Preface

Over the past decade, the methods available to analyze the genetic basis of behavioral phenotypes have changed enormously. Early enthusiasm about the role that genetics would play in our understanding of mental disorders was tempered in the late 1980s by the failure to replicate a number of well-publicized claims of linkage for bipolar affective disorder and schizophrenia. It is now well recognized that the gene-hunt battle will be long and difficult. Discrepant results have stimulated the development of the genetic epidemiological and statistical strategies used to study complex genetic disorders, and have led to refinements in phenotype analyses.

Our objective in *Psychiatric Genetics: Methods and Reviews* is to provide a comprehensive overview of the tools and methods that are currently available in psychiatric genetics, as well as archetypical examples obtained using these strategies. In particular, Part II of this book tackles the following methodological issues: study design, molecular techniques, clinical interviews, and population sampling methods. Part III of this book focuses on alternative methods for characterizing phenotypes with the aim of identifying entities with better genetic validity.

In the first chapter of Part II, Elena Grigorenko and David Pauls point out the advantages and limits of each study design for genetic epidemiology. They offer a clear vision of some much-debated problems, such as the power of detection, sensitivity, and specificity of each of the methods. Thomas Bourgeron and Bruno Giros provide an overview of the classical and novel gene-identification strategies available for the study of complex diseases. The future development of relevant molecular techniques is also usefully described.

Philip Gorwood gives a comprehensive review of the available clinical interviews for the assessment of classical nosographical entities. This chapter should enable psychiatrists to choose the most relevant clinical instrument for a particular research purpose. Sampling procedures for patients, subjects at risk, and controls are critical issues for the analysis of genetic vulnerability and protective factors implicated in psychiatric disorders. The advantages and limits of these sampling procedures, as well as potential sources of bias, are considered in the chapter written by Frank Bellivier.

The third part of *Psychiatric Genetics: Methods and Reviews* is introduced by Marion Leboyer in a comprehensive review of new phenotypic strategies, i.e., candidate symptoms and endophenotypes, and their scientific rationale. Athough

research in this field started only recently in psychiatry, the applications of these strategies have already provided exciting results. The results of the leading groups in cognition, brain imaging, and biochemical endophenotypes are summarized in this section. Michael Egan and Terry Goldberg give a brilliant and comprehensive review of their own work and the literature on cognitive intermediate phenotypes in schizophrenia. They provide the first evidence of an association between a cognitive deficit and a genetic polymorphism. This result suggests that there are links between genes and behavioral phenotypes.

A review of the biochemical endophenotypes observed in personality disorders is then presented by Antonia New and Larry Siever. They hypothesize that personality disorders may form biologically mediated traits that predispose individuals to the full-blown disorders.

Robert Freedman, the leading figure in the field of physiological endophenotypes, then describes the data obtained in genetic analysis of eye-tracking dysfunction, P50 and P300 as endophenotypes in schizophrenia and alcoholism.

Joseph Callicott and Daniel Weinberger provide a thorough review of one of the most promising approaches for the identification of valid phenotypes, i.e., the union of neuropsychological experimental designs and in vivo physiological brain mapping.

In the concluding part of this book, Ming Tsuang, Levi Taylor, and Stephen Faraone give a brilliant perspective on the methodological and ethical challenges that psychiatric genetics will face in the future.

Psychiatric Genetics: Methods and Reviews tells the very beginning story of a complicated, yet promising, saga in the field of psychiatric genetics. The message is clear: it will not be possible to unravel the complexities of psychiatric genetics unless we can precisely identify the phenotypes involved.

Marion Leboyer, ***MD, PhD***
Frank Bellivier, ***MD, PhD***

2

Analytical Methods Applied to Psychiatric Genetics

Elena L. Grigorenko and David L. Pauls

1. Introduction

The development of gene-mapping methodology has not been a linear process. Instead, this development has been multidimensional, culminating in the creation of a powerful and heterogeneous collection of tools. A description of the history of the development of this would include words such as "opportunistic" (i.e., capitalizing on the newest developments in computer technology and genomics) and "problem-solving oriented" (i.e., constantly addressing issues (such as the spotted nature of linkage disequilibrium) that arose during the development of the methodology). Therefore, the following presentation is method-oriented rather than problem-oriented. In describing the modern methodology of gene mapping, attempts will be made to describe the origin of a given methodology, the problems it was designed to address, and its known strengths and weaknesses.

There are several ways to categorize current approaches to gene mapping. One possible subdivision is whether a given methodology is a linkage approach or an association approach. A second possible

From: *Methods in Molecular Medicine, vol. 77: Psychiatric Genetics: Methods and Reviews*
Edited by: M. Leboyer and F. Bellivier © Humana Press Inc., Totowa, NJ

division would focus on whether a methodology deals with related individuals (e.g., family members) or unrelated individuals. A third possible division would consider the approaches dealing with related individuals only, summarizing the methods on the basis of the unit of analysis employed (i.e., the type and size of family units—sib-ships, nuclear and extended families, distant relatives, and so on). By necessity, these subdivisions are not exact because of the nature of data collected from families. And as would be expected, there are modern methods that simultaneously evaluate linkage and association, combine information from samples of related and unrelated individuals, and utilize multiple types of relatives.

This chapter is organized as follows. First, linkage methods are reviewed. Then, association study methods are summarized. And finally, the strengths and weaknesses of both approaches (pitfalls unique and common to both) are discussed.

2. Linkage Methods

Newton Morton is generally credited with initiating modern gene-mapping methodology with the publication of the classic paper in which he first introduced the lod-score method ***(1)***. The lod-score method allowed an estimate of the position of a disease gene on a map of markers by examining the likelihood of linkage given a specific genetic model and a specific recombination fraction. In later modifications, it was possible to incorporate incomplete disease allele penetrance and/or the absence of some key individuals in the analyzed pedigrees. Lod (log of the odds) scores consist of the base 10 logarithm of the likelihood ratio of two hypotheses. The first hypothesis postulates that a hypothetical gene is linked to a genetic marker at a given distance determined by the recombination fraction. The second hypothesis postulates no linkage (i.e., the recombination fraction is assumed to be 0.5). The base 10 logarithm of the ratio of the likelihoods of these two hypotheses is defined as the lod score. A separate lod is calculated for a range of recombination fractions. The test for linkage is conducted by examining the maximum value of the lod score for this range of recombination fractions.

The first lod-score test took the form of a sequential probability ratio test ***(1)***. This test was ideally suited for a Mendelian, single-gene mode of inheritance. In the early seventies, the method was extended with the introduction of the Elston-Steward ***(2)*** algorithm that allowed for complex inheritance (e.g., reduced penetrance) in large extended pedigrees. This algorithm was incorporated into the computer program LIPED ***(3)***. The development of LIPED and the advent of faster computers transformed linkage analyses from a time-consuming sophisticated "ordeal" into a common research tool. A major limitation of LIPED was its capacity to deal with only one marker at a time. Thus, a new set of programs ***(4)*** was developed that allowed linkage analyses of multiple markers simultaneously.

At the present time, most linkage analyses utilize multipoint strategies. It is well-known that these methods increase power when analyzing both Mendelian ***(4)*** and non-Mendelian (so-called complex traits) ***(5)***. A number of additional methods have been developed that facilitate the analysis of the multipoint data that are generated by studies performed at today's accepted marker density (10–25 cm marker spacing) ***(6)***. These methods include the exact enumeration of multi-locus genotype probabilities in small pedigrees ***(7)***; estimation of such probabilities for pedigrees of any size and of some complexity ***(8–10)***; and approximation of such probabilities for pedigrees of arbitrary size ***(11)***.

Yet, the lod-score method is preferred for Mendelian traits with (approximately) known inheritance parameters. However, the power of lod-score methods is reduced (sometimes dramatically) when the mode of inheritance ***(12–14)***, penetrance ***(15)***, and disease allele frequency ***(16–17)*** are not known and therefore possibly misspecified. Although this is a potential shortcoming of this method, it has been shown that when lod-score methods are applied many times with different modes of inheritance (e.g., dominant and recessive), a correct approximation of the mode results in lod scores that are generally superior to those obtained through other types of analyses ***(15)***.

Moreover, researchers have developed statistical methods that appear to be robust to misspecification of selected parameters. For example, a likelihood-based efficient score statistic ***(18)*** permits

testing the null hypothesis of no trait locus in a given chromosomal region. This statistic is asymptotically equivalent to the lod score, and it generalizes to a class of statistics developed for a non-parametric approach that examines only affected members of a pedigree ***(7,19–21)***. One advantage of this approach is that in the absence of complete information about the genetic model parameters, this statistic is easier to compute than the exact lod score. It does not require likelihood maximization with respect to the unknown parameters.

Although parametric linkage approaches are continually developed and remain heavily used in the field, the main disadvantage of these methods is that genetic model parameters (i.e., disease allele frequency, mode of inheritance, and penetrance) must be specified. By definition, this is not possible for complex (non-Mendelian) traits. To overcome this dilemma, non-parametric linkage methods have been developed.

Non-parametric linkage methods allow for the study of linkage between a marker (or a set of markers) and a disease without the need to specify the genetic model parameters for the trait under investigation. In classical statistics, non-parametric methods refer to methods in which observed values are replaced by their ranks. In human linkage analysis, non-parametric methods refer to methods in which parameters of disease inheritance are replaced by parameters of inheritance of markers hypothesized to be close to disease loci. An entire constellation of computer software has been developed since the 1990s (for review, *see* http:\\linkage.rockefeller.edu). This development capitalized on and was stimulated by progress in methods for likelihood calculations ***(7,9,22,23)***. Considering that the development of non-parametric methods started significantly later than that of parametric methods, most of them have developed the capacity to analyze both single and multipoint linkage data. For example, methods implemented in programs such as ASPEX ***(24)***, GENEHUNTER ***(7,25)***, and ALLEGRO ***(26)*** can utilize information from all markers on a chromosome and render any point along the chromosome as informative as possible.

It is important to remember, however, that the distinction between parametric and non-parametric methods is not sharp. In fact, it has

been shown that the affected sib-pair paradigm, a clearly non-parametric method in which the only connection to the disease is through the ascertainment scheme (i.e., families are studied in which there are at least two affected siblings) and which bases all calculations on the sharing of markers between these two affected siblings (i.e., no assumptions about parameters such as mode of inheritance or disease penetrance are necessary), is equivalent to the lod-score method when the latter is carried out under assumptions of recessive inheritance with full penetrance and all parental phenotypes are taken to be unknown ***(27,28)***. This implicit similarity is apparent in the use of the ANALYZE program, which emulates affected sib-pair analysis through lod-score analysis.

Whether parametric or non-parametric, linkage approaches utilize family data (its various configurations—siblings, nuclear families, or extended families) with the purpose of estimating the relevant parameters such as recombination fractions (map distances) in intervals between gene loci given certain sets of allele frequencies. These estimations are accomplished by maximum likelihood methods with recursive, family-based calculations of likelihood.

The most common procedures for numerical likelihood evaluation are the Elston-Steward ***(2)*** and second the Lander-Green (1987) ***(29)*** algorithms. The Elston-Steward algorithm (and its extensions) is based on pedigree traversing ("peeling") algorithms. With this approach, pedigrees are split into portions that are handled recursively, resulting in the evaluation of the full pedigree likelihood. Procedures of this type have been implemented in such programs as LIPED, LINKAGE, MENDEL, and VITESSE. The Lander-Green algorithm carries out peeling over loci; this algorithm is implemented in MAPMAKER, CRI-MAP, and GENEHUNTER. Thus, the methods have reciprocal profiles—the first method allows for the analysis of large pedigrees, but the number of gene loci that can be analyzed simultaneously is currently limited (the computational burden increases linearly with family size but exponentially with the number of loci), whereas the second method allows for the analysis of a relatively large number of loci in small pedigrees (the computational burden increases linearly with the number of loci and

exponentially with pedigree size). In addition, the development of the Markov chain Monte-Carlo methods of estimation of likelihoods ***(9,30)*** has allowed the analysis of large families and large numbers of markers (disease genes).

The common assumption for all these methodologies is that there are genes of major effect that "cause" the disease in question. Although this assumption has been modified to some degree in some of the software packages (e.g., the assumptions of heterogeneity within families (for example, as implemented in HOMOLOG and HOMOGM) and varying penetrance), it has limited investigators in the range of genetic systems that can be examined. For the most part, all analytic models are restrained to isolated chromosomes, treating multiple disease loci as if they were independent of each other.

This limitation has been recently addressed by a number of researchers interested in understanding the genetic etiology of complex traits. As noted here, by definition, complex traits are non-Mendelian, and thus are most likely influenced by multiple genetic and non-genetic factors. It is hypothesized that susceptibility to disease results from gene-gene and gene-environment interactions. In fact, the majority of medically and developmentally interesting traits are complex traits that are best conceptualized as quantitative rather than categorical. Methods developed to facilitate the identification of genomic locations of loci contributing to quantitative traits attempt to estimate the variance components associated with individual loci. Usually, such estimations are carried out using the concept of measured-locus heritability. There has been some debate in the literature as to whether there is a universally unbiased estimate of heritability and whether this estimate can be obtained ***(31–33)***. At the present time, there are no universally accepted measured-locus heritability estimates. The choice of an ideal estimator is a function of the sample size and magnitude of the locus-specific contribution to the overall phenotypic variance. Fortunately, the observed biases resulting from the use of different estimators are small, and, thus, this shortcoming should not be viewed as endangering overall outcomes of quantitative trait-linkage analyses.

There are two major classes of methods used for the identification of quantitative trait loci (QTLs), although arguably, the dividing line is artificial. The first class of methods is based on the regression of trait differences between sib-pairs on the number of alleles shared identical by descent (IBD) at a locus being tested ***(34)***. As noted, this approach is confined to sib-pairs and is not applicable to data collected from larger pedigrees.

The second class of approaches is based on classical variance-component analysis. This technique simply separates the total variance into components because of genetic and environmental effects ***(35)***. The first application of this approach to linkages analysis was developed by Hopper and Matthews (1982) ***(36)***. The focus of the method is in modeling an additional variance component for a hypothesized QTL near a marker site and establishing linkage to the marker in the presence of a statistically significant nonzero value for the QTL component (a relative size of the component is interpreted as an indicator of the magnitude of the effect of a detected locus).

Early implementations of the variance-component methodology were based on analysis of only one or two markers at a time ***(37–39)***. Then the methodology was extended to multipoint applications ***(11)*** and further strengthened by the added power of an exact multipoint approach ***(40)***. A number of simulation studies have demonstrated that the variance-components approach appears to be more powerful than the Haseman-Elston regression approach ***(11,41–44)***.

Demonstrating linkage between a disease gene and a marker is only the first (and, sometimes the smallest) step in the process of cloning the gene of interest. Traditionally, after establishing linkage, further recombination mapping techniques have been applied to narrow the region of interest. However, recombination mapping has not yielded significant success for complex traits in refining the region once it has been reduced to one or two megabases, since it is improbable that recombinants will be observed in extant family material ***(45)***. To address this challenge, researchers have developed a number of other methods. One successful approach is based on the observation that ancestral recombinants can produce a

predictable pattern of linkage disequilibrium between the disease gene and a set of markers spanning the critical region ***(46–48)***.

3. Association Methods

Whereas linkage analysis focuses merely on the position of a tested marker, association methodology tests whether a particular allele of a marker, a specific genotype, or a haplotype is enriched in (or statistically associated with) affected individuals compared with unaffected controls. In other words, genetic association studies evaluate the relationship between genetic variants and trait differences in a general population.

Association is observed either because the genetic variant being examined is a functional variant of a gene or the marker is in linkage disequilibrium with a susceptibility gene. When two markers are in linkage disequilibrium (LD), alleles at one locus will show a strong statistical association with alleles at a nearby locus, whereas alleles at distant loci will show no association. If one of these loci is a susceptibility gene, an association between an allele at the first locus and the disease being investigated will be observed. This circumstance forms the basis of LD mapping. The intuitive basis of this method is that specific alleles at loci that were immediately adjacent to the disease locus when it arose (through mutation) will tend to remain on the same chromosome as the disease locus (because of the paucity of recombination events), and thus will be transmitted together with the disease locus from generation to generation.

The genetic association study design has a controversial history in genetic research. Nevertheless, its popularity has grown remarkably during the last few years. The major reason for this growth is the increased number of genetic polymorphisms available to investigators. Ten years ago, the paucity of markers available to researchers made association studies tenuous at best. However, technological advances over the last 2–3 yr have resulted in the identification of nearly 2,000,000 DNA polymorphisms ***(49–50)*** and LD mapping studies are now becoming more feasible. Furthermore, with the

development of more efficient high-throughput genotyping methods, a growing understanding of the underlying structure of the complex phenotypes and the continued development of statistical methods, association approaches have become even more attractive.

The analysis of LD has been widely used for fine-genome mapping and has proven to be fruitful (*see* **ref. *51*** for theoretical support for the empirical success). These successful applications have included (but have not been limited to) simple disequilibrium mapping, examination of the pattern of pairwise disequilibrium between the disease gene and each of a set of markers ***(48,52)***, likelihood-based analyses ***(46,53,54)***, and haplotype fine mapping ***(55)***.

The goal of all these methods is to identify the precise disease-causing DNA variant(s) in a region that is known to be linked and associated with a disease. Within a targeted region, two association strategies are common: a positional candidate approach and a positional cloning approach. Within the positional candidate approach, specific genes or variants are examined on the basis of proposed relationships with the phenotype. Within the positional cloning approach, markers are selected for evaluation purely on the basis of their proximity to one another on a chromosome. These two types of positional searches are usually preceded by replicated linkage data, which typically narrow a region of interest to 1–10 cm. Both positional strategies have been successfully employed in the searches for genes in fully penetrant gene disorders such as cystic fibrosis and Huntington's disease ***(48,56,57)***. However, the application of these strategies has been less useful in complex disorders. A possible reason for this lack of success is that complex disorders are likely to be caused by multiple genes of moderate/small effects, making identification of the underlying genes more difficult. One of the pitfalls of the research on complex disorders using the LD method is our limited understanding of the extent to which LD occurs across the genome ***(58)***. Specifically, there may be a region in which only one functional variant may be relevant to the disorder, but LD could be present across multiple markers in the region,

making the task of "closing in on" the variant of interest much more challenging ***(59)***.

Two design strategies are employed in most association linkage-disequilibrium studies: population case-control designs and family-based association designs.

3.1. Case-Control Studies

The case-control design is the most frequently used design of association studies. The advantage of this design lies in the fact that cases are readily obtained, and can be efficiently genotyped and compared with control populations. The disadvantage of this approach is the difficulty in identifying an appropriate group of matched control cases. It is essential to establish an appropriate control sample, because any systematic allele frequency differences between cases and controls can appear as disease associations—although these may actually result from a number of other factors including but not limited to evolutionary history, group (e.g., ethnicity and gender) differences, and cultural traditions (e.g., mating customs).

The case-control design has been widely used, and its weaknesses are well-known. Specifically:

1. Association studies are often characterized by high rates of Type I (false-positive) errors—a statistically significant association between a phenotype and a polymorphism resulting from randomness in ascertainment of the case and control individuals. The danger of Type I error is increased in situations of multiple tests and relatively small sample sizes of case and control individuals. One reason for a Type I error is population stratification—a characteristic of a population in which cases and controls differ, not only with respect to the phenotype of interest and its genetic etiology, but also with respect to their overall population genetic ancestry (i.e., their general range and frequency of polymorphisms). The result of population stratification is that many irrelevant markers appear to be disease-associated.
2. In the presence of genetic heterogeneity, in which there may be many distinct and potentially interacting environmental and genetic risk factors, it is likely that no single tested genetic marker will pre-

dict disease accurately enough to be statistically apparent within the cost-effective limitations of a single study. Thus, at the present time, sample sizes may be too small to detect real associations.

3. Since association studies usually test many polymorphisms, the majority of them utilize conservative multi-test corrections (e.g., Bonferroni correction for *N* tests with a target per-test statistical threshold of *p*-value). However, there is no clear understanding of the magnitude of the Type II error (missed signal error) imposed by such corrections. These corrections may be especially detrimental for alleles with small main but large interactive effects.
4. Another source of false-positive findings is "cryptic relatedness" ***(60)***—an association between affected individuals sharing a genetic disorder. In the presence of cryptic relatedness, test statistics for case-control studies are likely to be inflated, relative to expectations, under the assumption of an independent sample and no genetic association with the disease.
5. Since LD appears to be variable over the genome, the current statistical procedures may not be sensitive enough to allow for the adequate evaluation of statistical significance of specific regions of interest.

Although the limitations of association studies are well-recognized, the association design represents an essential step in the identification and description of disease-mediating genetic variants. In the last several years, a number of proposals in the literature have been made, which should help to overcome some of the limitations of case-control studies. These are summarized here.

Cardon and Bell ***(59)*** suggest that the most appropriate way to ascertain a control sample is through a prospective cohort study. This approach requires the ascertainment of a large population sample of individuals, selected before the onset of disease, who are then followed prospectively until onset of the disease of interest. After the disease has manifested in some individuals, a group of affected individuals would be chosen and matched to a group of unaffected individuals who are part of the same original population sample. Although this approach may be feasible for disorders with relatively early onset, it would be prohibitively expensive for diseases of late onset.

Another possible way to approach the problem of stratification would be the recruitment of several control populations reflecting the various substructures that may exist in the case population. For example, one control population could be matched with the case population for age (to account for cohort-specific mating, migration, and other effects), whereas another control population could be matched with the case population for geographic location. The results of such multiple matching would be the comparison of the case population with a panel of subpopulations representative of the observed stratification.

Another very important consideration in designing an association study is that of power. Simply stated, for association studies to succeed, the samples should be large. This point has recently been vividly demonstrated in studies on the role of polymorphisms around the angiotensin l-converting enzyme (*ACE*) locus and its contribution to the risk of cardiovascular disease. One of the early publications on the role of this gene was conducted on samples of hundreds of men who had survived myocardial infarction and matched controls ***(61)***; it was reported that the *ACE* locus played a role in the risk of particular subgroups to cardiovascular disease. A series of replications, carried out with even smaller sample sizes, produced variable results ***(62)***. The hypothesis was then tested on samples involving thousands of individuals, and was not verified ***(63)***. Thus, for association studies aimed at identifying genes of moderate effects, samples should be comprised of thousands or even tens of thousands of individuals (also *see* **ref. *64***, for research on diabetes). There are very few association studies in which sample sizes approach the ones cited here. If samples of this magnitude were studied, it is likely that the number of unreplicated results would probably decrease ***(59)***.

One important advantage of case-control association studies is that DNA samples from cases and controls can be pooled and genotypes can be grouped together to determine differences in allele frequency across groups of affected and unaffected individuals. This technological advancement, recently applied in a number of contexts ***(65–67)***, must be extremely precise—the difference in

allele frequencies can be quite small and an experimental error of 1–2% can be high enough to jeopardize the outcome. When it is accurate, this technology allows rapid processing of samples from many individuals. However, its application is limited because it does not lend itself to direct haplotype assessment.

Although much work has been devoted to the development of research designs and analytic strategies to minimize Type I errors, it should be noted that the best way to confirm results is through independent replication. For example, Emahazion et al. ***(68)*** argue that Type I errors should be accepted as inevitable. These researchers suggest that association studies should be viewed as a way to screen large numbers of genes or markers, and that statistical thresholds should be chosen that would help identify genes of moderate-to-large effects. They further propose that there should be widespread efforts to replicate these findings. In addition, in an attempt to minimize the false-positive load, the association studies should be designed to minimize the clinical and population heterogeneity and to maximize the utilization of markers with known functional importance.

Although it is inevitable that there will be false-positive results, efforts should be made to attempt to minimize them. One recent approach has been suggested by Devlin and Roeder ***(60)***. These investigators have described a population-based association method using what they describe as a "genomic control" (GC). This method should help to minimize Type I errors that are caused by inappropriate matching of cases and controls. This method is designed to address two major problems that are characteristic of association studies—population stratification and cryptic relatedness. The method requires the additional genotyping of markers that are unlikely to affect liability (null loci). Chi-square statistics are calculated for both null and candidate loci. Utilizing the information on the variability and magnitude of the test statistics observed at the null loci, which are inflated by the impact of population stratification and cryptic relatedness, a multiplier is derived to adjust the critical values for significance tests for candidate loci, permitting analysis of stratified case-control data without an increase rate of

false-positives. If population stratification and cryptic relatedness are not detected from null loci, then the GC method is identical to a standard test of independence for a case-control design.

As previously mentioned, there are limitations to the case-control design. Yet it is clear that this paradigm can be a powerful tool to demarcate the genetic region of a disease-predisposing gene. As Jorde et al. ***(69)*** have argued, the application of association methodologies is especially useful in the case of markers that are tightly linked to a disease gene, when other mapping techniques become difficult. Yet given the variability of LD across the genome, once recombination distances between marker and disease genes become very small, accurate estimates of map position may become very difficult or impossible ***(70)***.

In summary, case-control studies should be considered to be one of several tools that may be useful in identifying susceptibility loci. It is unlikely that they will allow the identification of all genes of interest without other tools. Yet they may be very helpful in combination with other approaches, and they could be particularly helpful in situations in which the disorder under investigation has relatively late onset, making it difficult to obtain the family materials that are essential for other strategies.

For investigators who are considering case-control design, certain recommendations should be considered. First, the study should be designed to minimize population substructure. Second, when highly stratified populations are chose, every effort should be made to describe the substructures as much as possible and account for them in the ensuing statistical analysis. Third, if there is any doubt as to whether the sample being investigated is stratified, investigators should select null loci with common alleles and genotype them so that the GC approach can be utilized.

3.2. Family-Based Studies

An alternative approach for association studies that uses nuclear-family data to estimate control-marker allele frequencies was introduced by Rubinstein and colleagues ***(71)***, Field and colleagues ***(72)***,

and Falk and Rubinstein ***(73)***. The main objective for the development of this approach was to address the problem of population stratification caused by the ethnic mismatching between patients and randomly ascertained controls.

This approach is sometimes referred to as AFBAC (affected family-based controls), and is based on the assumption that the parental marker alleles that are not transmitted to an affected child can be used as control alleles. This matched design for patient (parental transmitted) and "control" (parental non-transmitted) marker alleles avoids ethnic confounding in the case of a stratified population ***(74–75)***. Thomson ***(76)*** demonstrated that for any single-locus model of disease susceptibility and for any nuclear family-based ascertainment scheme, the family-based association tests are an appropriate method for mapping disease genes.

If the "control population" is constructed from the non-transmitted parental alleles, a statistic known as "haplotype relative risk" (HRR—the family-based equivalent of the odds ratio or relative risk for rare diseases in a case-control study) can be computed if it can be assumed that there is random mating and that the population is in Hardy-Weinberg equilibrium ***(71,73,75,77–83)***.

Ott ***(78)*** discussed the statistical properties of the HRR in relation to the null hypothesis being tested. When random mating is assumed, the HRR statistic is equal to 1.0 when (*1*) there is no association between the marker and disease loci at the population level, (*2*) the marker and disease loci are unlinked, or (*3*) both (*1*) and (*2*) are true. However, when HRR = 1, the application of the conventional chi-square test is valid only under the assumption of random mating and when both (*1*) and (*3*) are true. If mating is nonrandom, the valid test for the condition (*2*) is the McNemar test, a statistic used in the evaluation of the "the transmission/disequilibrium test" (TDT) discussed here.

There has been considerable debate in the literature as to whether tests by HRR, contingency table, or McNemar statistics are tests of linkage or association ***(84–86)***. Thomson ***(76)*** has argued that none of these tests are association or linkage tests, according to the traditional definitions of these terms. He stated that these family-based

analyses allow detection of associations of marker genes in the presence of linkage to a disease gene, and therefore necessitate both association and linkage. A number of researchers ***(69,87)*** have noted that the requirement of association at the population level is usually a much more stringent condition than a requirement of linkage. Moreover, when there is no recombination in a randomly mating population, the quantities evaluated by HRR and contingency-table statistics can be compared to those obtained in case-control association studies. Terwilliger and Ott ***(79)*** demonstrated that when random-mating assumptions can be made, the contingency-table statistic is slightly more powerful than the HRR or McNemar tests. Only with large population stratification effects is the power of the McNemar test larger than that of the contingency-table test ***(76)***.

The family-based association paradigm has been extended to allow the incorporation of additional family members. For example, Field ***(88)*** and Thomson et al. ***(89)*** extended this approach to nuclear pedigrees ascertained for the presence of at least two affected siblings. In this design, the alleles that are not transmitted to either sib in the affected sib-pair are used as "control" alleles. Using the AFBAC approach for families with two affected siblings, Thomson and colleagues ***(89)*** showed a significant association between the class 1 allele of the 5' flanking polymorphism of the insulin gene and insulin-dependent diabetes (IDDM). Notably, affected-sib-pair-haplotype-sharing data showed no evidence of linkage to this marker ***(90)***.

Another application of this general approach is the transmission disequilibrium test (TDT) ***(81–82)***. The development of the TDT was motivated by the need to have a test of linkage in the presence of LD. However, it has been primarily used as a test of LD ***(91–92)***. The TDT has gained tremendous popularity because of its low computational demand and the fact that it is applicable to the most common study design used in complex diseases—that of affected and discordant sibling pairs ***(93–98).*** Further developments in TDT approaches resulted in inclusion of a number of additional statistical tests allowing investigation of maternal vs paternal marker association effects; marker associations that are genotype-dependent,

and maternal/fetal interaction effects, both allele- and genotype-specific ***(76)***.

Seltman, Roeder, and Devlin ***(99)*** have developed a strategy known as "evolutionary tree-TDT" (ET-TDT) by combining the theory of TDT with that of measured haplotype analysis (MHA) ***(100)***. MHA utilizes the evolutionary relationships among haplotypes to produce a limited set of hypotheses with regard to a subset of haplotypes. Thus, ED-TDT screens available haplotypes, clusters them, and points to the ancestral ones, which are especially useful for the determination of which polymorphisms within the haplotype are related to disorder liability. Finally, another very recent extension of the TDT for discrete traits includes the genome-wide analyses of SNPs ***(101)***.

Researchers ***(102)*** have compared the efficiency of the GC approach and the TDT method in the presence and absence of population stratification. When population substructure is absent, GC is found to be more efficient than TDT. In the presence of stratification, the GC method is an effective way to control for false-positives. Yet another advantage of GC is its applicability to the data obtained from small isolated populations, in which cryptic relatedness is often present (kinship is often established even between apparent non-relatives).

One disadvantage of the TDT is its reliance on heterozygous parents. Because not all parents will meet this criterion, many may have to be eliminated from the analyses, and this can result in a substantial loss of statistical power. In addition, these family-based approaches (including the TDT) require parental data that may not always be available, especially for disorders with late onset. Thus, although they are more robust in the presence of population stratification, the family-based methodologies are often less practical. Furthermore, in the presence of high homozygosity in families of affected individuals, these approaches could require sample sizes even larger than those for case-control studies to achieve adequate power.

Another disadvantage of the family-based approaches in general is that transmissions are sometimes difficult to resolve when parents

and offspring are all heterozygous for the same bi-allelic marker. To address this problem and increase definitive transmissions, several authors have proposed the use of haplotypes ***(103-108)***. With the exception of cases in which the markers being tested are functional variants of the susceptibility gene, transmissions from parents to offspring are more informative for haplotypes than single markers. However, it should be noted that using haplotypes increases the degrees of freedom of the test and thus reduces the power of the test.

In addition to the HRR and TDT, researchers have developed a number of statistical techniques to test for a marker/disease association by using nuclear-family data. In all of these approaches, contingency table analyses are used to examine the distribution of specific parental alleles among affected individuals.

Assuming random mating and no marker association with disease, a contingency table of parental transmitted vs non-transmitted alleles can be compared by means of the chi-square statistic ***(72,79,81,88,89).*** However, when there is evidence for non-random mating, the McNemar test can be applied to test deviations from the expected 50% transmission ratios of marker alleles from heterozygous parents ***(74,75,79,81,82,88,109–111)***.

Ott ***(78)*** and Knapp et al. ***(77)*** have demonstrated that the utilization of nuclear family-based data in the framework of association studies confounds tests of association and linkage. Family-based association studies will detect marker/disease associations only if the marker and disease genes are in LD. A number of comprehensive statistical packages have been developed that combine parametric and non-parametric linkage and disequilibrium analyses ***(112)***. For example, Göring and Terwilliger ***(16-17)*** estimate a test statistic that consists of three components: (1) linkage within sib-ships, (2) linkage between sib-ships, and (3) association between pedigrees. Unfortunately, at the present time, most of these methods are limited to studies in which the phenotypes are categorical.

As is the case for other analytic methods, the development of the association methodology for quantitative traits has lagged behind ***(32,113)***. Yet several developments should prove helpful in the

study of complex quantitative phenotypes. Allison ***(114)*** proposed a method for detecting linkage disequilibrium in proband/parent pairs for quantitative traits, and Rabinowitz ***(115)*** has extended this method to incorporate data from families. Subsequently, Fulker and colleagues ***(116)*** described a variance component model for the analyses of quantitative data generated from sib-pairs (in the absence of parental data). This method provides tests of linkage and association separately. Cardon ***(117)*** extended the model developed by Fulker et al. by describing a regression model for the analysis of LD in quantitative traits. One advantage of this extension is its relative ease and speed of application. And finally, Abecasis, Cardon, and Cookson ***(118)*** have extended Fulker's method to allow for sib-ships of any size, with or without parental data. With this approach, association is partitioned into two categories: between and within family components. One advantage of this method is that using families with multiple siblings can increase power. This extension is quite useful from a practical point of view. It is to be expected that in any study there will be families of variable sib-ship sizes and occasional missing parents. This method allows the use of all data collected.

In sum, association studies (whether case-control material or family-based) have both strengths and weaknesses. The eventual success of such studies is dependent on a more complete understanding of the distribution of LD across the genome, among other things. Given the information that has become available from the Human Genome Project, it is clear that more challenges remain in our attempts to identify genes of import for complex psychiatric traits. It is quite possible that new discoveries may challenge or strengthen some assumptions regarding association methodology. Nevertheless, association studies can be a valuable tool in identifying susceptibility genes, and can also help us to understand how the genome is organized and how it functions. However, as with any approach, this method must be applied with care. Investigators must be aware of the potential weaknesses in the results obtained and interpret their data accordingly. Caution and careful interpretation should be the mantra of all scientists, and this is especially true for researchers who study the genetics of complex psychiatric disorders.

3.3. Association Approaches Using Single-Nucleotide Polymorphisms (SNPs)

As noted, in order for association studies to be successful, a large number of closely linked markers spanning the regions of interest must be genotyped in order to demonstrate LD with the susceptibility gene. And this must be done inexpensively. Single-nucleotide polymorphisms (SNPs) ***(119–120)*** are a recently discovered class of polymorphisms that have been suggested as the markers of choice for such endeavors. SNPs are the most frequent type of variation in the human genome; the SNP refers to a position at which two alternative bases occur at appreciable frequency (>1%) in the human population. SNPs can be powerful tools for a variety of medical genetic studies (although individual SNPs, which have only two alleles, are less informative than currently used genetic markers (SSLPs—simple sequence-length polymorphisms), which are mostly multi-allelic), since they are much more abundant and the automatization of their processing can be done more easily than that of SSLPs ***(121)***.

SNP-based studies can be completed on either case-control or family data. The typical design of such a study relies on genotyping of a number of SNPs from candidate genes or regions (particularly those with hypothesized functional importance) in relatively large samples of affected and control participants gathered from families or specific populations. By genotyping many SNPs in a small region (or gene), it is likely that LD will be observed. It has been suggested that this approach should have the potential to identify common alleles that confer a twofold increased risk of disease. However, a number of investigators have suggested that this may be an optimistic prediction ***(122–127)***. The major concerns are: whether such common pathogenic variants exist for diseases of interest, and if so, whether sufficiently dense and powerful scans could be conducted given the diverse nature of human populations and the variability in the nature and extent of linkage disequilibrium across the genome ***(68)***.

As mentioned here, a generally accepted strategy in the mapping of a disease gene is to initially apply linkage analysis for an approx-

imate estimate of the location of the trait gene and to subsequently make use of linkage disequilibrium (association) for a more accurate localization. This general strategy is based on the assumption that disequilibrium extends over much shorter distances from a disease gene than linkage. The efficacy of this strategy has recently been challenged by the suggestion that, with a large number of SNPs available, it would be possible to localize disease genes with the disequilibrium mapping approach alone (e.g., by means of case-control studies). This assumption has not yet been empirically supported—no studies have used SNP LD strategy to map a disease gene. However, a number of theoretical investigations have explored efficiency, cost-effectiveness, and methods for this strategy.

One of the lines of such theoretical investigations involves the question of how many such markers exist on a genome-wide basis. This question can be reformulated in terms of the extent of LD in the genome—how rapidly does disequilibrium decay with the distance from the disease gene growing longer? An early estimate ***(128)*** was that, in large outbred populations, disequilibrium should be detectable within 100 kb of a disease locus. A later study that was based on a review of the published literature presented a more positive approach, suggesting that the distance is 300–500 kb ***(129)***. A recent computer simulation predicted an extremely short range of useful disequilibrium—3 kb ***(124)***. Such dramatic differences can be directly translated into associated costs—according to the first two estimates the required number of SNPs would be 30,000–100,000, and results from the third study suggest that 500,000 of SNPs would be needed.

One possible solution to the problem of not knowing the number of markers necessary to map a gene may be to select affected individuals from populations in which the extent of disequilibrium is greater than average. The literature contains some evidence suggesting that isolated populations are more advantageous for association mapping ***(130–131)***. However, this assumption has been challenged. Several examples have been published in which it appears that the extent of LD is either the same or only slightly higher in small, isolated populations as compared to large, outbred

populations ***(132–133)***. Although many factors may contribute to variability of the extent of LD in isolated populations (e.g., their histories, size, and current status), Ott ***(134)*** suggests that they appear to be of great inportance in association studies, especially when candidate genes are available.

3.4. Pitfalls of Current Gene-Mapping Methodologies

Whether conducting linkage or association studies, a number of factors make modern methodologies vulnerable to error.

3.4.1. Genotyping Errors

With the increased availability of markers (both SNPs and more conventional markers), the impact of genotyping error on the outcomes of different analytic methods is significant. Several authors have proposed methods to identify pedigrees and/or individuals with marker errors ***(135–140)***. Usually, an error is identified if it leads to a Mendelian inconsistency. However, Gordon et al. ***(141–142)*** have shown that under these conditions error detection rates are quite low, ranging between 25% and 30% (the detection rate is lowest when the two marker alleles have equal frequencies) when the true error rate is actually 3–4× higher. Clearly, better error detection is needed. On the other hand, a more economic approach may be to incorporate the allowance for errors into the analysis ***(143)*** as originally proposed by Keats et al. ***(144)***.

3.4.2. Map Misspecification

One of the major difficulties in the field at present is the degree of uncertainty in estimates of between-marker distances. Moreover, even when the distance is known, its estimate usually comes from a single source, and thus is usually a sex-averaged estimate. It is well-known that recombination rates differ in males and females ***(145)*** and they vary across different regions of the genome ***(146)***. Obviously, map misspecification can lead to lod score bias. Several studies have investigated the impact of map misspecification on linkage

and LD analyses ***(147–149)***. Most recently, Daw, Thompson, and Weijsman ***(150)*** have investigated map-misspecification bias (the discrepancy between the true lod score and the score estimated under incorrect map) in multipoint linkage analysis. These investigators reported that, in the presence of true linkage, any map misspecification causes a negative bias in lod scores, resulting in a loss of power to detect linkage. In the absence of linkage, map misspecification can cause positive or negative bias. Specifically, the utilization of the sex-average map results in a positive bias; so does overestimation of the distance. Underestimation of the distance results in a negative bias.

3.4.3. Allele-Frequency Misspecification

Genetic linkage and association analyses are highly sensitive to estimates of allele frequencies. Specifically, underestimation of allele frequencies can lead to false linkages, whereas overestimation can lead to reduced power ***(13,151–152)***. Allele frequency misspecification is especially dramatic for association and linkage studies performed on conglomerates of families (or case/control participants) of different ethnic origins. Several approaches have been proposed to address this problem. Specifically, when the analyzed sample is comprised of participants/families of different origins, one solution is to utilize published allele frequencies for each source population (i.e., populations whose representatives are present in the stratified sample), and, assuming that these published frequencies contain some error, "shrink" subpopulation estimates toward some common values ***(153–154)***. Lange's approach utilizes the empirical Bayes estimator for allele distributions and estimates the degree of allele frequency heterogeneity for a locus of interest, shrinking subpopulation-specific allele frequencies toward their pooled estimates as a function of the estimated subpopulation heterogeneity. Lockwood et al. ***(155)*** have extended Lange's approach by incorporating prior information about allele frequencies and interpopulation divergence into empirical Bayes analysis. This approach is implemented in the program ALLDIST.

3.5. Proportion of SNPs to Other Markers in Genome Screens

Chapman and Wijsman ***(156)*** have carried out a simulations analysis comparing diallelic markers and multiallelic markers in terms of sample sizes required for detection of LD (with the utilization of a single-marker locus in a case-control study for rare monophyletic diseases with Mendelian inheritance). These authors have demonstrated that multiallelic markers are more powerful for the detection of LD compared to diallelic markers, and that the ratio of the number of diallelic to the number of multiallelic markers, needed for equivalent power increases with mutation age and complexity in the mode of inheritance. In short, it takes many more diallelic markers than multiallelic markers to detect LD in a reasonable sample size.

3.6. Multiple Comparisons

The issue of multiple comparisons has long been the focus of attention of statisticians and epidemiologists, and there has been considerable debate about which corrections are appropriate. At one extreme is the position that the concern over multiple comparisons should not be an issue at all, and therefore does not require consideration ***(157)***. At the other extreme is the position that correction for multiple comparisons should be performed in all analyses, whether or not there were multiple comparisons in the reported analyses ***(158–159)***. And, of course, there are many positions between the two extremes.

This issue is important because of the rapidly approaching period when thousands of markers (SNPs and others) will be typed on the same sample. Thus, although it will be possible to significantly increase the number of markers typed, there are only a finite number of cases available for research. And as Lander ***(122)*** has pointed out, significantly increasing the number of comparisons requires a significant increase in the sample size studied. Although the magnitude of the increase in sample size is being debated ***(160)***, it is clear that more cases will be needed as the number of markers being genotyped increases.

3.7. Significance Level

Because a large number of markers are tested in genome-wide screens, Lander and Kruglyak ***(158)*** suggested that the point-wise (marker-specific) significance should be set at 0.000023 to correct for multiple comparisons. This *p*-value corresponds to a lod score of 3.6. The threshold results from the assumption that there is no disease gene when calculating the expected rate of false-positives. Thus, this approach has been challenged as being over-conservative, since it has been argued that researchers do not undertake linkage analyses unless there is strong evidence that genetic factors are important in the expression of the disorder being studied.

Several investigators have proposed methods for more accurately determining significance levels. Lucek and colleagues ***(161)*** introduced a methodology to investigate the inheritance of all markers jointly over the whole genome. The methodology unfolds as follows. For each parent in a set of affected sib-pair families, it is determined whether the parents pass on the same or a different allele to the two offspring. Under Mendelian inheritance, without influence of any disease loci, these two events have equal probabilities. However, under the assumption that the disease marker is located close to a disease locus underlying the trait in the two siblings, allele sharing is expected to occur with a probability higher than $^1/_2$. The goal of the methodology is to compare two sets of data, the observed allele-sharing data and randomly generated data that are known not to contain disease loci. The comparison is carried out by means of nonlinear discriminant analysis. The resulting weights are then used to construct a measure identifying the set of marker loci that jointly show deviations from random allele sharing.

Hoh and Ott ***(162)*** have developed a so-called scan statistic that is designed to combine information on multiple contiguous genetic markers used in a genome screen for susceptibility loci for various types of patients (e.g., sib-pairs, nuclear families, or case and control participants). This statistic can be calculated for a given length, and its significance can be assessed by a Monte Carlo permutation test. Multiple significance values are computed for statistics of given

lengths and then compared so that the smallest observed *p*-value is treated as the statistic of interest (for which the overall level can be determined). Illustrating this statistic, Hoh and Ott analyzed a 324-marker dataset obtained through a 10-cm-wide genome screen of autism affected-affected and affected-unaffected sibpairs. The initial single-marker screen did not result in any significant *p*-values. Thus, having used a set of statistics of a number of lengths (10, 20, 30, ..., 100 cm), the smallest observed value was obtained for the screen statistics calculated at 60 cm ($p = .015$). The overall significance level for the statistic of this length was .038. The scan statistic provides additional support for linkage above and beyond what is conveyed by the maximum lod score; it is especially useful when a susceptibility locus appears to be associated with multiple-marker loci (a situation frequently observed in genome-wide searchers). The scan statistic appears to be a useful tool in a number of designs, but its application may vary depending on the population investigated and/or the analysis utilized (e.g., it can be carried out with larger genomic regions in the context of linkage analyses and smaller genomic regions in the context of association analyses). The statistic appears to have more power as a method to detect linkage; however, once linkage has been established, it does not appear to be as useful for narrowing a candidate region.

4. Back to the Beginning

As stated at the outset, the development of appropriate methodologies for the genetic analysis of complex neuropsychiatric disorders has been, and will continue to be, challenging. This challenge arises because over the last decade the technology for gene mapping has developed exponentially, and the volumes of genomic data now available have required the development of new and innovative computing capabilities. There is considerable ongoing development in this area at the present time. Thus, it is very likely that the methods reviewed in this chapter will become outdated very quickly. A review of the major genetic journals—the *American Journal of Human Genetics*, *Neuropsychiatric Genetics*, and

Genetic Epidemiology, among others—reveals that there is at least one significant methodological advance each month. It is anticipated that there will now be an exponential increase in the analytic tools necessary to understand the genotypic and phenotypic data that will be generated.

References

1. Morton, N.E. (1955) Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7,** 277–318.
2 Elston, R.C. and Steward, J. (1971) A general model for the analysis of pedigree data. *Hum. Hered.* **21,** 523–542.
3. Ott, J. (1974) Estimation of the recombination fraction in human pedigrees: Efficient computation of the likelihood for human linkage studies. *Am. J. Hum. Genet.* **26,** 588–597.
4. Lathrop, G.M., Lalouel, J.M., Julier, C., and Ott, J. (1984) Strategies for multilocus linkage analysis in humans. *PNAS* **81,** 3443–3446.
5. Wijsman, E.M. and Amos, C. (1997) Genetic analysis of simulated oligogenic traits in nuclear and extended pedigrees: summary of GAW10 contributions. *Genet. Epidemiol.* **14,** 719–735.
6. Yuan, B., Vaske, D., Weber, J.L., Beck, J., and Sheffield, V.C. (1997) Improved set of short-tandem-repeat polymorphisms for screening the human genome. *Am. J. Hum. Genet.* **60,** 459–460.
7. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P., and Lander, E.S. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* **58,** 1347–1363.
8. Daw, E.W., Health, S.C., and Wijsman, E.M. (1999) Multipoint oligogenic analysis of age-at-onset data with applications to Alzheimer's disease pedigrees. *Am. J. Hum. Genet.* **64,** 839–851.
9. Heath, S.C. (1997) Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am. J. Hum. Genet.* **61,** 748–760.
10. Health, S.C., Snow, G.L., Thompson, E.A., Tseng, C., and Wijsman, E.M. (1997) MCMC segregation and linkage analysis. *Genet. Epidemiol.* **14,** 1011–1016.
11. Almasy, L. and Blangero, J. (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* **62,** 1198–1121.
12 Martinez, M., Khlat, M., Leboyer, M., and Clerget-Darpoux, F. (1989) Performance of linkage analysis under misclassification error when the genetic model is unknown. *Genet. Epidemiol.* **6,** 253–258.

13. Risch, N. and Giuffra, L. (1992) Model misspecification and multipoint linkage analysis. *Hum. Hered.* **42,** 77–92.
14. Schork, N.J., Boehnke, M., Terwilliger, J.D., and Ott, J. (1993) Two-trait-locus linkage analysis: a powerful strategy for mapping complex genetic traits. *Am. J. Hum. Genet.* **53,** 1127–1136.
15. Abreu, P.C., Greenberg, D.A., and Hodge, S.E. (1999) Direct power comparisons between simple lod scores and NPL scores for linkage analysis in complex diseases. *Am. J. Hum. Genet.* **65,** 847–857.
16. Göring, H.H.H. and Terwilliger, J.D. (2000) Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. *Am. J. Hum. Genet.* **66,** 1298–1309.
17. Göring, H.H.H. and Terwilliger, J.D. (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.
18. Whittemore, A.S. (1996) Genome scanning for linkage: an overview. *Am. J. Hum. Genet.* **59,** 704–716.
19. Fimmers, R., Seuchter, S.A., Neugebauer, M., Knapp, M., and Baur, M.P. (1989) Identity-by-descent analysis using all genotype solutions, in *Multipoint mapping and linkage based on affected pedigree members: Genetic Analysis Workshop 6,* (Elston, R.C., Spence, M.A., Hodge, S.E., and MacCluer, J. W., eds.), Alan R. Liss, New York, NY, pp. 123–128.
20. Kruglyak, L. and Lander, E.S. (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.* **57,** 439–454.
21. Whittemore, A.S. and Halpern, J. (1994) A class of tests for linkage using affected pedigree members. *Biometrics* **50,** 117–127.
22. Irwin, M., Cox, N., and Kong, A. (1994) Sequential imputation for multilocus linkage analysis. *Proc. Natl. Acad. Sci. USA* **91,** 11,684–11,688.
23. O'Connell, J.R. and Weeks, D.E. (1995) The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. *Nat. Genet.* **11,** 402–408.
24. Schwab, S.G., Albus, M., Hallmayer, J., Honig, S., Borrmann, M., Lichtermann, D., 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.

25. Friddle, C., Koskela, R., Ranade, K., Hebert, J., Cargill, M., Clark, C.D., et al. (2000) Full-genome scan for linkage in 50 families segregating the bipolar affective disease phenotype. *Am. J. Hum. Genet.* **66,** 205–215.
26. Gudbjartsoon, D.F., Jonasson, K., Frigge, M.L., and Kong, C.A. (2000) Allegro, a new computer program for multipoint linkage analysis. *Nat. Genet.* **25,** 12–13.
27. Hyer, R.N., Julier, C., Buckley, J.D., Trucco, M., Rotter, J., Spielman, R., et al. (1991) High-resolution linkage mapping for susceptibility genes in human polygenic disease: insulin-dependent diabetes mellitus and chromosome 11q. *Am. J. Hum. Genet.* **48,** 243–257.
28. Knapp, M., Seuchter, S.A., and Baur, M.P. (1994) Two-locus disease models with two marker loci: the power of affected-sib-pair tests. *Am. J. Hum. Genet.* **55,** 1030–1041.
29. Lander, E.S. and Green, P. (1987) Construction of multilocus genetic maps in humans. *Proc. Natl. Acad. Sci. USA* **84,** 2363–2367.
30. Guo, S.W. and Thompson, E.A. (1992) A Monte Carlo method for combined segregation and linkage analysis. *Am. J. Hum. Genet.* **51,** 1111–1126.
31. Borewinkle, E., Chakraborty, R., and Sing, C. (1986) The use of measured genotype information in the analysis of quantitative phenotypes in man. *Ann. Hum. Genet.* **50,** 181–194.
32. Boerwinkle, E.R.C. and Sing, C.F. (1986) Bias of the contribution of single-locus effects to the variance of a quantitative trait. *Am. J. Hum. Genet.* **39,** 663–676.
33. Wijsman, E.M. and Nur, N. (2001) On estimating the proportion of variance in a phenotypic trait attributable to a measured locus. *Hum. Hered.* **51,** 145–149.
34. Haseman, J.K. and Elston, R.C. (1972) The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2,** 3–19.
35. Lange, K., Westlake, J., and Spence, M.A. (1976) Extensions to pedigree analysis. III. Variance components by the scoring method. *Ann. Hum. Genet.* **39,** 485–491.
36. Hopper, J.L. and Matthews, J.D. (1982) Extensions to multivariate normal models for pedigree analysis. *Ann. Hum. Genet.* **46,** 373–383.
37. Amos, C.I. (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* **54,** 535–543.
38. Goldgar, D.E. (1990) Multipoint analysis of human quantitative genetic variation. *Am. J. Hum. Genet.* **47,** 957–967.

39. Schork, N.J. (1993) Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *Am. J. Hum. Genet.* **53,** 1306–1319.
40. Pratt, S.C., Daly, M.J., and Kruglyak, L. (2000) Exact multipoint quantitative-trait linkage analysis in pedigrees by variance components. *Am. J. Hum. Genet.* **66,** 1153–1157.
41. Amos, C.I., Zhu, D.K., and Boerwinkle, E. (1996) Assessing genetic linkage and association with robust components of variance approaches. *Ann. Hum. Genet.* **60,** 143–160.
42. Amos, C.I., Krushkal, J., Thiel, T.J., Young, A., Zhu, D.K., Boerwinkle, E., et al. (1997) Comparison of model-free linkage mapping strategies for the study of a complex trait. *Genet. Epidemiol.* **14,** 743–748.
43. Pugh, E.W., Jaquish, C.E., Sorant, A.J.M., Doetsch, J.P., Bailey-Wilson, J.E., and Wilson, A.F. (1997) Comparison of sib-pair and variance components methods for genomic screening. *Genet. Epidemiol.* **14,** 867–872.
44. Williams, J.T., Duggirala, R., and Blangero, J. (1997) Statistical properties of a variance components method for quantitative trait linkage analysis in nuclear families and extended pedigrees. *Genet. Epidemiol.* **14,** 1065–1070.
45. Boehnke, M. (1994) Limits of resolution of genetic linkage studies: implications for the positional cloning of human genetic disease. *Am. J. Hum. Genet.* **55,** 379–390.
46. Hästbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A., and Lander, E. (1992) Linkage disequilibrium mapping in isolated founder populations: diastrophis dysphasia in Finland. *Nat. Genet.* **2,** 204–211.
47. Hästbacka, J.K., de la Chapelle, A., Mahanti, M. M., Clines, G., Reeve-Daly, M.P., Daly, M., et al. (1994) The diastrophic dysphasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell* **78,** 1073–1087.
48. Kerem, B., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., et al. (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science* **245,** 1073–1080.
49. Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten W.J., Baldwin, J., Linton, L., et al. (2000a) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* **407,** 513–516.

50. Mullikin, J.C., Hunt, S.E., Cole, C.G., Mortimore, B.J., Rice, C.M., Burton, J., et al. (2000) An SNP map of human chromosome 22. *Nature* **407,** 516–520.
51. Devlin, B. and Risch, N. (1995) A comparison of linkage disequilibrium measures for fine scale mapping. *Genomics* **29,** 311–322.
52. Feder, J.N., Gnirke, A., Thomas, W., Tsuchihashi, Z., Ruddy, D.A., Basava, A., et al. (1996) A novel MHC class I-like gene is mutation in patients with hereditary haemochromatosis. *Nat. Genet.* **13,** 399–408.
53. Kaplan, N., Hill, W.G., and Weir, B.S. (1995) Likelihood methods for locating disease genes in nonequilibrium populations. *Am. J. Hum. Genet.* **56,** 18–32.
54. Devlin, B., Risch, N., and Roeder, K. (1996) Disequilibrium mapping: composite likelihood for pairwise disequilibrium. *Genomics* **36,** 1–16.
55. Lam, J.C., Roeder, K., and Devlin, B. (2000) Haplotype fine mapping by evolutionary trees. *Am. J. Hum. Genet.* **66,** 659–673.
56. Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., et al. (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245,** 1066–1073.
57. Rommens, J.M., Iannuzzi, M.C., Kerem, B., Drumm, M.L., Melmer, G., Dean, M., et al. (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245,** 1059–1065.
58. Abecasis, G.R., Nogushi, E., Heinzmann, A., Traherne, J.A., Bhattacharyya, S., Leaves, N.I., et al. (2001) Extent and distortion of linkage disequilibrium in three genomic regions. *Am. J. Hum. Genet.* **68,** 191–197.
59. Cardon, L.R. and Bell, J.I. (2001) Association study designs for complex diseases. *Nature Reviews, Genetics* **2,** 91–99.
60. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics* **55,** 997–1004.
61. Cambien, F., Poirier, O., Lecerf, L., Evans, A., Cambou, J.P., Arveiler, D., et al. (1992) Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. *Nature* **359,** 641–644.
62. Jeron, A., Hengstenberg, C., Engel, S., Lowel, H., Riegger, G.A., Schunkert, H., et al. (2001) The D-allele of the ACE polymorphism is related to increased QT dispersion in 609 patients after myocardial infarction. *Eur. Heart J.* **22,** 663–668.

63. Keavney, B., McKenzie, C., Parish, S., Palmer, A., Clark, S., Youngman, L., et al. (2000) Large-scale test of hypothesized associations between the angiotensin-converting-enzyme insertion/deletion polymorphism and myocardial infarction in about 5000 cases and 6000 controls. International Studies of Infarct Survival (ISIS) Collaborators. *Lancet* **355,** 434–442.
64. Altshuler, D., Hirschhorn, J.N., Klannemark, M., Lindgren, C.M., Vohl, M.C., Nemesh, J., et al. (2000b) The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26,** 76–80.
65. Barcellos, L.F., Klitz, W., Field, L.L., Tobias, R., Bowcock, A.M., Wilson, R., et al. (1997) Associating mapping of disease loci by use of a pooled DNA genomic screen. *Am. J. Hum. Genet.* **61,** 734–747.
66. Daniels, J., Holmans, P., Williams, N., Turic, D., McGuffin, P., Plomin, R., et al. (1998) A simple method for analyzing microsatellite allele image patterns generated from DNA pools and its application to allelic association studies. *Am. J. Hum. Genet.* **62,** 1189–1197.
67. Shaw, S.H., Carrasquillo, M.M., Kashuk, C., Puffenberger, E.G., and Chakravarti, A. (1998) Allele frequency distributions in poled DNA samples: applications to mapping complex disease gene. *Genome Res.* **8,** 111–123.
68. Emahazion, T., Feuk, L., Jobs, M., Sawyer, S.L., Fredman, D., St. Clair, D., et al. (2001) SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. *Trends Genet.* **17,** 401–407.
69. Jorde, L.B., Watkins, W.S., Carlson, M., Groden, J., Albertson, H., Thliveris, A., et al. (1994) Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. *Am. J. Hum. Genet.* **54,** 884–898.
70. Hill, W.G. and Weis, B.S. (1994) Maximum-likelihood estimation of gene location by linkage disequilibrium. *Am. J. Hum. Genet.* **54,** 705–714.
71. Rubinstein, P., Walker, M., Carpenter, C., Carrier, C., Krassner, J., Falk, C., et al. (1981) Genetics of HLA disease association: the use of the haplotype relative risk (HRR) and the "Haplo-Delta" (Dh) estimates in juvenile diabetes from three racial groups. *Hum. Immunol.* **3,** 384.
72. Field, L.L., Fothrgill-Payne, C., Bertrams, J., and Baur, M.P. (1986) HLD-DR effects in a large German IDDM data set. *Genet. Epidemiol.* (suppl.) **1**, 323–328.

73. Falk, C.T. and 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.
74. Khoury, M.J. (1994) Case-parental control method in the search for disease-susceptibility genes. *Am. J. Hum. Genet.* **55,** 414–415.
75. Schaid, D.J. and Sommer, S.S. (1994) Comparison of statistics for candidate-gene association studies using cases and parents. *Am. J. Hum. Genet.* **55,** 402–409.
76. Thomson, G. (1995) Mapping disease genes: family-based association studies. *Am. J. Hum. Genet.* **57,** 487–498.
77. Knapp, M., Seuchter, S.A., and Baur, M.P. (1993) The haplotype-relative risk (HRR) method for analysis of association in nuclear families. *Am. J. Hum. Genet.* **52,** 1085–1093.
78. Ott, J. (1989) Statistical properties of the haplotype relative risk. *Genet. Epidemiol.* **6,** 127–130.
79. Terwilliger, J.D. and Ott, J. (1992) A haplotype based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* **42,** 337–346.
80. Schaid, D.J. and Sommer, S.S. (1993) Genotype relative risks: methods for design and analysis of candidate-gene association studies. *Am. J. Hum. Genet.* **53,** 1114–1126.
81. Spielman, R.S. and Ewens, W.J. (1993) Transmission/disequilibrium test (TDT) for linkage and linkage disequilibrium between disease and marker. *Am. J. Hum. Genet.* (suppl.) **53,** 863.
82. Spielman, R.S., McGinnis, R.E., and Ewens, W.J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52,** 506–516.
83. Tiret, L., Nicaud, V., Ehnholm, C., Havekes, L., Menzel, H.J., Ducimetiere, P., et al. (1993) Inference of the strength of genotype-disease association from studies comparing off-spring with and without parental history of disease. *Ann. Hum. Genet.* **57,** 141–149.
84. Hodge, S.E. (1993) Linkage analysis versus association analysis: distinguishing between two models that explain disease-marker associations. *Am. J. Hum. Genet.* **53,** 367–384.
85. Hodge, S.E. (1994) Reply to Suarez and Hampe and Spielman et al.: cosegregation, association, and linkage. *Am. J. Hum. Genet.* **54,** 560–563.
86. Suarez, B.K. and Hampe, C.L. (1994) Linkage and association. *Am. J. Hum. Genet.* **54,** 554–559.

87. Klitz, W., Thomson, G., Borot, N., and Cambon-Thomsen, A. (1992) Evolutionary and population perspective of the human HLA complex, in *Evolutionary Biology* (vol. 26), (Hecht, M. K., ed.), Plenum, New York, NY, pp. 35–72.
88. Field, L.L. (1989) Genes predisposing to IDDM in multiplex families. *Genet. Epidemiol.* **6,** 101–106.
89. Thomson, G., Robinson, W.P., Kuhner, M.K., Joe, S., and Klitz, W. (1989) HLA and insulin gene associations with IDDM. *Genet. Epidemiol.* **6,** 155–160.
90. Spielman, R.S., Baur, M.P., and Clerget-Darpoux, F. (1989) Genetic analysis of IDDM: summary of GAW 5 IDDM results. *Genet. Epidemiol.* **6,** 43–58.
91. Bennet, S.T., Lucassen, A.M., Gough, S. C., Powell, E.E., Undlien, D.E., Pritchard, L.E., et al. (1995) Susceptibility to human type diabetes at IDDM2 is determined by tandem repeat variation at the insulin gene minisatellite locus. *Nat. Genet.* **9,** 284–292.
92. Merriman, T., Twells, R., Merriman, M., Eaves, I., Cox, R., Cucca, F., et al. (1997) Evidence by allelic association-dependent methods for a type I diabetes polygene (IDDM6) on chromosome 18q21. *Hum. Mol. Genet.* **6,** 1003–1010.
93. Boehnke, M. and Langefeld, C.D. (1998) Genetic association mapping based on discordant sib pairs: the discordant-alleles test. *Am. J. Hum. Genet.* **61,** 319–333.
94. Curtis, D. (1997) Use of siblings as controls in case-control association studies. *Ann. Hum. Genet.* **61**, 319–333.
95. Martin, E.R., Kaplan, N.I., and Weir, B.S. (1997) Tests for linkage and association in nuclear families. *Am. J. Hum. Genet.* **61,** 439–448.
96. Sham, P.C. and Curtis, D. (1995) An extended transmission/disequilibrium test (TDT) for multiallelic marker loci. *Ann. Hum. Genet.* **59,** 323–336.
97. Spielman, R.S. and Ewens, W.J. (1996) The TDT and other family-based tests for linkage disequilibirum and association (Editorial). *Am. J. Hum. Genet.* **59,** 983–989.
98. Spielman, R.S. and Ewens, W.J. (1998) A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am. J. Hum. Genet.* **62,** 450–458.
99. Seltman, H., Roeder, K., and Devlin, B. (2001) Transmission/Disequilibrium Test meets measured haplotype analysis: family-based association analysis guided by evolution of haplotypes. *Am. J. Hum. Genet.* **68,** 1250–1263.

100. Templeton, A.R. (1995) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping of DNA sequencing. V. Analysis of case/control sampling designs: Alzheimer's disease and the apoprotein E locus. *Genetics* **140,** 403–409.
101. Gordon, D., Heath, S.C., Liu, X., and Ott, J. (2001) A transmission/disequilibrium test that allows for genotyping errors in the analysis of single-nucleotide polymorphism data. *Am. J. Hum. Genet.* **69,** 371–780.
102. Bacanu, S.-A., Devlin, B., and Roeder, K. (2000) The power of genomic control. *Am. J. Hum. Genet.* **66,** 1933–1944.
103. Clayton, D.G. and Jones, H. (1999) Transmission/disequilibrium tests for extended marker haplotypes. *Am. J. Hum. Genet.* **65,** 1161–1169.
104. Clayton, D.G. (1999) A generalization of the transmission/disequilibrium tests for uncertain-haplotype transmission. *Am. J. Hum. Genet.* **65,** 1170–1177.
105. Lazzeroni, L.C. and Lange, K. (1998) A conditional inference framework for extending the transmission/disequilibrium test. *Hum. Hered.* **48,** 67–81.
106. Merriman, T.R., Eaves, I.A., Twells, R.C., Merriman, M.E., Danoy, P.A., Muxworthy, C.E., et al. (1998) Transmission of haplotypes of miscrosatellite markers rather than single marker alleles in the mapping of a putative type 1 diabetes susceptibility gene (IDDM6). *Hum. Mol. Genet.* **7,** 517–524.
107. Rabinowitz, D. and Laird, N. (2000) A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum. Hered.* **50,** 211–223.
108. Zhao, H., Zhang, S., Merikangas, K.R., Trixter, M., Wildenauer, D.B., Sun, F., et al. (2000) Transmission/disequilibrium tests using multiple tightly linked markers. *Am. J. Hum. Genet.* **67,** 936–946.
109. Jin, K., Speed, T., Klitz, W., and Thomson, G. (1994) A statistical test of segregation distortion. *Biometrics* **50,** 1189–1198.
110. Maxwell, A.E. (1970) Comparing the classification of subjects by two independent judges. *Br. J. Psychiatry* **116,** 651–655.
111. Rothman, K.J. (1986) *Modern Epidemiology*, Little & Brown, Boston, MA.
112. Trembath, R.C., Clough, R.L., Rosbotham, J.L., Jones, A.B., Camp, R.D.R., Frodsham, A., et al. (1997) Identification of a major

susceptibility locus on chromosome 6p and evidence for further disease loci revealed by a two stage genome-wide search in psoriasis. *Hum. Mol. Genet.* **6,** 813–820.

113. George, V.T. and Elston, R.C. (1987) Testing the association between polymorphic markers and quantitative traits in pedigrees. *Genet. Epidemiol.* **4,** 193–201.
114. Allison, D.D. (1997) Transmission-disequilibrium tests for quantitative traits. *Am. J. Hum. Genet.* **60,** 676–690.
115. Rabinowitz, D. (1997) A transmission disequilibrium test for quantitative trait loci. *Hum. Hered.* **47,** 342–350.
116. Fulker, D.W., Cherny, S.S., Sham, P.C., and Hewitt, J.K. (1999) Combined linkage and association sib-pair analysis for quantitative traits. *Am. J. Hum. Genet.* **64,** 259–267.
117. Cardon, L.R. (2000) A sib-pair regression model of linkage disequilibrium for quantitative traits. *Hum. Hered.* **50,** 350–358.
118. Abecasis, G.R., Cardon, L.R., and Cookson, W.O.C. (2000) A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.* **66,** 279–292.
119. Brookes, A.J. (1999) The essence of SNPs. *Gene* **234,** 177–186.
120. Kwok, P.Y., Deng, Q., Zakeri, H., Taylor, S.L., and Nickerson, D.A. (1996) Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. *Genomics* **31,** 123–126.
121. Wang, D.G., Fan, J.B., Siao, C.J., Berno, A., Young, P., Sapolsky, R., et al. (1988) Related articles, nucleotide large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280(5366),** 1077–1082.
122. Lander, E.S. (1996) The new genomics: global views of biology. *Science* **274,** 536–539.
123. Jorde, L.B. (2000) Linkage disequilibrium and the search for complex disease genes. *Genome Res.* **10,** 1435–1444.
124. Kruglyak, L. (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat. Genet.* **22,** 139–144.
125. Risch, N. and Merikangas, K. (1996) Related Articles. The future of genetic studies of complex human diseases. *Science* **273(5281),** 1516–1517.
126. Terwilliger, J.D. and Weiss, K.M. (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? *Curr. Opin. Biotechnol.* **8,** 578–594.

127. Weiss, K.M. and Tewilliger, J.D. (2000) How many diseases does it take to map a gene with SNPs? *Nat. Genet.* **26,** 151–157.
128. Bodmer, W.F. (1986) Human genetics: the molecular challenge. *Cold Spring Harbor Symp. Quant. Biol.* **51,** 1–13.
129. Collins, A., Lonjou, C., and Morton, N.E. (1999) Genetic epidemiology of single-nucleotide polymorphisms. *PNAS* **95,** 15,173–15,177.
130. Laan, M. and Pääbo, S. (1997) Demographic history and linkage disequilibrium in human populations. *Nat. Genet.* **17,** 435–438.
131. Peltonen, L. (2000) Positional cloning of disease genes: advantages of genetic isolates. *Hum. Hered.* **50,** 66–75.
132. Jorde, L.B., Watkins, W.S., Kere, J., Nyman, D., and Eriksson, A.W. (2000) Gene mapping in isolated populations: new roles for old friends? *Hum. Hered.* **50,** 57–65.
133. Lonjou, C., Collins, A., and Morton, N.E. (1999) Allelic association between marker loci. *PNAS* **96,** 1621–1626.
134. Ott, J. (2000) Predicting the range of linkage disequilibrium. *PNAS* **97,** 2–3.
135. Brzustowicz, L.M., Mérette, C., Xie, X., Townsend, L., Gilliam, T.C., and Ott, J. (1993) Molecular and statistical approaches to the detection and correction of errors in genotype databases. *Am. J. Hum. Genet.* **53,** 1137–1145.
136. Ehm, M.G., Kimmel, M., and Cottingham, R.W., Jr. (1996) Error detection for genetic data, using likelihood methods. *Am. J. Hum. Genet.* **58,** 225–234.
137. Lathrop, G.M., Hooper, A.B., Huntsman, J.W., and Ward, R.H. (1983) Evaluating pedigree data. 1. The estimation of pedigree error in the presence of marker mistyping. *Am. J. Hum. Genet.* **35,** 241–262.
138. Lincoln, S.E. and Lander, E.S. (1992) Systematic detection of errors in genetic linkage data. *Genomics* **14,** 604–610.
139. Sasse, G., Müller, H., Chakraborty, R., and Ott, J. (1994) Estimating the frequency of nonpaternity in Switzerland. *Hum. Hered.* **44,** 337–343.
140. Stringham, H.M. and Boehnke, M. (1996) Identifying marker typing incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **59,** 946–950.
141. Gordon, D., Heath, S., and Ott, J. (1999) True pedigree errors more frequent than apparent errors for single nucleotide polymorphisms. *Hum. Hered.* **49,** 65–70.

142. Gordon, D., Leal, S.M., Heath, S.C., and Ott, J. (2000) An analytic solution to single nucleotide polymorphism error-detection rates in nuclear families: implications for study design. *Pacific Symposium on Biocomputing* **1,** 663–674.
143. Ott, J. (2001) Major strengths and weaknesses of the lod score method. *Adv. Genet.* **42,** 125–132.
144. Keats, B.J., Sherman, S.L., and Ott, J. (1990) Human gene mapping 10.5—report of the Committee on Linkage and Gene Order. *Cytogenet. Cell Genet.* **55,** 387–394.
145. Broman, K.W., Murray, J.C., Sheffield, V.C., White, R.L., and Weber, R.L. (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63,** 861–869.
146. Li, W., Fann, C., and Ott, J. (1998) Low-order polynomial trends of female-to-male map distance ratios along human chromosomes. *Hum. Hered.* **48,** 266–270.
147. Cleves, M.A. and Elston, R.C. (1997) Alternative test for linkage between two loci. *Genet. Epidemiol.* **14,** 117–131.
148. Halpern, J. and Whittemore, A.S. (1999) Multipoint linkage analysis. A cautionary note. *Hum. Hered.* **49,** 194–196.
149. Ott, J., Lehner, T., and Sandkuyl, L.A. (1988) Fallacies of the technique of "ahplotyping" two apparently linked loci in multipoint linkage analysis. *Am. J. Hum. Genet.* **43,** A154.
150. Daw, E.W., Thompson, E.A., and Wijsman, E.M. (2000) Bias in multipoint linkage analysis arising from map misspecification. *Genet. Epidemiol.* **19,** 336–380.
151. Friemer, N., Sandkuijl, L., and Blower, S. (1993) Incorrect specification of marker allele frequencies: effects on linkage analysis. *Am. J. Hum. Genet.* **52,** 1102–1110.
152. Ott, J. (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. *Am. J. Hum. Genet.* **51,** 293–290.
153. Devlin, B., Risch, N., and Roeder, K. (1991) Estimation of allele frequencies for VNTR loci. *Am. J. Hum. Genet.* **48,** 662–676.
154. Lange, K. (1995) Applications of the Dirichlet distribution to forensic match probabilities. *Genetica* **96,** 107–117.
155. Lockwood, J.R., Roder, K., and Devlin, B. (2001) A Bayesian hierarchical model for allele frequencies. *Genet. Epidemiol.* **20,** 17–33.
156. Chapman, N.H. and Wijsman, E.M. (1998) Genome screen using linkage disequilibrium tests: optimal marker characteristics and feasibility. *Am. J. Hum. Genet.* **63,** 1872–1885.

157. Rothman, K. (1990) No adjustment are needed for multiple comparisons. *Epidemiology* **1,** 43–46.
158. Lander, E. and Kruglyak, L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* **11,** 241–247.
159. Witte, J.S., Elston, R.C., and Schork, N.S. (1996) Genetic dissection of complex traits (correspondence). *Nat. Genet.* **12,** 355–356.
160. Witte, J.S., Elston, R.C., and Cardon, L.R. (2000) On the relative sample size required for multiple comparisons. *Stat. Med.* **19,** 369–372.
161. Lucek, P., Hanke, J., Reich, J., Solla, S., and Ott, J. (1998) Multilocus nonparametric linkage analysis of complex trait loci with neural network. *Hum. Hered.* **48,** 275–284.
162. Hoh, J. and Ott, J. (2000) Scan statistics to scan markers for susceptibility genes. *PNAS* **17,** 9615–9617.