npg

ARTICLE

Multilocus statistics to uncover epistasis and heterogeneity in complex diseases: revisiting a set of multiple sclerosis data

Stefan Böhringer[*,1,2], Cornelia Hardt[3], Bianca Miterski[4], Ansgar Steland[2] and Jörg T Epplen[4]

[1]*Institut für Humangenetik, Universitätsklinikum Essen, Germany;* [2]*Department of Statistics, Faculty of Mathematics, Ruhr-Universität Bochum, Germany;* [3]*Institut für Immunologie, Universitätsklinikum Essen, Germany;* [4]*Human Genetics, Ruhr-Universität Bochum, Germany*

New statistics are developed to gather the contribution of many alleles at different loci to common diseases. Both inferential and descriptive statistics are included in order to uncover epistatic effects as well as heterogeneity. The problem of multiple testing is circumvented by considering a global null hypothesis. Global testing is supplemented by descriptive methods that make use of measures like odds ratio or the *P*-value of individually tested allele combinations. Visualization helps to reflect complex data sets. The methods described here have been scrutinized by statistical simulations, and we show that power gains can be substantial as compared to single locus statistics. Typing data of multiple sclerosis patients and controls are investigated, representing an example of larger scale information in screening candidate genes for their impact on complex diseases. New insights emerge from this data set demonstrating genetic heterogeneity and evidence for epistasis.

Introduction

Association studies have received substantial interest in the advent of the genomic era. In complex diseases, association studies can offer advantages over linkage studies, both, from a statistical and practical point of view.[1–5] However, problems with regard to population stratification and failure to reproduce results have drawn criticism and have led to several guidelines on how to conduct statistical analyses of association studies.[6–8] A major issue in association studies is that of studywide statistical significance (ie problem of multiple testing[9–11]). Association

studies are often performed by screening many loci in patients and controls and evaluating the frequencies for each locus separately. Patients are then stratified for the predisposing allele of a certain locus and analysed for allele frequencies of a second locus. By this approach, the number of comparisons increases rapidly and has to be considered in the statistical analysis. A stringent correction of *P*-values (to attain a studywide *P*-value by procedures such as Bonferroni's or Holm's[12]) can lead to reduction in power since correction procedures tend to be conservative and may, therefore, hamper reproducibility.

Single locus statistics to test for genetic association with disease make use of several statistical models. Choices are comparisons of allele, phenotype (allele carrier) or genotype frequencies. All these models have the advantage of being analysed in contingency tables where a χ^2 or Fisher statistics can be applied. The use of general linear models has been proposed,[13] for which most models have no direct

Correspondence: Dr S Böhringer, Institut für Humangenetik, Universitätsklinikum Essen, Hufelandstr. 55, IG1, 45122 Essen, Germany. Tel: +49 201 723 4681; Fax: +49 201 723 5900;
E-mail: stefan.boehringer@uni-essen.de

correspondence in classical genetic association studies. We have developed a statistic to test the following null hypothesis: no allele combinations (ACs) comprising alleles from a fixed number of loci are associated with a given disorder. Given such a set of loci, we define an AC of an individual to be a set of alleles which contains exactly one allele from each genotype of each locus. Therefore, if l loci are considered 2^l different ACs can be formed if the individual is heterozygous at each locus. ACs represent potential haplotypes if the loci are confined to the same chromosomal region. The number of loci considered for these sets is subject to practical considerations that are discussed below. The disease-modifying effects of ACs can be modelled by considering extensions of several genetic models to many loci. Rejection of the null hypothesis shows the existence of an AC to be associated with disease, that is, no particular AC is designated. Descriptive methods are implemented, including tests of individual ACs with the possibility of performing Bonferroni correction. Odds ratios (ORs) for all ACs are computed and highlighted graphically. These ORs are computed in terms of AC frequencies, which can be computed in several ways in the multilocus case. We define these frequencies in the Material and methods section. The graphical representation of ORs and other measures of association gives a concise summary of complex data sets and has proven useful for both simulated and real data sets. We apply a normalization of ORs of combinations by comparing each OR with its expectation based on ORs of single alleles. We used a simulated data set to illustrate the problems involved and show that power gains can be substantial when comparing our statistics to a conventional analysis. Descriptive methods may be effective in revealing interaction between several loci. We then applied the test statistics and descriptive methods to a case–control study for multiple sclerosis (MS).

Material and methods

When deriving ACs for an individual that is homozygous at some of the loci being investigated several identical ACs are observed. The identical ACs can be weighted in different ways. The possibilities investigated here are illustrated in Figure 1. In the additive case, all ACs are formed for each individual irrespective of whether they are identical or not and are used to compute AC frequencies in patients and controls. In the single count model ACs derived for each individual are counted at most once. These two models differ in how homozygous individuals are weighted. If evaluated for single loci the additive and single count models correspond to allele and phenotype frequencies, respectively. Genotypes are directly analogous in the single and multilocus cases. To quantify epistatic
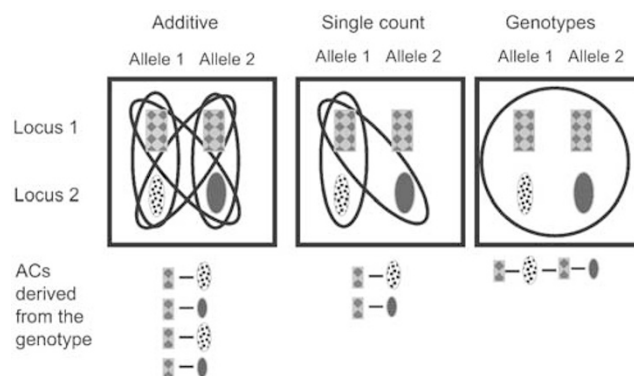

Figure 1 Genetic models as applied and extended to multiple loci. The additive model takes into account all possible combinations of alleles. The single-count model counts any particular AC at most once for an individual and the genotype model takes into account full genotypes at each locus.

effects we consider the quotient of observed and predicted ORs which is called normalized OR (nOR) hereafter (cf Appendix A). The statistical test can be performed for any of the genetic models. As a result of multiple interdependencies in the data set, a bootstrap simulation procedure was applied (cf Appendix A). The data set used for re-evaluation contains unrelated MS patients and healthy control subjects. These groups are matched ethnically and have been used in several association studies so far.[14–16] Details of clinical MS parameters and other group characteristics have been given previously.[14] The most basic information is summarized in Table 1. The data set contains 77 markers, 1187 MS patients and 524 control subjects. 60992 and 55498 alleles have been typed in the groups, respectively.

Simulation study

We have simulated our statistics for several scenarios to evaluate its power. A data set of variable size was generated according to a coalescent process.[17,18] A total of 20 unlinked loci were simulated mimicking the situation of an association study. To consider different alternative hypotheses, we picked an AC comprising an allele from the first and second locus and divided the data set into patients and controls generating deterministic frequencies in two groups (cf Appendix A). Frequencies were determined according to a single count model. The differences in the frequencies between the groups as well as the sizes of the cohorts were varied giving rise to several different simulated situations. In each case, a test was applied to detect association of ACs comprising two alleles.

Table 1 Properties of the MS patients and the control group

	All MS patients	Subgroup with primary progressive course	Subgroup with relapsing remitting or secondary progressive course	Control group
Number	1187	178 (15%)	738 (62.2%)	524
Age of onset/Age	30.12 ($\pm$9.71; 11–74)**	34.1 ($\pm$10.5; 11–57)**	29 ($\pm$9; 11–69)**	39.3 ($\pm$11.5; 20–70)**
Enhanced disability status scale (EDSS)	4.2 ($\pm$2.3; 0–9.5)**	5.8 ($\pm$1.9; 0.5–9)**	3.9 ($\pm$2.2; 0–9.5)**	
EDSS per year (EDSSY)	0.9 ($\pm$1.22; 0–9)**	0.92 ($\pm$1.1; 0.04–7)**	0.86 ($\pm$1.2; 0–8)**	

**Standard deviations and the range are given within parantheses.

Results
Simulation study

Results from the simulation study are summarized in Figure 2. Differences in AC frequencies between patients and controls of 0.21 (Figure 2a) and 0.11 (Figure 2b) were evaluated. As can be seen, power is excellent for a difference of ≥ 0.21, when a number of 400 individuals is sufficient to reach power >0.8. Power drops below 0.8 when differences are smaller than 0.11 and sample sizes are smaller than 800 individuals (Figure 2b).

Furthermore, we have evaluated the effect of choosing the wrong formal genetic model. We have simulated data as above, but then used single locus statistics to evaluate the simulated data sets. Figure 2c shows that power is reduced dramatically. Power of >0.8 cannot be achieved for <1000 individuals when the difference of the AC frequency is 0.11 between the groups. Figure 3 shows descriptive analyses of a simulated data set with 60 loci. This data set was produced as for the power simulations. The size of each spot represents the maximal OR > 1 of all ACs at the respective locus pair in the lower left part, by including only ACs with at least 50 observations. The minimal OR < 1 is shown in the upper right part. As can be seen, a strong association between loci 1 and 2 is evident, together with spurious associations of loci 1 and 2 with other loci. Also, a considerable background of false positives is present (Figure 3a). The background can be reduced by selecting ACs more stringently for inclusion into the descriptive analysis. Figure 3b represents such an analysis, which requires at least 80 observations, and the causative AC from locus 1 and 2 can be readily identified. To address the problem of spurious associations of alleles from other loci with a predisposing allele at either locus 1 or 2, ORs were normalized as described in Material and methods. As shown in Figure 3c (cf Appendix A) spurious associations could be eliminated. In contrast, single-allele analysis shows strong associations for many loci (Figure 3d).

MS case–control study
Table 2 lists the loci included in this study. Table 3 lists results from global hypothesis testing. Clearly, the results

are highly significant. Descriptive results for two-way interactions are shown in Figure 4. Results for two different genetic models, additive (Figure 4a) and single count (Figure 4b) are presented. Results for the different models are similar, yet there are noteworthy differences. For example, associations of the markers NFKBIA-2 through NFKBIA-5 with the marker D18S35 seem to be stronger for the additive than for the single-count model. In contrast, the TNF-1/FGF1 association from the single-count model is absent for the additive model. In general, several ACs are striking and fit into pathogenetic hypotheses. We point to combinations which include the allele IFNA-1:07. Associations for the single alleles were reported previously.[14,15] ACs including this predisposing allele and a certain allele of another locus seem to be strongly associated with MS. However, the display of nOR (Figure 4c; single-count model) shows that most of these associations can be explained by the individual effect of the IFNA-1:07 allele. ACs with strong disease association for which single alleles have weak effects include TCRB-5/HLA-DRB1, TCRB-5/NFKB1-2, TNF-1/HLA-DRB1 and HLA-DRB1/D18S364. These ACs display presumed epistatic effects, which correspond to large nORs (Figure 4c). As defined by testing ACs in a contingency table individually, the most significant ACs are iterated in Table 4 for the single count model.

Results for the single-count model in primary progressive MS (PPMS) are shown in Figure 4d. Differences between the group of all MS patients and that of PPMS are apparent. Comparing Figures 4b and d, HLA-DRB1 appears to be more important in predisposing to PPMS. Also, the locus D18S41 is more strongly associated in PPMS patients than in all MS patients. However, no epistatic effects are present as revealed by nOR analysis (data not shown). Epistatic effects are demonstrable for the combinations TNFRSF1A-1/NFKBIA-11, TNFRSF1A-2/NFKBIA-11 and TNFRSF1A-2/NFKBIL1-1. These are only present for PPMS. More descriptive results including lists of ACs and graphical analyses for other MS groups/parameters are presented on a supplementary website (http://www.s-boehringer.de/cd). All software used for this paper is available for downloading.

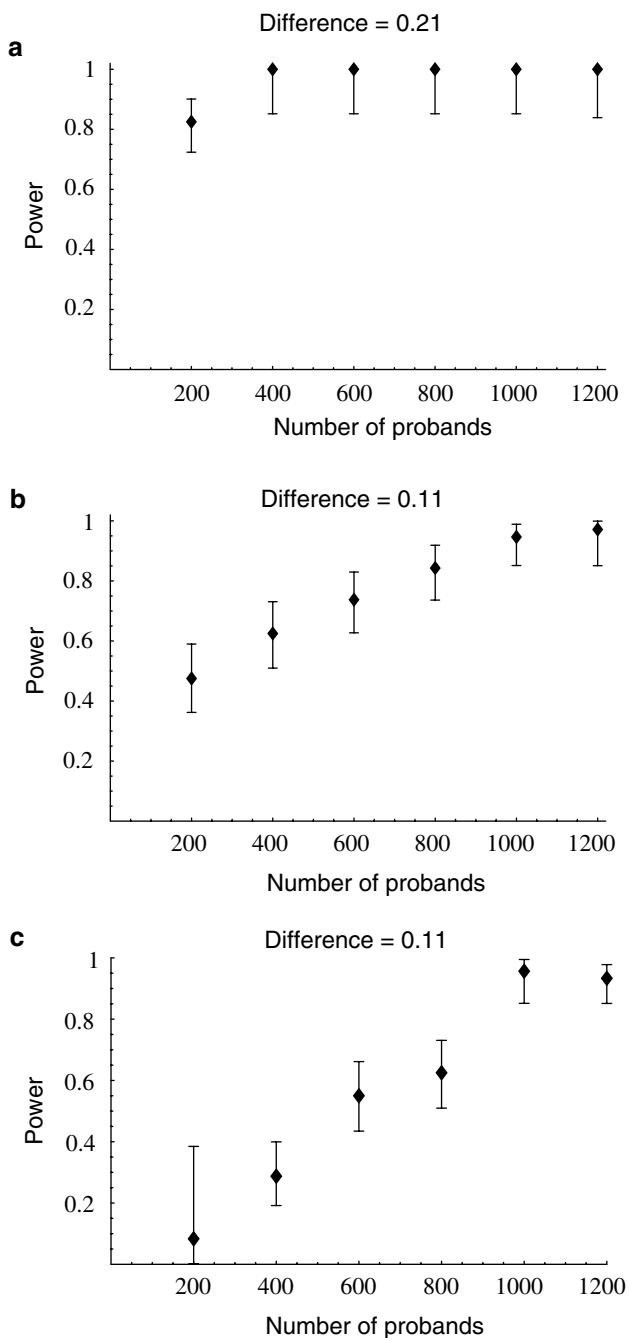**Multilocus statistics for complex diseases**
S Böhringer et al
576

Figure 2 Power simulations of a global null hypothesis. Power simulations for the test statistics under the alternative of one AC of size 2 being associated with the disease. Difference (either 0.21 or 0.11; a, b) denotes the surplus fraction of cases bearing the associated AC. Controls are fixed to a 19% frequency of the predisposing AC (**a, b**). Power simulations under misspecification of the inheritance model (**c**) (for further explanations see text). When a single AC of size 2 is associated with the disease, a test statistics for single alleles is employed.

Discussion

Our simulation study has shown that statistical power to detect genetic associations can be excellent whenever a global null hypothesis is considered. This approach circumvents the need for multiple testing at the cost of not readily identifying the presumably causative ACs. Therefore, descriptive methods are needed to weigh individual ACs. Several descriptive measures exist among which we used P-values of individual tests, ORs and nORs. Each measure generates a different aspect of the data set, which places a caveat on each descriptive measure. The simulated data have shown that ACs may be identified by descriptive measures if all differences between cases and controls are concentrated in a single AC. If several ACs have small contributions each, as is to be expected in a complex disorder, descriptive measures are less likely to give clear evidence for single factors. The test itself is robust against heterogeneity since a global null hypothesis is tested and all differences between the groups are summed up to produce a single test statistics. No exact rules can be given of how to identify single ACs relevant to disease predisposition. However, several patterns can hint at relevant ACs. The pattern exhibited by *INFA1:07* shows spurious associations with other alleles in the absence of epistatic effects, which is because of strong associations from the single allele. This effect could be revealed by considering nORs.

As shown previously,[15] the phenotype frequency for *IFNA1:07* was 6.3% (OR 12.41, $P_c < 8 \times 10^{-4}$). The combination *TCRB-5/HLA-DRB1*, which corresponds to *TCRBV6S3*2* and *HLA-DRB1*03*, has been described previously.[14] The OR for *TCRBV6S3*2* was 2.72 ($P_c < 0.006$), for *HLA-DRB1*03* it was 1.42 ($P_c < 0.8$) and for the AC *TCRBV6S3*2/HLA-DRB1*03* it was 22.03 ($P_c < 5 \times 10^{-3}$) comparing with an OR of 23.91 ($P_c = 3.64^{-2}$) in Table 4 (*HLA-DRB1:03/TCRB-5:02*; ie *HLA-DRB1:03/TCRB-5:02*) ($P_c = 3.64^{-2}$). The published data correspond to the single-count model (phenotype frequencies) and differs slightly due to inclusion of few extra probands. In this example, the additive model generates similar results, since the number of individuals homozygous for the relevant HLA-DRB1 allele or the relevant TCRB allele is extremely low. In addition, this example demonstrates that the statistical analysis presented here is in concordance with previous evaluations but it is more efficient, since multiple comparisons can be made in a single step. Moreover, allele combinations that may be overlooked because of borderline significance of certain alleles from single loci will be detected by comprehensive descriptive analysis. Taking into account P-values, ORs and nORs simultaneously may suggest interesting candidates.

In our statistics, a full model of a given complexity is considered (say all pairs of loci). This is desirable when sizable data sets are under scrutiny (say up to about hundred loci). However, the number of ACs increases
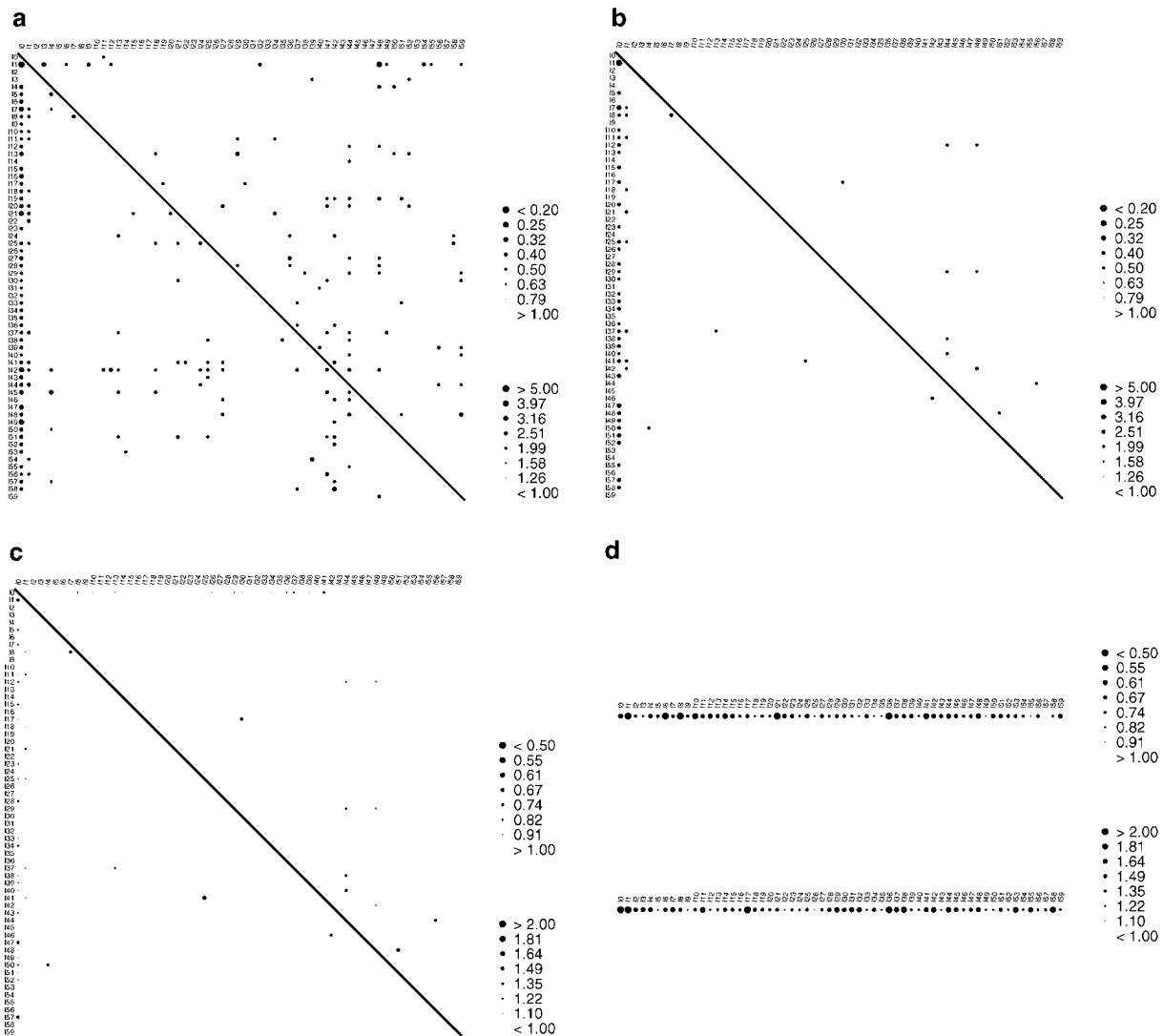
Figure 3 Two-way interaction in a simulated data set. Maximal (bottom left) and minimal ORs (top right, **a, b**) of ACs for a simulated data set according to the single-count model. The size of each point corresponds to the value of the OR according to a logarithmic scale. The minimal number of observations for an AC to be included is 50 (**a, c**) and 80 (**b**), respectively. nORs are shown in part (**c**) Minimal ORs (top) and maximal ORs (bottom) of single alleles for the data set are shown in part (**d**).

polynomially with the number of loci and exponentially with the number of alleles contained in each AC. This imposes a practical limitation for applying this test in terms of calculation time. The number of alleles contained in each AC is a parameter which is to be chosen carefully. The bootstrap procedures assume that AC frequencies can be estimated with high accuracy (ie the error in AC frequency estimation is only partially accounted for in the test statistics). We have, therefore, limited our analyses to combinations of two alleles. If SNPs are used throughout, ACs of size three could be realistically screened in case less than about 1000 probands are present in each group.

One issue raised in case/control designs is the problem of population stratification. Considering several loci simulta-neously can accentuate this confounder. In our simulation study we have assumed no stratification. The MS group set is closely matched with the control group, ethnically.[14] Nevertheless, population stratification remains an issue and certain applications may require our statistics to allow for stratification. Some authors have proposed solutions to the stratification problem.[19–22] Among these are general methods to estimate a degree of memberships to population substrata for each individual.[21] These values can be used to test for association in each substratum and combine these statistics to a global statistics. Another issue raised by association studies in complex diseases is that of inferential *vs* descriptive statistics. We have put forward the notion to look at association studies from a descriptive

Table 2 Results of statistical tests to infer genetic association in the MS data set

Model	Cardinality	Test statistics	E [T]	V [T]	P-Value
Codominant	1	850.191	265.328	761.022	< 5E-5
Additive	1	1006.45	309.813	1149.43	< 5E-5
Genotypes	1	1867.89	890.886	4284.58	< 5E-5
Codominant	2	90973.9	44669.6	1.18e+07	< 5E-5
Additive	2	145245	69453.5	3.21e+07	< 5E-5
Genotypes	2	207192	122809	5.41e+07	< 5E-5

Cardinality denotes the number of alleles considered in each AC. *Test statistics* results from calculating the test statistics from the MS data set. *E* and *V* are estimations for the distribution of the statistics as derived from drawing from the control data set. *P-value* is the empirical p-value derived from the bootstrap procedure.

Table 3 Loci included in this study of genetic MS association

Abbreviation	Type	Description	Localisation	MIM/Ref.
BCL2	SNP	B-cell leukaemia 2, exon 1 T7T (g/a)	18q21.3	151430
BCL3-1	Mic	B-cell leukaemia/lymphoma 3	19q13.1–q13.2	109560
BCL3-2	SNP	BCL3 exon 7 polymorphism N217N (c/t)	19q13.1–q13.2	109560
CD28	Mic	CD28 antigen (Tp44)	2q33	186760
CD3D	Mic	CD3D antigen	11q23	186760
CD4	Mic	CD4 antigen (p55)	12pter–12p	186940
CTLA4-1	SNP	Cytotoxic T-lymphocyte-associated protein 4, exon 1 T17A (c/t)	2q33	123890
CTLA4-2	SNP	CTLA4 promotor –318 (a/g)	2q33	123890
CTLA4-3	HT	Haplotype of loci CTLA4-1, CTLA4-2	2q33	123890
D18S35	Mic	Postulated association with IDDM	18q21	
D18S364	Mic	s. D18S364	18q21	
D18S41	Mic	s. D18S41	18q21	
D4S1628	Mic	Microsatellite NFKB region	4q22	
D4S1647	Mic	s. D4S1628	4p14	
D4S242	Mic	s. D4S1628	4q22	
FGF1	Mic	Fibroblast growth factor 1 (acidic)	5q31.3–q33.2	131220
HLA-DRB1	HT	Human leucocyte antigen, DRB1 locus	6p21.3	142857
ICAM4	SNP	Intercellular adhesion molecule 4	19p13.2–cen	111250
IFNA1	Mic	Interferon α1	9q22	147660
IFNA10	SNP	Interferon α10 C20X (t/a)	9q22	147577
IFNA17-1	SNP	Interferon α17 171insA	9q22	147583
IFNA17-2	SNP	IFNA17 I184R (t/g)	9q22	147583
IFNB1	SNP	Interferon β1 Y51Y (c/t)	9q22	147640
IFNG	Mic	Interferon γ	12q24.1	147570
IFNAR1	Mic	Interferon receptor 1	21q22.1	107450
IL1A	Mic	Interleukin 1α	2q12–q21	147760
IL10	Mic	Interleukin 10	1q31–q32	124092
IL1RN	Mic	Interleukin 1 receptor antagonist	2q14.2	147679
IL2	Mic	Interleukin 2	4q26–q27	147680
IL2RA	Mic	Interleukin 2 receptor, α chain	10p15–p14	147730
IL5RA	Mic	Interleukin 5 receptor, α chain	3p26–p24	147851
IRF1	Mic	Interferon regulatory factor 1	5q23–q31	147575
IRF2	Mic	Interferon regulatory factor 2	4q34.1–q35.1	147576
LST1	SNP	Lymphocyte specific transcript1 +38a/g	6q21.3	
LTA-1	SNP	Lymphotoxin α (LTA; TNF super family, member 1), +11 (g/a)	6p21.3	153440
LTA-2	SNP	LTA +81 (a/c)	6p21.3	153440
LTA-3	HT	Haplotype LTA-1/LTA-2	6p21.3	153440
LTA-4	SNP	LTA exon 1	6p21.3	153440
LTA-5	SNP	LTA H696P (a/c)	6p21.3	153440
LTA-6	SNP	LTA N723T (c/a)	6p21.3	153440
LTA-7	HT	Haplotype LTA-5/LTA-6	6p21.3	153440
NFKB1-1	SNP	Nuclear factor of κ light polypeptide gene enhancer in B-cells 1 (p105), exon 12 A380A (c/t)	4q24	164011
NFKB1-2	SNP	NFκB1 exon 17 L616F (g/t)	4q24	164011
NFKBIA-1	SNP	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha, -420 c/t	14q13	164008
NFKBIA-2	SNP	NFκBIA promotor -708ins8	14q13	164008

Table 3 (continued)

Abbreviation	Type	Description	Localisation	MIM/Ref.
NFKBIA-4	SNP	NFκBIA promotor polymorphism, -1001 a/g	14q13	164008
NFKBIA-5	SNP	NFκBIA promotor polymorphism, -1169 a/g	14q13	164008
NFKBIA-6	SNP	NFκBIA promotor polymorphism, -1256 c/t	14q13	164008
NFKBIA-7	SNP	NFκBIA D27D c/t	14q13	164008
NFKBIA-8	SNP	NFκBIA A102A c/t	14q13	164008
NFKBIA-9	SNP	NFκBIA intron 3 g49a	14q13	164008
NFKBIA-10	SNP	NFκBIA intron 3 g262a	14q13	164008
NFKBIA-11	SNP	NFκBIA exon 6 3′UTR t2c	14q13	164008
NFKBIL1-1	SNP	Nuclear factor of κ light polypeptide gene enhancer in B-cells inhibitor-like 1, exon 4 C225R (c/t)	6p21.3	601022
NFKBIL1-2	SNP	NFκBIL1 promotor polymorphism, g/a	6p21.3	601022
NFKBIL1-3	SNP	NFκBIL1 promotor polymorphism, a/t	6p21.3	601022
NFKBIL1-4	HT	Haplotype NFκBIL1 -2, NFκBIL-3	6p21.3	601022
TNFRSF1A-1	SNP	Tumour necrosis factor receptor super family, member 1A, exon 1 P12P (g/a)	12p13.2	191190
TNFRSF1A-2	SNP	TNFRSF1A promotor polymorphism, -609 g/t	12p13.2	191190
TNFRSF1B	SNP	Tumour necrosis factor receptor super family, member 1B, 15del	1p36.3–p36.2	191191
TCRB-9	Mic	TCR β, exon V26S1	1p36.3–p36.2	191191
SCA2	Mic	Spinocerebellar ataxia 2 (olivopontocerebellar ataxia 2, autosomal dominant, ataxin 2)	12q23–q24.1	601517
TEA	SNP	T cell early antigen	14q11.2	
TNF-1	Mic	Tumour necrosis factor (TNF super family, member 2)	6p21.3	191160
TNF-2	SNP	TNF promotor polymorphism, -862 (a/c)	6p21.3	191160
TNF-3	SNP	TNF promotor polymorphism, -805 (c/t)	6p21.3	191160
TNF-4	SNP	TNF promotor polymorphism, -238 (g/a)	6p21.3	191160
TNF-5	SNP	TNF promotor polymorphism, -308 (g/a)	6p21.3	191160
TCRB-1	SNP	TRCBV5S3	7q35	186930
TCRB-2	Mic	TRCBV6S1	7q35	186930
TCRB-3	Mic	TRCBV6S14	7q35	186930
TCRB-4	Mic	TRCBV6S1C	7q35	186930
TCRB-5	Mic	TRCBV6S3	7q35	186930
TCRB-6	Mic	TRCBV6S4	7q35	186930
TCRB-7	Mic	TRCBV6S7	7q35	186930
TCRB-8	SNP	TRCBV6S7C	7q35	186930

In column *Type* 'Mic' indicates a microsatellite locus, 'SNP' indicates single nucleotide polymorphisms and 'HT' indicates that haplotypes are considered.

point of view.[11] Although the statistic presented here provides a *P*-value (ie multiple testing is not involved), the topic of stringent inferential statistics is only deferred rather than answered satisfactorily. The issue is that association studies are continuously ongoing efforts. Also, tests for global null hypotheses are repeated when the data set grows. We rather consider this test as an additional useful descriptive measure to evaluate complex data sets. In our opinion, the burden of using stringent correction procedures for multiple testing (such as Bonferroni's or Holm's) can be relieved and should allow for more inclusive presentation of large association studies in terms of the amount of loci reported. We want to add that the test presented can be used in a more direct way to evaluate relative contributions from different loci. By excluding loci from the test, the contribution to the overall test statistics can be assessed (cf. Payami *et al*[23]). Descriptive and especially graphical methods can help to give comprehensive yet concise summaries of complex data sets. Especially data sets that are grown historically can be analysed by our methods, allowing to include the entire data set whenever the study is expanded. This can give more insights into data that have been gathered long before. So far, independent replications are the only means to verify any particular finding. ACs suggested in this paper to be associated with MS are yet to be replicated.

Still, there are very few examples that demonstrate complexity of a disorder which goes beyond single locus inheritance (eg Gabriel *et al*,[24] Hugot *et al*,[25] Ogura *et al*[26]). Clear evidence for more than a couple of loci increasing liability to disease through additive or epistatic effects is yet to be proven unequivocally. In part this is because of lack of appropriate statistics, which therefore should be considered a topic of paramount interest. In summary, we consider the statistical methods presented to improve the analysis of complex disease substantially when reasonable guidelines are respected.
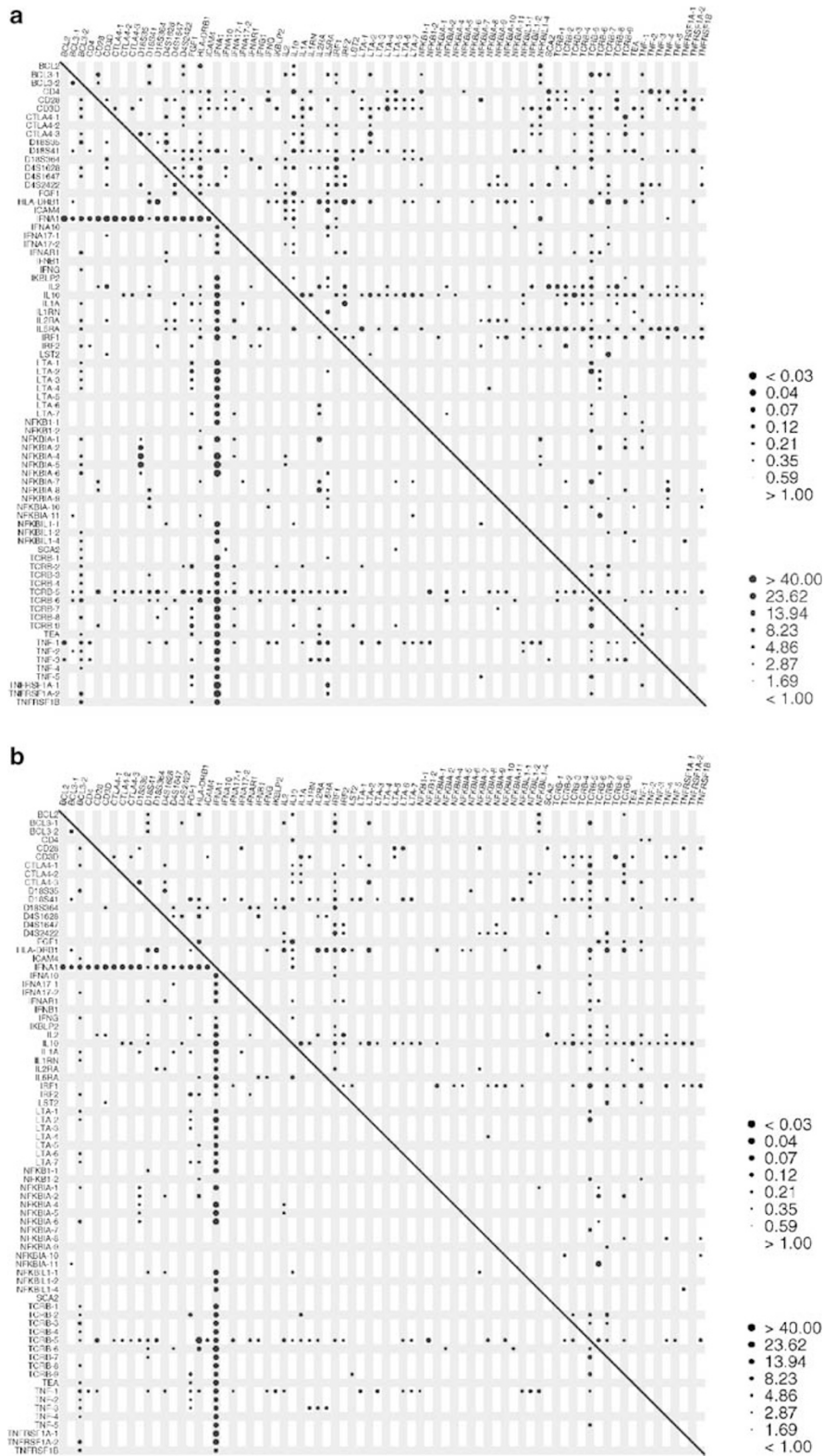
Figure 4 Two-way interaction for MS. Maximal (bottom left) and minimal ORs (top right) of ACs in the MS data set according to the additive model (**a**) and the single-count model (**b**). The size of each point corresponds to the value of the OR by a logarithmic scale. Minimal amount of observations for an AC to be included is 10. nORs are shown in part (**c**). ORs for ppMS are shown in part (**d**) (minimal count of observations is 30).
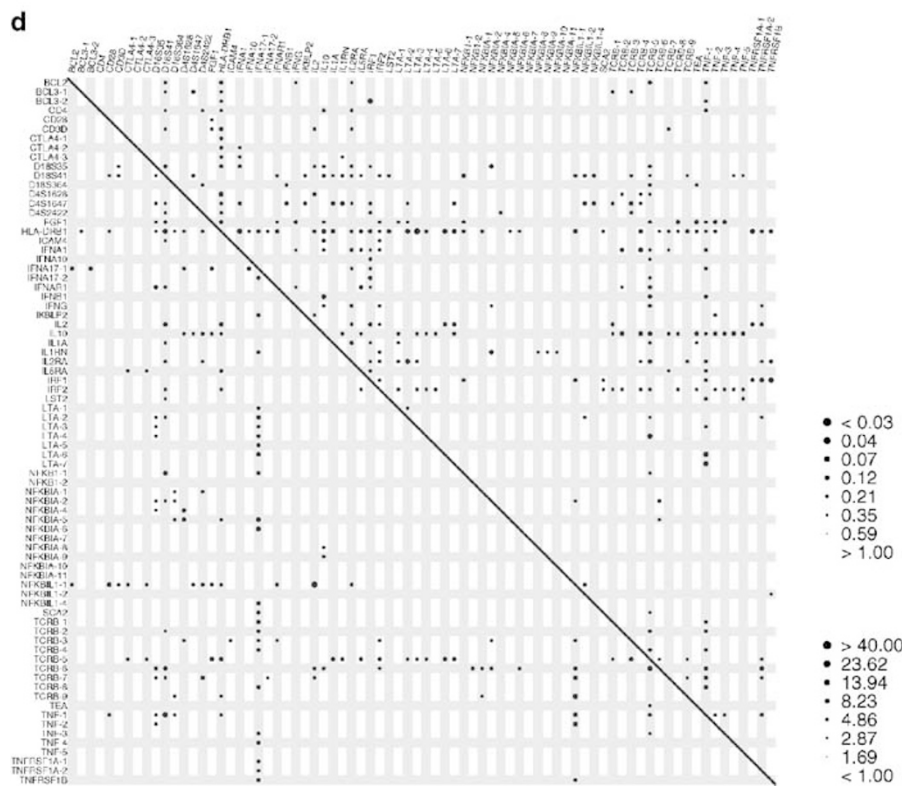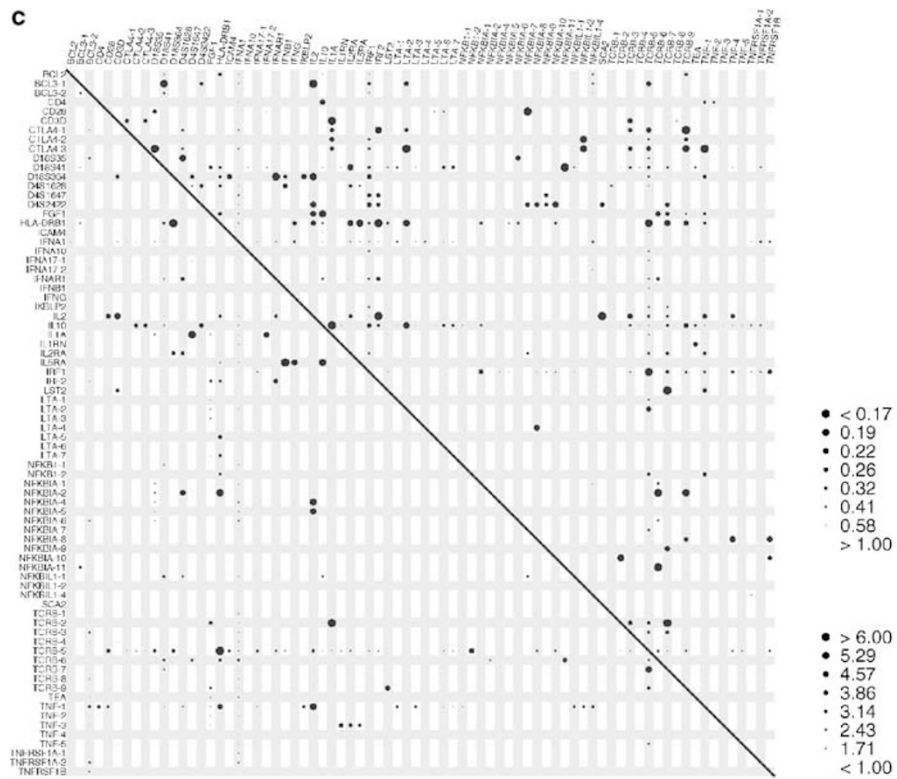
Figure 4 (*continued*)

Table 4 ACs most strongly associated with MS

AC	CU	NU	CA	NA	FU (%)	FA (%)	OR	nOR	P	Pc
IFNA1:07, TCRB-6:07	1	346	20	294	0.29	6.80	25.10	2.61	1.48E-06	4.50E-02
HLA-DRB1:03, TCRB-5:02	1	377	26	434	0.27	5.99	23.91	6.33	1.17E-06	3.56E-02
IFNA1:07, NFKBIA-6:01	2	331	22	162	0.60	13.58	25.69	1.89	1.19E-09	3.60E-05
IFNA1:07, TNFRSF1A-1:03	1	355	25	404	0.28	6.19	23.30	1.65	2.12E-06	6.43E-02
IFNA1:07, TNFRSF1A-2:02	1	340	25	395	0.29	6.33	22.85	1.78	2.26E-06	6.85E-02
IFNA1:07, NFKBIA-4:01	2	331	20	162	0.60	12.35	23.03	1.73	1.10E-08	3.35E-04
IFNA1:07, NFKBIA-5:01	2	328	20	170	0.61	11.76	21.61	1.63	2.41E-08	7.31E-04
IFNA1:07, TEA:02	1	337	22	416	0.30	5.29	18.72	1.54	2.13E-05	6.46E-01
IFNA1:07, LTA-2:06	1	353	25	497	0.28	5.03	18.61	1.37	1.94E-05	5.87E-01
CD3D:06, IFNA1:07	1	364	21	453	0.27	4.64	17.61	1.48	4.15E-05	1.26E+00
D18S35:12, IFNA1:07	1	316	14	273	0.32	5.13	16.98	1.64	2.28E-04	6.91E+00
CD3D:04, IFNA1:07	1	364	20	453	0.27	4.42	16.73	1.31	7.51E-05	2.27E+00
IFNA1:07, IL10:12	1	352	18	399	0.28	4.51	16.55	1.36	1.01E-04	3.07E+00
CTLA4-2:01, IFNA1:07	2	291	22	203	0.69	10.84	17.48	1.33	1.74E-07	5.27E-03
IFNA1:07, TCRB-1:02	1	364	17	393	0.27	4.33	16.38	1.48	1.47E-04	4.44E+00
IFNA1:07, NFKBIL1-1:01	2	303	29	284	0.66	10.21	17.07	1.30	6.68E-08	2.02E-03
IFNA1:07, IL1A:07	1	350	20	453	0.29	4.42	16.09	1.28	1.60E-04	4.84E+00
IFNA1:07, NFKB1-1:01	2	339	30	332	0.59	9.04	16.70	1.28	5.53E-08	1.68E-03
IFNA1:07, TCRB-7:14	1	366	19	461	0.27	4.12	15.66	1.28	1.45E-04	4.38E+00
IFNA1:07, TCRB-8:02	1	366	19	461	0.27	4.12	15.66	1.31	1.45E-04	4.38E+00
IFNA1:07, TCRB-3:07	1	316	18	378	0.32	4.76	15.71	1.32	2.10E-04	6.36E+00
D4S1628:03, IFNA1:07	1	331	15	330	0.30	4.55	15.67	1.17	2.33E-04	7.06E+00
IFNA1:07, NFKBIA-1:01	2	300	24	240	0.67	10.00	16.49	1.25	3.49E-07	1.06E-02
HLA-DRB1:02, IFNA1:07	1	362	20	483	0.28	4.14	15.57	0.71	1.69E-04	5.11E+00
BCL2:02, IFNA1:07	2	326	25	272	0.61	9.19	16.34	1.28	2.26E-07	6.83E-03
CTLA4-1:01, IFNA1:07	2	291	20	203	0.69	9.85	15.72	1.23	9.77E-07	2.96E-02
IFNA1:07, TNF-1:13	1	367	18	461	0.27	3.90	14.84	1.22	2.56E-04	7.76E+00
IFNA1:07, IL5RA:08	1	366	18	460	0.27	3.91	14.84	1.14	2.57E-04	7.78E+00
IFNA1:07, TEA:01	2	337	35	416	0.59	8.41	15.35	1.14	8.61E-08	2.61E-03
IFNA1:07, TNF-4:03	2	368	35	456	0.54	7.68	15.18	1.19	1.74E-07	5.26E-03
IFNA1:07, TCRB-6:05	1	346	12	294	0.29	4.08	14.63	1.20	8.97E-04	2.72E+01
IFNA1:07, TCRB-2:05	2	367	35	455	0.54	7.69	15.18	1.16	1.73E-07	5.25E-03
IFNA1:07, TNF-5:01	2	367	35	459	0.54	7.63	15.04	1.19	1.74E-07	5.28E-03
D18S364:07, HLA-DRB1:08	1	356	11	280	0.28	3.93	14.47	6.96	7.82E-04	2.37E+01
IFNA1:07, IL1A:06	1	350	18	453	0.29	3.97	14.41	1.15	2.97E-04	8.99E+00

The ACs are named *Locus1:Allele1- Locus2:Allele2*. The other columns list count of observations in unaffecteds (CU) out of a total (NU), these counts for affecteds (CA, NA), frequencies in controls (FU) and cases (FA), OR, nOR, *P*-value (*P*), and a Bonferroni-corrected *P*-value (*Pc*).

During the preparation of this manuscript similar methods have been developed. A method to identify genotype combinations was proposed.[27] To reduce the complexity of the set of relevant genotype combinations the authors group genotype combinations with similar effect into partitions which are then evaluated by cross validation for their predictive power. In another study, a global test statistic is computed in a two-stage process.[28] First, a sum of test statistics is calculated similar to the present study. A variable number of loci is considered at this point. The subset which displays the strongest association according to the initial test statistics is chosen and evaluated for significance by bootstrapping techniques. Simultaneous effects of loci are not considered.

References

1 Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–1517.
2 Tu IP, Whittemore AS: Power of association and linkage tests when the disease alleles are unobserved. *Am J Hum Genet* 1999; **64**: 641–649.
3 Tu IP, Balise RR, Whittemore AS: Detection of disease genes by use of family data. II. Application to nuclear families. *Am J Hum Genet* 2000; **66**: 1341–1350.
4 Risch N, Teng J: The relative power of family-based and case–control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. *Genome Res* 1998; **8**: 1273–1288.
5 Teng J, Risch N: The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. II. Individual genotyping. *Genome Res* 1999; **9**: 234–241.
6 Cardon LR, Bell JI: Association study designs for complex diseases. *Nat Rev Genet* 2001; **2**: 91–99.
7 Schulze TG, McMahon FJ: Genetic association mapping at the crossroads: which test and why? Overview and practical guidelines. *Am J Med Genet* 2002; **114**: 1–11.
8 Editorial: Freely associating. *Nat Genet* 1999; **22**: 1–2.
9 Perneger TV: What's wrong with Bonferroni adjustments. *BMJ* 1998; **316**: 1236–1238.

10 Bender R, Lange S: Multiple test procedures other than Bonferroni's deserve wider use. *BMJ* 1999; **318**: 600–601.

11 Boehringer S, Epplen JT, Krawczak M: Genetic association studies of bronchial asthma – a need for Bonferroni correction? *Hum Genet* 2000; **107**: 197.

12 Holm S: A simple sequentially rejective multiple test procedure. *Scand J Statist, Theory Appl* 1979; **6**: 65–70.

13 Whittemore AS, Tu IP: Detection of disease genes by use of family data. I. Likelihood-based theory. *Am J Hum Genet* 2000; **66**: 1328–1340.

14 Epplen C, Jackel S, Santos EJ *et al*: Genetic predisposition to multiple sclerosis as revealed by immunoprinting. *Ann Neurol* 1997; **41**: 341–352.

15 Miterski B, Jaeckel S, Epplen JT, Pohlau D, Hardt C: The interferon gene cluster: a candidate region for MS predisposition? Multiple Sclerosis Study Group. *Genes Immun* 1999; **1**: 37–44.

16 Miterski B, Epplen JT, Poehlau D, Sindern E, Haupts M: SCA2 alleles are not general predisposition factors for multiple sclerosis. *Neurogenetics* 2000; **2**: 235–236.

17 Hudson RR: Gene genealogies and the coalescent process; in Futuyma D and Antonovics J. (eds): *Oxford surveys in evolutionary biology*. Oxford, UK: Oxford University Press, 1990, pp 1–44.

18 Nordborg M: Coalescent theory; in Balding DJ, Bishop M, Cannings C (eds): *Handbook of Statistical Genetics*. Chichester, UK: Wiley, 2001, pp 179–212.

19 Bacanu SA, Devlin B, Roeder K: The power of genomic control. *Am J Hum Genet* 2000; **66**: 1933–1944.

20 Devlin B, Roeder K: Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.

21 Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: Association mapping in structured populations. *Am J Hum Genet* 2000; **67**: 170–181.

22 Pritchard JK, Rosenberg NA: Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999; **65**: 220–228.

23 Payami H, Joe S, Farid NR *et al*: Relative predispositional effects (RPEs) of marker alleles with disease: HLA-DR alleles and Graves disease. *Am J Hum Genet* 1989; **45**: 541–546.

24 Gabriel SB, Salomon R, Pelet A *et al*: Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nat Genet* 2002; **31**: 89–93.

25 Hugot JP, Chamaillard M, Zouali H *et al*: Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001; **411**: 599–603.

26 Ogura Y, Bonen DK, Inohara N *et al*: A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001; **411**: 603–606.

27 Nelson MR, Kardia SL, Ferrell RE, Sing CF: A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation. *Genome Res* 2001; **11**: 458–470.

28 Hoh J, Wille A, Ott J: Trimming, weighting, and grouping SNPs in human case–control association studies. *Genome Res* 2001; **11**: 2115–2119.

29 Cox DR, Hinkley DV: *Theoretical statistics*. London, UK: Chapman & Hall, 1974.

30 Efron B, Tibshirani RJ: *Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.

31 Merriman TR, Cordell HJ, Eaves IA *et al*: Suggestive evidence for association of human chromosome 18q12-q21 and its orthologue on rat and mouse chromosome 18 with several autoimmune diseases. *Diabetes* 2001; **50**: 184–194.

32 Barnes PJ, Karin M: Nuclear factor-kappaB: a pivotal transcription factor in chronic inflammatory diseases. *N Engl J Med* 1997; **336**: 1066–1071.

33 Holzinger I, de Baey A, Messer G, Kick G, Zwierzina H, Weiss EH: Cloning and genomic characterization of LST1: a new gene in the human TNF region. *Immunogenetics* 1995; **42**: 315–322.

34 de Chasseval R, de Villartay JP: Functional characterization of the promoter for the human germ-line T cell receptor J alpha (TEA) transcript. *Eur J Immunol* 1993; **23**: 1294–1298.

Appendix A

Statistical tests

Our statistics comprises sets of loci of fixed size. We denote the set of loci with $L = \{l_1, \ldots, l_n\}$, when n is the count of loci under scrutiny. The set of subsets of L, when each subset has t elements is denoted with L_t. We now consider a global test statistics, which is composed of single statistics gleaned from elements in L_t.

$$T = \sum_{l \in L_t} T(l)$$

Each statistics $T(l)$ is of the following form:

$$T(l) = F_l(\chi^2(l))$$

$\chi^2(l)$ is the test statistics resulting from a conventional χ^2 statistics from a contingency table[29] with the counts defined by the respective models as described above. Each individual statistics $T(l)$ tests for independence of AC frequencies and disease status. F_l denotes the cumulative distribution function of each statistics. Therefore, $T(l)$ is uniformly distributed from 0 to 1 if all observations are independent. Clearly, if all loci are unlinked, each statistics $T(l)$ will contribute equal weight to the summary statistics T.

Various interdependencies exist in the observations included in the individual statistics. For the additive and single-count models, observations at a single set of loci are dependent on each other. Also statistics $T(l)$ and $T(l')$ are dependent if l and l' overlap in general. Additionally, loci may be linked which may result in LD between certain alleles. This can lead to dependencies between certain test statistics $T(l)$ and $T(l')$ even if l and l' do not overlap. To account for these interdependencies we employ a bootstrap procedure to estimate the distribution of T under the null hypothesis.[30]

To perform one replication step of the bootstrap procedure, we draw with replacement individuals from either the patient or the control group. Two cohorts are produced which equal in size the number of probands contained in the patient and control group, respectively. Both these groups can therefore be considered to be drawn from the same empirical distribution. The tests statistics is applied on this data set. We use 20 000 replications to estimate the distribution of T in each case. To produce our simulated data sets, we use a coalescent process as described elsewhere.[17,18] We simulate microsatellite data with a stepwise mutation model. Case and control group were of equal size for each data set. The mutation rate chosen is $\theta = 5$ which is a realistic assumption for "real-world" populations.[21] To estimate power, we repeat the test 80 times for each scenario under the null hypothesis. We

have stopped this process early, if either the lower bound of the confidence interval (CI) at the 0.95 level for the power is $\geqslant 85\%$ or the upper bound of the CI is $\leqslant 40\%$.

Missing data are handled as follows. For each set of loci l the test statistics $T(l)$ is computed for the genotypes of these loci. If there are missing data in one AC the whole AC is ignored. If data are missing at random, the statistics remains valid. For real data sets missing data may be different for the patient and control groups. We compute the number of probands to draw in the bootstrap procedure as follows:

$$c_d = a_d \frac{c_s}{a_s}$$

when c_d is the count of probands to draw, a_d the count of alleles of the group for which a bootstrap data set is to be drawn, c_s the count of proband in the source cohort and a_s the count of alleles therein.

Descriptive statistics

The normalized OR is computed from a predicted distribution for ACs. This is assumed to be the product distribution of individual alleles, conditioned on the phenotype:

$$P(a_{l_1}, \ldots, a_{l_n}|Y) := \prod_1^n P(a_{l_i}|Y)$$

The OR is defined by

$$OR = \frac{\frac{P(a_{l_1}, \ldots, a_{l_n}|Y=1)}{1 - P(a_{l_1}, \ldots, a_{l_n}|Y=1)}}{\frac{P(a_{l_1}, \ldots, a_{l_n}|Y=0)}{1 - P(a_{l_1}, \ldots, a_{l_n}|Y=0)}}$$

and can be computed for, both, the predicted distribution (OR_p) and the observed joint distribution (OR) of alleles. In this paper we define nOR by

$$nOR = \frac{OR_p}{OR}$$