Haplotype Thinking in Lung Disease

Edwin K. Silverman

Channing Laboratory and Pulmonary and Critical Care Division, Brigham and Women's Hospital, Boston, Massachusetts

To identify the genetic etiology of a disease of interest, disease-related characteristics (phenotypes) are often tested for association with genetic variants (genotypes). Although genetic association studies of single genetic variants have been widely performed, there has been increasing interest in studies of multiple adjacent genetic variants on one chromosome, known as a haplotype. In this review, we will provide background about the origin of haplotypes and why they can be useful in genetic studies; we will discuss approaches to determining haplotypes and performing haplotype-based genetic association studies; and we will compare single variant and haplotype-based approaches.

Keywords: single nucleotide polymorphism; haplotype; genetic association analysis; linkage disequilibrium

HUMAN GENOME STRUCTURE AND HAPLOTYPES

A haplotype is the series of genetic variants on one chromosome that are inherited from one parent. In subsequent generations, the chromosomal haplotype is progressively broken up by crossing over events in meiosis. In practice, the term "haplotype" usually refers to closely linked genetic loci. The most common type of genetic variant, or polymorphism, is the single nucleotide polymorphism (SNP); an SNP is a location in the genome in which more than one nucleotide base (allele) is commonly observed in a population. SNPs that are located in close proximity tend to travel together—a phenomenon that is known as linkage disequilibrium (LD). In general, loci that are located more closely together on a chromosome will be in stronger LD than those loci that are far apart, but the correlation between LD and the physical distance separating two loci is modest: some loci that are separated by 20 bp will not be in LD, whereas other loci separated by 200,000 nucleotide bases will be in tight LD (1).

The recombination events that shuffle the components of a haplotype do not occur at random. Some locations in the genome have much higher recombination rates than others; these locations are often referred to as recombination hotspots. The occurrence of recombination hotspots has contributed to the limited haplotype diversity of much of the genome: even in a population of unrelated individuals, there are many fewer observed haplotypes in a genomic region than one would expect by random assortment. For example, let us consider the β_2 -adrenergic receptor gene, known as ADRB2, which is one of the most widely studied genes in airway disease genetics. Within the sequence of this intronless gene, there are multiple genetic variants; our research group has focused on eight SNPs in ADRB2, as shown

in Figure 1. At each of these SNPs, there are two alternate alleles that appear in the Childhood Asthma Management Program (CAMP) study population. Based on random assortment of these eight genetic variants, we would expect to observe $2^8 = 256$ different haplotypes. However, a much smaller number of haplotypes is typically observed. In CAMP, only three haplotypes were estimated to occur with frequencies above 5%.

The occurrence of recombination hotspots has generated a blocklike pattern to the LD structure of much of the human genome (2). In addition to recombination hotspots, additional population genetic factors, such as genetic drift and population subdivision, can produce a blocklike pattern of genetic variation (3). A recently developed method for estimating recombination rate variation using population genetic models of coalescent processes demonstrated substantial regional variation; in white European populations, approximately 50% of the recombination events occur in 10% of the genome (4). In addition, recombination events occur less frequently within genes than outside of genes.

The sizes of the "haplotype blocks," as they are referred to, vary greatly throughout the genome, as well as among ethnic groups. Populations of African origin, which are evolutionarily older than white and Asian populations, tend to have narrower haplotype blocks—with mean haplotype block sizes of approximately 9 kb (2). White and Asian populations have mean haplotype block sizes of approximately 18 kb. However, a substantial portion of the genome in each of these populations is in large blocks (> 100 kb), whereas other parts of the genome do not have a very blocklike structure.

The recently completed HapMap project used this blocklike structure of the genome to create a road map of genetic variation in three major ethnic groups: whites, Africans, and Asians. More than 4 million SNPs have been identified and genotyped in a modest number of individuals from these three groups, providing an atlas of haplotype block structure in the genome (5). This invaluable tool will assist in SNP selection for genetic association studies and provide guidance regarding localization of genetic association signals. However, the blocklike structure of much of the human genome structure also influences studies attempting to localize genetic determinants of a disease. When a series of genetic variants is highly correlated within a haplotype block, it is quite difficult to determine which of those variants is a functional variant in disease causation; testing other ethnic groups or functional studies are typically required to narrow the set of possible functional variants.

WHY CAN HAPLOTYPES BE USEFUL IN GENETIC STUDIES?

A new disease-causing mutation occurs on one particular haplotype; because this mutation is transmitted through the population, increased sharing of the haplotypic region around that mutational site will persist among individuals inheriting that mutation. Observing excess haplotype sharing among affected individuals has proven to be a useful approach for susceptibility gene localization in classic monogenic disorders, such as cystic fibrosis. The CFTR gene was localized as the causative locus for cystic fibrosis after genetic linkage analysis by observing

(Received in original form July 15, 2006; accepted in final form July 25, 2006)

Supported by R01 HL68926, R01 HL075478, and an American Lung Association Career Investigator Award.

Correspondence and requests for reprints should be addressed to Edwin K. Silverman, M.D., Ph.D., 181 Longwood Avenue, Boston, MA 02115. E-mail: ed.silverman@channing.harvard.edu

Proc Am Thorac Soc Vol 4, pp 4–8, 2007

DOI: 10.1513/pats.200607-145JG

Internet address: www.atsjournals.org

	SNP -709	SNP -654	SNP -47	SNP +46	SNP +79	SNP +252	SNP +491	SNP +523	Haplotype Frequency
Haplo 1	C	G	C	G	G	G	C	C	36.3%
Haplo 2	C	A	T	A	C	G	C	C	35.4%
Haplo 3	C	G	T	G	C	A	C	A	18.2%

Figure 1. Common haplotypes in the ADRB2 gene. In a study of the β_2 -adrenergic receptor gene (ADRB2), Silverman and colleagues estimated the haplotype frequencies in the Childhood Asthma Management Program using the E-M algorithm as implemented in the TRANSMIT program (21). Although eight SNPs were genotyped, with $2^8 = 256$ possible haplotypes, only three haplotypes occurred at frequencies above 5% (17).

haplotype sharing among affected individuals (6). In such monogenic disorders, haplotypes are often more informative about the parental origins of a particular genomic region than are individual SNPs. Overlap between haplotypes among affected individuals can identify a minimal shared region that likely contains a key disease genetic variant. When there are multiple mutational origins for disease-causing loci, the effects of haplotype sharing will be diluted.

Haplotypes provide a record of evolutionary history more accurately than individual SNPs, and they can capture the LD patterns of a genomic region more completely. Therefore, they may enable susceptibility gene identification in complex diseases, like asthma and chronic obstructive pulmonary disease (COPD), more effectively than individual SNPs. Haplotype-based association analysis can be used in studies of candidate genes or narrow genomic regions, in which minimal recombination is required, or in fine mapping studies of larger genomic regions, in which differences in recombination history between affected and unaffected individuals are used to localize key genetic loci (7). We focus primarily on candidate gene studies in this review.

Because SNPs comprising a haplotype tend to be inherited together, only a subset of those SNPs needs to be genotyped to provide the same amount of genetic information for genetic association studies. Sets of haplotype-tagging SNPs can be identified through a variety of algorithms—for example, the Best Enumeration of SNP Tags (BEST) algorithm developed by Sebastiani and colleagues (8). Haplotype-tagging SNPs should be distinguished from LD-tagging SNPs, which capture the LD information at an individual SNP level above a specified LD threshold of genomic coverage, using approaches such as the LD-Select algorithm developed by Carlson and colleagues (9). If one plans to perform genetic association analysis using haplotype-based methods (discussed *below*), then a haplotype-tagging approach is likely optimal; however, if individual SNP association analysis will be used, LD-tagging approaches will likely provide improved genomic coverage. Haplotype-tagging and LD-tagging SNPs may overlap, but they will not typically provide identical sets of SNPs.

A largely theoretical advantage of studying haplotypes is that the genetic variants on a particular haplotype may confer a unique phenotype when they occur together; for example, two genetic variants that both alter amino acid sequence and affect protein function could have a different functional effect if they appear together in the same transcribed messenger RNA and translated protein sequence. Key combinations of adjacent SNPs could be required to confer a particular phenotype; haplotype analysis may allow identification of such key SNP combinations. One potential example in respiratory disease relates to the ADRB2 gene in asthma; Drysdale and colleagues suggested that ADRB2 haplotypes had differential effects on bronchodilator responsiveness among subjects with asthma, which was not ob-

served in a single SNP analysis—potentially related to cis-acting effects within a haplotype (10).

HOW CAN HAPLOTYPES BE DETERMINED?

Determining the haplotypes that an individual possesses in a genomic region is challenging. Standard genotyping methods determine the alleles that an individual inherits at a particular genetic locus, but they do not provide information about whether particular alleles at adjacent loci occur in cis (same chromosomal strand) or trans (opposite chromosomal strand) orientation. If DNA samples from extended pedigrees are collected from all pedigree members up to the grandparental generation, then the “phase” or haplotypic organization of adjacent loci can often be inferred accurately. Because such complete families are rarely available for genetic studies, this is not practical for most complex disease investigations.

Molecular approaches to haplotype determination have been developed, but currently available approaches are not amenable to high throughput. For example, long-range polymerase chain reactions can be used to amplify one chromosome’s copy of a particular genomic region, which can be cloned and sequenced (10). Currently available molecular haplotyping methods are arduous and typically cost-prohibitive. New methods of molecular haplotype determination are under development (11, 12).

Because of the limitations of pedigree-based inference and molecular approaches for haplotype determination, statistical estimation approaches are typically used. An early approach involved sequential rules for haplotype inference (13). Subsequently, likelihood-based expectation–maximization (E–M) algorithms were developed (14), which have been incorporated into the SNPHAP program (15). In addition, approaches based on Gibbs sampling/Markov chain Monte Carlo methods have been widely used, as implemented in the PHASE program (16). In unrelated individuals, these approaches provide a probabilistic inference of the most likely haplotypes that an individual has inherited, as well as often providing the probability of alternate haplotypes being correct.

GENETIC ASSOCIATION ANALYSIS WITH HAPLOTYPES

Genetic association studies compare the distribution of genetic variants in cases and control subjects or assess the transmission of genetic variants within families. In either case-control or family-based designs, LD is the likely cause of significant associations—as long as genotyping error and population stratification (described *below*) have been avoided. Although initially performed with single SNPs, genetic association studies can also be performed with haplotypes. One approach is to assign the most likely haplotypes to each individual in a study population, and then to determine if the distribution of assigned haplotypes differs between cases and control subjects or within families.

Although this approach has been used in early studies, including our own analysis of ADRB2 haplotypes in asthma (17), it does not adjust for the uncertainty in haplotype assignment. Therefore, approaches that explicitly incorporate the relative probabilities of each haplotype for each individual are preferred. The statistical genetic issues in haplotype association studies of unrelated individuals have been recently reviewed by Schaid (18).

In a haplotype-based association analysis, one could test for association of each individual haplotype with a phenotype of interest. However, this requires adjustment for the multiple statistical testing involved with all of the individual haplotypes. If there is a single functional genetic variant that is located on a single haplotype, this approach may still have reasonable power. In general, however, a more robust approach is to perform a global test of haplotype association between the full complement of haplotypes and the phenotype.

A variety of haplotype-based association methods have been developed in both unrelated subjects (case-control or population-based) or in families; some of the more commonly used methods are listed in Table 1 (19, 22, 25, 27, 37, 38). For example, Schaid and colleagues developed a regression-based score test for haplotype association in unrelated subjects that allows for testing of both global haplotype association and individual haplotype association as implemented in the Haplo.Stats program (19). Such regression-based approaches have a number of advantages, including the inclusion of covariates for environmental and other nongenetic factors as well as the inclusion of haplotype-by-environment interactions (20).

Family-based haplotype association methods have also been developed. Although pedigree information can assist in haplotype assignment, haplotype uncertainty still must be considered in family-based haplotype association analysis. The TRANSMIT program, developed by Clayton and Jones, provides global and individual haplotype association analysis of qualitative traits (21). Subsequently, Horvath and colleagues extended the Family-Based Association Analysis Test (FBAT) to allow tests of global and individual haplotype associations for both qualitative and quantitative phenotypes (22). These family-based haplotype association analysis methods are also available in the Pedigree-Based Association Test (PBAT) program (23).

Application of haplotype-based association analysis requires judgment regarding the size of the haplotype to be included as well as the haplotype frequencies to be analyzed. Inclusion of a recombination hotspot within the haplotypic region studied will likely reduce power to detect significant associations (18). Candidate genes are often analyzed as a unit, but if a recombination hotspot occurs within a gene, this may not be appropriate. Haplotype-based analysis could be limited to haplotype blocks, although a large number of different algorithms for haplotype block definition have been devised, and results may vary considerably with the block definition algorithm used. Sliding window approaches may provide a reasonable compromise: a set of adja-

cent SNPs (e.g., 2, 3, or 4) are analyzed progressively across a region to identify the most significant region of association. Although a few haplotypes may account for most of the observed haplotypes in a population, there are often a large number of rare haplotypes as well. In addition to greater difficulty of accurately assessing these rare haplotypes, inclusion of rare haplotypes can increase the multiple statistical testing challenges. Although it is generally agreed that rare haplotypes are problematic, the optimal approach to dealing with them—lump all rare haplotypes together, shrink the effects of the rare haplotypes, or cluster them with phylogenetically similar haplotypes—has not been resolved (18, 24). Inclusion of rare haplotypes may change the association results in unpredictable ways. Although a significant association between ADRB2 variants and asthma diagnosis in the CAMP population of parent-child trios was not observed when we excluded haplotypes below 10% in frequency (17), evidence for association to asthma was observed when rare haplotypes were included (22).

An alternative to haplotype-based association analysis is to perform multimarker analysis—association analysis with multiple adjacent markers without regard to haplotypic phase. The relative power of multiple marker versus haplotype-based methods has not been definitively resolved. In a simulation study, Morris and colleagues suggested that inclusion of accurate haplotype information provided a modest improvement in efficiency at localizing a functional variant (~6%) compared with multimarker analysis with their shattered coalescent model (25). They also noted that the approach of inferring haplotypes from unphased genotypes in unrelated individuals (e.g., using SNPHAP or PHASE) and then using those inferred haplotypes in genetic association studies, led to a marked reduction in efficiency in localizing functional variants.

In addition to the approaches for haplotype-based association analysis described above, genetic studies in isolated populations, such as the Central Valley of Costa Rica, provide unique opportunities for haplotype-based analyses. For such populations, approaches based on determining association to presumed ancestral disease-carrying haplotypes have been developed (26). Cladistic approaches to reconstruct the phylogenetic ancestry of haplotypes in a genomic region and perform genetic association analysis have also been developed for nonisolated populations (27, 28).

RECENT HAPLOTYPE ANALYSES IN RESPIRATORY GENETICS

Haplotype-based association analyses are often included in genetic association studies of candidate genes and genomic regions. We will briefly review two recent respiratory genetics examples from our research group.

Raby and colleagues performed a genetic association study of SNPs in the TBX21 (T-Bet) gene and asthma (29). They

TABLE 1. SOME COMMONLY USED HAPLOTYPE ASSOCIATION ANALYSIS METHODS

Method	Population Type	Phenotypes	Software	Reference
Regression with score tests	Unrelated individuals (case-control or cohort)	Qualitative or quantitative	Haplo.Stats	19
Retrospective likelihood method	Case-control	Qualitative	CHAPLIN	37
Fine mapping with shattered coalescent	Case-control	Qualitative	COLDMAP	25
Family-based weighted conditional score test	Parent-child trios or pedigrees	Qualitative or quantitative	Haplotype FBAT	22
Family-based score test	Parent-child trios	Qualitative	TRANSMIT	38
Phylogenetic tree with family-based association analysis	Parent-child trios	Qualitative	Evolutionary Tree TDT	27

Definition of abbreviations: FBAT = Family-Based Association Analysis Test; TDT = Transmission/Disequilibrium Test.

resequenced the gene in 30 individuals to identify SNPs and characterize the haplotype block structure. Two haplotype blocks were identified using the Gabriel algorithm (2) for haplotype block definition in the Haploview program (30), although a reasonably high correlation between the haplotype blocks was found. They performed association analysis of 16 single TBX21 SNPs with a variety of asthma-related phenotypes in parent-child trios from the CAMP study, and significant associations of several SNPs to airway responsiveness were found. Haplotype association analysis was performed within each of the two haplotype blocks and across the entire TBX21 region. For asthma diagnosis, haplotype analysis was performed using the TRANSMIT program, and no significant associations were found. However, haplotype association analysis of airway responsiveness using the FBAT program revealed a significant global association test across the whole gene, which localized to the second haplotype block. In this second block, only one individual haplotype had a significant individual haplotype association test result. Thus, the haplotype analysis provided additional support for an association between TBX21 and airway responsiveness as well as guidance regarding the likely location of a susceptibility locus.

DeMeo and colleagues performed an association analysis between 48 SERPINE2 SNPs and COPD in Boston Early-Onset COPD Study extended pedigrees as well as in a case-control COPD study (COPD cases from the National Emphysema Treatment Trial and control subjects from the Normative Aging Study) (31). Single SNP association analysis was the primary approach; 16 SNPs demonstrated association to quantitative airflow obstruction phenotypes in Boston Early-Onset COPD Study families, and five of these SNPs replicated in the case-control association analysis. Haplotype association analysis was performed in the case-control population using sliding windows of 2, 3, and 4 adjacent SNPs with the HaploStats program. Two regions of the gene, in intron 1 and exon 3, provided the strongest haplotype associations. Significant individual haplotype associations were also found in these regions in Boston Early-Onset COPD Study extended pedigrees using the PBAT program. The replication of both single SNP and haplotype results in a family-based and case-control analysis, with the same direction of association in each population, increases the likelihood that these results are valid.

HAPLOTYPES OR SINGLE SNPs: WHICH APPROACH IS BETTER?

The relative power of using single SNPs or haplotypes for genetic association analysis depends on a variety of factors relating to the evolutionary history of the population and the genetic architecture of the disease gene variants in that population. If there is only a single genetic variant causing a particular disease in a population, the most powerful genetic association studies would involve a single SNP association analysis of that functional variant. If single SNPs in LD with a functional SNP are used rather than the functional variant itself, power for detecting that association will be reduced in a predictable manner—the equivalent population size is reduced in direct proportion to the degree of LD.

The power of haplotypes to detect genetic associations also depends on evolutionary history, which influences the extent of haplotype diversity. If a single functional mutation occurred on one haplotype, and most individuals that carry that particular haplotype also carry that mutation, power to detect a haplotype association to disease will be high. If functional mutations occurred on multiple haplotype backgrounds or if a functional variant occurs on only a small percentage of a particular haplotype, then power will be reduced. In a theoretical study, Morris

and Kaplan suggested that if there are multiple mutational origins for susceptibility alleles, power to detect associations will be decreased for both single SNP and haplotype-based approaches (32). However, haplotype-based approaches were more powerful than single SNP approaches in the setting of multiple susceptibility loci, especially when the SNPs that comprise the haplotype are not in strong LD—this LD pattern would be more likely to be seen in a fine mapping study of a large genomic region rather than a candidate gene investigation. In a particular complex disease genetic study, it is impossible to predict in advance whether a single SNP or haplotype approach will be more powerful. Both approaches are typically used.

A potential, but largely unexplored, complication of haplotype-based association studies relates to the impact of population stratification. Population stratification relates to differences in genetic ancestry between case and control populations; it can cause both false-positive and false-negative evidence for association in case-control studies. Methods to assess for, and adjust for, population stratification in single SNP studies have been developed (33, 34). For example, a panel of randomly selected SNPs can be genotyped in both cases and control subjects, and the extent of population stratification can be estimated using genomic control methods; adjustment of association analysis for this degree of estimated stratification will likely allow valid genetic association studies of single SNPs in most cases. However, because haplotypes capture evolutionary history more accurately than single SNPs, it is possible that haplotype-based association analyses may be more susceptible to population stratification than single SNPs. Appropriate assessment and adjustment for population stratification in haplotype-based methods have not yet been developed. Some family-based haplotype association methods (e.g., FBAT) are immune to population stratification effects, whereas other family-based methods (e.g., TRANSMIT) are not (35).

Another uncertain area in haplotype-based association analysis relates to the role of haplotype-based analysis in genomewide association studies. Progress in SNP genotyping technology and the identification of large numbers of SNPs through the HapMap project have allowed genomewide association studies to be designed (36), although they have not yet been reported for a respiratory disease. These studies typically involve genotyping 300,000 to 500,000 SNPs in case-control or family-based samples, which are analyzed for association. If single SNPs are analyzed, the challenges of adjusting for the multiple statistical testing involved are daunting. However, this multiple testing problem would be even greater if both single SNP and haplotype association analysis were performed. In addition, because the SNPs included in genomewide association panels are often selected to optimize LD coverage with the minimum number of SNPs, it is unclear how large a set of adjacent SNPs should be included in haplotype-based association analysis—even adjacent SNPs in a genomewide association analysis panel may occur in different haplotype blocks, reducing power to detect haplotype associations.

As noted above, in haplotype-based association analysis, localization of the related key functional variant or variants can be challenging. The associated haplotype is unlikely to be a functional element by itself, with possibly rare exceptions. Thus, it is essential to determine the haplotype block structure of the associated region to determine the potential locations of functional variants. Because the haplotype block structure could differ between reference subjects (e.g., HapMap subjects or study control subjects) and cases with a disease of interest, it may be important to determine the haplotype block structure within disease cases. Despite using multiple replication populations, including populations of different ethnicities, localization to a single functional variant may not be possible with genetic

epidemiologic approaches alone; at some point, molecular validation of functional impact is required.

Conflict of Interest Statement: E.K.S. received grant support, consulting fees, and honoraria from GlaxoSmithKline for studies of COPD genetics. He also received a speaker's fee from Wyeth for a talk on COPD genetics and received honoraria from Bayer.

Acknowledgment: The author thanks Drs. Dawn DeMeo, Scott Weiss, Nan Laird, Benjamin Raby, and Craig Hersh for helpful discussions.

References

- Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 2001;69:1–14.
- Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225–2229.
- Clark AG. The role of haplotypes in candidate gene studies. *Genet Epidemiol* 2004;27:321–333.
- 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.
- Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. *Nature* 2005;437:1299–1320.
- Kerem B, Rommens JM, Buchanan JA, Markiewicz D, Cox TK, Chakravarti A, Buchwald M, Tsui LC. Identification of the cystic fibrosis gene: genetic analysis. *Science* 1989;245:1073–1080.
- Laird N, Kraft P, Lange C, Van Steen K. Testing for association in genetic studies. In: Silverman EK, Shapiro SD, Lomas DA, Weiss ST, editors. *Respiratory genetics*. London: Hodder Arnold; 2005. pp. 27–46.
- Sebastiani P, Lazarus R, Weiss ST, Kunkel LM, Kohane IS, Ramoni MF. Minimal haplotype tagging. *Proc Natl Acad Sci USA* 2003;100:9900–9905.
- 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.
- Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc Natl Acad Sci USA* 2000;97:10483–10488.
- Kwok PY, Xiao M. Single-molecule analysis for molecular haplotyping. *Hum Mutat* 2004;23:442–446.
- Mitra RD, Butty VL, Shendure J, Williams BR, Housman DE, Church GM. Digital genotyping and haplotyping with polymerase colonies. *Proc Natl Acad Sci USA* 2003;100:5926–5931.
- Clark AG. Inference of haplotypes from PCR-amplified samples of diploid populations. *Mol Biol Evol* 1990;7:111–122.
- Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995;12:921–927.
- Clayton D. SNPHAP: a program for estimating frequencies of large haplotypes of SNPs. Available from: <http://www-gene.cimr.cam.ac.uk/clayton/software/> (accessed October 16, 2006).
- Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68:978–989.
- Silverman EK, Kwiatkowski DJ, Sylvia JS, Lazarus R, Drazen JM, Lange C, Laird NM, Weiss ST. Family-based association analysis of beta2-adrenergic receptor polymorphisms in the Childhood Asthma Management Program. *J Allergy Clin Immunol* 2003;112:870–876.
- Schaid DJ. Evaluating associations of haplotypes with traits. *Genet Epidemiol* 2004;27:348–364.
- 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.
- Lake SL, Lyon H, Tantisira K, Silverman EK, Weiss ST, Laird NM, Schaid DJ. Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. *Hum Hered* 2003;55:56–65.
- Clayton D, Jones H. Transmission/disequilibrium tests for extended marker haplotypes. *Am J Hum Genet* 1999;65:1161–1169.
- Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM. Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet Epidemiol* 2004;26:61–69.
- Lange C, DeMeo D, Silverman EK, Weiss ST, Laird NM. PBAT: tools for family-based association studies. *Am J Hum Genet* 2004;74:367–369.
- Clayton D, Chapman J, Cooper J. Use of unphased multilocus genotype data in indirect association studies. *Genet Epidemiol* 2004;27:415–428.
- Morris AP, Whittaker JC, Balding DJ. Little loss of information due to unknown phase for fine-scale linkage-disequilibrium mapping with single-nucleotide-polymorphism genotype data. *Am J Hum Genet* 2004;74:945–953.
- Service SK, Lang DW, Freimer NB, Sandkuijl LA. Linkage-disequilibrium mapping of disease genes by reconstruction of ancestral haplotypes in founder populations. *Am J Hum Genet* 1999;64:1728–1738.
- Seltman H, Roeder K, Devlin B. Transmission/disequilibrium test meets measured haplotype analysis: family-based association analysis guided by evolution of haplotypes. *Am J Hum Genet* 2001;68:1250–1263.
- Templeton AR. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping or DNA sequencing. V. Analysis of case/control sampling designs: Alzheimer's disease and the apolipoprotein E locus. *Genetics* 1995;140:403–409.
- Raby BA, Hwang ES, Van Steen K, Tantisira K, Peng S, Litonjua A, Lazarus R, Giallourakis C, Rioux JD, Sparrow D, et al. T-bet polymorphisms are associated with asthma and airway hyperresponsiveness. *Am J Respir Crit Care Med* 2006;173:64–70.
- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263–265.
- DeMeo D, Mariani TJ, Lange C, Srisuma SS, Litonjua AA, Celedon JC, Lake SL, Reilly JJ, Chapman HA, Mecham BH, et al. The SERPINE2 gene is associated with chronic obstructive pulmonary disease. *Am J Hum Genet* 2006;178:253–264.
- Morris RW, Kaplan NL. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. *Genet Epidemiol* 2002;23:221–233.
- Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999;55:997–1004.
- Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999;65:220–228.
- Cervino AC, Hill AV. Comparison of tests for association and linkage in incomplete families. *Am J Hum Genet* 2000;67:120–132.
- Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005;6:95–108.
- Satten GA, Epstein MP. Comparison of prospective and retrospective methods for haplotype inference in case-control studies. *Genet Epidemiol* 2004;27:192–201.
- Clayton D. A generalization of the transmission/disequilibrium test for uncertain- haplotype transmission. *Am J Hum Genet* 1999;65:1170–1177.