Nonparametric Evaluation of Quantitative Traits in Population-Based Association Studies when the Genetic Model is Unknown

Frank Konietschke[1], Ondrej Libiger[2], Ludwig A. Hothorn[2,3]*

1 Department of Medical Statistics, University of Göttingen, Göttingen, Germany, 2 Scripps Genomic Medicine, The Scripps Research Institute, La Jolla, California, United States of America, 3 Institute of Biostatistics, Leibniz-University Hannover, Hannover, Germany

Abstract

Statistical association between a single nucleotide polymorphism (SNP) genotype and a quantitative trait in genome-wide association studies is usually assessed using a linear regression model, or, in the case of non-normally distributed trait values, using the Kruskal-Wallis test. While linear regression models assume an additive mode of inheritance via equi-distant genotype scores, Kruskal-Wallis test merely tests global differences in trait values associated with the three genotype groups. Both approaches thus exhibit suboptimal power when the underlying inheritance mode is dominant or recessive. Furthermore, these tests do not perform well in the common situations when only a few trait values are available in a rare genotype category (disbalance), or when the values associated with the three genotype categories exhibit unequal variance (variance heterogeneity). We propose a maximum test based on Marcus-type multiple contrast test for relative effect sizes. This test allows model-specific testing of either dominant, additive or recessive mode of inheritance, and it is robust against variance heterogeneity. We show how to obtain mode-specific simultaneous confidence intervals for the relative effect sizes to aid in interpreting the biological relevance of the results. Further, we discuss the use of a related all-pairwise comparisons contrast test with range preserving confidence intervals as an alternative to Kruskal-Wallis heterogeneity test. We applied the proposed maximum test to the Bogalusa Heart Study dataset, and gained a remarkable increase in the power to detect association, particularly for rare genotypes. Our simulation study also demonstrated that the proposed non-parametric tests control family-wise error rate in the presence of non-normality and variance heterogeneity contrary to the standard parametric approaches. We provide a publicly available R library nparcomp that can be used to estimate simultaneous confidence intervals or compatible multiplicity-adjusted p-values associated with the proposed maximum test.

Introduction

Genome-wide association studies involving large population-based samples have become a common strategy employed for the identification of common variants that affect a particular trait or play a role in disease. A majority of these studies involve comparing allele frequencies of di-allelic markers (e.g., SNPs) in cases and controls (see e.g., [1]). Formally, this is often accomplished via the Cochran-Armitage trend test [2] as implemented in the publicly available software PLINK [3]. Since the mode of inheritance at a given locus is often unknown, a maximum-test based on three mode-specific standardized Cochran Armitage trend tests was proposed [4].

Alternatively, continuous endpoints (i.e., quantitative traits), such as uromodulin [1] or TNFa protein [5] are commonly analyzed via a linear regression model using genotype scores $x = (0,1,2)$ adjusted for covariates [6]. In order to maintain statistical validity, this approach requires that three important assumptions be met: 1. additive mode of inheritance; 2. normally distributed errors, and; 3. homogeneous variances. However, in reality one or more of these assumptions are often violated. The underlying mode of inheritance is often unknown. In addition, the assumption of normality is violated in studies that involve pQTL data [5], continuous endpoints with outliers (e.g., [6]), ordered categorical data (e.g., [7]), or phenotypes with values below the detection limit (e.g., [5]). Although transformation of the endpoints into an approximate normal distributed variable allows the use of standard approaches in the generalized linear model, the transformation is data-dependent, i.e. the choice of log-, log+constant, Box-Cox-transformation for pQTLs [6] might result in different conclusions. In particular, the re-transformation on the original scale is not unique. Nonparametric regression models, e.g. quantile regression [6], are an interesting option, however, up to now, only available for an additive mode of inheritance.

Nonparametric approaches do not require normality. However, the often used nonparametric Kruskal-Wallis test [1],[8],[9] achieves suboptimal power when the locus is governed by a specific mode of inheritance. This occurs because it is a global test of heterogeneity in the endpoint values among the three genotype groups. It is also not robust against variance heterogeneity. Jonckheere-Tepstra test [10], an analog of the Kruskal-Wallis test for near-to-linear ordered restricted alternatives, shares many

characteristics with the Kruskal-Wallis test, while being particularly sensitive to an additive mode of inheritance. Furthermore, Kruskal-Wallis and Jonckheere-Tepstra nonparametric procedures are global testing procedures based on global ranks whose distribution is only available under the global null hypothesis. Therefore, it is not possible to compute confidence intervals for the genetic effects of interest using these approaches. In summary, none of these classic nonparametric approaches i) allow the identification of the most likely mode of inheritance via estimation of related simultaneous confidence intervals, ii) are sensitive not only to an additive mode of inheritance, or iii) are robust against variance heterogeneity. Our proposed testing procedure, on the other hand, can be extended to provide this crucial information for interpreting the biological relevance of the association results.

Recommendations given in the 'Strengthening the Reporting of Genetic Association studies' report [11] include providing estimators of an adequate effect size and their confidence intervals. For example, reporting odds ratios for additive, recessive and dominant models and their marginal confidence limits (as in e.g., [12]) provides a percentage measure of clinical relevance (distance from the lower/upper confidence limit to one, the value associated with the null hypothesis). While traditional significance testing usually deals with differences between population means, there is an increasing focus in medicine on the probability of one treatment being more successful than another on a per-individual basis [13]. The relative effect size [14]

$$p = P(X < Y) + 0.5P(X = Y) \qquad (1)$$

represents *a measure of how often a randomly chosen subject receiving treatment X will outperform a randomly chosen subject receiving treatment Y* [13], i.e. the probability that a randomly selected subject in the control reveals a smaller response value than a randomly selected subject in the treatment group. In case of ordered categorical data, p is also called *ordinal effect size measure* [15].

We describe a Behrens-Fisher version of multiple contrast test for relative effects [16],[14] based on the maximum test principle [4]. This is a purely nonparametric testing procedure that is valid when the three assumptions mentioned in the previous paragraph are not met. Furthermore, our proposed approach simultaneously tests association under the assumption of the three basic modes of inheritance, additive (add), recessive (rec) and dominant (dom) for both continuous and discontinuous distributions. We generalize the relative effect p for an adequate formulation of genetic effects, and provide multiple contrast tests and simultaneous confidence intervals, which allow the simultaneous testing of the three genetic models of inheritance.

Methods

Motivating example

A real data example with the right skewed distributed phenotype *total cholesterol* was selected from the Bogalusa Heart Study (BHS) [17]. This longitudinal study included genotype information on 525 unrelated individuals of European descent at 545,821 SNPs where twelve clinically-relevant quantitative traits were measured for each study participant. We applied the nonparametric multiple contrast test to a one-way layout for SNP *rs7738656* in the gene C6orf170/GJA1 and the phenotype *total cholesterol*, which was published in Table 1 of the original paper as significant for an unimputed SNP [18]. The jittered boxplots in Figure 1 show an unbalanced design, variance heterogeneity and a rather skewed distribution with some extreme values, particularly for the homozygote minor allele genotype group. Therefore, the question arises whether the parametric analysis in the original publication using a linear regression model assuming an additive mode of inheritance and normally distributed errors with homogeneous variances is appropriate.

Nonparametric model and genetic effects

Let aa, aA and AA denote the genotypes, where A is the high risk allele and a is any of the other alleles. For convenience, abbreviate the genotypes with $aa = 1$, $Aa = 2$, and $AA = 3$. The related data are given by X_{ik}, where $i = 1,2,3$, and k denotes the subject within genotype level i, $k = 1,\ldots,n_i$. The data X_{ik} are assumed to be independent. The total sample size is $N = \sum_{i=1}^{3} n_i$. We assume that the phenotypes X_{ik} follow an arbitrary distribution F_i, i.e.

$$X_{ik} \sim F_i, \; i = 1,2,3; \; k = 1,\ldots,n_i. \qquad (2)$$

This general model (2) does not contain any parameters that could be used to describe a difference between the distributions. Therefore, the distribution functions $F_i(x)$ are used to define purely nonparametric treatment effects on an individual basis for each genotype level by

$$p_j = \frac{1}{3}\sum_{i=1}^{3}\left[P(X_{i1} < X_{j1}) + 0.5P(X_{i1} = X_{j1})\right], j = 1,2,3. \qquad (3)$$

These effects are also called unweighted relative effects [19,20]. If $p_i < p_j$, then the values from F_i tend to be smaller than those from F_j. In case of $p_i = p_j$, none of the observations tend to be smaller or larger. Therefore, these effects can be as easily interpreted as the usual means in parametric models. Let $\mathbf{p} = (p_1, p_2, p_3)'$ denote the vector of the unweighted relative effects.

For the formulation of nonparametric genetic effects, let

$$\mathbf{C} = \begin{pmatrix} \mathbf{c}'_{dom} \\ \mathbf{c}'_{add} \\ \mathbf{c}'_{rec} \end{pmatrix} = \begin{pmatrix} -1 & \dfrac{n_2}{n_2+n_3} & \dfrac{n_3}{n_2+n_3} \\ -1 & 0 & 1 \\ -\dfrac{n_1}{n_1+n_2} & -\dfrac{n_2}{n_1+n_2} & 1 \end{pmatrix}$$

Table 1. 95-%Simultaneous confidence intervals for p_{dom}, p_{add} and p_{rec} for the SNP rs7738656.

Model	Effect-Estimator	95%-Simultaneous Intervals	Adjusted p-Value
Dominant	−0.240	[−0.378; −0.082]	0.0058
Additive	−0.250	[−0.387; −0.092]	0.0043
Recessive	−0.053	[−0.119; 0.015]	0.13

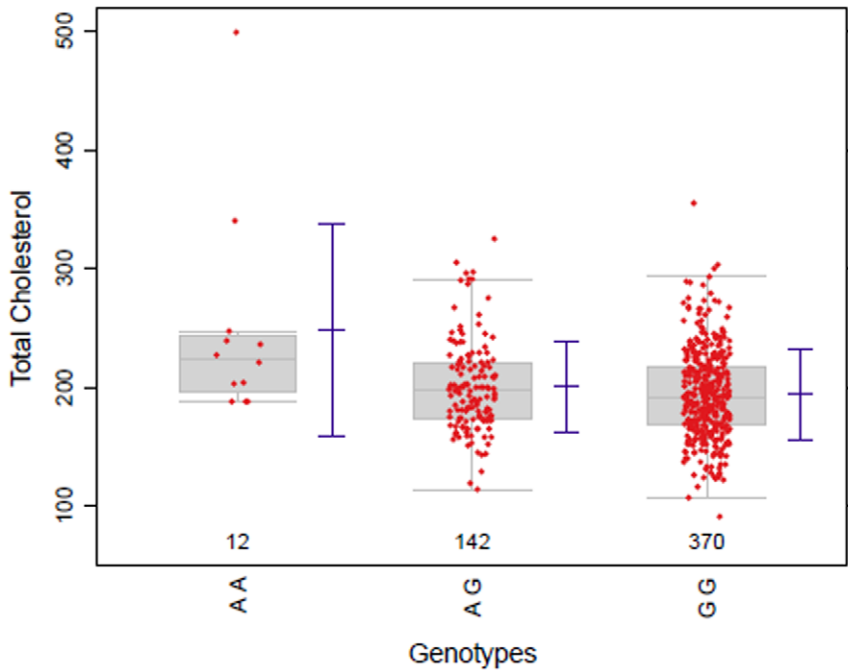doi:10.1371/journal.pone.0031242.t001

Figure 1. Boxplot: Total cholesterol vs. genotype rs7738656.
doi:10.1371/journal.pone.0031242.g001

denote the so-called Marcus-type contrast matrix [21]. In case of a balanced design, $\mathbf{C}$ reduces to

$$\mathbf{C} = \begin{pmatrix} \mathbf{c}'_{dom} \\ \mathbf{c}'_{add} \\ \mathbf{c}'_{rec} \end{pmatrix} = \begin{pmatrix} -1 & \frac{1}{2} & \frac{1}{2} \\ -1 & 0 & 1 \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix}.$$

Each row vector $\mathbf{c}'_\ell$ of $\mathbf{C}$ corresponds to one of the three genetic models. In case of a dominant mode of inheritance, the distributions F_2 and F_3 are identical, therefore a relative genetic effect for this mode can be expressed by

$$p_{dom} = \mathbf{c}'_{dom}\mathbf{p} = -p_1 + \frac{n_2}{n_2+n_3}p_2 + \frac{n_3}{n_2+n_3}p_3,$$

which denotes the difference between the pooled effect $\frac{n_2}{n_2+n_3}p_2$ $+\frac{n_3}{n_2+n_3}p_3$ among the samples 2 and 3 and p_1. Thus, in case of "no dominant effect", $\mathbf{c}'_{dom}\mathbf{p} = 0$, or equivalently, $\frac{n_2}{n_2+n_3}p_2 +$ $\frac{n_3}{n_2+n_3}p_3 = p_1$. Analogously, in case of a recessive mode of inheritance, the distributions F_1 and F_2 are identical, thus, a relative recessive effect can be expressed by

$$p_{rec} = \mathbf{c}'_{rec}\mathbf{p} = -\frac{n_1}{n_1+n_2}p_1 - \frac{n_2}{n_1+n_2}p_2 + p_3,$$

which denotes the difference between the pooled effect $\frac{n_1}{n_1+n_2}p_1 + \frac{n_2}{n_1+n_2}p_2$ and p_3. "No recessive mode of inheritance" means $\mathbf{c}'_{rec}\mathbf{p} = 0$, or, equivalently, $\frac{n_1}{n_1+n_2}p_1 + \frac{n_2}{n_1+n_2}p_2 = p_3$. In addition, the relative genetic effect for an additive mode of inheritance can be expressed by

$$p_{add} = \mathbf{c}'_{add}\mathbf{p} = -p_1 + p_3.$$

Thus, the case of *no global effect* is characterized by $p_{dom} = p_{rec} = p_{add}$, or, equivalently, $\mathbf{Cp} = 0$.

A multiple contrast test approach for the three genetic models

To test the individual hypothesis $H_0 : \mathbf{c}'_\ell\mathbf{p} = p_\ell = 0$, where $\ell \in \{dom,rec,add\}$, define a test statistic T_ℓ, which denotes, as usual, a studentized estimator $\hat{p}_\ell$ of p_ℓ with its estimated standard error (details see Technical details). The three test statistics T_{dom}, T_{rec} and T_{add} are collected in the vector

$$\mathbf{T} = (T_{dom}, T_{rec}, T_{add})'.$$

The multiple contrast test and the simultaneous confidence intervals for p_ℓ are based on the asymptotic multivariate normality of $\mathbf{T}$, i.e. the correlation among the three test statistics T_ℓ is accounted for. Instead of using critical values coming from a standard normal distribution (or t-distribution), we use critical values from the multivariate normal distribution $N(0,\hat{\mathbf{R}})$, where $\hat{\mathbf{R}}$ denotes the estimated correlation matrix. This means, the individual hypothesis $H_0 : \mathbf{c}'_\ell\mathbf{p} = 0$ is rejected at multiple level α of significance, if

$$|T_\ell| \geq z(1-\alpha,\hat{\mathbf{R}}), \tag{4}$$

where $z(1-\alpha,\hat{\mathbf{R}})$ denotes the $(1-\alpha)$-equicoordinate quantile from $N(0,\hat{\mathbf{R}})$. Simultaneous confidence intervals for the three genetic effects p_{dom}, p_{rec} and p_{add} are given by

$$\hat{p}_\ell \pm z(1-\alpha,\hat{\mathbf{R}})\cdot SE(\hat{p}_\ell), \tag{5}$$

i.e. point estimator $\pm$ quantile $\times$ estimated standard error. Note that the individual test decisions and the simultaneous confidence intervals are compatible, i.e. it can not occur that an individual hypothesis has been rejected, but the corresponding simultaneous confidence interval includes the value from the null hypothesis. These confidence intervals, however, may not be range preserving, i.e. the lower bounds may be smaller than -1 and the upper bounds can be larger than 1. Range preserving confidence intervals can be easily constructed by using the delta method [16,22], with the Fisher transformation. The global hypothesis $H_0 : \mathbf{Cp} = 0$ will be rejected, if

$$\max\{|T_{dom}|,|T_{rec}|,|T_{add}|\} \geq z(1-\alpha,\hat{\mathbf{R}}), \qquad (6)$$

i.e. if any of the three individual hypotheses have been rejected. For small sample sizes, the quantiles from the multivariate normal distribution are replaced by quantiles coming from a multivariate t-distribution (details see Technical details).

Evaluation of score phenotypes. Particularly in psychiatric epidemiology, different mental scores are often used as phenotypes, see e.g. [7]. Some mental scores are based on only few categories, e.g. 0,1,2, others represent sums of sub-scores with a wider range of count values. The definition of the unweighted relative effect p_j defined in (3) includes ordered categorical data. For an arbitrary monotone transformation ϕ of the data, it can be seen that

$$p_j^{\phi} = \frac{1}{3}\sum_{i=1}^{3}\left[P(\phi(X_{i1}) < \phi(X_{j1})) + 0.5P(\phi(X_{i1}) = \phi(X_{j1}))\right]$$
$$= \frac{1}{3}\sum_{i=1}^{3}\left[P(X_{i1} < X_{j1}) + 0.5P(X_{i1} = X_{j1})\right] = p_j.$$

Thus, the effect measure is invariant under monotone transformations of the data. On the other hand, if the data are transformed by a monotonic decreasing function ψ

$$p_j^{\psi} = \frac{1}{3}\sum_{i=1}^{3}\left[P(\psi(X_{i1}) < \psi(X_{j1})) + 0.5P(\psi(X_{i1}) = \psi(X_{j1}))\right]$$

$$= 1 - \frac{1}{3}\sum_{i=1}^{3}\left[P(X_{i1} < X_{j1}) + 0.5P(X_{i1} = X_{j1})\right] = 1 - p_j.$$

This means that the effect measure p_j is reflected at 0.5 in case of a monotonic decreasing transformation of the data. Therefore, p_j is an adequate measure for ordered categorical data, because the information is independent from the chosen scale of the scores.

Evaluation of phenotypes with values below a detection limit. Sometimes phenotypes with values below a detection limit occur, see e.g. [5]. Since the ordinal effect size is appropriate for tied values, all data below the detection limit should be fixed on a particular value. Note that this approach is only exact when a unique detection limit exists, i.e. the problem is more complex when different centers in a meta-analysis framework have different detection limits. However, due to the ranking of the groups, the problem of choosing the "best value" will not occur.

Results

Evaluation of the example

The new nonparametric multiple contrast test $\mathbf{T}$ was used for the statistical analysis of the motivating example above. Rank

estimators for the three unweighted relative effects are given by $\hat{p}_{AA} = 0.66$, $\hat{p}_{AG} = 0.44$ and $\hat{p}_{GG} = 0.41$, respectively. Assuming AA is the risk allele, a decrease to AG and GG occur. Table 1 summarizes the results of simultaneous Marcus-type comparisons.

The upper confidence limit of the additive model is most distant to $\mathbf{H}_0$, or, compatible to that, reveals the smallest p-value of $p = 0.0043$. The 95%-simultaneous confidence intervals indicate a positive association with the high risk allele A for the phenotype *total cholesterol*. The related parametric approach results a much smaller p-value of $p = 8.1^{-6}$ for the additive mode of inheritance ($p = 2.1^{-7}$ in the original publication [17] with an adjustment against covariates). This example illustrates the impact of the underlying assumptions being violated, in particular the assumption of normally distributed errors with homogeneous variances. The global rank Kruskal-Wallis test on heterogeneity reveals a p-value of only $p = 0.0062$.

In summary, using the multiple contrast tests yields specific information regarding the genetic mode of inheritance as well as simultaneous confidence intervals.

Simulations

We evaluated the empirical type-I error rates and the powers of nonparametric multiple contrast tests via extensive simulation studies. All simulations were performed using the publicly available software R (version 2.12.1; www.r-project.org). Every simulation step was repeated 10,000 times.

The trait genotypes for $N = 500,1000$ subjects were randomly drawn from a multinomial distribution with cell probabilities given by allele frequencies at trait locus $p = 0.5$, allele frequencies at trait marker $pm = 0.05,0.1,0.2,0.3,0.5$ and linkage disequilibrium delta $\delta = 0,0.01,0.02,0.03,0.04$. Phenotypic values for the quantitative traits were generated from normal and log-normal distributions, choosing 1 for the residual variance, and varying the percentage of variance explained by the quantitative trait $w = 0,0.2,0.4$ for an additive, dominant, or recessive mode of inheritance. Log-normal phenotypes were generated by first drawing normal phenotypic values X_{ik} and then by applying the transformation method $\Psi^{-1}(\Phi(X_{ik}))$, where $\Psi^{-1}(y)$ denotes the quantile function of the log-normal distribution, and $\Phi(x)$ denotes the standard normal distribution function. If $w = 0$, no variance is explained by the quantitative trait, thus, $\mathbf{Cp} = 0$ for all parameter settings. Low values of allele frequencies at trait marker (pm) result in strongly unbalanced designs. In addition, different values of pm, δ, and w form specific multimodal distributions under the alternative. Figure 2 displays examples of simulated normal and log-normal data for different values of w, $\delta = 0.04$, $p = 0.5$ and $pm = 0.5$.

Since the expectation of a multimodal distribution is the weighted sum of the single expectations, the parameter settings on p,pm and δ are an important issue in the investigation of power analyses.

Results. We simulated the nonparametric multiple contrast tests $\mathbf{T}$ as defined in (6) as well as its transformed approach by using the Fisher-transformation (*Fisher*). Two different types of contrasts will be examined throughout the simulation studies: (i) all-pairwise comparisons by using the contrast matrix

$$\mathbf{A} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}$$

to be sensitive against any heterogeneity (*All-Pairs*) and (ii) the Marcus-type contrast matrix $\mathbf{C}$ to be sensitive against exactly the three basic genetic modes of inheritance (*Marcus*). For each kind of
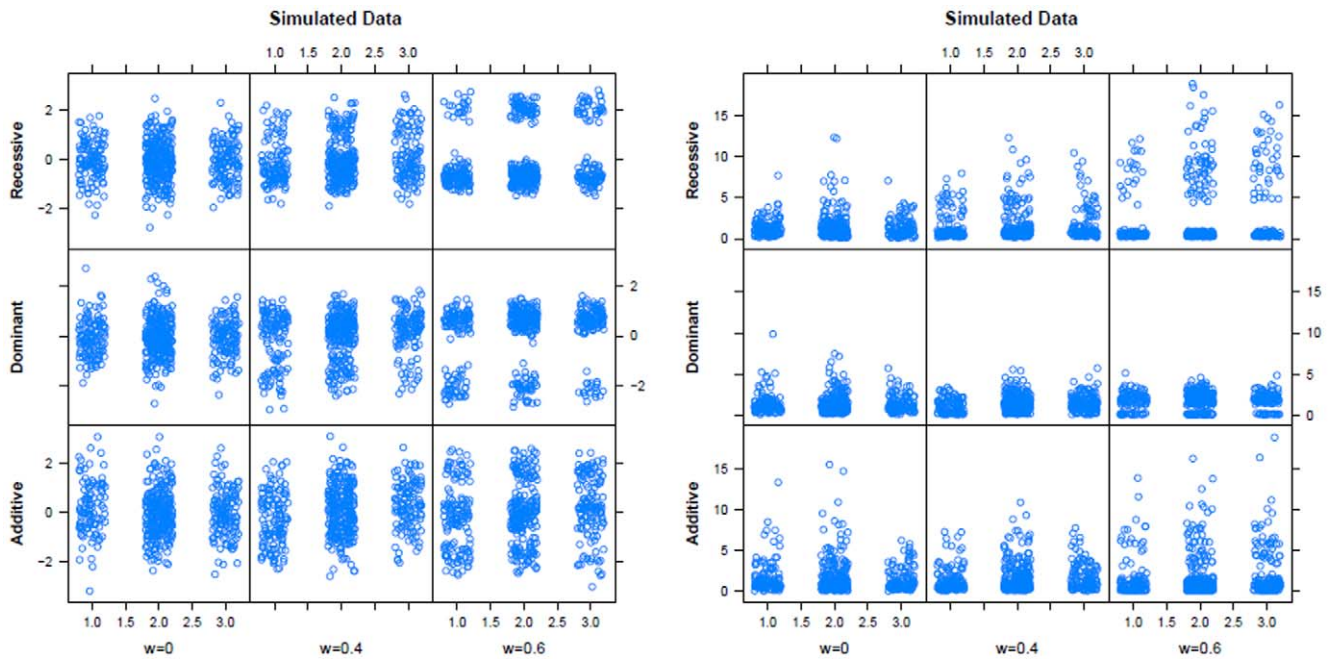
Figure 2. Simulated normal (left) and log-normal (right) data for different values of variance explained by the quantitative trait $w = 0, 0.4, 0.6$, $p = 0.5$, $pm = 0.5$, $\delta = 0.04$ **and an additive, dominant and recessive mode of inheritance.**
doi:10.1371/journal.pone.0031242.g002

contrast, the nonparametric multiple contrast tests are compared with the parametric multiple contrast tests for homoscedastic normal samples proposed by [23] as well as for heteroscedastic normal samples by [24] (denoted by *Bretz* and *Hasler*). For all-pairs comparisons, these four multiple contrast test procedures are compared with the nonparametric multiple test procedures by Steel [25] (*Steel*), the permutative Nemenyi-test [26] for all-pairs comparisons (*Nemenyi*), the Kruskal-Wallis test (*KW*) and the usual ANOVA-F-test (*ANOVA*). For Marcus-type comparisons, the four multiple contrast tests are compared with the nonparametric permutative Nemenyi-test for Marcus-type comparisons and the usual linear regression analysis (*Reg*). Figure 3 displays the type-I error simulation results (i.e. $w = 0$; $\alpha = 5\%$) for different values of pm, an additive, dominant, and recessive mode of inheritance, and linkage disequilibrium delta $\delta = 0, 0.01, 0.02, 0.03, 0.04$ for both normal and log-normal distributed phenotypes ($N = 500$).

It follows from Figure 3, that, under normality, all considered procedures control the type-I error at level $\alpha = 5\%$ for both *all-pairs* and *Marcus*-type comparisons. In the case of extremely unbalanced designs ($pm = 0.05$), however, the new multiple contrast tests **T** and *Fisher* tend to be quite liberal. This is due to the fact that these procedures do not use a pooled variance estimator. We observed this for both normally and log-normally distributed data. In the case of larger sample sizes ($N = 1000$), this effect disappears. The parametric multiple contrast test by *Hasler* tends to be very liberal when the normality assumption is violated.

To investigate the power of the different procedures mentioned above, different parameter settings on the variance explained by the quantitative trait ($w = 0.2, 0.4$) and different values on pm and δ were examined. Figure 4 displays the simulation results for $w = 0.2$.

Figure 4 shows that the power of all investigated procedures depends on the parameter settings of pm and δ. The combination of these parameters leads to specific values of weights in the multimodal distributions of the phenotypes as displayed in Figure 2. We observe that for a given w and δ, the power of the

tests is smaller for larger pm, although the data are almost balanced in such settings. This occurs, because the weighting parameters of the multimodal distributions are likely in case of smaller allele frequencies at trait marker. In case of $pm = 0.5$, the bimodal distributions consist of a dominated and a dominating part, which results in a smaller expectation of the phenotypic values in all considered cases. In case of extremely unbalanced designs ($pm = 0.05$), the power of the new procedures is quite low; in general, their power is not estimable due to their liberality in such settings. For normal distributions, the powers of all the parametric and nonparametric procedures are nearly identical in case of $pm \geq 0.2$. When the normality assumption is violated, the nonparametric procedures have a considerably higher power than the parametric procedures. The power of the new multiple contrast tests are likely to be identical to the power of the Kruskal-Wallis test. The Kruskal-Wallis test, however, can only be used for testing the global null hypothesis, and cannot provide any information regarding genetic association. Further, comparing the results of the all-pairs and Marcus-type comparisons, we observe that all the four multiple contrast tests exhibit higher power when using the Marcus-type contrast matrix compared to using the Tukey-type contrast matrix **A**. Simple linear regression analysis should not be used, because (i) the genotypic values are not metric numbers and thus the results depend on the chosen numbers for the three genotype scores and (ii), in all simulations the regression does not provide a considerably higher power than the multiple contrast test procedures. The same conclusions can be drawn for the simulation results obtained by $w = 0.4$, which are displayed in Figure 5.

Software

For a convenient application of the developed procedures, the R-software package *nparcomp* was developed and is available from CRAN. It contains various functions for the analysis of two independent samples (*npar.t.test*), as well as functions for the computation of nonparametric multiple contrast tests and simultaneous confidence intervals based on global ranks and
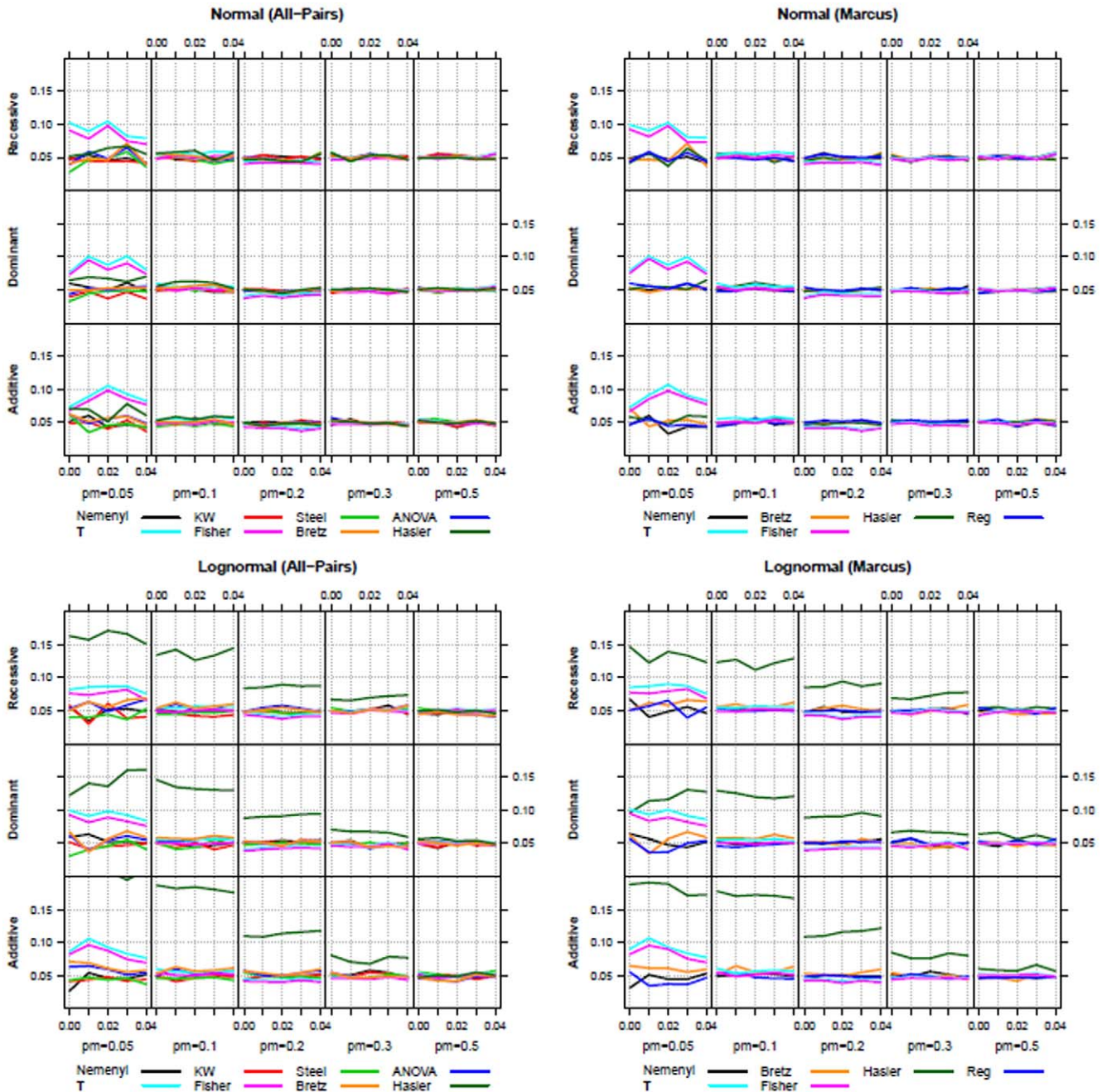
Figure 3. Type-I error ($\alpha = 5\%$) simulation results for all-pairs (left) and Marcus-type (right) comparisons using normal (upper row) and log-normal (lower row) distributions ($N = 500$).
doi:10.1371/journal.pone.0031242.g003

pairwise ranks. For example, the function *nparcomp* computes simultaneous confidence intervals and adjusted p-values for relative effects in arbitrary contrast settings based on pairwise ranks. Moreover, one-sided and two-sided confidence intervals and adjusted p-values are computed using multivariate normal-approximation, multivariate t_v-approximation, Logit-approximation and Probit-approximation described in [16].

Discussion

A nonparametric approach to evaluate the association between a di-allelic marker and a non-normal distributed quantitative trait

is proposed for simple population-based studies. Using a Marcus-type multiple contrast test for relative effects allows model-specific testing of either dominant, additive or recessive mode of inheritance. Furthermore, an all-pairwise comparisons contrast test is proposed as an alternative to the Kruskal-Wallis heterogeneity test. Procedures for obtaining related simultaneous confidence intervals or multiplicity-adjusted p-values are provided. The advantage of obtaining confidence intervals is their interpretability in terms of stochastic order for studies with individuals according to [13]. Although related software is freely available using the R library nparcomp, the routine analysis of hundreds of thousands of SNPs can not be recommended. The
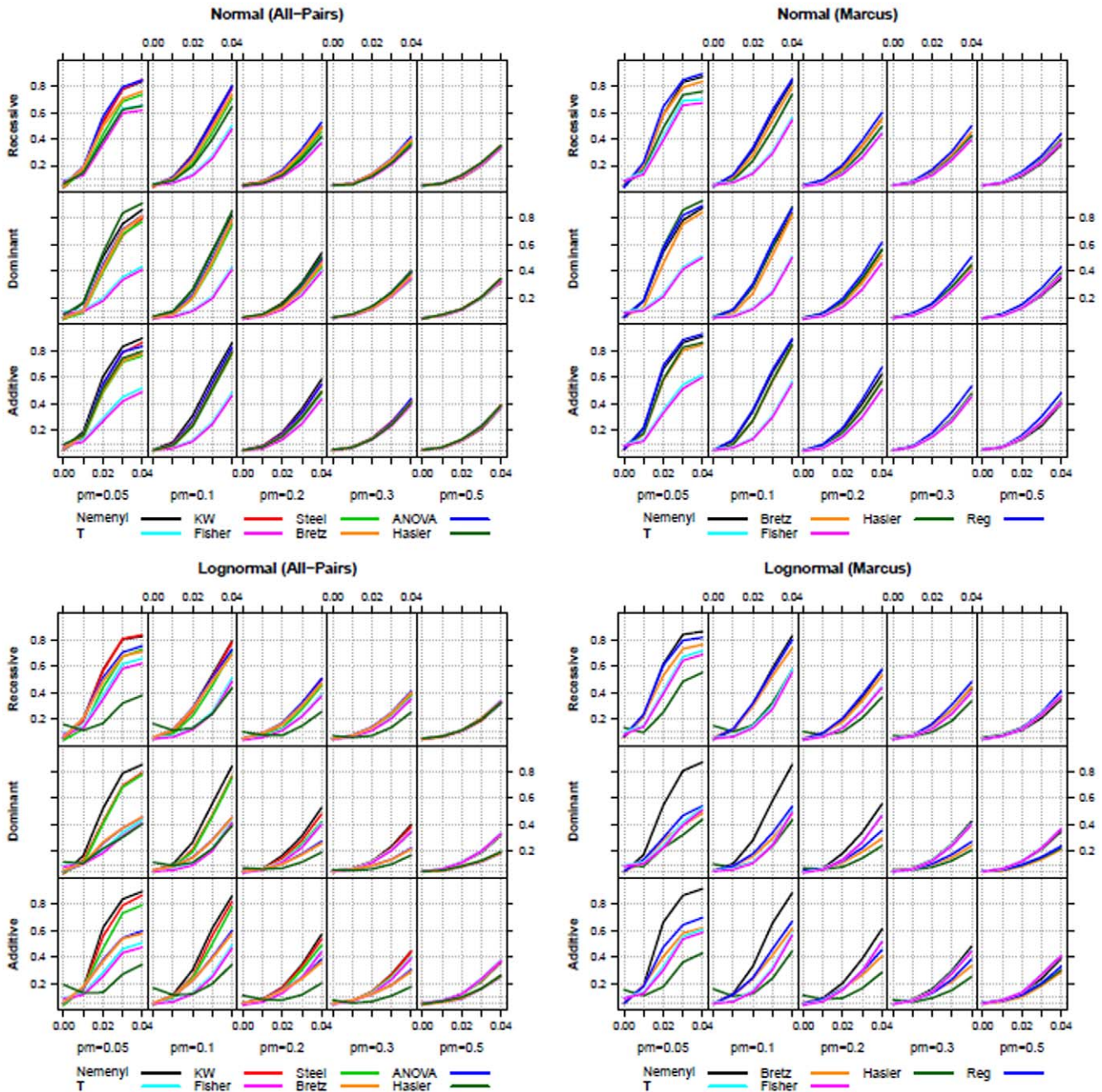
Figure 4. Power-simulation results ($\alpha = 5\%$) for all-pairs (left) and Marcus-type (right) comparisons using normal (upper row) and log-normal (lower row) distributions ($N = 500$). The variance explained by the quantitative trait was set to $w = 0.2$.
doi:10.1371/journal.pone.0031242.g004

computing time would be excessive and the amount of detailed information difficult to manage. For some selected candidate SNPs this approach can be easily performed for a number of phenotypes. If still an analysis on a genome-wide level is intended, an appropriate multiplicity adjustment of the simultaneous confidence is recommended, such as the false coverage statement rate [27].

Adjustment against multiple covariates is an important issue in unbiased testing association. The adjustment against population stratification, e.g. by principle components [28], or subject-specific baseline values, e.g. age, are relevant. For example [6] adjusted the relationship between an eQTL and the genotype scores against the

covariates age, kind of tissue (kidney cortex or medulla), ancestry (CEU or not) and gender (males or females). Nonparametric analysis of covariance is challenging [29], particularly to adjust against covariates due to possible population stratification. This is a topic of future work.

Technical details

To estimate the unknown relative effect p_j defined in (3), let

$$\hat{F}_i(x) = \frac{1}{n_i} \sum_{i=1}^{3} c(x - X_{ik}), \ i = 1,2,3,$$
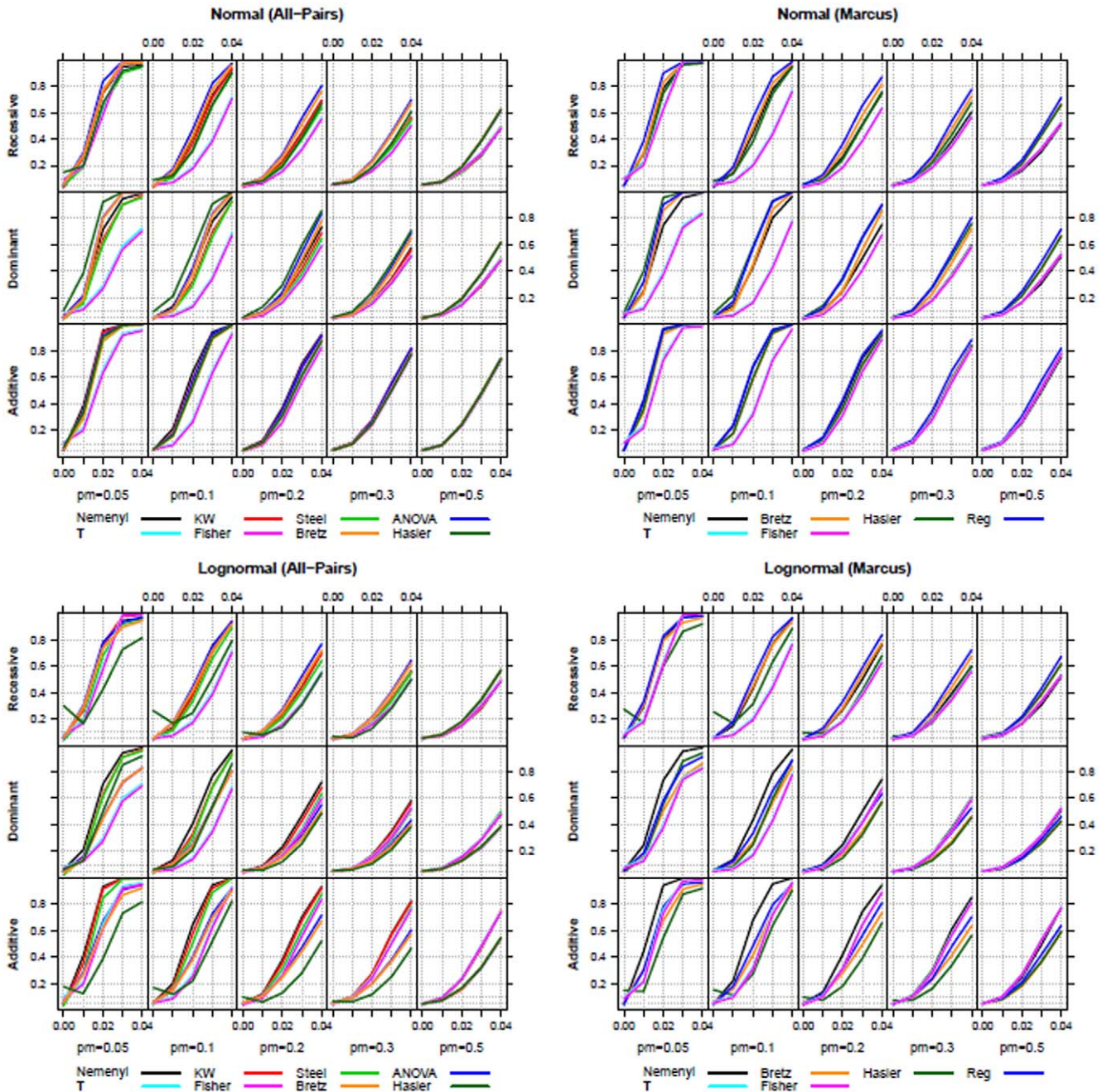
Figure 5. Power-simulation results ($\alpha = 5\%$) for all-pairs (left) and Marcus-type (right) comparisons using normal (upper row) and log-normal (lower row) distributions ($N = 500$). The variance explained by the quantitative trait was set to $w = 0.4$.
doi:10.1371/journal.pone.0031242.g005

denote the empirical distribution function, where $c(x) = 0, 0.5, 1$ according to $x < 0, x = 0, x > 0$, respectively. An unbiased estimator of $p_{ij} = \int F_i dF_j = P(X_{i1} < X_{j1}) + 0.5 P(X_{i1} = X_{j1})$ as used in (3) is obtained by replacing the unknown distributions F_i and F_j by their empirical counterparts $\hat{F}_i$ and $\hat{F}_j$. The estimators

$$\hat{p}_{ij} = \int \hat{F}_i d\hat{F}_j = \frac{1}{n_i} \left(\overline{R}_j^{(ij)} - \frac{n_j + 1}{2} \right), \quad \overline{R}_j^{(ij)} = \frac{1}{n_j} \sum_{k=1}^{n_j} R_{jk}^{(ij)}, \quad (7)$$

can be easily computed with the ranks $R_{jk}^{(ij)}$ of the observations $X_{j1}, \ldots, X_{jn_j}$. Here, $R_{jk}^{(ij)}$ denotes the rank of X_{jk} among all $n_i + n_j$ observations in the combined sample (i,j). Thus, an estimator of p_j is given by

$$\hat{p}_j = \frac{1}{3} \sum_{i=1}^{3} \hat{p}_{ij}.$$

The ranks used for the estimation of p_j are also called *pseudo-ranks* in the literature [20]. In case of a balanced design ($n_1 = n_2 = n_3$), the pseudo-ranks are identical to the usual global ranks. Let $\hat{\mathbf{p}} = (\hat{p}_1, \hat{p}_2, \hat{p}_3)'$ denote the vector of the three estimators. Thus, rank estimators of the three

relative genetic effects p_{dom}, p_{rec}, and p_{add} are given by $\hat{p}_{dom} = \mathbf{c}'_{dom}\hat{\mathbf{p}}$, $\hat{p}_{rec} = \mathbf{c}'_{rec}\hat{\mathbf{p}}$, and $\hat{p}_{add} = \mathbf{c}'_{add}\hat{\mathbf{p}}$, respectively.

It was shown that $\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p})$ is asymptotically multivariate normal with mean 0 and covariance matrix $\mathbf{V}$ [16,22]. Due to the quite involved structure of $\mathbf{V}$, let the estimator of $\mathbf{V}$ be denoted by $\hat{\mathbf{V}}$ [16,22]. To test each individual hypothesis $H_0 : \mathbf{c}'_\ell\mathbf{p} = 0$ on no genetic association, where $\ell \in \{dom, rec, add\}$, let $\hat{v}_\ell = \mathbf{c}'_\ell \hat{\mathbf{V}} \mathbf{c}_\ell$ denote the variance estimator of $\sqrt{N}\mathbf{c}'_\ell\hat{\mathbf{p}}$ and define the test statistic

$$T_\ell = \sqrt{N} \frac{\mathbf{c}'_\ell(\hat{\mathbf{p}} - \mathbf{p})}{\sqrt{\hat{v}_\ell}}.$$

The three test statistics T_ℓ are collected in the vector

$$\mathbf{T} = (T_{dom}, T_{rec}, T_{add})'.$$

The distribution of $\mathbf{T}$ can be approximated by a multivariate $T(v, 0, \hat{\mathbf{R}})$ distribution, where v denotes a Welch-Satterthwaite degree of freedom, non-centrality vector 0, and estimated correlation matrix $\hat{\mathbf{R}}$ [16]. The individual hypothesis $H_0 : \mathbf{c}'_\ell\mathbf{p} = 0$ will be rejected at multiple level α, if

$$|T_\ell| \geq t(1 - \alpha, v, \hat{\mathbf{R}}),$$

where $t(1 - \alpha, v, \hat{\mathbf{R}})$ denotes the $(1 - \alpha)$ equicoordinate quantile of $T(v, 0, \hat{\mathbf{R}})$. Approximate $(1 - \alpha)$ simultaneous confidence intervals for the three genetic effects p_{dom}, p_{rec}, and p_{add} are obtained from

$$\left[\mathbf{c}'_\ell\hat{\mathbf{p}} - t(1 - \alpha, v, \hat{\mathbf{R}})\sqrt{\hat{v}_\ell/N}; \mathbf{c}'_\ell\hat{\mathbf{p}} - t(1 - \alpha, v, \hat{\mathbf{R}})\sqrt{\hat{v}_\ell/N}\right]. \quad (8)$$

The global hypothesis $H_0 : \mathbf{Cp} = 0$ will be rejected, if

$$\max\{|T_{dom}|, |T_{rec}|, |T_{add}|\} \geq t(1 - \alpha, v, \hat{\mathbf{R}}).$$

Range preserving confidence intervals are given by

$$p_{\ell,L} = \frac{\exp(2p^*_{\ell,L}) - 1}{\exp(2p^*_{\ell,L}) + 1} \; p_{\ell,U} = \frac{\exp(2p^*_{\ell,U}) - 1}{\exp(2p^*_{\ell,U}) + 1}, \quad (9)$$

where

$$p^*_{\ell,L}, p^*_{\ell,U} = \frac{1}{2}\log\left(\frac{1 + \hat{p}_\ell}{1 - \hat{p}_\ell}\right) \mp z(1 - \alpha, \hat{\mathbf{R}})\frac{1}{1 - \hat{p}_\ell^2} SE(\hat{p}_\ell). \quad (10)$$

Alternatively, a pairwise rankings version is available which can be easily derived from two-sample tests. They behave similarly to the global rankings approach, but they can lead to paradoxical results.

Pairwise rankings version

As mentioned in the previous section, the three genetic effects $p_\ell = \mathbf{c}'_\ell\mathbf{p}$, $\ell \in \{dom, rec, add\}$, denote generalized two-sample relative effects, which were estimated with global ranks of the data X_{ik}.

Thus, the effects can be modified such that pairwise ranks are used for estimation. Let

$$p_{ij} = P(X_{i1} < X_{j1}) + 0.5P(X_{i1} = X_{j1}), i \neq j, \quad (11)$$

denote the two-sample relative effect between the genotype levels i and j. If $p_{ij} < 0.5$, then the values from F_i tend to be larger than those from F_i. In case of $p_{ij} = 0.5$, none of the observations tend to be smaller or larger. Thus, the case of *no association* can be expressed by $p_{ij} = 0.5$. The relative dominant genetic effect on association describes the difference between the distribution F_1 and the combined sample $F_{dom} = \frac{n_2}{n_2 + n_3}F_2 + \frac{n_3}{n_2 + n_3}F_3$. Thus, a two-sample relative dominant effect can be described by

$$q_{dom} = \frac{n_2}{n_2 + n_3}p_{12} + \frac{n_3}{n_2 + n_3}p_{13},$$

and denotes a linear combination of p_{1j}. The relative recessive effect describes the difference between the combined sample $F_{rec} = \frac{n_1}{n_1 + n_2}F_1 + \frac{n_2}{n_1 + n_2}F_2$ and F_3. Thus, a relative effect on a recessive mode of inheritance is given by

$$q_{rec} = \frac{n_1}{n_1 + n_2}p_{13} + \frac{n_2}{n_1 + n_2}p_{23}.$$

Finally, the relative two-sample effect on an additive mode of inheritance can be expressed by

$$q_{add} = p_{13}.$$

The effects q_ℓ can be estimated by using the pairwise rank estimators $\hat{p}_{ij}$ defined in (7) by

$$\hat{q}_{dom} = \frac{n_2}{n_2 + n_3}\hat{p}_{12} + \frac{n_3}{n_2 + n_3}\hat{p}_{13}$$

$$\hat{q}_{rec} = \frac{n_1}{n_1 + n_2}\hat{p}_{13} + \frac{n_2}{n_1 + n_2}\hat{p}_{23}$$

$$\hat{q}_{add} = \hat{p}_{13}.$$

Multiple contrast tests for the hypotheses $H_0 : q_\ell = 0.5$ and simultaneous confidence intervals for the effects q_ℓ, where $\ell \in \{dom, rec, add\}$, can be derived in the same way as described in the previous section. We note that the effects p_{ij} may be intransitive, i.e. it may occur that $p_{12} \leq p_{23} \leq p_{31}$ resulting in paradoxical results [30,31]. Therefore, we recommend using the global ranking version.

Author Contributions

Analyzed the data: FK OL LAH. Contributed reagents/materials/analysis tools: FK OL LAH. Wrote the paper: FK OL LAH.

References

1. Padmanabhan S, Melander O, Johnson T, Di Blasio AM, Lee WK, et al. (2010) Genome-wide association study of blood pressure extremes identifies variant near umod associated with hypertension. Plos Genetics 6: e1001177.
2. Sasieni P (1997) From genotypes to genes: Doubling the sample size. Biometrics 53: 1253–1261.
3. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) Plink: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81: 559–575.
4. Freidlin B, Zheng G, Li Z, Gastwirth J (2002) Trend tests for case-control studies of genetic markers: Power, sample size and robustness. Human Heredity 53: 146–152.
5. Melzer D, Perry JRB, Hernandez D, Corsi AM, Stevens K, et al. (2008) A genome-wide association study identifies protein quantitative trait loci (pqtls). Plos Genetics 4: e1000072.
6. Wheeler HE, Metter EJ, Tanaka T, Absher D (2009) Sequential use of transcriptional profiling, expression quantitative trait mapping, and gene association implicates mmp20 in human kidney aging. Plos Genetics 5: e1000685.
7. Fanous AH, Chen XN, Wang X, Amdur R, O'Neill FA, et al. (2009) Genetic variation in the serotonin 2a receptor and suicidal ideation in a sample of 270 irish high-density schizophrenia families. American Journal of Medical Genetics Part B-neuropsychiatric Genetics 150B: 411–417.
8. Kruglyak L, Lander ES (1995) A nonparametric approach for mapping quantitative trait loci. Genetics 139: 1421–1428.
9. Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. J Amer Stat Ass 47: 583–621.
10. Jonckheere A (1954) A Distribution-free k-sample Test Against Ordered Alternatives. Biometrika 41: 133–145.
11. Little J, Higgins JPT, Ioannidis JPA, Moher D, Gagnon F, et al. (2009) Strengthening the reporting of genetic association studies (strega): an extension of the strobe statement. Human Genetics 125: 131–151.
12. Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, et al. (2008) Genome-wide association study provides evidence for a breast cancer risk locus at 6q22–33. Proceedings of the National Academy of Sciences of the United States of America 105: 4340–4345.
13. Browne RH (2010) The t-test p value and its relationship to the effect size and p(x<y). American Statistician 64: 30–33.
14. Brunner E, Munzel U (2000) The nonparametric behrens-fisher problem: Asymptotic theory and a small-sample approximation. Biometrical Journal 1: 17–21.
15. Ryu E (2008) Simultaneous confidence intervals using ordinal effect measures for ordered categorical outcomes. Statistics in Medicine 28: 3179–3188.
16. Konietschke F (2009) Simultane Konfidenzintervalle für nichtparametrische relative Kontrasteffekte. Ph.D. thesis, Georg-August Universität Göttingen.
17. Smith EN, Chen W, KahonenMea (2010) Longitudinal genome-wide association of cardiovascular disease risk factors in the bogalusa heart study. Plos Genetics 6: e1001094.
18. Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, et al. (2009) Genome-wide association study of bipolar disorder in european american and african american individuals. Molecular Psychiatry 14: 755–763.
19. Brunner E, Puri ML (2001) Nonparametric methods in factorial designs. Statistical Papers 42: 1–52.
20. Gao XY, Stamier J, Martin ER (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. Genetic Epidemiology 32: 361–369.
21. Marcus R (1982) Some results on simultaneous confidence-intervals for monotone contrasts in oneway anova model. Communications In Statistics Part A-Theory And Methods 11: 615–622.
22. Pfeiffer S (2010) Simultane Konfidenzintervalle für nichtparametrische Effekte in faktoriellen Modellen. Master's thesis, University of Göttingen.
23. Bretz F, Genz A, Hothorn LA (2001) On the numerical availability of multiple comparison procedures. Biometrical Journal 43: 645–656.
24. Hasler M, Hothorn LA (2008) Multiple contrast tests in the presence of heteroscedasticity. Biometrical Journal 50: 793–800.
25. Steel RDG (1960) A rank sum test for comparing all pairs of treatments. Technometrics 2: 197–207.
26. Hothorn T, Hornik K, van de Wiel MA, Zeileis A (2007) coin: Conditional Inference Procedures in a Permutation Test Framework. URL http://CRAN.R-project.org/. R package version 0.6-2.
27. Benjamini Y, Yekutieli D (2005) False discovery rate-adjusted multiple confidence intervals for selected parameters. Journal of the American Statistical Association 100: 71–81.
28. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38: 904–909.
29. Bathke AC, Brunner E Recent advances and trends in nonparametric statistics, Elsevier, chapter A Nonparametric Alternative to Analysis of Covariance. pp 109–120.
30. Brown BM, Hettmansperger TP (2002) Kruskal-wallis, multiple comparisons and efron dice. Australian and New Zealand Journal of Statistics 44: 427–438.
31. Thangavelu K, Brunner E (2006) Wilcoxon mann-whitney test and efrons paradox dice. Journal of Statistical Planning and Inference 137: 720–737.