Molecular Psychiatry (2002) 7, 56–66

www.nature.com/mp

ORIGINAL RESEARCH ARTICLE

Regional meta-analysis of published data supports linkage of autism with markers on chromosome 7

JA Badner and ES Gershon

Department of Psychiatry, University of Chicago, Chicago, IL, USA

Although the concept of meta-analysis of multiple linkage scans of a genetic trait is not new, it can be difficult to apply to published data given the lack of consistency in the presentation of linkage results. In complex inheritance common diseases, there are many instances where one or two studies meet genome-wide criteria for significant or suggestive linkage but several other studies do not show even nominally significant results with the same region. One possibility for resolving differences between study results would be to combine an available result parameter of several studies. We describe here a method of regional meta-analysis, the multiple-scan probability (MSP), which can be used on published results. It combines the reported *P*-values of individual studies, after correcting each value for the size of the region containing a minimum *P*-value. Analyses of the power of MSP and of its type I error rates are presented. The type I error rate is at least as low as that for a single genome scan and thus genome-wide significance criteria may be applied. We also demonstrate appropriate criteria for this type of meta-analysis when the most significant study is included, and when that study is used to define a region of interest and then excluded. In our simulations, meta-analysis is at least as powerful as pooling data. Finally, we apply this method of meta-analysis to the evidence for linkage of autism susceptibility loci and demonstrate evidence for a susceptibility locus at 7q.
Molecular Psychiatry (2002) **7**, 56–66. DOI: 10.1038/sj/mp/4000922

Keywords: meta-analysis; autism; genetic linkage analysis; complex genetic diseases; statistical data interpretation

Introduction

Suarez *et al*[1] demonstrated a curious finding, which has since been widely cited by statistical analysts of common disease genetics as an explanation of inconsistencies in linkage results. They found that, for an oligogenic trait simulated under reasonable parameters, when sampling families sequentially, the first true linkage to be detected will not likely be replicated when a second pedigree series reaches the same size. One can interpret their sequential sampling simulation results into a fixed sample size, and conclude that acceptable power to replicate the first linkage was present only when the sample was several times larger than the initial sample. The reason was self evident—if there are 10 true linkages to be found, the probability of detecting *any* one of the 10 is higher than the probability of detecting one in particular. This finding led many research planners to demand megasamples for linkage studies of inherited common disease, when oligogenic inheritance is thought to exist. But this has its disadvantages—experiments that are so large they may never be tested for replication, and very long lead times between envisioning a study and receiving a result. In the meanwhile, for disease, like bipolar illness and schizophrenia, many linkage studies with modest sample sizes are reported with results that are hard to interpret. In a given chromosomal region there may be one or two significant or suggestive reports (by the guidelines of Lander and Kruglyak[2]), and other results that do not suggest linkage. There also have been instances where several studies show nominally significant results within a given chromosomal region but none exceeds the thresholds suggested by Lander and Kruglyak.[2] Although meta-analysis of linkage studies has been proposed and performed many times in recent years,[2–6] criteria for evaluating a *series* of linkage studies have not been developed. Intuitive interpretations have been widely offered. Some scientists hold that there is no replication unless a finding can be seen in every study, others believe that pooling of data is the only acceptable way to combine disparate studies, and others eschew statistics and believe a sophisticated scientist can tell by inspection of a series of results when there is a subtle finding and when there is not.

We have undertaken to develop a statistical criterion for evaluating a series of linkage studies. In several diseases where multiple studies have been done, significance criteria are not met consistently, yet the evidence

Correspondence: JA Badner, MD, PhD, Department of Psychiatry, University of Chicago, 5841 South Maryland Ave, MC3077 Chicago, Illinois 60637, USA. E-mail: jbadner@yoda.bsd.uchicago.edu
Received 12 December 2000; revised 8 March 2001; accepted 12 March 2001

is not easily dismissed. Recently, Horikawa *et al*[7] detected a susceptibility gene for non-insulin-dependent diabetes mellitus (NIDDM1) by following up on linkage results that were not, on inspection, inspiring.[8–10] Interpreting the significance of multiple genome scans with conflicting results within the same region, as is true of NIDDM1, is difficult. One possibility is to combine the different studies into one sample and test for linkage. However, this is often difficult to do and may give misleading results if there are biases as to which studies are included. Another possibility is to assess the probability of the different studies exceeding specified thresholds within a given linkage region by chance. Badner and Goldin[11] developed statistics for assessing the probability of when r genome scans out of a total of s genome scans exceed a particular threshold within an n cM region. They showed that it was possible to obtain highly significant results when nominal P-values of each genome scan exceeded relatively modest significance criteria. For example, the estimated probability of 4/4 genome scans demonstrating a P-value of less than 0.01 within a 30 cM region was 3×10^{-5}.

Hedges and Olkin[12] describe a method of combining P-values developed by Fisher.[13] This method has been applied to genetic analysis by Allison and Heo[6] and Guerra *et al*.[5] For k independent studies testing the same null hypothesis (eg linkage is not present) yielding P-values $P_1, \ldots P_k$, if the null hypothesis is true, then $-2\Sigma_{i=1,k}\ln(P_i)$ is distributed as a χ^2 with $2k$ degrees of freedom. We will call $P(\chi^2 > -2\Sigma_{i=1,k} \ln(P_i))$ MSP for Multiple Scan Probability. The advantage of this test is the simplicity of use and the applicability to a wide range of published data. Other techniques of meta-analysis using parameter estimates of linkage (eg Identity By Descent scores) might prove more accurate but would be difficult to apply to published studies where the information required may not be readily available for all studies.

Since there can be significant variation in location estimates for a true linkage finding,[14] it would be desirable to look at a linkage region rather than a single point in the genome. Also, when results from published studies are used, frequently only information on local minimum P-values is available. Therefore, we calculate the MSP by taking the minimum P-value over n cM in each study, correct this observed P-value by estimating the probability of it occurring within n cM and sum the logs of the corrected P-values as above. The null hypothesis is that the distribution of the P-values is uniform. If this null hypothesis is rejected, that implies that true linkage is present in at least one of the studies. Allison and Heo[6] advocated a similar method of using P-values within a region of linkage and correcting for the size of the linkage region but they did not estimate the type I error and power of this method.

In this paper, we discuss how to correct for combining P-values across regions rather than at a single point. Simulations are performed to demonstrate the type I error, and power of the MSP. This test is applied to published evidence for linkage of susceptibility loci for autism.

Methods

If the probabilities at the same point in a genome scan (ie observed P-values at the same marker across studies) are combined from multiple studies, the resulting probability can be calculated using the equation for MSP, ie

Given k independent studies with P-values $P_1, \ldots P_k$

$$Y^2 = -2\Sigma_{i=1,k} ln(P_i) \quad (1)$$

$$\text{MSP} = P(\chi^2 \text{ with } 2k \text{ degrees of freedom} > Y^2) \quad (2)$$

This would give a nominal probability which would need to be corrected for genome-wide testing. However, evidence for a linkage can occur over a broad region (20–30 cM). Therefore, it would be of interest to combine probabilities across regions rather than at single points. In order to do this, the observed minimum P-value from each study needs to be corrected for the size of the linkage region. Feingold *et al*[15] estimate the probability of a P-value being observed in a given sized region:

$$P^* = CP + 2\lambda GZ(P)\phi(Z(P))v[Z(P)\text{sqrt}(4\lambda\Delta)] \quad (3)$$

where P is the observed P-value from scan i, C is the number of chromosomes, λ is the rate of crossovers per Morgan and varies depending on the method of analysis and family structure analyzed, G is the size of the region in Morgans, $Z(P)$ is standard normal inverse of P, $\phi(Z(P))$ is the normal density function, and Δ is average marker spacing in Morgans. The function $v(x)$ is a discreteness correction for the distance Δ between markers and can be approximated as $\exp(-0.583x)$ when $x < 2$. For the case of continuous markers ($\Delta = 0$, $v(x) = 1$) and small P_i, the above equation is essentially the same as in Lander and Kruglyak.[2]

$$P^{LK} = (C + 2\lambda G(Z(P))^2)P \quad (4)$$

Allison and Heo[6] used Equation 4 for their correction. However, since P is not always small, Equation (3) tends to provide a better fit than Equation (4). It also allows the correction for the actual marker density for each scan rather than assuming an infinitely dense scan for all studies. Thus, the MSP for a chromosomal region can be calculated by substituting P^* for P_i in Equation 1. Since all the simulations and autism analyses involved affected sib pairs, λ was set to 2.

Simulations

Simulations were performed with three purposes: (1) To verify the probability distribution of the MSP using P^*, that is, its type I error. (2) Power analysis of the MSP method as developed here. (3) Analysis of heterogeneity, sample size differences, methods of analysis differences, and other issues.

In order to assess the type I error, and the comparison of the probability distribution across genomic

length, with that expected for a single study as predicted using Equation 4, pedigrees were simulated using SIMULATE.[16] In each replicate, four linkage studies of nuclear families with two affected offspring were simulated. A map of 200 markers, each with four equifrequent alleles, with 1 cM spacing was simulated. The markers were unlinked to the disease genotype. Table 1 shows the different types of simulations that were performed. For several of the simulations, the four studies were combined and analyzed as a single study with 400 nuclear families. In each simulation, 1000 replicates were simulated. Two different sets of simulation were performed. In the first set, the minimum *P*-value from each study was taken from regions of fixed size (n = 0, 10, 20, or 30 cM) and corrected for region size. In the second set of simulations, the *P*-value from the most significant study was not corrected for region size and the minimum *P*-value was taken from regions of varying sizes around the location of the minimum *P*-value of the most significant study.

For the power analysis, pedigrees were simulated using SLINK.[17–19] In each replicate, four sets of 100 nuclear families with two affected offspring were simulated. A single marker had four equifrequent alleles and was linked to the disease locus with a recombination fraction of 0.05. A multipoint map was not simulated due to limited computer resources. Although results simulated this way would not have the power of an 'infinitely dense map', the relative power of the different analyses should be similar to a multipoint analysis. Two genetic models seen in complex genetic disorders were simulated: (1) A locus with a common susceptibility allele (P = 0.1–0.3) and low penetrance (probability of affection given 2, 1, and 0 susceptibility alleles is 0.10, 0.05, 0.00 respectively) as would be seen in genes with epistatic interactions. (2) A locus with susceptibility allele frequency of 0.01 and low penetrance (same as for the previous model) that is not present in all affected individuals as would be seen in traits with genetic heterogeneity. The proportion of families transmitting the susceptibility allele ranged between 40% and 60%. The effects of study heterogeneity, ie, not all studies having a susceptibility locus in the region studied, were analyzed, as was the effect of different sample sizes (equivalent to the unequal sample sizes simulated in Table 1). For these analyses, a genetic model of a locus with a common susceptibility allele was used with penetrances as described above and allele frequency 0.1–0.3 for Figure 3a and allele frequency 0.1 for Figure 3b.

Autism data

Autism is a neurodevelopmental disorder characterized by impairment in social interaction, communication, and behavior. Evidence for a genetic component has been shown in twin studies and family studies.[20,21] Four genome scans for autism have been published. The International Molecular Genetic Study of Autism Consortium study (IMGSAC)[22] genotyped 99 multiplex families (87 families with affected sib pairs) in which the proband met criteria for autism and another relative met criteria for an autism spectrum diagnosis (Autism, Asperger Syndrome or Pervasive Developmental Disorder, NOS). All were Caucasian and were from the UK, Germany, the Netherlands, the USA, France and Denmark. The average intermarker spacing was 10 cM. A variety of analyses were performed, however results from ASPEX were presented for the whole genome, whereas for other methods only the results from nominally significant regions were presented. Therefore, the ASPEX results for the entire sample were used for the meta-analysis.

The Paris study[23] genotyped 51 multiplex families from Sweden, France, Norway, the USA, Italy, Austria and Belgium. Each family had at least an affected sib pair or half-sib pair where both sibs had autism. Two hundred and sixty-four markers were genotyped which led to an average intermarker spacing of approximately 10 cM. MAPMAKER/SIBS[24] was used for the multipoint analysis and these are the results used for this meta-analysis.

The Stanford study[25] genotyped in two stages. The

Table 1 Simulations performed to assess type I error and probability distribution across genome length

Simulation	*Corrected for region size*	*Number of families/study*	*Linkage analysis method*
Uncorrected	No	Each study with 100 families, four studies total	GENEHUNTER/PLUS[29]
Corrected	Yes	Each study with 100 families, four studies total	GENEHUNTER/PLUS
Unequal sample sizes	Yes	Study 1: 125 families Study 2: 75 families Study 3: 150 families Study 4: 50 families	GENEHUNTER/PLUS
Different genetic linkage analytic methods	Yes	Each study with 100 families, four studies total	(1) parametric HLOD[a] (2) NPL[a] (3) ASPEX sib_ibd[30] (4) GENEHUNTER/PLUS

[a]GENEHUNTER.[31]

first stage had 90 multiplex families and the second stage had 49 multiplex families for a total of 139 families, all from the USA. All families had at least two siblings with autism. There were 519 markers genotyped in the first stage of families and 149 markers genotyped in the second stage of families. Genotyping was more dense in regions of nominally significant results for this study and in regions identified by other studies including 6p, 7q, and 15q. ASPEX was used for the analysis. The results from the analysis using all the families are used for this meta-analysis.

The Collaborative Linkage Study of Autism (CLSA)[26] genotyped 75 families with at least two children with autism from the USA. Markers had an average spacing of 9 cM but 7q and 13q were more densely genotyped. Multipoint results using MMLS/het were presented for the genome. MMLS/het calculates a heterogeneity lod score with GENEHUNTER under both a dominant and a recessive model and takes the maximum of the two and corrects the significance for testing for two models. GENEHUNTER NPL *P*-values were also presented but not for the whole genome and thus, the MMLS/het scores were used for the meta-analysis.

Results

Type I error

For the simulation of unlinked regions, linkage statistics were generated for each cM of the 200 cM region in each replicate. For the combined (pooled data analyzed with GENEHUNTER/PLUS), the probability of observing the statistic was calculated for each cM. For the n cM analysis, where n = 0, 10, 20, or 30, at each cM, the probability (P) of the most significant statistic in the n cM region following each cM was calculated (ie, for cM 1, the most significant statistic in the 1 to n+1 cM region). This was done for each cM of the 200 cM. These *P*-values are presented in two ways: uncorrected for region size (P_i), and corrected for region size (P^*). In the analyses that were uncorrected for region size, these uncorrected *P*-values were used in Equations 1 and 2 to obtain the MSP (Figure 1a). In the other analyses of unlinked regions, the corrected *P*-value, P^* was calculated for each cM of each scan using Equation 3 (Figures 1b–d).

The empirical pointwise probabilities were estimated by calculating the probability that P_i or P^* was less than or equal to 0.05 at each cM for the 1000 replicates. When the *P*-values were not corrected for linkage region size, the type I error rate was 0.047, 0.19, 0.34, and 0.47 for a 0, 10, 20, 30 cM linkage region size respectively. When the *P*-values were corrected for linkage region size, the type I error rate was 0.046, 0.063, 0.054, and 0.044 for a 0, 10, 20, 30 cM linkage region size respectively. Thus, the correction gave an acceptable type I error rate.

Figure 1 shows the empirical probability which was estimated at each cM in each replicate by calculating whether P_i or P^* was less than or equal to 0.05 in the region scanned up to that point (ie, the probability at 17 cM is the probability that P_i or P^* was 0.05 at any of the points between 0 and 17 cM). These probabilities are compared to that which is expected for a single infinitely dense genome scan which forms the basis of the criteria suggested by Lander and Kruglyak,[2] calculating the probability as $1 - \exp(-P^{LK})$ where P^{LK} is derived from Equation 4 using $P = 0.05$. If the probabilities of the MSP are the same or decreased compared to that expected for a single genome scan, then that would mean that Lander and Kruglyak genome scan criterion for linkage could be applied to the results of the MSP analysis with a similar or lower rate of false positives as applying the criteria to a single genome scan.

Figure 1a shows that the *uncorrected* probabilities exceed those expected for a single infinitely dense genome scan, and this is true even for linkage region sizes as small as 10 cM, which could be considered well within the range of the confidence interval of the location of the putative susceptibility locus. For *corrected* probabilities, Figure 1b demonstrates that the larger the n cM region is, the smaller the probabilities become. This is likely due to the fact that there is increased autocorrelation between P^* at each cM as compared with that expected for Equation 4. Therefore, applying Lander and Kruglyak genome scan criterion to this type of analysis would lead to a lower number of false positives than predicted by the criteria. Figures 1c–d demonstrate that for the conditions we simulated, there is little to no effect of unequal sample sizes or using different genetic analytic methods on these conclusions.

In a collection of linkage studies for a particular trait, a region is usually interesting because at least one study obtained a very significant result in the area. The study with the most significant result is more likely to have the best localization of a susceptibility locus than other less significant studies, since the higher significance presumably reflects greater power to detect linkage, although exceptions to this hypothesis obviously can occur. Thus, rather than correcting for a fixed n cM region, each study is corrected for twice the distance between its minimum *P*-value value and the location of the minimum *P*-value in the most significant study. The reason why twice the distance is used is because local minimum *P*-values that are proximal or distal to the location of the minimum *P*-value in the most significant study would be of interest.

There is some question whether including the results of the most significant study in a meta-analysis will incur a bias. We hypothesized that when the most significant study is included in the MSP, the results need to be compared to genome-wide thresholds. When the most significant result is excluded in the meta-analysis and other results are still corrected for distance between the local minimum *P*-value and the minimum *P*-value of the most significant study, then the results can be compared to nominal thresholds. To test this, a simulation of a 200 cM map of markers unlinked to a susceptibility locus in four studies using the same conditions as 'corrected' in Table 1 was performed. For each replicate, the location of the most significant test

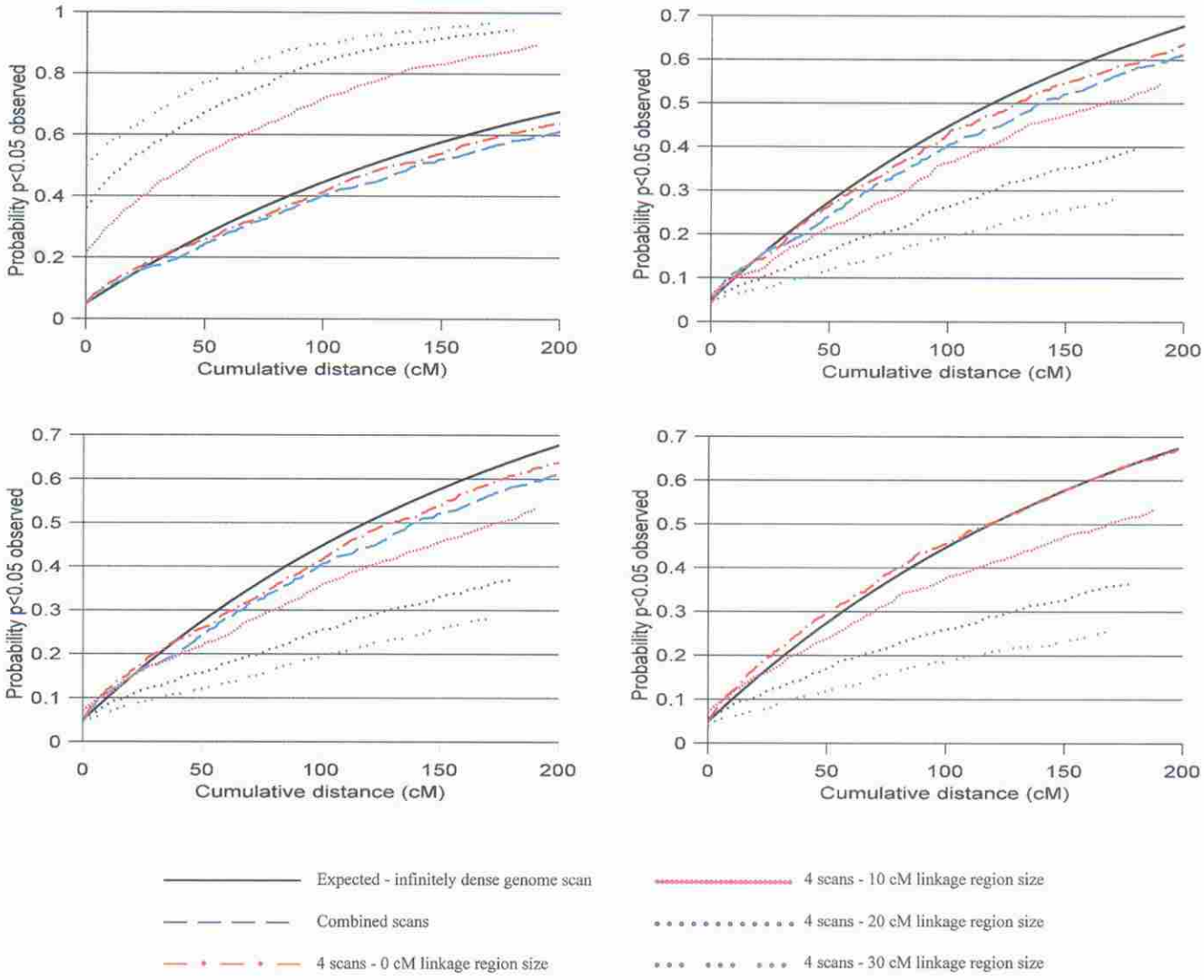

Figure 1 Probability distribution for the Multiple Scan Probability (MSP). One thousand replicates of four studies, each with 100 affected sib pairs and a 200 cM map with 200 markers unlinked to the susceptibility locus were simulated. The empirical probability is estimated at each cM in each replicate by calculating whether P (observed P-value) or P^* (P-value corrected for linkage region size) was less than or equal to 0.05 in the region scanned up to that point (ie the probability at $m = 17$ cM is the probability that P or P^* was 0.05 at any of the points between 0 and 17 cM). These probabilities are compared to that which is expected for a single infinitely dense genome scan. The n cM, where $n = 0$, 10, 20, or 30 cM, refers to the size of the region that the most significant result is taken from. The combined scans are when all four studies are combined and analyzed as a single study. (a) Probabilities that are uncorrected for the n cM region size. (b) Probabilities that are corrected for the n cM region size. (c) Probabilities from four studies with four different sample sizes and corrected for the n cM region size. (d) Probabilities from four studies each with a different genetic linkage analytic method and corrected for the n cM region size (see Table 1).

statistic of the four studies was taken as the location of the putative susceptibility locus (LPSL). In the other three studies, the P-value of the most significant test statistic for each of these studies (local minimum P-value) was corrected for twice the distance away from the LPSL. The observed local minimum was estimated in two ways. The first way was to take the minimum P-value of the study, regardless of the distance from the LPSL. The second way was to take the minimum P-value within 30 cM of the LPSL. MSP was calculated both including and excluding the most significant study. For nominal probabilities of 0.05, 0.01, 0.001, 0.0001, the empirical P-value was calculated by counting the number of replicates for which either MSP was equal or less than the nominal probability. Table 2 shows these results. When all four scans are included in the MSP, the expected probability of observing a P-value at or below a certain threshold was calculated using Equation (4) (P = nominal probability, G = 200 cM, $C = 1$) to estimate $P^{L,K}$ and then calculating $1-\exp(-P^{L,K})$. When the most significant scan was excluded, the expected probability was equal to the

Table 2 Analysis of simulated pedigrees with no linkage: type I error rates when the most significant result is included or excluded

Nominal P-value for significance	*All four scans included*			*Excluding most significant result*		
	Expected probability[a]	*Observed*		*Expected probability*[a]	*Observed*	
		Entire region	*Within 30 cM*		*Entire region*	*Within 30 cM*
0.05	0.68	0.51	0.70	0.05	0.036	0.058
0.01	0.36	0.23	0.32	0.01	0.004	0.007
0.001	0.074	0.048	0.060	–	–	–
0.0001	0.011	0.010	0.011	–	–	–

[a]Expected probability of observing a *P*-value at or below the nominal *P*-value anywhere with the 200 cM region.
A simulation of a 200 cM map of markers unlinked to a susceptibility locus in four studies using the same conditions as 'corrected' in Table 1 was performed. For each replicate, the location of the most significant test statistic of the four studies was taken as the location of the putative susceptibility locus (LPSL). In the other three studies, the *P*-value of the most significant test statistic for each of these studies (local minimum) was corrected for twice the distance away from the LPSL. The observed local minimum was estimated in two ways. The first way was to take the minimum *P*-value of the study, regardless of the distance from the LPSL ('Entire region'). The second way was to take the minimum *P*-value within 30 cM of the LPSL. MSP was calculated both including and excluding the most significant study. For varying nominal probabilities, the number of replicates for which either MSP was equal or less than the nominal probability was calculated. When all four scans are included in the MSP, the expected probability of observing a *P*-value at or below a certain threshold was calculated using Equation (4) (P = nominal probability, G = 200 cM, C = 1) to estimate P^{LK} and then calculating $1-\exp(-P^{LK})$. When the most significant scan was excluded, the expected probability was equal to the nominal probability.

nominal probability.

Table 2 demonstrates that, when calculated in this way, the observed probability is more conservative than the expected probability for both conditions when local minima are taken from anywhere in the region. Also, the observed probability is close to the expected probability when the local minima are taken from within 30 cM of the LPSL. This is consistent with Figure 1 which demonstrates that the larger the linkage region size that is corrected for, the more conservative the MSP becomes. Therefore, our hypothesis that when the most significant study is included, genome-wide criteria should be applied and when it is excluded, nominal criteria may be applied is supported by this simulation. Although it intuitively is not reasonable to incorporate results from regions more than 30 cM away of the LPSL, doing so does not appear to inflate the type I error in these simulations.

Power analysis

For the power analysis, the MSP was calculated by incorporating the simulation probabilities from two, three or four studies into Equations 1 and 2. Each study was analyzed with GENEHUNTER/PLUS on affected sib pair data. Analyses involving only one study are also presented. When only one to three studies were analyzed, the first one to three of simulated studies were incorporated into the analysis. The combined analysis was performed pooling the raw data from all four studies and analyzing as a single study using GENEHUNTER/PLUS. The power for these analyses was calculated by estimating the proportion of the 100 replicates that has a *P*-value less than or equal to 2.2×10^{-5}.

Power of the MSP test is compared to the power of using *modified* Lander and Kruglyak criteria for significant and replicated linkage. Lander and Kruglyak criteria were not designed for the interpretation of results from multiple genome scans. For example, the probability of a single study meeting criteria for significance or replication by chance would be much higher if 20 different genome scans are performed than if only two are performed. However, in practice, significant evidence of linkage is frequently claimed when one genome scan exceeds Lander and Kruglyak criteria, regardless of how many other studies have been negative in the same region. Therefore, in our analyses we modified their criteria as follows: 'LK significant' is defined as any one of the four studies showing a probability less than or equal to 2.2×10^{-5}. 'LK replicated' was defined using the criterion of 'LK significant' and also requiring that one of the remaining three studies show a probability less than or equal to 0.01. The power for 'LK significant' and 'LK replicated' was calculated as the proportion of replicates that met the criteria. These criteria were modified from Lander and Kruglyak[2] where a *P*-value of 2.2×10^{-5} in a single study of affected sib pairs was the criterion for significant linkage and confirmed linkage was if a single study met the criteria for significant linkage and a second study showed a *P*-value of 0.01. However, this does not take into account the results of additional studies that may have been done. It could be argued that results should not meet 'LK significant' or 'LK replicated' criteria if the remaining studies are not also nominally significant. This requirement would decrease the power for these criteria compared to what is presented here. Therefore, the power for the 'LK' cri-

teria presented here may be too optimistic.

Figure 2 shows the power to detect linkage to a susceptibility locus. For both genetic models, the power of the MSP for three and four scans is significantly greater than the power of either of the 'LK' criteria. The power of the MSP for four scans is very similar to the power of the combined analysis for both genetic models. The power to detect linkage using a single study is very low and reflects the power of applying Lander and Kruglyak criteria to a single study for these genetic models. Figure 3a shows that unequal sample sizes have little effect on the power of the MSP for four studies. The power for 'LK significant' is slightly increased which is probably secondary to the larger sample sizes in some of the studies in the unequal sample size as compared with the equal sample size. Figure 3b demonstrates the effect of heterogeneity between studies in terms of the presence of a susceptibility locus within the genomic region studied. The results of the MSP for four studies are similar to the combined analysis when 0 or 1 study did not have genetic linkage and are more powerful than 'LK' criteria. When two of the four studies did not have genetic linkage, the MSP for four studies was more powerful than the combined analysis (pooled data) or the 'LK' criteria. When only one out of four studies showed linkage, power was very low for all analyses.

Analysis of autism data

Table 3 shows the application of the MSP to four published genome scans for autism. In this analysis, any region that demonstrated a *P*-value less than 0.01 for any of the four studies was included. The distances were estimated using the Marshfield maps.[27] Equation 3 was used to correct each *P*-value for twice the distance away from the most significant result. The marker density of each scan was incorporated into Equation 3, thereby accounting for the fact that subsequent studies may have had denser genotyping in regions that earlier studies found significant. The most significant regions from this analysis are at 7q ($P = 0.00014$) and the more distal region of 13q ($P = 0.0006$). These results were suggestive by the genome-wide criterion. When the MSP was calculated using the same location in each genome scan rather than local minimum *P*-value, the results for 7q and 13q were 0.0007 and 0.00088. Other regions, which by inspection might have seemed to have similar results, did not have significant or suggestive MSPs (Table 3). Replication MSPs were calculated for all the chromosomal regions for demonstration purposes although they are generally only meaningful when the MSP is significant. The only significant replication MSP is for 7q ($P = 0.02$).

Discussion

MSP offers a means of analyzing the results of multiple linkage studies to determine if the overall results are significant. We have shown that these tests are more powerful than *modified* Lander and Kruglyak criteria without increasing the type 1 error rate. (The modifi-

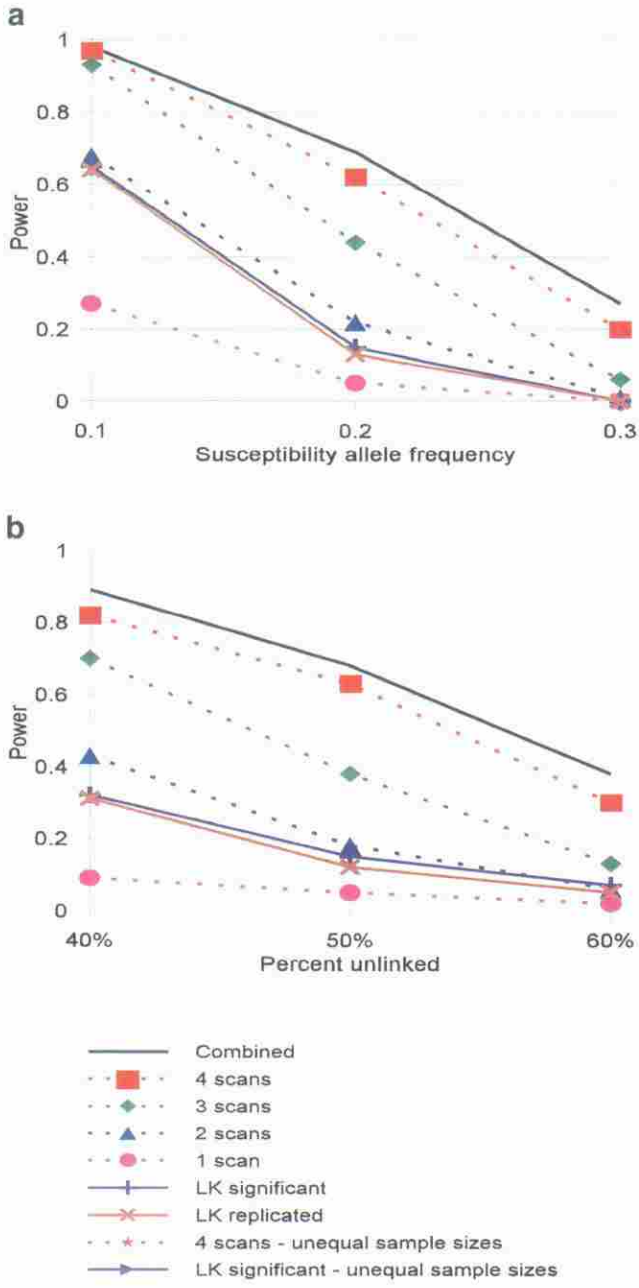

Figure 2 Power of the MSP for different genetic models. One hundred replicates of four studies, each with 100 affected sib pairs and a single marker with four equifrequent alleles linked to the susceptibility locus with recombination fraction 0.05. Power for 'LK significant' was calculated by estimating the proportion of replicates in which at least 1/4 studies had a *P*-value less than 2.2×10^{-5}. 'LK replicated' required that at least 1/4 studies have a *P*-value less than 2.2×10^{-5} and an additional study had a *P*-value less than 0.01. For the other analyses, power was calculated by estimating the proportion of replicates that had a *P*-value less than 2.2×10^{-5}. 'Combined' refers to when the data from the four studies were combined and analyzed as a single study. '*n* studies' refers to the number of studies included in the MSP analysis. (a) A single gene with a common susceptibility allele was simulated. This is similar to what would be expected in epistatic interactions. (b) A single gene with allele frequency of 0.01 which is present in only a fraction of those with the trait or disease which simulates genetic heterogeneity.

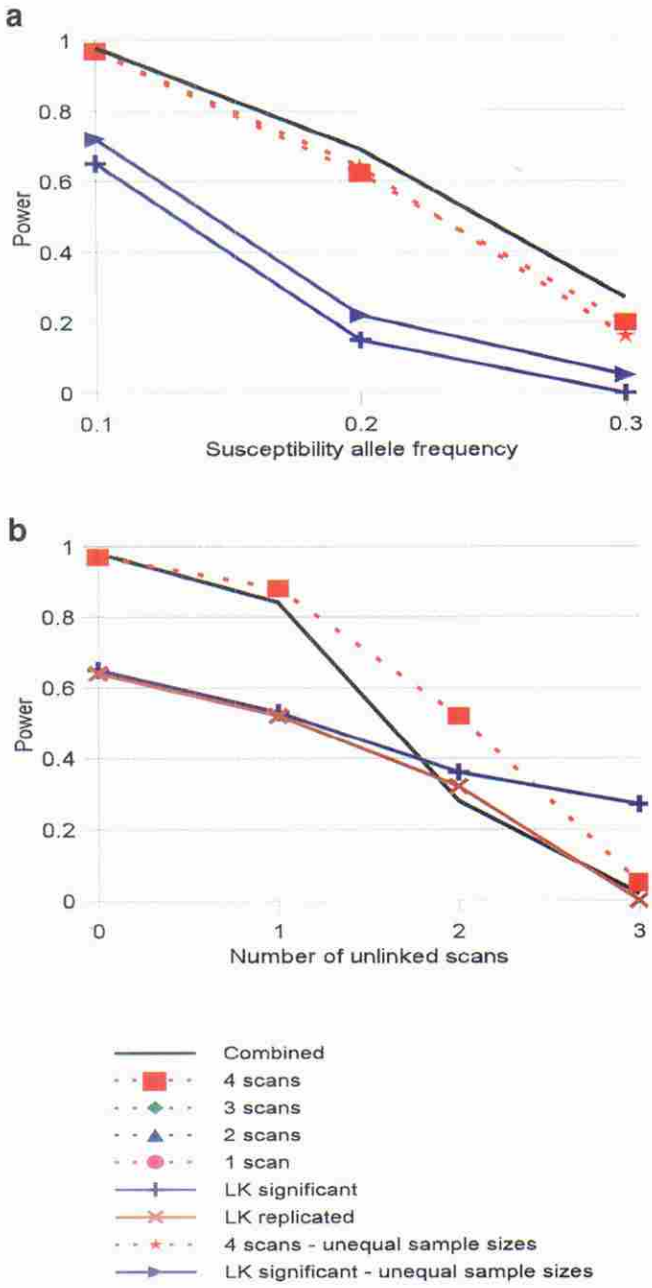

Figure 3 The effect on power of the MSP of differing study sample sizes and study heterogeneity. The same simulations were used as were used in Figure 2. Power is calculated the same way. The same genetic model as Figure 2a is used. (a) Comparison between the power for four studies with equal sample sizes with the power for four studies with four different sample sizes. The data for only the combined analysis, four studies, and the 'LK significant' are presented. The power for 'LK replicated' was virtually identical to 'LK significant' for equal and unequal sample sizes. (b) In the four studies, some studies were simulated to have no linkage between the susceptibility locus and the marker. The number of these studies is varied between 0 and 3. This is to analyze the effect of heterogeneity between the studies. The data for only the combined analysis, four studies, the 'LK significant', and the 'LK replicated' are presented.

cation, defining 'LK significant' as any one of the four studies showing a probability less than or equal to 2.2×10^{-5} and 'LK replicated' as using the criterion of 'LK significant' and also requiring that one of the remaining three studies show a probability less than or equal to 0.01, is a very liberal interpretation of the Lander and Kruglyak criterion for two studies applied to multiple studies. A more conservative interpretation would necessarily give lower power, making the MSP even more appealing by comparison.)

For the power analyses, we chose to simulate genes of small effect because these are the genes most likely to give conflicting results across linkage studies and be candidates for meta-analyses. The power to detect these genes will be low in general unless the sample size is large. It is true that power will be related to the significance criteria used. Here, we used significance criteria that are designed to give a low rate of false positives. But if the cost of missing a true locus is greater than the cost of detecting a false positive, a less stringent significance criterion may be desirable. Determining the most appropriate significance criteria is beyond the scope of this paper.

Some general problems of meta-analysis are applicable here. Meta-analysis would be susceptible to publication bias if positive findings are more likely to be published than negative findings. By selecting the results of whole genome scans, we expect to decrease the likelihood that positive findings are over-represented, although genome scans that are entirely negative may not be reported. However, nominally significant results ($P < 0.05$) are likely to occur in any genome scan by chance and are therefore likely to be reported.

For the linkage analysis of complex genetic traits, multiple analyses with different affection status models, different analytic methods, and subdividing of linkage samples are frequently performed. If the most significant results from each study are selected, this will lead to an increased type I error. There are different ways to deal with this. One way is this; before performing the analysis, it is best to develop *a priori* criteria of what sorts of results will be included. For example, results for the same affection status model should be included across studies if possible. A hierarchy of analytic methods to be included can be developed, eg, use the result of Affected Sib Pair methods if available, if not, then use non-parametric pedigree analysis results, and if that is also not available, then use parametric methods. The specific ordering of the hierarchy is not as important as the fact that it exists *a priori* and is used consistently. If results from dividing the sample are used, then the same sample subdivision should be used across studies, eg, looking at IDDM linkage results when both affected sibs have *HLA* DR3/DR4 or looking at families in which inheritance is through the paternal line. The problem with this method is that different meta-analyses on the same data could give very different results depending on which affection status model and hierarchy of analytic methods were used. Loci that were more easily

Table 3 Meta-analysis of four autism genome scans—regions in which at least one study had a nominal *P*-value less than 0.01

Region	*cM*[a]	*IMGSAC*		*Stanford*		*Paris*		*CLSA*		*MSP*	*Rep MSP*
		local min[b]	*dist (cM)*[c]	*local min*	*dist (cM)*	*local min*	*dist (cM)*	*local min*	*dist (cM)*		
1p	149	0.5	0	0.00083	0	0.5	24	0.3	26	0.021	0.7
4p	5	0.0038	0	0.5	0	0.6	0	0.3	5	0.044	0.6
4q	165	0.5	0	0.065	25	0.035	30	0.017	0	0.06	0.4
6q	109	0.5	0	0.5	0	0.0013	0	1.0	0	0.041	0.8
7q	142	0.00032	0	0.019	5	0.040	18	0.0029	38	0.00014	0.02
10p	52	0.0062	0	0.5	0	0.3	14	1.0	0	0.1	0.8
13q	19	0.5	0	0.2	0	0.6	0	0.0023	0	0.020	0.4
13q	50	0.050	23	0.038	0	0.4	5	0.00040	0	0.00060	0.07
16p	18	0.0042	0	0.5	0	0.051	5	0.3	12	0.014	0.2
19q	42	0.016	9	0.2	18	0.010	0	1.0	9	0.029	0.2
22q	11	0.0057	0	0.2	20	0.6	20	1.0	20	0.2	0.9

[a]Distance from the p-telomere–location of the most significant result of the four scans.
[b]Minimum *P*-value in the region.
[c]Distance from the most significant result of the four scans.

detected using a different affection status model than the one chosen could also be missed. Another alternative is to choose the model and analytic method with the most significant results for each study and chromosomal region and make a Bonferroni correction to the observed *P*-value, correcting for the number of affection status models and analytic methods. This correction is likely to be conservative since the results of all these different analyses will not be independent of each other within a single study.

False negatives could occur when linkage is only detectable in specific populations and hence will not show up in most studies. It can also occur if the studies included did not have the power to detect linkage for specific loci because of small sample size, low density of markers, or a method of analysis with low power.

This type of meta-analysis tests only whether the observed statistical results could have occurred by chance if there was no genetic linkage in any of the studies. If this null hypothesis is rejected, that suggests that linkage is present in one or more of the studies. Rejection of the null hypothesis does not mean that evidence of linkage is present in each of the studies that were included in the analysis. Since it is theoretically possible for the MSP to be significant because evidence for linkage was present in only one out of a large number of studies, the issue arises as to whether this is a meaningful result. For susceptibility genes in complex genetic traits, it is known that evidence for linkage can vary widely depending on the degree of genetic heterogeneity, the proportion of parents homozygous for the susceptibility gene, ethnic composition of the pedigree sample, and ascertainment of pedigrees. Thus, the presence of linkage heterogeneity between studies does not necessarily invalidate a finding. On the other hand, a result is more believable if it does occur multiple times in independent studies. The replication MSP (an analysis which excludes the most significant study) does offer a method of determining if the MSP result is primarily due to one significant study or if at least one other study is also contributing. Methods of meta-analyses which look for a consistent effect across studies and can directly assess heterogeneity between studies, for example looking at the Identity by Descent (IBD) score[3,4] are generally thought to be more accurate.[12] However, while tests of significance (*P*-value or Lod score) are usually available for each study, the same linkage parameter (eg, IBD score) is not always available in published studies, which limits the usefulness of this type of meta-analysis.

While it may appear preferable to combine data from several studies for a unified analysis rather than perform a meta-analysis, there are some problems with this approach. Often, not all data will be available for this type of analysis and there may be biases in terms of included studies being either more or less likely to show evidence of linkage than the entire collection of studies. Also, the results of the analysis involving study heterogeneity suggest that the MSP may be more powerful than a combined analysis when there is significant (~50%) study heterogeneity. This is likely due to two factors: (1) The combined analysis in this paper did not make any allowances for heterogeneity. (2) The null hypotheses of each method are different. The null hypothesis for the combined analysis is that linkage is not present in the entire data. The null hypothesis for the MSP is that linkage is not present in any of the individual studies. Therefore, evidence of linkage in only a small proportion of studies would violate the null hypothesis for the MSP but perhaps not for the combined analysis. On the other hand, one potential advantage of a combined analysis would be to narrow the confidence interval of the location estimates for a susceptibility locus and to perform more sophisticated methods of analysis with raw data. Analyses involving MSP and pooling data need not be mutually exclusive.

If an MSP analysis shows a significant result, that may give more impetus to combining data to improve localization of a susceptibility locus.

The MSP using corrected *P*-values from an *n* cM region where *n* is between 10 and 30 cM may be more conservative in terms of genome-wide significance than the MSP using *P*-values from the same point in each genome scan. But it is not clear whether it is more powerful. For autism, the former test was more significant than the latter test for 7q and 13q but this may not always be the case.

In the MSP analysis of autism, 7q ($P = 0.00014$) and 13q ($P = 0.0006$) were the most significant findings. If Lander and Kruglyak genome-wide criteria were applied to these MSP results, both 7q and 13q would meet criteria for suggestive significance ($P < 0.00074$). There is also further linkage evidence of an autism susceptibility locus on 7q by Ashley-Koch *et al.*[28] This study was not included in the original MSP analysis because the results were not reported as part of a genome scan and hence may have added positive bias to the results since such results might not have been available were they not significant. In an analysis of 76 families with at least two individuals with autism, genotyping nine markers over a 35 cM region, a maximum ASPEX LOD score of 1.77 was found 129 cM from the p-telomere. An MSP analysis incorporating the Ashley-Koch study demonstrates a *P*-value of 1.5×10^{-5}, which exceeds Lander and Kruglyak criteria for significant linkage. If the most significant study (IMGSAC) is excluded from the analysis, the *P*-value of the 'replication' MSP analysis is 0.0022. For 13q, the 'replication' MSP analysis (excluding the CLSA study) is 0.07 which suggests that most of the evidence for linkage comes from one study. This does not necessarily invalidate the MSP result but suggests that evidence for replication of the result is weak.

MSP analysis offers a method of looking at published genome scans to determine if the body of evidence suggests the presence of a susceptibility locus in a particular region. It is more powerful than requiring that one of several genome scans meets Lander and Kruglyak criteria for significant linkage without increasing false positives. It is also robust to a considerable amount of heterogeneity. The analysis has been applied to autism data and evidence of significant and replicated linkage has been found to 7q despite the fact that none of the individual studies exceeded Lander and Kruglyak criteria for significant linkage.

Acknowledgements

This work was supported by a seed grant to JAB from the Howard Hughes Medical Institute.

References

1 Suarez BK, Hampe CL, Van Eerdewegh P. Problems of replicating linkage claims in psychiatry. In: Gershon ES, Cloninger CR (eds). *Genetic Approaches to Mental Disorders*. American Psychiatric Press: Washington, DC, 1994, pp 23–46.
2 Lander ES, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet* 1995; **11**: 241–247.
3 Dorr DA, Rice JP, Armstrong C, Reich T, Blehar M. A meta-analysis of chromosome 18 linkage data for bipolar illness. *Genet Epidemiol* 1997; **14**: 617–622.
4 Rice JP. The role of meta-analysis in linkage studies of complex traits. *Am J Med Genet* 1997; **74**: 112–114.
5 Guerra R, Etzel CJ, Goldstein DR, Sain SR. Meta-analysis by combining p-values: simulated linkage studies. *Genet Epidemiol* 1999; **17**: S605–S609.
6 Allison DB, Heo M. Meta-analysis of linkage data under worst-case conditions: a demonstration using the human OB region. *Genetics* 1998; **148**: 859–865.
7 Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M *et al.* Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nature Genet* 2000; **26**: 163–175.
8 Hanis CL, Boerwinkle E, Chakraborty R, Ellsworth DL, Concannon P, Stirling B *et al.* A genome-wide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nature Genet* 1996; **13**: 161–166.
9 Hani EH, Hager J, Philippi A, Demenais F, Froguel P, Vionnet N. Mapping NIDDM susceptibility loci in French families: studies with markers in the region of NIDDM1 on chromosome 2q. *Diabetes* 1997; **46**: 1225–1226.
10 Ghosh S, Hauser ER, Magnuson VL, Valle T, Ally DS, Karanjawala ZE *et al.* A large sample of Finish diabetic sib-pairs reveals no evidence for a non-insulin-dependent diabetes mellitus susceptibility locus at 2qter. *J Clin Invest* 1998; **102**: 704–709.
11 Badner JA, Goldin LR. Meta-analysis of linkage studies. *Genet Epidemiol* 1999; **17**: S485–S490.
12 Hedges LV, Olkin I. *Statistical Methods for Meta-analysis*. Academic Press: San Francisco, 1985.
13 Fisher RA. *Statistical Methods for Research Workers*. Oliver & Boyd: London, 1932.
14 Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS. Replication of linkage studies of complex traits: an examination of variation in location estimates. *Am J Hum Genet* 1999; **65**: 876–884.
15 Feingold E, Brown PO, Siegmund D. Gaussian models for genetic linkage analysis using complete high-resolution maps of identity by descent. *Am J Hum Genet* 1993; **53**: 234–251.
16 Terwilliger JD, Ott J. SIMULATE 2.24. ftp://linkage.cpmc.columbia.edu/software/simulate
17 Ott J. Computer-simulation methods in human linkage analysis. *Proc Natl Acad Sci USA* 1989; **86**: 4175–4178.
18 Weeks DE, Ott J, Lathrop GM. SLINK: a general simulation program for linkage analysis. *Am J Hum Genet* 1990; **47**: A204.
19 Cottingham RW Jr, Idury RM, Schaffer AA. Faster sequential genetic linkage computations. *Am J Hum Genet* 1997; **53**: 252–263.
20 Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E *et al.* Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* 1995; **25**: 63–77.
21 Bolton P, Macdonald H, Pickles A, Rios P, Goode S, Crowson M *et al.* A case-control family history study of autism. *J Child Psychol Psychiatry* 1994; **35**: 877–900.
22 International Molecular Genetic Study of Autism Consortium. A full genome screen for autism with evidence for linkage to a region on chromosome 7q. International Molecular Genetic Study of Autism Consortium. *Hum Mol Genet* 1998; **7**: 571–578.
23 Philippe A, Martinez M, Guilloud-Bataille M, Gillberg C, Rastam M, Sponheim E *et al.* Genome-wide scan for autism susceptibility genes. Paris Autism Research International Sibpair Study. *Hum Mol Genet* 1999; **8**: 805–812.
24 Kruglyak L, Lander ES. Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am J Hum Genet* 1997; **57**: 439–454.
25 Risch N, Spiker D, Lotspeich L, Nouri N, Hinds D, Hallmayer J *et al.* A genomic screen of autism: evidence for a multilocus etiology. *Am J Hum Genet* 1999; **65**: 493–507.
26 Barrett S, Beck JC, Bernier R, Bisson E, Braun TA, Casavant TL *et al.* An autosomal genomic screen for autism. *Am J Med Genet* 1999; **88**: 609–615.
27 Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. Com-

prehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* 1998; **63**: 861–869.
28 Ashley-Koch A, Wolpert CM, Menold MM, Zaeem L, Basu S, Donnelly SL *et al.* Genetic studies of autistic disorder and chromosome 7. *Genomics* 1999; **61**: 227–236.
29 Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet* 1997; **61**: 1179–1188.
30 Hauser ER, Bochnke M, Guo SW, Risch N. Affected sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. *Genet Epidemiol* 1996; **13**: 117–137.
31 Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified approach. *Am J Hum Genet* 1996; **58**: 1347–1363.