A detailed view on Model-Based Multifactor Dimensionality Reduction for detecting gene-gene interactions in case-control data in the absence and presence of noise

TOM CATTAERT[1,2], **M. LUZ CALLE**[3], **SCOTT M. DUDEK**[4], **JESTINAH M. MAHACHIE JOHN**[1,2], **FRANÇOIS VAN LISHOUT**[1,2], **VICTOR URREA**[3], **MARYLYN D. RITCHIE**[4], and **KRISTEL VAN STEEN**[1,2,*]

[1]Systems and Modeling Unit, Montefiore Institute, University of Liege, Grande Traverse 10, 4000 Liège, Belgium

[2]Bioinformatics and Modeling, GIGA-R, University of Liege, Avenue de l'Hôpital 1, 4000 Liège, Belgium

[3]Department of Systems Biology, University of Vic, Carrer de la Sagrada Família 7, 08500 Vic, Spain

[4]Ritchielab, Center for Human Genetics Research, Vanderbilt University, 519 Light Hall, 2215 Garland Avenue, Nashville, TN 37232-0700, USA

SUMMARY

Analyzing the combined effects of genes and/or environmental factors on the development of complex diseases is a great challenge from both the statistical and computational perspective, even using a relatively small number of genetic and non-genetic exposures. Several data mining methods have been proposed for interaction analysis, among them, the Multifactor Dimensionality Reduction Method (MDR), which has proven its utility in a variety of theoretical and practical settings. Model-Based Multifactor Dimensionality Reduction (MB-MDR), a relatively new MDR-based technique that is able to unify the best of both non-parametric and parametric worlds, was developed to address some of the remaining concerns that go along with an MDR-analysis. These include the restriction to univariate, dichotomous traits, the absence of flexible ways to adjust for lower-order effects and important confounders, and the difficulty to highlight epistasis effects when too many multi-locus genotype cells are pooled into two new genotype groups. Whereas the true value of MB-MDR can only reveal itself by extensive applications of the method in a variety of real-life scenarios, here we investigate the empirical power of MB-MDR to detect gene-gene interactions in the absence of any noise and in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. For the considered simulation settings, we show that the power is generally higher for MB-MDR than for MDR, in particular in the presence of genetic heterogeneity, phenocopy, or low minor allele frequencies.

Keywords

---

*Corresponding author kristel.vansteen@ulg.ac.be.

INTRODUCTION

Many common human diseases and traits are believed to be influenced by several genetic and environmental factors, each factor potentially having a modifying effect on the other. Understanding the interplay between genetic and non-genetic factors that underlies these complex diseases and traits is one of the major goals of genetic epidemiology. In genetic association studies for common complex diseases, single nucleotide polymorphisms (SNPs) are the most commonly used type of genetic markers (Marnellos, 2003). This is in part understood by their dense distribution across the genome and their low mutation rate. Genome-wide association analysis (GWA), using a dense map of SNPs, has become one of the standard approaches for disentangling the genetic basis of complex genetic diseases (Hardy & Singleton, 2009). Despite the fact that GWAs have provided convincing evidence for identifying important genetic variants influencing a wide variety of common diseases and traits (Manolio *et al*., 2008, Seng & Seng, 2008), a lot of the genetic heritability cannot be explained by the (major) genetic loci discovered so far (Manolio *et al*., 2009). This may be attributed to the fact that reality shows multiple small associations, whereas common statistical techniques in this context only exhibit sufficient power to detect moderate to large associations. Also, looking beyond singular genetic effects and beyond the boundaries of additive inheritance of SNP polymorphisms should better reflect biological pathways that are involved in disease etiology (Dixon *et al*., 2000).

Standard methods to analyze the simultaneous evaluation of a large pool of predictors (whether genetic or not) broadly fall into two classes: parametric and non-parametric methods. For instance, in a classic logistic modeling framework, in which case-control status is taken as the outcome variable, the search for functional variants can be carried out by constructing a model for the probability of disease. Quantifying the effects of a single locus is achieved by interpreting the corresponding regression coefficients, conditional on the fixed status at the remaining loci. However, if the single locus is involved in complex multi-collinearity patterns with other loci included in the model, it is questionable how much value can be placed on this interpretation (Van Steen & Molenberghs, 2004). This issue becomes even more relevant as the number of terms increases and interaction terms are considered as well. In addition, traditional parametric approaches have severe limitations when there are too many independent variables in relation to the number of observed outcome events. This is also referred to as the curse of dimensionality (Bellman, 1961). Therefore, alternative methods have been proposed to deal with elevated dimensionality and related problems when investigating interactions, including penalized logistic regression (Park & Hastie, 2008), (bagged) logic regression (Ruczinski *et al*., 2004), and non-parametric multi-locus techniques based on machine learning and data mining. The latter comprise tree-based methods (e.g., Recursive Partitioning and Random Forests), pattern recognition methods (e.g., Symbolic Discriminant Analysis, Mining Association rules, Neural Networks and Support Vector Machines), and data reduction methods (e.g., Multifactor Dimensionality Reduction). Nice overviews have been given by Onkamo and Toivonen (Onkamo & Toivonen, 2006), by Motsinger *et al*. (Motsinger *et al*., 2007) and by Cordell (Cordell, 2009).

Several of the aforementioned strategies have been implemented, within the context of genetic association studies that specifically aim to identify and characterize gene-gene interactions (Cordell, 2002, Liang & Kelemen, 2008, Musani *et al*., 2007), with variable success. Often inadequate solutions are given to complex statistical hurdles such as acknowledging different modes of interaction, higher-order (>2) interactions or threshold effects (Altshuler *et al*., 2008, Moore, 2003). The observation that subtle variation in allele frequency can either introduce an interaction effect, or likewise remove an interaction effect from a particular dataset, further complicates the process (Greene *et al*., 2009). In general,

whatever strategy is chosen for epistasis detection, the analysis is complicated by the fact that many interactions are investigated (from the same data set), and that chance alone can elevate the possibility of finding a significant association. Methods that adequately find equilibrium between controlling false positive findings, yet have sufficient power to identify interactions, within a reasonable amount of computation time, will survive the test of time and serve promising for large-scale genome-wide interaction screening.

In this study, we focus on the recently introduced Model-Based Multifactor Dimensionality Reduction (MB-MDR) technique (Calle *et al*., 2008b). It was developed to address some of the remaining shortcomings of a classical MDR-analysis with univariate binary traits. While introducing several sources of noise that may distort the identification of epistasis signals, we evaluate the power of MB-MDR and compare its performance with MDR under the same simulated scenarios. Although both MB-MDR and MDR are applicable to identify higher-order interactions (>2), the simulation study restricts attention to identifying SNP-SNP interactions of order 2.

MATERIALS AND METHODS

In what follows, we briefly outline the key features of MDR and MB-MDR. More details about important differences between MDR and MB-MDR are referred to the discussion section.

MDR: Multifactor Dimensionality Reduction

Several publications exist that fully describe the general procedures of the MDR method(Hahn *et al*., 2003, McKinney *et al*., 2006, Moore *et al*., 2006, Ritchie *et al*., 2003, Ritchie *et al*., 2001). In summary, the main idea behind MDR is to reduce dimensionality by pooling multi-locus genotypes into two groups. For binary traits, these two groups can be viewed as risk groups for disease, and are usually referred to as high-risk and low-risk categories.

In particular, for each k-tuple of markers (in this study, k ranges from 1 to 5), the ratio of the number of cases to controls is evaluated within each multifactor cell and compared with the global ratio of cases over controls in the particular genotype combination being evaluated. Those cells with a case/control ratio equal to or above the global ratio are labeled as 'high-risk' and the remaining cells as 'low-risk'. This leads to a one-dimensional association model for disease with k loci, by pooling the high-risk cells into one group (H) and the low-risk cells into another group (L). The ability of the simplified model to correctly classify subjects as cases or controls is evaluated through the 'balanced accuracy' ((sensitivity + specificity) / 2) computed on training data (training accuracy) and test data (predictive accuracy), derived from a 10-fold cross-validation procedure. In particular, within a cross-validation data partition, but for every k-tuple of SNPs, the balanced predictive accuracy for the model with maximum training balanced accuracy is stored. The best k-locus model, over the 10 cross-validation sets, is subsequently selected as the single model that has the highest cross-validation consistency with average balanced predictive accuracy breaking any ties. Among the best k-factor models, the MDR best model is the model with maximum cross-validation consistency. Where ties are present, maximum average prediction accuracy is used and then, if still present, the rule of parsimony is adopted and the smallest model is selected. Statistical significance of the final model is determined by comparing the observed average balanced predictive accuracy for the final model with the empirically derived values for the best MDR model under the null hypothesis of no association. The latter is achieved by creating 1000 permutation data sets, while randomly permuting case/control labels of study subjects and running the entire MDR analysis procedure on all of the permuted

datasets. More information can be found at the URL
http://chgr.mc.vanderbilt.edu/ritchielab/MDR.

Although MDR is being widely used for interaction detection (Ma *et al.*, 2010, Pae *et al.*,
2010, Sonoda *et al.*, 2010, VanCleave *et al.*, 2010), it suffers from some major drawbacks
including that important interactions could be missed due to pooling too many multi-locus
genotype cells together and that it cannot adjust for covariates (confounding factors, lower-
order interaction or main effects). Hence, extensions to MDR have been proposed in order to
improve its performance, such as the Odds Ratio based Multifactor Dimensionality
Reduction method (OR-MDR) (Chung *et al.*, 2007) and the Generalized MDR (GMDR)
method (Lou *et al.*, 2007). These extensions copy the key principle of an MDR analysis,
namely selection of one best model via selection criteria based on a cross-validation
strategy. An alternative method is MB-MDR, its implementation offering a flexible
framework to encompass different study designs.

MB-MDR: Model-Based Multifactor Dimensionality Reduction

The key steps underlying an MB-MDR analysis have first been described by Calle *et al.*
(Calle *et al.*, 2008a, Calle *et al.*, 2008b) and are graphically displayed in Figure 1.

In Step 1, the possible multi-factor classes of k factors ($k = 2$ in Figure 1) are represented in
a k-dimensional space, and each multi-locus genotype cell, denoted by c_j ($j = 1,…,3^k$ for
diallelic markers), is tested for association with the response variable Y. These association
tests T_j can consistently be carried out within a parametric or a non-parametric paradigm. In
this study, the null hypothesis of no association between the binary trait Y and G_j (a
membership indicator variable for the multi-locus genotype cell c_j), is tested via a chi-square
test with 1 degree of freedom. In general, the cell c_j-specific association test statistics T_j can
be either positive or negative, depending on the direction of the effect (Figure 1). Because,
in our case, the chi-square test is always positive, we will actually assume that T_j is equal to
the square root of the chi-square test, with the sign depending on the derived odds ratio OR_j.
More specifically, $T_j > 0$ if $OR_j > 1$ and $T_j < 0$ if $OR_j < 1$.

In Step 2, the p-values p_j obtained for the association tests T_j are then compared to a
reference critical value p_c (Figure 1), usually taken to be $p_c = 0.1$. For more details about the
effect of alternative choices of p_c on power performance of MB-MDR, we refer to the
results and discussion section. Although the MB-MDR implementation is flexible in the way
the labeling is done, we have created 3 possible labels in the following way: high risk (H) if
$p_j < p_c$ and $T_j > 0$, low risk (L) if $p_j < p_c$ and $T_j < 0$ or no evidence for risk (O) if $p_j > p_c$,
respectively. This process is also illustrated in Figure 1. Pooling alike cells will establish 3
multi-locus genotype classes, and hence a new one-dimensional categorical variable X with
possible values H (high), L (low), O (no evidence), that can again be tested for its
association with Y. Once more, MB-MDR allows for different testing strategies, such as
computing the maximum test result $T = \max(|T_{H/LO}|,|T_{L/HO}|)$ of contrasting H versus $\{L, 0\}$
and L versus $\{H, O\}$ or computing the maximum test result $T = \max(|T_{H/L}|,|T_{H/LO}|,|T_{L/HO}|)$.
For the purpose of this study, we will focus primarily on a single chi-squared test $T = |T_{H/L}|$
with 1 degree of freedom, while testing X, now with possible values H and L, and ignoring
the multi-locus genotype category O, for its association with Y. Steps 1 and 2 are repeated
for every selection of k factors to be studied for their potential synergetic effects on the trait
Y.

Finally, in Step 3, a significance assessment is made. Special care needs to be taken, since
the pooling of multi-locus genotypes in Step 2 uses information about disease status, and
therefore leads to overly optimistic test results and inflated false positive rates. As in our
recent work (Cattaert *et al.*, 2010), but extended to multiple model selection and different

testing approaches, we compute resampling-based Westfall and Young step-down maxT adjusted p-values (Westfall & Young, 1993). The beauty of this method is that it allows drawing conclusions about the joint significance of several marker-tuples. The adopted procedure will always weakly control the family-wise error rate (FWER) at 5%. Moreover, under the assumption of subset pivotality, a reasonable assumption in the absence of Linkage Disequilibrium (LD) between the markers, even strong control of the FWER applies. The latter implies that the FWER is controlled under whatever configuration of true and false null hypotheses.

An implementation of the MB-MDR approach is available as the R package "mbmdr", and can be retrieved from the URL http://cran.r-project.org/web/packages/mbmdr/index.html. This R package calculates marginal permutation p-values, for the test approach $T = \max (|T_{H/LO}|, |T_{L/HO}|)$, but leaves the multiple testing correction for the different pairs to the user. A C++ implementation of the present approach is available from the authors upon request and will be described elsewhere.

Simulation study

We replicated a simulation study performed by (Ritchie *et al*., 2003), by using the simulated data available from http://chgr.mc.vanderbilt.edu/ritchielab/projects/MDR/DataSimulationFiles.zip, and by applying both MDR and MB-MDR to the available data, with and without added noise. In particular, 100 case-control data sets (200 cases and 200 controls) were simulated using 6 different two-locus epistasis models that harbor interaction effects (SNP5 × SNP10) in the absence of main effects. Genotypes for 10 SNPs, including the functional loci, were generated according to Hardy-Weinberg proportions, with minor allele frequency (MAF) for both functional and non-functional SNPs set to 0.5 in models 1 and 2, 0.25 in models 3 and 4 and 0.1 in models 5 and 6. An overview of the model-dependent allele frequencies and the corresponding penetrance functions is given in Figure 2. In addition, data were generated under the 6 epistasis models of Figure 2, in the presence of commonly encountered sources of noise: 5% genotyping error (GE), 5% missing data (MS), 50% phenocopy (PC), and 50% genetic heterogeneity (GH). In the presence of GH, the two functional SNP pairs were (SNP5, SNP10) and (SNP3, SNP4). 1000 null data sets were also generated under the null hypothesis of no association at all, with 10 SNPs having a MAF of 0.5 and independent of case-control status.

Although these data have been analyzed before using MDR (Ritchie *et al*., 2003), and GH results obtained with MDR have been reinterpreted (Ritchie *et al*., 2007), all data were re-analyzed using the latest MDR software that exploits insights and knowledge acquired since its conception. This acquired knowledge implies using balanced accuracy instead of simple accuracy as an evaluation measure, and only performing a single cross-validation run instead of multiple runs (see also foregoing subsection '*MDR: Multifactor Dimensionality Reduction*'). Since MDR only identifies one best model, model selection is based on screening over 1-5-order models in order to be able to detect two functional pairs. The same screening algorithm was adopted, even for those simulated scenarios without GH.

In analyzing the available data with MB-MDR, targeting two-order gene-gene interactions, we considered different Step 2 p-value cut-offs $p_c = 0.05, 0.1, 0.2, 0.5$ and 1. In addition, the Step 2 choices $T = \max (|T_{H/L}|, |T_{H/LO}|, |T_{L/HO}|)$, $\max (|T_{H/LO}|, |T_{L/HO}|)$ and $|T_{H/L}|$ were investigated for their performance on the power of MB-MDR for detecting the causal interacting pair(s).

For both MDR and MB-MDR, p-values of the final results (permutation-based p-values based on 1000 replicates for MDR, resampling-based Westfall and Young step-down maxT

adjusted p-values for MB-MDR based on 999 permutations) were compared to 0.05 to assess significance of the findings. This constitutes another difference with the MDR results in the original publications where no significance assessment was involved, but selection probabilities were obtained rather than power.

Power, specific power and false positive rate

For the settings without GH, the power of MB-MDR to identify the actual causal pair was defined as the proportion of the 100 datasets for which the functional pair (SNP5, SNP10) was found significant at the 5% level. Similarly, the power of MDR to identify the actual causal pair was defined as the proportion of the 100 simulated datasets for which the best model included the pair (SNP5, SNP10) and was found significant at the 5% level after permutation testing.

For scenarios involving GH, several useful definitions of power can be introduced, as there are, the power to find both functional pairs (SNP5, SNP10) and (SNP3, SNP4), the power to retrieve the first functional pair, and the power to find at least one of the functional pairs. As long as respectively both pairs, pair (SNP5, SNP10), or at least one of the two functional pairs are assessed significant for MB-MDR, or are included in the best model for MDR and this best MDR model is significant in permutation testing, the power is elevated with 1%. An overview of the various power definitions is given in Table S1 as Supporting Information. Note that these definitions obviously imply that the power to find both functional pairs is smaller than the power to find the first functional pair, which is in turn smaller than the power to find at least one functional pair. As a remark, slightly different definitions for the power to find the first and at least one of the pairs have been used in the original re-analysis (Ritchie *et al*., 2007), invalidating the first of these logical orderings.

Each of these definitions allow for additional non-functional pairs to be found significant (in the case of MB-MDR) or non-functional loci to be included in the significant best model (for MDR). Therefore, more specific power definitions for both MB-MDR and MDR have been explored. These specific power evaluations were defined analogously to the aforementioned power evaluations, but hold information about specifically detecting the functional pair(s) and not detecting also non-functional pairs or loci. An overview of the different resulting specific power definitions can be found in Table S1. By construction, the same logical ordering of the different power evaluations defined for GH also applies for the corresponding specific power evaluations.

When exploring false positive rates we distinguished between false positive rates for null data, i.e. when no association was present, and false positive rates under the alternative of epistasis, i.e. when one or more interacting pairs exist. For null data, false positive rates were computed for MB-MDR as the proportion of null data sets that highlight at least one significant pair, and for MDR as the proportion of null data sets for which the best model is found significant. Note that this false positive rate is evaluated family-wise (FWER) for MB-MDR which can select multiple significant pairs, while it is a simple rate for MDR which proposes only one best model.

For data under the alternative of epistasis, false positive rates for MB-MDR were defined as the proportion of data sets in which at least one non-functional pair was wrongly found significant (FWER). For MDR and in the absence of GH, the false positive rate was defined as the proportion of data sets for which the best model was assessed significant but did not exactly coincide with the loci of the actual pair. In the presence of GH the situation is more complex and the MDR false positive rate was defined in terms of obtaining a significant best model that either did not contain at least one of the functional pairs, or contained at least one non-functional locus. Again, an overview of the various false positive rate definitions is

given in Table S1. Note that, as specific power is defined in terms of finding the significant pair(s) while at the same time not making any error, and false positive rates in terms of making at least one error, the power loss observed when adopting the more stringent specific power definition is bounded by the corresponding false positive rate. In the presence of GH, this applies to each of the three different power definitions.

RESULTS

We first compared (Figures 3 and 4) MDR screening over 1-5 order and MB-MDR with different p-value cut-offs $p_c = 0.05, 0.1, 0.2, 0.5$ and 1, and different test approaches $T = |T_{H/L}|$, max ($|T_{H/LO}|$, $|T_{L/HO}|$) and max ($|T_{H/L}|$, $|T_{H/LO}|$, $|T_{L/HO}|$). For $p_c = 1$, there is no O category and the three methods coincide. MDR results were visualized by bullets at $p_c = 1$ because MDR does not use the O category.

Figure 3 shows MB-MDR and MDR power to correctly identify the interacting pair for data with different sources of noise excluding GH. First, MB-MDR clearly outperforms MDR in all scenarios, especially for models 5 and 6, and to a lesser extent for models 3 and 4. Second, the presence of PC drastically reduces the power of both methods, especially for models 3-6. However, MB-MDR power estimates are always larger than MDR power estimates. Third, different choices of test approaches and p-value cut-offs p_c do not seem to have a large effect on MB-MDR power for most models. Exceptions are model 1 in PC scenario's, for which power for MB-MDR tends to be higher for higher p-value cut-offs p_c, Also, for model 5 and error-free data or data with induced missingness MB-MDR testing using $T = $ max ($|T_{H/LO}|$, $|T_{L/HO}|$)seems to perform worse than the other two test approaches for low values of p_c. In general, since power estimates are often very similar, some curves and bullets in Figure 3 are superimposed or difficult to distinguish.

The power improvement of the new methodology is especially relevant in the presence of GH, where MDR performs rather poorly (Figure 4). Figure 4 shows the power to identify both functional interacting pairs (black), the power to find the first interacting pair (red), and the power to retrieve at least one of the interacting pairs (green). As is to be expected, the more stringent the power definition, the lower the corresponding observed power. As in the absence of GH, power is generally higher for MB-MDR than for MDR. For models 1 and 2, even to identify both interacting pairs, MB-MDR has excellent power. The impact of different choices of test approaches and p-value cut-offs p_c is more explicit than could be observed from Figure 3. Preferred settings depend on the underlying genetic model. For instance, for model 1, high p-value cut-offs p_c lead to the highest power. For models 3 and 4, the $T = $ max ($|T_{H/LO}|$, $|T_{L/HO}|$) test approach is to be preferred. Finally, for models 5 and 6, a combination of low p_c and $T = |T_{H/L}|$ performs best.

Favoring a substantial power increase in settings with limited power to start with (models 5 and 6), we will now further investigate false positive error rate, power and specific power in the presence of combinations of error sources, with $p_c = 0.1$ and $T = |T_{H/L}|$ when MB-MDR analyses are involved. The false positive rate for the simulated null data (no association between genetic markers and trait) is 5.7% for MB-MDR and 5.5% for MDR. Table 1 gives the false positive rates of MB-MDR and MDR under several alternatives. For MB-MDR, we observe that these error rates average to the theoretical 5% level, an important check of the validity of our method. Interestingly, for MDR, false positive rates often exceed 10% (going even up to 24%), with the highest false positive rates being generated for models 3, 5 and 6.

Table 2 shows MB-MDR and MDR power to identify the true interacting pair(s) for data with different combinations of error sources. For simulated scenarios including GH, power refers to identifying both interacting pairs (see Table S1). In general, MB-MDR outperforms

MDR, with an up to 10-fold increase in power for model 1 in the presence of GH. For all models, the power increase is most dramatic in the presence of GH. Interestingly, for models 5 and 6, additional power is gained over MDR by following an MB-MDR strategy.

Tables 3 and 4 show the results of the alternative power definitions for data with GH. More specifically, Table 3 shows the power to correctly identify the first interacting pair, and Table 4 gives the power to correctly identify at least one of the interacting pairs (see Table S1). Again, power is seen to be much higher for MB-MDR than for MDR.

Specific power results are given in Supporting Information Tables S2-S4. Table S2 reports the specific power to detect the functional pair(s), both with and without GH. Tables S3 and S4 give alternative specific power evaluations under GH: the specific power to detect the first functional pair, and the specific power to detect at least one of the functional pairs. The results show that the different forms of specific power are indeed smaller than their non-specific counterparts in Tables 2-4. Furthermore, as to be expected, the observed decrease is bounded above by the false positive rates listed in Table 1. For completeness, we have also considered MB-MDR specific power for different p-value cut-offs $p_c = 0.05, 0.10, 0.20, 0.50, 1$ and different test approaches $T = |T_{H/L}|$, max $(|T_{H/LO}|, |T_{L/HO}|)$ and max $(|T_{H/L}|, |T_{H/LO}|, |T_{L/HO}|)$. Results for simulated scenario's not involving GH are visualized in Figure S1, while the GH results are depicted in Figure S2.

DISCUSSION

MB-MDR has a different way of combining multi-locus genotype cells using disease status than MDR. In particular, the concept of "no evidence" cells O is introduced, when no evidence for labeling the multi-locus genotype cells as high risk or low risk is found. The latter can be caused by a genuine lack of signal, or by insufficient power to make any reliable statements. MDR - and also extensions thereof, such as GMDR (Lou *et al.*, 2007) - do not allow for these indeterminate multi-locus genotype cells, whereas our results show that this category of cells is worthwhile being accounting for (Figure 2-3). Indeed, MB-MDR with $p_c = 1$, contrasting high risk cells versus low risk cells (test $T = |T_{H/L}|$) resembles a classical MDR setting in that all multi-locus genotypes are assigned to one of both risk groups, using disease status information. Not ignoring the "no evidence" O category seems to be particularly relevant for those epistasis models with low MAFs (models 5 and 6) and induced GH, giving rise to reduced power for p_c criterions tending to 1 (Figure 4). We have also introduced different test approaches $T = $ max $(|T_{H/L}|, |T_{H/LO}|, |T_{L/HO}|)$, max $(|T_{H/LO}|, |T_{L/HO}|)$ and $|T_{H/L}|$, and observed that the choice of test approach in combination with the p-value criterion p_c may affect power. The impact on power may slightly depend on the true underlying epistasis model. As stated already in the Results section, we favor a substantial power increase in settings with limited power to start with (for instance, models 5 and 6 with relatively low MAFs), and therefore we recommend to use MBMDR test approach $|T = T_{H/L}|$ with $p_c = 0.1$. The flexible framework of MB-MDR also allows introducing alternative definitions to cluster multi-locus genotype cells, which may increase MB-MDR power even further.

MB-MDR aims to identify the most significant associations (possibly more than one) between groups of markers and the trait of interest. In contrast, MDR identifies a single best model on the basis of measures of prediction accuracy and cross-validation consistency. Besides making it possible to detect multiple models, the use of association models in MB-MDR, rather than prediction accuracy and cross-validation consistency as in MDR, seems to be beneficial also in itself, in that it leads to a better performance, both in terms of controlling false positives and in terms of achieving adequate power, in most of the considered simulated settings (e.g., Tables 1-2 and S2). Certainly in the presence of GH, it is

essential to have a tool available that is able to identify several networks of markers that are significantly associated with the disease trait. The consequences of the somewhat restrictive property of MDR to only identify one best interaction model are also reflected in the simulation results. The outperformance of MB-MDR in the presence of GH (Tables 2-4 and S2-S4), is an important MB-MDR characteristic in the context of complex diseases that are likely to be driven by several interacting susceptibility genes, each with a mixture of rare and common alleles and genotypes. We emphasize that there is a conceptual difference in the way MB-MDR and MDR search for the 4 functional loci in case of GH. Whereas MB-MDR will retrieve this model by finding two significant pairs of loci, MDR will retrieve this model as a significant k-locus epistasis model ($k \geq 4$), even though no 4-order interactions are present. Hence, MB-MDR enables to better distinguish between different genetic models than MDR does, while recognizing two functional pairs rather than a more general 4-locus model. Also, MDR is more specific, in that finding the pairs (10, 4) and (5, 3) would not be considered a success in MB-MDR screening for the functional pairs (10, 5) and (4, 3), whereas MDR would not be able to make this distinction.

Different disease traits can be accommodated within the same framework offered by MB-MDR. Moreover, confounding factors, as well as lower-order genetic effects, can be accounted for in the interaction screening. MB-MDR can perform covariate corrections either a priori, by regressing out the covariates and taking the residuals to be the newly defined traits, or a posteriori, in the process of risk category assessment. MDR and MB-MDR inherently assume that the analysis is carried out in a sufficiently homogeneous population. However, population stratification is always a point of concern for case-control studies. Testing genetic effects may be biased by population admixture and stratification and may therefore affect the power and false positive rate of any proposed testing strategy. Because MB-MDR allows for covariate adjustment, population substructure characteristics can in principle be accounted during an MB-MDR screen (Devlin & Roeder, 1999).

In conclusion, the presented simulation results have illustrated that MB-MDR has increased power over MDR to identify gene-gene interactions for most considered genetic models, even in the presence of error sources. The presence of MS and/or GE hardly impact MB-MDR power, whereas PC and GH largely deteriorate power (Tables 2 and S2). Despite the power increase achieved by MB-MDR, it is hoped that alternative risk cell definitions will be able to better deal with PC, especially when external information (other than observed phenotypes) are used to label or "order" multi-locus genotype cells.

Both MDR and MB-MDR control false positive error rates to 5%, by permutation testing, under the null hypothesis of no association at all. In addition, MB-MDR controls false positives under any configuration of true and false null hypotheses, if the condition of subset pivotality is fulfilled. This assumption holds in the absence of LD between markers. Hence, it is not surprising that our results indeed demonstrate FWER control at 5%, also under the alternative hypothesis of epistasis, with or without genetic GH (Table 1). In contrast, MDR does not adequately control errors under the alternative hypothesis. Indeed, consider for simplicity a true underlying genetic epistasis model with one functional pair. Then it is hoped that the best MDR model is the one involving both functional loci and no others. Whenever a significant model does not contain both functional loci or contains a non-functional locus, this is a false positive result. With MDR, it is rather common that the functional pair is present in the best model, but not exclusively. The probability of this to occur is not controlled at 5% by the MDR permutation procedure. This explains the elevated false positive rates for MDR (Table 1), and also the apparent unbounded reduction in power when comparing, specific power (Tables S2-S4) with non-specific power (Tables 2-4). In contrast, for MB-MDR such a reduction is bounded to at most 5% on the average, by construction of the method.

The results of this study support the MB-MDR framework as a promising tool for detecting gene-gene interactions. MB-MDR applications to continuous traits (Cattaert *et al.*, 2010, Mahachie John *et al.*, 2009) and time-to-event data (in preparation) are just emerging, and power studies for a variety of scenarios with alternative outcome types, either univariate or multivariate, are on the way. However, in this post-genomic era, the genetic epidemiology community is most interested in having tools available that allow the researcher to screen hundreds of thousands of genetic markers for interactions with the trait(s) of interest. Although this study has restricted attention to 10 markers only, analyzing hundreds of markers with MB-MDR is feasible within a reasonable amount of time. For the present 200 cases and 200 controls, an MB-MDR 2-order screening of our 10 bi-allelic genetic markers used 0.82 MB memory and 0.26 seconds CPU time on an Intel(R) Xeon(R) CPU L5420 @ 2.50GHz processor, with the MB-MDR standard choices of critical value $p_c = 0.1$ and test approach $T = |T_{H/L}|$. An MDR screen for 1-5 order models needed 0.59 MB memory and 45.67 seconds CPU time on an Intel(R) Xeon(R) CPU E7330 @ 2.40GHz processor. On the other hand, an MDR screen restricted to 2-order models only used 0.35 MB memory and 1.02 seconds CPU time on the same platform. The latter constitutes a more honest comparison with MB-MDR in terms of computational resources, but obviously, by construction, such an MDR analysis will always fail to detect GH. Although pre-screening interesting clusters of markers for epistasis analysis have proven to be successful in large-scale genetic studies (Calle *et al.*, 2008a, Elbers *et al.*, 2009, Moore & White, 2007), a parallel version of MB-MDR, which enables to scale up the current implementation of MB-MDR to the GWA level without pre-screening, is on the way.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.
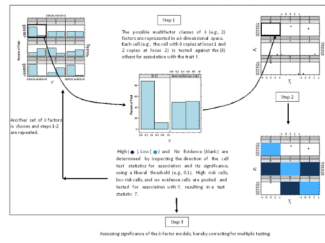
Acknowledgments

References

Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science. 2008; 322:881–8. [PubMed: 18988837]

Bellman, RE. Adaptive control processes: A guided tour. Princeton: Princeton University Press; 1961.

Calle ML, Urrea V, Vellalta G, Malats N, Van Steen K. Improving strategies for detecting genetic patterns of disease susceptibility in association studies. Stat Med. 2008a; 27:6532–46. [PubMed: 18837071]

Calle ML, Urrea V, Vellalta G, Malats N, Van Steen K. Model-Based Multifactor Dimensionality Reduction for detecting interactions in high-dimensional genomic data. U. O. V. Department of Systems Biology (ed). 2008b

Cattaert T, Urrea V, Naj AC, De Lobel L, De Wit V, Fu M, Mahachie John JM, Shen H, Calle ML, Ritchie MD, Edwards TL, Van Steen K. FAM-MDR: a flexible family-base multifactor

dimensionality reduction technique to detect epistasis using related individuals. Public Library of Science ONE. 2010

Chung Y, Lee SY, Elston RC, Park T. Odds ratio based multifactor-dimensionality reduction method for detecting gene-gene interactions. Bioinformatics. 2007; 23:71–6. [PubMed: 17092990]

Cordell HJ. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Human Molecular Genetics. 2002; 11:2463–2468. [PubMed: 12351582]

Cordell HJ. Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet. 2009; 10:392–404. [PubMed: 19434077]

Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999; 55:997–1004. [PubMed: 11315092]

Dixon MS, Golstein C, Thomas CM, Van Der Biezen EA, Jones JD. Genetic complexity of pathogen perception by plants: the example of Rcr3, a tomato gene required specifically by Cf-2. Proc Natl Acad Sci U S A. 2000; 97:8807–14. [PubMed: 10922039]

Elbers CC, Van Eijk KR, Franke L, Mulder F, Van Der Schouw YT, Wijmenga C, Onland-Moret NC. Using genome-wide pathway analysis to unravel the etiology of complex diseases. Genet Epidemiol. 2009; 33:419–31. [PubMed: 19235186]

Greene CS, Penrod NM, Williams SM, Moore JH. Failure to replicate a genetic association may provide important clues about genetic architecture. PLoS ONE. 2009; 4:e5639. [PubMed: 19503614]

Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. Bioinformatics. 2003; 19:376–82. [PubMed: 12584123]

Hardy J, Singleton A. Genomewide association studies and human disease. N Engl J Med. 2009; 360:1759–68. [PubMed: 19369657]

Liang Y, Kelemen A. Statistical advances and challenges for analyzing correlated high dimensional SNP data in genomic study for complex diseases. Statistics Surveys. 2008; 2:43–60.

Lou XY, Chen GB, Yan L, Ma JZ, Zhu J, Elston RC, Li MD. A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence. Am J Hum Genet. 2007; 80:1125–37. [PubMed: 17503330]

Ma DQ, Rabionet R, Konidari I, Jaworski J, Cukier HN, Wright HH, Abramson RK, Gilbert JR, Cuccaro ML, Pericak-Vance MA, Martin ER. Association and Gene-Gene Interaction of SLC6A4 and ITGB3 in Autism. American Journal of Medical Genetics Part B-Neuropsychiatric Genetics. 2010; 153B:477–483.

Mahachie John JM, Baurecht H, Rodríguez E, Naumann A, Wagenpfeil S, Klopp N, Mempel M, Novak N, Bieber T, Wichmann HE, Ring J, Illig T, Cattaert T, Van Steen K, Weidinger S. Analysis of the high affinity IgE receptor genes reveals epistatic effects of FCER1A variants on eczema risk. Allergy. 2009

Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest. 2008; 118:1590–605. [PubMed: 18451988]

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, Mccarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, Mccarroll SA, Visscher PM. Finding the missing heritability of complex diseases. Nature. 2009; 461:747–53. [PubMed: 19812666]

Marnellos G. High-throughput SNP analysis for genetic association studies. Curr Opin Drug Discov Devel. 2003; 6:317–21.

Mckinney BA, Reif DM, Ritchie MD, Moore JH. Machine learning for detecting gene-gene interactions: a review. Appl Bioinformatics. 2006; 5:77–88. [PubMed: 16722772]

Moore JH. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum Hered. 2003; 56:73–82. [PubMed: 14614241]

Moore JH, Gilbert JC, Tsai CT, Chiang FT, Holden T, Barney N, White BC. A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. J Theor Biol. 2006; 241:252–61. [PubMed: 16457852]

Moore JH, White BC. Tuning ReliefF for Genome-Wide Genetic Analysis. Lecture Notes in Computer Science. 2007; 4447:166–175.

Motsinger AA, Ritchie MD, Reif DM. Novel methods for detecting epistasis in pharmacogenomics studies. Pharmacogenomics. 2007; 8:1229–41. [PubMed: 17924838]

Musani SK, Shriner D, Liu N, Feng R, Coffey CS, Yi N, Tiwari HK, Allison DB. Detection of gene x gene interactions in genome-wide association studies of human population data. Hum Hered. 2007; 63:67–84. [PubMed: 17283436]

Onkamo P, Toivonen H. A survey of data mining methods for linkage disequilibrium mapping. Hum Genomics. 2006; 2:336–40. [PubMed: 16595078]

Pae CU, Drago A, Forlani M, Patkar AA, Serretti A. Investigation of an Epistatic Effect Between a Set of TAAR6 and HSP-70 Genes Variations and Major Mood Disorders. American Journal of Medical Genetics Part B-Neuropsychiatric Genetics. 2010; 153B:680–683.

Park MY, Hastie T. Penalized logistic regression for detecting gene interactions. Biostatistics. 2008; 9:30–50. [PubMed: 17429103]

Ritchie MD, Edwards TL, Fanelli TJ, Motsinger AA. Genetic heterogeneity is not as threatening as you might think. Genetic Epidemiology. 2007; 31:797–800. [PubMed: 17654613]

Ritchie MD, Hahn LW, Moore JH. Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genet Epidemiol. 2003; 24:150–7. [PubMed: 12548676]

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet. 2001; 69:138–47. [PubMed: 11404819]

Ruczinski I, Kooperberg C, Leblanc ML. Exploring interactions in high-dimensional genomic data: an overview of Logic Regression, with applications. Journal of Multivariate Analysis. 2004; 90:178–195.

Seng KC, Seng CK. The success of the genome-wide association approach: a brief story of a long struggle. Eur J Hum Genet. 2008; 16:554–64. [PubMed: 18285837]

Sonoda T, Suzuki H, Mori M, Tsukamoto T, Yokomizo A, Naito S, Fujimoto K, Hirao Y, Miyanaga N, Akaza H. Polymorphisms in estrogen related genes may modify the protective effect of isoflavones against prostate cancer risk in Japanese men. European Journal of Cancer Prevention. 2010; 19:131–137. [PubMed: 19952760]

Van Steen, K.; Molenberghs, G. Multicollinearity. In: Encyclopedia of Biopharmaceutical Statistics. In: Chow, S-C., editor. Encyclopedia of Biopharmaceutical Statistics. London: Informa Healthcare; 2004.

Vancleave TT, Moore JH, Benford ML, Brock GN, Kalbfleisch T, Baumgartner RN, Lillard JW, Kittles RA, Kidd LCR. Interaction Among Variant Vascular Endothelial Growth Factor (VEGF) and Its Receptor in Relation to Prostate Cancer Risk. Prostate. 2010; 70:341–352. [PubMed: 19908237]

Westfall, PH.; Young, SS. Resampling-based multiple testing. New York: Wiley; 1993.

Figure 1. Graphical overview of major MB-MDR steps

Figure 2. Penetrance functions of simulated data of (Ritchie et al., 2003)
Multilocus penetrance functions and MAFs used to simulate case-control data exhibiting
gene-gene interactions in the absence of main effects.
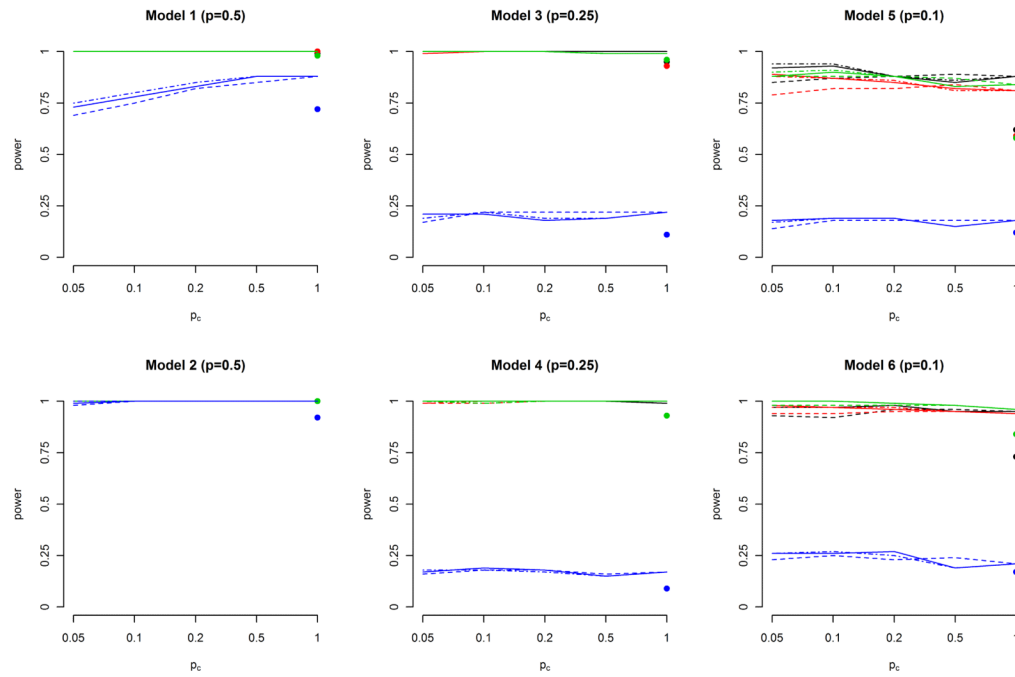
Figure 3. MB-MDR and MDR power with different sources of noise, excluding genetic heterogeneity

The 6 plots display MB-MDR power estimates to identify the correct interacting pair for models 1-6, for different p-value cut-offs p_c = 0.05,0.1,0.2,0.5 and 1. The color coding is as follows: error-free data (black), data with induced missingness (red), genotyping errors (green) and phenocopy (blue). The line types refer to the different MB-MDR testing strategies used: $T = |T_{H/L}|$ (solid line), max ($|T_{H/LO}|,|T_{L/HO}|$) (dashed line) and max ($|T_{H/L}|,|T_{H/LO}|,|T_{L/HO}|$) (dot-dashed line). MDR power estimates of screening over 1-5 order models are also shown (bullets at p_c=1).
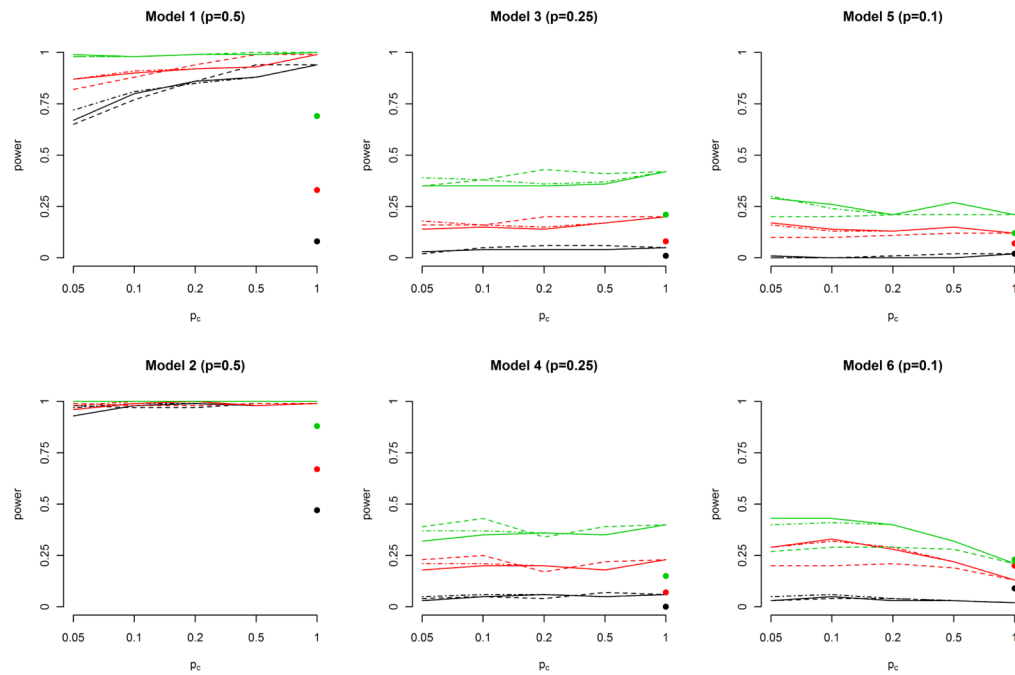
Figure 4. MB-MDR and MDR power in the presence of genetic heterogeneity

The 6 plots display MB-MDR power estimates for models 1-6, for different p-value cut-offs $P_c = 0.05, 0.1, 0.2, 0.5$ and 1. The color coding is as follows: power to identify both interacting pairs (black), the first interacting pair (red), and at least one of the interacting pairs (green). The line types refer to the different MB-MDR testing strategies used: $T = |T_{H/L}|$ (solid line), max $(|T_{H/LO}|, |T_{L/HO}|)$ (dashed line) and max $(|T_{H/L}|, |T_{H/LO}|, |T_{L/HO}|)$ (dot-dashed line). MDR power estimates of screening over 1-5 order models are also shown (bullets at P_c=1).

Table 1

MB-MDR and MDR false positive rates (%) for data subject to different error sources.

Error	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	MB	MDR	MB	MDR	MB	MDR	MB	MDR	MB	MDR	MB	MDR
None	6	9	4	5	6	17	5	13	5	21	5	23
GE	2	14	2	3	6	18	11	11	5	21	4	23
GH	4	7	9	2	7	14	5	6	4	8	2	17
PC	6	8	6	13	3	4	3	3	5	8	3	11
MS	7	16	4	3	6	18	8	9	8	21	7	24
GE+GH	5	9	11	2	7	14	5	6	5	8	2	8
GE+PC	5	6	10	12	3	9	5	4	5	7	6	11
GE+MS	4	12	2	3	6	19	13	14	4	22	4	23
GH+PC	4	4	7	10	5	2	3	6	5	5	9	6
GH+MS	4	14	9	4	5	0	4	11	5	4	6	11
PC+MS	1	7	2	6	7	10	3	5	6	11	7	13
GE+GH+PC	7	5	8	6	8	8	2	5	6	3	7	5
GE+GH+MS	3	15	9	5	6	8	6	5	6	9	1	4
GH+PC+MS	2	9	5	4	6	7	3	6	3	3	6	4
GE+PC+MS	6	10	5	12	5	9	10	10	7	10	4	12
GE+GH+PC+MS	3	3	4	6	5	4	8	8	5	7	7	5

Family-wise error rates (FWER) are shown for MB-MDR (MB) with $p_C = 0.1$ using the T $= |T_H/L|$ test approach and MDR screening over 1-5 order models.

GE= genotype error; GH=genetic heterogeneity; PC=phenocopy; MS=missing data

Table 2

MB-MDR and MDR power (%) to identify the correct interacting pair(s) with different errors.

Error	Model 1 MB	Model 1 MDR	Model 2 MB	Model 2 MDR	Model 3 MB	Model 3 MDR	Model 4 MB	Model 4 MDR	Model 5 MB	Model 5 MDR	Model 6 MB	Model 6 MDR
None	100	99	100	100	100	95	100	93	93	62	97	73
GE	100	98	100	100	100	96	100	93	90	58	100	84
GH	80	8	98	47	4	1	5	0	0	2	5	9
PC	78	72	100	92	21	11	19	9	19	12	26	17
MS	100	100	100	100	100	93	100	93	87	59	97	84
GE+GH	75	5	98	47	3	1	5	0	0	2	7	9
GE+PC	80	69	99	95	17	16	18	7	21	7	35	15
GE+MS	100	100	100	100	100	95	100	95	86	60	96	71
GH+PC	2	0	3	0	0	0	0	0	1	0	1	0
GH+MS	74	11	95	42	3	0	4	2	1	2	3	5
PC+MS	82	71	98	95	21	13	19	9	17	12	24	14
GE+GH+PC	1	0	3	1	0	0	0	0	0	0	0	0
GE+GH+MS	64	8	91	42	2	1	4	1	1	1	4	6
GH+PC+MS	0	0	0	0	0	0	0	0	0	0	0	0
GE+PC+MS	84	70	98	95	21	16	20	25	9	7	16	13
GE+GH+PC+MS	1	0	3	0	0	0	0	0	0	1	0	1

Results are shown for MB-MDR (MB) with $p_C = 0.1$ using the T = $|T_H/L|$ test approach, and for MDR screening over 1-5 order models.

GE= genotype error; GH=genetic heterogeneity; PC=phenocopy; MS=missing data

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

MB-MDR and MDR power (%) to identify the first interacting pair in the presence of genetic heterogeneity.

Error	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	MB	MDR	MB	MDR	MB	MDR	MB	MDR	MB	MDR	MB	MDR
GH	90	33	99	67	15	8	20	7	14	7	33	20
GE+GH	89	36	99	67	16	8	21	7	13	7	26	12
GH+PC	6	5	18	11	1	1	0	3	2	1	2	4
GH+MS	89	28	97	68	14	0	14	10	13	6	17	10
GE+GH+PC	5	3	17	12	1	2	0	1	0	0	0	1
GE+GH+MS	81	31	97	66	16	8	21	9	11	7	23	9
GH+PC+MS	7	7	16	7	2	2	1	4	0	1	0	0
GE+GH+PC+MS	5	5	14	5	1	1	2	0	0	2	0	2

Results are shown for MB-MDR (MB) with $p_C = 0.1$ using the T = $|T_H/L|$ test approach, and for MDR screening over 1-5 order models.

GE = genotype error; GH = genetic heterogeneity; PC = phenocopy; MS = missing genotypes

Table 4

MB-MDR and MDR power (%) to identify at least one of the two interacting pairs in the presence of genetic heterogeneity.

Error	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	MB	MDR	MB	MDR	MB	MDR	MB	MDR	MB	MDR	MB	MDR
GH	98	69	100	88	35	21	35	15	26	12	43	30
GE+GH	98	69	100	88	36	21	36	15	24	12	43	23
GH+PC	12	6	30	22	4	3	1	6	5	3	5	5
GH+MS	98	62	100	92	33	0	24	18	25	11	34	18
GE+GH+PC	10	7	28	21	2	4	0	4	5	2	3	4
GE+GH+MS	95	63	100	93	33	17	35	18	20	15	38	10
GH+PC+MS	13	13	36	16	3	2	3	5	3	3	4	1
GE+GH+PC+MS	12	11	27	15	4	3	6	2	0	4	2	3

Results are shown for MB-MDR (MB) with p_C =0.1 using the T = $|TH/L|$ test approach, and for MDR screening over 1-5 order models.

GE= genotype error; GH=genetic heterogeneity; PC=phenocopy; MS=missing genotypes