

## INVITED EDITORIAL

# Statistical Approaches to Gene Mapping

Jurg Ott and Josephine Hoh

Laboratory of Statistical Genetics, Rockefeller University, New York

### Introduction

In this brief primer, we hope to provide a general overview on statistical methods for disease-gene mapping. Of course, this cannot be complete—our apologies to researchers whose methods are not mentioned below. More-detailed information may be found in relevant textbooks (Ott 1999) and at the Web Resources of Genetic Linkage Analysis site (Laboratory of Statistical Genetics, Rockefeller University). The main purpose of this primer is to present, in a nontechnical manner, the methodological background and rationale of genetic mapping and to relate the various approaches to each other. In addition, current analysis methods for analysis of microarray data are discussed. Microarray data represent a new type of information that can provide important insight about the interaction of genes and that thus can complement the statistical approaches to gene mapping.

Statistical genetic-mapping methods all rest on one biological phenomenon, recombination (crossing-over), which is exploited for the purposes of determining the genetic distance—or at least the closeness—between two loci. Crossovers between homologous chromosome strands occur semirandomly. Loci in close proximity to each other will rarely be separated by a recombination, whereas, for distant loci, recombinations occur as often as not. This phenomenon is used to derive a statistical measure of genetic distance. In family pedigrees, recombinations may be seen more or less directly; on the other hand, the consequences of recombinations in past generations can be observed in the form of linkage disequilibrium—that is, the preferential occurrence, in one gamete, of specific alleles at different loci.

### Genetic-Linkage Analysis

Throughout the human chromosomes, genetic maps have been created that consist of a dense set of genetic

marker loci—that is, loci with a known Mendelian mode of inheritance. In the early days of genetic mapping, enzyme and blood-group polymorphisms served as genetic markers. They now have all but been replaced by DNA polymorphisms, which have the advantage of lacking functional significance. The newest type of genetic marker is the single-nucleotide polymorphism (SNP), of which many thousands are likely to be identified.

To localize on the human gene map a simple (recessive or dominant) Mendelian disease gene, the most straightforward approach is to investigate haplotypes passed from parents to offspring—that is, sequences of alleles, at different loci, that are transmitted from a parent in one gamete. For example, a parent may be heterozygous for a recessive disease whose gene resides on a given chromosome. One of his chromosomes then carries a disease allele and the other (homologous) chromosome carries a normal allele. Inspection of haplotypes in the family may reveal which of the two chromosomes (haplotypes) carries the disease allele. When a chromosome transmitted from parent to child exhibits a crossover, the disease locus must be above or below the crossover. Because an affected child must have received that chromosomal portion containing the disease locus, the location of the disease gene is then known to be on one or the other side of the crossover; for an example, see figure 1 in an article by Plášilová et al. (1998). Multiple such observations eventually allow one to localize the disease gene to a small interval between marker loci.

Because of incomplete penetrance of disease genes, missing individuals, etiologic heterogeneity, and various other complications, the haplotype approach to disease-gene mapping is often not applicable. The method of choice is then an estimation of the disease locus, an estimation based on marker and disease phenotypes. The statistical principle employed is that of maximum likelihood estimation, where the likelihood is the probability of occurrence of the data, given assumed parameter values. On the basis of some or all markers on a chromosome, computer programs such as LINKAGE (Lathrop et al. 1984), MENDEL (Lange et al. 1988), VITESSE (O'Connell and Weeks 1995), ASPEX (Schwab et al. 1995), GENEHUNTER (Kruglyak et al. 1996), or ALLEGRO (Gudbjartsson et al. 2000) can compute this likelihood for any as-

Received June 12, 2000; accepted for publication June 15, 2000; electronically published July 6, 2000.

Address for correspondence and reprints: Dr. Jurg Ott, Rockefeller University, 1230 York Avenue, Box 192, New York, NY 10021-6399. E-mail: ott@rockvax.rockefeller.edu

This article represents the opinions of the authors and has not been peer reviewed.

© 2000 by The American Society of Human Genetics. All rights reserved.  
0002-9297/2000/6702-0007\$02.00

sumed position of the disease locus. The likelihood is of intrinsic value only in comparison with the likelihood that is obtained under the assumption that the disease locus is off the marker map. For this reason, likelihoods are generally transformed into so-called LOD scores and are graphed across the human gene map. A LOD score is the decimal logarithm of the likelihood ratio, with an assumed disease position in the numerator and with assumed absence of the disease locus in the denominator. Peaks on the LOD-score curve that exceed some threshold, such as 3, identify potential locations of disease genes. For background information on this methodology, see the review by Nyholt (2000 [in this issue]) and the work of Ott (1999).

Linkage analysis methods fall into two broad classes—parametric and nonparametric approaches. Methods in the first class require specification of the mode of trait inheritance—for example, values for penetrances, recombination fractions, heterogeneity parameters, and so on. If good estimates for these parameters are known, then parametric linkage analysis is very powerful. For many diseases (particularly in the case of complex traits; see below), it may not be desirable or possible to specify parameter values for Mendelian inheritance; so nonparametric methods are appealing for these traits. Methods in this class do not make assumptions about the mode of inheritance of the trait but rely entirely on the known modes of marker inheritance. Rather sophisticated nonparametric approaches have been implemented in several computer programs. Some of these approaches *estimate* allele sharing and represent results in the form of LOD scores, which may be graphed along the genome, as in the case of parametric linkage analysis. Many nonparametric methods are equivalent to parametric methods. For example, a particular form of the affected-sib-pair analysis has a 1:1 correspondence of LOD-score analysis for a fully penetrant recessive disease (Knapp et al. 1994). Thus, nonparametric methods often can be emulated by standard linkage programs (Goring and Terwilliger 2000).

Likelihood (LOD score) calculations are generally carried out recursively; that is, the data are split into suitable subsets, and calculations are performed on one subset at a time, with the results attached to the next subset, and so on. This allows for large data sets to be processed in a sequential manner. Two types of recursive procedures are in common usage: (1) recursion over family members in a pedigree (a procedure in which all loci are considered at once) and (2) recursion over loci (a procedure in which all family members are considered jointly). The first type of procedure (the Elston-Stewart [Elston and Stewart 1971] algorithm and related procedures) can handle large pedigrees but only up to a handful of loci at a time and is implemented in programs

such as LINKAGE. The second approach (the Lander-Green [Lander and Green 1987] algorithm) can work on large numbers of loci but only on small numbers of family members at a time; it is implemented, for example, in the GENEHUNTER and ALLEGRO programs. Currently, there is no exact likelihood-calculation method that can process both large numbers of loci and large numbers of family members. Approximate methods to do this do exist and are based on Monte Carlo Markov-chain methods (e.g., see Heath 1997, 1998). Rather than evaluating all genotypes compatible with the observations, these methods infer, on the basis of phenotypes, a suitable number of underlying genotypes and work on this reduced set of genotypes.

Most methodological developments in recent years have focused on more-efficient likelihood calculations and on the development of nonparametric methods based on allele sharing. Approaches of the latter type have found much use for complex traits (see below). Another new development makes use of a peculiar property of genomewide LOD-score curves for families selected to contain affected individuals (Terwilliger et al. 1997)—this property is that true peaks tend to be wider than false peaks. In likelihood theory, the only relevant quantity from this analysis is the peak height, but the article discussing this approach suggests that the width of the LOD-score curve should provide extra power for localization of disease genes. Several ad hoc methods have been proposed to make use of this phenomenon (Goldin and Chase 1997; Goldin et al. 1999). A comprehensive approach for capturing the effects of the height *and* the width of the LOD score may be effected by means of scan statistics (authors' unpublished data). In this statistic, one considers  $n$  consecutive markers and calculates the sum of the LOD scores at these markers. The maximum sum over all possible sets of  $n$  consecutive marker loci is called the "linear scan statistic of length  $n$ ." Scan statistics provide greatly improved power over regular LOD-score analysis (authors' unpublished data) and, presumably, will see much use in the near future.

### **Disequilibrium Mapping**

Linkage analysis is based on recombination and requires the analysis of family data. Other methods of gene mapping do not assess recombination in observed data but, rather, exploit the consequences of recombination that has occurred between a mutation and a marker some time ago. These methods can work on unrelated individuals and make use of unobserved recombinations in previous generations.

Consider a protein-encoding gene and an SNP (with alleles 1 and 2) in its immediate vicinity. If a mutation occurs at this locus, the mutated allele (often called "the

disease allele” or “the mutation”) will be in coupling with a specific marker allele—for example, the 1 allele. Offspring of the individual in whom the mutation occurred may then receive the disease allele along with the 1 allele. Through recombination between the disease and marker loci, the disease allele may eventually occur in coupling with the 2 allele, but, for many generations to come, the disease allele will tend to be associated with the 1 allele much more frequently than with the 2 allele. This situation is referred to as “linkage disequilibrium” (LD) or “allelic association.” When an equilibrium situation is reached, the proportion of 1 alleles in coupling with the disease allele is the same as the corresponding proportion of 2 alleles, which is just the population frequency of the disease allele. In the absence of such disturbing forces as mutation and selection, under random mating, genotype frequencies are entirely determined by allele frequencies. This situation is referred to as “Hardy-Weinberg equilibrium” (HWE). Genotypes of marker loci generally are in HWE, and deviations from it are, in many laboratories, taken as evidence for a technical problem. However, when no such problems exist, deviations from HWE may be indicative of genetic association of the given marker with a disease gene, or LD (Nielsen et al. 1998).

LD may be detected, for example, in a case-control study. Consider a sample of individuals affected with a particular trait (“cases”) and a number of individuals without the trait (“controls”). Allele frequency differences between the populations, for a given marker, are evidence for LD. LD may occur as a consequence of population admixture, but strong effects of LD are generally interpreted as being indicative of tight genetic linkage. Depending on the history of a population, LD is seen over varying distances from a disease gene. For large outbred populations, LD generally extends only up to 0.3 cM (Collins et al. 1999; Ott 2000). LD may span much larger distances in special populations—that is, populations with a small number of founders and subsequent rapid expansion—or in small stable populations (Terwilliger et al. 1998). One such population is the Afrikaans-speaking population of South Africa, in whom the LD between marker loci can extend over multiple centimorgans (Gordon et al. 2000). Because of its short range in large populations, LD is often advocated as a method for fine mapping once an approximate location for a disease gene has been found by linkage analysis. Clearly, it is important that control individuals be carefully matched to case individuals in these studies; without proper matching of populations, spurious results may be produced by this analysis, because of population stratification.

Rather than such population controls, family-based methods for LD analysis include internal controls (Falk and Rubinstein 1987; Ott 1989; Thomson et al. 1989;

Terwilliger and Ott 1992; Spielman and Ewens 1996). These approaches rest on a comparison of marker alleles transmitted by a parent to an affected offspring (“case” alleles) versus those alleles not transmitted by the parent (“control” alleles). The typical family structure for these types of analyses is a family “trio” consisting of two parents and an affected offspring. An important difference between linkage analysis and LD analysis is that, in the former, no attention is paid to the identities of the alleles that occur in recombinant or nonrecombinant haplotypes, whereas, in LD analysis, association between disease status and specific marker alleles is investigated.

Special analysis methods for family-based control data have been developed—in particular, the transmission/disequilibrium test (TDT) (Spielman and Ewens 1996). The TDT allows for multiple affected offspring and is a test for linkage in the presence of LD; without LD, it has no power. Currently, however, many scientists are of the opinion that the potentially damaging effect of population stratification may have been overestimated and that family-based control data are less efficient for detection of LD than are case-control studies (Morton and Collins 1998). Several additional methods for LD mapping have been developed. For example, Boehnke and Langefeld (1998) have introduced several family-based tests of association that use discordant sib pairs, in which one sib is affected with a disease and the other sib is not.

### Localization of Complex-Trait Genes

Many traits are clearly heritable yet do not follow a known Mendelian pattern of inheritance. In contrast to Mendelian diseases, these complex traits are rather common and presumably are due to multiple interacting genes, thus making genetic analysis more difficult. Because the inheritance pattern is not understood, researchers often prefer nonparametric methods to search for genes underlying complex traits. On the other hand, one may apply standard parametric linkage-analysis methods under a suitable assumed-inheritance model for the trait. Then, the question concerns what consequences incorrect model assumptions have on the results. Much methodological research has gone into investigation of such questions. In general, incorrect penetrance assumptions are not serious as long as recessive-like traits are analyzed under a recessive mode and dominant-like traits are analyzed under a dominant mode of inheritance. Thus, parametric analyses often are analyzed under a small number of different model assumptions and may then be at least as powerful as nonparametric analyses (Abreu et al. 1999).

The threshold for significance in the analysis of complex disease is currently under debate. In a genome scan,

a statistical test for linkage or association is essentially carried out for each marker. There is then a question of the appropriate marker-specific significance level such that the overall (genomewide) false-positive rate (type 1 error) is not, say, >5%. For a time, it appeared that this question had to be answered differently for Mendelian and complex diseases. However, under statistical testing theory, the false-positive rate is the proportion of significant results, given absence of any disease genes, and Lander and Kruglyak (1995) have proposed corresponding rigorous locus-specific significance levels. As it turns out, obtaining significant evidence for the localization of genes in complex traits is very difficult. It appears that sample-size requirements for localization of genes underlying complex traits are much higher than those for Mendelian disease genes (Suarez et al. 1994; Lernmark and Ott 1998). This seems to have resurrected previous demands that, when it comes to the establishment of significance levels, complex traits be treated differently than Mendelian diseases, the main argument being that no researcher will search for genes in a trait that is not known to be genetic. But, under statistical test theory,  $P$  values are independent of the presence of disease genes. Both sides of this argument have good reasons for their beliefs. However, if researchers investigate traits only when they are known to be genetic, and if this allows them to lower the threshold for significance, then the threshold should apply to both complex and Mendelian traits; but this argument is not generally invoked for Mendelian disorders.

As discussed so far, analysis methods are not intrinsically different for complex and Mendelian traits. Genome screens still proceed essentially by testing one marker at a time, where informativeness at a marker may be enhanced by information gleaned from neighboring markers (in “multilocus analysis,” “multi” generally refers to multiple markers, not to multiple disease genes). The specific multigenic nature of complex traits is not usually a factor in the statistical analysis. It is only recently that methods of multigene analysis have begun to be proposed—for example, by Cordell et al. (2000).

The prospect of the availability of thousands of SNPs densely spaced on the human gene map has led several research groups to collect large numbers of individuals affected with a complex trait. Massive case-control studies and studies with quantitative outcomes are planned. This poses interesting statistical problems. In principle, for a complex trait, marker genotypes should be analyzed jointly with disease phenotypes, but the potentially very large number of markers renders this impossible. Thus, two-stage procedures have been proposed such that an initial stage of marker selection is followed by a statistically sophisticated multivariate analysis of marker *and* trait phenotypes (Hoh et al., in

press). For that second stage, pattern-recognition methods may be applied—for example, logistic regression or artificial neural networks (Lucek et al. 1998).

### Microarray Expression Data

Linkage and association mapping have their limits in dealing with complex traits. To find interactions among large numbers of genes is highly desirable yet difficult to accomplish with these methods. A new type of data, generated using microarray(biochip)-based technologies, promises to provide an extensive view of biological information on the interplay of genes in the entire genome. Research into analytical methods and computer algorithms to facilitate the interpretation of microarray data is currently a very active area.

High-throughput microarray data usually consist of expression levels of thousands of genes, each possibly measured under various experimental conditions. Several clustering algorithms are now available to classify the entire set of genes into hierarchical subsets. As a result, genes with similar patterns of expression are grouped together, where similarity may be based on the Euclidean distance of the observations in the multivariate data space, a sample standard correlation coefficient (Eisen et al. 1998), or other appropriate metric systems. In general, clustering methods can be divided into two classes: unsupervised and supervised. The key difference between the two approaches is that the supervised method classifies genes on the basis of some reference information, such as groups of genes known to be coregulated. An example of the supervised method is support-vector machines, a computer-learning method (Brown et al. 2000). In contrast, the unsupervised method relies on no prior knowledge of biological functions. Since little is known about the biological properties of the genes assayed in most microarray experiments, unsupervised clustering is often chosen as the first step of the analysis. One can follow this by a supervised method, which may be referred to as “hybrid clustering,” to increase resolution (Getz et al. 2000).

Another important aspect of organizing the microarray expression data by clustering methods is that it may be done in two directions (two-way clustering), both having biological meaning: clustering with respect to the genes across the experimental conditions and clustering with respect to the experimental conditions over all the genes (Alon et al. 1999). The existing algorithms for two-way clustering separately perform clustering of genes and of conditions and then combine the results, for a graphic representation. A more challenging task is the clustering on genes and on conditions at the same time (two-dimensional clustering). To this end, a procedure proposed by Hastie et al. (2000) is the only method available at the time of the writing of the present

review (the computer algorithm is available on the Software site of the Laboratory for the Statistical Analysis of Microarray Data, Stanford University).

Clustering is employed under the presumption that data are “clean”—that is, background noise has been eliminated, outliers are detected, variance is stable across arrays for each gene, intensities are measured on the same scale, missing data are imputed correctly, and so on. It is difficult to overemphasize the importance of scrutinizing the data before the application of analysis methods. In doing so, one must understand the nature of the data, which is outlined briefly. Basically, there are currently two types of DNA microarrays—(1) cDNA arrays and (2) oligonucleotide arrays. In cDNA arrays, thousands of short-length DNAs are spotted onto glass microscope slides by a robotic spotter. These DNAs represent fragments of known or unknown genes. From cells of interest, mRNA is extracted and converted to cDNA, with the introduction of a fluorescent label. The fluorochrome-tagged cDNAs are then hybridized to the DNA spots on the slide. The intensity of fluorescence on each spot is detected with a laser and is recorded with a microscope. On the other hand, oligonucleotide arrays contain chemically synthesized 25-mer (this number may have a range of 20–50) DNA oligonucleotides whose sequences are known. The oligonucleotides are deposited on silicon chips and are used for both the subsequent hybridization with the experimental samples and the data-recording process. Data handling for these two types of arrays is similar, mainly because variations in expressions among oligonucleotides for each gene are overlooked. However, ignoring this variation could potentially be detrimental, so future analysis methods will have to examine its effects further.

## Outlook

Despite major efforts by researchers and good support by the National Institutes of Health, success in the localization of complex-trait genes has been rather modest. Although there may have existed some unwarranted enthusiasm in the initial phases of these efforts (Terwilliger and Weiss 1998), we are convinced that eventually the current large investments of time and effort will pay off.

## Acknowledgments

Support by National Institutes of Health grants MH44292 and HG00008 is gratefully acknowledged.

## Electronic-Database Information

The URLs for data in this article are as follows:

Software (Laboratory for the Statistical Analysis of Microarray

Data, Stanford University), <http://www-stat.stanford.edu/~tibs/lab/software.html>  
 Web Resources of Genetic Linkage Analysis, <http://linkage.rockefeller.edu>

## References

- Abreu PC, Greenberg DA, Hodge SE (1999) Direct power comparisons between simple LOD scores and NPL scores for linkage analysis in complex diseases. *Am J Hum Genet* 65:847–857
- Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, Levine AJ (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci USA* 96:6745–6750
- Boehnke M, Langefeld CD (1998) Genetic association mapping based on discordant sib pairs: the discordant-alleles test. *Am J Hum Genet* 62:950–961
- Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, Ares M Jr, et al (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci USA* 97:262–267
- Collins A, Lonjou C, Morton NE (1999) Genetic epidemiology of single-nucleotide polymorphisms. *Proc Natl Acad Sci USA* 96:15173–15177
- Cordell HJ, Wedig GC, Jacobs KB, Elston RC (2000) Multilocus linkage tests based on affected relative pairs. *Am J Hum Genet* 66:1273–1286
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–14868
- Elston RC, Stewart J (1971) A general model for the analysis of pedigree data. *Hum Hered* 21:523–542
- Falk CT, Rubinstein P (1987) Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann Hum Genet* 51:227–233
- Getz G, Levine E, Domany E, Zhang MQ (2000) Super-paramagnetic clustering of yeast gene expression profiles. *Physica A* 279:457–464
- Goldin LR, Chase GA (1997) Improvement of the power to detect complex disease genes by regional inference procedures. *Genet Epidemiol* 14:785–789
- Goldin LR, Chase GA, Wilson AF (1999) Regional inference with average p values increase the power to detect linkage. *Genet Epidemiol* 17:157–164
- Gordon D, Simonic I, Ott J (2000) Significant evidence for linkage disequilibrium over 5 cM region in Afrikaners. *Genomics* 66:87–92
- Goring HH, Terwilliger JD (2000) Linkage analysis in the presence of errors. IV. Joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. *Am J Hum Genet* 66:1310–1327
- Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000) Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 25:12–13
- Hastie T, Tibshirani R, Eisen M, Brown P, Ross D, Scherf U, Weinstein J, et al (2000) Gene shaving: a new class of clus-

- tering methods for expression arrays. Tech rep, Stanford University, Palo Alto, CA
- Heath SC (1997) Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet* 61:748–760
- (1998) Generating consistent genotypic configurations for multi-allelic loci and large complex pedigrees. *Hum Hered* 48:1–11
- Hoh J, Wille A, Zee R, Lindpaintner K, Ott J. Selecting SNPs in two-stage analysis of disease association data: a model-free approach. *Ann Hum Genet* (in press)
- Knapp M, Seuchter SA, Baur MP (1994) Linkage analysis in nuclear families. II. Relationship between affected sib-pair tests and lod score analysis. *Hum Hered* 44:44–51
- Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 58:1347–1363
- Lander ES, Green P (1987) Construction of multilocus genetic maps in humans. *Proc Natl Acad Sci USA* 84:2363–2367
- Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11:241–247
- Lange K, Weeks D, Boehnke M (1988) Programs for pedigree analysis: MENDEL, FISHER, and dGENE. *Genet Epidemiol* 5:471–472
- Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 81:3443–3446
- Lernmark A, Ott J (1998) Sometimes it's hot, sometimes it's not. *Nat Genet* 19:213–214
- Lucek P, Hanke J, Reich J, Solla S, Ott J (1998) Multi-locus nonparametric linkage analysis of complex trait loci with neural networks. *Hum Hered* 48:275–284
- Morton NE, Collins A (1998) Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci USA* 95:11389–11393
- Nielsen DM, Ehm MG, Weir BS (1998) Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am J Hum Genet* 63:1531–1540
- Nyholt DR (2000) All LODs are not created equal. *Am J Hum Genet* 67:282–288 (in this issue)
- O'Connell JR, Weeks DE (1995) The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. *Nat Genet* 11:402–408
- Ott J (1989) Statistical properties of the haplotype relative risk. *Genet Epidemiol* 6:127–130
- Ott J (1999) *Analysis of human genetic linkage*, 3d ed. Johns Hopkins University Press, Baltimore
- (2000) Predicting the range of linkage disequilibrium. *Proc Natl Acad Sci USA* 97:2–3
- Plášilová M, Feráková E, Kádasi L, Poláková H, Gerinec A, Ott J, Ferák V (1998) Linkage of autosomal recessive primary congenital glaucoma to the GLC3A locus in Roms (Gypsies) from Slovakia. *Hum Hered* 48:30–33
- Schwab SG, Albus M, Hallmayer J, Honig S, Borrmann M, Lichtermann D, Ebstein RP, et al (1995) Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. *Nat Genet* 11:325–327
- Spielman RS, Ewens WJ (1996) The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 59:983–989
- Suarez BK, Hampe CL, Van Eerdewegh P (1994) Problems of replicating linkage claims in psychiatry. In: Genetic approaches to mental disorders. Gershon ES, Cloninger CR (eds) American Psychiatric Press, Washington, DC, pp 23–46
- Terwilliger JD, Ott J (1992) A haplotype-based “haplotype relative risk” approach to detecting allelic associations. *Hum Hered* 42:337–346
- Terwilliger JD, Shannon WD, Lathrop GM, Nolan JP, Goldin LR, Chase GA, Weeks DE (1997) True and false positive peaks in genomewide scans: applications of length-biased sampling to linkage mapping. *Am J Hum Genet* 61:430–438
- Terwilliger JD, Weiss KM (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr Opin Biotechnol* 9:578–594
- Terwilliger JD, Zollner S, Laan M, Pääbo S (1998) Mapping genes through the use of linkage disequilibrium generated by genetic drift: “drift mapping” in small populations with no demographic expansion. *Hum Hered* 48:138–154
- Thomson G, Robinson WP, Kuhner MK, Joe S, Klitz W (1989) HLA, insulin gene, and Gm associations with IDDM. *Genet Epidemiol* 6:155–160