89

Haseman and Elston Revisited

R. C. Elston, S. Buxbaum, K .B. Jacobs, J.M. Olson
Case Western Reserve University, Cleveland, Ohio, U.S.A.

Haseman and Elston (Behav Genet, 2:3-19, 1972) proposed a method to detect quantitative trait loci by linkage to a marker. The squared sib-pair trait difference is regressed on the proportion of marker alleles the pair is estimated to share identical by descent: a significantly negative regression coefficient suggests linkage. It has been shown that a maximum likelihood method that models the sib-pair covariance directly has more power. This increase in power can also be obtained using the Haseman and Elston regression procedure by changing the dependant variable from the squared difference to the product of the sibs' trait values. Multiple sibs in a sibship can be accommodated by weighted least squares, allowing for correlations between pairs of products. Multiple markers can be the basis of a multipoint analysis, using the fast method due to Fulker et al. (Am J Hum Genet, 56:1224-1233, 1995). Multiple trait loci, including epistatic interactions, involve only multiple linear regression. Multivariate traits can use the method of Amos et al (Am J Hum Genet, 47: 247-254) to find the linear function of the traits that maximizes the evidence for linkage; an approximation due to Mangin et al (Biometrics, 54:88-99, 1998) makes this method computationally fast. The same general scheme can be used to study affected sib pairs, testing whether their identity by descent sharing probabilities are greater than expected in the absence of linkage. Results of simulation studies will be presented that investigate Type I error and power.

90

Detecting influential families in linkage and association studies

S.B. Bull, D. Pinnaduwage, C. Toma.
Samuel Lunenfeld Research Institute, University of Toronto, Toronto, Ontario, Canada.

Using jackknife methods, we formulate influence measures for a nuclear family, and for an individual within a family, for inference concerning genetic linkage and association. Different graphical displays allow different aspects of the data to emerge more clearly, particularly for multiple markers. These include plots of case influence measures ordered by family characteristics, such as diagnostic subclass and ethnicity, for each of several markers. Simultaneous plots for several pedigrees ordered by marker location emphasize marker similarities and differences. The methods are illustrated in several datasets from studies of complex disease that include multiple markers and some covariate information.

Case influence measures are data-driven and hence exploratory in nature, and may be best suited to sensitivity analyses. They can facilitate the detection of unusual families and individuals for further examination, and can indicate heterogeneity among pedigrees and covariate-defined groups of pedigrees.

Supported by the Natural Sciences and Engineering Research Council of Canada.

91

Susceptibility to celiac disease in Tunisian children and GM immunoglobulin allotypes

F. Bouguerra[1], J.M. Dugoujon[2], M.C. Babron[3], L. Greco[4] J. Clayton[2], F Khaldi F[1], A. Debbabi[1], B. Bennaceur[1] and F. Clerget-Darpoux[3]
[1]Hôpital d'Enfants de Tunis, 1007 Jebbabi - Tunis, Tunisia, [2]CNRS UPR 8291, Toulouse , France, [3] INSERM U155, Paris, France, [4]University of Naples, Naples, Italy.

Celiac disease is a malabsorption disorder of the small intestine resulting from ingestion of gluten. The immunogenetic component is clearly demonstrated by the association of the disease with human leukocyte antigens (HLA). Among the other candidate genes are the GM allotypes, which are the markers of the constant parts of heavy chains of the subclasses IgG1, IgG2 and IgG3.

GM immunoglobulin allotypes have been analyzed in 131 unrelated Tunisian children with celiac disease. All patients and their parents were tested for G1M(1, 2, 3, 17), G2M(23), G3M(5, 6, 10, 11, 13, 14, 15, 16, 21, 24, 28) by the classical hemagglutination method. Genotypes and haplotypes were deduced from phenotypes in patients and their parents. Transmission disequilibrium tests have been performed in 79 informative families. The GM*3;..;5* haplotype was transmitted more often (23) than not (8) by heterozygous parents ($\chi^2 = 7.26$ and $p = 0.007$). This difference remained significant after correction for multiple testing.

This study provides evidence for association and linkage between GM and celiac disease. It suggests that GM or genes close to GM play a role in the development of the disease.

92

A Multivariate and Multilocus Variance Components Approach using Structural Relationships to Assess Quantitative Trait Linkage via SEGPATH

M.A. Province, T. Rice, I.B. Borecki, C. Gu, and D.C. Rao
Washington Univ., St. Louis, USA

A general purpose model and a flexible computer program performing path and segregation analysis jointly, has been extended to cover "model-free" ro-

bust linkage analysis based on IBD sharing estimates. The SEGPATH model, based upon the variance components approach, can be used to analyze linkage to a single marker or to perform multipoint linkage analysis, with a single phenotype or multivariate vector of phenotypes, in pedigrees which can be defined as arbitrarily complex. The computer program retains its flexible model-specification syntax so that SEGPATH models can perform segregation analysis, path analysis, linkage analysis or combinations using any user-specified model. SEGPATH models can incorporate environmental or other measured covariate fixed effects (including measured genotypes), genotype-specific covariate effects, population heterogeneity models, repeated-measures models, longitudinal models, auto-regressive models, developmental models, gene by environment interaction models, etc., with or without linkage components. Data analyzed via SEGPATH can have any missing value structure, with entire individuals missing, or missing on one or more measurements. Corrections for ascertainment can be made on a vector of phenotypes and/or other measures. Because the model specification syntax is general, SEGPATH can also be used in non-genetic applications where there is a hierarchical structure, such as longitudinal, repeated-measures, time series, or nested models. SEGPATH also comes with a set of SAS Macros which allow easy interface between the two, so that SEGPATH analyses can be conducted and managed in the very rich and flexible SAS environment. Specific models are provided as well as some comparisons with other linkage analysis programs, which demonstrate that the extended SEGPATH approach appears to perform quite well.

93

Uses of Optimum Experimental Design theory in the design of gene mapping studies

A.A. Todorov, P.A.F. Madden and A.C. Heath.
Washington University, St. Louis, USA.

The theory of optimum experimental design is concerned with determining the most efficient allocation of a fixed number of observations to experimental conditions defined by the levels of one or more treatments. In the present, we illustrate the use of this theory in the planning of a study of gene-environment interactions. We consider the case where nuclear families are selected from strata defined by the level of exposure to an environmental factor. We assume that the probability of affection where, given the genotype (g) and the level of exposure (ε), the probability of affection is given by $\exp(\alpha_g + \beta_g\varepsilon) / \{1 + \exp(\alpha_g + \beta_g\varepsilon)\}$. For the case of sib-pairs, we determined that the optimum sampling scheme selects pairs around three nodes: both sibs with low exposure levels, both sibs with high exposure levels, and sibs highly discordant with respect to exposure. We surmise that the allocation scheme naturally gravitated toward one where genotypes were predicted with high probability. Other cells are represented to insure model identification and for the estimation of parameters associated with the genotype AB. We also observed that the information matrices were either singular or nearly so when sib pairs are selected where both sibs have the same level of exposure. This held regardless of how many levels of exposure were represented.

Supported in part by grants AA-00728 and DA-00272.

94

A Meta-Analysis Methodology for Combining Results of Family-Based Genetic Association Studies

Chi Gu, M.A. Province and D.C. Rao
Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA.

The methods of genetic association analysis have attracted much attention recently for their potential advantage in achieving better statistical power for mapping complex diseases of small effects, and for their ability to enhance mapping resolution in genomic scans. Family-based association tests that rely on linkage disequilibrium have shown noticeable enhancement in power and large-scale testing by association analysis appears to be a popular design. As several genome-wide association studies are in planning stages, methodological development of quantitative methods for systematic analysis of such studies is well in order. Herein, we develop a meta-analysis methodology to pool the results of family-based association studies of different study designs and correlated phenotypes, under the assumption that individual studies share the same underlying genetic model of the disease. The random effects model of DerSimonian and Laird is applied to allow for variation among the true genetic effects of individual studies, and the popular transmission-disequilibrium-test (TDT) and its variants are used as a prototype of the statistics for pooling. A heterogeneity test is given to determine poolability and the model for pooling (e.g., fixed effects or random effects). Both the case of pooling published statistics and of pooling results from individual centers of a collaborative study are considered. Monte Carlo methods are used to estimate the posterior distribution of the parameters of interest when closed-form formula is not available. Important practical issues when performing meta-analysis of genetic studies, such as heterogeneity, publication bias, and pre- and post-processing of individual studies, are discussed in details.

95

Linkage of Dyslexia to Chromosome 6p23-p21.3 Not Confirmed

L.L. Field and B.J. Kaplan
Univ. of Calgary and Alberta Children's Hospital, Calgary, Canada

Previous studies have suggested that a locus conferring susceptibility to specific reading disability (dyslexia) resides on chromosome 6p, in the HLA region at 6p21.3 or telomeric to that region at 6p23-p21.3. We have investigated 79 nuclear and extended families having a minimum of two siblings affected with phonological coding disorder, the most common form of dyslexia (617 people typed, 294 affected), and performed linkage analysis with the same genetic markers that were reported to be linked in those previous studies. No evidence for linkage was found using parametric lod score analysis or nonparametric affected sibpair methods. However, using the nonparametric affected-pedigree-member method (APM program), significant evidence for increased marker allele sharing in affected individuals (reflecting either linkage or association) was detected with D6S299 and TNFB — but only when using published marker allele frequencies with weighting of rarer alleles. Results were not significant when marker allele frequencies estimated from parents (two per pedigree) were employed in the APM analyses. Furthermore, results were not significant using the more robust nonparametric SIMIBD method with either published or parental marker frequencies. Finally, association analysis using the AFBAC program to compare marker alleles transmitted and not transmitted from parents to affected children showed no evidence for association with any marker. We conclude that the affected-pedigree-member (APM) method should be employed only with extreme caution, as it appears to have generated false positive results in our study and possibly those of others. In summary, using a larger dataset than examined in any previous investigation, we find no evidence for linkage or association between dyslexia and chromosome 6p markers.

Supported by Alberta Mental Health Research and the Canadian Genetic Diseases Network of Centres of Excellence programme.

96

Increasing Incidence of type 1 diabetes - A Role for Genes?

P. Onkamo[1,2], J. Pitkäniemi[1,2], E. Tuomilehto-Wolf[2], J. Tuomilehto[2] and E. Arjas[1,2]
[1]Rolf Nevanlinna Institute, Univ. Helsinki, Finland
[2]National Public Health Institute, Helsinki, Finland

Type I diabetes incidence is increasing fast in many populations. The reasons for the increasing incidence are not known, although an increase in the penetrance of the diabetes associated alleles, through changes in the environment, might be the plausible mechanism for the observed increase in incidence. After the introduction of insulin treatment in the 1920s, an increase in the pool of genetically susceptible individuals has been suggested as one cause contributing to the increase in the incidence of type 1 diabetes.

Based on basic population genetic principles, we show that there is no inherent tendency of diabetic allele frequency to increase. Only assuming biological selection mechanism favouring diabetic allele(s), as suggested by observations on transmission distortion of type 1 diabetes associated HLA-alleles may the gene pool change. A simple genetic model for genotype frequency and incidence change, in the presence of transmission distortion of susceptibility alleles, has been constructed. In the model, the HLA-alleles have been divided into two categories, the alleles showing transmission distortion and conferring susceptibility to diabetes, and those with Mendelian inheritance. Penetrances are determined according to genotype and age group. Theoretical behaviour of the model was explored, with transmission probability varying from 0.5 to 0.8, and differing penetrance parameters. The gene pool change appears to be very slow with reasonably low values of transmission distortion. As a consequence, transmission distortion can increase the incidence only gradually. Hence, the observed steep increase in the incidence of type I diabetes cannot be properly explained with transmission distortion effect. Other genetic effects and thus far unidentified environmental factors must play an important role in the increase.

97

Transmission of alleles at the HLA A, B and DR loci in Finnish families with IDDM

J.Pitkäniemi[1,2], P. Onkamo[1,2], E. Arjas[1,2], E. Tuomilehto-Wolf[2], J. Tuomilehto[2] and the DiMe Study Group
[1]Univ. Helsinki, Finland, [2]National Public Health Institute, Finland

We studied transmission of alleles at the HLA A, B and DR loci in 801 Finnish families ascertained through a child with IDDM under the age 15 and time of diagnosis in the recruitment period from September 1986 to April 1989. Ascertainment was corrected using transmissions from the parent to the offspring born after the proband. Proband was the first born child dignosed with IDDM during the recruitment period. Statistical analysis was based on the log-linear model. A global χ^2-test for single allele effects and standard normal test

for each allele effect were performed. There was some evidence of transmission distortion at the A locus both in the maternal allele (χ^2=40.1, 26 df., p=0.04) and paternal allele (χ^2=37.4, 24 df.,p=0.04). Paternal A26 (p<0.01) and maternal A32 (p<0.01) were transmitted less than expected. There was a statistically significant distortion in the maternal allele at the B locus (χ^2=97.6, 63 df. ,p<0.01). There was some evidence of transmission distortion in the paternal allele at the B locus (χ^2=71.9 ,54 df., p=0.05). Both paternal B38 (p<0.01) and maternal B62 (p<0.01) were transmitted less than expected. There was no evidence of transmission distortion at the DR locus of either maternal alleles (χ^2=42.0, 39 df. , p=0.34) or paternal alleles (χ^2=43.4, 29 df., p=0.25). Interestingly, maternal DR2 was statistically significantly differently transmitted (p<0.01) and none of the paternal single allele effects were statistically significant. Results from this study support hypothesis of transmission distortion in the HLA region in the Finnish IDDM families. If the "protective" allele, DR2, is indeed transmitted less than expected and therefore replaced with alleles with higher risk, one can speculate that this phenomen is one possible mechanism for increasing incidence of IDDM in Finland.

98

Search for Statistical Interactions in Data from Genome-Wide Screen for Asthma Susceptibility Loci in Three U.S. Populations

Collaborative Study on the Genetics of Asthma
University of Chicago University of Maryland, University of Minnesota, Johns Hopkins University, and NHLBI
(Presenting author: Jianfeng Xu)

A genome-wide screen for asthma susceptibility loci using 552 affected sib-pairs from three racial groups within the U.S. identified multiple regions with nominal evidence for linkage. To examine the evidence for statistical interactions among these regions, we estimated the correlation coefficient matrix for the multipoint NPL scores from those regions with at least nominal (p<0.05) evidence for linkage. Several correlation coefficients were significant even after correcting for the number of comparisons, including 10q and 2q in the Hispanic population and 11q and 2q in the African American population. In the genome wide screen, a different region provided the strongest evidence for linkage in each of the three racial/ethnic groups. The evidence for linkage to these regions was taken into account by weighting the contributions from families according to their evidence for linkage (multipoint NPL score) at the given location. In this conditional analysis, chromosomal regions 1p, 10q and 2q show the most consistent increase in evidence for linkage in all ethnic groups. While these analyses are clearly secondary, a systematic examination of the evidence for statistical interactions between regions may provide useful additional information for mapping genes for complex traits.

99

Apolipoprotein E E4 is associated with dementia and peripheral neuropathy in HIV infection

E.H. Corder[1,2], K. Robertson[3], L. Lannfelt[4], N. Bogdanovic[4], G. Eggertsen[5], J. Wilkins[3], C. Hall[3.]
[1]Genetic Epidemiology, Odense Univ., Denmark, [2]Demographic Studies, Duke Univ., USA, [3]Neurology, Univ. of North Carolina, Chapel Hill, USA, [4]Clinical Neuroscience and [5]Clinical Chemistry, Huddinge Univ. Hospital, Sweden.

Activation of microglia near amyloid plaques in Alzheimer's disease (AD) and microglia infected by HIV-1 suggested that shared inflammatory processes might be important for both disorders. We investigated whether the major known inherited risk factor for AD — the E4 isoform for apolipoprotein E (apoE) implicated in inflammation and lack of neuronal repair — might also predict more cognitive-motor symptoms in HIV-1 infection. ApoE isoform information was available for 44 subjects in an HIV-1 cohort previously prospectively investigated to define the neurologic natural history of the infection; 207 semiannual examinations over 2.5 years. Subjects were most often young white homosexual men. Eleven (25%) were E4(+), near the expected frequency for Caucasians. None was homozygous E4. E4(+) subjects in an HIV-1 cohort were more often found to be demented (30% vs 15% subjects; 13% vs 3% exams) and to have peripheral neuropathy (70% vs 39% subjects; 42% vs 14% exams) over 2.5 years. These differences, and those for AIDS dementia complex stage, were highly significant in repeated measures statistical models adjusted for CD4(+) count (p<0.001). Neither attrition of CD4+ cells nor progression to AIDS was strongly associated with E4. These results support the belief that AD and the cognitive-motor syndrome found HIV-1 infection represent in part the consequences of apoE-related inflammatory processes.

100

The hunt for candidate genes in families with a missing parent

A. C. L. Cervino[1], A.V. S. Hill[1] and P. Donnelly[2]
[1] The Wellcome Trust Centre for Human Genetics, Oxford University, UK
[2] Department of Statistics, Oxford University, UK

When investigating association between a disease trait and a candidate gene, one possible approach is to use family based association tests. These family based tests have the major advantage of being resistant to population stratification. One problem of these tests is that they are conditional on the parental genotype and therefore rely on the availability of both parent's genotype, which is often extremely difficult, if not impossible, to achieve. In the recent literature alternative tests have been presented to overcome this missing data problem by either working with 'horizontal' transmissions and using healthy sibs instead of the parents or by reconstructing the missing parent and still work with the 'vertical' transmissions.

A new likelihood ratio test the LRAT, Likelihood Ratio Association Test, will be presented. Some of these new tests and available programs will be compared on simulated data sets. These comparisons will particularly focus on power to detect association, resistance to population stratification and resistance to misclassification in the biallelic case.

101 [Invited Speaker]

Lessons from a "Second-Generation"Genome Screen for Type 1 Diabetes

R. S. Spielman
Univ. Pennsylvania, Philadelphia, USA

As a consortium of 4 groups, we have carried out a "second-generation" screen of the genome for susceptibility to IDDM. Our results indicate some of the difficulties frequently encountered in work of this kind. The maximum sample size is 679 ASPs, appreciably more than in previous screens of this type. As expected, we found strong evidence for linkage in the HLA region (multipoint lod [mlod] was >30). We also found suggestive evidence for linkage in a region on chromosome 1q, not previously reported, where the maximum mlod is 3.3 and the evidence for a contribution is stronger than that for any other gene or region, except HLA.

Apart from HLA, the only regions containing previously reported loci where we found mlods >1 are on chromosomes 2q and 6q; the increased mlods are very modest, however. These mlods are 1.07 (near IDDM 7, 12, and 13), 1.80 (for IDDM5), 1.3 (for IDDM8), and 1.7-2.3 (depending on map distance from HLA), for IDDM15. For the remaining IDDM loci, the mlods were <1; thus we found no support for IDDM susceptibility genes in these regions, and only modest support on chromosomes 2q and 6q.

In summary, excluding HLA, we found little or no support for linkage in most (6) of the previously reported chromosomal regions. We did, however, find suggestive evidence for linkage on chromosome 1q, in a region not previously reported. Additional results from large samples of families, and other approaches, will be required to establish which, if any, of the proposed "linkages" with IDDM (including that for chromosome 1q) actually reflect the presence of genes contributing to IDDM susceptibility.

102 [Invited Speaker]

Genetic epidemiology of infectious diseases.

Laurent Abel
INSERM U.436, Paris, France.

There is now accumulating evidence that genetic factors play a major role in the response to various infectious pathogens in humans. The genetic epidemiology of human infectious diseases raises the same difficulties as similar studies performed on other multifactorial traits, with some specificities: (1) the contact with the infectious agent is required to get the disease (but often not sufficient), (2) the environmental factors influencing the risk of contamination are generally well known, (3) for a same pathology, it is of particular interest to study a large number of complementary traits as clinical phenotypes, biological phenotypes measuring infection intensities (eg fecal egg counts in schistosomiasis, and immunological phenotypes (eg levels of antibodies or cytokines involved in the immune response). The usual statistical methods of genetic epidemiology are used to investigate the role of genetic factors in the control of these phenotypes. Among these methods, parametric approaches (segregation analysis and linkage analysis by the lod-score method) are based on maximum likelihood principles and need to construct a model specifying the relationship between the phenotype and factors that can be involved in its expression, mainly a putative gene and environmental covariates. On the other hand, nonparametric approaches (association studies, sib-pair methods...) allow to test the association and/or the linkage between a phenotype and a genetic marker without specifying the model relating these two factors. As detailed in the presentation, the use of these different methods allowed to begin the genetic dissection of human infectious diseases such as schistosomiasis and leprosy.