Indian J. Genet., 79(1) Suppl. 276-283 (2019)
DOI: 10.31742/IJGPB.79S.1.19

Statistical challenges in genetic association analyses: Population-based versus family-based studies

Saurabh Ghosh* and Tanushree Haldar[1]

Human Genetics Unit Indian Statistical Institute, Kolkata 700 108; [1]Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA

(Received: February 2019; Revised: March 2019; Accepted: April 2019)

Abstract

Genetic association tests provide clues on chromosomal locations of putative genes underlying complex traits, both qualitative (such as disease status) and quantitative (such as measurable precursors of clinical outcomes). One can adopt either a population-based or a family-based study design to generate genotype and phenotype data necessary to carry out the tests of association. In this article, we provide an overview of the two contrasting study designs, the statistical issues pertaining to each of these study designs as well as their relative advantages and disadvantages. We also outline the current statistical challenges in interpreting association findings in the presence of population stratification and in carrying out appropriate comparisons of the powers of the association tests based on the two study designs.

Key words: Association mapping, quantitative traits, binary traits, population startification

Introduction

Association mapping of susceptible genes underlying complex disorders is an active area of current research in genetic epidemiology. Compared to Mendelian disorders, there has been limited success in identifying genes involved in complex disorders as these traits are believed to be controlled by multiple loci, some with minor gene effects, and genetic variation at any one locus does not completely determine the trait. Moreover, epistatic as well as gene-environment interactions often modify the risk of developing the disease. While linkage analyses (Ott 1999) have been traditionally successful in identifying rare variants with large genetic effect sizes characterizing Mendelian disorders, they have been relatively unsuccessful in detecting common variants with moderate effect sizes characterizing complex disorders. There is evidence that association studies, which measure the extent of linkage disequilibrium (LD) between alleles of two loci (Weir 1996), are statistically more powerful than linkage studies in gene mapping of complex traits (Risch and Merikangas 1996). This is because linkage disequilibrium (LD) exists over smaller distances on the genome compared to linkage. Thus, a positive association finding provides a more precise location of a locus modulating the underlying trait.

The aim of genetic association studies is to relate genetic information to a clinical outcome or a phenotype, which could be either qualitative or quantitative in nature. Qualitative traits are predominantly, though not exclusively, binary in nature and denote the affection status with respect to a clinical outcome. In contrast, quantitative traits are measurable physiological or bio-chemical quantities such as height, blood pressure, serum cholesterol concentration and body mass index (BMI) that are often, though not necessarily, major precursors of clinical outcomes. For example, total serum cholesterol and triglyceride levels are examples of quantitative traits that characterize cardiovascular disorder, a binary trait. These quantitative precursors carry more information on inter-individual trait variability compared to the binary end-point traits governed by them and hence, it has been argued that analyzing these quantitative traits may statistically be a more prudent strategy to detect association.

There are two popular classes of study designs for carrying out genetic association analyses: (i) population based and (ii) family based. While the first design involves collection of genotypic and phenotypic

*Corresponding author's e-mail: saurabh@isical.ac.in

Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org; indianjournals.com

data on a unrelated set of individuals selected randomly from a population, the latter deals with data on multiple individuals within families (Foulkes 2009). Thus, there are fundamental differences in the statistical issues pertaining to tests for association based on the two designs. Unlike population based designs which result in independent observations, family based designs induce observations that have higher within family correlations compared to between family correlations and hence, statistical tests based on identically and independently distributed observations need to be modified for analyzing family data. Moreover, family-based genetic data provides the flexibility to detect recombinations via haplotype construction but such detection is not possible based on unrelated individuals as in population-based designs. Thus, statistical evidence of association based on family-based designs are valid only in the presence of linkage but that based on population-based designs are, in general, independent of linkage and hence, may turn out to be genetically spurious.

Population-based genetic analyses

Binary traits

Population-based designs include case-control studies, cross-sectional studies and prospective or retrospective cohort studies. The individuals included in the study should be representative of all individuals in the *a priori* de- fined specific population. The most popular among the above designs based on qualitative data is the case-control test due to the ease of data collection and statistical methodology of testing for association. A comprehensive overview of the case-control study design is provided in Lewis and Knight [2012]. A random sample of cases (individuals affected with the disease of interest) is collected along with an independent random set of control indi-viduals (in general, those unaffected with respect to the disease) and the test for association is based on a comparison of the allele frequencies at a SNP between these two groups. It is important to ascertain the cases based on appropriate clinical criteria in order to ensure phenotypic homogeneity. While the usual mechanism of selecting controls is through screening based on absence of the disease condition, an alternative scheme is to select a random sample of individuals from the populations without any clinical ascertainment. It is intuitively clear that tests for association based on both types of controls yield comparable powers for rare diseases. However, for common diseases, the power of a test based on controls ascertained for absence of disease condition may be substantially higher than that based on controls without clinical ascertainment.

Statistical tests for association based on the case-control design can be carried out at two levels: genotypic and allelic. The genotype-level test is essentially a test of homogeneity of distributions of the different genotypes among cases and controls at a locus. For a single nucleotide polymorphism (SNP), there are three genotypes and hence, the test statistic is asymptotically distributed as chi-squares with two degrees of freedom under the null hypothesis of no genotype association. However, the test does not use any genetic information in the sense that the test statistic is invariant under the order of the genotypes in the 2 x 3 contingency table and hence, does not identify risk genotypes. On the other hand, the allele-level test is based on counts of the two alleles among cases and controls that can be obtained by decomposing the 2 x 3 genotype table to a 2 x 2 table. The test statistic is asymptotically distributed as chi-squares with one degree of freedom under the null hypothesis of no allelic association. This test is equivalent to the genotype level test when the true model of association is multiplicative (or log additive), that is, the genotype relative risk of the major homozygous genotype compared to the heterozygous genotype is equal to that of the heterozygous genotype compared to the minor homozygous genotype (Sasieni 1997). The Cochran-Armitage trend test (CATT) that explores for a trend in the risks corresponding to the ordered genotypes in the 2 × 3 table (Cochran 1954; Armitage 1955) is asymptotically equivalent to the allele-level test but is less affected by departures from Hardy-Weinberg Equilibrium (Sasieni 1997). While the allele level test is used in candidate gene approaches, CATT has gained popularity [O'Donovan et al. 2008] for genome-wide association studies (GWAS). If one is interested in more specific genetic hypotheses such as the mode of inheritance (dominant or recessive) at a SNP, one can compare the combined risk conferred by two of the genotypes with that conferred by the third. However, it is important to note that the test statistics corresponding to the different analytic methods described above are highly correlated and hence, one should apply a multiple testing correction when several of the analyses are simultaneously carried out on the same set of data (Lewis and Knight 2012).

Quantitative traits

In order to carry out association analyses of continuous or other quantitative traits, data are collected on genotypes and phenotypes data for a set of individuals selected randomly from the population without any ascertainment. The tests for association are based on detecting differences in phenotype characteristics across the different genotypes at a SNP. The popular statistical tools for this purpose are analysis of variance (ANOVA) and linear regression. ANOVA is analogous to the Pearson two degree of freedom genotype-level test in the case-control framework as it compares the null hypothesis of no association (equal means across the three genotypes) with a general alternative, while the linear regression approach assumes an additive allelic effect resulting in a reduction in the degrees of freedom from two to one (Balding 2006). Both the tests are valid under the assumption that the distribution of the underlying trait conditioned on each genotype is normal with the same variance. For violations in the above assumptions, approximate normality can be induced using logarithmic transformations of the observed trait values.

Many quantitative traits (e.g., symptom counts in a psychiatric diagnosis) do not follow a normal distribution, even after certain transformations (Li et al. 2013). Analyses based on ANOVA or a standard linear regression model may lead to misleading inferences. An alternative approach is to use nonparametric tests that depend on ranks and hence are not sensitive to violations in underlying distributional assumptions. The most popular among them is the Kruskal Wallis test (Kruskal and Wallis 1952) that compares the distribution of the quantitative trait across different genotype groups. In fact, it is analogous to standard ANOVA with the actual quantitative values replaced by their ranks. On the other hand, the Kruskal-Wallis test is less powerful than ANOVA when the underlying assumptions such as normality and homoscedasticity across genotype groups are indeed valid. Moreover, if prior knowledge is available on the ordering of the median trait values across the three genotype groups, the Jonckheere-Terpstra test (Jonckheere 1954; Terpstra 1952) is more optimal (though less robust) compared to the Kruskal Wallis test.

It is possible to analyze quantitative traits in the case-control framework by dichotomizing the sample based on some threshold. However, this leads to reduction in information on inter-individual variability resulting in sub- stantial loss in power of the association tests. However, Slatkin [1999] showed that the power of these tests may be increased by ascertaining individuals only from the extremes of the trait distribution. A detailed discussion on analyses based on transforming quantitative traits into binary traits is avail- able in Lewis and Knight (2012).

The Caveat: Population stratification

One of the major limitations of population-based genetic case-control studies as well as quantitative trait association analyses is the problem of population stratification. It is well known that allele frequencies vary widely within and between populations, irrespective of disease status (Cavalli-Sforza et al. 1994; Perez-Lezaun et al. 1997). This disparity across different populations can be attributed to unique genetic and social histories in terms of ancestral patterns of geographical migration, mating practices, reproductive expansions and bottlenecks as well as stochastic variation between individuals (Slatkin 1991). While none of the above factors is necessarily associated with any particular disease, tests for association based on a sample comprising genetically or phenotypically heterogeneous subpopulations are susceptible to inflated rates of false positives and hence, result in spurious evidence of association. The classical problem of population stratification is characterized by genetic differences between cases and controls attributable to diversity in background populations unrelated to the trait under study. Studies have shown the presence of allelic heterogeneity within many genes related to clinical traits (Goddard et al. 2000; Stephens et al. 2001). Hence, population substructure is a serious confounder in genetic association studies. A comprehensive discussion on population stratification has been provided in Cardon and Palmer [2003]. It is important to note that in addition to genetic heterogeneity, there needs to be differences in the disease prevalance across subpopulations and the presence of any one of these phenomenon is not sufficient to result in population stratification (Wacholder et al. 2000). The adverse effect of population stratification is also relevant in the con- text of quantitative traits. (Haldar and Ghosh 2012) theoretically studied the marginal as well as joint effects of genetic heterogeneity (differences in marker allele frequencies) and phenotypic heterogeneity (differences in standardized phenotypic means) across subpopulations on the false positive rates of the three popular population-based tests of association for quantitative traits : ANOVA, linear regression with an additive allelic effect and the Kruskal Wallis test. Population stratification is probably the major reason behind the failure to replicate many genetic association results (Tabor et al. 2002; Weiss and Terwilliger 2000; Terwilliger and Goring 2000). This problem is of specific relevance for studies on Indian populations due to increasing evidence of genetic heterogeneity among different ethnic populations in India (Basu et al. 2003; Thangaraj et al. 2005; Brahmachari et al. 2008).

Family-based genetic analyses

Family-based designs may vary from simple cases of parent-offspring trios, concordant/discordant sib-pairs to large multigenerational pedigrees. While most family-based tests for association are based on the transmission bias of an allele within informative families (families with at least one heterozygous parent) (Ewens et al. 2008), a few methods such as Abecasis's 'total' association test (Abecasis et al. 2000) have been developed to analyze all available data.

Trios comprising two parents and one offspring are the most popular family- based design. The classical Transmission Disequilibrium Test, more commonly known as TDT [Spielman et al., 1993] is the standard approach to test for linkage disequilibrium in the presence of linkage when at least one of the parents is heterozygous at the SNP and the offspring is affected with the disease of interest. The major difference between TDT and the case-control association test lies in the estimation of allele counts under the null hypothesis of no association. While the case-control design involves the computation of unconditional expectation of allele counts assuming the background frequencies to be same in cases and controls, TDT is based on the conditional expectation of allele counts among affected offspring given parental genotypes under Mendelian segregation [Laird and Lange, 2008].

The classical TDT (Spielman et al. 1993) based on the trio design is a valid test for both linkage and association. However, it is a valid test only for linkage in the presence of multiple sibs in a family. In the absence of allelic association between a marker and a disease locus, a parent heterozygous at the marker locus has equal chance of transmitting any one of the two alleles to an affected offspring. On the other hand, a marker allele that is in positive linkage disequilibrium with the risk-predisposing allele at the disease locus is likely to be preferentially transmitted by a heterozygous parent to the affected offspring. The test is equivalent to a McNemar's test when all parents are considered and a binomial equality of proportion test when only heterozygous parents are considered. Under the null hypothesis of no linkage or no association, the test statistic is asymptotically distributed as chi-squares with 1 degree of freedom. Since TDT requires genotype information on parents, the trio design is often infeasible for diseases with older age at onset. It can be analytically shown that even if it is possible to decipher the identity of the allele transmitted from a heterozygous parent when the data on the other parent is missing, inclusion of such families can increase the false positive rate of TDT [Curtis and Sham, 1995]. An alternative to TDT is Sib TDT (Spielman and Ewens 1998) that uses genotype data on sibships, comprising both affected and unaffected siblings, instead of parents. The test statistic compares the counts of a specific allele among affected and unaffected sibs. Since TDT explicitly uses information on parental trans- missions while Sib TDT attempts to decipher the same based on genotypes of sibs, TDT is expected to be more powerful compared to Sib TDT, espe- cially for smaller sibships. Whittaker and Lewis [1999] have shown that the number of families required for Sib TDT based on one affected and two un- affected sibs per family is one and a half times the number of trios required for TDT to yield equivalent powers to detect association. However, the test statistics corresponding to TDT and Sib TDT may be combined to perform a single test, thus allowing the flexibility of incorporating both trios as well as sibships in the analyses.

Weinberg et al. (1998) proposed a log-linear framework for qualitative traits to model the frequencies of the fifteen possible trio types comprising the different parental mating types and the feasible offspring genotypes. The case/pseudocontrol approach based on case-parent trios proposed by Cordell and Clayton (2002) is similar in principle, except that it models off-spring genotypes conditioned both on parental genotypes and ascertainment through the affected offspring. A modification of the above model has been presented in Cordell et al. (2004) that can be viewed as a generalization of the classical TDT and other methods (Schaid and Sommer 1993; Schaid 1996).

The model provides the flexibility of considering more complex models involving multiple linked or unlinked predisposing loci with possible epistatic or gene-environment interactions.

There have been various extensions of the classical TDT for binary traits to analyze quantitative trait data. The simplest approach is to transform the quantitative trait values to a dichotomous variable based on some threshold. For quantitative precursors of diseases, these thresholds are often determined by clinical manifestations. The transformed trait can then be analyzed using traditional methods like TDT (Allison 1997). However, an arbitrary choice of the threshold leads to substantial loss of information on trait variability. Allison (1997) modeled the quantitative trait of an offspring as a function of his/her genotype conditioned on the parental mating type and used standard ANOVA to test for association under the assumption that the trait values are normally distributed. On similar lines, Abecasis et al. (2000) proposed QTDT that models the trait values based on linear effects of the offspring genotype and the average of the parental genotypes. However, unlike the classical TDT, these models do not assume independence between parental transmissions. Moreover, since these approaches model a quantitative trait as a function of genotypes, they are essentially prospective analyses (Wheeler and Cordell 2007). On the other hand, modeling offspring genotypes in terms of quantitative trait values and possibly parental genotypes result in retrospective analyses (Wheeler and Cordell 2007) as is the case in the classical TDT (Spielman et al. 1993).

Among the retrospective approaches, Waldman et al. (1999) developed a logistic regression approach that models the log odds of transmission of a specific allele at a locus by a heterozygous parent conditioned on the quantitative trait value of the offspring. The model assumes the transmissions from two heterozygous parents to their offspring to be independent. Using a polytomous logistic regression extension of the log-linear likelihood approach for qualitative traits, Kistner and Weinberg (2004, 2005) modeled the log odds of an offspring having two copies of a specific allele versus one copy and that of having no copy of the allele versus one copy as functions of the trait value of the offspring conditioned on the parental genotypes. Wheeler and Cordell (2007) extended the case/psuedocontrol approach for qualitative traits. (Cordell and Clayton 2002) to deal with quantitative traits. (Haldar and Ghosh 2015) proposed a computationally simple logistic regression model *TBAT* to test for association for quantitative traits based on the traditional trio design which can be viewed as a direct extension of the classical TDT (Spielman et al. 1993) to quantitative traits. The method can also be easily modified to incorporate a multivariate phenotype vector possibly comprising both quantitative as well as qualitative traits.

In addition to the parametric and semi-parametric methods mentioned above, nonparametric extensions of the TDT have been suggested for quantitative traits. These methods do not make any assumptions about the distribution of the quantitative trait and are also applicable in situations with multiple offspring per nuclear family. The general approach involves estimating the covariance between the quantitative trait and the indicator for parental transmission (Rabinowitz 1997; Monks and Kaplan 2000). Similarly, another class of family-based association tests, referred to as FBAT, is also based on the covariance between a function of the genotype and a function of the trait (Laird et al. 2000).

Discussion

Given the susceptibility of population-based case-control studies to yield false positive findings due to population stratification, it has been of interest to develop family-based study designs, which despite being more demanding with respect to genotype requirements, circumvents the above problem. However, family-based designs are not cost-effective, often not feasible (for example, in pharmacogenetic studies) and grossly underpowered to detect genes with modest effects (Risch and Merikangas 1996). Thus, it has been argued that correction for population stratification in case-control studies using statistical adjustments may be more optimal than modifying the study design to a family-based framework. There have recently been promising developments in statistical methodologies in this regard (Pritchard et al. 2000; Devlin et al. 2001; Price et al. 2006; Majumdar et al. 2013), though it remains a statistical challenge to determine the optimal number of genome- wide markers required to evaluate the level of stratification and the extent to which the statistical corrections are able to reduce the inflated rate of false positives. However, there is increasing belief that population stratification probably does not have that adverse an effect as originally postulated, except when a replication study is carried out in an ethnically diverse population (Cardon and Bell 2001; Morton and Collins 1998).

The major analytical challenge in evaluating the relative performances of the population-based and family-based study designs lies in the fact that a direct and straight forward power comparison between allele-level tests of association based on the case-control design in the absence of population stratification and the classical TDT (Spielman et al. 1993) based on the trio design is not possible in the strict statistical sense because the study designs are different with respect to data requirements. A possible analytical frame- work to address this issue is to determine the number of cases (or controls) in a case-control design with equal number of cases and controls as well as the number of transmissions from heterozygous parents to affected offspring (along with the total number of families to be sampled to obtain the requisite number of transmissions) in a trio design that yield a pre-assigned power. We would also like to emphasize that while population stratification has an adverse effect on the false positive rates of only population-based association tests and not family-based tests of transmission disequilibrium, it can adversely affect the powers of association tests based on both types of data. Thus, one needs to be cautious that while one can evaluate both the marginal as well as the joint effects of genetic and phenotypic heterogeneities on the powers of family-based tests, one can ideally study only the marginal effect of either phenotypic or genetic heterogeneity on the powers of population-based tests so as to ensure controlled false positive rates.

Declaration

The authors declare no conflict of interest.

References

Abecasis G. R., Cardon L. R. and Cookson W. O. 2000. A general test of association for quantitative traits in nuclear families. American J. Human Genet., **66**: 279-292.

Allison D. B. 1997.Transmission-disequilibrium tests for quantitative traits. American J. Human Genet., **60**: 676-690.

P. Armitage. 1955. Tests for linear trends in proportions and frequencies. Biometrics, **11**(3): 375-386.

Balding D. J. 2006. A tutorial on statistical methods for population association studies. Nature Reviews Genet., **7**(10): 781-791.

A. Basu A., Mukherjee N., Roy S., Sengupta S., Banerjee S., Chakraborty M., B. Dey B., Roy M., Roy B., Bhattacharyya N. P., Roychoudhury S. and Majumder P. P. 2003. Ethnic India: A genomic view, with special reference to peopling and structure. Genome Res., **13**: 2277-2290.

Brahmachari S. K., Majumder P., Mukerji M., Habib S., Dash D., Ray K., Bahl S., Batra J., Consortium I. G. V., et al. 2008. Genetic landscape of the people of india: a canvas for disease gene exploration. J. Genet., **87**(1): 3-20.

Cardon L. R. and Bell J. I. 2001. Association study designs for complex diseases. Nature Reviews Genet., **2**(2): 91-99.

Cardon L. R. and Palmer L. J. 2003. Population stratification and spurious allelic association. The Lancet, **361**(9357): 598-604.

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

Cochran W. G. 1954. Some methods for strengthening the common χ^2 tests. Biometrics.

Cordell H. J. and Clayton D. G. 2002. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: Application to hla in type 1 diabetes. The American J. Human Genet., **70**(1): 124-141.

Cordell H. J., Barratt B. J. and Clayton D. G. 2004. Case/pseudocontrol analysis in genetic association studies: A unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. Genetic Epidemiol., **26**(3): 167-185.

Curtis D. and Sham P. 1995. A note on the application of the transmission dis- equilibrium test when a parent is missing. American J. Human Genet., **56**(3): 811.

Devlin B., Roder K. and Wasserman L. 2001. Genomic control, a new approach to genetic-based association studies. Theor. Population Biol., **60**: 155-166.

Ewens W. J., Li M. and Spielman R. S. 2008. A review of family-based tests for linkage disequilibrium between a quantitative trait and a genetic marker. PLoS Genet., **4**(9).

Foulkes A. S. 2009. Applied statistical genetics with R. Springer.

Goddard K. A., Hopkins P. J., Hall J. M. and Witte J. S. 2000. 2000. Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. The American J. Human Genet., **66**(1): 216-234.

Haldar T. and Ghosh S. 2012. Effect of population stratification on false positive rates of population-based association analyses of quantitative traits. Annals Human Genet., **76**: 237-245.

Haldar T. and Ghosh S. 2015. Statistical equivalent of the classical tdt for quantitative traits and multivariate phenotypes. J. Genet., **94**(4): 619-628.

Jonckheere A. R. 1954. A distribution-free k-sample test against ordered alternatives. Biometrika, 133-145.

Kistner E. O. and Weinberg C. R. 2004. Method for using complete and incomplete trios to identify genes related to a quantitative trait. Genetic Epidemiol., **27**(1): 33-42.

Kistner E. O. and Weinberg C. R. 2005. A method for identifying genes related to a quantitative trait, incorporating multiple siblings and missing parents. Genetic Epidemiol., **29**(2):155-165.

Kruskal W. H. and Wallis W. A. 1952. Use of ranks in one-criterion variance analysis. J. American Statistical Assoc., **47**(260): 583-621.

Laird N. M. and Lange C. 2008. Family-based methods for linkage and association analysis. Adv. Genet., **60**: 219-252.

Laird N. M., Horvath S. and Xu X. 2000. Implementing a unified approach to family-based tests of association. Genetic Epidemiol., **19**(S1): S36-S42.

Lewis C. M. and Knight J. 2012. Introduction to genetic association studies. Cold Spring Harbor Protocols, **2012**(3): pdb-top068163.

Li Q., Li Z., Zheng G., Gao G. and Yu K. 2013. Rank-based robust tests for quantitative-trait genetic association studies. Genetic Epidemiol., **37**(4): 358-365.

Majumdar A., Bhattacharya S., Basu A. and Ghosh S. 2013. A novel bayesian semiparametric algorithm for inferring population structure and adjusting for case-control association tests. Biometrics, **69**(1): 164-173.

Monks S. A. and Kaplan N. L. 2000. Removing the sampling restrictions from family-based tests of association for a quantitative-trait locus. American J. Human Genet., **66**: 576-592.

Morton N. and Collins A. 1998. Tests and estimates of allelic association in complex inheritance. Proc. Nat. Acad. Sci., **95**(19): 11389-11393.

O'Donovan M. C., Craddock N., Norton N., Williams H., Peirce T., Moskvina V., Nikolov I., Hamshere M., Carroll L., Georgieva L., et al. 2008. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nature Genet., **40**(9): 1053-1055.

Ott J. 1999. Analysis of Human Genetic Linkage. Johns Hopkins University Press, Baltimore, 3rd edition.

Perez-Lezaun A., Calafell F., Mateu E., Comas D., Bosch E. and Bertranpetit J. 1997. Allele frequencies for 20 microsatellites in a worldwide population survey. Human Hered., **47**: 189-196.

Price A. L., Patterson N. J., Plenge R. M., Weinblatt M. E., Shadick N. A. and Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association analysis. Nature Genet., **38**: 904-909.Pritchard J. K., Stephens M., Rosenberg N. A. and Donnelly P. 2000. Association mapping in structured populations. American J. Human Genet., **67**: 170-181.

Rabinowitz D. 1997. A transmission disequilibrium test for quantitative trait loci. Human Hered., **47**(6): 342-350.

Risch N. and Merikangas K. 1996. The future of genetic studies of complex human disorders. Science, **273**: 1516-1517.

Sasieni P. D. 1997. From genotypes to genes: doubling the sample size. Biometrics, **53**(4): 1253-1261.

Schaid D. and Sommer S. 1993. Genotype relative risks: methods for design and analysis of candidate-gene association studies. American J. Human Genet., **53**(5): 1114.

Schaid D. J. 1996. General score tests for associations of genetic markers with disease using cases and their parents. Genetic Epidemiol., **13**(5): 423-449.

Slatkin M. 1991. Inbreeding coefficients and coalescence times. Genetical Res., **58**(02): 167-175.

Slatkin M. 1999. Disequilibrium mapping of a quantitative-trait locus in an expanding population. American J. Human Genet., **64**(6): 1765-1773.

Spielman R. S. and Ewens W. J. 1998. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. American J. Human Genet., **62**(2): 450-458.

Spielman R. S., McGinnis R. E. and Ewens W. J. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (iddm). American J. Human Genet., **52**: 506-516.

Stephens J. C., Schneider J. A., Tanguay D. A., Choi J., Acharya T., Stanley S. E., Jiang R., Messer C. J., Chew A., Han J.-H., et al. 2001. Haplotype variation and linkage disequilibrium in 313 human genes. Science, **293**(5529): 489-493.

Tabor H. K., Risch N. J. and Myers R. M. 2002. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nature Reviews Genet., **3**(5): 391-397.

Terpstra T. 1952. The asymptotic normality and consistency of kendalls test against trend, when ties are present in one ranking. Indagationes Mathematicae, **14**(1952): 327-333.

Terwilliger J. and Goring H. 2000. Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Human Biol., **72**(1): 63-132.

Thangaraj J., Sridhar V., Kivisild T., Reddy A. G., Chaubey G., Singh V. K., Kaur S., Agarawal P., Rai A., Gupta

J., Mallick C. B., Kumar N., Velavan T. P., Suganthan R., Udaykumar D., Kumar R., Mishra R., Khan A., Annapurna C. and Singh L. 2005. Different population histories of the mundari- and mon-khmer-speaking austro-asiatic tribes inferred from the mtdna 9-bp deletion/insertion polymorphism in indian populations. Human Genet., **116**: 506-517.

Wacholder S., Rothman N. and Caporaso N. 2000. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J. National Cancer Institute, **92**(14): 1151-1158.

Waldman I. D., Robinson B. F. and Rowe D. C. 1999. A logistic regression based extension of the tdt for continuous and categorical traits. Annals Human Genet., **63**(4): 329-340.

Weinberg C., Wilcox A. and Lie R. 1998. A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. American J. Human Genet., **62**(4): 969-978.

Weir B. S. 1996. Genetic Data Analysis 2 : Methods for discrete population genetic data. Sinauer, Sunderland, MA, 3rd edition.

Weiss K. M. and Terwilliger J. D. 2000. How many diseases does it take to map a gene with SNPs? Nature Genet., **26**(2): 151-158.

Wheeler E. and Cordell H. J. 2007. Quantitative trait association in parent off-spring trios: Extension of case/pseudocontrol method and comparison of prospective and retrospective approaches. Genetic Epidemiol., **31**(8): 813-833.

Whittaker J. C. and Lewis C. M. 1999. Power comparisons of the transmission/disequilibrium test and sib-transmission/disequilibrium-test statistics. American J. Human Genet., **65**(2): 578.