European Journal of Human Genetics (2013) 21, 680–686

www.nature.com/ejhg

ARTICLE

Statistical tests for detecting associations with groups of genetic variants: generalization, evaluation, and implementation

John Ferguson[1], William Wheeler[2], YiPing Fu[3], Ludmila Prokunina-Olsson[3], Hongyu Zhao[1] and Joshua Sampson*[,4]

With recent advances in sequencing, genotyping arrays, and imputation, GWAS now aim to identify associations with rare and uncommon genetic variants. Here, we describe and evaluate a class of statistics, generalized score statistics (GSS), that can test for an association between a group of genetic variants and a phenotype. GSS are a simple weighted sum of single-variant statistics and their cross-products. We show that the majority of statistics currently used to detect associations with rare variants are equivalent to choosing a specific set of weights within this framework. We then evaluate the power of various weighting schemes as a function of variant characteristics, such as MAF, the proportion associated with the phenotype, and the direction of effect. Ultimately, we find that two classical tests are robust and powerful, but details are provided as to when other GSS may perform favorably. The software package CRaVe is available at our website (http://dceg.cancer.gov/bb/tools/crave).
European Journal of Human Genetics (2013) **21,** 680–686; doi:10.1038/ejhg.2012.220; published online 24 October 2012

Keywords: rare variants; score test; GWAS; association test

INTRODUCTION

The search for rare variants associated with common diseases, and traits in general, has already started.[1–5] As these variants are rare, most studies will be inadequately powered to detect an association with any single variant.[6–8] When only a handful of minor alleles are observed for any single-nucleotide variant (SNV), obtaining statistical significance, especially at traditional genome wide levels of 10^{-8}, can be near impossible. Therefore, instead of trying to identify associations with individual variants, the goal has been to identify associations with a group of rare variants in a shared region (eg, exons, genes) or pathway, effectively increasing power by pooling information across SNVs.[9] Numerous statistical tests that can search for these regional associations have already been introduced, developed, and compared.[7,10–19]

Our three goals in this paper are to (1) Unify (2) Identify, and (3) Modify association tests for rare variants. First, we introduce a simple statistical framework and show that the majority of rare-variant association tests can be reformulated within this framework. Second, we show that within this framework, we can easily identify the relationship between a statistic's performance and the genetic characteristics of the tested SNVs, such as the proportion of SNVs associated with an outcome, direction of effects, and the relationship between effect size and MAF. Third, we show that the standard test statistics can be further tailored to the specifics of a given study or an investigator's prior beliefs.

To achieve our objective of unification, we first revisit the standard association test for a single uncommon SNV. The standard approach would be to divide study participants into two groups, those with and without the minor allele, and then measure the difference in the average phenotype between those two groups. This approach applies equally for continuous and dichotomous variables (eg, disease), where for the latter case, we would measure the difference in disease prevalence. The resulting difference, at least for many scenarios, completely captures the available information for detecting an association. When testing a group of SNVs, the relevant information from each SNV is still only this difference, and, therefore, joint tests only vary by how they combine this information.

Most test statistics combine these differences in a very specific way. For describing the unified framework, let individual i be in the study, Y_i be the phenotype (eg, weight, height, and disease status) and let G_{ij} be the number of rare variants at SNV j. Therefore, for SNV j, we can calculate the difference between the average of the phenotype values in subjects with a minor allele, $\bar{Y}_{G_j \geq 1}$, and the average in subjects without a minor allele, $\bar{Y}_{G_j = 0}$.

$$\Delta_j = \bar{Y}_{G_j \geq 1} - \bar{Y}_{G_j = 0} \quad (1)$$

The majority of statistics proposed for testing associations are a weighted sum of the squared differences, Δ_j^2, and their cross-products, $\Delta_{j_1}\Delta_{j_2}$. As the weights are allowed to depend on Δ, all statistics cannot necessarily be formulated as a second-degree polynomial. We refer to this broader class of statistics as Data-Adaptively Weighted Generalized Score Statistics (DAWGSS, pronounced *dogs*). The full definition of DAWGSS, which can accommodate covariates and population stratification, will be provided later.

Standard rare-variant tests, such as the Sum test,[11] Hotelling's T^2-test,[20] Stouffer's Z-test,[15] Data Adaptive Sum test,[21] C-alpha,[16]

[1]Division of Biostatistics, Yale School of Public Health, New Haven, CT, USA; [2]Information Management Services, Rockville, MD, USA; [3]Laboratory of Translational Genomics, DCEG, National Cancer Institute, Bethesda, MD, USA; [4]Biostatistics Branch, DCEG, National Cancer Institute, Rockville, MD, USA
*Correspondence: Dr J Sampson, Biostatistics Branch, DCEG, National Cancer Institute, 6120 Executive Boulevard, 8038, Rockville, MD 20852, USA. Tel: +1 301 443 8207; Fax: +1 301 402 0081; E-mail: joshua.sampson@nih.gov
Received 26 January 2012; revised 31 May 2012; accepted 21 July 2012; published online 24 October 2012

similarity regression,[17] variance components,[17] CMC,[22] and SKAT,[23] to name a few, only vary by their chosen set of weights. We specify five global features of the weights that vary among common test statistics. We then identify the properties of these global features that are desirable, or provide high power, when genetic variants have certain behaviors. For example, one feature is the relationship between the weights and the magnitude of Δ_j. We show that the power of association tests can be increased by shrinking the weights when Δ_j is small if only a minority of rare variants are truly associated with the outcome. By observing this connection, we develop a new test statistic, ROVER, that performs well in such scenarios.

The idea of a general framework for rare-variant tests has been proposed previously.[12,13] Our objective is to propose a similar framework, but one where a rare-variant test can be recast as a function of the single SNV test statistics. This framework illuminates similarities with classical statistical tests, offers a clear means for performing meta-analyses when studies genotype different sets of SNVs, and facilitates interpretable modifications.

We use these test statistics to better understand the association between bladder cancer and the *UGT1A* gene locus on chromosome 2q37. UGTs facilitate cellular detoxification of multiple exogenous and endogenous substrates.[24] Specifically, UGTs participate in the removal of aromatic amines, which are the main risk factors for bladder cancer found in tobacco smoke and industrial chemicals.[24] This locus has been associated with colorectal cancer,[25,26] pancreatic cancer,[27] liver cancer,[28] and most recently bladder cancer.[29] Following-up the GWAS hit, rs1189203, for bladder cancer, we performed targeted resequencing to identify variants within the UGT1A region and then a focused association study in 9319 individuals.[30] After the study identified one variant that was highly associated with bladder cancer, we now use ROVER and the other test statistics described here to determine if the remaining variants, many with low MAF, were enriched for associations.

In the methods section, we provide a more complete definition of DAWGSS and provide the details about our simulations and the study of bladder cancer. In the results section, we compare the performance of these statistics for our two types of data. Importantly, we compare the statistics within their DAWGSS framework so we can demonstrate how the choice of weights determines the properties of the test statistics. Note, the common framework allows rapid computation of multiple-test statistics simultaneously, enabling our broad comparison. Our software, CRaVe, is available as both a stand-alone Unix program and an *R* function. In the discussion section, we summarize our conclusions. The Supplementary material shows how to reformulate many commonly used test statistics in their DAWGSS format.

METHODS

DAWGSS: definition

We let n be the number of subjects, Y_i be the normalized phenotype ($\sum_i Y_i = 0$, $\frac{1}{n}\sum_i Y_i^2 = 1$), and G_{ij} be the number of rare variants at SNV j, $j \in \{1,\ldots,J\}$. We denote the vector of all genotypes by $G_i = [G_{i1}\ G_{i2} \ldots G_{iJ}]^t$.

We then define the genetic covariance, V, and correlation, σ, matrices from a group, S, containing n_S individuals by

$$V = \frac{1}{n_S}\sum_{i\in S}(G_{i\cdot} - \hat{\mu}_G)(G_{i\cdot} - \hat{\mu}_G)^t \tag{2}$$

$$\sigma = D(V)^{-\frac{1}{2}} V D(V)^{-\frac{1}{2}} \tag{3}$$

where $\hat{\mu}_G = [\hat{\mu}_{G1}\ \hat{\mu}_{G2}\ \ldots\ \hat{\mu}_{GJ}]^t$, $\hat{\mu}_{Gj} = \frac{1}{n_S}\sum_{i\in S} G_{ij}$, and D(V) is the matrix containing only the diagonal elements of V, with the off-diagonal elements set to 0. We will denote the element in row j_1, column j_2 of V and σ by $V_{j_1 j_2}$ and $\sigma_{j_1 j_2}$, respectively. Furthermore, let σ^{-1} be the inverse of σ so that $\sigma\sigma^{-1} = I$, where I is the $J \times J$ identity matrix.

Our definition of Δ_j in the introduction was slightly simplified, in that it only allowed two genotypes. We redefine it here as the score statistic (or, equivalently, as the correlation between Y and $G_{\cdot j}$ multiplied by a normalizing factor of $\sqrt{n}$),

$$\Delta_j = \frac{1}{\sqrt{n}}\frac{\sum_i Y_i G_{ij}}{\sqrt{V_{jj}}} \tag{4}$$

and let $\Delta = [\Delta_1\ \Delta_2 \ldots \Delta_J]^t$. Note, for any SNV, Δ_j is asymptotically distributed as a normal variable with mean 0 and variance 1 under the null hypothesis.

As stated in the introduction, DAWGSS have the form

$$\sum_j w_j \Delta_j^2 + \sum_{j_1, j_2} w_{j_1 j_2} \Delta_{j_1} \Delta_{j_2} \tag{5}$$

When there are covariates, we extend the definition of Δ_j as follows. Let us expand our notation. For subject i, let $Y_i^\dagger$ be the outcome and $X_i = [X_{i1}\ X_{i2} \ldots X_{iT}]^t$ be a set of T-covariates. We estimate $\mu_i \equiv E[Y_i^\dagger \mid X_{i\cdot}]$ by either linear or logistic regression as appropriate. We define $Y_i^\ddagger \equiv Y_i^\dagger - \hat{\mu}_i$, $\mu_Y^\ddagger = \frac{1}{n}\sum_{i=1}^n Y_i^\ddagger$ and $\sigma_Y^{2\ddagger} = \frac{1}{n}\sum_{i=1}^n (Y_i^\ddagger - \mu_Y^\ddagger)^2$. Usually the values of $\hat{\mu}_i$ are defined so $\mu_Y^\ddagger = 0$ and can be ignored.

$$Y_i \equiv \frac{Y_i^\ddagger - \mu_Y^\ddagger}{\sqrt{\sigma_Y^{2\ddagger}}} \tag{6}$$

We can then use this redefined Y_i in equation (4).

DAWGSS: global features

Within the DAWGSS framework, most tests can be distinguished by five global features.

1. The magnitude of $w_{j_1 j_2}$ for uncorrelated SNVs (ie, are cross-product terms included when SNVs are in linkage equilibrium?)
2. The relationship between w_j and minor allele frequency (ie, are the contributions of SNVs weighted by their MAF?)
3. The relationship between the weights and the signs of Δ (ie, is preference given to variants that appear harmful?)
4. The relationship between w_j and Δ_j (ie, are small Δ_js shrunk on the presumption that most SNVs have no association?)
5. The relationship between $w_{j_1 j_2}$ and linkage disequilibrium (ie, are independent and correlated SNVs treated equally?)

The results section will discuss how the behavior of rare variants determines the desirable properties of features 1–4. Feature 5 is listed here for completeness, but will only be addressed briefly in the discussion.

Statistics

We compare the performance of multiple-test statistics, including

- SUM: Sum test[11]
- HOI: Hotelling's T^2-test assuming independence[20]
- STZ: Stouffer's Z-test[15]
- MDF: MultiDegree of Freedom test[16,17]

These four tests are fully defined in Table 1. Note that HOI sets σ to be the identity matrix, allowing for more direct comparisons with the other tests, which omit any reference to σ. We focus on these four tests because they highlight the importance of the first two global features and because they are the most ubiquitous, noting that the recommended versions of tests using C-alpha, similarity regression, variance components, and kernels are all equivalent to MDF. However, for the larger simulations and bladder cancer

Table 1 Four test statistics – MultiDegree of Freedom Test (MDF), modified Hotelling's *T*-test (HOI), Stouffer's *Z*-Test (STZ), and the Sum test differ only by whether the statistics include cross-product terms, $\Delta_{j_1}\Delta_{j_2}$, and whether the weights incorporate $\sqrt{V}$, where V_{jj} is the variance of the genotype (ie, $\propto$MAF(1 −MAF))

$\sqrt{V}$ \ $\Delta_{j_1}\Delta_{j_2}$	Yes	No
Yes	SUM $\sum_j V_{jj}\Delta_j^2+\sum_{j_1\neq j_2}\sqrt{V_{j_1j_1}V_{j_2j_2}}\Delta_{j_1}\Delta_{j_2}$	MDF $\sum_j V_{jj}\Delta_j^2$
No	STZ $\sum_j \Delta_j^2+\sum_{j_1\neq j_2}\Delta_{j_1}\Delta_{j_2}$	HOI $\sum_j \Delta_j^2$

The independent version of Hotelling's *T*-test (HOI) sets the covariance matrix in the standard statistic to the identity matrix, allowing for more direct comparisons with the other tests. MDF has also been referred to as the *C-alpha*, similarity regression, and variance components tests.

data, we show the results for six other tests that can be described within the DAWGSS framework.

- HOT: Hotelling's T^2-test

HOT accounts for the correlation between SNVs,

$$\mathrm{HOT}=\sum_j \sigma^{-1}_{j,j}\Delta_j^2+\sum_{j_1\neq j_2}\sigma^{-1}_{j_1,j_2}\Delta_{j_1}\Delta_{j2} \quad (7)$$

- STZ+: Positive Stouffer's *Z*-test
- MDF+: Positive MultiDegree of Freedom test

STZ+ [31] and MDF+ tests modify their original test statistics by including only those variants with a positive Δ_j. We define the binary function, $1(\cdot)$ by $1(\cdot)=1$ if the enclosed statement is true, 0 otherwise.

$$\mathrm{STZ}+=\sum_j 1(\Delta_j>0)\Delta_j^2+\sum_{j_1\neq j_2}1(\Delta_{j_1},\Delta_{j_2}>0)\Delta_{j_1}\Delta_{j_2} \quad (8)$$

$$\mathrm{MDF}+=\sum_j 1(\Delta_j>0)V_{jj}\Delta_j^2 \quad (9)$$

- THR: Threshold test

THR is designed to reflect the statistic from Hoffmann *et al*,[12] only including variants with $|\Delta_j|>\Phi^{-1}(1-\frac{\alpha}{2})$. In practice, $\alpha=0.05$.

$$\mathrm{THR}=\sum_j 1(|\Delta_j|>\Phi^{-1}(1-\frac{\alpha}{2}))\Delta_j^2+\sum_{j_1\neq j_2}1(|\Delta_{j_1}|>\Phi^{-1}(1-\frac{\alpha}{2}),|\Delta_{j_2}|>\Phi^{-1}(1-\frac{\alpha}{2}))\sigma^{-1}_{j_1,j_2}\Delta_{j_1}\Delta_{j_2} \quad (10)$$

- DAS: Data Adaptive Sum test

DAS, defined by Han and Pan,[21] also requires the definition of a threshold, which we similarly set at $\alpha_0=0.05$.

$$\mathrm{DAS}=\sum_j V_{jj}\Delta_j^2+\max\Big(\sum_{j_1\neq j_2}w_{j_1}w_{j_2}\Delta_{j_1}\Delta_{j_2},\sum_{j_1\neq j_2}W_{j_1}W_{j_2}\Delta_{j_1}\Delta_{j_2}\Big) \quad (11)$$

Table 2 Summary of the global features

Test	$\Delta_{j_1}\Delta_{j_2}$	V_{jj}	*LD*	*Dir*	*DAW*
SUM[11]	•	•		*s*	
HOI[20]					
STZ[15]	•			*s*	
MDF[17]		•			
HOT[20]			•		
STZ+[31]	•			*p*	•
MDF+		•		*p*	•
THR					•
DAS[21]	•	•			•
ROV					•

A • in column 1, $\Delta_{j_1}\Delta_{j_2}$, indicates the presence of a cross-product term for independent SNVs. A • in column 2, V_{jj}, indicates that the weights increase with MAF. A • in column 3, LD, indicates linkage disequilibrium alters $w_{j_1j_2}$. A 's' in column 4, Dir, indicates the statistic increases when Δ_js have the *same* sign and '*p*' indicates the statistic increases when Δ_js are *positive*. A • in column 5, DAW, indicate the weights depend on Δ_j.

where

$$w_j=\sqrt{V_{jj}}(1-2\times 1(\Delta_j<\Phi^{-1}(\frac{\alpha_0}{2}))) \quad (12)$$

$$W_j=\sqrt{V_{jj}}(1-2\times 1(\Delta_j>\Phi^{-1}(1-\frac{\alpha_0}{2}))) \quad (13)$$

- ROV: Rover

ROVER shrinks the weights for SNVs with low signal

$$\mathrm{ROV}=\sum_j(1-\exp(-0.2\Delta_j^2))\Delta_j^2 \quad (14)$$

Table 2 summarizes the global features for each of these ten statistics, with Supplementary tables summarizing additional statistics.

Simulations

Independence. We simulated genes from a case/control study with a total of $n=2000$ individuals, equally divided between the two groups, under the null and multiple alternative hypotheses. For each simulated gene, containing 40-independent SNVs with MAF equally spaced between 0.005 and 0.02, we calculated values for six-test statistics: SUM, HOI, STZ, MDF, MDF+, and ROV. Genes simulated under the null hypothesis were used to estimate the threshold for rejection, given a particular significance level, α. Power was then estimated as the proportion of genes simulated under an alternative, which exceeded that threshold. Under the alternative, the relationship between the probability of disease and genotype was defined by

$$\mathrm{logit}(P(Y_i=1\mid G_{i\cdot}))=\beta_0+\sum_j\beta_jG_{ij} \quad (15)$$

The total number of influential SNVs ($N_I=\sum_j 1(\beta_j\neq 0)$), the proportion of influential SNVs that reduced risk ($\sum_j 1(\beta_j<0)/N_I$), and the log odds ratio (β_j) were varied. The values of β_j, constant across all SNVs, were chosen by one of two methods. In the main text, we primarily discuss the example where $\beta_j=\exp(2/N_I^{0.6})$. In the Supplementary material, we discuss different effect sizes.

Linkage disequilibrium. We simulated genes from a case/control study with a total of 2000 or 10 000 individuals under the null and alternative hypotheses. Here, haplotypes were chosen to mimic those observed in the 1000 genomes project,[32] and each gene contained 80 SNVs with MAF between 0.005 and 0.10. The MAF was increased because most SNVs with MAF<0.02 appeared independent of each other, while the number of SNVs was increased because, in the presence of linkage disequilibrium (LD), one influential SNV creates multiple-associated SNVs. For each alternative hypothesis, power was defined as the proportion of simulated genes with a permutation-based *P*-value below

10^{-3}. Null simulations were only used to confirm that α-levels were correctly calibrated. Under the alternative, the relationship between the probability of disease and genotype was again defined by equation (15).

For the alternative hypotheses, simulations varied by the following parameters. (1) The number of subjects (n). (2) The number of influential SNVs (N_I). (3) The relationship between MAF and β_j. Either β_j was constant for all associated variants or inversely proportional to MAF. (4) Effect direction: either 50% or all of the associated variants increased risk ($\sum_j 1(\beta_j < 0)/N_I \in \{0, 0.5\}$). (5) The magnitude of the effect size. Although the odds ratios were generally set so the power of the MDF test was $\approx$0.4–0.5, a 'large' effect size, where the power of the MDF test was inflated to $\approx$0.8, was also examined. Details about the simulation process are available in the Supplementary material.

Bladder cancer study

The association studies have been described elsewhere, so we only offer a brief summary here.[27,30] The original GWAS contained 3532 cases and 5120 controls of self-described European descent.[29] Among the 591 637 SNVs passing quality-control, 166 SNVs were within the UGT1A region, defined as the 158 Kb of the UGT1A cluster + /100 Kb (chr2:234 091 000–234 447 000, hg18). A promising association between SNV rs11892031 in the cluster and bladder cancer ($P = 7.7 \times 10^{-5}$) suggested additional examination. In the first step of the follow-up study, we generated highly specific long-range amplicons and sequenced the alternative first exons in each of the *UGT1A* genes in 44 bladder cancer cases and 30 trios from the HapMap CEU set, detecting 43 known exonic SNVs. In the second step, we selected 18 SNVs, based on LD and functional annotation, and genotyped these in a set of 1055 cases and 962 controls from the Spanish Bladder Cancer Study (SBCS), a component of the original GWAS. In step 3, based on the SBCS data enriched across the region, we imputed these additional exonic variants for the remaining samples from the stage 1 GWAS (2477 cases/4158 controls). The final data set contained 49 exonic SNVs in 3532 cases and 5120 controls. Using imputation based on the combined reference panels of HapMap 3 CEU and 1000 Genomes data, an extended data set was created with 1170 SNVs in the same group of individuals.

We performed the previously described association tests on subgroups of the 49 and 1170 SNVs, including only SNVs with a MAF below 0.1. As rs17863783 is significant by itself, we repeated these tests on the remaining SNVs after adjusting for rs17863783, using a simplified approach of treating that SNV like a covariate. All analyses were also repeated after adjusting for the covariates age, gender, smoking status, and study center.

RESULTS

Simulations

We first compared the power to detect associations between disease status and simulated genes containing either 40-independent or 80-dependent SNVs.

(1) The magnitude of $w_{j_1 j_2}$ for uncorrelated SNVs. Statistics that include cross-product terms are more advantageous when a large proportion of SNVs are associated with the outcome. In our simulations assuming independence, the power for the SUM test, which includes cross-product terms, exceeds that for the MDF test when at least 15 of the 40 SNVs influence disease risk (Figure 1). If we increase β in equation (15), then that intersection point, when the study power becomes higher for the SUM test, increases slightly (Supplementary material). If we increase the total number of SNVs beyond 40, the absolute number of influential SNVs needed for SUM to have higher power will increase, but the proportion of SNVs, relative to the total number, decreases (Figure 3).

Our simulations where SNVs within a gene are in linkage disequilibrium show that the SUM test becomes the more powerful test at a comparatively smaller number of influential SNVs. Table 3 shows that the advantage for those tests without cross-product terms

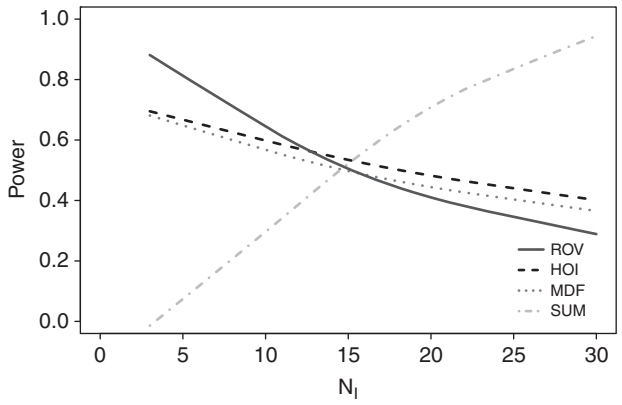

Figure 1 The power (y-axis) to detect an association between a gene with 40-independent SNPs and a disease is illustrated for four different test statistics (ROV, HOI, MDF, SUM), as a function of the number (x-axis) of those SNPs that increase disease risk.

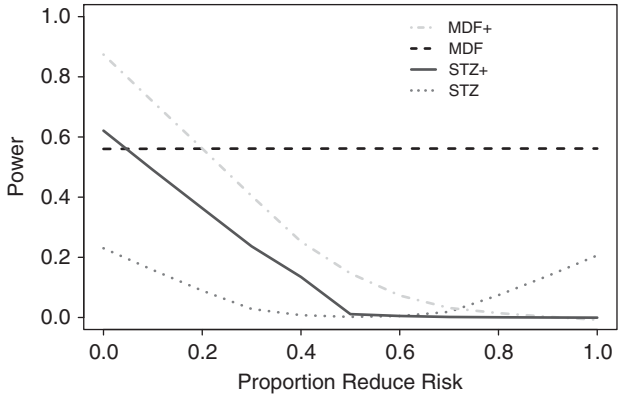

Figure 2 The power (y-axis) to detect an association between a gene with 40-independent SNPs and a disease is illustrated for four different test statistics (MDF+, MDF, STZ+, STZ), as a function of the proportion (x-axis) of the 10-associated SNPs that reduce disease risk. The remaining associated SNPs increase risk.

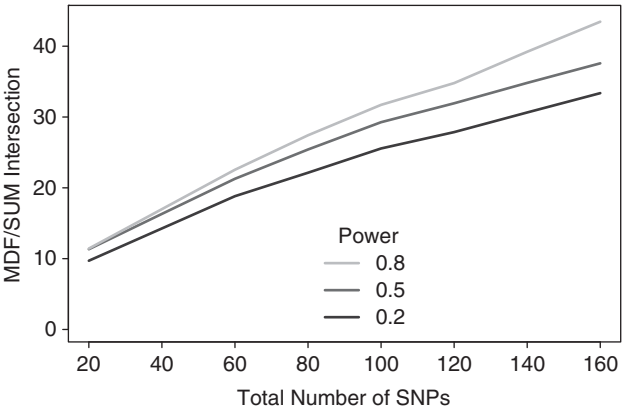

Figure 3 The lines illustrate the number of SNVs that need to be influential in order for the SUM test to have higher power than the MDF test, as a function of the total number of independent SNVs in the tested region. Results are based on simulations where SNVs have MAF equally spaced between 0.005 and 0.02, and the influential SNVs are randomly distributed, have equal effect size, and increase risk. The effect size (β in equation (15)) was chosen so the MDF test had a specified power: 0.2 (brown), 0.5 (red), or 0.8 (orange).

Table 3 The statistical tests were compared when 4 and 16 out of the 80-dependent SNVs were influential

	4 Influential SNPs					*16 Influential SNPs*				
		2000 Subjects	*1.5β*	*β∝1/MAF*	*50% Protective*		*2000 Subjects*	*1.5β*	*β∝1/MAF*	*50% Protective*
ROV	0.88	0.83	0.93	0.95	1	0.71	0.69	0.78	0.68	0.92
HOI	0.87	0.83	0.91	0.96	0.98	0.73	0.70	0.81	0.70	0.96
MDF	0.87	0.85	0.92	0.87	0.96	0.74	0.73	0.81	0.65	1
SUM	0.67	0.65	0.70	0.69	0.21	0.90	0.89	0.92	0.86	0.21
STZ	0.65	0.60	0.67	0.69	0.19	0.88	0.84	0.90	0.89	0.18
HOT	0.66	0.53	0.88	0.83	0.97	0.36	0.30	0.49	0.34	0.79
DAS	0.72	0.69	0.78	0.76	0.72	0.79	0.76	0.85	0.74	0.81
THR	0.78	0.75	0.81	0.85	0.81	0.72	0.70	0.77	0.68	0.83
MDF+	1	1	1	1	0.55	0.94	0.96	0.96	0.84	0.53
STZ+	0.86	0.83	0.86	0.93	0.29	1	1	1	1	0.29

Each column shows the relative power under a different scenario. Relative power is the actual power divided by the maximum power for that column or scenario. For the first column, the simulated study assumed 10 000 subjects, effect size provided power near 0.40 for MDF, effect size was independent of MAF, and all influential SNVs were harmful. For columns 2–5, one parameter was changed, and either only 2000 individuals were included in the study, effect size was increased by 50%, effect size was inversely proportional to MAF, or 50% of the SNVs were protective. Test definitions for the Sum test (SUM), Hotelling's T^2-test (HOT), an independent Hotelling's T^2-test (HOI), Stouffer's *Z*-test (STZ), the MultiDegree of Freedom test (MDF), Positive STZ (STZ+), Positive MDF (MDF+), Data Adaptive Sum test (DAS), Threshold test (THR), ROVER (ROV) are provided in the Methods section.

is only slight when 4 out of the 80 SNVs are influential. Because of linkage disequilibrium, the number of SNVs associated with the disease is considerably larger than four. Even after varying other parameters, such as the number of subjects in the study, the odds ratio for the disease alleles, and the relationship between MAF and the magnitude of the odds ratio, these general trends still hold.

Until now, all SNVs have increased disease risk. In simulations where the strength of the association is identical among all influential SNVs, but 50% reduce disease risk, the two tests that include cross-product terms, SUM and STZ, will have nearly no power to detect an association. The cost can be understood by either noting that the cross-product terms, $\Delta_{j_1}\Delta_{j_2}$, are negative when the effects go in opposite directions, thereby reducing the test statistic, or by rewriting the test statistic as the square of a variable with an expected value of 0 (eg, STZ, $(\sum_j \Delta_j)^2$).

(2) The relationship between w_j and MAF. The relationship between the weights and MAF is mediated by $\sqrt{V_{jj}}$, where V_{jj} is the variance of the genotype at SNV *j* and is proportional to MAF (1-MAF). Clearly, compared with the STZ and HOI statistics, the statistics that include $\sqrt{V_{jj}}$ are upweighting SNVs with a higher MAF. In our simulations where SNVs are independent, the HOI statistic, which omits $\sqrt{V_{jj}}$ from the weights, slightly outperforms the MDF test after adding a small stabilizing constant, 0.005, to the denominator of Δ_j (Figure 1). Whereas the magnitude of Δ is effectively independent of MAF, multiplying by $\sqrt{V_{jj}}$ makes the contributions proportional to the number of rare variants: $\sqrt{V_{jj}}\Delta_j \approx \sum Y_i G_{ij}$ (see equation (4)). In our simulations where SNVs are in linkage disequilibrium, the two statistics, HOI and MDF, performed nearly identically. When we allow the odds ratio (ie β_j in equation (15)) to be inversely proportional to MAF, reflecting the hypothesis that low MAF SNVs will have larger effects, the loss of power by including V_{jj} must, therefore, be larger (Table 3).

To explore the data-adaptive weights, we consider statistics introduced more recently. The STZ+ [31] and MDF+ tests modify their original test statistics by including only those variants with a positive Δ_j. These statistics were designed to gain power when mutations increase disease risk. ROV includes weights, w_j, that decrease with Δ_j, $w_j = (1 - \exp(-0.2\Delta_j^2))$, and was designed to gain power when only a small proportion of SNVs are influential.

(3) The relationship between the weights and the signs of Δ. All simulations show that without exception, STZ+ and MDF+ must outperform their all-inclusive counterparts when all SNVs increase the risk of disease (Figure 2, Table 3). Figure 2 shows, in our example where 10 out of 40-independent SNVs are influential, that the power for STZ+ and MDF+ decrease quickly as the proportion of SNVs, which reduce risk increases. If three SNVs reduce disease risk, then MDF and MDF+ statistics perform similarly, whereas if more than three SNVs reduce risk, the MDF test outperforms MDF+. Comparing STZ and STZ+ tests confirms that, in general, STZ+ will outperform STZ because as we discussed in section 1, STZ already performs poorly when SNVs have opposing effects. Clearly, when all variants are protective, the power for the MDF+ and STZ+ tests are no larger than the α-level. Table 3 shows that trends observed in the independent simulations still hold when there is LD. Although the MDF+ and STZ+ have the highest, or nearly the highest, power when all 4 or 16 influential SNVs increase risk, the power for either the MDF+ or STZ+ is only $\approx$10–20% higher than their counterparts. Table 3 further shows that when half of the SNVs reduce risk, their power is $\sim$50% of that for the MDF test.

(4) The relationship between w_j and Δ_j. In our simulations with 40-independent SNVs, Figure 1 shows that ROV outperformed all other statistics so long as fewer than $\sim$10 of those SNVs were influential. The exact intersection depends on odds ratios of the influential SNVs (see Supplementary material). In our simulations with linkage disequilibrium, ROV only outperforms other statistics when <4 out of the 80 SNVs are influential (Table 3). With linkage disequilibrium, non-influential SNVs will still be associated with the outcome and shrinkage is less helpful (Table 3). Table 3, however, shows that ROV, a form of soft-thresholding, may still slightly outperform methods that remove all Δ with a *P*-value below a hard threshold of 0.05 (ie, DAS and THR).

Bladder cancer

Among the sets of 49 exonic and 1170 total SNVs covering the UGT1A region, 24 and 556 had MAF below 0.1. In the fine-mapping study, which looked only at SNVs individually, the variant at rs17863783 was found to decrease the risk of bladder cancer ($P = 5 \times 10^{-7}$). Table 4 shows that had we not examined each SNV individually and only examined the region as a whole, no test would

Table 4 Testing the association between the UGT1A region and bladder cancer with multiple DAWGSS

	24 SNVs					*556 SNVs*				
	All		*w/o rs17863783*			*All*		*w/o rs17863783*		
	—	*Cov*	—	*SNV*	*SNV/Cov*	—	*Cov*	—	*SNV*	*SNV/Cov*
ROV	0.0054	0.0027	0.015	0.13	0.10	0.00013	0.00011	0.00016	0.11	0.17
HOI	0.0051	0.0027	0.015	0.12	0.10	0.00010	0.00012	0.00015	0.11	0.16
MDF	0.028	0.042	0.04	0.13	0.16	0.0032	0.0069	0.0034	0.073	0.11
SUM	0.69	0.88	0.51	0.7	0.99	0.77	0.81	0.76	0.54	0.83
STZ	0.83	0.31	0.85	0.93	0.68	0.62	0.23	0.66	0.87	0.91
HOT	0.066	4.2e −05	0.062	0.22	0.16	3.1e −05	0.11	0.00088	0.00042	0.58
DAS	0.024	0.13	0.042	0.13	0.12	0.0043	0.025	0.0042	0.11	0.075
THR	0.067	0.064	0.1	0.13	0.12	0.00034	0.00033	0.00037	0.11	0.092
MDF +	0.047	0.1	0.045	0.12	0.21	0.016	0.062	0.016	0.063	0.14
STZ +	0.11	0.21	0.1	0.21	0.32	0.075	0.17	0.073	0.24	0.33

The first set of columns considers only the 24-exonic SNVs with MAF below 0.1, whereas the second set considers all 556 SNVs with MAF below 0.1 within that region. For each set of SNVs, we list five columns of *P*-values. The first column measures the association between all SNVs and bladder cancer without adjusting for covariates, the second column measures the same association adjusting for covariates, the third column measures the association between all SNVs excluding rs17863783 and bladder cancer, the fourth column measures the same association adjusting for rs17863783, and the fifth column measures the same association adjusting for rs17863783 and covariates.

have found the regional association to be statistically significant after adjusting for testing 20 000 genes and using a significance threshold of $5 \times 10^{-7} = 0.01/20\,000$. However, among all tests, ROV, HOI, and HOT resulted in the lowest *P*-values. Impressively, these statistics generally performed better when the 556 SNVs were examined jointly, as opposed to only 24.

The purpose of the joint examination was to determine whether the remaining SNVs, as a group, show an association with bladder cancer. First, all tests still suggested a possible association after removing rs17863783, because surrounding SNVs were in LD with rs17863783 (Table 4). Therefore, only a test statistic that can condition on the associated SNV can be used for this analysis. Even after conditioning on an individual's genotype at rs17863783, a few tests generally suggested an association, with, for example, the HOT and MDF having *P*-values of 0.0004 and 0.073. Here, after identifying a single gene, adjusting for multiple comparisons is unnecessary. However, by adjusting for the study center, a surrogate for population structure, tests suggested that there was no additional association between the *UGT1A* cluster and bladder cancer, with the HOT and MDF now having *P*-values of 0.58 and 0.11. Therefore, in this region, which is highly associated with multiple cancers, only a single SNV appears to directly influence the risk of bladder cancer.

DISCUSSION

We examined a class of statistics, DAWGSS, that test for an association between a group of genetic variants and a phenotype and show that the majority of statistics that are currently available for testing associations with groups of variants, despite the diversity of their original presentations, can be rewritten as DAWGSS. Even the linear kernel methods, which generalize a large subset of statistics,[23] are, in turn, generalized by DAWGSS when we restrict to their preferred metric of IBS. Within this shared framework, it is clear that the differences among these statistics are wholly encompassed by the weights multiplying the Δ_j^2 and $\Delta_{j_1}\Delta_{j_2}$ terms in the sum from equation (5). In the case of independent SNVs, the four classical and most ubiquitous statistics (STZ, SUM, HOI, MDF) differ only by whether they include cross-product terms and/or incorporate the SD, $\sqrt{V}$, of the number of variants into their weights. When comparing a broader range of statistics, we found five global features that distinguish many of the known tests. Furthermore, we demonstrated that the desirable properties for these features depend on the behavior of rare variants.

This paper was focused on showing the similarity of rare-variant test statistics, and in the Supplementary material, we show how specific tests can be reformulated as DAWGSS. However, our discussions are not intended to be inclusive. Foremost, DAWGSS are limited to describing statistics that have an additive effect across the SNVs. Therefore, rank-based approaches,[7] methods that allow for interactions,[6] methods with non-additive effects,[18] and methods that compare all subsets of SNVs[33] are outside of the DAWGSS framework. However, we believe these limitations have minimal practical implications. Currently, studies are underpowered to detect most interactions, especially among rare variants, and, if desired, DAWGSS can easily be extended to account for interaction terms. Moreover, among methods based on distance matrices, such as similarity regression, kernel-based approaches, and variance components, the additive model outperformed other options.[17,34]

The association between bladder cancer and the *UGT1A* region appears to be caused by a single genetic variant. Considering the importance of the UGT family in a number of cancers, we expected that variants, which affected one cancer would also affect others, even if to a lesser extent, resulting in multiple variants associated with disease risk. However, in our tests of association, there was no evidence to support such a hypothesis as there was no regional association after adjusting for rs17863783.

The DAWGSS framework offers a practical means for performing association tests. To facilite its use, we have provided the software, CRaVe, on the author's website that can efficiently perform all tests that can be described with the DAWGSS framework. The program inputs standard data formats (eg, vcf, tped) and can accomodate covariates and bioinformatic weights. CRaVe can perform all tests defined within this manuscript and allows the user to new tests within this framework. We are currently using this software to study the fifth global feature, the relationship between $w_{j_1 j_2}$ and LD, and aim to provide results evaluating tests, such as HOT, that adjust for LD in the near future.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The research of Dr Joshua Sampson was supported by the intramural program of the National Institute of Cancer. The research of this study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. (http://biowulf.nih.gov).

1 Ahituv N, Kavaslar N, Schackwitz W *et al*: Medical sequencing at the extremes of human body mass. *Am J Hum Genet* 2007; **80**: 779–791.
2 Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH: Sequence variations in pcsk9, low ldl, and protection against coronary heart disease. *N Engl J Med* 2006; **354**: 1264–1272.
3 Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH: Multiple rare alleles contribute to low plasma levels of hdl cholesterol. *Science* 2004; **5685**: 869–872.
4 Nejentsev S, Walker N, Riches D, Egholm M, Todd JA: Rare variants of ifih1, a gene implicated in antiviral responses, protect against type 1 diabetes. *Science* 2009; **5925**: 387–389.
5 Romeo S, Pennacchio LA, Fu Y *et al*: Population based resequencing of angptl4 uncovers variations that reduce triglycerides and increase hdl. *Nat Genet* 2007; **4**: 513–516.
6 Liu DJ, Leal SM: A novel adaptive method for the analysis of next generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. *PLoS Genet* 2010; **10**: e1001156.
7 Madsen BE, Browning SR: A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* 2009; **5**: e1000384.
8 Morgenthaler S, Thilly WG: A strategy to discover genes that carry multiallelic or monoallelic risk for common diseases: a cohort allelic sums test (cast). *Mutat Res* 2007; **12**: 28–56.
9 Wang K, Li M, Hakonarson H: Annovar: functional annotation of genetic variants from highthroughput sequencing data. *Nucleic Acids Res* 2010; **38**: e164–e164.
10 Basu S, Pan W, Shen X, Oetting WS: Multilocus association testing with penalized regression. *Genet Epidemiol* 2011; **35**: 755–765.
11 Chapman J, Whittaker J: Analysis of multiple snps in a candidate gene or region. *Genet Epidemiol*, **32**: 2008; **6**: 560–566.
12 Hoffmann TJ, Marini NJ, Witte JS: Comprehensive approach to analyzing rare genetic variants. *PLoS ONE* 2010; **11**: e13584.
13 Lin DY, Tang ZZ: A general framework for detecting disease associations with rare variants in sequencing studies. *Am J Hum Genet* 2011; **3**: 354–367.
14 Luedtke A, Powers S, Petersen A, Sitarik A, Bekmetjev A, Tintle N: Evaluating methods for the analysis of rare variants in sequence data. *BMC Proc* 2011; **5**: S119.
15 Mosteller F, Fisher RA: Questions and answers. *Am Stat* 1948; **5**: 30–31.
16 Neale BM, Rivas MA, Voight BF *et al*: Testing for an unusual distribution of rare variants. *PLoS Genet* 2011; **7**: e1001322.
17 Tzeng JY, Zhang D, Chang SM, Thomas DC, Davidian M: Genetrait similarity regression for multimarkerbased association analysis. *Biometrics* 2009; **65**: 822–832.
18 Wessel J, Schork NJ: Generalized genomic distance based regression methodology for multilocus association analysis. *Am J Hum Genet* 2006; **79**: 792–806.
19 Xu X, Tian L, Wei LJ: Combining dependent tests for linkage or association across multiple phenotypic traits. *Biostatistics* 2003; **2**: 223–229.
20 Hotelling H: The generalization of student's ratio. *Ann Math Stat* 1931; **3**: 360–378.
21 Han F, Pan W: A data adaptive sum test for disease association with multiple common or rare variants. *Hum Hered* 2010; **70**: 42–54.
22 Li B, Leal SM: Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 2008; **83**: 311–321.
23 Wu M, Lee S, Cai T, Li Y, Boehnke M, Lin X: Rare variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011; **89**: 82–93.
24 Tukey RH, Strassburg CP: Human udpglucuronosyltransferases: metabolism, expression, and disease. *Annu Rev Pharmacol Toxicol* 2000; **40**: 581–616.
25 Chan AT, Tranah GJ, Giovannucci EL, Hunter DJ, Fuchs CS: Genetic variants in the ugt1a6 enzyme, aspirin use, and the risk of colorectal adenoma. *J Natl Cancer Inst* 2005; **6**: 457–460.
26 Strassburg CP, Vogel A, Kneip S, Tukey RH, Manns MP: Polymorphisms of the human udpglucuronosyltransferase (ugt) 1a7 gene in colorectal cancer. *Gut* 2002; **60**: 851–856.
27 Ockenga J, Vogel A, Teich N, Keim V, Manns MP, Strassburg CP: Udp glucuronosyltransferase (ugt1a7) gene polymorphisms increase the risk of chronic pancreatitis and pancreatic cancer. *Gastroenterology* 2003; **7**: 1802–1808.
28 Vogel A, Kneip S, Barut A *et al*: Genetic link of hepatocellular carcinoma with polymorphisms of the udpglucuronosyltransferase ugt1a7 gene. *Gastroenterology* 2001; **121**: 1136–1144.
29 Rothman N, GarciaClosas M, Chatterjee N *et al*: A multistage genomewide association study of bladder cancer identifies multiple susceptibility loci. *Nat Genet* 2010; **11**: 978–984.
30 Tang W, Fu YP, Figueroa J *et al*: An uncommon synonymous humanspecific coding variant within the ugt1a6 gene affects mrna expression and protects from bladder cancer. *Genome Biol* 2011; **12**: 1–27.
31 IonitaLaza I, Buxbaum JD, Laird NM, Lange C: A new testing strategy to identify rare variants with either risk or protective effect on disease. *PLoS Genet* 2011; **7**: e1001289.
32 1000 Genome Consortium. A map of human genome variation from population scale sequencing. *Nature* 2010; **467**: 1061–1073.
33 Yu K, Li Q, Bergen AW *et al*: Pathway analysis by adaptive combination of P-values. *Genet Epidemiol* 2009; **33**: 700–709.
34 Mukhopadhyay I, Feingold E, Weeks DE, Thalamuthu A: Association tests using kernel-based measures of multi-locus genotype similarity between individuals. *Genet Epidemiol* 2010; **34**: 213–221.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)