Hypothesis/Commentary

Perils in the Use of Linkage Disequilibrium for Fine Gene Mapping: Simple Insights from Population Genetics

Prakash Gorroochurn

Division of Statistical Genetics, Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, New York

Abstract

It is generally believed that genome-wide association (GWA) studies stand a good chance for finding susceptibility genes for common complex diseases. Although the results thus far have been somewhat promising, there are still many inherent difficulties and many initial associations do not get replicated. The common strategy in GWA studies has been that of selecting the most statistically significant single nucleotide polymorphisms with the hope that these will be very physically close to causal variants because of strong linkage disequilibrium (LD). Using simple ideas from population genetics, this commentary explains why this strategy can be misleading. It argues that there is an intrinsic problem in the way LD is currently used for fine-mapping. This is because most of the metrics that are currently used to measure LD are inadequate, as they do not take into account evolutionary variables that shape the LD structure of the human genome. Recent research on another metric, based on Malécot's model for isolation by distance, holds considerable promise for GWA studies and merits more serious consideration by geneticists. (Cancer Epidemiol Biomarkers Prev 2008;17(12):3292–7)

Introduction

Genome-wide association (GWA) studies are well under way and, in their first couple of years, have identified variants for several complex diseases, including at least four types of cancer (1). This seems promising, although it remains to be seen how many of these will be consistently replicated in the future. Once a GWA study is done, there are usually two types of subsequent association studies (2): those aimed at replicating the original significant markers (the so-called ''exact'' approach) and those aimed at fine-mapping other loci in the surrounding region in addition to the original significant loci (the so-called ''local'' approach). Both the original GWA and any follow-up association study depend crucially on the concept of linkage disequilibrium (LD; also known as gametic disequilibrium, allelic association, or two-locus Hardy-Weinberg disequilibrium). It is unsurprising therefore that, well before GWA studies were under way, a vital task was to elucidate the LD structure of the human genome. This structure could shed considerable light on the prospects and limitations of LD mapping, especially for GWA.

The indiscriminate use of LD can, however, lead to inaccurate results, especially when doing fine-mapping. Association studies are usually criticized on design and methodologic grounds (3-6) while underestimating the difficulties inherent in the conventional use of LD. By using simple population genetics ideas, this commentary aims to enumerate specific instances where the results of LD fine-mapping can be misinterpreted.

Of the several theoretical studies on LD, an important simulation-based work is that of Kruglyak (7). Kruglyak argued that LD between the common variant and other markers was essentially short-ranged and was unlikely to extend beyond a genetic distance of 3 kb. A more recent study (8) found that LD could extend to genetic distances of 1 Mb. Empirical studies have shown that LD is highly ''patchy'' in nature and does not decrease monotonically with physical distance. Chromosomes tend to contain blocks of long-range LD (which is different from Kruglyak's predictions) separated by small region of recombination hotspots (9-11). LD is high within blocks but low between blocks.

Insights from Population Genetics

Insight 1: Population History Is a Key Determinant of LD in Small DNA Regions and There Is Usually No Correlation between LD and Physical Distance when Doing Fine-Mapping. The cystic fibrosis gene was found by climbing up a gradient until a peak was reached. However, when the same principles are applied to the Huntington's disease (HD) gene, something unexpected happens (see Fig. 1).

The HD gene shows weak LD to some physically close markers (e.g., D4S180) but strong LD to a more physically distant marker (D4S98). Thus, if both markers were tested for association with disease, one (D4S98) would show strong association through LD with the HD

Received 8/4/08; revised 8/29/08; accepted 9/15/08.

Grant support: National Institute of Mental Health grant MH-48858.

Requests for reprints: Prakash Gorroochurn, Division of Statistical Genetics, Department of Biostatistics, Mailman School of Public Health, Columbia University, Room 620, 722 West 168th Street, New York, NY 10032. Phone: 212-342-1263; Fax: 212-342-0484. E-mail: pg2113@columbia.edu

doi:10.1158/1055-9965.EPI-08-0717

gene, whereas the other (D4S180) would show weak association. This could mislead many to believe that the HD gene was physically much closer to the D4S98 locus than to the D4S180 locus when in fact the opposite is true.

Is such erratic LD behavior surprising? Not when looked from a population genetics perspective. Indeed, the noncorrelation between LD and physical distance is the norm rather than the exception when dealing with small DNA regions (e.g., <100 kb; ref. 12). To understand why, consider the simple equation for LD:

$$D_t = D_0(1-\theta)^t, \quad (1)$$

where D_t is the LD after t generations, D_0 is the initial LD, and θ is the recombination fraction. (D_t is actually the deviation at generation t from random association of two-locus haplotype frequencies.) When θ is small (e.g., $\theta < 10^{-3}$), $D_t \approx D_0$, the value of LD is primarily determined by its initial value (which is a function of haplotype and allele frequencies) and does not decay with either physical distance or time. Different markers at various short distances from a given markers will have different values of D_0 depending on the evolutionary history of the population.

For example, consider three biallelic loci A, B, and C across a small DNA region (e.g., <100 kb), with B closer than C to A (see Fig. 2). Suppose A has both *A* and *a* alleles, but B and C have homozygotes *B/B* and *C/C*, respectively. We shall consider the normalized LD measure D' (which equals D divided by its theoretical maximum value), because it is more indicative of the relative strength of LD. Because there is no allelic variation at B and C, $D' = 0$ for alleles at A and B and for alleles at A and C. Suppose now an initial LD is created between alleles *A* and *B* through gene flow from several founders carrying copies of *b*. At about the same time in the past, an initial LD is created between alleles *A* and C through a single mutation of a *C* allele to *c*. Now, several copies of *b* are introduced with some on the same chromosome as *A* and others on the same chromosome as *a*, whereas the single copy of *c* will always be associated with *A* (or *a* depending on which chromosome the mutation occurred). Therefore, the initial LD D_0' will be larger in magnitude for *AC* than for *AB*, although locus C is further away from A than B is. Thus, in general, D_t' is no indicator of the relative proximity of a locus. When doing fine-mapping in small DNA regions, there is usually no correlation between the magnitude of LD and physical distance (or recombination fraction): the fact that a particular marker has a very high disease association does not guarantee it must be physically close to the causative locus. Strachan and Read (ref. 13, p. 448) provide an accurate description of the issue at hand: ''. . .the patchy nature of LD, with some long-range correlations coexisting with short-range lack of correlation means that despite the theoretical high resolution of association studies, in reality one can seldom be confident that one is looking in exactly the right place for the susceptibility determinant.''

Even for moderate to large DNA regions (~1-10 Mb), the use of D_t (or D_t') can still be problematic. In that case, the value of θ in Eq. (1) is no longer inconsequential, and both D_0 and $(1 - \theta)^t$ influence the value of D_t. A locus L relatively far from a given locus could give a large value of D_t (or D_t') if mutant alleles at L originated very recently back in time or if L is located in a recombination cold spot (14). Another locus could be relatively close to the given locus and result in small values of D_t (or D_t').

Insight 2: Genetic Drift Generates LD between Disease and Marker Alleles in a Random Fashion, so a Particular Disease-Marker Association Found in One Population Need Not Exist in Another Population or Might Even Be in the Opposite Direction. Templeton (12) provides an example for two linked loci. Both the human populations living in southern Italy west of the Apennine mountain range and on the nearby island of Sardinia were formed from a founder effect. In the Italian population, all *Med1* G6PD-deficient males had red/green color blindness, whereas in the Sardinian population almost none of the *Med1* G6PD-deficient males had red/green color blindness. Thus, in both cases, *Med1* G6PD-deficient males are strongly associated with red/green color blindness but in complete opposite directions. Furthermore, if we had selected cases with red/green color blindness by sampling from both Italian and Sardinian populations, it is very likely that no association whatsoever would be found.

All of this makes sense because populations with small effective sizes undergo extensive amounts of genetic drift. Genetic drift is the random fluctuation of allele frequencies across generations because of finite population size. Let the mutant allele at the G6PD locus be denoted by *a* and the allele for color blindness by *b*, with both alleles having relatively small frequencies. Let the wild-type alleles at each locus be *A* and *B*, respectively. Now, for these two loci, we have four haplotypes, with frequencies even smaller than those of the corresponding alleles. The combination of small effective population size and rare haplotypes means that each one is subject to considerable random fluctuation, so that any one of them could increase in frequency. In our example, the frequency of the *ab* haplotype increased in the Italian population, whereas the frequency of the *Ab* haplotype increased in the Sardinian population. Therefore, because genetic drift is a stochastic phenomenon, it can create an association between the alleles at a disease locus and a marker locus in a completely random fashion. This explains the completely opposite associations in the two cases. Furthermore, when the two populations are pooled, the LD between the alleles at the loci is destroyed because now, in the total population, the *b* allele can occur on the same chromosome as either an *a* or an *A* allele.

The haphazard effect of genetic drift on LD raises questions on the legitimacy of replicating a significant association found in one population in a different population. Emphasis has often been placed on replicating initial disease-marker associations. However, if we are willing to accept that LD is generally population-specific, then failure to replicate using different populations should be no reason for alarm. Thus, if an initial association is found in one population but no association in a different population, then it is not necessary for the first association to be a false-positive. *Both results* can be equally valid because they are specific for the populations in which they were done. Whereas this point about the specificity of LD has been well taken by some (3, 15-17), it is perhaps not universally appreciated.

Insight 3: The Act of Combining Two or More Genetically Distinct Subpopulations Creates LD between Any Two Alleles (at Different Loci) with Different Allele Frequencies Even when the Loci Are Unlinked. A well-known example is reported in Knowler et al. (18), whereby a strong negative association was found between non-insulin-dependent diabetes and a haplotype at the IgG locus in the Pima and Papago tribes. When the study was repeated by stratifying on reported ancestry, the initial association vanished.

This a classic case of confounding by population stratification, which arises when cases and controls are sampled from genetically distinct subpopulations or when they are sampled from a single population made up of genetically distinct subgroups (19). In fact, genetic drift is one of the major evolutionary forces that can create differentiation among subpopulations. There have been several recent methods to combat PS in association studies (20-23), with the principal components method of Price et al. (22) being especially suited for GWA studies. PS gives rise to a spurious disease-marker association even if a true causal gene never existed (even if the disease was nongenetic). In the example above, the association is an artifact of the controls having a higher proportion of European ancestry than the cases. In Gorroochurn et al. (24), we gave the necessary and sufficient condition for such false disease-marker associations (due to PS) to arise.

$$\sum_{i=1}^{K} \pi_i d_i r_i - \sum_{i=1}^{K} \pi_i d_i \sum_{i=1}^{K} \pi_i r_i = 0, \qquad (2)$$

where π_i, d_i, and r_i are, respectively, the relative subpopulation size, the disease prevalence, and the allele frequency (for *any* allele) in subpopulation i, and K is the number of discrete subpopulations. From Eq. (2), consider for example two subpopulations of the same size. PS would cause the null hypothesis of no disease-marker association to be always false, for *any* marker allele as long its frequency is different in the two subpopulations, and the disease prevalences are also different. There is no necessity for a true causal gene to exist.

We can also consider the situation when a true causal marker exists. Let M be any marker allele at a locus and N be an *unassociated* disease allele at a different locus in each of, say, two genetically distinct subpopulations. Let the frequency of M in subpopulation $i (i = 1,2)$ be r_i and that of N be S_i, where $r_1 \neq r_2$ and $s_1 \neq s_2$. Then, when the subpopulations are combined, LD is created between the two alleles with variable (12)

$$D_{\text{comb}} = \pi(1-\pi)(r_1 - r_2)(s_1 - s_2), \qquad (3)$$

where π is the size of subpopulation 1 relative to subpopulation 2. Because of this (negative) disequilibrium in our example, the cases will be overrepresented in terms of the N allele, with the reverse being true in the controls, thus implying a disease-marker association. However, such an association provides no indication as to the *physical proximity* of the true causal locus, which could potentially be ~100 Mb away or even on different chromosomes.

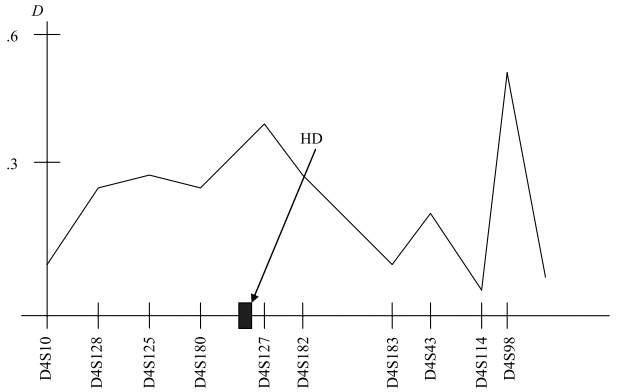

Figure 1. HD locus (total DNA stretch is 2,500 kb). Adapted from Krawczak and Schmidtke (47).

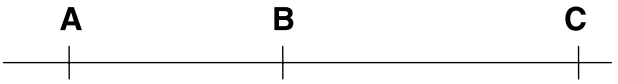

Figure 2. At locus A, alleles are both of the A and a types. At locus B, all alleles are of the B type and several b alleles are introduced by gene flow. At locus C, all alleles are of the C type and a single C allele mutates to c.

The LD in Eq. (3) is an example of spurious LD, because the disequilibrium created is not the result of linkage (and physical proximity), but other population artifacts, in this case gene flow between two genetically distinct subpopulations. Indeed, when any of the conditions (random mating, infinite population size, no selection, no mutation, and no gene flow) necessary for one-locus HWE to hold are violated, two-locus HWE is also distorted and spurious LD can be created (unless the LD is due to close physical proximity).

Insight 4: When Two or More Genetically Distinct Subpopulations Are Admixed, Assortative Mating Helps Maintain Spurious LD in the Total Population. A typical example is the U.S. population. Early colonization of North America resulted in the admixture of European and sub-Saharan African populations that had allele frequency differences at numerous loci. Along with admixture between these two subpopulations for centuries, there also exists strong assortative mating based on skin color (12). A key observation here is that, in spite of the long history of admixture between these two subpopulations, significant allele frequency differences are maintained at several loci both for loci influencing skin color and for other loci having no effect on skin color.

This is because when genetically distinct populations are combined, assortative mating ensures that initial allelic differences are maintained for genes influencing the genetic trait. Suppose allele A at locus L_1 influences a trait on which assortative mating occurs and suppose allele B at locus L_2 has no such influence. If alleles A and B each initially had different frequencies in the African and European subpopulations, Eq. (3) predicts that LD is created between A and B. Because of this LD, assortative mating on the trait influenced by A causes the differences in B to be also maintained, although B has no effect on the trait whatsoever. For example, the African and

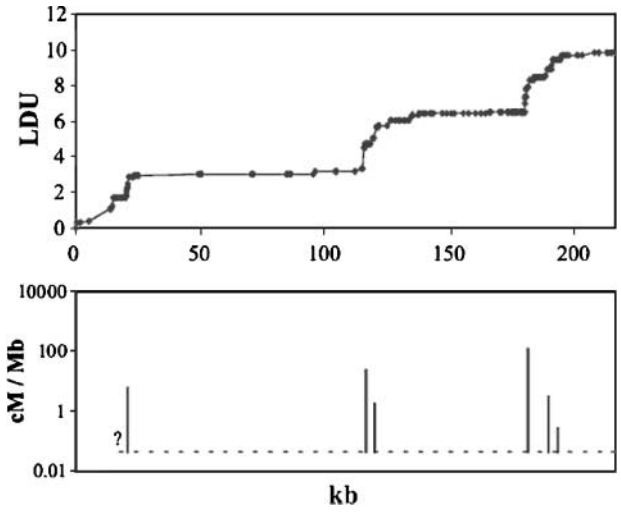

Figure 3. *Top,* LDU map for a 216-kb segment of class II of MHC. *Bottom,* corresponding recombination hotspots. From Jeffreys et al. (9).

European populations are characterized by differences at several blood group loci, although there is no tendency for nonrandom mating to occur based on blood group. The differences are due to initial differences being maintained because of LD between blood group genes and genes for skin color. The LD created between alleles *A* and *B* above has little chance of being broken due to assortative mating. This explains the potential for extensive amounts of LD to be maintained in such a population.

The interaction of admixture with assortative mating adds further wrinkles to LD fine-mapping. The sustained allele frequency differences at several loci, in spite of generations of admixture, implies that a case-control association study done from the total population could be considerably prone to PS, thus leading to many false-positives. Moreover, the extensive LD created because of allele frequency differences is again a case of spurious LD. We also note that, even for homogenous populations, Redden and Allison (25) have shown that assortative mating can lead to spurious associations.

LD Mapping Using Population Genetics Theory: The Malécot's Model

Various evolutionary forces can create strong LD between alleles at loci that can be physically far away from each other, and the use of such LD for fine gene mapping can be misleading. Selecting a candidate marker that is highly associated with a disease usually implies that the disease marker is in strong LD with the candidate marker, but this fact in itself gives no indication as to the physical whereabouts of the disease locus and even as to the latter's very existence!

The problem arises because of our conceptualization of the association (or LD) between alleles at two loci, that is, as the correlation r between the alleles or a function of the correlation. Almost all pairwise metrics (such as D, D', and r^2) that are conventionally used to measure LD are functions of r and vanish when $r = 0$. These metrics do not explicitly take into account the various evolutionary forces that shape the current LD structure. Instead, they are functions of only the current haplotype and allele frequencies. Because they ignore population history, these LD metrics are unable to capture the extensive variations of the actual LD across the human genome, making them prone to the very problems we have outlined in Insights from Population Genetics. The following concluding remarks from Lewontin (26) are quite sobering: ''...we *may* be able to find a measure of association that is preserved under particular conditions, but the search for a 'pure' measure of gametic disequilibrium is doomed to failure.''

However, Collins, Morton, and others (27-34) have recently put forward a LD metric (known as the association probability, ρ_x) based on population genetics theory, more specifically on Malécot's model for isolation by distance (35, 36):

$$\rho_x = (1 - L)Me^{-\epsilon x} + L \qquad (4)$$

In the above, the variable L is spurious LD (any LD that is not due to linkage), ε (>0) is a constant for each interval between two loci, x is the physical distance between the two loci, and M is the association probability at $x = 0$. Morton et al. (28) have shown that ρ_x is the most efficient for modeling LD as a function of distance compared with other commonly used metrics. The Malécot's model has a major advantage in that it incorporates the various evolutionary forces that shape the LD structure in a particular population. More specifically, the variable M depends on whether the susceptibility gene has a monophyletic or polyphyletic origin; it is equal to 1 if there is a unique susceptibility gene (as in Mendelian diseases) and it is <1 otherwise (as in complex diseases). The variable ε is a function of the number of generations for the equilibrium probability in Eq. (4) to be achieved and of various evolutionary forces such as mutation, selection, and recombination (27). By using composite likelihoods (37), the Malécot's model is fitted to each marker of interest in a particular DNA region. Each marker has its own ε estimate, but there are single values of L and M for the given region (38). A LD map is then built with additive distances in terms of LD units (LDU). One LDU is defined as one swept radius and corresponds to $\varepsilon x = 1$; it is the physical distance over which ''useful LD'' [the first term in the right in Eq. (4)] decays to $1/e \approx 37\%$ of its starting value. Figure 3 gives an example of a LDU map for a 216-kb segment of class II region of MHC, with corresponding hot spots of recombination (9). Because they are defined in terms of εx, the LDUs are negatively correlated with LD and positively correlated with recombination. More technical details on the construction of LD maps can be found in Maniatis et al. (39) and a discussion of optimal statistical properties of ρ_x can be found in Shete (40).

Eq. (4) indicates that the LD between markers at two loci is decomposed into two components: the ''true'' LD (due to linkage) and the spurious LD (due to population artifacts). However, only the ''true'' LD contributes to the LD map. Thus, two markers physically far apart could be in high LD but only a few LDUs from each other if the

association was due to population artifacts. If we were to rely on D (or D', r^2, etc.), these would be reasonably large and wrongly suggest that the markers were physically close; however, the low value for LDU would correctly indicate that the two markers were indeed physically far apart in spite of being strongly associated! This gives an example of how the use of LD maps can directly address the problems associated with the conventional methods used for LD fine-mapping as explained in Insights from Population Genetics.

LD maps are thus better able to physically locate genes of interest and thus less prone to most of the problems mentioned in Insights from Population Genetics. The LD maps built by Maniatis et al. (33) and others (e.g., refs. 41-43) support this statement. For example, by using standard LD mapping techniques, Hosking et al. (44) located the CYP2D6 gene (which is associated with poor drug-metabolizing activity) to within a DNA interval of 390 kb. Using LDU locations, Maniatis et al. (33) were able to predict the location of the gene only 14.9 kb from its true location, surrounded by a 95% confidence interval of 172 kb. LD maps can also be constructed on genome-wide scales for GWA studies through selected map assembly from DNA segments (39). Kuo et al. point out that LD maps are achievable even at the highest marker densities, including the HapMap data with >3 million single nucleotide polymorphisms.

Conclusion

Understanding the major limitations of the current use of LD for fine-mapping is essential lest the technique will be regarded with skepticism and distrust by many. Moreover, recent research on metrics based on population genetics theory represents a positive step toward alleviating some of the current disenchantment with association studies. In spite of the advantages offered by the Malécot's model and LD maps, the more popular strategy in GWA studies has been simply the selection of the statistically most significant single nucleotide polymorphisms (45, 46). However, selecting significant single nucleotide polymorphisms and then trying to find the causal gene in a neighboring region based on strong LD will likely lead to many mapping problems in the future.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Drs. S.E. Hodge, G.A. Heiman, R. Fan, M. Durner, and W.J. Ewens for helpful comments and the two anonymous reviewers for useful suggestions.

References

1. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 2008;118:1590–605.
2. Clarke GM, Carter KW, Palmer LJ, Morris AP, Cardon LR. Fine mapping versus replication in whole-genome association studies. Am J Hum Genet 2007;81:995–1005.
3. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. Genet Med 2002;4:45–61.
4. Ott J. Association of genetic loci: replication or not, that is the question. Neurology 2004;63:955–8.
5. Cardon LR, Bell JI. Association study designs for complex diseases. Nat Rev Genet 2001;2:91–9.
6. Gorroochurn P, Hodge SE, Heiman GA, Durner M, Greenberg DA. Non-replication of association studies: ''pseudo-failures'' to replicate? Genet Med 2007;9:325–31.
7. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 1999;22:139–44.
8. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. Am J Hum Genet 2001;69:1–14.
9. Jeffreys AJ, Kauppi L, Neumann R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 2001;29:217–22.
10. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nat Genet 2001;29:229–32.
11. Petes TD. Meiotic recombination hot spots and cold spots. Nat Rev Genet 2001;2:360–9.
12. Templeton AR. Population genetics and microevolutionary theory. John Wiley & Sons; 2006.
13. Strachan T, Read A. Human molecular genetics. KY: Garland Science/Taylor & Francis Group; 2004.
14. Jobling MA, Hurles ME, Tyler-Smith C. Human evolutionary genetics. New York: Garland Science; 2003.
15. Vieland VJ. The replication requirement. Nat Genet 2001;29:244–5.
16. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. Nat Genet 2001;29:306–9.
17. Palmer LJ, Cardon LR. Shaking the tree: mapping complex disease genes with linkage disequilibrium. Lancet 2005;366:1223–34.
18. Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. Am J Hum Genet 1988;43:520–6.
19. Gorroochurn P, Hodge SE, Heiman GA, Greenberg DA. A unified approach for quantifying, testing and correcting population stratification in case-control association studies. Hum Hered 2007;64:149–59.
20. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997–1004.
21. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. Am J Hum Genet 2000;67:170–81.
22. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.
23. Gorroochurn P, Heiman GA, Hodge SE, Greenberg DA. Centralizing the non-central chi-square: a new method to correct for population stratification in genetic case-control association studies. Genet Epidemiol 2006;30:277–89.
24. Gorroochurn P, Hodge SE, Heiman G, Greenberg DA. Effect of population stratification on case-control association studies. II. False-positive rates and their limiting behavior as number of subpopulations increases. Hum Hered 2004;58:40–8.
25. Redden DT, Allison DB. The effect of assortative mating upon genetic association studies: spurious associations and population substructure in the absence of admixture. Behav Genet 2006;36:678–86.
26. Lewontin RC. On measures of gametic disequilibrium. Genetics 1988;120:849–52.
27. Collins A, Morton NE. Mapping a disease locus by allelic association. Proc Natl Acad Sci U S A 1998;95:1741–5.
28. Morton NE, Zhang W, Taillon-Miller P, Ennis S, Kwok PY, Collins A. The optimal measure of allelic association. Proc Natl Acad Sci U S A 2001;98:5217–21.
29. Lonjou C, Collins A, Morton NE. Allelic association between marker loci. Proc Natl Acad Sci U S A 1999;96:1621–6.
30. Zhang W, Collins A, Maniatis N, Tapper W, Morton NE. Properties of linkage disequilibrium (LD) maps. Proc Natl Acad Sci U S A 2002;99:17004–7.
31. Maniatis N, Collins A, Gibson J, Zhang W, Tapper W, Morton NE. Positional cloning by linkage disequilibrium. Am J Hum Genet 2004;74:846–55.
32. Collins A, Lau W, De La Vega F. Mapping genes for common diseases: the case for genetic (LD) maps. Hum Hered 2004;58:2–9.
33. Maniatis N, Morton NE, Gibson J, Xu CF, Hosking LK, Collins A. The optimal measure of linkage disequilibrium reduces error in association mapping of affection status. Hum Mol Genet 2005;14:145–53.

34. Morton NE. Linkage disequilibrium maps and association mapping. J Clin Invest 2005;115:1425–30.
35. Malécot G. Les Mathématiques de l'Hérédité. Paris: Masson et Cie; 1948.
36. Malécot G. The mathematics of heredity. San Francisco: Freeman; 1969.
37. Devlin B, Risch N, Roeder K. Disequilibrium mapping: composite likelihood for pairwise disequilibrium. Genomics 1996;36:1–16.
38. Tapper W. Linkage disequilibrium maps and location databases. In: Collins AR(ed). Linkage Disequilibrium and Association Mapping, pp 23–45. New Jersey: Humana Press; 2007.
39. Kuo T-Y, Lau W, Collins AR. LDMAP. In: Collins AR(ed). Linkage Disequilibrium and Association Mapping, pp 47–57. New Jersey: Humana Press; 2007.
40. Shete S. A note on the optimal measure of allelic association. Ann Hum Genet 2003;67:189–91.
41. Tapper WJ, Maniatis N, Morton NE, Collins A. A metric linkage disequilibrium map of a human chromosome. Ann Hum Genet 2003; 67:487–94.
42. Tapper W, Collins A, Gibson J, Maniatis N, Ennis S, Morton NE. A map of the human genome in linkage disequilibrium units. Proc Natl Acad Sci U S A 2005;102:11835–9.
43. De La Vega F, Isaac H, Collins A, et al. The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. Genome Res 2005;15:454–62.
44. Hosking LK, Boyd PR, Xu CF, et al. Linkage disequilibrium mapping identifies a 390 kb region associated with CYP2D6 poor drug metabolising activity. Pharmacogenomics J 2002;2: 165–75.
45. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308: 385–9.
46. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–78.
47. Krawczak M, Schmidtke J. DNA fingerprinting. Oxford: BIOS Scientific Publishers; 1998.