Human Heredity

Hum Hered 2011;72:228–236
DOI: 10.1159/000334381

Published online: December 23, 2011

Linkage Analysis in the Next-Generation Sequencing Era

Joan E. Bailey-Wilson Alexander F. Wilson

Inherited Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, Md., USA

Key Words
Linkage · Genetics · DNA sequence · Whole-genome sequence · Whole-exome sequence

Abstract
Linkage analysis was developed to detect excess co-segregation of the putative alleles underlying a phenotype with the alleles at a marker locus in family data. Many different variations of this analysis and corresponding study design have been developed to detect this co-segregation. Linkage studies have been shown to have high power to detect loci that have alleles (or variants) with a large effect size, i.e. alleles that make large contributions to the risk of a disease or to the variation of a quantitative trait. However, alleles with a large effect size tend to be rare in the population. In contrast, association studies are designed to have high power to detect common alleles which tend to have a small effect size for most diseases or traits. Although genome-wide association studies have been successful in detecting many new loci with common alleles of small effect for many complex traits, these common variants often do not explain a large proportion of disease risk or variation of the trait. In the past, linkage studies were successful in detecting regions of the genome that were likely to harbor rare variants with large effect for many simple Mendelian diseases and for many complex traits. However, identifying the actual sequence variant(s) responsible for these linkage signals was challenging because of difficulties in sequencing the large regions implicated by each linkage peak. Current 'next-generation' DNA sequencing techniques have made it economically feasible to sequence all exons or the whole genomes of a reasonably large number of individuals. Studies have shown that rare variants are quite common in the general population, and it is now possible to combine these new DNA sequencing methods with linkage studies to identify rare causal variants with a large effect size. A brief review of linkage methods is presented here with examples of their relevance and usefulness for the interpretation of whole-exome and whole-genome sequence data.

Introduction

The identification of genes that contribute to the risk of a disease or the variation of a quantitative trait has long been one of the goals of human and medical genetics. Historically, both linkage and association methods have been used to narrow the location of such genes to small regions of the genome with the hope that the gene, and eventually the causal variant, may be identified and char-

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

Accessible online at:
www.karger.com/hhe

Joan E. Bailey-Wilson, PhD
NIH/NHGRI
333 Cassell Drive, Suite 1200
Baltimore, MD 21224 (USA)
Tel. +1 443 740 2921, E-Mail jebw@mail.nih.gov

acterized. Recently, the focus in genetic epidemiology has been on genome-wide association studies (GWAS) because of the availability and affordability of dense sets of polymorphic genetic markers that can be genotyped on large numbers of individuals. GWAS examine common alleles for association with disease or trait phenotypes and have identified many regions of the genome that show such associations for a large number of important traits (http://www.genome.gov/26525384). It is important to note that these associations do not identify specific causal variants per se, but rather relatively short candidate regions that include the associated allele/variant and all of the variants that are highly correlated with it in a linkage disequilibrium block. Both theoretical and applied studies have shown that most of these associated loci have small effects on the disease or trait in question. Linkage studies, on the other hand, had been the study design of choice for many years, mostly because they were feasible with much less dense sets of genetic markers, and because linkage methods had the power to detect co-segregation over a much larger range. Linkage methods are particularly powerful for the detection of variants with a large effect size, which often are rare in the population. Power to detect such loci using linkage methods can be enhanced by ascertaining families with aggregation of the trait of interest ('loaded families'). Like tests of association, linkage methods are also able to identify candidate regions, but the regions are much larger, sometimes spanning 40 Mb. Interest in these methods is undergoing a renaissance due to the availability of 'next-generation' DNA sequencing and its promise to allow identification of the rare variants underlying linkage signals.

Linkage studies have led to the identification of genes that cause or substantially increase the risk of many diseases and birth defects. For example, linkage analyses led to the identification of the genes that cause many Mendelian disorders such as cystic fibrosis *(CFTR)* [1, 2] and Huntington disease *(HTT)* [3, 4]. Linkage studies of families selected because of very strong aggregation of specific complex diseases have also led to the identification of rare, high-penetrance risk alleles in certain genes that cause large increases in susceptibility to complex diseases, for example the *BRCA1* and *BRCA2* genes and breast cancer. This study design has also led to the identification of genes with rare risk alleles that cause moderate increases in the risk for complex diseases such as the *NOD2* gene and inflammatory bowel disease [5–8]. Linkage methods have been successfully applied to quantitative traits as well. In a series of papers spanning almost 30 years, the specific activity of dopamine-beta hydroxylase activity, an enzyme that catalyzes the conversion of dopamine to norepinephrine, was localized to the chromosomal 9q34 region, and specific variants that were responsible, at least in part, for the variation of the trait activity have been identified [9–14]. It is important to note that in this case the phenotype, the specific activity of the enzyme, is functionally closely related to the underlying structural locus. Other linkage studies, where the phenotype is not closely related to the underlying genotype, have not been as straightforward, or as successful.

However, for other Mendelian and complex disorders, linkage signals have been detected in family studies, but the causal genes and risk alleles have not yet been discovered. It is thought that this may be due to a variety of reasons, including (1) the previous high cost of DNA sequencing that precluded sequencing all genes under broad linkage peaks; (2) sequencing studies that included only exons of genes under linkage peaks, ignoring changes in regulatory regions; (3) clinical and genetic heterogeneity, and (4) false-positive evidence of linkage. While some significant linkage signals reported in the literature are almost certainly false-positive results [15], those that have been confirmed in independent sets of families are more likely to be true [16]. Advances in our understanding of the complexities of the human genome have made it clear that sequencing only exons will not detect all DNA variants that contribute to disease risk or to variation in quantitative traits. New, cost-effective DNA sequencing methods have recently made it possible and economically feasible to combine linkage information with whole-exome or whole-genome sequence data to identify the causal variants that contribute to the linkage signals. For example, Sobreira et al. [17] combined linkage information for the Mendelian disease metachondromatosis (OMIM 156250) with whole-genome DNA sequence in a single proband to identify an 11-bp deletion in exon 4 of *PTPN11*, that alters the reading frame, resulting in premature translation termination, and that co-segregates with the phenotype. They confirmed this result by finding a different nonsense mutation in exon 4 of this gene that segregates with disease in another family. Bowden et al. [18] have used a similar strategy to identify a gene *(ADIPOQ)* contributing to variation of a quantitative trait, serum adiponectin level, by (1) identifying families responsible for a linkage signal to plasma adiponectin levels [19]; (2) performing whole-exome sequencing of 3 individuals in the two most strongly linked families with attention targeted to the region of the linkage peak, and (3) performing sequencing of the candidate gene *ADIPOQ* in additional samples from these two families

and from unrelated individuals, showing that the risk allele is rare in the population (less than 2%) but that it accounts for the linkage signal in the two most strongly linked families. Several groups are now using similar approaches involving linkage in conjunction with whole exome sequencing, whole-genome sequencing or targeted next-generation sequencing, leading to a resurgence of interest in linkage analysis methods. Therefore, we briefly review classic linkage analysis methods here.

Brief Overview and History of Linkage Analysis

Linkage analysis refers to a group of statistical methods that are used to map a gene to the region of the chromosome in which it is located. These methods take advantage of the fact that many more genes exist than chromosomes, and thus many genes are transmitted together from parents to offspring during meiosis. Linkage is the tendency of two or more genetic loci to be transmitted together during meiosis because they are physically close together on a chromosome. As such, linkage represents a violation of Mendel's law of independent assortment.

The concept that chromosomal segregation could explain the physical basis of Mendelian inheritance was first put forward by Sutton [20, 21] in the early 1900's. Most early linkage studies were performed in plants and experimental animals. Correns [22] reported the first linkage analysis in plants, with Bateson and Punnett [23] observing the presence of recombinations between syntenic loci (i.e. genetic loci on the same chromosome). During the first meiotic prophase, pairing of the duplicated homologous chromosomes (synapsis) occurs. At this stage, a physical exchange of chromosomal material occurs between homologues. These exchanges are called chiasmata and lead to a 'crossover' of the DNA between the two homologues. These chiasmata occur frequently, but it is well known that the presence of one chiasma at a specific chromosomal location will decrease the chances that other chiasmata will form nearby (chiasma interference) [24]. Thus, the probability that crossovers will occur between two syntenic loci is dependent on the distance between the loci [25, 26], but the probability of double crossovers is disproportionately low between very close loci due to chiasma interference [24]. Phase is a term that refers to which alleles at two syntenic loci are physically located together on the same homologue. Consider two syntenic loci, *A* and *B*, each with two alleles, A_1 and A_2, and B_1 and B_2, respectively. A person with genotypes A_1/A_2 and B_1/B_2 is a double heterozygote. There are two possible phases: (1) the A_1 and B_1 alleles reside together on one member of the chromosome pair and the A_2 and B_2 alleles on the other, or (2) the A_1 and B_2 alleles reside together on one homologue and the A_2 and B_1 alleles on the other. Only odd numbers of crossovers between the two loci can be detected by examining the genotypes of the parents and offspring because an even number will result in the original alleles at the two loci being transmitted together, maintaining the parental phase with respect to these two loci. When an odd number of crossovers occurs between two syntenic loci, then the alleles at these loci are recombined, i.e. transmitted to the offspring in a new combination or new phase. Two loci that are far apart on the chromosome (syntenic loci) have a high probability of recombination in any meiosis, such that they assort independently to offspring. Syntenic loci that are very far apart experience recombination about 50% of the time, and thus appear to be assorting independently, just as loci on different chromosomes do.

The recombination fraction measures the proportion of recombinations observed between two loci in a group of offspring. Linkage occurs when two loci are physically close enough so that alleles on the same homologous chromosome tend to be transmitted together, and no or very few recombinations are observed among the offspring. The recombination fraction, often represented as θ, is estimated by counting the number of offspring that show recombination for a given pair of loci, divided by the total number of offspring (the number of recombinants plus the number of non-recombinants). If two loci are physically next to one another, there is very little chance that a crossover will occur between them and the recombination fraction is close to zero. When the loci are on separate chromosomes or are far apart on the same chromosome, the recombination fraction is 1/2, with values between these two extremes indicating some degree of linkage.

Linkage analysis in humans is more difficult than in experimental organisms because of limitations in family size, the inability to do test crosses, the long generation time and lack of knowledge of phase in parents who are heterozygous at both loci being studied. Many approaches have been used over the years that aim to test, directly or indirectly, for lower than expected observed recombinations between two loci. These statistical approaches are of two basic types, often termed 'parametric' and 'nonparametric' linkage analysis.

Parametric or model-based or model-dependent linkage analysis (often called LOD score linkage analysis) assumes that the genetic models underlying both the trait

and marker loci are known. Thus, assumed values (parameters) for qualitative traits that must be specified for use in the analysis include the allele frequencies at the trait and marker loci, dominance relationships among the alleles, and relationships between genotypes and phenotypes at both the trait and marker loci (penetrance). For quantitative traits, the parameters that must be specified include allele frequencies at the trait and marker loci, the means and variances of the phenotype for each genotype, and the relative frequencies of the genotypes. The main difference between parametric linkage analysis for qualitative and quantitative traits is that definitive recombinants can be identified for qualitative trait linkage analysis but not for the linkage analysis of quantitative traits. This is due to the nature of the models underlying each type of trait. Because normal probability densities are used to model the genotypic distributions in quantitative linkage analysis, and these densities asymptotically approach, but never reach, zero in both tails, every individual has a non-zero probability for having each genotype. This is problematic when trying to identify recombination events that help to localize candidate regions, but methods have been developed to classify individuals based on their most probable genotype [27].

Non-parametric or model-free (or model-independent or weakly parametric) linkage methods make fewer assumptions about the underlying trait genetic model, although these methods still assume that the marker locus model(s) is known. These methods of analysis were first developed in the 1930's, with Fisher's [28] publication of maximum-likelihood scoring procedures called *u*-scores (parametric) and Penrose's [29] development of the sib-pair method (non-parametric). Fisher's *u*-scores and Finney's [30–35] extensions assumed specific models for the mating types at a trait locus and further assumed that the resulting score was normally distributed. Haldane and Smith [36] developed an 'inverse probability' ratio test, now known as a likelihood ratio test, that is the basis of modern parametric likelihood ratio tests for linkage. In this test, given a particular set of data, the likelihood of a hypothesis of linkage with some specific recombination fraction ($\theta < 1/2$) is compared to a hypothesis of no linkage, i.e. the independent assortment of the alleles at the two loci ($\theta = 1/2$). Smith [37] proposed taking the log of this test, and in 1955, Morton applied Wald's [38] sequential probability ratio test to combine results from a series of families and to determine appropriate significance levels for this sequential test [39]. Morton [39] coined the term LOD score, although the term 'LODs' was originally defined by Bernard [40] as the logarithm of the backward odds (the likelihood ratio). The two-point LOD score between a trait and a single marker locus is typically calculated over several recombination fractions between 0 and 1/2, and the recombination fraction that maximizes the likelihood (the maximum LOD score) is considered to be the best estimate of the recombination fraction. Traditionally, when the maximum LOD score is greater than 3 (a backward odds ratio of 1,000:1), the null hypothesis of independent assortment is rejected and linkage between the trait and the marker locus is assumed. Conversely, for those recombination fractions where the LOD score is less than –2, the null hypothesis of independent assortment is not rejected and linkage is assumed to be excluded. LOD scores can be converted to p values; a LOD score of 3 corresponds to a large-sample significance level of 0.0001 [39, 41, 42] and a reliability of 0.991 [43]. Morton subsequently extended the test to nuclear families, multiple allelic loci, sex linkage and genetic heterogeneity [44–46].

Elston and Stewart [47] developed a method (commonly called the Elston-Stewart algorithm) to compute the likelihood of a simple extended pedigree recursively and incorporated a general trait model that allowed for decreased penetrance and quantitative traits. Many types of trait models can be used with this algorithm. These are outside the scope of this overview, but comprehensive reviews are available in several articles and texts [27, 48–64]. Ott [65] implemented the Elston-Stewart algorithm to calculate the likelihood ratio test for linkage in human families of arbitrary size in LIPED, the first widely available computer program for this purpose. Many additional extensions to these methods have been published, including multipoint linkage analysis that uses information from multiple genetic markers, incorporation of variable age at onset and genetic heterogeneity, and methods that can analyze pedigrees with marriage or inbreeding loops [49, 66–75]. However, the computation time for multipoint linkage using the Elston-Stewart algorithm is prohibitive. Computation time for this algorithm scales linearly with the number of meioses but exponentially with the number of marker loci. Another major development was the Lander-Green algorithm for rapidly performing maximum-likelihood multilocus linkage computations [67, 76, 77]. The computation time for this algorithm scales linearly with the number of markers; however, it is only suitable for small pedigrees since the amount of computer memory required becomes prohibitive in pedigrees with a large number of meioses. Algorithms that calculate approximations to the likelihood of a pedigree for multipoint linkage, such as SIMWALK2 [78], offer a middle ground between these two options. Excellent treatments of these

subjects are found in several reviews and texts [79–87]. With the advent of dense maps of marker loci and multipoint linkage analysis (where the hypothesis of no linkage is tested assuming a recombination fraction of zero at thousands of locations along the chromosomal map), Lander and Kruglyak [88] proposed alternative significance thresholds based on an 'infinitely dense' map of marker loci to control the genome-wide probability of observing a false-positive linkage at 5%. Their proposed 'genome-wide significant' threshold of a LOD of 3.3 ($p = 4.9 \times 10^{-5}$) for parametric maximum-likelihood multipoint linkage analysis generated substantial controversy and methods development [41, 89–95] but has become a fairly standard guideline, as have their suggested significance thresholds for non-parametric allele-sharing linkage analyses (e.g. 2.2×10^{-5} in sibling pairs). Other factors that affect significance levels in linkage analyses are testing multiple parametric models [96–101], utilizing heterogeneity LOD scores [102–105], and the presence of inter-marker linkage disequilibrium when using a linkage method that assumes linkage equilibrium [106–108].

Non-parametric or model-free linkage methods do not require the specification of parameters for the mode of inheritance for the trait being linked to marker loci. These methods are based on testing whether relatives with similar trait phenotypes are also more similar than expected at a specific marker locus, implying low recombination rates between the unobserved trait locus and the specific marker locus. Non-parametric methods have also undergone substantial development since Penrose's introduction of the sib-pair test for qualitative and quantitative traits [29, 109]. These early tests were based on the proportion of alleles that a sib pair shared identical-by-state (IBS), which is also sometimes called identical-in-state (IIS). The number or proportion of alleles at a locus that are shared IBS by a pair of individuals is based solely on sharing the same allele(s) at the marker locus. More recent methods of model-free linkage are usually based on identity-by-descent (IBD) sharing among relatives, that is, the number or estimated proportion of alleles at a locus that are shared by a pair of relatives because they are copies of the same ancestral allele (inherited from a common, recent ancestor). Haseman and Elston [110] developed a model-free sib-pair linkage test based on estimates of IBD sharing among the sibling pairs for quantitative traits, and Suarez et al. [111] developed a similar IBD-based sib-pair linkage test for a qualitative trait. Amos et al. [112] extended these methods to other relative pairs in addition to sibs. Multipoint estimates of IBD sharing in sibling pairs at any genomic location were developed by Kruglyak et al. [113] and Kruglyak and Lander [114] based on the Lander-Green algorithm and later extended to additional types of relative pairs [77].

These IBD estimates are utilized somewhat differently in model-free tests of linkage for quantitative and qualitative traits. For quantitative traits, Haseman and Elston [110] proposed regressing the square of the difference of the trait values in the sibling pair against the estimated proportion of alleles shared IBD at a single marker locus with an extension to several loci without epistatic interaction. Amos and Elston [115] extended this to the squared trait difference for various other types of relative pairs. The slope of this regression line is expected to be zero under the null hypothesis of no linkage, inferring that the estimated proportion of alleles shared IBD has no effect on the trait difference. Similarly, the slope of the regression is non-zero in the presence of linkage, so a one-sided t test for a non-zero slope is the test of interest. Further extensions were also made to allow for dominance variance and epistatic interactions [116–118]. Variance components analysis has also been used for linkage for quantitative traits [119, 120] by modeling the variance of the quantitative trait into components due to a causal gene linked to a specific location on the marker map and residual polygenic and environmental components. These methods have been extended to allow for analyses of large pedigrees [121, 122]. Elston et al. [123] introduced a revised Haseman-Elston regression method that has similar power to variance components methods. Several reviews of these methods exist [124–127].

For qualitative or dichotomous traits, one can utilize the methods for quantitative traits by simply coding affected individuals as '1' and unaffected individuals as '0' to create a quantitative phenotype and testing the difference between the means of the two groups. However, other approaches are often taken for qualitative traits, where the IBD sharing at marker loci is studied conditional on affection status. These methods include the 'affected pairs' methods. In 1953, Penrose [128] introduced an affected sib-pair linkage test that tests whether the proportion of alleles IBD at a marker was larger than expected, and many other methods building on this concept have been proposed [111, 129–139]. Tests for linkage when the trait is caused by multiple loci have also been developed [140–144]. Tests have also been developed that allow all affected pairs in a pedigree to be tested for excess IBD sharing together [135, 145–148].

Parametric and non-parametric methods have different strengths and weaknesses [149]. Parametric linkage analysis is more powerful than non-parametric linkage

methods if the genetic model for both trait and marker loci are correctly specified; however, for complex traits where such correct model specification is difficult, nonparametric methods may be more powerful.

Linkage and Complex Diseases and Traits

Linkage studies have been successful in leading to the discovery of genetic loci that contribute to the risk of diseases or variation of quantitative traits. For example, the *BRCA1* gene with hundreds of rare, high-penetrance risk alleles that cause major increases in the risk of breast and ovarian cancer was first discovered [150] after its location was identified via a linkage study [151]. Since then, more genes have been identified with inherited mutations that predispose to breast or ovarian cancer, but most risk alleles in these loci are individually rare in the population. Walsh et al. [152] recently showed that genomic capture and massively parallel sequencing of these genes can detect a wide array of known mutations in 21 of these breast-ovarian cancer risk loci.

However, for many traits, existing genome-wide significant [88] and replicated linkage results have not resulted in the identification of the genes responsible for the linkage signals. There are several possible reasons for this phenomenon. First, as with any statistical test, false-positive results (type I errors) are to be expected. The more linkage studies that are carried out for a specific disease or trait, the higher the chance that a highly significant false-positive linkage will be observed [15]. While genome-wide significant linkage results have high reliability [41, 43, 100], linkage results with 'suggestive' significance levels are not as reliable and have a higher probability of being false-positive results [43]. However, performing linkage analyses appropriately (without violation of the assumptions of the analysis methods), calculating significance levels appropriately for the specific analysis methodology [100, 105], requiring stringent significance levels to declare 'genome-wide significance' [88] and also requiring replication of a significant linkage in an adequately powered [88, 153] independent dataset (a practice that is also common in GWAS) before considering a linkage result to have strong support, will help to control this [16]. However, such stringent control of false-positive rates will decrease power to detect linkage. A second reason that significant linkage results have not resulted in the identification of the genes responsible for the linkage signals is that the correct gene has been identified, but its function and the effect of mutations on this function are not yet well enough understood for researchers to realize that it is the causal locus. In addition, gene-gene or gene-environment interactions may cause inconsistent associations when mutations discovered in linked families are subsequently evaluated in population-based association studies. However, for many linkage findings, the reason that a causal locus has not been found may be that adequate DNA sequencing has not yet been performed. In the past, when only Sanger sequencing methods were available, DNA sequencing of the entire region under a linkage peak was prohibitively expensive because these linked candidate regions can often cover 100–200 megabases. Sequencing in these regions has often been limited to only a few exons in a few candidate genes. As the Human Genome Project has progressed, our understanding of DNA structure and function has grown, such that we now realize that we must sequence not just exons but also promoters, splice sites, 3′ UTRs, microRNAs, long non-coding RNAs and other non-coding regulatory elements. For many candidate linkage regions, the failure to identify the causal disease gene may simply mean we have not yet sequenced enough DNA in the region on a large enough sample of people. Next-generation DNA sequencing holds the promise to allow us to eliminate the last two possibilities by making it economically feasible to thoroughly sequence the DNA of an adequate number of affected individuals for many diseases. However, we must recognize that these methods are not a panacea, and that complex diseases and traits are indeed complex.

Benefits of Combining Linkage and DNA Sequence Information

As large samples of whole-exome and whole-genome sequence data have been accumulated, certain issues have become clear. First, rare variants are individually rare, but each person will have thousands of such rare variants across their genome. It can be difficult to determine whether a novel variant is a sequencing artifact or whether it is a true variant when only a single individual in a sample exhibits this variant. However, one expects that even rare variants should segregate within a family. Thus, family studies of DNA sequence data can be useful for determining which rare variants are likely to be real variants and also which variants segregate with a disease or trait within the family, and analogously whether copy number variants are likely to be inherited or novel, although identifying variants based on repeated sequences is still somewhat problematic at this point in time. This

method of measuring the co-segregation of any sequence variant with disease is simply linkage analysis. Linkage analysis results can be used to identify families that are most likely to segregate genetic variants and to guide interpretation of whole-exome and whole-genome sequencing results or to choose regions for targeted DNA sequencing. Results from the recent Genetic Analysis Workshop 17 suggested that analyses of rare variants in whole-exome sequence data would require much larger sample sizes in studies of unrelated individuals than in family studies, since family studies allow amplification of effect of the rare variants because many family members carry the same rare variant [154]. Combining linkage studies with sequencing can allow the identification of important genes and gene pathways, which can then become candidates for sequencing in much larger samples of individuals with the pertinent disease or trait.

References

1 Tsui LC, et al: Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. Science 1985;230:1054–1057.
2 Riordan JR, et al: Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 1989;245: 1066–1073.
3 Gusella JF, et al: A polymorphic DNA marker genetically linked to Huntington's disease. Nature 1983;306:234–238.
4 The Huntington Disease Collaborative Research Group: A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 1993;72:971–983.
5 Brant SR, et al: American families with Crohn's disease have strong evidence for linkage to chromosome 16 but not chromosome 12. Gastroenterology 1998;115:1056–1061.
6 Ogura Y, et al: A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001;411:603–606.
7 Hugot JP, et al: Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411:599–603.
8 Hampe J, et al: Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet 2001;357:1925–1928.
9 Goldin LR, et al: Segregation and linkage studies of plasma dopamine-beta-hydroxylase (DBH), erythrocyte catechol-O-methyltransferase (COMT), and platelet monoamine oxidase (MAO): possible linkage between the ABO locus and a gene controlling DBH activity. Am J Hum Genet 1982;34: 250–262.
10 Asamoah A, et al: Segregation and linkage analyses of dopamine-beta-hydroxylase activity in a six-generation pedigree. Am J Med Genet 1987;27:613–621.
11 Wilson AF, et al: Linkage of a gene regulating dopamine-beta-hydroxylase activity and the ABO blood group locus. Am J Hum Genet 1988;42:160–166.
12 Craig SP, et al: Localization of the human dopamine beta hydroxylase (DBH) gene to chromosome 9q34. Cytogenet Cell Genet 1988;48:48–50.
13 Zabetian CP, et al: A quantitative-trait analysis of human plasma-dopamine beta-hydroxylase activity: evidence for a major functional polymorphism at the DBH locus. Am J Hum Genet 2001;68:515–522.
14 Cubells JF, et al: Linkage analysis of plasma dopamine-B-hydroxylase in families of patients with schizophrenia. Hum Genet 2011; 130:635–643.
15 Ioannidis JP: Why most published research findings are false. PLoS Med 2005;2:e124.
16 Moonesinghe R, Khoury MJ, Janssens AC: Most published research findings are false – but a little replication goes a long way. PLoS Med 2007;4:e28.
17 Sobreira NL, et al: Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. PLoS Genet 2010;6:e1000991.
18 Bowden DW, et al: Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. Hum Mol Genet 2010;19: 4112–4120.
19 Guo X, et al: Genome-wide linkage of plasma adiponectin reveals a major locus on chromosome 3q distinct from the adiponectin structural gene: the IRAS family study. Diabetes 2006;55:1723–1730.
20 Sutton WS: On the morphology of the chromosome group in Brachystola magna. Biol Bull 1902;4:24–39.
21 Sutton WS: The chromosomes in heredity. Biol Bull 1903;4:231–251.
22 Correns C: Uber Vererbungsgesetze. Berlin, G. Borntrager, 1905.
23 Bateson W, Punnett RC: Experimental studies in the physiology of heredity. Reports of the Evolution Committee. Roy Soc 1906;3:1–53.
24 Muller HJ: The mechanism of crossing over. Am Nat 1916;50:193–221.
25 Morgan TH: Random segregation versus coupling in Mendelian inheritance. Science 1911;34:384.
26 Sturtevant AH: The linear arrangement of six sex-linked factors in Drosophila as shown by their mode of association. J Exp Zool 1913; 14:43–59.
27 Wilson AF, et al: Stepwise oligogenic segregation and linkage analysis illustrated with dopamine-beta-hydroxylase activity. Am J Med Genet 1990;35:425–432.
28 Fisher RA: The detection of linkage. Ann Eugen 1935;6:187–201.
29 Penrose LS: The detection of autosomal linkage in data which consists of pairs of brothers and sisters of unpecified parentage. Ann Eugen 1935;6:133–138.
30 Finney DJ: The detection of linkage. Ann Eugen 1940;10:171–214.
31 Finney DJ: The detection of linkage. II. Further mating types, scoring of Boyd's data. Ann Eugen 1941;11:10–30.
32 Finney DJ: The detection of linkage. III. Incomplete parental testing. Ann Eugen 1941; 11:115–135.
33 Finney DJ: The detection of linkage. IV. Lack of parental records and the use of empirical information. J Hered 1942;33:157–160.
34 Finney DJ: The detection of linkage. V. Supplementary tables. Ann Eugen 1942;11:224–232.
35 Finney DJ: The detection of linkage. VII. Combination of data from matings of known and unknown phase. Ann Eugen 1943;12: 31–43.
36 Haldane JB, Smith CA: A new estimate of the linkage between the genes for colourblindness and haemophilia in man. Ann Eugen 1947;14:10–31.
37 Smith CAB: The detection of linkage in human genetics. J Roy Stat Soc B 1953;15:153–184.
38 Wald A: Sequential Analysis. New York, Wiley, 1947.
39 Morton NE: Sequential tests for the detection of linkage. Am J Hum Genet 1955;7:277–318.
40 Bernard GA: Statistical inference. J R Stat Soc B 1949;11:115–140.
41 Morton NE: Significance levels in complex inheritance. Am J Hum Genet 1998;62:690–697.
42 Morton NE: Erratum, significance levels in complex inheritance. Am J Hum Genet 1998; 63:1252.
43 Rao DC, et al: Variability of human linkage data. Am J Hum Genet 1978;30:516–529.

44 Morton NE: The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. Am J Hum Genet 1956;8:80–96.
45 Morton NE, Steinberg AG: Sequential test for linkage between cystic fibrosis of the pancreas and the MNS locus. Am J Hum Genet 1956;8:177–189.
46 Morton NE: Further scoring types in sequential linkage tests, with a critical review of autosomal and partial sex linkage in man. Am J Hum Genet 1957;9:55–75.
47 Elston RC, Stewart J: A general model for the genetic analysis of pedigree data. Hum Hered 1971;21:523–542.
48 Mather K, Jinks JL: Biometrical Genetics; the Study of Continuous Variation, ed 2, revised. Ithaca, Cornell Univ Press, 1971, p 382.
49 Lange K, Elston RC: Extensions to pedigree analysis I. Likehood calculations for simple and complex pedigrees. Hum Hered 1975;25: 95–105.
50 Mather K, Jinks JL: Introduction to Biometrical Genetics. Ithaca, Cornell Univ Press, 1977, p 231.
51 Elston RC: Segregation analysis. Adv Hum Genet 1981;11:63–120, 372–373.
52 Mather K, Jinks JL: Biometrical Genetics: The Study of Continuous Variation, ed 3. London, New York, Chapman and Hall, 1982, p 396.
53 Lalouel JM, et al: A unified model for complex segregation analysis. Am J Hum Genet 1983;35:816–826.
54 Elston RC: Segregation and linkage analysis. Anim Genet 1992;23:59–62.
55 Bonney GE: Regressive logistic models for familial disease and other binary traits. Biometrics 1986;42:611–625.
56 Bonney GE: A note on the basis of regressive models for genetic analysis. Genet Epidemiol Suppl 1986;1:37–42.
57 Bonney GE, Lathrop GM, Lalouel JM: Combined linkage and segregation analysis using regressive models. Am J Hum Genet 1988;43: 29–37.
58 Bonney GE, Dunston GM, Wilson J: Regressive logistic models for ordered and unordered polychotomous traits: application to affective disorders. Genet Epidemiol 1989;6: 211–215.
59 Demenais FM, Bonney GE: Equivalence of the mixed and regressive models for genetic analysis. I. Continuous traits. Genet Epidemiol 1989;6:597–617.
60 Bonney GE: Compound regressive models for family data. Hum Hered 1992;42:28–41.
61 Stricker C, Fernando RL, Elston RC: Segregation analysis under an alternative formulation for the mixed model. Genet Epidemiol 1993;10:653–658.
62 Falconer DS, Mackay TFC: Introduction to Quantitative Genetics, ed 4. Essex, Longman, 1996, p 464.
63 Elston RC, Olson JM, Palmer L: Biostatistical Genetics and Genetic Epidemiology. Wiley reference series in biostatistics. Chichester, New York, Wiley, 2002, p 831.
64 Wilson AF, Elston RC: Linkage analysis in the study of the genetics of alcoholism; in Begleiter H, Kissin B (eds): The Genetics of Alcoholism. Oxford University Press, 1995, pp 353–376.
65 Ott J: Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. Am J Hum Genet 1974;26:588–597.
66 Lander E, Botstein D: Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 1989;121:185–199.
67 Lander ES, Green P: Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 1987;84:2363–2367.
68 Ott J: Linkage analysis and family classification under heterogeneity. Ann Hum Genet 1983;47:311–320.
69 Smith CAB: Homogeneity test for linkage data. 2nd International Congress on Human Genetics, 1961.
70 Cannings C, Thompson EA, Skolnick MH: Probability functions on complex pedigrees. Adv Appl Prob 1978;10:26–61.
71 Meyers DA, et al: Linkage group I: the simultaneous estimation of recombination and interference. Cytogenet Cell Genet 1976;16: 335–339.
72 Lathrop GM, et al: Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 1984;81:3443–3446.
73 Lathrop GM, et al: Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am J Hum Genet 1985;37:482–498.
74 Cottingham RW Jr, Idury RM, Schaffer AA: Faster sequential genetic linkage computations. Am J Hum Genet 1993;53:252–263.
75 Hodge SE, et al: Age-of-onset correction available for linkage analysis (LIPED). Am J Hum Genet 1979;31:761–762.
76 Lander ES, et al: MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1987;1: 174–181.
77 Kruglyak L, et al: Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 1996;58:1347–1363.
78 Sobel E, Lange K: Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. Am J Hum Genet 1996;58:1323–1337.
79 Forabosco P, Falchi M, Devoto M: Statistical tools for linkage analysis and genetic association studies. Expert Rev Mol Diagn 2005; 5:781–796.
80 Ott J: Analysis of Human Genetic Linkage, ed 3. Baltimore, Johns Hopkins University Press, 1999.
81 Dudbridge F: A survey of current software for linkage analysis. Hum Genomics 2003;1: 63–65.
82 Elston RC, Anne Spence M: Advances in statistical human genetics over the last 25 years. Stat Med 2006;25:3049–3080.
83 Terwilliger JD, Goring HH: Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 2000;72:63–132.
84 Strug LJ, Hodge SE: An alternative foundation for the planning and evaluation of linkage analysis. I. Decoupling 'error probabilities' from 'measures of evidence'. Hum Hered 2006;61:166–188.
85 Strug LJ, Hodge SE: An alternative foundation for the planning and evaluation of linkage analysis. II. Implications for multiple test adjustments. Hum Hered 2006;61: 200–209.
86 Freimer N, Sabatti C: The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology. Nat Genet 2004;36:1045–1051.
87 Conneally PM, Rivas ML: Linkage analysis in man. Adv Hum Genet 1980;10:209–266.
88 Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 1995;11:241–247.
89 Simonsen KL, McIntyre LM: Using alpha wisely: improving power to detect multiple QTL. Stat Appl Genet Mol Biol 2004;3:1.
90 Cheverud JM: A simple correction for multiple comparisons in interval mapping genome scans. Heredity 2001;87:52–58.
91 Curtis D: Genetic dissection of complex traits. Nat Genet 1996;12:356–357.
92 Risch N, Botstein D: A manic depressive history. Nat Genet 1996;12:351–353.
93 Witte JS, Elston RC, Schork NJ: Genetic dissection of complex traits. Nat Genet 1996; 12:355–356.
94 Rao DC: CAT scans, PET scans, and genomic scans. Genet Epidemiol 1998;15:1–18.
95 Fernando RL, et al: Controlling the proportion of false positives in multiple dependent tests. Genetics 2004;166:611–619.
96 Risch N: A note on multiple testing procedures in linkage analysis. Am J Hum Genet 1991;48:1058–1064.
97 Hodge SE, Abreu PC, Greenberg DA: Magnitude of type I error when single-locus linkage analysis is maximized over models: a simulation study. Am J Hum Genet 1997; 60:217–227.
98 Greenberg DA, et al: Power, mode of inheritance, and type I error in LOD scores and affecteds-only methods: reply to Kruglyak. Am J Hum Genet 1998;62:202–204.
99 Greenberg DA, Abreu PC: Determining trait locus position from multipoint analysis: accuracy and power of three different statistics. Genet Epidemiol 2001;21:299–314.
100 Hodge SE, et al: Multipoint lods provide reliable linkage evidence despite unknown limiting distribution: type I error probabilities decrease with sample size for multipoint lods and mods. Genet Epidemiol 2008;32:800–815.
101 Camp NJ, Farnham JM: Correcting for multiple analyses in genomewide linkage studies. Ann Hum Genet 2001;65:577–582.

102 Vieland VJ, Logue M: HLODs, trait models, and ascertainment: implications of admixture for parameter estimation and linkage detection. Hum Hered 2002;53:23–35.
103 Abreu PC, Hodge SE, Greenberg DA: Quantification of type I error probabilities for heterogeneity LOD scores. Genet Epidemiol 2002;22:156–169.
104 Hodge SE, Vieland VJ, Greenberg DA: HLODs remain powerful tools for detection of linkage in the presence of genetic heterogeneity. Am J Hum Genet 2002;70: 556–559.
105 Xing C, Morris N, Xing G: Distribution of model-based multipoint heterogeneity lod scores. Genet Epidemiol 2010;34:912–916.
106 Huang Q, et al: Examining the effect of linkage disequilibrium on multipoint linkage analysis. BMC Genet 2005;6(suppl 1):S83.
107 Boyles AL, et al: Linkage disequilibrium inflates type I error rates in multipoint linkage analysis when parental genotypes are missing. Hum Hered 2005;59:220–227.
108 Kim Y, et al: Examining the effect of linkage disequilibrium between markers on the Type I error rate and power of nonparametric multipoint linkage analysis of two-generation and multigenerational pedigrees in the presence of missing genotype data. Genet Epidemiol 2008;32:41–51.
109 Penrose LS: Genetic linkage in graded human characters. Ann Eugen 1938;8:233–237.
110 Haseman JK, Elston RC: The investigation of linkage between a quantitative trait and a marker locus. Behav Genet 1972;2:3–19.
111 Suarez BK, Rice J, Reich T: The generalized sib pair IBD distribution: its use in the detection of linkage. Ann Hum Genet 1978; 42:87–94.
112 Amos CI, Dawson DV, Elston RC: The probabilistic determination of identity-by-descent sharing for pairs of relatives from pedigrees. Am J Hum Genet 1990;47:842–853.
113 Kruglyak L, Daly MJ, Lander ES: Rapid multipoint linkage analysis of recessive traits in nuclear families, including homozygosity mapping. Am J Hum Genet 1995; 56:519–527.
114 Kruglyak L, Lander ES: Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am J Hum Genet 1995; 57:439–454.
115 Amos CI, Elston RC: Robust methods for the detection of genetic linkage for quantitative data from pedigrees. Genet Epidemiol 1989;6:349–360.
116 Elston RC: The genetic dissection of multifactorial traits. Clin Exp Allergy 1995; 25(suppl 2):103–106.
117 Tiwari HK, Elston RC: Linkage of multilocus components of variance to polymorphic markers. Ann Hum Genet 1997;61: 253–261.
118 Tiwari HK, Elston RC: Restrictions on components of variance for epistatic models. Theor Popul Biol 1998;54:161–174.
119 Goldgar DE: Multipoint analysis of human quantitative genetic variation. Am J Hum Genet 1990;47:957–967.
120 Schork NJ: Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. Am J Hum Genet 1993;53: 1306–1319.
121 Amos CI: Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet 1994;54:535–543.
122 Blangero J, Almasy L: SOLAR: Sequential Oligogenic Linkage Analysis Routines. Technical notes. 1996, Southwest Foundation for Biomedical Research. Population Genetics Laboratory: San Antonio, TX.
123 Elston RC, et al: Haseman and Elston revisited. Genet Epidemiol 2000;19:1–17.
124 Elston RC, Cordell HJ: Overview of model-free methods for linkage analysis. Adv Genet 2001;42:135–150.
125 Almasy L, Blangero J: Contemporary model-free methods for linkage analysis. Adv Genet 2008;60:175–193.
126 Almasy L, Blangero J: Human QTL linkage mapping. Genetica 2009;136:333–340.
127 Almasy L, Blangero J: Variance component methods for analysis of complex phenotypes. Cold Spring Harb Protoc 2010; 2010:pdb.top77.
128 Penrose LS: The general purpose sib pair linkage test. Ann Eugen London 1953;18: 120–124.
129 Day NE, Simons MJ: Disease susceptibility genes – their identification by multiple case family studies. Tissue Antigens 1976;8: 109–119.
130 Fishman PM, et al: A robust method for the detection of linkage in familial disease. Am J Hum Genet 1978;30:308–321.
131 Blackwelder WC, Elston RC: A comparison of sib-pair linkage tests for disease susceptibility loci. Genet Epidemiol 1985;2:85–97.
132 Knapp M, Seuchter SA, Baur MP: Linkage analysis in nuclear families. 1. Optimality criteria for affected sib-pair tests. Hum Hered 1994;44:37–43.
133 Schaid DJ, Nick TG: Sib-pair linkage tests for disease susceptibility loci: common tests vs. the asymptotically most powerful test. Genet Epidemiol 1990;7:359–370.
134 Feingold E, Siegmund DO: Strategies for mapping heterogeneous recessive traits by allele-sharing methods. Am J Hum Genet 1997;60:965–978.
135 Whittemore AS, Halpern J: A class of tests for linkage using affected pedigree members. Biometrics 1994;50:118–127.
136 Whittemore AS, Halpern J: Probability of gene identity by descent: computation and applications. Biometrics 1994;50:109–117.
137 Whittemore AS, Tu IP: Simple, robust linkage tests for affected sibs. Am J Hum Genet 1998;62:1228–1242.
138 Holmans P: Asymptotic properties of affected-sib-pair linkage analysis. Am J Hum Genet 1993;52:362–374.
139 Faraway JJ: Improved sib-pair linkage test for disease susceptibility loci. Genet Epidemiol 1993;10:225–233.
140 Knapp M, Seuchter SA, Baur MP: Two-locus disease models with two marker loci: the power of affected-sib-pair tests. Am J Hum Genet 1994;55:1030–1041.
141 Dupuis J, Brown PO, Siegmund D: Statistical methods for linkage analysis of complex traits from high-resolution maps of identity by descent. Genetics 1995;140:843–856.
142 Farrall M: Affected sibpair linkage tests for multiple linked susceptibility genes. Genet Epidemiol 1997;14:103–115.
143 Olson JM: Likelihood-based models for genetic linkage analysis using affected sib pairs. Hum Hered 1997;47:110–120.
144 Cordell HJ, et al: Multilocus linkage tests based on affected relative pairs. Am J Hum Genet 2000;66:1273–1286.
145 Davis S, et al: Nonparametric simulation-based statistics for detecting linkage in general pedigrees. Am J Hum Genet 1996;58: 867–880.
146 Kong A, Cox NJ: Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 1997;61:1179–1188.
147 McPeek MS: Optimal allele-sharing statistics for genetic mapping using affected relatives. Genet Epidemiol 1999;16:225–249.
148 Olson JM: A general conditional-logistic model for affected-relative-pair linkage studies. Am J Hum Genet 1999;65:1760–1769.
149 Bailey-Wilson JE: Parametric versus nonparametric and two-point versus multipoint: controversies in gene mapping; in Dunn M, et al. (eds): Encyclopedia of Genomics, Proteomics and Bioinformatics. John Wiley & Sons, 2005.
150 Miki Y, et al: A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 1994;266:66–71.
151 Hall JM, et al: Linkage of early-onset familial breast cancer to chromosome 17q21. Science 1990;250:1684–1689.
152 Walsh T, et al: Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. Proc Natl Acad Sci USA 2010; 107:12629–12633.
153 Suarez BK, Hampe CL, Van Eerdewegh P: Problems of replicating linkage claims in psychiatry; in Gershon ES, Cloninger CR (eds): Genetic Approaches to Mental Disorders. Washington, American Psychiatric Press, 1994, pp 23–46.
154 Wilson AF, Ziegler A: Lessons learned from genetic analysis workshop 17: transitioning from genome-wide association studies to whole-genome statistical genetic analysis. Genet Epidemiol 2011, in press.