginal model at each SNP (logit $Pr(Y_i = 1) = \beta_{0j} + X_{ij}\beta_j$) and depends heavily on the association directions; there is a large power loss if the coefficient signs are quite different. The data-adaptive sum test adapts the coding of each SNP such that SNP codings are optimal. This is done by first fitting marginal logistic regression models at each SNP j and selecting a threshold α_0. If the jth SNP coefficient is negative and has a p-value below α_0, then the jth SNP coding is reversed. The final set of SNP codings is then used in the fitting of the common-effect model. A normal-based p-value p can be found from the normalized score statistic under $H_0 : \beta_c = 0$, but because of the data-driven coding approach that shifts the distribution of the score statistic, it will yield inflated type I error rates. A proper permutation-based p-value is found in the usual manner of permuting the case-control indicators, then fitting the marginal and common-effect models as before, and finally finding the proportion of permutations with smaller p-values than the original data. The authors refer to this test as aSum-P, and they also consider the empirical distribution of the score statistics from the B permutations to construct the aSum test statistic, which has a null distribution that is a linear function of a χ_1^2 random variable.

The authors also consider a test in which they partition the variants into two groups: common and rare. Then, in a logistic regression, the two corresponding regression coefficients are tested. They refer to this approach as aSumC, and when a permutation-based p-value is implemented it is called aSumC-P (8).

In a simulation study, the authors compare variants of their data-adapted sum test with the usual sum test, as well as the CMC method (using Hotelling's T^2) and weighted sum test (8). The sum test and weighted sum test have the highest power when all causal associations are in the same direction and there are not any noncausal SNPs. With the addition of noncausal SNPs with low MAF (0.02 or 0.05), the CMC test tends to achieve the highest power, closely followed by the aSumC-P and aSumC tests, but this is not the case for the inclusion of nonfunctional rare variants, where the sum test variants perform better than the CMC test. In the situations where the direction of association differs among the SNPs, the sum test and weighted sum test have the lowest powers, with or without the inclusion of nonfunctional variants; generally the various versions of the aSum tests attain the highest powers, but the CMC test occasionally performed better.

ISSUES AND DISCUSSION

The analysis of rare variants is not only complicated by low power, but also by factors such as the difficulty of calling rare genotypes and of uncertainty associated with sequence calls. Individual samples are resequenced many times and the combined reads are used to calculate SNP-specific quality scores, which evaluate how likely a polymorphism truly exists at a particular location and how confident the genotype assignment is. Rare variants with lower quality scores are not as reliable and may need to be downweighted in analyses.

In the search for causal rare variants, the probability of detection is higher when genes are sequenced only in cases, or when the proportion of cases is larger than controls, but this approach is problematic. When variant discovery is based on an excess of cases and the remaining samples are genotyped at these variants, there tends to be inflated false-positive rates, but the type I error rate is well-controlled when both cases and controls are used for discovery (12).

Genotype imputation is often used to predict unobserved/missing genotypes in order to obtain a larger set of SNPs over a finer grid for analysis in genome-wide association studies. Several imputation algorithms have been developed and include IMPUTE [v1 (17) and v2 (9)], BEAGLE (3), and MACH (13). Imputation methods estimate genotypes by combining information from a reference panel (e.g., HapMap or 1,000 Genomes Project), consisting of genotypes for a dense set of SNPs, and a study sample in which the SNPs are genotyped at a subset of the reference panel SNPs. Following the notation of Howie et al. (9), SNPs can be categorized into one of two disjoint sets: the set $\mathcal{T}$ of SNPs typed in both the study sample and reference panel, or the set $\mathcal{U}$ of SNPs untyped in the study sample but typed in the reference panel. In most imputation methods, the best match is searched for between the resulting haplotypes from phasing the SNPs in $\mathcal{T}$ and the corresponding partial haplotypes from the reference panel. This is based on the assumption that haplotypes that match at the SNPs in $\mathcal{T}$ will also match at the SNPs in $\mathcal{U}$. Imputed genotypes may also be obtained at typed SNPs by removing the information for the SNP of interest and using only the remaining SNPs to estimate the genotypes at the SNP for every individual. This probability distribution is (should be) then taken into account for association analysis. In IMPUTE v2 (9), multiple reference panels genotyped on different sets of SNPs can be used in the imputation so that the reference panel is effectively increased. An expanded panel is shown to improve the imputation of rare SNPs rather than using only a single reference set.

By increasing the set of SNPs for an association analysis, imputation increases the power for detecting associations with disease. In comparison to an association analysis based on a

tagSNP design, there is a clear increase in power by including imputed causal SNPs that were not tagged, with the largest increases occurring for variants with MAF < 0.05 (17). In the case of common causal variants, similar powers are achieved by an imputation-extended tagSNP approach and by resequencing all individuals (30), suggesting that imputation is a cost-effective alternative to resequencing in this case. However, for rare variants, despite the increase in power by adding in imputed SNPs to the analysis, this does not attain a power as high as when complete resequencing data are used (30). Care also needs to be taken to ensure that an appropriate genetically similar reference panel is used.

Imputation clearly has its advantages, but the low MAF of rare variants and their low LD with other variants make it difficult to impute them. A related issue in imputing rare variants is the difficulty in assessing imputation accuracy. Imputation accuracy is typically defined by the proportion of correctly classified genotypes (15), or equivalently by the discordance between imputed and observed genotype calls (9). Using this measure of accuracy, there is a higher confidence in making calls at a rare SNP than at a common one, giving the impression of higher accuracy than at common SNPs because for most rare SNPs the genotype will be homozygous for the common allele. By this same reasoning, rare SNPs will have a smaller proportion of missing genotypes than those of higher frequency (9). Even by randomly assigning the two alleles of a rare SNP to a sample, using only the MAF (<5%), an apparent accuracy greater than 90% can be achieved if only the concordance is considered (15). This indicates that alternate accuracy measures are needed for imputation at rare SNPs.

In evaluating imputation methods, Bryan et al. (9) also examine the minor allele calls at rare SNPs (MAF < 5%) in terms of false positives (heterozygous call when homozygous common) and false negatives (homozygous common call when heterozygous). In doing so, they find that IMPUTE v2 generally has higher accuracy at rare SNPs than competing methods. An imputation quality score (IQS) based on Cohen's kappa statistic for inter-rater agreement adjusts the observed proportion of agreement by the agreement that would occur simply due to chance (15). An IQS of 1 indicates a perfect match, and negative values occur when the imputation method performed worse than a random genotype assignment. In a comparison of the concordance proportion and IQS for imputations (using IMPUTE v1) of European Americans (EA) using the CEU reference panel of HapMap and for African Americans (AA) using the YRI reference panel, the authors illustrate that IQS is a better measure of imputation accuracy. The mean concordance proportion for AA (97.1) is almost identical to EA (98.8), whereas the difference is much larger between the mean IQSs, with AA (78.3) having a much lower average score than EA (90.2), which reflects the fact that it is more difficult to impute African populations because of their low LD structure. In a plot of the relationship between MAF and the two imputation accuracy measures, the concordance measure increases as MAF decreases, whereas the IQS drops as the MAF decreases, which is the expected relationship since rare SNPs do not impute well.

When testing for associations using imputed genotypes, there is a higher uncertainty than in experimental genotype calls, and the imputation accuracy needs to be accounted for. Methods for incorporating this uncertainty have been developed for single SNP association testing, e.g., implemented in the software SNPTEST (17). For rare variants, a different strategy is required because of the low power single SNP tests have in detecting disease-associated rare variants.

Genotype imputation can also be used to facilitate meta-analysis so that there are common SNPs in the study samples to be combined. When meta-analyzing across different populations, differences in LD patterns, as well as directions of association at the individual SNPs, become issues. If an untyped functional polymorphism is in strong LD with a typed SNP in one population but not in the others, the

meta-analysis will not be effective to identify the association. One way of quantifying how different the LD patterns are between populations is varLD (32). For rare variant signal meta-analysis, summary statistics can be combined at the locus rather than at the SNP level, thus alleviating problems associated with allelic heterogeneity. However, as effect size estimates within individual strata can be unstable due to low rare allele numbers, p-value–based meta-analysis may be preferable. The latter can also overcome direction of effect differences across studies.

In rare variant analyses, the need for stringent quality control procedures is highlighted by the loss of power due to genotype misspecification, which causes a rise in the false-positive error rate. Rare variants are difficult to genotype and challenging to impute, so meticulous quality checks are essential before declaring association. The field of complex trait genetics is undergoing a shift in focus from common to rare variants, primarily driven by advances in next generation sequencing. As with GWA studies a few years ago, technology has once again outstripped analytical capacity. Streamlined, powerful rare variant analysis methods that take diverse study designs under account (e.g., using family-based samples, samples drawn from the extremes of a distribution, etc.) are urgently needed. Finally, improved annotation of the human genome will undoubtedly enable the interpretation of rare variant signals and will help move the field forward, enhancing our understanding of complex trait genetics.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

JA and EZ are supported by the Wellcome Trust (WT088885/Z/09/Z).

LITERATURE CITED

1. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57:289–300
2. Bodmer W, Bonillna C. 2008. Common and rare variants in multifactorial susceptibility to common diseases. *Nat. Genet.* 40:695–701
3. Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* 81:1084–97
4. Deleted in proof.
5. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. 2004. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 305:869–72
6. Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, et al. 2006. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc. Natl. Acad. Sci. USA* 103:1810–15
7. Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amos CI. 2008. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am. J. Hum. Genet.* 82:100–12
8. Han F, Pan W. 2010. A data-adaptive sum test for disease association with multiple common or rare variants. *Hum. Hered.* 70:42–54
9. Howie BN, Donnelly P, Marchini J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 5:e1000529
10. Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, et al. 2008. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat. Genet.* 40:592–99

11. Li B, Leal SM. 2008. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.* 83:311–21
12. Li B, Leal SM. 2009. Discovery of rare variants via sequencing: implications for the design of complex trait association studies. *PLoS Genet.* 5(5):e1000481. doi:10.1371/journal.pgen.1000481
13. Li Y, Abecasis GR. 2006. Mach 1.0: rapid haplotype reconstruction and missing genotype inference. *Am. J. Hum. Genet.* S79:2290
14. Lichtarge O, Bourne HR, Cohen FE. 1996. An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* 257:342–58
15. Lin P, Hartz SM, Zhang Z, Saccone SF, Wang J, et al. 2010. A new statistic to evaluate imputation reliability. *PLoS ONE* 5:e9697
16. Madsen BE, Browning SR. 2009. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.* 5(2):e1000384. doi:10.1371/journal.pgen.1000384
17. Marchini J, Howie B, Myers S, McVean G, Donnelly P. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* 39:906–13
18. Deleted in proof.
19. Morgenthaler S, Thilly WG. 2007. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutat. Res.* 615:28–56
20. Morris AP, Zeggini E. 2010. An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet. Epidemiol.* 34:188–93
21. Morris RW, Kaplan NL. 2002. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. *Genet. Epidemiol.* 23:221–33
22. Mukhopadhyay I, Feingold E, Weeks DE, Thalamuthu A. 2010. Association tests using kernel-based measures of multi-locus genotype similarity between individuals. *Genet. Epidemiol.* 34:213–21
23. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. 2009. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 324:387–89
24. Ng PC, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31:3812–14
25. Pritchard JK, Cox NJ. 2002. The allelic architecture of human disease genes: common disease–common variant . . . or not? *Hum. Mol. Genet.* 11:2417–23
26. Ramensky V, Bork P, Sunyaev S. 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* 30:3894–900
27. Schaid DJ, McDonnell SK, Hebbring SJ, Cunningham JM, Thibodeau SN. 2005. Nonparametric tests of association of multiple genes with human disease. *Am. J. Hum. Genet.* 76:780–93
28. Schork NJ, Murray SS, Frazer KA, Topol EJ. 2009. Common vs rare allele hypotheses for complex diseases. *Curr. Opin. Genet. Dev.* 19:212–19
29. Service RF. 2006. Gene sequencing: the race for the $1,000 genome. *Science* 311:1544–46
30. Servin B, Stephens M. 2007. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet.* 3(7):e114
31. Slager SL, Schaid DJ. 2001. Case-control studies of genetic markers: power and sample size approximations for armitage's test for trend. *Hum. Hered.* 52:149–53
32. Teo YY, Fry AE, Bhattacharya K, Small KS, Kwiatkowski DP, Clark TG. 2009. Genome-wide comparisons of variation in linkage disequilibrium. *Genome Res.* 19:1849–60
33. Xiong M, Zhao J, Boerwinkle E. 2002. Generalized T^2 test for genome association studies. *Am. J. Hum. Genet.* 70:1257–68
34. Van Liere JM, Rosenberg NA. 2008. Mathematical properties of the r^2 measure of linkage disequilibrium. *Theor. Popul. Biol.* 74:130–37
35. Wang T, Elston RC. 2007. Improved power by use of a weighted score test for linkage disequilibrium mapping. *Am. J. Hum. Genet.* 80:353–60
36. Wessel J, Schork NJ. 2006. Generalized genomic distance-based regression methodology for multilocus association analysis. *Am. J. Hum. Genet.* 79:792–806
37. Zeggini E, Rayner W, Morris AP, Hattersley AT, Walker M, et al. 2005. An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets. *Nat. Genet.* 37:1320–22

Annual Review of Genetics

Volume 44, 2010

Contents

Errata

An online log of corrections to *Annual Review of Genetics* articles may be found at http://genet.annualreviews.org/errata.shtml