Original Paper

Human Heredity

Hum Hered 2005;59:118–124
DOI: 10.1159/000085226

Received: January 25, 2005
Accepted after revision: February 22, 2005
Published online: April 18, 2005

Linkage Disequilibrium: Ancient History Drives the New Genetics

Gonçalo R. Abecasis Debashis Ghosh Thomas E. Nichols

Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, Mich., USA

Key Words
Linkage disequilibrium · Natural selection · Association mapping · Recombination rate

Abstract
This brief review provides a summary of the biological causes of genetic association between tightly linked markers – termed linkage disequilibrium – and unlinked markers – termed population structure. We also review the utility of linkage disequilibrium data in gene mapping in isolated populations, in the estimation of recombination rates and in studying the history of particular alleles, including the detection of natural selection. We discuss current understanding of the extent and patterns of linkage disequilibrium in the genome, and its promise for genetic association studies in complex disease. Finally, we highlight the importance of using appropriate statistical procedures, such as the false discovery rate, to maximize the chances of success in large scale association studies.

Introduction

Genetic variation can determine disease susceptibility [Botstein and Risch, 2003], elucidate the history of human populations [Underhill et al., 2000], and provide the tools for understanding basic biological processes [Jeffreys et al., 2001]. Recently, studies of genetic variation have focused on single nucleotide polymorphisms (SNPs), an abundant class of DNA variants that can be cheaply genotyped in very large numbers – the most recent version of dbSNP [Sachidanandam et al., 2001] includes about 10 million such polymorphisms, and it is likely that many more exist [Carlson et al., 2003].

Like other genetic polymorphisms, these SNPs are not independent, but exhibit complex relationships. For most of them, a single mutation event in the distant past resulted in the two alleles that exist today. The original copy of the mutant allele was confined to a specific haplotype belonging to a specific individual in a specific population. Over time, this allele may have spread to other haplotypes through recombination, gene-conversion, or recurrent mutation; it may have spread to other populations by migration of its carriers; and it may have changed in frequency due to natural selection (which can even change the frequencies of neutral alleles through a process known as hitch-hiking) or genetic drift. Together, these phenomena have resulted in specific patterns of linkage disequilibrium and population structure in modern populations.

Linkage Disequilibrium in the Genome

Linkage disequilibrium refers to association between tightly linked SNPs. This type of association can exist, for example, because one allele is preferentially flanked by a

KARGER

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

0001–5652/05/0592–0118$22.00/0

Accessible online at:
www.karger.com/hhe

Dr. G.R. Abecasis
Department of Biostatistics, School of Public Health
University of Michigan
Ann Arbor, MI (USA)
Tel. +1 734 763 4901, Fax +1 734 615 8322, E-Mail goncalo@umich.edu

short stretch of the ancestral haplotype where it first originated through mutation or entered a particular population through migration (fig. 1). For the majority of non-selected alleles, linkage disequilibrium is gradually broken down by recombination and gene-conversion, and regenerated by genetic drift. In most human populations, LD extends for relatively short distances, on the order of 10s to 100s of kb, in most genomic regions [Abecasis et al., 2001; Reich et al., 2001]. In a few cases, LD may extend for longer distances, for example, when the effective population size is small or in populations that have undergone recent admixture [Zhu et al., 2005].

While population structure can result in strong association between SNPs that differ in frequency between the individual subpopulations, it is important to note that its effects are not restricted to linked SNPs. Association signals that are due to population structure occur when particular variants are sequestered in a geographically or culturally isolated populations. Over time, mutation and genetic drift produce unique patterns of variation in these isolated populations – for example, alleles that are rare in other populations can become common in isolated populations, and some other alleles may be lost altogether. Thus, allelic association due to population structure will occur between both linked and unlinked markers, and care must be taken when using these populations for gene mapping. Although most genetic polymorphisms are shared between many populations, a fraction shows substantial differences in frequency between populations [Rosenberg et al., 2002].

Our understanding of linkage disequilibrium and population structure has advanced greatly in recent years. For example, early studies of linkage disequilibrium focused on HLA [Tomlinson and Bodmer, 1995], single isolated genes [Chakravarti et al., 1984a, b; Jorde et al., 1994], and later on describing the average correlation between nearby SNPs [Abecasis et al., 2001; Reich et al., 2001]. More recent studies examined regional variation in linkage disequilibrium patterns and showed that 'step-like' patterns in the decay of disequilibrium are common [Daly et al., 2001; Dawson et al., 2002; Gabriel et al., 2002]. Together with the results of new sperm-typing studies [Jeffreys et al., 2001], these observations have suggested that most recombination events in humans may occur in tightly localized hotspots. Future linkage disequilibrium studies will allow the construction of very fine-scale maps of recombination [Zhang et al., 2002; McVean et al., 2004] in the genome, and provide insight into the underlying biological processes.

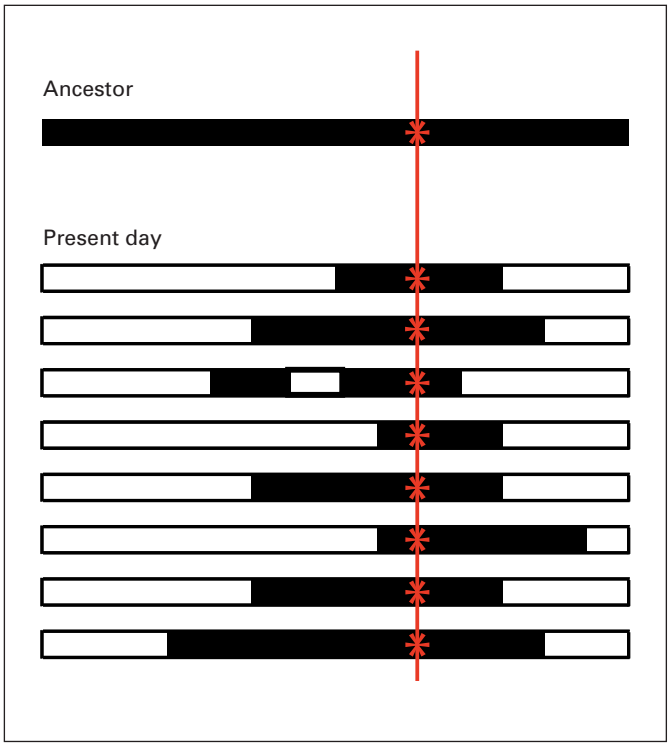

Fig. 1. Pictorial representation of the origin of linkage disequilibrium. Most SNP alleles originated with the mutation of a single ancestral chromosome. In the figure, this hypothetical ancestral chromosome is colored in black, and the mutant allele is indicated by the red star (*). Present-day carriers of this mutant allele will also carry a small surrounding stretch of the ancestral chromosome. Alleles within these short stretches are not associated randomly, but instead often appear in the same configuration as in the original mutant chromosome, and are said to be in linkage disequilibrium. The exact length of this conserved stretch will depend on local rates of recombination, gene conversion, and mutation, since all of these phenomena erode the original ancestral chromosome. Population history, especially through genetic drift and natural selection, also plays a role as it shapes the genealogy of individual alleles and determines the number of rounds of mutation, recombination and natural selection to which each allele is subjected.

Uses of Linkage Disequilibrium Data

One of most significant applications of linkage disequilibrium to date has been the study of rare diseases in isolated populations. In addition to greater phenotypic homogeneity and reduced genetic diversity, these populations often exhibit high levels of linkage disequilibrium because of genetic drift [Peltonen et al., 2000]. In these populations, rare disease alleles are often surrounded by extensive (up to several megabases) haplotypes that are shared between many affected individuals – a phenom-

enon that has facilitated the localization and identification of many disease genes [Peltonen et al., 2000].

Linkage disequilibrium data has also been useful in studies of the history of individual loci or alleles, within a population. Because linkage disequilibrium is expected to be most extensive surrounding relatively young alleles (which are either the result of recent mutation or migration events), the pattern of linkage disequilibrium around specific alleles can be used to date their origin and to identify alleles that have undergone positive selection – positively selected alleles increase in frequency rapidly and are surrounded by more linkage disequilibrium than other alleles of similar frequency. Examples include studies of the CCR5-Δ32 allele [Stephens et al., 1998], which confers resistance to AIDS, and alleles at the G6PD and CD40 loci [Sabeti et al., 2002], which confer resistance to malaria. In both cases, there are very long conserved haplotypes surrounding the disease resistance alleles. The example of malaria resistance is particularly interesting, since extended linkage disequilibrium is observed surrounding alleles that confer resistance to infection in humans [Sabeti et al., 2002] and surrounding alleles that confer resistance to commonly used treatments in the malaria parasite *Plasmodium falciparum* [Wootton et al., 2002]. The conserved haplotypes and extended linkage disequilibrium are the signature of positive selection – which rapidly increased the frequency of alleles conferring resistance (to infection in humans and to treatment in *P. falciparum*) without allowing recombination to erode the ancestral haplotypes where these alleles originated.

Despite many important applications, such as the study of recombination rates, gene identification in isolated populations and the demonstration of natural selection, the use of linkage disequilibrium data that has generated the most interest is in the detection of association between common variants and common disease [Cardon and Abecasis, 2003]. Specifically, knowledge of patterns of linkage disequilibrium in the genome means that carefully selected sets of SNPs [Johnson et al., 2001; Cardon and Abecasis, 2003; Carlson et al., 2004] can provide useful information on much larger sets of unobserved variants. This means that comprehensive surveys of human polymorphism can be carried out even when it is too costly, or otherwise impractical, to survey all known variants (at this point, >10 million). This possibility provided the impetus for the soon to be completed Human Haplotype Map Project [The International HapMap Consortium, 2003], which has already provided information on the patterns of variation and linkage disequilibrium for >1,000,000 SNPs (genotype data publicly available at http://www.hapmap.org), and has led to the development of many statistical methods that attempt to use linkage disequilibrium to learn about the properties and effects of unobserved alleles [see Durrant et al., 2004 for a recent example]. Results for a similar genome-wide survey of linkage disequilibrium including ~1.5 million SNPs have just been reported [Hinds et al., 2005]. Hinds and colleagues [2005] have placed their data in the public domain and it is likely that it will quickly be used to help design many gene mapping studies.

Linkage Disequilibrium in Genome-Wide Association Studies

Linkage disequilibrium between markers examined in a particular association study and unexamined markers, means that it is possible to survey substantial fractions of human variation in a cost-effective manner. This is true whether randomly selected SNPs, evenly spaced SNPs or carefully selected tag SNPs are genotyped. While the latter approach is likely to be most efficient on a per SNP basis, it is possible that designs based on randomly selected SNPs might be more efficient on dollar basis. This dichotomy results because some platforms can assay very large numbers of unselected SNPs for extremely low cost [Matsuzaki et al., 2004]. While we are still waiting for results of the first genome-wide association studies, results of many smaller studies show the approach is likely to be effective. For example, genetic association studies can demonstrate the impact of polymorphism in specific genes on mRNA transcript levels [Morley et al., 2004], protein levels [McKenzie et al., 2001] and protein activity levels [Zabetian et al., 2001; Zabetian et al., 2003] even when causal variants are not genotyped. In each of these cases, linkage disequilibrium between trait-modifying alleles and other nearby variants meant that genotyping a relatively small number of markers [McKenzie et al., 2001; Zabetian et al., 2001; Morley et al., 2004] near the trait locus was sufficient to demonstrate a genetic effect, before more comprehensive studies of genetic variation were undertaken. Even for complex traits, there is convincing evidence that linkage disequilibrium will often lead to detectable association at alleles that surround a trait locus but are not themselves causal [Martin et al., 2000] – again leading to the prospect of cost-effective, large-scale association studies.

Linkage disequilibrium between variants and haplotypes tested within a particular study, however, presents

important analytical challenges. Specifically, we might expect to detect association with multiple SNPs (or haplotypes), even when a single disease allele exists. However, conventional statistical techniques are designed to assess significance in situations where a single alternative hypothesis is true (i.e., a single marker shows association). The problem is even more severe when scans of candidate regions are expanded to a genome-wide scale, at which multiple disease susceptibility alleles might be presented. As these large scale association studies are carried out in the next few years, it will be important to accurately assess the information for association they provide while taking into account the dependence between measured SNPs and haplotypes, as well as the possibility of multiple associated markers. The Bonferroni correction and other conventional statistical techniques that assume independent data are likely to be inefficient in these settings, and in the next section we discuss some promising alternatives. An emerging, and potentially very useful, class of methods based on the coalescent [Morris et al., 2002; Zollner and Pritchard, 2004] is not discussed here.

Statistical Approaches to Large-Scale Association Studies

A large number of statistical methods have been proposed to evaluate association between individual markers, haplotypes and chromosomal regions (for examples of each of these approaches, [see Abecasis et al., 2000b; Schaid et al., 2002; Zaykin et al., 2002; Durrant et al., 2004; Lin et al., 2004]). Whatever analytical approach is selected, one important consideration will be evaluating the significance of findings while accounting for a large number of potentially correlated tests.

The standard approach to controlling statistical significance is to define a threshold for the test statistic of interest which is exceeded by chance very rarely and control the total number of false positives per study, even after accounting for testing all available SNPs, haplotypes, or chromosomal regions. This threshold can be calculated by either simulating the null distribution of the data, usually by permuting case and control labels [for an efficient implementation and discussion of related issues, see Dudbridge and Koeleman, 2004] or by using a stringent Bonferroni threshold which becomes increasingly conservative when markers are in linkage disequilibrium. This approach is most appropriate when we expect that a single hypothesis among the set of tested hypothesis is true. This is unlikely to be the case in large scale association studies, since multiple susceptibility alleles are likely to exist for any complex disease and, through linkage disequilibrium, each susceptibility allele can result in genetic association between the trait and multiple SNPs. Instead of searching for a single extreme statistic, there is growing interest in methods that search for a set of unusual results, but less extreme, results. It is likely that these methods will provide more power for complex disease association scans [Wille et al., 2003].

Many of these approaches can be implemented through simulation (for example, to compare expected and actual number of tests to exceed some threshold; or to evaluate the empirical distribution of composite statistics combining information from many markers). We believe that a useful complement to these simulations will be the false discovery rate [FDR, Benjamini and Hochberg, 1995], a computationally inexpensive calibration procedure for multiple testing that has gained popularity in recent years. The method (illustrated in fig. 2) is appropriate when a large number of hypothesis tests are carried out, and we expect to reject the null hypothesis for several of them. Rather than accepting or rejecting individual hypothesis, the method identifies a subset of hypotheses were the null is unlikely to be true, and controls the proportion of false positives within this subset. For example, if we set the false discovery rate to 0.20, and identify 100 associated SNPs in a particular association study, we expect that about 80 of them are truly associated with the trait (directly or through LD) whereas the other 20 might be false positives.

The method has been widely applied in gene-expression studies [see Tusher et al., 2001; Rhodes et al., 2002; Rhodes et al., 2004 for examples]. Work by Genovese and Wasserman [2002] and by Storey [2002] has shown that control of the false discovery rate leads to a gain in power relative to a standard multiple testing adjustment using Bonferroni's correction, which controls the total number of false positives. There is extensive statistical literature on estimation of false discovery rates and related quantities [Storey, 2002; Storey and Tibshirani, 2003], and the original procedure by Benjamini and Hochberg [1995] remains popular. Of crucial importance for genetic studies is recent work [Benjamini and Yekutieli, 2001] showing that the original method, developed under the assumption of independence, can be applied in situations where the results of hypothesis tests are positively correlated (as is the case with nearby SNPs that are in linkage disequilibrium).

We believe that statistical methods based on the FDR could help investigators evaluate interim results of ge-

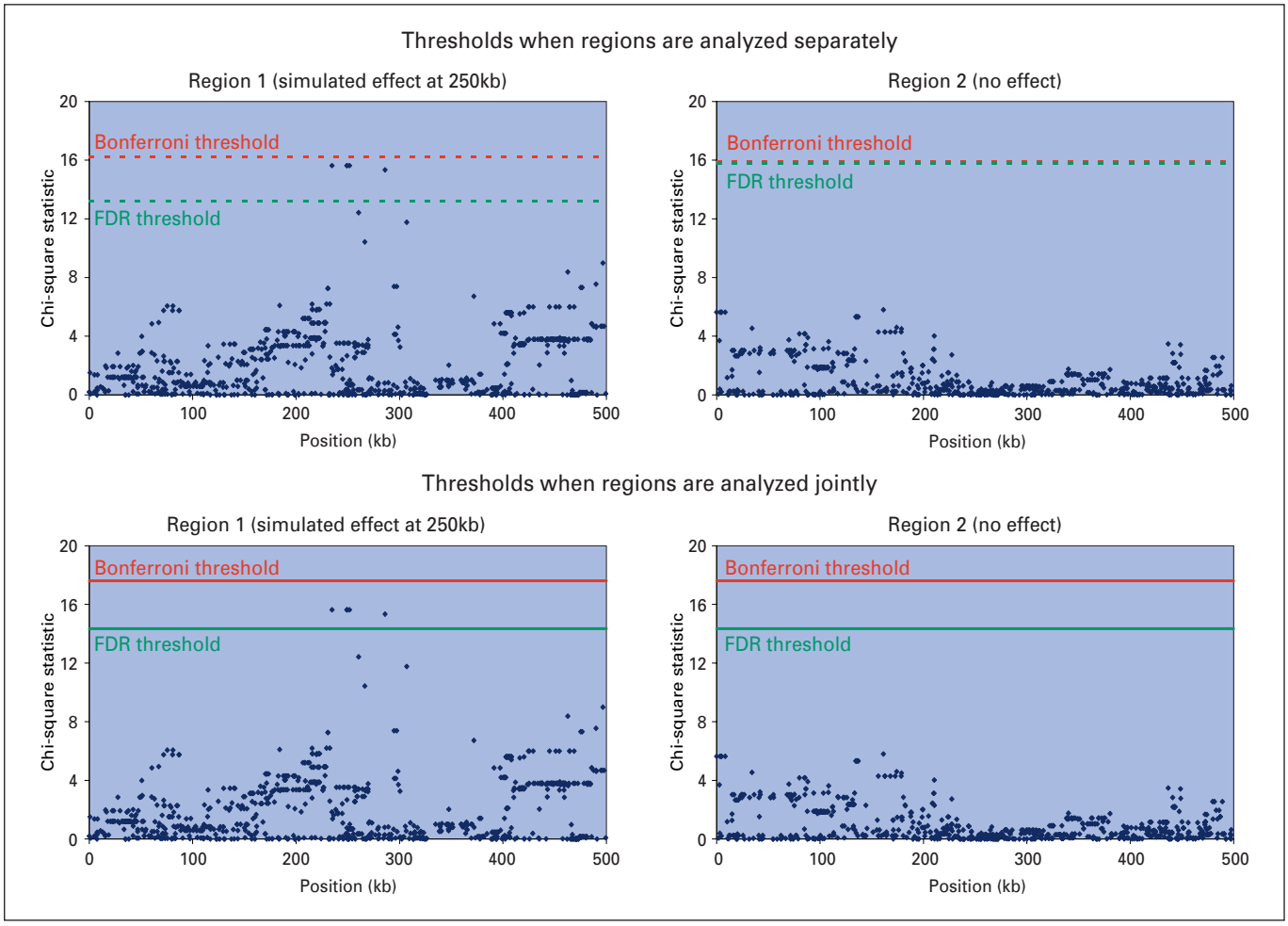

Fig. 2. To illustrate the advantages of the FDR approach, we simulated a simple genetic association study using SNPs. In order to mimic patterns of linkage disequilibrium in real data, publicly available data from the two densely typed regions [The International HapMap Consortium 2003] was used to generate simulated genotypes. The two regions (ENCODE 113 and ENCODE 131) include a total of 1,643 SNPs with minor allele frequency of >5% in the HapMap CEU samples. We simulated a simple quantitative trait whose levels depended on genotypes for marker rs2389380, a SNP approximately in the middle of region 1 (ENCODE 113) and tested for association between the trait and each of the other individual SNP polymorphisms using standard methods [Abecasis et al., 2000a]. Next, we determined significance thresholds for an overall error rate $\alpha = 0.05$ using the Bonferroni and FDR approaches [Benjamini and Hochberg, 1995], either when each region was analyzed separately (top) or for a joint analysis of the two regions (bottom). Notice that, even when the two regions are analyzed separately (top), none of the statistics exceeds the Bonferroni threshold for significance ($\alpha = 0.05/1,642$). The Bonferroni threshold ensures no more than one false positive every 20 studies but it is much too stringent because the test statistics are correlated – that is, although 1,642 SNP polymorphisms were tested, there are many fewer than 1,642 independent tests. The correlation between test statistics is illustrated by the runs of similar test statistics along the graphs. In contrast to the Bonferroni threshold, which depends only on the number of tests performed, the FDR threshold is adaptive and adjusts to the observed results. We set the false discovery rate at $\alpha = 0.05$, ensuring that only 1 in 20 among results exceeding the significance threshold should be a false-positive. In region 1, which shows an excess of large test statistics, the threshold is set at $\chi^2 = 16.42$. This threshold should be exceeded by chance once in ~3,500 tests, but it is exceeded 5 times in the data and all five associated SNPs are within 50 kb of the simulated trait alleles. In contrast, in region 2, there is no excess of large test statistics and the FDR and Bonferroni thresholds are the same. The FDR approach correctly identifies a set of associated SNPs even when we analyze two regions jointly (bottom).

nome-wide association studies before extensive simulations can be carried out. Compared to techniques that define a stringent threshold to control the total number of false positives, the FDR should lead to gains in power that could be important for the success of large-scale genetic association studies. The FDR appraoch has been evaluated in the context of genome-wide association studies by Sabatti et al. [2003] and also Devlin et al. [2003].

Conclusions

Human population history has resulted in substantial, but varying, levels of linkage disequilibrium throughout the genome. Advances in genotyping technology have made possible genetic association studies on a very large scale, and we expect that these will enable lead to the identification of alleles involved in susceptibility to many complex diseases. It is likely that these endeavors will be aided by the incorporation of novel statistical approaches, such as the false-discovery rate approach, in the analysis of gene-mapping data, a process that is already underway.

References

Abecasis GR, Cardon LR, Cookson WOC: A general test of association for quantitative traits in nuclear families. Am J Hum Genet 2000;66: 279–292.

Abecasis GR, Cookson WO, Cardon LR: Pedigree tests of transmission disequilibrium. Eur J Hum Genet 2000;8:545–551.

Abecasis GR, Noguchi E, Heinzmann A, Traherne JA, Bhattacharyya S, Leaves NI, Anderson GG, Zhang Y, Lench NJ, Carey A, Cardon LR, Moffatt MF, Cookson WO: Extent and distribution of linkage disequilibrium in three genomic regions. Am J Hum Genet 2001;68: 191–197.

Benjamini Y, Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc (B) 1995;57:289–300.

Benjamini Y, Yekutieli D: The control of the false discovery rate in multiple testing under dependency. Ann Stat 2001;29:1165–1188.

Botstein D, Risch N: Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 2003;33(suppl):228–237.

Cardon LR, Abecasis GR: Using haplotype blocks to map human complex trait loci. Trends Genet 2003;19:135–140.

Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, Nickerson DA: Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat Genet 2003;33:518–521.

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA: Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 2004;74: 106–120.

Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH: Non-uniform recombination within the human beta-globin gene cluster. Am J Hum Genet 1984;36:1239–1258.

Chakravarti A, Phillips JA, 3rd, Mellits KH, Buetow KH, Seeburg PH: Patterns of polymorphism and linkage disequilibrium suggest independent origins of the human growth hormone gene cluster. Proc Natl Acad Sci USA 1984;81:6085–6089.

Daly MJ, Rioux JD, Schaffner SE, Hudson TJ, Lander ES: High-resolution haplotype structure in the human genome. Nat Genet 2001; 29:229–232.

Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, et al: A linkage disequilibrium map of chromosome 22. Nature 2002;418:544–548.

Devlin B, Roeder K, Wasserman L: False discovery or missed discovery? Heredity 2003;91: 537–538.

Dudbridge F, Koeleman BP: Efficient computation of significance levels for multiple associations in large studies of correlated data, including genomewide association studies. Am J Hum Genet 2004;75:424–435.

Durrant C, Zondervan KT, Cardon LR, Hunt S, Deloukas P, Morris AP: Linkage disequilibrium mapping via cladistic analysis of single-nucleotide polymorphism haplotypes. Am J Hum Genet 2004;75:35–43.

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D: The structure of haplotype blocks in the human genome. Science 2002;296:2225–2229.

Genovese C, Wasserman L: Operating characteristics and extensions of the false discovery rate procedure. J R Stat Soc (B) 2002;64:499–517.

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR: Whole-genome patterns of common DNA variation in three human populations. Science 2005;307: 1072–1079.

Jeffreys AJ, Kauppi L, Neumann R: Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 2001;29:217–222.

Johnson GCL, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RCJ, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SCL, Clayton DG, Todd JA: Haplotype tagging for the identification of common disease genes. Nat Genet 2001;29:233–237.

Jorde LB, Watkins WS, Carlson M, Groden J, Albertsen H, Thliveris A, Leppert M: Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. Am J Hum Genet 1994;54:884–898.

Lin S, Chakravarti A, Cutler DJ: Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. Nat Genet 2004;36:1181–1188.

Martin ER, Lai EH, Gilbert JR, Rogala AR, Afshari AJ, Riley J, Finch KL, Stevens JF, Livak KJ, Slotterbeck BD, Slifer SH, Warren LL, Conneally PM, Schmechel DE, Purvis I, Pericak-Vance MA, Roses AD, Vance JM: SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease. Am J Hum Genet 2000;67: 383–394.

Matsuzaki H, Dong S, Loi H, Di X, Liu G, Hubbell E, Law J, Bernsten T, Chadha M, Hui H, Yang G, Webster T, Cawley S, Walsh P, Jones K, Mei R: Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. Nature Methods 2004; 1:109–111.

McKenzie CA, Abecasis GR, Keavney B, Forrester T, Ratcliffe PJ, Julier C, Connell JM, Bennett F, McFarlane-Anderson N, Lathrop GM, Cardon LR: Trans-ethnic fine mapping of a quantitative trait locus for circulating angiotensin I-converting enzyme (ACE). Hum Mol Genet 2001;10:1077–1084.

McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P: The fine-scale structure of recombination rate variation in the human genome. Science 2004;304:581–584.

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG: Genetic analysis of genome-wide variation in human gene expression. Nature 2004;430:743–747.

Morris AP, Whittaker JC, Balding DJ: Fine-scale mapping of disease loci via shattered coalescent modeling of genealogies. Am J Hum Genet 2002;70:686–707.

Peltonen L, Palotie A, Lange K, Department of Medical Genetics UoH, National Public Health Institute FLmue: Use of population isolates for mapping complex traits. Nature reviews Genetics 2000;1:182–190.

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES: Linkage disequilibrium in the human genome. Nature 2001;411:199–204.

Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM: Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. Cancer Res 2002;62:4427–4433.

Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM: Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci USA 2004;101:9309–9314.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW: Genetic structure of human populations. Science 2002;298:2381–2385.

Sabatti C, Service S, Freimer N: False discovery rate in linkage and association genome screens for complex disorders. Genetics 2003;164:829–833.

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES: Detecting recent positive selection in the human genome from haplotype structure. Nature 2002;419:832–837.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, et al: A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001;409:928–933.

Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA: Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 2002;70:425–434.

Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, Winkler C, et al: Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. Am J Hum Genet 1998;62:1507–1515.

Storey JD: A direct approach to false discovery rates. J R Stat Soc (B) 2002;64:479–498.

Storey JD, Tibshirani R: Statistical significance for genomewide studies. Proc Natl Acad Sci USA 2003;100:9440–9445.

The International HapMap Consortium: The International HapMap Project. Nature 2003;426:789–796.

Tomlinson IP, Bodmer WF: The HLA system and the analysis of multifactorial genetic disease. Trends Genet 1995;11:493–498.

Tusher VG, Tibshirani R, Chu G: Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 2001;98:5116–5121.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ: Y chromosome sequence variation and the history of human populations. Nat Genet 2000;26:358–361.

Wille A, Hoh J, Ott J: Sum statistics for the joint detection of multiple disease loci in case-control association studies with SNP markers. Genet Epidemiol 2003;25:350–359.

Wootton JC, Feng X, Ferdig MT, Cooper RA, Mu J, Baruch DI, Magill AJ, Su XZ: Genetic diversity and chloroquine selective sweeps in Plasmodium falciparum. Nature 2002;418:320–323.

Zabetian CP, Anderson GM, Buxbaum SG, Elston RC, Ichinose H, Nagatsu T, Kim KS, Kim CH, Malison RT, Gelernter J, Cubells JF: 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.

Zabetian CP, Buxbaum SG, Elston RC, Kohnke MD, Anderson GM, Gelernter J, Cubells JF: The structure of linkage disequilibrium at the DBH locus strongly influences the magnitude of association between diallelic markers and plasma dopamine beta-hydroxylase activity. Am J Hum Genet 2003;72:1389–1400.

Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG: Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 2002;53:79–91.

Zhang W, Collins A, Maniatis N, Tapper W, Morton NE: Properties of linkage disequilibrium (LD) maps. Proc Natl Acad Sci USA 2002;99:17004–17007.

Zhu X, Luke A, Cooper RS, Quertermous T, Hanis C, Mosley T, Gu CC, Tang H, Rao DC, Risch N, Weder A: Admixture mapping for hypertension loci with genome-scan markers. Nat Genet 2005;37:177–181.

Zollner S, Pritchard JK: Coalescent-based association mapping and fine mapping of complex trait loci. Genetics 2005;169:1071–1092.