Am. J. Hum. Genet. 76:1–7, 2005

REVIEW ARTICLE
Prospects for Admixture Mapping of Complex Traits

Paul M. McKeigue

Conway Institute, University College Dublin, Dublin

Admixture mapping extends to human populations the principles that underlie linkage analysis of an experimental cross. For detecting genes that contribute to ethnic variation in disease risk, admixture mapping has greater statistical power than family-linkage studies. In comparison with association studies, admixture mapping requires far fewer markers to search the genome and is less affected by allelic heterogeneity. Statistical-analysis programs for admixture mapping are now available, and a genomewide panel of markers for admixture mapping in populations formed by West African–European admixture has been assembled. Some of the remaining technical challenges include the ability to ensure that the statistical methods are robust and to develop marker panels for other admixed populations. Where admixed populations and panels of markers informative for ancestry are available, admixture mapping can be applied to localize genes that contribute to ethnic variation in any measurable trait.

Introduction

Gene flow between subpopulations generates chromosomes made up of segments that have ancestry from different subpopulations. The admixture proportions of an individual are defined as the proportions of the individual's genome that have ancestry from each subpopulation. Suggestions that the genetic structure of admixed human populations could be exploited to localize genes that underlie ethnic variation in diseases or traits of interest date to Rife (1954). At first, writers who explored this approach viewed it as an extension of classic linkage-disequilibrium mapping, based on detection of allelic association with the disease or trait of interest (Chakraborty and Weiss 1988; Risch 1992; Briscoe et al. 1994; Stephens et al. 1994; McKeigue 1997). This approach, named "mapping by admixture disequilibrium" (Stephens et al. 1994), does not fully exploit the information available (Hoggart et al. 2004). The problem of how to exploit admixture to localize genes is more easily understood as an extension of linkage analysis of a cross (McKeigue 1998; McKeigue et al. 2000). In an experimental cross, inbred strains that differ with respect to a trait of interest are crossed for at least two generations to generate hybrid individuals that are typed at marker loci where different alleles have been fixed in the two ancestral strains. Linkage is detected by testing for association of the trait with ancestry—defined as the number of gene copies inherited from each ancestral strain—at each locus. In principle, this approach can be extended to detect genes in which trait-influencing alleles are distributed differentially between human subpopulations. If we type marker loci to infer ancestry at each locus, we can test for association of locus ancestry with any traits that have been measured in a sample of admixed individuals. Since inference is based on locus ancestry rather than on linkage disequilibrium, this approach is named simply "admixture mapping."

Received August 10, 2004; accepted for publication October 7, 2004; electronically published November 11, 2004.

Address for correspondence and reprints: Dr. Paul M. McKeigue, Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland. E-mail: paul.mckeigue@ucd.ie

0002-9297/2005/7601-0002$15.00

Theory of Admixture Mapping

To extend the methods used for linkage analysis of an experimental cross to admixed human populations, three main problems must be overcome. First, individual histories of admixture cannot be under experimental control and are usually unknown. We cannot, for instance, design a study that is based on sampling only F2 intercrosses. Gene flow from ancestral subpopulations into an admixed population generates variation of admixture proportions between individuals. This generates associations of the disease or trait with states of ancestry (or alleles) at loci that are not linked to a trait locus. This problem can be reduced to one of controlling for confounding by parental admixture proportions. To eliminate this confounding, we condition on parental admixture proportions (which can be estimated from the individual's own genotype data) when testing for association between trait and ancestry (McKeigue 1998). Since some confusion has arisen on this point (Zhu et al. 2004), the argument is spelled out here. Associations between states of ancestry at different loci are generated,

because ancestry at each locus on a gamete depends on the proportion of the parent's genome that has ancestry from each subpopulation. Because genes at unlinked loci segregate independently, locus-ancestry associations that are independent of parental admixture proportions cannot be generated unless the loci are linked (McKeigue 1998). This argument holds regardless of the history of admixture: it does not matter whether there has been continuous gene flow or isolation since the first generation of admixture. A similar argument underlies the "structured-association" approach to controlling for population stratification in ordinary genetic-association studies (Pritchard and Donnelly 2001).

The second problem is that human ethnic groups are not inbred strains, between which Wright's fixation index (F_{ST}) is 1 by definition. F_{ST} distances between human subpopulations originating on different continents are typically in the range of 0.1–0.2 (Cavalli-Sforza et al. 1994), which implies that only 10%–20% of shared allelic diversity has been lost since these subpopulations separated. Marker loci such as *FY,* at which different alleles have been fixed in West Africans and Europeans, are rare. Thus, we cannot infer the number of gene copies inherited from each ancestral subpopulation simply by typing a single marker locus. This problem can be overcome by combining data from all markers on each chromosome in a multipoint analysis to infer ancestry at each locus. As with multipoint methods for classic linkage analysis, in which segregation indicators at each locus are inferred by combining information from all markers on each chromosome (Lander and Green 1987), the proportion of information extracted can be increased to any required level by increasing the density of the marker map.

The third problem is that the ancestral groups that underwent admixture may be unavailable for study or may not be known precisely. Estimates of the allele frequencies in each subpopulation are thus subject to uncertainty. For instance, we cannot sample the exact mix of West African subpopulations that contributed genes to the modern African American population. Allele frequencies in modern unadmixed West African and European populations may vary from the ancestry-specific allele frequencies—the allele frequencies given African and European ancestry at the locus—within the African American population. This problem can be overcome by combining data from unadmixed and admixed populations to re-estimate the ancestry-specific allele frequencies within the admixed population under study (Hoggart et al. 2004; Patterson et al. 2004). This also yields estimates of the "dispersion" between allele frequencies in the unadmixed populations and the corresponding ancestry-specific allele frequencies within the admixed population under study (Hoggart et al. 2004; Patterson et al. 2004).

Design Considerations

The design of admixture-mapping studies depends on the size of the effect associated with locus ancestry and the number of generations since admixture. For a binary trait, the size of the effect associated with locus ancestry can be quantified by the ancestry-risk ratio—the ratio of risk in individuals with two gene copies who have ancestry from the high-risk subpopulation to risk in individuals with zero gene copies who have ancestry from this subpopulation—that the locus contributes. Under the simplifying assumption of a model in which risk increases multiplicatively with each copy of the high-risk allele, the effect associated with locus ancestry depends on only the ancestry-risk ratio (McKeigue 1998). To calculate the required sample size for a given statistical power, under the assumption of a perfectly informative marker map, the investigator need specify only the ancestry-risk ratio and the population admixture proportions. A practical lower limit for the size of effect that can be detected by admixture mapping is an ancestry-risk ratio of ~1.5, which would require a few thousand affected individuals to detect (Hoggart et al. 2004; Patterson et al. 2004).

The average number of generations since admixture can be estimated from marker data for the admixed population under study. This determines the density of markers required to extract a given proportion of information about locus ancestry, the mapping resolution (for a given effect size and sample size), and the number of independent hypotheses that are tested in a search of the entire genome (Hoggart et al. 2004).

Comparison with Established Approaches

In the specific situations in which it can be applied, admixture mapping has several advantages over established approaches to localization of disease-susceptibility genes. In comparison with family linkage studies, admixture mapping has higher statistical power to detect genes of modest effect if risk alleles in these genes are distributed differentially between subpopulations. Illustrative comparisons are given by McKeigue (1998) and by Montana and Pritchard (2004). Thus, for instance, a locus at which the haplotype-risk ratio is 2 can contribute an ancestry-risk ratio as high as 4 (when low-risk and high-risk alleles are differentially fixed in the two subpopulations) but can contribute a sibling-recurrence–risk ratio only as high as 1.13 (when the frequency of the high-risk allele is 34%). In this extreme case, <200 affected individuals are required for detection of the locus by admixture mapping, but at least 4,000 pairs are required for detection of it in an affected–sib-pair design at the same statistical power. This statistical-power advantage has a fundamental statistical basis: linkage analysis of a cross is based on a direct (fixed-effects) comparison, whereas family linkage studies are based on an

indirect (random-effects) comparison that is less efficient (Lander and Schork 1994).

In comparison with association studies, admixture mapping has two key advantages. First, it typically requires only 2,000–3,000 ancestry-informative markers for a genome search (McKeigue 1998; Smith et al. 2004), compared with estimates of at least 250,000 markers for whole-genome association studies (Carlson et al. 2004). Second, admixture mapping is less susceptible to allelic heterogeneity. The ability to detect a disease locus depends only on whether the pool of high-risk alleles is distributed differentially between subpopulations; it does not matter whether there are a few common risk-associated alleles or many rare risk-associated alleles at the locus under study (Terwilliger and Goring 2000). In contrast, SNP association studies of common diseases depend critically on the "common disease–common variant" hypothesis (Terwilliger and Weiss 1998).

What Problems Have Been Solved?

Although the theory that underlies admixture mapping was outlined several years ago (McKeigue 1998), its application has awaited the availability of genomewide panels of markers informative for ancestry between continental groups and statistical methods that combine information from these markers to infer ancestry.

Statistical and Computational Methods

Realistic statistical models for genotype data from admixed populations are too complex to be easily fitted by classic methods but can be handled with Bayesian computationally intensive methods. Three Bayesian programs—STRUCTURE (Falush et al. 2003), ADMIXMAP (Hoggart et al. 2003, 2004), and ANCESTRYMAP (Patterson et al. 2004)—have been described for multipoint statistical modeling of genotype data from admixed populations. Classic likelihood-based programs (Zhang et al. 2004; Zhu et al. 2004) have also been developed, but those cannot exploit the hierarchical dependence of individual-level parameters on population-level parameters, nor do they allow for uncertainty in model parameters, such as allele frequencies.

The Bayesian programs and one of the classic programs (Zhang et al. 2004) are based on a statistical model in which K subpopulations contribute to the gene pool of the admixed population, and variation of ancestry on each gamete is generated by K independent Poisson arrival processes, one for each subpopulation. In this model, it is as if states of ancestry from each subpopulation "arrive" at random as we progress along the chromosome, and locus ancestry is determined by the last arrival. The ratios between the intensities of these arrival processes are specified by the admixture proportions of the parent. The sum of intensities of the arrival processes can be interpreted as the average number of generations since admixture: this is estimated to be approximately six per morgan in the African American population (Hoggart et al. 2004; Patterson et al. 2004). The density of markers required to achieve a given map-information content is directly proportional to this sum-of-intensities parameter. This basic model can be extended to include regression models for binary or quantitative traits (Hoggart et al. 2004). Two elegant computational innovations that have made feasible the analysis of large data sets with dense marker maps are the introduction of a hidden Markov model forward-backward algorithm for sampling ancestry states on each chromosome (Falush et al. 2003) and a Rao-Blackwellized estimator that exploits this algorithm to calculate test statistics (Patterson et al. 2004).

To test for linkage in an affected-only design, we compare the observed and expected proportions of gene copies that have ancestry from the high-risk subpopulation at the locus under study. To test the null hypothesis that the ancestry-risk ratio is 1, either a score test (McKeigue et al. 2000; Hoggart et al. 2004) or a likelihood-ratio test (Patterson et al. 2004) can be calculated. These tests are asymptotically equivalent in large samples. The affected-only test statistic proposed by Montana and Pritchard (2004) is equivalent to the score test but with the score variance estimated by simulation rather than calculated from the likelihood.

Similar approaches can be used to construct tests that compare cases and controls or to construct tests for linkage with a quantitative trait. Other types of outcome variable, such as survival times, could be handled within the framework of generalized linear models. To detect modifier loci that influence the severity of a Mendelian disease such as sickle-cell anemia, we could sample affected individuals of mixed descent and test for linkage with severity of disease, measured as a quantitative trait.

Identification of Markers Informative for Ancestry

In principle, any type of marker—STRs, insertion-deletion polymorphisms, or SNPs—can be used for admixture mapping if their allele frequencies differ between the ancestral subpopulations. Marker information content for ancestry can be measured by the score variance (Fisher information) (McKeigue 1998; Molokhia et al. 2003; Pfaff et al. 2004) or by the expected log-likelihood ratio (Kullback-Leibler information) (Rosenberg et al. 2003; Smith et al. 2004) contributed by typing a gamete at the marker locus. These measures can be calculated from the ancestry-specific allele frequencies, expressed as a proportion of the information about ancestry that would be extracted by a perfectly informative marker, given a uniform prior distribution. The symbols f (Mc-

Keigue 1998) and I_n (Rosenberg et al. 2003) have been used for Fisher and Kullback-Leibler ancestry-information content, respectively. These measures rank markers similarly, with respect to information content for ancestry, although their absolute values differ. For two subpopulations, the average f value of stable diallelic marker polymorphisms is related to the F_{ST} distance, as $f = F_{ST}/(2 - F_{ST})$. Thus, if the average F_{ST} distance between European and West African populations is ~0.15 (Cavalli-Sforza et al. 1994), the average f value of randomly chosen SNPs between these two subpopulations would be ~0.08.

As in classic multipoint-linkage analysis, the adequacy of the marker panel for admixture mapping can be evaluated by an information-content map. This measures, at each locus, the ratio of the observed information (about the ancestry-risk ratio) extracted by an affected-only study that uses this marker panel to the complete information that would be extracted if locus-ancestry and parental-admixture proportions were observed directly. The observed information and complete information are evaluated when the affected-only score test is calculated (Hoggart et al. 2004). This calculation can be extended to partition the missing information into two components: (1) uncertainty about locus ancestry arising from inadequate marker coverage of the region containing the locus under study and (2) uncertainty about model parameters such as parental admixture and allele frequencies.

To assemble panels of markers that are informative for ancestry between populations originating on different continents, it is necessary to screen large numbers of marker loci. Smith et al. (2004) have assembled such a panel for admixture mapping in populations of mixed West African and European descent. They screened 450,000 SNPs for which allele-frequency data were available and chose 3,075 informative markers to achieve a target level of coverage of the genome. The average information content of these markers for West African versus European ancestry was 35% (measured by the f value) and 28% (measured by I_n), which corresponds to an average–allele-frequency differential of 0.56. From this panel, they selected a subset of 2,154 SNPs optimized for admixture mapping. They calculated the average map-information content of this panel in African Americans to be 71% when there was no uncertainty about model parameters but only ~50% when estimated from real data. They attributed this discrepancy to uncertainty about the ancestry-specific allele frequencies. As genotype data for these markers in African Americans accumulate, it should be possible to reduce this uncertainty.

Montana and Pritchard (2004) suggest that, as multiplexed assays capable of scoring 10,000 SNPs at low cost become available, it may be unnecessary to preselect markers that are highly informative for ancestry. Their simulations show that, as one would expect, the density of markers required to achieve a given map-information content is inversely proportional to the ancestry-information content of the individual markers. On this basis, in populations of mixed West African and European ancestry, a panel of 10,000 unselected SNPs, which would have an average f value of 0.08, would be equivalent to a panel of 2,000 ancestry-informative SNPs with an average f value 0.40. Apart from the lower genotyping workload, however, there are other advantages in the use of a panel of markers preselected to be informative: markers for which allele frequencies vary within continental groups can be excluded, the marker spacing can be large enough to ensure no allelic association within subpopulations, and the computational burden is reduced. To reduce costs further, a two-stage genotyping strategy can be used (Hoggart et al. 2004)

What Problems and Challenges Remain?

How Robust Will Admixture Mapping Be to Violation of Model Assumptions?

The first methods to be developed for linkage analysis of complex traits in human families were highly susceptible to false-positive results when model assumptions were violated (Ott 1992; Babron et al. 1993; Freimer et al. 1993; Göring and Terwilliger 2000). Can we avoid a similar experience with admixture mapping? The Bayesian statistical methods developed for admixture mapping are more flexible than the approaches used in early linkage analysis and allow diagnostic tests for violation of the assumptions that underlie the statistical model to be constructed (Hoggart et al. 2004). When these diagnostics indicate violation of model assumptions, the statistical model can be respecified to relax these assumptions.

The most serious problems are likely to arise with affected-only tests, which assume that the frequencies of locus-ancestry states do not vary across the genome within the admixed population under study. As Montana and Pritchard (2004) note, this assumption may be false when the admixed population has been small enough for locus-ancestry–state frequencies to drift. The effect of such drift is to inflate the variance of the affected-only test statistic. Use of a case-control design overcomes this problem but increases fourfold the number of individuals who have to be genotyped to detect an effect of given size. Instead, one could test the assumption of homogeneity of locus-ancestry–state frequencies across the genome by evaluating the affected-only test in a control group. When this test yields evidence for drift, data from unlinked loci could be used to correct the score variance by a "genomic control" approach, as suggested by Zhu et al. (2004). As in ordinary association studies

(Devlin and Roeder 1999), such a correction also would account for cryptic relatedness between cases.

The statistical model for variation of ancestry on chromosomes that is specified by current programs for modeling admixture does not correspond exactly to any biological model, even when admixture occurs in a single generation, followed by random mating (McKeigue 1998). The fit of data to this simple statistical model is likely to be especially poor when there has been continuous gene flow. It is possible, in principle, to fit a more-complex model for ancestry on chromosomes. Intuitively, however, one would expect that increasing the density of markers would make inference about locus ancestry more robust to violation of the model assumptions, just as increasing the density of markers used in multipoint family-linkage analysis makes inference about segregation robust to misspecification of allele frequencies.

Assembling Marker Panels for Admixture between Less Genetically Distant Populations

The cost of screening large numbers of SNP loci to identify those with extreme allele-frequency differentials can be reduced by measuring allele frequencies in pooled DNA samples, especially if multiplexed assays are available. Between populations that are separated only by a small genetic distance, such as Europeans and South Asians, SNPs that are highly informative for ancestry may be difficult to identify by this approach. A possible supplemental strategy is to construct compound marker loci that consist of two or more tightly linked SNPs and to model the unobserved haplotypes to extract information about ancestry (Hoggart et al. 2004).

Extension to Pedigrees

Although sampling unrelated individuals is the most efficient design for admixture mapping, many existing collections of DNA and clinical data are based on simple pedigrees, such as affected sib pairs. Developing ways to exploit these collections for admixture mapping most efficiently will require extension of current statistical-analysis programs to deal with pedigrees. For this, we can combine the model developed for admixture in unrelated individuals with the hidden Markov model developed for multipoint analysis of linkage in pedigrees (Lander and Green 1987). Thus, we can specify a model in which the variation of ancestry at marker loci on each founder chromosome and the variation of segregation indicators at these loci in each meiosis arise from independent Markov processes. This is straightforward in principle but, in practice, will present some statistical and computational challenges for implementation.

Fine Mapping Where Linkage Has Been Detected

Calculations suggest that the resolution of admixture mapping for effects of the size that the study is powered to detect will typically be <5 cM (Hoggart et al. 2004). The next step will be fine mapping. Standard tests for allelic association are not reliable for fine mapping of a gene that contributes to ethnic variation in disease risk because, in an admixed population, allelic associations with the disease will typically extend over long distances. This problem can be overcome by constructing a test for allelic association that conditions on locus ancestry rather than on parental-admixture proportions. Thus, for instance, in an African American sample that has been typed with ancestry-informative markers across the region under study, we can stratify the sample of gametes by ancestry (West African or European) at the locus under test, and we can test for association at alleles at this locus with disease within each stratum. Another possible approach to fine mapping of genes that underlie ethnic differences in disease risk is to test for evidence of recent selection pressure (Sabeti et al. 2002), given that, when risk alleles at a locus have become differentially distributed between populations, this is often a consequence of differential selection pressure, as with loci that affect malaria susceptibility.

Range of Applications

Populations

Admixture mapping can be applied only when admixture has been occurring at least two generations. The most obvious applications are to populations formed by admixture between groups originating on different continents that occurred as a result of European maritime expansion during the past few hundred years. These include populations formed by two-way and three-way admixture between Europeans, West Africans, and Native Americans in the Americas and populations formed by two-way admixture of Europeans with indigenous populations in Australia, the Pacific Islands, and polar regions.

In other parts of the world, admixture between populations that are less genetically distant (as measured by F_{ST}) and that are separated only by land barriers has occurred more slowly over a longer time period. Examples include the Tibeto-Burman populations of southern China, formed by admixture of northern Chinese immigrants with indigenous populations during the past 2,600 years (Wen et al. 2004), and the Roma populations of southeastern Europe, formed by admixture of migrants from southern Asia with indigenous Europeans during the past 1,000 years (Gresham et al. 2001). Admixture mapping in these populations will be more challenging, because markers informative for ancestry are

more difficult to identify when F_{ST} distances are small and because higher marker densities are required to extract information about ancestry when admixture has occurred over a long period of time.

Diseases and Traits

There are relatively few diseases (McKeigue 1997; Patterson et al. 2004) for which epidemiological criteria (based on migrant studies and the relationship of risk to individual admixture proportions) support genetic explanations for ethnic variation in risk. However, these diseases include some leading causes of morbidity and mortality, such as type 2 diabetes, hypertension, obesity, coronary disease, and prostate cancer. The most widespread ethnic variation in disease risk is for type 2 diabetes: in comparison with Europeans, high-risk groups include Native Americans (Knowler et al. 1978), Pacific Islanders (Zimmet et al. 1977), indigenous Australians (Wise et al. 1976), South Asians (McKeigue et al. 1991), and Peninsular Arabs (Al-Mahroos and McKeigue 1998). Although the most obvious applications of admixture mapping are to diseases for which risk varies between ethnic groups, the approach is not necessarily limited to such diseases. It is possible that loci at which risk alleles are distributed differentially between populations exist even when no difference in overall disease risk can be detected: for instance, when two or more loci have effects in opposite directions.

Conclusion

Most of the technical problems in admixture mapping have now been solved. When admixed populations and panels of markers informative for ancestry are available, admixture mapping can be applied to localize and, ultimately, identify genes that contribute to ethnic variation in any measurable trait. This raises some issues about informed consent and the interests of research participants. For instance, when admixture mapping leads to the discovery of an allele that is unique to a group defined by biogeographical ancestry and that is associated with a trait perceived as undesirable, misuse of this result could jeopardize the interests of the group in which the allele occurs. When studying the genetic basis of disease susceptibility, the value of discoveries that lead to advances in the control of disease may outweigh such objections. Application of admixture mapping to traits that are not of direct medical relevance, such as psychological traits, is likely to be more controversial and may undermine people's willingness to participate in research. Individuals who donate DNA for admixture-mapping studies should be made aware that their DNA can be used to study the genetic basis of ethnic variation in any trait that has been measured in them. Although community consultation is good practice in any research that depends on the active participation of a population group (see the Bioethics Resources on the Web site), a requirement for additional consent at the group level would raise further difficulties (Juengst 1998; Weijer et al. 1999). The groups under study in admixture mapping do not necessarily correspond to those whose interests might be jeopardized by misuse of research results, and the demarcation, for purposes of consent, of groups defined by biogeographical ancestry is contrary to current notions of equality between individuals (Juengst 1998). The ability of researchers to gain acceptance for research on the genetic basis of ethnic variation in disease risk may depend on how discoveries in this area are exploited.

Acknowledgment

This work was supported by National Institutes of Health grant MH60343 (to the author).

Electronic-Database Information

The URL for data presented herein is as follows:

Bioethics Resources on the Web, http://www.nih.gov/sigs/bioethics/named_populations.html

References

Al-Mahroos F, McKeigue PM (1998) High prevalence of diabetes mellitus in Bahrainis: associations with ethnicity and raised plasma cholesterol. Diabetes Care 21:936–942

Babron MC, Martinez M, Bonaitipellie C, Clergetdarpoux F (1993) Linkage detection by the affected-pedigree-member method—what is really tested? Genet Epidemiol 10:389–394

Briscoe D, Stephens JC, O'Brien SJ (1994) Linkage disequilibrium in admixed populations: applications in gene mapping. J Hered 85:59–63

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

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ

Chakraborty R, Weiss KM (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci USA 85:9119–9123

Devlin B, Roeder K (1999) Genomic control for association studies. Biometrics 55:997–1004

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Freimer NB, Sandkuijl LA, Blower SM (1993) Incorrect specification of marker allele frequencies: effects on linkage analysis. Am J Hum Genet 52:1102–1110

Göring HHH, Terwilliger JD (2000) Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. Am J Hum Genet 66:1298–1309

Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kučinskas V, Perez-Lezaun A, Marushiakova E, Popov V, Kalaydjieva L (2001) Origins and divergence of the Roma (Gypsies). Am J Hum Genet 69:1314–1331

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM (2003) Control of confounding of genetic associations in stratified populations. Am J Hum Genet 72:1492–1504

Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM (2004) Design and analysis of admixture mapping studies. Am J Hum Genet 74:965–978

Juengst ET (1998) Groups as gatekeepers to genomic research: conceptually confusing, morally hazardous, and practically useless. Kennedy Inst Ethics J 8:183–200

Knowler WC, Bennett PH, Hamman RF, Miller M (1978) Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minnesota. Am J Epidemiol 108:497–505

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84:2363–2367

Lander ES, Schork NJ (1994) Genetic dissection of complex traits. Science 265:2037–2048

McKeigue PM (1997) Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. Am J Hum Genet 60:188–196

——— (1998) Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. Am J Hum Genet 63:241–251

McKeigue PM, Carpenter J, Parra EJ, Shriver MD (2000) Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. Ann Hum Genet 64:171–186

McKeigue PM, Shah B, Marmot MG (1991) Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. Lancet 337:382–386

Molokhia M, Hoggart C, Patrick AL, Shriver M, Parra E, Ye J, Silman AJ, McKeigue PM (2003) Relation of risk of systemic lupus erythematosus to West African admixture in a Caribbean population. Hum Genet 112:310–318

Montana G, Pritchard JK (2004) Statistical tests for admixture mapping with case-control and case-only data. Am J Hum Genet 75:771–789

Ott J (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. Am J Hum Genet 51:283–290

Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, Hauser SL, Smith MW, O'Brien SJ, Altshuler D, Daly MJ, Reich D (2004) Methods for high-density admixture mapping of disease genes. Am J Hum Genet 74:979–1000

Pfaff CL, Barnholtz-Sloan J, Wagner JK, Long JC (2004) Information on ancestry from genetic markers. Genet Epidemiol 26:305–315

Pritchard JK, Donnelly P (2001) Case-control studies of association in structured or admixed populations. Theor Popul Biol 60:227–237

Rife DC (1954) Populations of hybrid origin as source material for the detection of linkage. Am J Hum Genet 6:26–33

Risch, N (1992) Mapping genes for complex diseases using association studies with recently admixed populations. Am J Hum Genet Suppl 51:13

Rosenberg NA, Li LM, Ward R, Pritchard JK (2003) Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 73:1402–1422

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 (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837

Smith MW, Patterson N, Lautenberger JA, Truelove AL, McDonald GJ, Waliszewska A, Kessing BD, et al (2004) A high-density admixture map for disease gene discovery in African Americans. Am J Hum Genet 74:1001–1013

Stephens JC, Briscoe D, O'Brien SJ (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am J Hum Genet 55:809–824

Terwilliger JD, Göring HHH (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 72:63–132

Terwilliger JD, Weiss KM (1998) Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr Opin Biotechnol 9:578–594

Weijer C, Goldsand G, Emanuel EJ (1999) Protecting communities in research: current guidelines and limits of extrapolation. Nat Genet 23:275–280

Wen B, Xie X, Gao S, Li H, Shi H, Song X, Qian T, Xiao C, Jin J, Su B, Lu D, Chakraborty R, Jin L (2004) Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. Am J Hum Genet 74:856–865

Wise PH, Edwards FM, Craig RJ, Evans B, Marchland JB, Sutherland B, Thomas DW (1976) Diabetes and associated variables in the South Australian aboriginal. Aust N Z J Med 6:191–196

Zhang C, Chen K, Seldin MF, Li H (2004) A hidden Markov modeling approach for admixture mapping based on case-control data. Genet Epidemiol 27:225–239

Zhu X, Cooper RS, Elston RC (2004) Linkage analysis of a complex disease through use of admixed populations. Am J Hum Genet 74:1136–1153

Zimmet P, Taft P, Guinea A, Guthrie W, Thoma K (1977) The high prevalence of diabetes on a Central Pacific island. Diabetologia 13:111–115