Genetic Epidemiology 17:190–229 (1999)

ABSTRACTS FROM THE

EIGHTH ANNUAL MEETING OF THE INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY

ST. LOUIS, MISSOURI
SEPTEMBER 9–10, 1999

With financial support from:
Integrated DNA Technologies
Division of Biostatistics,
Washington University School of Medicine

1

The Additive Gamma Genetic Model For Linkage Analysis of Age-of-Onset Variation for Complex Diseases.

Hongzhe Li
UC Davis School of Medicine, Davis, CA 95616

Analysis of age of onset is a key factor in the linkage analysis of many complex diseases. Affected individuals with different ages of onset may be the result of different genetic etiologies, and unaffected individuals may develop disease later. Current methods in linkage analysis are mainly concentrated on affected family members, and age of onset information is either ignored or is taken into account by specifying age-dependent penetrances for liability classes. In this paper, multiple markers are used to calculate the probability distribution of the inheritance vector at the locus of interest. Using this probability distribution, we define genetic frailties for each individuals within a nuclear family, and conditioning on these frailties, we use the proportional hazards model to model the risk of developing disease. We define a hazard ratio statistic to measure familial aggregation, and show that the test of linkage can be formulated as a test of zero variance of the additive gamma frailty. Two LOD score statistics are proposed for testing linkage. Results from simulation study indicated that the proposed methods can be more powerful than the NPL score statistic of GENEHUNTER.

2

In Search of Genes for Type 2 Diabetes in West Africa: the Design and Results of the First Phase of the Africa America Diabetes Mellitus (AADM) Study: For the AADM Investigators.

Presenter Charles Rotimi, PhD
Howard University, College of Medicine,
Washington, DC

The purpose of this study is to map type 2 diabetes susceptibility genes in West African ancestral populations of African-Americans, through an international collaboration between West African and US investigators. Affected sib-pairs along with unaffected spouse

controls are being enrolled from two study sites in Ghana (Accra and Kumasi) and three in Nigeria (Enugu, Ibadan and Lagos). Participants are invited to study clinics to obtain detailed epidemiologic, family, and medical history information. Blood samples are drawn to measure glucose, insulin, c-peptide, total cholesterol, LDL, HDL, triglycerides, albumin, creatinine, urea, uric acid, total calcium, and to detect autoantibodies to glutamatic acid decarboxylase. High quality DNA is isolated for genome analyses. Quality control procedures for isolated DNA, including PCR amplification with a core set of genomic markers, have been implemented and routinely performed by the coordinating center at Howard University. With full informed consent, 162 individuals from 78 families have been enrolled and examined since field activities began in June of 1997. At the end of the third year (September, 2000), the AADM study will have enrolled 400 affected sib-pairs and 200 spouse controls. This will provide a comprehensive epidemiologic and genetic resource for a powerful genome-wide search for West African susceptibility genes for type 2 diabetes. This presentation describes the design and results of the first phase of the AADM study.

3

Glutathione S-transferases and susceptibility to peripheral arterial disease (PAD) given exposure to cigarette smoking: The ARIC Study

R. Li,[1] A.R. Folsom,[3] R. Sharrett,[4] D. Couper[1], M. Bray,[2] H.A. Tyroler[1]
[1]Univ. North Carolina, Chapel Hill, NC, USA; [2]Univ. Texas, Houston, USA; [3]Univ. Minnesota, Minneapolis, MN, USA; [4]NIH/NHLBI, USA

Glutathione S-transferases M1 or T1 (GSTM1/GSTT1) affect the body's ability to detoxify or to activate chemicals in cigarette smoke. Cigarette smoking increases the risk of PAD. We conducted a cross-sectional study to evaluate a hypothesized interaction between the genetic polymorphisms, GSTM1 and T1, and cigarette smoking on the risk of PAD in the ARIC study. A stratified-random sample including 212 PAD cases (ankle-brachial index <0.9 in men or <0.85 in women) and 1,277 non-cases were selected from the ARIC cohort of 12,041 middle-aged participants free of CHD, transient ischemic attack and stroke at baseline (1986–89). Smoking was more prevalent among PAD cases than non-cases (45% vs. 25%, current smoking, and 58% vs. 29%, smoked 20+ pack-years (pkyrs)). Overall, the difference in the proportions of GSTM1-0 and GSTT1-0 (the homozygous deletion genotype) was not statistically significant between cases and non-cases (44% vs. 41% and 28% vs. 18%). There was no indication that GSTM1 modified the association between smoking and PAD. In contrast, GSTT1-1 (the non-deletion genotype) was associated with greater risk of PAD given exposure to smoking after adjustment for other risk factors. The ORs (95% CIs) of PAD were 3.6 (1.4, 9.0) for current smoking and 5.0 (1.9, 13.0) for 20+ pkyrs in individuals with GSTT1-1. Given GSTT1-0, the ORs were 0.8 (0.2, 2.8) for current smoking and 0.6 (0.1, 2.1) for 20+ pkyrs. The interaction was significant ($p<0.05$) on the additive scale for current smoking, and on both the additive and multiplicative scales for 20+ pkyrs. Among non-smokers, GSTT1-1 was not associated with PAD. The results suggest that the GSTT1-1 polymorphism may be a susceptibility factor modifying the risk of PAD associated with cigarette smoking.

4

Likelihood Alternatives to the TDT

M.M.Iles, D.T.Bishop
ICRF Genetic Epidemiology, Leeds, UK

The TDT (Spielman *et al.* 1993, AJHG 52:506–516) tests for both association and linkage simultaneously. Its main advantage is that it protects against population effects by conditioning on parental genotype. However, information from homozygous parents is wasted. Although homozygous parents provide no information about linkage, they provide evidence of population association.

Various tests are proposed using the same nuclear family data as the TDT, but including homozygous parents. It is shown that extra power is gained by using all parents to detect deviation from Hardy Weinberg. For instance, one maximum likelihood ratio test is constructed using all the data from nuclear families with affected offspring, making the assumption that transmissions from parents are independent (as does the TDT). Using this test, power can be increased by up to 10%, although results depend on modes of inheritance, population frequencies, etc.

It is shown that association at a marker leads to more heterozygous parents of affected children than expected under Hardy Weinberg. This is detected by tests which uses all parents, but not the TDT. However more heterozygotes means extra power for the TDT.

The main concern is that spurious association caused by either stratification or admixture may lead to an excess of false positive results. However, any increase in type 1 error caused by population effects is minimal. In the test above, under realistic admixture models, the type 1 error is increased from a nominal value of 5% to no more than 8%, under stratification the same test becomes conservative.

5

A regression-based transmission/disequilibrium test for binary traits using logit link function

V. George[1], H.K. Tiwari[2]
[1]Medical College of Wisconsin, Milwaukee, WI, USA; [2]Case Western Reserve University, Cleveland, OH, USA.

The transmission/disequilibrium test (TDT), originally proposed by Spielman, et al for binary traits, is a powerful method for detecting linkage between a marker locus and a disease locus in the presence of allelic association. The TDT uses information on the parent-to-offspring transmission status of the associated allele at the maker locus to assess linkage or association in the presence of the other. Recently, George, et al proposed a regression-based TDT for linkage between a marker locus and a quantitative trait locus by modeling the trait as the dependent variable and the transmission status as one of the independent variables along with other predictors and confounders in a linear regression model. This test is very powerful, especially when the marker locus is close to the trait locus. In this presentation, we extend this idea to develop a TDT for linkage between a binary trait locus and an associated marker locus in nuclear families using linear mixed models with logit link function. The method allows for flexible correlation structure among the members of the nuclear family, and, allows us to estimate the effects of other relevant covariates simultaneously with the detection of linkage. We investigate the statistical power and validity of the test by simulating markers at various recombination fractions from the trait locus.

6

Multipoint extension of the WPC linkage method for large pedigrees with application to breast cancer and alcoholism familial data.

A. Zinn-Justin[1], A. Ziegler[2], L. Abel[1]
[1]INSERM U.436, Paris, France; [2]Institute of Medical Biometry & Epidemiology, Philipps-University of Marburg, Germany.

The non parametric (model free) method of linkage analysis WPC (Weighted Pairwise Correlation) proposed by Commenges [Genet Epidemiol 1994, 11:189–200] allows to analyze any kind of phenotype and to consider all pairs of relatives in large pedigrees. We recently extended the WPC method to introduce Identical by Descent (IBD) information [Genet Epidemiol 1999, 17:35–50]. Here, we propose a multipoint WPC method suitable for extended pedigrees with a large number of markers. For each pair of relative, the multipoint IBD sharing is computed using the SOLAR package as described by Almasy and Blangero [Am J Hum Genet 1998, 62:1198–1211]. A valid WPC test is obtained using a new within-family Monte Carlo (MC) permutation procedure in which phenotypic resemblances are permuted with IBD sharing for pairs of relatives from the same degree. This procedure allows to compute fast MC p-values and an empirical variance for the WPC statistic. Application of the method to the 214 pedigrees from the Breast Cancer Linkage Consortium provided for the Genetic Analysis Workshop (GAW) 9 shows that multipoint WPC statistic values were not far from lod score values obtained by the classical parametric linkage method. The multipoint WPC method was also used to analyze the familial COGA data on alcoholism released for GAW11, and allowed a better specification of the linkage results previously obtained within the chromosome 4 region [Zinn-Justin and Abel, Genet Epidemiol, in press].

7

Evaluation of average lod score and estimated size of modeled effects in a variance components approach

Yin Yao Shugart[1], Jeff O'Connell[2], Alexander Wilson[3]
[1]Johns Hopkins University, Baltimore, [2]University of Pittsburgh, Pittsburgh, [3]NIH/NHGRI, Baltimore, USA

The variance components approach implemented in the Beta version of GENEHUNTER 2 (GH2) was used to determine average lod scores (ALS) and to estimate the size of single locus (SL) and polygenic (PG) components for four different genetic models. Each generating model contained 100 families of sibship size 5. For each model, 2000 replicates were simulated using G.A.S.P. (Wilson et al 1997). In model 1, 30% of the variance was due to a major locus and 70% was due to random environment. In models 2–4, 30% of the variance was attributed to a major locus, 40% to random environment, and 30% to other factors. More specifically, model 2 contained an unlinked second major locus, model 3 had a polygenic component and model 4 had a common environmental component. The ALS generated under model 1 was 1.4, compared to 1.74 and 1.72 for model 2 and 3. The ALS nearly doubled (2.57) when a common sibship environment was simulated in model 4. In model 1–3, GH2 underestimated the variance attributed to the SL component, while inflating the variance attributed to the PG component . In model 4, however, the environmental variance was also underestimated (11%) while the PG variance was severely overestimated (64%). The variance components approach in GH2 appears to underestimate variance attributed to a SL component and overestimate variance attributed to a PG component in all four models regardless the source of variation.

8

Bias and efficiency in family-matched gene association studies: Conditional, prospective, retrospective, and joint likelihoods

P Kraft, DC Thomas
University of Southern California, Los Angeles, CA, USA.

We revisit the standard conditional likelihood for stratum-matched case-control studies and consider three alternatives that may be more appropriate for family-matched gene association studies: First, the ascertainment-corrected prospective likelihood, that is, Pr(D|G,A); second, the retrospective likelihood, Pr(G|D); and third, the ascertainment-corrected joint likelihood, Pr(D,G|A). We show that these likelihoods provide unbiased estimators of genetic relative risk parameters, as well as population allele frequencies and baseline risks. The parameter estimates based on the retrospective likelihood remain unbiased even when the ascertainment scheme cannot be modeled, as long as ascertainment only depends on families' phenotypes.

Despite the need to estimate additional parameters, the prospective, retrospective, and joint likelihoods can lead to considerable gains in efficiency relative to the standard conditional likelihood. This is true if baseline risks and allele frequencies can be assumed to be homogeneous.

In the presence of heterogeneity, however, the parameter estimates assuming homogeneity can be seriously biased. We discuss the extent of this problem and present a mixed models approach for providing consistent parameter estimates when baseline risks and allele frequencies are heterogeneous. We also compare the efficiency of the mixed-model prospective, retrospective, and joint likelihoods to the efficiency of standard conditional likelihood.

9

Effects of Stratification in the Analysis of Affected Sib-Pair Data.

Suzanne M. Leal and Jurg Ott
Laboratory of Statistical Genetics, The Rockefeller University, New York, NY

The benefits and costs of stratifying affected sib-pair (ASP) data are examined in three situations where. 1.) There is no difference in identity-by-descent (IBD) allele sharing between stratified and unstratified ASP data sets. 2.) There is an increase in IBD allele sharing in one of the stratified groups. 3.) The data is stratified based upon IBD allele sharing status at one locus, and the stratified ASPs are then analyzed for linkage at a second locus.

Where there is no difference in IBD sharing between strata, a penalty is always paid for stratifying the data. The loss of power to detect linkage in the stratified ASP data sets is due to multiple testing and the smaller sample size within individual strata.

In the case where etiologic heterogeneity (i.e. severity of phenotype, age of onset) represents genetic heterogeneity, the power to detect linkage can be increased by stratifying ASP data. This benefit is obtained where there is sufficient IBD allele sharing and sample sizes.

Once linkage has been established for a given locus, data can be stratified based upon IBD status at this locus, and tested at a second locus for linkage. In the case where the relative risk is in the vicinity of 1, the power to detect linkage at the second locus is always greater for the unstratified ASP data set. Even for values of the relative risk which sufficiently diverge from 1, with adequate sample sizes and IBD allele sharing, the benefits of stratifying ASP data are minimal.

Although stratification can be advantageous, it should be carried out with caution in order to avoid a potential loss in power to detect linkage.

10

Detection of a dominant major gene predisposing to HTLV-I infection in a Noir-Marron population of French Giuana.

S. Plancoulaine[1,2], A. Gessain[2], R.P. Buigues, P. Tortevoye[2], M. Joubert[4], I. Jeanne[3], A. Talarmin[3], G. de Thé[2], L. Abel[1].
[1]INSERM U 436, Paris, France ; [2]Institut Pasteur de Paris, France ; [3]Institut Pasteur de Guyane, Cayenne, Guyane française; [4]Centre de Santé, Maripasoula, Guyane française.

To investigate whether familial aggregation of HTLV-I seropositivity could be explained in part by genetic factors, we conducted a large genetic epidemiology survey in an endemic population of French Guiana. All families of two Noir-Marron villages were included representing 94 pedigrees with 1567 subjects (730 females, 837 males) of whom 151 (9.6%) were HTLV-I positive. The familial segregation analysis was performed with regressive logistic models which are testing for the presence of a major gene taking into account simultaneously covariates influencing HTLV-I infection (e.g. age and gender) and other sources of familial dependencies (e.g. due to virus transmission routes). Results show the presence of a dominant major gene predisposing to HTLV-I infection in addition to the expected familial correlations (mother-offspring, spouse-spouse) due to the transmission routes of the virus. The Mendelian transmission hypothesis was borderline rejected (p=0.03) when using a χ^2 distribution with 3 dfs. However, simulation studies show that, for this latter test, the use of an asymptotic distribution was inappropriate and the empirical p value was >0.05, indicating that HTLV-I familial data are compatible with the segregation of a dominant major gene. Under this

dominant model, about 2% of the population is predicted to be highly predisposed to HTLV-I infection, and seropositive children <10 years are genetic cases whereas most HTLV-I seropositive adults are sporadic cases. Linkage and association studies with genetic markers are ongoing to confirm and identify these genetic factors.

11

Multilocus linkage tests based on affected relative pairs

Heather J. Cordell, Geoffrey C. Wedig, Kevin B. Jacobs, Robert C. Elston
Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio, USA

For complex diseases, methods that take into account joint effects at several interacting loci are of particular interest. Conditioning on effects of disease loci at known locations can lead to increased power to detect effects at other loci. Moreover, use of multilocus models allows modeling of etiological mechanisms that may be involved in the disease. Here we present a method for simultaneously analyzing joint genetic effects at several linked or unlinked loci using affected relative pairs. We express the relative risk, λ_R, to a relative of an affected individual, in terms of the additive and epistatic components of variance at an arbitrary number of disease loci, and show how this can be used to fit a likelihood model to the multilocus identity-by-descent sharing among pairs of affected relatives in extended pedigrees. We implement the method using a stepwise strategy in which, given evidence of linkage to disease at *m*-1 locations on the genome, we calculate the conditional likelihood curve across the genome for an *m*th disease locus, using multipoint methods similar to those implemented in MAPMAKER/SIBS. In analysis of simulated data, our method shows increased power to detect non-multiplicative effects compared to single-locus methods. We apply the method to real data from a genome screen for type 1 diabetes, and find increased evidence for disease loci on chromosomes 6, 8, 11, 15, 16 and 18, when multilocus models are used.

12

Demography of Genotypes: The Italian Centenarians Study

Yashin[1] A.I., De Benedictis[2] G., Vaupel1 J,W., Andreev[1] K.F., Tan[1] Q, Iachine[3] I.A., Bonafe[4] M., Valensin[4] S., Carotenuto[5] L., C. Franceschi[4].
[1]Max Planck Inst. for Dem. Res., Rostock, Germany; [2]Dept. of Cell Biol., Univ. of Calabria, Rende, Italy; [3]Dept. of Stat. and Dem., Odense Univ., Denmark; [4]Dep. of Biomed. Sci., Univ. of Modena, Italy; [5]Dept. of Syst. Sci., Univ. of Calabria, Italy.

In population studies of aging, the data on genetic markers are often collected for individuals from different age groups. We show that the traditional methods of analysis of such data may lead to erroneous conclusions. We show how such factors as changes in the initial frequencies of candidate genes in subsequent cohorts or secular trends in cohort mortality may influence the results of an analysis. The unexpected results observed in genetic studies of centenarians are reviewed. The new approaches to the analysis of cross-sectional data on genetic markers are developed. These approaches allow us to combine genetic data with demographic information about population under study. The parametric and semiparametric methods of analysis which may control for observed and unobserved heterogeneity are discussed. The methods are tested on simulated data and then applied to the analysis of data on genetic markers obtained in the study of centenarians in Italy. The relative risks and initial frequencies of candidate genes from six loci (5 nuclear loci and mitochondrial DNA) are estimated. The interaction of these genes with covariates describing the effects of area and sex are tested. The results of statistical analysis of data and directions of further research are discussed.

13

Adding further power to the Haseman and Elston Method for detecting linkage

S. Shete, K. Jacobs, R. Elston
Department of Epidemiology and Biostatistics
Case Western Reserve University, Cleveland, OH.

Abstract: Haseman and Elston (H-E) proposed a robust test to detect linkage between a quantitative trait and a genetic marker. In their method the squared sib-pair trait difference is regressed on the estimated proportion of alleles shared identical by descent by the sib pairs. This method has recently been improved by changing the dependent variable from the squared difference to the mean corrected product of the sib-pair trait values, a significantly positive regression indicating linkage. Here we propose a further improvement of the H-E method in which generalized least squares estimators of the regressions are obtained separately for the squared sib-pair trait differences and the squared sib-pair mean corrected sums. The test for linkage is then based on a weighted average of these two estimators, the weights being inversely proportional to the residual variances obtained from these two different regressions. Control over the type I error probabilities and the increased power is investigated by simulation. A similar strategy of weighing estimators based on the residual variances is suggested for multivariate and binary traits.

14

A general conditional logistic regression model for affected-relative-pair linkage studies.

Jane M. Olson
Department of Epidemiology and Biostatistics, MetroHealth Campus, Case Western Reserve University, Cleveland, Ohio

Model-free lod-score methods are often employed to detect linkage between marker loci and common diseases, using samples of affected sib pairs (ASPs). Although extensions of the basic one-disease-locus model have been proposed that allow separate inclusion of additional disease loci, covariates, other types of affected relative pairs (ARPs) or discordant pairs, a common framework that can handle simultaneously all of these features has been lacking. I propose a conditional logistic parameterization that generalizes easily to include all of these features. The Risch single-locus ASP model is reparameterized in terms of the logarithms of the offspring and monozygotic twin relative risks $\lambda_o = e^{\beta 1}$, $\lambda_m = e^{\beta 2}$ and written in a form that is generally applicable to all types of ARPs, provided one can compute the ARP's marker allele-sharing probabilities. Multilocus models are parameterized in terms of the logarithms of joint-allele-sharing-specific relative risks and are easily written in general, multiplicative, additive, and mixed forms. Discordant pairs can be included with the addition of a single parameter. Covariates can be included in both the single- and multilocus models under the assumption that the covariate acts multiplicatively on the relative risks, i.e., additively on the log relative risks. Examples are given to demonstrate these extensions.

15

Trimmed-haplotype analysis for fine-scale mapping of complex traits

MacLean CJ and Martin RB
Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA

With the advent of myriad tightly linked microsatellite markers, we have an opportunity to extend linkage disequilibrium analysis from single markers to multiple marker haplotypes that have greater statistical power to disclose the presence of a disease susceptibility locus. Since any particular haplotype may well be rare, related haplotypes are also needed as evidence. Trimmed-haplotype analysis encompasses the entire cluster of partly related, partly unrelated haplotypes that would result from historical recombination and mutation in markers surrounding an ancestral disease locus. The probability of preservation of each category of trimmed haplotypes is a function of its inter-marker distances and of time, whereas the probability of random similarity in absence of a disease susceptibility locus is a product of haplotype frequencies. Therefore, a statistical test for the existence of a disease susceptibility locus is based on trimmed-haplotype category frequencies. Rather than standard parent-offspring triads, trimmed-haplotype analysis exploits multiplex pedigrees by considering the pattern of haplotype transmission and affection status in the pedigree as a whole. Trimmed-haplotype analysis has been incorporated in a software package called HAL, that is available to all investigators on the internet.

16

The Impact of Racial Admixture on Traditional Linkage Analysis

J.S. Barnholtz[1,2], M. de Andrade[1], R. Chakraborty[2]
[1]MD Anderson Cancer Center, Dept. of Epidemiology, Houston, TX; [2]UT School of Public Health, Houston, TX

Racially admixed families are routinely excluded from traditional (LOD score) linkage analysis or analyzed as racially homogeneous using the proband's race. Through simulation, we investigated the effect of admixture on the LOD score of two racial groups under various conditions.

Four-generation racially homogeneous and admixed families were simulated with 27 markers and two linked, bi-allelic disease loci. One locus was linked to a marker with correct allele frequencies for both groups and the other was linked to a marker with correct information for only one group. Two different types of admixture were tested: admixture within a family unit and a mixture of homogeneous families within a data set. The mixing was done at the founder level in three different proportions: 30/70, 50/50 and 70/30.

We observed that the LOD scores under both models of admixture were closest to the homogeneous family scores of the racial group having the highest mixing proportion. In the 50/50 case, the LOD scores were in-between the homogeneous scores. Random sampling of families or ascertainment of families with disease affection status did not affect this observation, nor did the mode of inheritance (dominant/recessive) or disease/marker allele frequencies.

LOD scores in admixed family data were further affected due to departures from Hardy-Weinberg expectations of genotypic frequencies and presence of linkage disequilibrium in admixed populations. Subroutines performing these tasks with relaxed assumptions are being implemented in current LOD score analysis programs. (Research supported by NCI-R25CA57730-07 and NIH-GM41399.)

17

Power Comparisons Between the TDT and Two Likelihood-Based Tests for Complex Traits

Slager SL[1], Huang J[2], Vieland VJ[1,3].
[1]Department of Preventive Medicine, Division of Biostatistics, [2]Department of Statistics and Actuarial Science, and [3]Department of Psychiatry, College of Medicine, University of Iowa, Iowa City.

Currently, linkage tests are used for genomic screens and association tests are used for fine mapping. However, advancements in molecular technology, promising maps on the order of a million markers using single nucleotide polymorphisms, could change this. We compare the statistical power of the TDT with two likelihood-based linkage tests, the classical LOD score, and a modified LOD score in which a linkage disequilibrium (LD) parameter is incorporated into the likelihood (LD-LOD). We hypothesize that the LD-LOD will have the greatest power of the three tests when LD is present, since the TDT is a score test based on a pseudolikelihood rather than the correct likelihood when multiplex families are ascertained, and the LOD score has previously been shown to be underestimated when LD is present (Clerget-Darpoux, 1982). We test this claim using a simulation study in which we generate ASP pedigrees under a range of genetic models, varying the genotypic relative risk (GRR) from 6 to 16. Since the likelihood-based tests require that a genetic model be specified, we compare the tests under two scenarios, (i) we assume the true genetic model in the analysis, and (ii) we compare the tests when the LD-LOD (LOD) is maximized over two wrong genetic models. Based on the generating models we considered, we find that the LD-LOD tends to have greater power than the TDT even when the genetic model is mis-specified and the results corrected for multiple tests. The most extreme difference occurs under the multiplicative and dominant models, for which the difference in power could be as high as 40% at maximum LD. Only for the additive model we find the TDT to have slightly greater power than the LD-LOD (7% difference). The LOD score provides the lowest power in the presence of LD for the range of GRR considered here.

18

High-density Genome Scans with Quantitative Traits – Resolving critical issues

S. Nath[1], J. Gardner[2], A. Aviv[2], N. Schork[1]
[1]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, [2]New Jersey School of Medicine and Dentistry, Newark, NJ

Linkage analyses for complex traits are beset by a number of problems, including 1) accommodation of covariates in the analyses, 2) possibility of heterogeneity, 3) correct and appropriate modeling of multiple genetic effects, and 4) assessment of the statistical significance of linkage results. We investigated each of these problems in the context of a variance component analysis of quantitative cellular phenotypes using a highly dense map of 7745 markers. Our aim was to identify loci for genes controlling two cellular phenotypes associated with store-operated calcium entry while considering some theoretical issues in linkage analyses. Stable phenotypes were collected for five densely mapped CEPH families. Evidence for linkage was obtained at several chromosomal regions. Our results suggest that variance component linkage analysis of densely mapped markers can reveal chromosomal regions likely to harbor loci influencing complex traits but care must be taken when conducting linkage-based genome scans to accommodate covariate, heterogeneity, multiple locus, and statistical significance.

19

Haplotyping in pedigrees via a genetic algorithm

Pradip Tapadar, Saurabh Ghosh, Partha P. Majumder
Indian Statistical Institute, Calcutta, India.

Genome-wide scanning for localization of disease genes has become very popular. Analyses of data on families generated by genome-wide scans necessitate the reconstruction of haplotypes for identifying, among others, the smallest genomic region containing the disease gene and also genotyping errors. Several systematic methods for haplotype reconstruction, either likelihood-based or rule-based, are currently available. We propose a rule-based method for haplotype reconstruction in pedigrees using a “genetic algorithm”, which a member of the class of computational algorithms that use certain principles of biological evolution to find optimal solutions to complex problems. Compared to the currently-available methods for haplotyping, the proposed method uses much fewer assumptions and is much less data-demanding. The optimality criterion used in the proposed method is the minimum number of recombinations over possible haplotype configurations of members of a pedigree. Using simulations, we show that the proposed method results in virtually error-free determination of multilocus haplotypes in extremely short computational time. It also provides multiple optimal haplotype configurations of a pedigree, if such multiple optima exist.

20

Evidence for further Breast Cancer susceptibility genes in addition to BRCA1 and BRCA2: the ABC study.

A.C. Antoniou[1], P.D. Pharoah[2,3], N.E. Day[3], B.J.Ponder[2], D.F. Easton[1]
[1]CRC Genetic Epidemiology Unit, IPH, [2]CRC Human Cancer Genetics Research Group, [3]Dept of Community Medicine, University of Cambridge, UK.

We used data from a population based series of breast cancer cases to investigate genetic models that can best explain familial breast cancer outside BRCA1 and BRCA2 families, and evaluate the evidence for risk modifiers in BRCA1/2 carriers. The dataset consisted of 1484 cases diagnosed with breast cancer under the age 55 between 1991 and 1996 in the region served by the Anglian cancer registry. BRCA1 and BRCA2 mutations were detected in 21 cases. The genetic models were constructed using information on the first-degree relatives with the computer program MENDEL. We estimated the effects of BRCA1, BRCA2, a third hypothetical gene "BRCA3" and a polygenic effect. The cancer incidence rate for an individual i was modelled as $\lambda_i(t) = \lambda_0(t)exp(G_i+P_i)$, where $\lambda_0(t)$ is the baseline incidence rate, G_i depends on the major genotype and P_i on the polygenotype of the individual. The Hypergeometric Polygenic Model[1] was used to approximate polygenic inheritance. The models were assessed by likelihood comparisons and by comparison of the observed number of mutations and affected relatives with the predicted numbers. The best fitting model for BRCA3 was a recessive model with a disease allele frequency 24% and penetrance 42% by age 70. However a polygenic model also fitted well. The estimated population frequencies for BRCA1 and BRCA2 mutations were 0.024% and 0.041% respectively. The fit was improved when the penetrance of BRCA1/2 was assumed to be lower than the Breast Cancer Linkage Consortium estimates. We are currently combining the results from this population series with those obtained from mutation testing in multiple case families to evaluate the effect of risk modifiers.

[1]Lange (1997), Genetics 147:1423–1430

21

An extended transmission/disequilibrium test (TDT) for two multi-allele marker loci

M. Knapp, A. Hahn
Institute for Medical Statistics, University of Bonn, Germany

The transmission/disequilibrium test (TDT) as introduced by Spielman et al. (Am J Hum Genet 52: 502–516, 1993) is a test for linkage between a complex disease and a diallelic marker. The original TDT has been extended by a number of authors in several aspects. Surprisingly, there currently is no practicable approach available which allows for simultaneous inclusion of more than one marker locus in the TDT. Wilson (Ann Hum Genet 61: 151–161, 1997) described a two-locus extension of the TDT, but she assumed that marker haplotypes can be observed in parents and children. Usually, however, the data will consist of marker genotypes at both loci, and the reconstruction of haplotypes will not be possible in some families, even if one is willing to assume that both marker loci are completely linked. The beta-release of the new version of the GENEHUNTER program contains a two-, three-, and four-locus extension of the TDT, but this program simply discards families for which phase is ambiguous. It can be shown that this approach may give biased results due to an inflation of the type I error rate.

Here we propose a two-locus extension of the TDT which is based on the conditional likelihood approach similar to the one considered by Sham and Curtis (Ann Hum Genet 59: 323–336, 1995) for a single marker locus. It will be assumed that the sample consists of simplex families and that all individuals in all families have been typed at both marker loci. Finally, the implementation of this two-locus TDT (2L-TDT) by means of SAS-IML software (SAS Institute, Cary, NC, 1995) is described.

22

Localization of prostate cancer aggressiveness genes

JS Witte[1], KAB Goddard[1], RC Elston[1], W Catalona[2], and the Prostate Cancer Genetics Study
[1]Case Western Reserve Univ., Cleveland, OH;
[2]Washington Univ., St. Louis, MO, USA

A measure of prostate cancer aggressiveness is the Gleason score, which may be a strong predictor of survival. Localizing genes that predispose men to present with higher Gleason scores—and who may require more active screening and treatment—is a significant research priority. Therefore, we conducted a genome-wide scan among 441 Caucasian men with prostate cancer (i.e., the equivalent of 282 sib-pairs). We evaluate the evidence for linkage with Gleason score using a new, more powerful, version of the Haseman-Elston (HE) statistical method (Elston et al., 1999; *Genetic Epi*: in press). The modified HE method is essentially linear regression where the dependent variable, the mean-corrected cross product, is regressed on the estimated proportion of marker alleles shared among brothers identical-by-descent.

Our multipoint analyses indicated that there was suggestive evidence for linkage ($p < 0.01$) in five genomic regions on chromosomes: 5, 7 (two regions), 10, and 19. There is broad support for linkage in many of these regions. We are presently adding more closely spaced markers around the peaks in each region to try and determine whether the peaks remain, and if so, to further refine the regions. We will present results from this analysis, and discuss our next step in attempting to localize gene(s) for prostate cancer aggressiveness.

23

Linkage disequilibrium and allele frequency distributions in 114 SNPs across the genome in five populations

KAB Goddard[1], JM Hall[2], JS Witte[1]
[1]Case Western Reserve Univ., Cleveland, OH, USA,
[2]PPGx, Inc, La Jolla, CA, USA

Single nucleotide polymorphisms (SNPs) are commonly observed in the genome, and may be less expensive to genotype compared to conventional markers. Thus, some researchers suggest using SNPs for both linkage and association mapping studies of complex traits. However, little is known about the distribution of SNP characteristics that are important in mapping studies—such as allele frequencies and inter-marker linkage disequilibrium (LD)—particularly in the context of population differences. Therefore, we evaluated 114 SNPs distributed in 33 genes on 16 chromosomes. Genes were selected based on having potential pharmacogenomic effects. The sample included 727 individuals from African American, Caucasian, Chinese, Hispanic, and Japanese populations. We used the EM algorithm to estimate haplotype frequencies for SNPs within a gene, and we considered several measures of LD.

Both alleles were observed in 105 SNPs for the African Americans compared to only 68 and 70 SNPs for the Japanese and Chinese respectively. Comparing allele frequencies among populations, the Japanese and Chinese had the highest correlation (R=.99), the Hispanics had a high correlation (R>.87) with all others, and the remaining pairs of populations had R<.83. The correlation in LD among populations appeared to distinguish two groups (Japanese and Chinese vs. the others), where R>.87 for populations within groups, and R<.65 for populations between groups. In most cases, pairwise LD was detected (p-value≤05) if the common allele frequency was <.95 in both SNPs. Exceptions included 22/74 and 15/78 pairs of SNPs for the Chinese and Japanese, respectively. These results suggest that one must carefully select SNPs for use in a genome screen using LD or linkage mapping methods. Also, the inter-marker LD we observed may illustrate the distribution of marker-trait LD for complex diseases.

24

Detecting allelic association in nuclear families.

K.D. Siegmund, W.J. Gauderman, D.C. Thomas.
University of Southern California, Los Angeles, USA
.

We present methods for testing allelic association in the presence of linkage from sibships with multiple cases and/or controls. We describe how recent 'allele-counting' methods can be applied using a likelihood that conditions on the phase and identity-by-descent (ibd) structure of the sibling genotypes, an extension of the conditional likelihood approach proposed by Self et al. (1991). Under an additional assumption of equal male and female recombination fractions, the conditioning statement can be relaxed to condition only on the number of alleles shared identical by descent between sib pairs. In a simulation study, we compare score tests from likelihoods that condition on (1) parental mating type only, (2) parental mating type and ibd structure, and (3) parental mating type, ibd structure, and phase. We consider a single diallelic susceptibility gene with a disease gene freqeuncy of 5% and population attributable risk of 20%, and a linked marker with four alleles; the population frequency of disease is 10%. Under tight linkage ($\theta = 0$), the test size from a likelihood conditioning only on parental mating type is high. For affected sib triplets, the false-positive rate under the multiplicative gene model is twice the nominal value at the 5% significance level and 4-times the rate at the 0.1% level. Score tests from likelihoods that condition on (2) or (3) achieve the nominal level, however tests from the latter appear conservative. We estimate power when the disease allele is in positive disequilibrium with one marker allele (δ_1). At 50% of $\max(\delta_1)$, the power for a sample of 100 affected sib triplets is 81% when conditioning on ibd compared to 74% when conditioning on ibd and phase ($\alpha = 0.05$). In summary, this likelihood approach provides a flexible alternative to the 'allele-counting' methods proposed in the literature; it will be useful for the fine-mapping of disease genes in sibships of arbitrary size, when linkage has already been established.

25

The use of conditional logistic regression and family-based controls to investigate association of a candidate autoimmune gene with rheumatoid arthritis

J H Barrett, S John, A Myerscough, J Worthington
ARC Epidemiology Unit, University of Manchester, Manchester, UK

Rheumatoid arthritis (RA) is an autoimmune disease. Apart from an association with HLA, little is

known of the genetic basis of the disease. We investigated association of D6S446, a microsatellite marker previously shown to be linked to other autoimmune diseases, with RA, using family-based controls from different types of family unit.

A conditional logistic regression framework was used to combine data on RA patients and either their parents or unaffected and affected siblings (Schaid and Rowland[1]). Once evidence for association with a particular allele was established, possible differences in association according to HLA genotype, age at disease onset and sex of the case were investigated by including interaction terms in the model. After excluding uninformative families, data consisted of 80 parent-affected-child trios, 32 parent-child duos and 32 sibships, consisting of between 3 and 8 siblings (with at least one affected and one unaffected).

There was overall evidence of association between D6S446 and RA ($p = 0.002$). The strongest evidence was for preferential transmission of allele 4, with an estimated odds ratio for RA of 2.7 (95% confidence interval 1.4, 5.3) for subjects with the allele. There was no evidence of interaction of this allele with HLA, sex or age at disease onset.

The results provide evidence of association between D6S446, a candidate autoimmune locus, and RA in a family-based design, ruling out a result due to population stratification. The conditional logistic regression framework provides a flexible method for combining data from different types of families and incorporating information on other factors of interest.

[1]Schaid DJ and Rowland C. Am. J. Hum. Genet. 63:1492–1506, 1998

26

Random effect models for combination of association and sib-pair analysis applied to osteoarthritis (OA)

J.J. Houwing-Duistermatt[1], C. Bijkerk[1,2], I. Meulenbelt[2], Th. Stijnen[1], P.E. Slagboom[2], C.M. van Duijn[1]
[1]Dept. of Epidemiology & Biostatistics, Erasmus University Rotterdam; [2]TNO, Institute of Prevention and Health, Gaubius Laboratory, Leiden, Netherlands.

Aim: Investigating the effect of Col2A1 on various OA outcomes in a population based cohort (749) with additional information on siblings (230) in a subset of this cohort (120). The observed outcomes are knee, hip, spine OA (binary) and generalized and hand OA (quantitative). Evidence exists that several genes play a role in OA [1]. In the population cohort the effect of Col2A1 on OA for women was significant. However, a sib-pair analysis did not confirm this result, necessitating an analysis that combines both types of information.

We propose random effect models with logit link for the binary outcomes and with identity link for the quantitative outcomes. Col2A1 is incorporated as a fixed effect. Two causes of correlation within sibships are modelled namely due to 1) sharing unlinked genes and 2) sharing genes located near Col2A1. Testing the null hypothesis of no correlation versus a certain type of correlation is straightforward [2]. For the logit link and correlation 2, modelling, assuming normal distributions for the random effects is complex [2].

Our hypothesis is that if Col2A1 has a causal effect the model with correlation 1 will give the best fit, while if Col2A1 is in linkage disequilibrium with a causal gene model 2 will give the best fit. At the meeting the results will be presented.

[1] Bijkerk C et al. To appear in Arthritis Rheum
[2] Houwing-Duistermaat JJ et al. (1998) Stat. in Med. 2939–2954

27

HLA and mate selection in North Carolina populations

L.B. Pollak Jr.[1], E.G. Reisner[2], D.D. Kostyu[2], D.V. Dawson[1]
[1]Case Western Reserve Univ., Cleveland, OH; [2]Duke Univ., Durham NC

HLA-A and HLA-B phenotypes collected from paternity litigation cases (1980–1997) were analyzed to see if the degree of antigen sharing between mates differed from the degree of sharing expected in the absence of selective factors. Those pairs with putative fathers whose HLA phenotype or blood type was incompatible with paternity were not included in the analysis. The remaining data consist of 287 African-American pairs and 326 Caucasian pairs. Race was self-reported. Data for Caucasians and African-Americans were permuted within race so that all possible pairs of mates were created, resulting in 326 samples of 326 pairs each and 287 samples of 287 pairs each, respectively. The average number of shared HLA antigens was calculated for each permuted sample. These values were used to construct empiric distributions of antigen sharing corresponding to the hypothesis of no mate selection, to which the observed data were compared. Of 287 generated samples, only 8 (2.78%) had equal or greater sharing at the A locus than did the observed data for African-Americans. Only 7 of the permuted samples (2.44%) had equal or greater sharing at the B locus. When the numbers of matches at the two loci were combined, no other sample had equal or greater HLA sharing than the observed data. In the Caucasian data, only 7 of 326 generated samples (2.15%) had equal or decreased sharing compared to the observed data for the A locus. Twenty-five of the samples (7.67%) had equal or greater sharing than that

seen in the sample for the B locus. Findings for the two groups were thus quite disparate. While findings for HLA-A in Caucasians parallel previous studies which suggested that mate choice might lead to decreased HLA sharing, the reverse was true for African-Americans. This finding, combined with other negative reports, suggests that negative assortative mating with respect to HLA is not universal.

28

Ramifications of HLA class I polymorphism and population genetics for vaccine development.

D.V. Dawson[1], M. Ozgur[1], M. Ghanayem[2], K. Sari[1], D.D. Kostyu[2].
[1]Case Western Reserve Univ., Cleveland, OH;
[2]Duke Univ., Durham, NC; USA.

The remarkable level of polymorphism associated with the class I loci of the HLA (human major histocompatibility) complex is well known, with the less than complete molecular typing to date yielding over 300 HLA-A, -B and -C alleles. Further, these HLA alleles are not randomly distributed across populations, but show considerable ethnic differences. These factors may pose a serious concern for the design of peptide vaccines due to HLA restriction of antigen recognition: It is quite possible that a vaccine effective in one population may be ineffective in another, due to differing distributions of HLA class I alleles. The types of vaccines potentially affected include those for prevention of infectious diseases and for immunotherapy of tumors. In order to begin to assess the potential problems in vaccine design due to HLA class I polymorphism, we have applied an algorithm to predict levels of favorable response in 15 populations embracing a spectrum of ethnic and racial groups, based upon population genetic models and known allele and haplotype frequencies for class I HLA loci. Approaches developed provide for the identification of alleles that should be targeted when designing vaccines, and permit prediction of responder status in any given population for vaccines with known HLA restricted epitope(s). They may also facilitate analysis of a protein sequence in order to identify potential epitopes for use in immunotherapeutic vaccines. Further, results from peptide binding studies suggest that, in some cases, peptide binding differs among DNA-defined subtypes of serologically defined HLA antigens; this is shown to have potentially significant impact on the proportion of persons theoretically capable of responding to certain vaccines. These findings therefore may have implications for the design of both vaccines and vaccine trials.

29

The relationship between the sibling recurrence risk-ratio and genotype relative risk

B. Rybicki[1] and R.C. Elston[2]
[1]Henry Ford Health System, Detroit, MI;
[2]Case Western Reserve University, Cleveland, OH.

The recurrence risk-ratio of disease in sibs, λ_S, has become a standard parameter used to estimate the statistical power of affected-sib-pair linkage studies. However, λ_S serves only as a surrogate for the genotype relative risk, γ, the parameter that directly determines the statistical power to detect a specific disease loci. We demonstrate that based on the definition of the two parameters, λ_S varies significantly more with respect to γ and the disease allele frequency for two-locus multiplicative interaction models compared with other two-locus and single gene models. Furthermore, for any given disease allele frequency and inheritance model, λ_S has an upper limit as γ increases. Under most inheritance models we tested, for a disease allele with frequency of 20% or greater, λ_S cannot exceed 10. Under dominant inheritance, if one or more susceptibility genotypes are known, restricting sibships to those ascertained by a proband with the high risk genotype will inflate the estimate of λ_S and can serve as an indirect way of testing if the putative susceptibility genotype increases disease risk. While disease genes with values of γ less than 4 cannot be mapped under most genetic models, λ_S in the range of 2 to 3 portend good success in gene mapping under a wide variety of genetic models. Investigators need to be cognizant of the basic differences between λ_S and γ in the planning of affected-sib-pair linkage studies.

30

The genetics of regional fat distribution phenotypes in the population-based National Heart, Lung, and Blood Institute Family Heart Study (NHLBI-FHS).

M. Feitosa[1], I. Borecki[1], S. Hunt[2], D. Arnett[3], P. Schreiner[3], D.C. Rao[1], M. Province[1].
[1]Washington Univ., St. Louis, MO, [2]Univ. Utah, Salt Lake City, UT; [3]Univ. Minnesota, Minneapolis, MN – USA.

Intra-abdominal fat is associated with increased risk of cardiovascular disease (CVD). Waist to hip ratio (WHR) and subscapular to triceps skinfold thickness ratio (STR) are measures of relative central fat distribution; they have been shown to be predictors of CVD morbidity and mortality, but the genetics are less well known. The inheritance patterns of WHR and STR, both before and after adjustment for body mass index (BMI), were investigated in the random sample of the

NHLBI-FHS, a multicenter population-based family study (N=2,716 subjects). Segregation analysis results for age adjusted WHR suggested an additive major gene that accounts for 35% of the phenotypic variance, with approximately 30% of the sample homozygous for the "high" genotype. The results for age and BMI adjusted WHR were also compatible with a major gene, however the multifactorial model provided the most parsimonious fit to the data. It is unclear if WHR was picking up a BMI gene. On the other hand, there is no evidence that the familial resemblance for STR phenotype is influenced by a Mendelian gene, whether controlling for BMI or not. Therefore, while both of these obesity measurements are potent predictors of chronic diseases, WHR also may be informative for studying the genetics of obesity, especially as a cardiovascular risk factor.

31

On a Randomization Procedure in Linkage Analysis

H. Zhao[1,2], K.R. Merikangas[1], K.K. Kidd[2]
[1]Department of Epidemiology and Public Health, and [2]Department of Genetics, Yale University School of Medicine, New Haven, CT 06520

Although much theoretical work has been undertaken to derive thresholds for statistical significance in genetic linkage studies, real data are often complicated by many factors, such as missing individuals or uninformative markers, which make the validity of these theoretical results questionable. In a typical genome-wide scan, many markers on the same chromosome are typed. Because of the dependence among these markers and the uncertainties in inferring the allele sharing status at a given position along the genome, the determination of statistical significance levels based on asymptotic results for these statistical procedures is also problematic. Many simulation-based methods have been proposed in the literature to determine empirically the statistical significance of the observed test statistics. However, these methods are either not generally applicable to complex pedigree structures, or too time consuming. In this presentation, we propose a general and computationally efficient randomization procedure that is applicable to arbitrary pedigree structures. This randomization procedure can be combined with any statistical test to assess the statistical significance for genetic linkage between a locus and a qualitative or quantitative trait. Furthermore, the genome-wide significance level can be appropriately controlled when many linked markers are studied in a genome-wide scan. Simulated data and a diabetes data set are analyzed to demonstrate the usefulness of this novel simulation method.

32

Association tests with marker haplotypes.

D. Clayton[1] & H. B. Jones[1,2]
[1]Cambridge Univ., Cambridge, UK; [2]Stanford Univ., Stanford, CA, USA.

With the increasing availability of single nucleotide polymorphism's (SNPs), methods for determining association between haplotypes composed of several adjacent markers are needed. The transmission/disequilibrium test (TDT) can be extended to multiple markers by considering the transmission of all distinct haplotypes. This has the disadvantage of ignoring any "similarity" between the haplotypes that are derived from a common ancestral chromosome. Furthermore, there may be a large number of haplotypes leading to relatively low statistical power. A method that takes account of the similarity between susceptibility haplotypes is described. This is based on techniques developed in spatial statistics and tests for a greater similarity between haplotypes that are transmitted than between those that are not transmitted to the affected offspring. Similarity is measured as the size of the contiguous section of the chromosome that is identical measured from a focal point. Such similarity will arise when mutation and recombination occur on the ancestral predisposing haplotype leading to several different but related haplotypes, each conferring susceptibility to the disease. Application of this method to data from auto-immune diseases show that it can detect associations that are missed by single marker methods.

33

Selecting contingency table in population-based association study: Allele frequency or positivity?

Jun Ohashi and Katsushi Tokunaga
Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Hongo, Tokyo 113, Japan.

In population-based association studies, the significance of the association between a candidate gene and a disease is usually examined by a simple chi-square test. Such studies require a 2 by 2 contingency table made up of either allele frequencies or positivities in affected and control groups. In order to investigate the influence of each 2 by 2 table on the power of the chi-square test, P values were calculated for two penetrance models (multiplicative and additive). When the value of penetrance was small and not markedly different among genotypes, a large difference in the power of the chi-square test was observed between the two tables. In a multiplicative model, the allele frequency table was superior to the positivity table for detecting significance. In contrast, in an additive model, the positivity table was most suitable. Selecting a

contingency table was especially important for detecting true association, when the required significance level was corrected to avoid a problem of multiple hypothesis testing. The chi-square test, therefore, should be performed for both tables to identify a susceptible gene with a low penetrance, even when no significant difference is observed in one table.

34

Homozygosity mapping and heterogeneity: a challenge.

Emmanuelle Génin and Françoise Clerget-Darpoux
INSERM U155, Genetic Epidemiology, France

To find genes involved in recessive diseases, homozygosity mapping has proven its power. The method focuses on inbred affected individuals and looks for regions of the genome where they are homozygous with the idea that the gene involved will be found in one of these regions.

For many recessive diseases however, the mutated gene may differ from one family to another. This heterogeneity considerably reduces the power of the method and in most situations, it becomes impossible to detect any of the genes involved using samples of unrelated inbred affected individuals. For instance, with a sample of 20 first cousin progeny, the probability to detect a gene responsible for half the disease cases is only 25% and falls below 6% for a gene involved in 40% of cases.

To reduce the level of heterogeneity, one strategy consists in using large inbred families. We have investigated this possibility and determined the information provided by related inbred affected individuals and their unaffected sibling in different situations. Based on these computations, guidelines regarding which individuals to type and what markers to use for the genome screen are proposed.

35

Setting the prevalence constraint in linkage analysis of multifactorial diseases.

P. Margaritte-Jeannin[1], M.C. Babron[1], H. Quesneville[2], F. Clerget-Darpoux[1]
[1]INSERM U155; [2]Institut Jacques Monod; Paris, France

For multifactorial diseases, “model free” linkage analyses are frequently applied to detect linkage with a set of markers. It is often argued that it may be more powerful to employ the classical lod score analysis (Morton, 1955) while maximizing on the underlying unknown genetic model. Because of the large number of parameters to be varied, the disease prevalence constraint is often set to limit the space of models considered. However, this constraint is not valid if the familial correlation for the disease is not explained by a single susceptibility locus, which is likely to be the case for multifactorial diseases.

The question addressed in this study is the impact of setting this constraint on linkage conclusion in different situations. Family data is simulated under different disease models and marker maps. Among the possible models, we consider a) two susceptibility loci with and without interaction; and b) a possible familial environment interacting with a disease susceptibility locus. Lastly, the impact of the prevalence constraint will be evaluated on the GAW11 simulated datasets (Greenberg, 1999), in which several familial correlation factors (genes and environment) interplay.

36

Reproductive factors and genetic predisposition for breast cancer: main effects and gene-envionment interactions. Results from a case-control study.

Jenny Chang-Claude[1], Silke Schmidt[1], Heiko Becher[2].
[1]German Cancer Research Center, Heidelberg,
[2]University of Heidelberg

A case-control study to quantify reproductive factors (number of full-term pregnancies, induced abortions, breast-feeding) and genetic predisposition on breast cancer risk including gene-environment interactions was carried out in two geographical areas in Germany with 706 incident, histologically verified breast cancer cases diagnosed between 1992 and 1995, 1391 population controls and 252 sister controls. Data was collected with a self-administered questionnaire. Main effects were estimated with multiple logistic regression. For estimation of gene-environment interaction we used different approaches: (1) inclusion of a variable for familial predisposition into the logistic model (number of first degree relatives with breast/ovarian cancer, or family risk score, or a modified version of the gene carrier probability estimated from a genetic model using the complete pedigree information), (2) analysis based on the case-only-design and (3) mixture-likelihood-model, in which the estimated gene carrier probabilities were used.

In the multiple logistic model comparing cases with population controls we found a highly significant protective effect for duration of breastfeeding ($p < 0.01$) and number of births ($p < 0.05$). Increased risk was found for abortions ($p < 0.05$) and genetic predisposition ($p < 0.001$). The comparison with the sister control group yielded similar results. Some indication for positive interaction between predisposition and abortions was found. A detailed comparison of the results from the different methods of analysis will be presented.

For interpretation it must be taken into account that assessment of genetic predispositon is based on surrogate variables rather than direct assessment of mutations in the susceptibility genes such as BRCA1/BRCA2 which limits the power of the analysis.

37

Family ascertainment using a population-based registry of myocardial infarction

S. Engel[1], H. Bickeböller[1], C. Hengstenberg[2], S. Holmer[2], H. Löwel[1]
[1]GSF-Institute of Epidemiology, Neuherberg, Bavaria, Germany [2]Univ. of Regensburg, Regensburg, Bavaria, Germany.

Myocardial infarction (MI) is a late onset disease with a high case fatality. This study reports experiences with the ascertainment of MI affected families when starting from a population-based registry which is an optimal setting, monitoring 310,000 subjects aged 25 to 59 years for 11 years (1985–95). **Methods**: 1381 MI survivors under age 60 were asked for participation in a family study and for information on vital status and history of MI of their first degree relatives. Index probands (IPs) and living siblings were invited for a standardized interview and offered an extensive cardiovascular examination. **Results**: In 1996, 371 IPs were lost to follow-up, mainly due to death (266 IPs). Of the remaining 1007 IPs, 504 (50%) had to be excluded for unavailable family members. Out of the 503 eligible IPs, 293 (21% of the MI survivors) were examined with at least 1 sibling (536 sibs). 31 affected sib-pairs were ascertained. Over 90% of the parents were deceased. Using baseline characteristics of the registry, loss to follow-up was associated with recurrent MI, diabetes, cigarette smoking, application of thrombolysis at the acute event and type of discharge medication. Exclusion from the study was associated with cigarette smoking and negative parental history of MI. **Conclusion**: The sample ascertained from the registry appears to be suited for discordant sib pair analyses. It is selected by survival, but apparently also by positive parental history of MI. This latter overrepresentation counteracts the suggested elevated hereditary proportion in early fatal events.

38

Incorporating Genetic Marker Information into the Analysis of Twin Survival Data: A Simulation Study

I.A. Iachine[1], K. Christensen[2], A.I. Yashin[3]
[1]Department of Statistics and Demography, Odense University, Odense, Denmark; [2]Institute of Public Health, Odense University, Odense, Denmark; [3]Max Planck Institute for Demographic Research, Rostock, Germany and Duke University, Durham, NC, USA.

Traditional methods of linkage analysis of sib-pair data for quantitative traits are not well suited for the analysis of survival data. In particular, analysis of censored and truncated survival times requires a completely parametric model. To address the censoring and truncation problem, semiparametric frailty models have been developed and successfully applied in the genetic analysis of survival data on Danish Twins without information on genetic markers. In this paper we suggest a method for analysis of sib-pair survival data with genetic marker information that combines both approaches. We develop a bivariate frailty model where individual frailty is treated as a quantitative trait with an underlying genetic locus that may be in genetic linkage with one or more genetic markers, with the location of the frailty gene to be estimated. The model includes additional frailty components associated with effects of polygenes, shared and non-shared environment. We suggest a semiparametric maximum likelihood estimation procedure and study the properties of the new model using simulated survival data on MZ and DZ twins. In particular, we investigate the sample size required for detecting linkage under different censoring mechanisms and consider the situation where a genetic marker is a mortality risk factor in addition to being linked to the frailty gene.

39

Comparison of Type I error rates of Haseman-Elston sib-pair linkage method in various quantitative trait models.

D. M. Mandal[1], A. F. Wilson[2], J. E. Bailey-Wilson[2].
[1]Louisiana State University Medical Center, New Orleans, USA; [2]National Human Genome Research Institute, Baltimore, USA.

It is well known that Haseman-Elston (H-E) sib-pair linkage method does not assume that the genetic model underlying the trait phenotype is known without error although this assumption is made for marker loci. However, misspecification of allele frequencies at the marker locus decreases power when some or all parental genotypes are unknown. In this study, the Type I error rates of the H-E sib-pair methods were compared for different types of traits when some or all parental data were missing and allele frequencies at the marker loci were misspecified.

Data were generated for a quantitative trait and marker loci in nuclear families using G.A.S.P. (V3.3). Three types of traits were simulated: two due to an unlinked additive major locus with two equifrequent alleles with a random environmental effect (50%, 90%) and one that was completely environmentally determined. The simulated data were analyzed using (i) one of the parent's marker data, and (ii) no parental marker data,

with both correct and incorrect marker allele frequencies. When the trait was completely environmentally determined, the Type I error rate of the H-E sib-pair linkage analysis was robust to misspecification of marker allele frequencies even if the allele frequencies were severely misspecified (i.e., ±0.8 of true frequencies). Thus, this method appeared to be remarkably robust to even very large misspecifications of allele frequencies.

40

Testing for Allelic Associations in Structured Populations

N. Chapman, E. Thompson
University of Washington, Seattle, WA, USA.

Understanding disequilibria within and between loci is important for disease mapping from population data. In genetic isolates, the population structure has a major impact on allelic associations both within and between loci. We study disequilibria using the North American Hutterites as an example isolate, focusing first on a single locus.

Disequilibrium at a single locus is described by the departure of observed heterozygote proportions from those expected under Hardy Weinberg Equilibrium (HWE). The observed deficiency or excess of heterozygotes depends on the extent to which mates are more or less closely related than a randomly chosen pair. Excess homozygosity is of particular interest to epidemiologists, since it can result in higher prevalence of rare recessive disorders. Testing for HWE is done using a $\chi 2$ statistic, to compare the observed genotype counts with HW expectations. The distribution of the statistic under the null hypothesis depends on having independently sampled genotypes, which is not the case if the sample consists of groups of siblings. In this situation, the type I error of the test can be grossly inflated over the nominal significance level.

The Hutterite population is divided into three separate sub-populations (which are called leut), and into colonies within leut. We test for departures from HWE within leut, where the genotype data consist of samples of groups of siblings, and the sibships are also more distantly related. We use simulation to find the distribution of the test statistic under the null hypothesis of random mating within leut.

41

Mapping disease genes using haplotype shared lengths

D. Qian, D.C. Thomas
Univ. of Southern California, Los Angeles, CA, USA.

We describe an approach to mapping disease genes based on the sharing of haplotypes between unrelated parents in a set of nuclear families. A disease score $d_i(x)$ at a candidate location x is assigned to each parental haplotype i based on the numbers of transmissions to affected and unaffected offspring. The lengths $L_{ij}(x)$ of the segments surrounding x that are shared between all possible pairs i and j of unrelated parental haplotypes are computed. The summary statistic $\Lambda(x) = \Sigma_{\{i \neq j\}}\ d_i(x)\ d_j(x)\ L_{ij}(x)$ is compared with a null distribution obtained by randomizing the disease scores against the lengths. In 4 out of 10 simulations of an isolated population with age of 100 generations on a chromosomal segment of length 100 cM with 99 diallelic markers, the strongest peak occurred flanking the true disease location, and in 4 others, the second highest peak occurred there. Strong signals at flanking disease locations are observed in populations with more than 80% disease haplotypes sharing only 1 or 2 same ancestral mutations. To estimate the location x0 of the disease gene, a generalized estimating equations approach is proposed to regress haplotype-pair scores $D_{ij} = (d_i - \mu)(d_j - \mu)$ on a mean function of the form $\Sigma x\ L_{ij}(x)\ \exp[-\beta(x-x_0)^2]$. A weighted variance estimator using the shared length of the four members in each pair of paired haplotypes is constructed in a similar way to that proposed by Lumley and Heagerty (1999). The performance of these estimating equations is under investigation.

42

A two-locus model for nonsyndromic congenital dysplasia of the hip

C. Scapoli[1], G. Bertolani[2], G. Atti[3], V. Sollazzo[2]
[1]Biology Dep., [3]Paediatric Div., [2]Biomed. Science and advanced therapies - Ferrara University, Italy

Congenital dysplasia of the hip (CDH) is one of the most common skeletal congenital anomalies. For Caucasian populations an incidence of CDH of 1 per 1,000 can be assumed. In Italy, the incidence is about 10 per 1,000 and in Ferrara is still higher, equal to 18.4 per 1,000 live births. Several examples of familial transmission of isolated CDH are well known; however, the cause of CDH is not yet completely clear, even if the importance of genetic and environmental factors is evident.

Among all patients treated at the Ferrara's Center for the study of CDH in the period 1991–96, a sample of 171 patients with positive familial recurrence of nonsyndromic CDH was collected. The 171 available pedigrees were partitioned into 507 nuclear families and complex segregation analysis was performed by applying the mixed model of inheritance, expanded as the unified model, implemented in the computer programs POINTER and COMDS.

Using POINTER, the hypotheses of sporadic, multifactorial and polygenic transmission of nonsyndromic CDH were rejected. Among the models postulating a major locus only the recessive was accepted.

Since the evaluation of CDH severity was made according to Graf's echographic classification, assuming 4 pathological levels with increasing gravity, we tested the two-locus model including this information in COMDS analysis. Under this model, a significant improvement of the general likelihood was observed ($\chi^2_{[4]}$=26.2, P<0.001), supporting the presence of at least a second locus. Amongst the various two-locus models, the one assuming recessive transmission for the major gene (frequency of the deleterious allele = 0.2) showed the best likelihood. For the modifier locus, all mendelian genetic hypotheses were accepted on LRT basis; however, since the recessive-recessive two-locus model was the most parsimonious, we considered it the most appropriate for our data.

43

Significance and power estimates of genome scans in relation to follow up strategies.

S. Loesgen, H. Bickeböller
Institute of Epidemiology, GSF-National Research Center for Environment and Health, Neuherberg, Germany

Simulation studies are widely used to evaluate the performance of statistical methods for identifying susceptibility genes, including whole genome simulations for multipoint methods. Here, the classification into 'true' and 'false' regions is a special problem hardly discussed in the literature where the variety of used definitions is vast.

A fast and flexible genome scan simulation program GESIMS is presented. Results are given for 1000 sets of 2x200 affected sib pair families generated for a multiplicative disease model with a genotype relative risk of 4. Various classification schemes are applied to a 10cM screen on the first 200 families. Power estimates to a screening level of p=0.01 vary between 77% and 90% if only the marker next to the disease locus or the whole chromosome is counted. The difference in power estimates which take subsequent finemapping into account is less than 2% if a 1cM mapping in an area of 20cM or 60cM around peaks is planned, while the mean number of typed markers increases from 42 to 78.5 and 2.2 to 5.2 on the linked respectively an unlinked chromosome (of length 150cM). With the second set of families the results of different fine mapping strategies are evaluated.

For realistic power estimates not only the nominal p-value but the width of the region one is prepared to follow up has to be considered. The differences in estimates obtained by the various definitions underline the need to set a standard, which must adapt to the change in cost of marker typing.

44

HLA associations with subtypes of juvenile idiopathic arthritis based on a latent class analysis of symptoms

Barrett JH[1], Thomas E[1], Thomson W[1], Donn R[1], Woo P[2], Southwood TR[3]
[1]ARC Epidemiology Unit, University of Manchester, [2]Department of Molecular Rheumatology, University College London, [3]Department of Rheumatology, University of Birmingham

Juvenile idiopathic arthritis (JIA) is a childhood rheumatic condition with known HLA associations. It encompasses a spectrum of diseases that vary widely in onset characteristics, clinical course and associated manifestations. Existing widely used systems for classifying JIA into subtypes are based on clinicians' perceptions of disease patterns, and may not define biologically homogeneous subgroups. We have used latent class analysis (LCA) to identify underlying classes which best explain the relationships between clinical variables and investigated HLA associations with these subgroups. Clinical data were obtained from 572 subjects in the British Paediatric Rheumatology Group National Repository for JIA. The analysis was based on the age at onset of arthritis (< 5 years, 5–9 years, 10–16 years), and 9 other key clinical variables, each of which was recorded as present or absent. The most appropriate solution statistically was based on 7 latent classes, 4 with approximately 20% of the patients each and 3 smaller classes. Of the JIA patients 461 (81%) could be assigned to a particular class with probability greater than 0.7. There was some correspondence between the latent classes and currently identified subgroups, but these were by no means coincident. Marked differences in HLA phenotype frequencies were observed in the different classes. For example DRB1*08, which shows an overall association with JIA, was found to be associated with four of the JIA subgroups, but not the other three. This analysis provides an alternative way of identifying homogeneous groups of patients. Further work examining other genetic associations and prognosis in the latent class subgroups

45

Design and analysis options for the estimation of gene-environment interaction from epidemiological case-control studies

Silke Schmidt[1], Heiko Becher[2], Jenny Chang-Claude[3]
[1]Duke University Medical Center, Durham, NC;

[2]University of Heidelberg, Heidelberg, Germany; [3]German Cancer Research Center, Heidelberg, Germany.

The assessment of gene-environment interaction with data from large epidemiologic studies continues to offer methodological challenges, especially when rare and highly penetrant susceptibility genes are concerned. When measured genotypes of study probands are available, it is of interest to compare the following design options with simulated case-control data: a pair-matched case-control study, either using a pair-matched or an age-stratified analysis; an unmatched case-control study, either including age in the model or using a post-hoc age stratification; a pair-matched case-control study using sibling rather than population controls; and a case-only analysis discarding the controls, allowing for better precision of the interaction estimate, but assuming a different scale of measurement than a case-control analysis.

In large case-control studies, it is often cost-prohibitive to obtain measured genotypes on all study probands. Some approaches for assessing gene-environment interaction with surrogate genetic measures based on family history information from cases and controls may then be compared on simulated data. These methods are also illustrated on a German case-control study of premenopausal breast cancer.

46

Extending family based tests of association: using unaffected sibs, covariates, and testing for gene-gene interaction.

KL Lunetta[1], S Faraone[2], J Biederman[2], D Rabinowitz[3], NM Laird[1]
[1]Harvard School of Public Health and [2]Medical School, Boston, MA; [2]Columbia Univ., New York, NY. USA

We extend the methodology of Self (1991) and Schaid (1996) to incorporate arbitrary phenotypes and covariates into general score tests of association. We first use standard models of association for a phenotype Y and any number of predictors, which may include indicators for gene alleles and environmental factors. We then construct the score by assuming likelihoods for the distribution of Y, given genotype. The distribution of the score is computed as a function of offspring genotypes conditional on parental genotypes and trait values for offspring and parents (Rabinowitz & Laird 1999). Our approach provides a natural extension of the TDT (Spielman et al. 1993) to any phenotype and multiple genes or environmental factors, and allows the examination of gene-gene and gene-environment interaction.

When the trait Y varies among subjects, or when there are covariates in the model, the score statistic depends on one or more nuisance parameters. Often these parameters do not depend on genotype, and population estimates may be available. Sometimes the parameters may be estimated from the data. We suggest an alternative approach in which we optimize the χ^2 statistic over a nuisance parameter, and then apply a correction to the p-value suggested by Davies (1977).

We illustrate our methods with a sample of 43 nuclear families having ≥1 member with Attention Deficit Hyperactivity Disorder (ADHD), with 44 affected and 34 unaffected children. We examine associations and interactions between ADHD or a related measured phenotype and the dopamine D4 and dopamine transporter genes, using sex and parental affection status as covariates. For tests without covariates, we compare results using population and sample estimates of one nuisance parameter and the optimized χ^2 statistic with Davies (1977) correction.

47

Testing Gene-environment Interaction Using Affected Sibpairs

FZ Sun[1], WD Flanders[1], QH Yang[2], HY Zhao[3], MJ Khoury[2]
[1]Emory Univ., Atlanta, GA, USA; [2]CDC, Atlanta, GA, USA; [3]Yale Univ., New Haven, CT, USA

Testing for gene-environment interactions is an important problem in genetic epidemiology. Several designs are available to test for gene-environment interactions such as case-control studies, case-only studies, and family based case-control studies. Families with affected sibpairs are routinely collected to find the genetic bases of complex diseases. Here we compare several methods for detecting gene-environment interactions using affected sibpairs. We considered the following tests: (1) comparing the fraction of affected sibpairs sharing 0 alleles identity by decent (IBD) when the environment is present with the fraction of sibpairs sharing 0 alleles IBD when the environment is absent; (2) comparing the average allele sharing IBD when the environment is present with the average allele sharing IBD when the environment is absent; (3) regress the average allele sharing probability with respect to the number of sibs having the environment. We compare the tests under two models. In the first model, the environment is independently distributed among individuals. We show that generally the third test is the most powerful. In the second model, the environment is the same within a family. We show that the third and the second tests have essentially the same power and are generally more powerful than the first test. Even for the most powerful test, the

sample size needed to have a reasonable power to detect the gene-environment interaction is usually too large to be realistic. Alternative methods should be used to detect gene-environment interactions.

48

A Note on the Relative Powers of TDT, S-TDT and 1-TDT

Dai Wang and Fengzhu Sun
Emory University, Atlanta, GA, USA

The transmission/disequilibrium test (TDT) is a widely used method to detect the linkage disequilibrium between a candidate locus (a marker) and a disease locus. TDT is a family-based design and has the advantage that it is a valid test when population stratification exists. Standard TDT requires the marker genotypes of affected individuals and their parents. For diseases with late age of onset, it is difficult or impossible to obtain the marker genotypes of the parents. When both parents' marker genotypes are unavailable, Ewens and Spielman extended the standard TDT to S-TDT for use in sibships with at least one affected individual and one unaffected individual. When only one parent's genotype is available, Sun et al. proposed a test, the 1-TDT, for use with marker genotypes of affected individuals and only one available parent. Here we study the relative powers of TDT, S-TDT and 1-TDT. We show that the sample size needed for 1-TDT is roughly the same as the sample size needed for S-TDT with two sibs and is about twice the sample size needed for standard TDT. We are currently comparing the powers of S-TDT and 1-TDT with the corresponding statistics for DNA pooling data studied by Risch and Teng.

49

Tree-based Methods for Mapping Genes in Complex Diseases

Heping Zhang[1] and George Bonney[2]
[1]Department of Epidemiology and Public Health, Yale University School of Medicine, New Have, CT 06520, and [2]National Human Genome Center at Howard University, 2216 sixth Street, N.W., Washington, DC 20059

We propose the use of tree-based methods for mapping genetic markers for complex diseases. These methods are known for their capability to incorporate high dimensional data and have great potential in selecting candidate genes among many markers for complex diseases, particularly in the presence of epistasis. Our methods combine the tree-based model described in Zhang (1998) with the regressive models of Bonney (1986) by generalizing the goodness of node split and the tree cost-complexity measure. Both simulations and real applications will be used to test the usefulness of these methods.

50

Meta-analysis for interpreting results from multiple genome scans

J.A. Badner, E.S. Gershon
University of Chicago, Chicago, IL, USA

Lander and Kruglyak gave guidelines for interpreting linkage results based on estimating how often a particular threshold for significance would be exceeded by chance in a single genome scan. Criteria for significant linkage was recommended to be a threshold that would be exceeded 0.05 times by chance in a single genome scan. These criteria do not always enable interpretation of the results of multiple genome scans of a genetic trait. In some cases, there may be several scans that show significant results within the same region but none of the individual results would meet criteria for significant linkage. In other cases, two or more studies exceed criteria for significant linkage but several other studies do not show nominally significant results with the same region. This may occur due to low power to detect genes of small effect using the criteria recommended by Lander and Kruglyak. One possibility is to combine the results of these studies. However, this is difficult to do in many instances. We propose a type of meta-analysis which involves combining p values across the studies. We make recommendations for criteria of linkage using this analysis. Comparisons of the power of this meta-analysis with Lander and Kruglyak criteria are presented. We apply this method of meta-analysis to the evidence for linkage of IDDM susceptibility loci.

51

Segregation analysis of cancer in families of glioma patients.

M de Andrade[1], J S Barnholtz[1,2], C I Amos[1], P Adatto[1], C Spencer[1], M Bondy[1]
[1]Department of Epidemiology, UT M D Anderson Cancer Center, Houston, Texas; [2]UT School of Public Health, Houston, Texas

The probands were patients diagnosed with glioma between June 1992 and June 1995 registered at the University of Texas M. D. Anderson Cancer Center. All

probands were younger than 65 years at diagnosis and resided in the United States or Canada. We used a sequential sampling scheme where second-degree relatives were included only if a first-degree relative had cancer. The study included 5088 relatives (3810 first-degree and 1278 second-degree) of 639 probands. We performed segregation analyses using PAP (Pedigree Analysis Package). Gliomas are a rare type of brain cancer (~1–2% of all cancers are gliomas), therefore analyses were divided in three categories: 1) using all 639 families; 2) using families with at least one first-degree relative with cancer; 3) using families whose probands have p53 information (brain cancer is part of the Li-Fraumeni syndrome). In all analyses, we demonstrated that a multifactorial model was favored and a model postulating a purely environmental cause of cancer was rejected.

52

Familial Aggregation and Patterns of inheritance of Resting Metabolic Rates in African-American Families.

Luke A*, Rotimi C, Guo X, Zhu X, Cooper R.
Loyola University Medical Center, Maywood, Illinois,USA

Epidemiologic studies have consistently reported higher rate of obesity among African Americans relative to their white counterparts. The reason for the increased rate is complex and poorly understood. Although lacking consistent data, several investigators have suggested that the variation observed in resting metabolic rate (RMR - which can represent up to 70% of total daily energy expenditure) may explain in part the increased rate of obesity observed in this ethnic group. This study was designed to address the issue of heritability of RMR in a cohort of African-American families enrolled and examined in the Chicago area. RMR was measured using Delta Trac II metabolic cart (SensorMedics, Anaheim, CA) on a population sample of 639 persons from159 families. Body composition was measured by bioelectric impedance analysis (BIA; RJL, Inc., Clinton Township, MI). Familial correlation was estimated for adjusted (age, sex and fat-free mass FFM, fat mass FM) and unadjusted RMR using the FCOR sub-routine available in S.A.G.E. Heritability (h^2) estimates for adjusted-RMR were 30% and 52% for unadjusted-RMR. FFM was the most important determinant of RMR accounting for about 63% of observed variance. The heritability estimate for FFM was 49%. The presence of a moderate spouse correlation (0.15 vs 0.11) respectively for RMR and FFM suggests the importance of shared environmental factors. In summary, the findings of this study suggest the combined influence of genetic and environmental factors in the determination of measured RMR phenotype.

53

Segregation analysis of late onset Alzheimer's disease: evidence for a major gene

E.Drigalenko[1,2], S.E. Poduslo[1], R.C. Elston[2]
[1]Texas Tech University Health Sciences Center, Lubbock, TX; [2]Case Western Reserve University, Cleveland, OH

Alzheimer's disease was considered as a dichotomous trait determined by a diallelic locus with (geno)type-, age-, and sex-dependent penetrances. The subjects comprise 171 Caucasian families, each ascertained via a patient in Texas. A total of 3615 family members, including 308 affected, having age of onset not less than 65 years, were analyzed. Age of onset was known for each patient, and age at examination was known for each unaffected relative.

Segregation analysis was performed under the logistic regressive model using the program REGTL (S.A.G.E. 1998). Model I (Elston et al 1978), which assumed type-dependent age of onset distribution, resulted in the penetrance of heterozygotes being smaller than those of both homozygotes in all samples. This model was rejected in favor of Model II, which allows for type- (and sex-) dependent susceptibilities. The hypothesis of "No major gene" was rejected in all cases. Hardy-Weinberg disequilibrium, familial effects, age transformation, and sex-dependence of age of onset parameters were tested and found to be non-significant.

Families in which the mean age of onset for patients was between 65 and 74 years (the last was the median of the family means) were found to have a Mendelian recessive inheritance with an allele frequency of 20%. Dominant inheritance was rejected with P=0.012. For families with the mean age of onset for patients over or equal to 75 years dominant inheritance was preferable; but recessive type of inheritance was not rejected. Hence, using age 75 as a dividing point between early and late onset yielded more genetically homogeneous samples.

Supported by funds from Texas and grants P41RR-03655, GM28356, and F05TW05285.

54

Least Squares Estimation of Variance Components for Genetic Linkage Analysis

Christopher I. Amos and Xiangjun Gu
UT M D Anderson Cancer Center, Houston, Texas, USA

Variance components methods provide an efficient method to detect genetic linkage for quantitative data. Available procedures for variance components estimation either require multivariate normal assumptions and use iterative maximum likelihood methods, or use

computationally intensive generalized estimation equations methods. Here, we develop a rapid and simple estimation procedure using a least squares approach. Using this approach, we provide closed-form noniterative expressions for the variance components estimates. Using simulated data, we have compared least squares and maximum likelihood estimation approaches for normal and skewed data. For normally distributed data, the least squares approach provided estimates of the variance components at least 80 times faster than the maximum likelihood procedure. For data generated with a chi-squared residual, the least squares procedure was between 100 and 900 times faster than the maximum likelihood procedure. The least squares methods yielded unbiased estimates of the variance components, but with the efficiency (relative to maximum likelihood) ranging from 40–112% with lower efficiencies for the normal data and higher efficiencies for the nonnormal data.

55

The Influence of Environment on Heritability of ACE and Angiotensinogen: A Comparison of US Blacks and Nigerians

X. Guo[1], R. Cooper[1], C. Rotimi[1], A. Luke[1], R. Ward[2], A. Adeyemo[3], S. Danilov[4]
[1]Loyola University Chicago, IL; [2]Oxford University, UK; [3]University of Ibadan, Nigeria; [4]University of Illinois, Chicago, IL.

Both angiotensinogen (AGT) and angiotensin-I converting enzyme (ACE) have been shown to be heritable traits and genetic markers have been identified that are associated with variation in their blood levels. The extent to which the environmental context determines heritability has not been examined, however. The African diaspora provides an opportunity to compare the expression of these traits in genetically related populations that live under widely contrasting lifestyles. As part of an on-going study on the genetics of hypertension we examined nuclear families that included 2,093 Nigerians and 1,993 African Americans. Body mass index (BMI; weight[kg]/height[m]^2) was 21 kg/m^2 in Nigeria kg/m^2 and 29 in the US. AGT was considerably higher among African Americans (1919 vs.1396, $p < .01$), while mean ACE was higher in Nigerians (630 vs. 517, $p < .01$). A substantial household effect was observed among the Nigerian families, with spousal correlations of 0.35 for AGT, 0.20 for ACE, and 0.20 for BMI, and correlations among first degree relatives were very substantial, viz, 0.5 for AGT and 0.4 for ACE. Among African Americans, in contrast, no clear patterns of familial aggregation were observed for AGT and ACE. The familial resemblance of BMI was similar among African Americans and Nigerians (r=0.2). Additional analyses eliminating persons treated for hypertension and adjusting traits by covariates showed that there is no spouse correlation for AGT and ACE in the US. In summary, a substantially lower degree of familial aggregation was observed for AGT and ACE among African Americans, most likely reflecting greater random environmental effect at the individual level.

56

Testing quantitative traits for association and linkage in the presence or absence of parental data

X. Zhu[1], R.C. Elston[2]
[1]Department of Preventive Medicine and Epidemiology, Loyola University Medical Center, Maywood, IL, USA; [2]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA.

Zhu and Elston developed a TDT method for quantitative traits by defining a linear transformation to condition out founder information. The method tests the joint null hypothesis of no linkage or association and can be applied to general pedigree structures. However, this method requires both parental genotype and phenotype information, which may sometimes be difficult to obtain. In this paper, we describe a method to overcome this problem, for the case where nuclear families are sampled, by regressing the sib's phenotype differences on their genotype difference. We show that the method is not affected by population stratification. The statistical power and validity of the test are investigated by simulation. In practice, the data may contain some families with parental phenotype and genotype information and some without such information. We show how all the data may be analyzed together in a single analysis.

57

Colorectal cancer with defective DNA mismatch repair is associated with alpha-1 antitrypsin deficiency heterozygosity and history of cigarette smoking

P. Yang, J.M. Cunningham, J. Katzmann, T. Lesnick, K. Halling, L. Burgart, E. Wiegert, A. Knutson, E. Christensen, R. Kyle, N. Lindor, S.N. Thibodeau
Mayo Clinic and Foundation, Rochester, MN, USA

Microsatellite instability (MSI) is a type of genomic alteration observed in approximately 30% of colorectal cancer (CRC). Three MSI phenotypes have been defined for CRC. The MSI-H is characterized by MSI at ≥30% of the loci, the MSI-L by MSI at 1–30% of the loci, and the MSS by an absence of MSI at any of the loci examined. The MSI-H is the result of defective

DNA mismatch repair (MMR). We report that MSI-H colorectal tumors are associated with alpha-1 antitrypsin deficiency carrier (α1AD-ht) status and cigarette smoking history of the patients. Among 55 CRC patients with MSI-H tumors, the α1AD-ht rate was 21.8% whereas among 120 patients with MSI-L/MSS tumors, the rate was 9.2% (p=0.02). A similar difference in the α1AD-ht rate was also observed between the 55 MSI-H CRC patients and 191 general population controls without CRC. The relative risk of having MSI-H CRC among α_1AD carriers was 3.0 (95% CI 1.1-6.7) compared to non-carriers, after adjusting for age, gender, and cigarette smoking history. Cigarette smoking, past and current, was associated with a 2- and 5-fold elevated risk respectively for MSI-H CRC, but not for MSI-L/MSS CRC. Our finding of an >3.0-fold increase in α_1AD-ht among MSI-H CRC patients over MSI-L/MSS CRC patients was consistent in proximal and distal colon and rectum. These data suggest a possible etiologic link between α_1AD-ht and the development of CRC with defective MMR.

58

Recent Genetic Findings in Bipolar Disorder

[1]JINurnberger, [2]JRDePaulo, [3]MCBlehar, [4]ES Gershon, [5]T Reich, [1]T Foroud, NIMH Genetics Initiative Bipolar Group
[1]Indiana University School of Medicine, [2]Johns Hopkins University, [3]National Institute of Mental Health, [4]University of Chicago, [5]Washington University of St. Louis

Bipolar Affective Disorder (BP) clearly aggregates within families. The genetics of BP is complex, however, and not consistent with single major locus inheritance; no specific genes have yet been located and confirmed. A collaborative study involving four sites was supported as part of the NIMH Genetics Initiative. A structured interview (DIGS) was developed to provide a comprehensive phenotypic assessment of patients and relatives. Families included were required to have at least two affected subjects with bipolar I (BPI) disorder or one with BPI and a second with schizoaffective disorder, bipolar type (SA/BP). Probands and relatives were interviewed and provided a blood sample for transformation and storage at a national data bank. We present results from 540 subjects selected from 97 families. This group included 282 affected sibling pairs, (BP & UP), as well as 412 affected relative pairs. A survey was completed with 319 markers. Analysis was carried out using SIBPAL and Genehunter Plus. A number of candidate areas are supported especially areas on chromosomes 6, 10 and 16. A summary of results of linkage analysis on individual families from this dataset will also be presented.

59

A low level of response to alcohol and its relation to alcoholism

L. Flury[1], M. Schuckit[2], H. Edenberg[1], Kalmijn[2], T. Reich[3], L. Bierut[3], A. Goate[3], T. Smith[2], T. Foroud[1].
[1]Dept of Medical & Molecular Genetics, Indiana Univ School of Medicine, Indianapolis, IN [2]Dept of Psychiatry, Veterans Affairs Medical Center, Univ of California, San Diego, San Diego, CA, [3]Dept of Psychiatry, Washington Univ School of Medicine, St. Louis, MO.

Alcoholism is a complex genetic disorder with an estimated heritability of 40–60%. Although attempts to identify genes contributing to the risk for alcoholism have yielded several regions with evidence of linkage, genes have not yet been identified. A more powerful method to detect susceptibility genes for alcoholism may be to use the biological endophenotype of a low level of response to alcohol (LR), which is a significant predictor of alcoholism. LR is exhibited as a need for higher doses of alcohol to produce effects of intoxication.

Data were generated from families collected as part of a Collaborative Study on the Genetics of Alcoholism. LR was estimated using the Self-Rating of the Effects of Alcohol (SRE), a questionnaire which obtains the number of drinks required to achieve various effects of intoxication during the first five drinking episodes (FIRST 5), the period of heaviest drinking, and the most recent three months of drinking. The average number of drinks necessary to achieve alcohol-related effects across all time periods was used as a composite variable (TOTAL). FIRST 5 and TOTAL were analyzed as both quantitative (400–700 pairs) and qualitative (70–100 pairs comprised of individuals in the lowest 1/3 of the SRE distribution) variables. Multipoint sibpair linkage analysis was performed. Evidence for genes affecting LR was found on two chromosomes that had previously shown evidence for genes affecting risk for alcoholism: chromosomes 1 and 7 (lod ≈1.5–2.0). In addition novel evidence of linkage to the LR endophenotype was identified on chromosomes 15 and 16 (lod≈2.5) as well as 21 (lod≈3.0).

60

Evidence for a shared genetic susceptibility to epilepsy and febrile convulsions: a familial aggregation analysis

C. Barker-Cummings, R. Ottman, Columbia Univ., New York, NY, USA.

Febrile convulsions and epilepsy are each known to be influenced by genetic factors, and children with febrile convulsions have increased risk for later development of epilepsy. Thus we tested the hypothesis of a

shared genetic susceptibility to these two disorders. The study population comprised 2404 full siblings of adult probands with either idiopathic/cryptogenic or postnatal symptomatic epilepsy whose mothers had been interviewed to obtain a history of febrile convulsions, and who had survived to age 5 or older. Epilepsy was defined as a lifetime history of ≥2 unprovoked seizures, and a febrile convulsion as a seizure occurring in the setting of fever at age 5 or younger, in the absence of a central nervous system infection or other acute cause. Siblings with idiopathic/cryptogenic epilepsy were more likely to have had a febrile convulsion than siblings without [RR=2.6 (95% CI 1.01, 6.66)]. In siblings without idiopathic/cryptogenic epilepsy, febrile convulsions were associated with the proband's history of a febrile convulsion [RR=2.2 (1.34, 3.72)], indicating familial aggregation of febrile convulsions. Febrile convulsions in siblings were not associated with the proband's etiology of epilepsy (idiopathic/cryptogenic vs. remote symptomatic) [RR=1.0 (0.55, 1.76)], or with a history of idiopathic/cryptogenic epilepsy in one first-degree relative other than the proband [RR=1.2 (0.52, 2.89)]. However, they were independently associated with a history of idiopathic/cryptogenic epilepsy in ≥2 first-degree relatives in addition to the proband [RR= 13.7 (5.40, 34.73)]. These findings support the hypothesis of a shared genetic susceptibility to febrile convulsions and epilepsy, and suggest that this shared susceptibility may be restricted to families containing many affected individuals.

61

Evidence that diastolic blood pressure links to two loci in Mexican Americans

LD Atwood[1], PB Samollow[2], JE Hixson[2], MP Stern[3], JW MacCluer[2]
[1]University of Minnesota, Minneapolis, [2]Southwest Foundation for Biomedical Research, San Antonio, [3]University of Texas Health Science Center, San Antonio

We performed a genome scan of diastolic blood pressure (DBP) on approximately 441 individuals in 10 randomly ascertained families of the San Antonio Family Heart Study. A likelihood-based Mendelian model incorporating genotype-specific effects of sex, age, age^2, BMI, and systolic blood pressure as covariates, gave strong evidence that DBP is oligogenic. Using model parameters that give the global maximum likelihood we performed two-point linkage analysis using 401 highly polymorphic markers. DBP was significantly linked to D2S1790 ($Z=3.91$, $\theta=0.00$). There was also linkage to D8S373 ($Z=1.92$, $\theta=0.00$) which is within 2cM of the candidate genes that cause glucocorticoid-remediable aldosteronism, a rare form of hypertension.

We also performed four additional genome scans using model parameters of the local maxima with the four highest likelihoods that were less than the global maximum. Maximum lodscores (and recombination fractions) for the two markers of interest were:

Maximum	D2S1790	D8S373
Global	3.91 (0.00)	1.92 (0.00)
2	2.27 (0.03)	2.71 (0.01)
3	3.31 (0.00)	1.29 (0.00)
4	1.38 (0.05)	3.19 (0.00)
5	3.79 (0.00)	2.33 (0.00)

Over all five genome scans, there was only one lodscore greater than 2 outside these two regions (D16S3253, $Z=2.28$, $\theta=0.01$). While it is not clear what significance level should be used for local maxima, it is clear that loci in the two regions around D2S1790 and D8S373 dominate the likelihood surface.

Supported by NIH HL54707 and HL45522.

62

A Single, Sequential, Genome-Wide Test to Simultaneously Identify All Promising Areas in a Linkage Scan

M.A. Province, Washington Univ., St. Louis, USA

Inflation of type I error occurs when conducting a large number of statistical tests in genome-wide linkage scans. Some have advocated stringent α-levels to protect against the high numbers of expected false-positives at the cost of more false-negatives. A more balanced tradeoff is provided by the theory of sequential analysis, which can be used in a genome scan even when the data are collected using a fixed sample design. Classical sequential tests give complete, simultaneous control of both the type I and type II errors of each individual test while using the smallest possible sample size for analysis. For fixed samples, the excess N "saved" can be used in a confirmatory, replication phase of the original findings. Using the theory of Sequential Multiple Decision Procedures (SMDP) (Bechhoffer, Kiefer, and Sobel, 1968) we can replace the series of individual marker tests with a new single, simultaneous genome-wide test, which has multiple possible outcomes, that partitions all markers into two subsets: the "signal" vs. the "noise," with an apriori specifiable genome-wide error rate. These tests are demonstrated for the Haseman-Elston approach, are applied to real data, and are contrasted with traditional fixed sampling tests in Monte Carlo simulations of repeated genome-wide scans. The sequential tests allow tighter control of both types of error, and achieve fewer total misclassifications at smaller average sample numbers than do the fixed sample ones. The SMDP works

particularly well, making only one error in all 100 genome scans of 400 markers each, as compared to 24 using Lander-Kruglyak α levels, and 13 using Bonferroni corrections. The method allows efficient identification of the true signals in a genome scan, uses the smallest possible sample sizes, saves the excess to confirm those findings, controls both types of error, and provides one elegant solution to the debate over the best way to balance between false positives and negatives in genome scans.

63

False positives and false negatives in variance components linkage analysis of quantitative traits using SEGPATH

I.B. Borecki, M.A. Province, D.C. Rao
Washington University School Medicine, St. Louis, MO

Simulation studies were used to explore the power characteristics and type I error rates in variance components linkage anlaysis as implemented in the program SEGPATH (Province et al. 1999). A diallelic quantitative trait locus (QTL) was simulated along with a marker locus with eight equally-frequent alleles. For assessment of empiric type II error rates, we assumed θ=0; for assessment of type I error rates, θ=0.5. We simulated 100 replications under each condition, calculating the proportion of samples in which the null hypothesis – that the variance attributable to the QTL is zero – is rejected.

Several issues were explored in this framework. First, since a common unit of analysis is sibpairs, we attempted to determine the smallest detectable QTL effectsize. It appears that even with unrealistically large samples, a QTL accounting for less than 10% of the phenotypic variation may not be detectable with decent power using this design and methodology. Second, simulating an increasing degree of skewness in the phenotypic distribution while maintaining a constant effectsize reveals a modest increasing trend in the power for QTL discovery, as well as good robustness against type I errors for modest to moderate degrees of non-normality. Finally, we will quantify the effect on power when pooling data from heterogeneous samples.

64

Apolipoprotein E frequencies in an elderly population

C.M.van Duijn, A.J.C. Slooter, M. Cruts, C. Van Broeckhoven, A.Hofman[1].
Erasmus UniversityRotterdam, The Netherlands.

The apolipoprotein E (APOE) gene plays an important role in lipid metabolism and the occurrence of cardiovascular disease at early age. APOE also is a genetic determinant for Alzheimer's disease (AD), an important cause of mortality in the elderly. Due to the relationship to these major disorders, APOE frequencies are expected to change with age. However, findings up to date have been controversial. In a population-based study of 6000 subjects aged 55 years or older we have studied APOE allele frequencies of the three common alleles, APOE*2, APOE*3 and APOE*4, in different age categories. Further, the potential influence of the change in frequencies with regard to bias in non age-matched case-control studies was examined. The APOE*4 allele frequency decreased from 0.17 in 55-years old to 0.12 in subjects 90 or over. The decrease in APOE*4 was to the benefit of APOE*3 because also the frequency of the APOE*2 allele slightly decreased with age. Although the decrease in the APOE*4 allele by age appeared to be limited, it resulted in false positive findings when conducting association analysis unmatched for age. The pattern of false positive findings was that of an increasing effect of the allele with increasing with age. Our study implies that age differences between cases and controls must always be adjusted for when studying late-onset disorders. Allele effects that show an increasing effect with age may be explained by age-related selection bias.

65

Statistical Strategies to Distinguish the Effects of Maternal and Offspring Genes in the Etiology of Nonsyndromic Oral Clefts

DF Wyszynski, MH Khoshnevisan, T Wu, A Miller-Chisholm, SR Diehl
Craniofacial Epidemiology and Genetics Branch, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, Maryland

For disorders in which the mother's genotype might play a critical role, such as birth defects, the classic Transmission Disequilibrium Test (TDT) can be extended to separate the effects of maternal genetic variation from that of the affected child. We developed a likelihood-based method of analysis, following the approach proposed by Weinberg et al. (1998) and Wilcox et al. (1998), to discriminate these genetic effects using case-parent triad data. Our method may be used with multiallelic markers. Since it can incorporate adjustment for environmental exposures, it is a useful tool to explore gene-environment interactions. We used this method to analyze four candidate gene polymorphisms for nonsyndromic cleft lip with/without cleft palate (CL/P): MTHFR C677T, MTHFR A1298C, TGFα TaqI, and TGFα BAMHI. Additionally, maternal and child

gene-environment interactions were tested using an extension of the case-only method of Khoury and Flanders (1996). Periconceptional vitamin use and cigarette smoking during the first trimester of gestation were incorporated into the case-only analyses of MTHFR and TGFα, respectively. Strong evidence of both fetal and maternal effects were found for all four genetic markers, with higher odds ratios when two copies of the high-risk alleles were carried versus only one or no copies. Our results suggest that there is a differential effect of these polymorphisms on risk of nonsyndromic CL/P depending on whether the high risk genotypes are carried by the mother, the child or both. This finding reinforces the view that the mother is not only a genetic parent, but also provides an important component of the fetal environment.

66

Shared genes influence quantitative variation in plasma levels of nitric oxide and LDL-related phenotypes in the San Antonio Family Heart Study.

S.A. Czerwinski, M.C. Mahaney, D.L. Rainwater, J. Blangero and J.W. MacCluer. Southwest Foundation for Biomedical Research, San Antonio, TX USA.

Nitric oxide (NO) has been shown to regulate vascular tone and influence lipoprotein oxidation. To determine the extent to which shared genes influence normal quantitative variation in plasma NO and LDL-related phenotypes, we conducted a multivariate quantitative genetic analysis. NO, APOB, APOE, total LDL, median LDL diameter and triglycerides (TG) were measured in a sample of 364 Mexican-Americans. After simultaneously adjusting for age and sex using a maximum likelihood-based variance components approach, we found significant ($p<0.05$) heritabilities (h^2) for all of the phenotypes: NOx ($h^2=0.25$), APOB ($h^2=0.49$), APOE ($h^2=0.45$), total LDL ($h^2=0.51$), median LDL diameter ($h^2=0.33$), and TG ($h^2=0.39$). We found significant positive genetic correlations indicative of pleiotropy between NOx and APOB ($\rho_G=0.58$), APOE ($\rho_G=0.37$), total LDL ($\rho_G=0.29$) and TG ($\rho_G=0.66$). We infer from these results that the genes that influence increases in plasma NO also influence increases in these LDL-related phenotypes. Consistent with the inverse relationship between particle size and LDL concentration, we found a significant negative genetic correlation ($\rho_G=-0.59$) between NO and median LDL diameter. These results show that genes shared with NO are responsible for between 8.4% and 43.5% of the additive genetic variance in LDL-related traits, with the strongest evidence of pleiotropy occurring with TG and median LDL diameter.

67

Evaluation of Candidate Genes in Family Studies: GEE and Bootstrap Approaches

J. Shin[1], S. Bull1,[2], G. Darlington[1], M. Corey[1,3]
[1]University of Toronto, Toronto, Canada; [2]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada; [3]Research Institute, Hospital for Sick Children, Toronto, Canada.

Cohort studies can be used to examine the association of genetic factors, such as candidate genes, and other risk factors with the presence or absence of complex disorders. If randomly ascertained families are investigated, possible familial correlation among observations must be considered. Two statistical approaches for analyzing correlated binary data from nuclear families are evaluated and compared. The generalized estimating equations approach (GEE), an extension of a generalized linear model, can be used to adjust for familial correlation. The relationship between covariates and the response is modeled, and the correlations among observations are treated as nuisance parameters. For comparison, two strategies from a hierarchical nonparametric bootstrap approach are proposed. One strategy (S1) bootstraps distinct families, preserving the structure and correlation within each family. A second strategy (S2) uses bootstrapped families from S1 but randomly samples offspring with replacement in each family. Each strategy accounts differently for within-family variability. A one-step logistic regression model is applied to the bootstrap data, assuming that individuals within a family are independent, thus allowing the bootstrap to determine the variation.

A simulation study is conducted in which family data are generated from an underlying multifactorial genetic model. The impact of sample size, degree of correlation and genetic model are investigated. Results from this study indicate that the bootstrap approach outperforms the GEE in terms of confidence interval coverage probabilities for small sample sizes. In terms of efficiency and bias, the bootstrap is only a slight improvement over the GEE. One drawback with the bootstrap approach is found in its computational demands.

68

Genome wide linkage analysis using genetic variance effects of asthma-associated binary, censored and continuous traits.

LJ Palmer, KB Jacobs, S Bhattacharrya, NI Leaves, AW Musk, AL James, RC Elston, WOCM Cookson, PR Burton.
TVWT Institute for Child Health Research & University of Western Australia, Australia. University

of Oxford, UK. Case Western Reserve University, Cleveland, USA.

We used novel statistical techniques to analyse a whole genome screen for asthma-associated phenotypes (n=364 Caucasian subjects in 80 nuclear families). Methods: Variance component models were fitted to the traits of interest using Gibbs sampling methods in order to generate new phenotypes - additive genetic variance residual effects (at the level of the individual) - for linkage analysis. Linkage analyses for quantitative traits, right censored survival times and binary traits were then based upon a novel implementation of the Haseman-Elston IBD sib-pair method using the sib-sib covariance as the dependent variable (HE II). In separate analyses, a focussing technique was also used to weight the IBD regression by the results of Gibbs sampling based segregation analyses. Results: The use of HE II together with additive genetic variance residual effects for asthma-associated traits greatly strengthened evidence of linkage versus use of standard HE regression and raw phenotypes. Mutual adjustment strategies allowed definition of linkages to the component of genetic variance in each phenotype that was shared or unshared with the other phenotypes (e.g., asthma independent of atopy). Use of the weighted regression approach greatly strengthened evidence of linkage to each phenotype in specific chromosomal regions. Conclusions: This study suggests that these novel methods of phenotype definition and linkage analysis may offer both the ability to correctly analyze right-censored traits for linkage and increased efficiency and power to detect linkage in studies of asthma and other complex diseases. The use of weighted focussing may allow efficient localisation of specific major genes.

69

Influence of individuals or sibships in QTL linkage studies

D.L. Koller[1], M.J. Econs[1], J.C. Christian[1], S.L. Hui[1], P.A. Morin[2], M.E. Curran[2], G. Joslyn[2], C.C. Johnston[1], M. Peacock[1], P.M. Conneally[1], T. Foroud[1]
[1]Indiana Univ. School of Medicine, Indianapolis, IN, USA; [2]Axys Pharmaceuticals, Inc., La Jolla, CA, USA.

Current searches for genes underlying complex disorders typically focus on non-parametric linkage analysis in sibships or families. Analyses of these data often rely on large-sample approximations for determination of the significance level associated with the linkage result(s). We report several situations where individuals or sibships exerted a disproportionate influence on linkage results, as computed by the ML variance estimation method implemented in Mapmaker/SIBS, in a QTL mapping study involving a genome screen of 425 sibling pairs.

1) Phenotypic outliers: Inclusion or exclusion of an individual 4 SD below the mean altered the maximum LOD score from 3.4 to 1.8, respectively, in this large sample. Inclusion of this individual also produced rapid increases/decreases in the multipoint LOD score in intermarker intervals.

2) Half-siblings: Erroneous inclusion of two discordant half-sibling pairs as full siblings resulted in a spurious increase in the lod score (to 3.49 from 2.82).

3) Unlinked subsets: A LOD score of 3.50 in 425 sibling pairs was reduced to 2.16 upon addition of 84 sibling pairs with little evidence of linkage to the region in question. The position of the peak LOD score was shifted by 18 cM as well.

4) Marker informativeness: A highly phenotypically variable five-sibling pedigree without genotyped parents resulted in discrepant two-point (LOD=1.31) and multipoint (LOD=3.38) results at the same marker position, due to the uninformativeness in this pedigree of a single marker in this linked region.

70

Detection of Interaction between Loci in Genome Scans

J. Biernacka[1,2], S.B. Bull1,[2], P. Nikolakakos[2]
[1]Univ. of Toronto, Toronto, Canada; [2]Samuel Lunenfeld Research Inst., Mount Sinai Hospital, Toronto, Canada

Genome scans are commonly used in the identification of susceptibility genes for complex disorders. However, such methods have had limited success, with many studies pointing out multiple regions with suggestive evidence for linkage, but few resulting in definite implication of specific loci. Furthermore, results cannot always be replicated by different working groups looking for susceptibility genes for the same disease. One reason for this may be that multipoint linkage analysis fails to account for the interaction of multiple genetic factors. Simultaneous consideration of susceptibility for multiple regions may increase the power to detect genes involved in the predisposition to complex disorders. Several methods to examine multiple regions simultaneously have been proposed, including the methods of Cox et al. (1999), who suggested using standard linkage analysis based on allele-sharing, with family weighting based on the evidence of linkage at an unlinked locus, believed to interact with the locus of interest. We investigated two issues in the application of this approach. To identify loci that are potentially interacting, Cox et al. suggested examining the correlation of family NPL scores at the two loci. As an adjunct to this, we investigated the value of scatterplot matrices and other more complex data visualization techniques such as 3-d plots using XGOBI (Swayne et

al, 1998). Secondly, the method proposed by Cox et al. focuses on evaluating linkage at a specified susceptibility locus identified in previous studies. However, because genome-wide scans may only provide a rough localization of susceptibility genes, we propose evaluating chromosomal regions around a locus for potential interactions. We use simulation to evaluate the significance of the maximum regional increase in the LOD score, associated with weighting by other loci.

71

Power of the affected-sib pair with assortative mating and cultural transmission

A. A. Todorov, N. L. Saccone, P.A.F. Madden, A.C. Heath
Washington University, St. Louis, MO, USA

Typically, studies of the effect of assortative mating on linkage power assume that mating is assortative with respect to the phenotype under study and that there is no residual (e.g., cultural) transmission from parent to offspring. We have implemented a general algorithm that relaxes these two assumptions. The joint parental phenotypic distribution is modeled from the marginal distributions using a copula, which permits the imposition of a constant rank correlation between mates' phenotypes, independently of the marginals. As with other approaches to this problem, we obtain $f^{(k+1)}(x_0, g_0)$, the joint distribution of oligogenotype and phenotype in generation *k+1*, by integrating/summing $f^{(k)}(x_1, x_2, g_1, g_2)\, f(x_0, g_0 \mid x_1, x_2, g_1, g_2)$ over all parental phenotypes (x_1, x_2) and oligogenotypes (g_1, g_2). The program uses adaptive quadrature integration (A. Genz). It computes $f^{(k+1)}(x_0, g_0)$ for all g_0 (including phase) and for a number of values of x_0. At the moment, intermediate values are recovered using Akima splines. Users provide three separate functions that specify (a) the genetic model under study, (b) the copula function, and (c) the transmission function. All three of which can be made dependent upon k. We are currently running analyses to determine whether the power of the affected-sib pair method can be adequately described in terms of the basic parameters of the model.

Supported in part by AA-00728, AA-12232, and MH-17104

72

Age-dependent penetrances among carriers of NAT2 polymorhisms with mutations in DNA mismatch repair genes

Frederick T. O'Donnell, Marsha L. Frazier, Christopher I. Amos
The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030

Background Mutations in N-acetyltransferase (NAT2), a highly polymorphic drug metabolizing enzyme expressed in the colorectal mucosa, may affect risk for colorectal cancer (CRC), especially among individuals with germline mutations in DNA mismatch repair (MMR) genes.

Methods NAT2 genotypes and allele frequencies were determined for 79 individuals with colorectal adenocarcinoma who also had mutations in *h*MLH1, *h*MSH2, or *h*PMS1. After each individual was tested for 3 possible NAT2 polymorphisms M1 (341 T > C); M2 (590 G > A); M3 (857 G > A), Kaplan-Meier survival curves were prepared to compare time to onset for carriers of each polymorphism. In addition, NAT2 genotype was used to determine acetylator phenotype for all MMR mutation carriers. We then compared time to onset between rapid and slow acetylators.

Results When individuals were grouped according to acetylator phenotype, no significant difference in time to onset was observed between rapid and slow acetylators. However, when individuals were stratified by NAT2 polymorphism (M1, M2, M3), those who were heterozygous for the mutant M3 allele (WT/M3) had a significantly higher risk of CRC (n=7, mean age = 41) than individuals homozygous for the wild type M3 allele (WT/WT) (n=70, mean age = 48) (p = 0.0186).

Conclusions These findings suggest NAT2 genotype to be an important factor in tumorigenesis of CRC. Further studies may be needed to better characterize relationships between NAT2 polymorphisms and age-dependent penetrances among HNPCC mutation carriers.

73

Childhood antisocial behavior and dopamine genes: Testing for mediation via regression-based extensions of the TDT

ID Waldman[1], JH Mohr[1], A Abramowitz[2], SL Sherman[3], HH Cleveland[4], JMC Gard[5], LN Giedinghagen[5], C Stever[5], DC Rowe[4,6]
Psychology[1], Psychiatry[2], and Genetics and Molecular Medicine[3], Emory University; Family Studies[4], Molecular and Cell Biology[5], Graduate Committee on Genetics[6], University of Arizona

Understanding the etiology and development of antisocial behavior is an important, albeit controversial, goal of psychiatric genetic research. Behavior genetic studies have found substantial heritability for childhood and adolescent antisocial behavior, as well as for Attention Deficit Hyperactivity Disorder (ADHD). Given the considerable overlap between ADHD and disorders

representing childhood antisocial behavior (e.g., Oppositional Defiant Disorder [ODD], Conduct Disorder [CD]), it is reasonable to hypothesize that these disorders share common causes. A number of previous studies have found association and linkage between childhood ADHD and the dopamine transporter gene (DAT1) and the dopamine receptor D4 gene (DRD4). In this study, we examine the etiological role of DAT1 and DRD4 in childhood antisocial behavior. We sampled DNA and symptoms of ADHD, ODD, and CD from clinic-referred children, their parents, and their siblings in 123 families. We examined linkage disequilibrium between DAT1 and DRD4 and symptoms of these disorders using regression-based extensions of the Transmission Disequilibrium Test (TDT) designed to accommodate multiple categorical and continuous traits. We found evidence suggesting association and linkage of DAT1, but not DRD4, with increasing levels of ODD symptoms, CD symptoms, and the hyperactivity-impulsivity symptoms of ADHD. After controlling for level of hyperactivity-impulsivity symptoms, the relation of DAT1 with both ODD and CD symptoms was non-significant. These findings suggest that the relation of DAT1 with childhood antisocial behavior appears to be totally mediated by hyperactivity-impulsivity.

74

Measured haplotype analysis of the Angiotensin-I converting enzyme (ACE) gene in Jamaican subjects

Zhu X[1], Rieder M[2], McKenzie C[3], Forrester T[3], Wilks R[3], Cooper R[1], Guo X[1], Rotimi C[1], Luke A[1].
[1]Loyola University Medical Center, Maywood, Illinois, USA; [2]Department of Molecular Biotechnology, University of Washington, Seattle, Washington, USA; [3]Tropical Metabolism Research Unit, University of the West Indies, Kingston, Jamaica

The definition of the relationship between multiple polymorphisms in a small genomic region and an underlying quantitative trait locus (QTL) represents a major challenge. Pedigree analyses have shown that angiotensin-I converting enzyme (ACE) levels are influenced by a QTL located within or close to the ACE gene. In addition, measured haplotype analyses in Caucasian families appear to have excluded upstream sequences including the ACE promoter from harboring the major ACE-linked variant. It is possible that similar analyses in unrelated individuals might be used to further refine the position of the ACE-linked QTL. We have studied a random sample of 178 individuals of Afro-Caribbean descent from Jamaica using linkage disequilibrium (LD) analysis as well as a measured haplotype analysis. An examination of 5 polymorphisms, including the insertion/deletion (I/D) polymorphism in intron 16, spanning 22 kilobases shows a graded increase in LD, from 5´ to 3´, between the polymorphisms and the putative ACE QTL. A series of nested, measured haplotype analyses was performed which showed that the I/D polymorphism is unlikely to be the ACE-linked QTL. In fact the variant which influences ACE levels is likely to lie further downstream. Our results suggest that analysis of haplotypes, which summarize linkage disequilibrium relationships between polymorphisms in small genomic regions, is a useful approach to localizing variants which influence quantitative traits.

75

A new explicit algorithm to calculate identity by descent probabilities for pairs of relatives with respect to two linked loci

H. Tiwari, R. Elston
Department of Epidemiology and Biostatistics
Case Western Reserve University, Cleveland, OH.

Denniston (1975) gave a method to calculate ibd sharing probabilities for any non-inbred pair of relatives with respect to two linked loci. However, the computation of ibd sharing probabilities by his method is cumbersome and is not a recursive algorithm. The present paper first gives a simple recursive algorithm for determining the joint probabilities that alleles are identical by descent at two linked autosomal loci for any pair of relatives in a non-inbred pedigree. We then indicate how the method can be extended to arbitrary pairs of relatives and we give several examples to illustrate the method.

76

Using Categorical Path Analysis to Estimate Direct and Indirect Effects of Genetic and Environmental Risk Factors in Epidemiologic Studies

T.G. Lesnick, W. Liu, D. Miller, R. Marks, P. Yang.
Mayo Clinic and Foundation, Rochester, MN USA

Most epidemiologic studies of complex diseases use techniques such as multiple regression to investigate relationships between outcomes of interest and several risk factors. Often, we are interested in providing causal interpretations to the estimated associations. Path analysis has been used as a method to provide these interpretations. A conceptual model of causation based on, e.g., biological principles or temporal information, is specified in the first step. Regular path analysis then uses hierarchical systems of linear models to estimate direct, indirect, and total causal effects. Spurious effects are removed in the process. Considering categori-

cal variables in the analysis involves additional challenges, and is only incompletely analogous to the regular method. Whereas the regular method partitions correlations between variables, categorical path analysis uses systems of loglinear models to estimate effects. We present the concepts, procedures, and pictorial representations used in this approach, and discuss potential obstacles. We illustrate the methodology using two analyses of factors predicting short-term survival of lung cancer patients. A combination of genotypes at glutathione-related loci (*GSTP1* and *γ-GCS*) is examined in the first, demonstrating a direct genetic effect. Another combination of loci (*GSTT1* and *GSTM1*) is in the second, demonstrating null effects. Chemotherapy, stage of disease at diagnosis, and pack years smoked are considered in both examples.

77

Identifying false linkage signals from neural networks

NL Saccone[1], JP Rice[1], TJ Downey Jr[2], DJ Meyer[2], RJ Neuman[1], N Rochberg[1], A Goate[1], HJ Edenberg[3], T Foroud[3], T Reich[1]
[1]Washington Univ., St. Louis, MO; [2]Partek Inc., St. Peters, MO; [3]Indiana Univ., Indianapolis, IN

Neural networks have drawn attention as tools for linkage analysis of complex traits. Current methods use neural nets to classify subjects by phenotype, using genetic information as input. Loci important for classification are potentially linked to the trait. These methods can produce high false positive rates, leading to concerns about their usefulness.

We note that neural net classifiers do not inherently look for segregation of a genetic locus with a trait. For example, we have used neural nets to classify sib pairs as phenotypically concordant or discordant, using IBD sharing. Any locus at which sharing is markedly different for concordants versus discordants may lead to a signal; this may include loci where discordants share more often due to chance. To rule out such signals, we suggest a simple screen of the sharing patterns at signaled loci.

We illustrate using the COGA (Collaborative Study on the Genetics of Alcoholism) data. On chromosome 1, neural nets indicated two signals; only the second was supported by traditional methods. When we examined the first locus, 24% of the concordant pairs showed low sharing, 49% medium and 27% high, while discordants gave 18% low, 57% medium and 24% high. In contrast, the second signal corresponded to patterns of 22%, 45%, 33% in concordants versus 34%, 47%, 19% in discordants, which appears more indicative of linkage. We are refining these tests, as neural net models may detect interacting loci that may not appear linked to the trait when examined alone.

78

Transmission disequilibrium analysis to identify a modifier gene for cystic fibrosis (CF) in the chromosome 19q13.2 region.

Corey M[1,2], Zielenski J[1], Markiewicz D[1], Rozmahel R[1], Aznarez I[1], Shin J[1], Durie P[1,2], Tsui LC1,[2], Collaborative CF Modifier Study Group.
[1]Hospital for Sick Children, [2]University of Toronto, Toronto, Canada.

Clinical phenotype in CF is highly variable, depending on CFTR genotype as well as secondary factors. We found evidence for a candidate locus (Zielenski et al. Nature Genetics in press) that modifies the severity of the intestinal phenotype, characterized by the presence or absence of meconium ileus (MI), a severe bowel obstruction affecting 15–20% of CF patients at birth. The region with highest linkage was a 3.75 Mb interval between markers D19S408 and D19S412 with maximum at the D19S112 marker. Since the region was still too large for systematic transcript analysis, transmission disequilibrium testing (TDT) of markers and known genes in the region was undertaken. The study group consisted of 142 families with at least one child with MI. We used GASSOC (Schaid, Genet Epidemiol 1996;13:423–49) to provide a general score test of association for each marker tested, and GENEHUNTER (Spielman et al, Am J Hum Genet 1993;52:506–16) for allele-specific TDT's. To date we have performed TDT analysis for 13 microsatellite markers in the linked region, and 2 single nucleotide polymorphisms (SNP) in candidate genes CALM3 and STD. The most promising detection of transmission disequilibrium was obtained for the APOC2 marker (GASSOC $p=0.02$, with GENEHUNTER pointing to two 2 specific alleles with $p<0.01$). These results were based on a subset of 33 families and must be confirmed in the larger sample. No association was suggested by SNP analysis for candidate genes CALM3 and STD (93 families, $p>0.3$). TDT analysis continues with emphasis on markers surrounding the APOC2 locus.

Collaborative CF Modifier Group: T.Casals, S.Larriba, B. Mercier, G.Cutting, M. Macek, A. Krebsowa, E. Bjorck, B. Strandvik, A. Palacio, E. Langfelder, B. Marshall, M. Claustres, D. Bozzn, A. Nowakowska, J. Bal, C. Ferec, X. Estivill.

79

Evaluating The Evidence For A Relationship Between Fatty Acid Levels And Atopy In An Isolated Caucasoid Population.

R. Mathias[1], F. Chilton[2], L. Freidhoff[1], R. Nickell, M. Beaver[2], B. Barham[2], C.Bickel[1], T. Beaty[1], K. Barnes[1].

[1]Johns Hopkins Univ., Baltimore,MD,USA; [2]Wake Forest Univ., Winston-Salem, NC, USA.

A number of mediators that play a role in allergic airway disease are synthesized from Arachadonic Acid (AA), which belongs to the n-6 series of polyunsaturated fatty acids (PUFAs). Epidemiologic data suggest that an increase in the prevalence of asthma and atopy in developed countries parallels an increase in n-6:n-3 PUFAs, the former giving rise to potent inflammatory mediators. We studied the relationship between n-6 PUFA levels and atopy in the isolated Tangier Island, VA population, where the prevalence of self-reported allergic rhinitis (AR) is 59%. We tested for linkage of skin test confirmed AR and 'high' IgE to 16 microsatellite markers spanning 70cM on chromosome 12q. Using Transmission Disequillibrium Test, we found evidence of linkage to AR (n=33) and 'high' total IgE (n=43) at D12S1052 (AR: global P=0.01; allelic, P=0.01; 'high'IgE: allelic, P=0.04). Categorizing PUFA levels into extreme quartiles, we observed modest evidence for differences in mean Log[IgE] (AA p=0.016; Linoleic acid p=0.053; Dihomo-gamma-linoleic acid p= 0.020) (n=217). No such trends were observed for any of the n-3 PUFAs examined. Association analyses were done for the extreme quartiles of the PUFAs and 7 markers on 12q. Modest evidence for associations between the same marker alleles in linkage disequillibrium with AR and 'high' IgE were observed for several of the n-6 PUFAs (n=153) (D12S58, allele 83: AA, p=0.035; LA, p=0.053; allele 89: GLA, p=0.046; D12S1052, allele 157: AA, p=0.049). This indicates an association between elevated n-6 PUFAs and atopy in residents from Tangier and supports the putative role of diet in atopy.

80

Effect of apoE polymorphism on β-lipoproteins in Mexican Americans

W.-C. Hsueh, B.D. Mitchell, D.L. Rainwater, J.E. Hixson
Southwest Foundation for Biomedical Research, San Antonio, TX

The apolipoprotein E (apo E) isoforms are associated with variation in serum lipids and apolipoproteins and may influence risk of CVD. We evaluated the impact of these isoforms on lipid metabolism relative to that of other known CVD risk factors (e.g., lifestyle factors and other genetic effects) in the San Antonio Family Heart Study, a population-based study of atherosclerosis and its determinants in Mexican American families. This study is based on 859 subjects, aged 16–92 yr. from 41 families examined between 1991–1996, who possessed at least one ε3 allele. We measured concentrations of apoE and the major constituents of HDL (cholesterol, apoAI, and apoAII) and β-lipoproteins (cholesterol and apoB).

The heritability of these traits ranged from 34–55%. The distribution of apoE isoforms was E2/E3: 6.3%, E3/E3: 81.4%, and E3/E4: 12.3%. After adjustment for age, sex, and lifestyle risk factors, subjects with E2/E3 had significantly higher concentrations of apoE and significantly lower concentrations of non-HDLC and apoB than E3/E4 individuals; there was no effect of apoE isoforms on HDL measures. The effect of apoE isoforms was significantly greater on apoB levels in women than in men (sex × apoE interaction p value < 0.05). Using maximum likelihood methods, we partitioned variation in lipid and lipoproteins into components attributable to the effects of age, sex, lifestyle variables, apoE isoforms, and residual genetic factors. Together, these factors accounted for 40–60% of the total variability in these traits. The apoE isoforms accounted for 4.4%, 12.4%, and 4.8% of the total genetic variation in apoB, apoE, and non-HDLC, respectively. We conclude that apoE isoforms accounts for a modest, albeit significant, proportion of phenotypic variation in β-lipoprotein in this population and that these isoforms have a greater effect on apoB levels in women than in men.

81

Estimation of variance components in an inbred pedigree

M. Abney[1], M.S. McPeek[2], C. Ober[1]
[1]Dept of Human Genetics, [2]Dept of Statistics, University of Chicago, Chicago, IL USA

Inbred pedigrees present special challenges for variance component estimation. First, there are more genetic variance components to be estimated in the inbred case because of the non-zero probability of being homozygous by descent, even for a relatively simple genetic model. Second, there are more identity coefficients that must be calculated in an inbred pedigree in order to perform the estimation, and these are computationally more difficult to obtain in the inbred than in the outbred case. As a result, inbreeding effects have generally been ignored in practice. We describe here the estimation of variance components of quantitative traits in large inbred pedigrees, using the example of the Hutterites. We use additional simulated examples to give an indication of under what conditions one has the power to detect the additional variance components and to examine their impact on variance component estimation. We discuss the implications for mapping and heritability estimation.

M.A. was supported by the Alfred Sloan/DOE postdoctoral fellowship in computational biology.

82

Multipoint disequilibrium mapping by the decay of haplotype sharing

M.S. McPeek[1], A. Strahs[1]
[1]Dept of Statistics, University of Chicago, Chicago, IL, USA

We propose a multilocus model, the decay of haplotype sharing model, for linkage disequilibrium around a mutation of relatively recent origin. This allows a likelihood approach to disequilibrium mapping explicitly incorporating chance sharing, mutations, heterogeneity, and multiple origin of the variant. Our implementation uses a hidden Markov method that permits application not only when complete haplotype data are available, but also allows for missing data, partially determined haplotypes from nuclear family data, and for the case when only genotype data, rather than haplotype data, are available. While the dependence among nearby loci is explicitly modeled in the likelihood, dependence among haplotypes due to population structure is taken into account using a quasi-likelihood approach, based on first and second moments. We implement the quasi-likelihood approach with a conditional coalescent assumption, which should arguably yield conservative results in a growing or stable-size population. Extensive simulation results and application to published data sets show that this approach works extremely well for gene localization, and that the coverage of confidence intervals for gene localization is excellent.

83

Linkage disequilibrium between IL-4 receptor polymorphisms in various samples of populations

J Xu[1], TD Howard[1], C Ober[2], G Koppelman[3], DS Postma[3], ER Bleecker[1], DA Meyers[1], and CSGA
[1]Univ. of Maryland, Baltimore, MD; [2]Univ. of Chicago, Chicago, IL, [3]Univ. Hospital Groningen, Groningen, The Netherlands

Interleukin-4 receptor (IL-4R) is a key component in the induction of IgE production. Multiple polymorphisms of the IL-4R gene have been identified, and some have been reported to be associated with a higher risk of atopy or atopic asthma. We studied four bi-allelic polymorphisms (E400A, C431R, S503P, and Q576R) within a single exon of IL-4R in four different ethnic samples. The distance between the polymorphisms is 93bp, 216bp, and 219bp, respectively. The Dutch sample was from Groningen in Northern Holland, ascertained through an asthmatic proband. The Hispanic sample was ascertained in New Mexico, the Caucasian and African American samples were ascertained in the Baltimore area, all through asthmatic sib-pairs, as part of the CSGA (NHLBI) study on the genetics of asthma. Standardized pair-wise linkage disequilibrium (LD) coefficient (D´) between polymorphisms was estimated from unrelated pedigree founders and its significance was evaluated using the Fisher exact test. The Dutch sample (2n=516) showed highest LD between all pairs of polymorphisms, with D´ between 0.85 to 1, p<1E-5. The Hispanic (2n=130) and Caucasian (2n=141) samples showed less LD between polymorphisms, with D´ from 0.72 to 1.00, p<1E-5. The African-American sample (2n=77) generally did not show significant LD, with D´ from 0.19 to 0.58. The D´ between E400A and C431R (93bp) was estimated to be 0.35, p=0.43. These results provide useful information in designing LD mapping studies using different populations. Relatively homogeneous populations such as Dutch, Hispanic, and US Caucasians can be used to show whether a specific candidate gene is associated with a phenotype, while the African American population may be used to pinpoint the causal polymorphism.

84

Evaluating evidence for association of IL-4 and IL-4 receptor (IL-4r) polymorphisms with occupational allergy

DG Wiesch, J Xu, PJ Amelung, ER Bleecker, DA Meyers
Univ. Maryland, Baltimore MD

A case-control study of genetic susceptibility and gene-environment interaction to Laboratory Animal Allergy (LAA), allergy from occupational exposure to animals, was performed in employees of The Jackson Laboratory, Bar Harbor, Maine. LAA affords an excellent opportunity to study the genetics of allergy as a simplified disease model, especially in these workers, due to the ubiquitousness of the environmental mouse exposure and the relative ethnic homogeneity of the population. Association of genotypes in the IL-4(-590) promoter polymorphism and three IL-4r polymorphisms (Q576R, E400A, and I75V) were compared in 46 LAA cases, employees with LAA symptoms and sensitization to mice, and 85 LAA controls, workers with no LAA symptoms or sensitization. No evidence of association with LAA was found for any of the genotypes. Association was also investigated in an expanded sample of 192 employees with atopic phenotypes including skintest reactivity, total serum IgE, and mouse specific IgE. Marginally significant associations were observed for codominant genotypes in IL-4(-590) and quantitative total serum IgE (p=0.05) and for IL-4r(576) and the presence of mouse specific IgE (p=0.02 for AA vs GG or GA genotypes). However, Bonferroni's

correction for multiple comparisons resulted in non-significant p-values. Haplotype association analyses with IL-4r polymorphisms did not show significance for LAA or any atopic phenotype. No evidence was found for gene-gene or for gene-environment interactions considering different mouse exposure levels. The sample size limits definitive conclusions regarding these polymorphisms. Thus, more employees will be collected to increase power and additional candidates will be tested.

85

Maternal Multivitamin Use, Genetic Variation in Folate Metabolism and Risk of Oral Clefts

M. H. Khoshnevisan,[1] D. F. Wyszynski,[1] T. Wu ,[1] A. Miller-Chisholm,[1] D. Ruggles,[1] E. Turon,[1] M. Mazaheri,[2] R. Long,[2] E. O'Connell,[2] L. Lukes,[1] P. Le,[1] J. Jones,[1] S.R. Diehl[1]
[1]Craniofacial Epidemiology and Genetics Branch, National Institute of Dental and Craniofacial Research, National institutes of Health, Bethesda, Maryland; [2]Lancaster Cleft Palate Clinic, Lancaster, Pennsylvania

Purpose: The purpose of this study was to evaluate the association between single nucleotide polymorphisms (SNPs) at the methylenetetrahydrofolate reductase (MTHFR) gene, maternal multivitamin consumption and risk of nonsyndromic cleft lip with or without cleft palate (CL/P) in a Caucasian population.

Methods: We studied 234 families (933 individuals) with one or more CL/P affected member. A questionnaire was used to collect clinical, family history and risk factors information. Blood samples were obtained from both affected and unaffected family members. Two SNPs in the MTHFR gene (C677T and A1298C) were assayed by standard PCR-RFLP techniques. We performed a transmission disequilibrium test (TDT) and a case-case analysis to evaluate linkage disequilibrium (association) and interactions with maternal multivitamin consumption.

Results: Without considering multivitamin consumption, we found significant ($p<0.05$) evidence of disequilibrium with risk of cleft lip for the C677T SNP, but no association with the A1298C polymorphism for all subjects. However, when we incorporated multivitamin consumption into our analyses, we found a much stronger association of CL/P risk for the C677T MTHFR SNP for mothers who did use multivitamins during their first trimester of pregnancy ($p< 0.005$). We also found a weak association with the A1298T MTHFR SNP for mothers who did not use multivitamins.

Conclusion: These findings indicate the presence of a strong gene by environment interaction for MTHFR SNPs and maternal multivitamin consumption.

86

Genetic Epidemiology of the GENNID study I: covariate-adjustment and linkage analysis in the Mexican-American cohort.

M. M. Barmada, C. E. Aston
Dept. of Human Genetics, Univ. of Pittsburgh, Pittsburgh, PA, USA.

We have begun an analysis of the GENNID (Genetics of Non-Insulin Dependent Diabetes) study by comparing the results of univariate linkage analysis in the GENNID Mexican-American cohort with previously reported genome-wide analyses in other type 2 diabetes studies. The significant amount of data collected in the GENNID study enabled quantitative trait analysis to be undertaken. The specific quantitative traits analyzed were fasting insulin and glucose measurements, baseline-adjusted 1- and 2-hour glucose and insulin measurements (two of the time points measured during an oral glucose tolerance test), and two derivative measures of insulin resistance and secretion. Quantitative traits were adjusted before analysis to compensate for the effects of environmental variables: smoking status, age, sex, BMI (body-mass index), dietary measures, and physical activity. Due to the observed collinearity between the environmental variables, principal component transformation was applied to allow standard regression methods to be employed for covariate-adjustment. This method additionally allowed us to assess the relative importance of each environmental variable in accounting for variation in the given quantitative trait. Following covariate-adjustment, the residual values in each trait and the genotype data provided in the GENNID study were analyzed using Haseman-Elston regression. The results highlight a number of regions previously implicated in type 2 diabetes, as well as some interesting regions unique to this study.

87

The estimation of family correlations under nonrandom sampling

J. P. Rice, N. L. Saccone, J. Corbett
Washington Univ., St. Louis, MO

We assume a set of variables $X_1,...,X_n$ have a standard normal distribution. If values of X_n are selected, then the correlations before selection may be estimated from the (normal) conditioned distribution in the selected sample. Although this is well known in the literature, it has not been widely recognized in human family studies where probands are selected (either directly on the measure or for related disease status), and their relatives subsequently examined. We discuss this problem and consider an estimator

that can be used in situations where the proband's score is not available, or only the variance-covariance matrix in the selected population is available for analysis. We performed a set of simulations where observations were sampled from a trivariate normal distribution and compared the efficiency of the method omitting the proband's score to the full iterative likelihood method. The average efficiency ranged from 64 to 86%.

88

Robustness of the single-locus model for a two-locus quantitative trait

C.M. Vachon[1] and L.D. Atwood[2]
[1]Mayo Clinic, Rochester, MN, USA. [2]University of Minnesota, Minneapolis, MN, USA.

In the investigation of a quantitative trait (QT), a linkage analysis is often performed using initial parameter estimates from a segregation analysis of the QT. However, the segregation analysis assumes a one-locus, two-allele underlying model and is often misspecified with multiple genes responsible for the variability in the QT. We address the question of robustness of the single-locus model with respect to linkage information when the QT is actually affected by two genes. Specifically, using maximum likelihood parameter estimates from a single-locus model, we performed two-point linkage analysis of markers flanking two genes that had an additive effect on the QT. We examined 24 generating models comprising all possible combinations of: allele frequency (q) set to .1 or .3, the mode of transmission set to additive or dominant, and the percentage of variation affecting the trait due to the two genes varied in six ways i.e. 60-0, 50-10, 40-20, 30-30, 30-30polygenic and 30-0. For each generating model, 100 replications of 200 nuclear families with three offspring were generated using GASP. Maximum likelihood estimates were found using REGCHunt and linkage was performed using LODLINK. Our results show that the single-locus model is robust and that linkage was detectable when the model is misspecified. Among the two gene models (50-10, 40-20, 30-30), the worst case for detecting any single gene was the 30-30 case; the average lodscores ranged from 1.35 to 2.87. Unlinked markers had an average maximum lodscore of 0.09. In general, across models, those with smaller allele frequencies had greater LOD scores, with the dominant loci higher than the codominant loci. Power curves for the two gene models showed decreasing power to detect linkage from the 50-10 to 40-20 to 30-30 models. The polygene model (30% to one gene and 30% polygenes) had slightly lower power than the 30-30 case.

89

Linkage disequilibrium of 10 SNPs within a 10 kb region of the AI-CIII-AIV gene cluster in Familial Combined Hyperlipidemia (FCH) probands and spouses

R.M. Cantor[1], M. Groenendijk[2], G.M. Dallinga-Thie[2]
[1]Univ. CA, Los Angeles, USA; [2]Univ. Med Ctr., Utrecht, Netherlands.

Successful application of SNP technology to identify complex disease gene variants is contingent upon multiple factors, including informative patterns of linkage disequilbrium (ld). Disease mutation haplotypes increase in frequency and narrow with recombination in affecteds, which may lead to distorted patterns of ld on disease chromosomes. We have reported increases in specific AI-CIII-AIV haplotypes in FCH probands and are examining ld patterns within this region in probands and informative controls (marrying-in spouses with at least two haplotyped offspring). Simwalk was used to haplotype 10 SNPs in a 10 kb region of the AI-CIII-AIV gene cluster. From these haplotypes, 45 measures of pairwise ld were estimated separately in 52 disease and 192 control chromosomes, using a statistic that summarizes ld for all possible haplotypes. Fishers' exact test assessed the significance of ld. As expected, ld did not correlate with physical distance and was increased in FCH chromosomes; ld estimates ranged between .02 and .92 (median = .21) in controls and .18 and 1.0 (median = .49) in probands. For 11 control and 10 proband ld tests, $p>.05$. SNP pairs having non-significant ld were fairly consistent across both samples. However, one pair of adjacent SNPs had significant ld in the proband sample ($p=.003$), while the p-value was .85 in controls. Comparison of SNP ld in disease chromosomes with background ld in controls may be used to help focus the search for complex disease gene variants.

90

Evidence for genetic influences on the change in percent body fat over time in Mexican Americans.

A.G. Comuzzie[1], B.D. Mitchell[1], J. Blangero[1], J.W. MacCluer[1], M.P. Stern[2].
[1]Southwest Foundation for Biomedical Research, San Antonio, TX, and [2]The University of Texas Health Science Center, San Antonio, TX.

As part of the ongoing work in the San Antonio Family Heart Study (SAFHS), follow-up data is being collected on a variety of phenotypes in order to look for genes related to their change over time. Here we have examined changes in percent body fat (Δ%FM) using data on 395 SAFHS participants (244 M; 151 F). Delta values were obtained by subtracting the value

at time 1 from the value at time 2 (mean time span between measurements = 4.7 years). A highly significant ($p < 0.0001$) additive genetic heritability of 47% was detected for this trait. Subsequently, complex segregation analysis of this phenotype revealed evidence of an anonymous major gene. The environmental model, which allows for no genetic transmission, was rejected as being a worse fit to the data than the general model ($p = 0.007$) while the mixed Mendelian model could not be rejected ($p = 0.58$). Based on the Mendelian model, homogygous recessive individuals (aa) gained approximately 9 to 10 times the percent body fat of homogygous dominant individuals (AA) (Males means: AA = 4%, Aa = 17%, aa = 30%; Female means: AA = 4%, Aa = 4%, aa = 36%). Based on the analysis of this preliminary data set we conclude that changes in body composition appear to be significantly influenced by genetic effects. These findings suggest that these phenotypes could be informative for use in future linkage analyses in order to identify specific genes involved in changes in body composition over time. Supported by NIH grant HL45522.

91

Sibling Correlations of Echocardiographic Left Ventricular (LV) Measurements Derived from Factor Models: The HyperGEN Study

W Tang[1], R Devereux[2], M Province[3], L Atwood[1], D Arnett[1]
[1]Univ. of Minnesota, MN; [2]Cornell Univ. Medical College, NY; [3]Washington Univ., MO.

We evaluated whether factor analysis scores for echocardiographic measures of LV structure and function exhibited greater familial correlation than individual measures of wall thickness, LV dimension, and systolic and diastolic function. Factor models were derived for Caucasian (C) and African-American (AA) hypertensive sibships (907 sibpairs) and 476 random subjects. Sibling correlation coefficients (SCCs) for individual measurements and the factors derived from the models are listed in the table. Factor 1 (in AA) and factor 2 (in C) exhibit higher SCCs than the individual components, but SCCs for factor 3 and 4 are similar to those of individual measures.

	SCC	SCC for Individual Variables Loading to Factor		
	Factor 1	**PWT**	**IVSD**	**LVID**
C	0.17	0.20	0.16	0.22
AA	0.30	0.17	0.15	0.29
	Factor 2	**ESSM**	**FS**	**LVID**
C	0.29	0.13	0.15	0.22
AA	0.14	0.15	0.12	0.29
	Factor 3	**MVI3**	**MVA**	
C	0.15	0.11	0.33	
AA	0.26	0.20	0.30	
	Factor 4	**MVE**	**MVA**	
C	0.32	0.32	0.33	
AA	0.25	0.28	0.30	

Structural measures: PWT, posterior wall thickness; IVSD, interventricular septum dimension; LVID, left ventricular internal dimension. Functional measures: ESSM, meridional end-systolic stress; FS, fractional shortening; MVE, mitral valve (MV) early filling phase peak; MVA, MV atrial filling phase peak; MVI3, MV 1st third filling fraction.

92

Kinship Inference Using Identity-By-State Information

Jeff T. Williams and John Blangero
Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas USA.

A measure of relationship between individuals is required for many kinds of genetic analysis. For example, in variance component analysis the relationship matrix 2Φ, where Φ is the kinship matrix, is the structuring matrix for the component of variance due to additive genetic factors. If pedigree relationships are known, the relationship matrix 2Φ can be computed easily using established algorithms. Common practice, however, is to assume that founders are unrelated. For many populations the assumption of noninbred founders is probably suspect, but the extent to which the assumption is violated is difficult to quantify. We describe a bias-corrected relationship matrix T based on identity-by-state information that estimates accurately the relationship matrix 2Φ. Knowledge of pedigree relationships is not required, consequently the method is especially useful for genetic studies of inbred and wild populations. Results from simulation studies and real data sets illustrate the accuracy with which the expected relationship matrix can be recovered from identity-by-state information at a series of genetic markers. Genotypic information was simulated for 5–90 microsatellite loci using pedigree structures with various degrees of inbreeding. Markers were simulated based on human genetic data to be highly polymorphic. The correlation between the estimated (T) and expected (2Φ) relationship matrices increases rapidly with number of typed markers, and nearly complete relationship information can be inferred from identity-by-state information when only 30–40 microsatellite markers are used.

93

Identifying a functional polymorphism through linkage analysis conditional on measured genotype

L. Almasy, J. Blangero
Southwest Foundation for Biomed. Res., San Antonio TX

Association analyses in regions showing linkage may help to narrow the candidate region around a QTL as linkage disequilibrium is likely to extend over a shorter distance than linkage. Several markers within a candidate region may be in disequilibrium with a QTL, making it necessary to differentiate between a functional polymorphism and variants in disequilibrium with it. If there is a single functional polymorphism in a QTL, we expect that linkage analyses conditional on a fixed-effect measured genotype analysis of the polymorphism would yield a LOD score of zero. Alternatively, if there are multiple polymorphisms or if the genotyped marker is merely in disequilibrium with the sole functional variant, we expect linkage analyses conditional on measured genotype to produce a non-zero LOD score. We have tested these hypotheses in simulated data from Genetic Analysis Workshop 10, re-binning the 20 alleles used to produce the diallelic QTL to form new markers in varying degrees of disequilibrium with the QTL. In analyses with the QTL itself, the conditional LODs ranged from 0 - 1.91, with 92% under 0.588 and 43% being zero. Conditional LOD scores increased rapidly for markers in disequilibrium with the QTL. For a marker with a correlation of 0.73 with the QTL, only 17% of conditional LOD scores were under 0.588 and only 1.5% were zero. These results suggest that linkage analyses conditional on measured genotype will be a useful tool in differentiating between functional polymorphisms and variants in disequilibrium with them.

Supported by NIH grants GM18897 and MH59490.

94

A survey of genome scan linkage results from human complex diseases and traits: do loci cluster and why?

A.D. Paterson, A. Petronis
Neurogenetics Section, CAMH, University of Toronto, Toronto, Ontario, Canada.

Numerous genome-wide linkage studies of human complex disease and traits have been performed with the aim of assisting in the identification of susceptibility loci. Previously it has been suggested that loci for a variety of autoimmune diseases cluster, suggesting either closely linked loci, or pleiotropy. A confound that may also produce clustering of loci is that of transmission ratio distortion (meiotic drive or segregation distortion). We have searched the literature for genome scans in human complex diseases and traits that have used > 150 microsatellite markers. We documented any lod scores ≤ 2 (or equivalent statistic), and listed their position on a sex-averaged genetic map. 169 genome scans in complex diseases and traits were identified – eight did not produce a single lod score ≤ 2. We observed the pattern of lod scores ≤ 2 across all chromosomes. On many chromosomes no clear pattern of clustering of positive lod scores was observed. As expected, a wide variety of autoimmune diseases clustered on chromosome 6p around the HLA region. Interestingly, the pericentromeric region of chromosome 10 also demonstrated a cluster of linkage results in schizophrenia, bipolar, type 1 diabetes, Crohn's disease, obesity and blood pressure. Analysis of 40 CEPH families showed weak excess allele sharing in females at the same region on chromosome 10 suggesting possible transmission ratio distortion. Transmission ratio distortion should be excluded for regions that demonstrate clustering of linkages since it may confound mapping studies.

95

Gene-environment interaction models for complex traits with molecular discrete data: application of structural modelling to Mood Disorders.

Fabio Macciardi[1], Alberto Morabito[2]
[1]CAMH, Univ. of Toronto, Toronto, Canada; [2]Univ. of Milan , Milan , Italy.

It is well known that both genes and environment factors may play a role in the aetiology of psychiatric disorders, like schizophrenia or bipolar mood disorder. Advancement in molecular genetic techniques and the contemporary development of statistical genetics allowed formulating a theory suitable for the hypothesis testing of specific transmission mechanisms from the single major locus to polygenic models. Moreover, epidemiological investigations showed that also environmental risk factors seem to be correlated with major psychiatric disorders. Thus moving toward a more integrated approach of genetic and non-genetic risk factors is now the major task that the investigators have to accomplish in the search for the causes of psychiatric disorders. Usually the investigations in this field are restricted to situations in which the genetic trait and the environmental variables are continuous and asymptotically gaussian. Our current approach extends the possibility of modelling the interaction under the hypothesis that the genetic mechanism would be discrete, using genotypes or alleles at potential candidate loci. We genotyped 100 cases (patients with Mood Disorders, bipolar) and 100 healthy controls for several genes potentially related to bipolar disorders and we investigated dimensional non-genetic variables also potentially related to the disease expression, i.e. the degree of social adjustment and the self esteem. Information about the number of depressive and manic events and symptoms were also available. A causal model with observable and latent variables describing environmental and genetic components and a latent ordinal variable repre-

senting the bipolar trait was adopted and analyzed with structural modelling. The generalized linear model has been used to check the validity of the model.

96

Power of Quantitative Transmission Disequilibrium Test in evaluating the genetic components of pharmacological response in psychiatric disorders.

Fabio Macciardi, Emanuela Mundo
CAMH, Univ. of Toronto, Toronto, Canada

One of the problems in genetic studies on complex traits, and particularly on psychiatric disorders, is the definition of the phenotype. In most cases, the phenotype has been identified with the clinical diagnosis, but there are alternative possibilities for a better definition of a psychiatric phenotype, considering also biological elements (e.g. hormonal response to challenge studies, determination of peripheral markers of neurotransmitters etc.). Using the clinical response to psychotropic compounds as a way to define a phenotype of interest in psychiatric diseases represents an innovative strategy. In such an approach, the analytical methods should include both qualitative and quantitative measures of the pharmacological response and should include the polymorphisms of the genes potentially implicated in the pharmacological response to the compound of interest. The qualitative analysis of the pharmacological response by genotypes and alleles can be performed with classical association and Transmission Disequilibrium Test (TDT), and it does not present methodological problems. On the other hand, the quantitative analysis of the pharmacological response would require a more complex strategy. The quantitative Transmission Disequilibrium Test (Q-TDT) is a model recently developed by some Authors and is a TDT procedure modified for quantitative phenotype (e.g. severity of symptoms, percentage of improvement of symptoms induced by treatment). Current available quantitative methods, however, only approximate the qualitative TDT design. We present a comparison of the relative power of TDT versus Q-TDT in a simulated set of data, where genotypes for a candidate gene and both a qualitative and a quantitative measure for the phenotype (i.e., pharmacological response to a given psychotropic drug) are available.

97

Comparison of Generalized Estimating Equations, Logistic Regression and Haseman-Elston Regression for a Dichotomous Trait.

M. Nicolaou, A. DeStefano, L. Farrer, L. Adrienne Cupples.
Boston Univ. Schools of Public Health and Medicine, USA.

Multiple sibling families are common in linkage studies, resulting in correlated data. If this issue is not addressed in the analysis, an increased number of false positive linkages may be observed. We used simulated data to compare the Type I error and power of generalized estimating equations (GEE), logistic regression and Haseman-Elston regression (H-E). The data sets generated consisted of nuclear families with 2, 3 and 5 siblings. Different models were studied with varying penetrance and heterogeneity among families resulting in different within family correlations. The three correlations studied were 0, 0.25 and 0.4. We found that logistic and H-E regression exhibited nominal Type I error with the independent data while GEE was slightly conservative. With larger families and increased correlation, the Type I error of logistic and H-E regressions increased to higher values than expected while GEE with both independent and exchangeable correlation structures performed at the nominal Type I error rate. The power was similar among these methods, although the higher Type I error of logistic and the H-E with larger correlations suggest GEE is more appropriate in this situation.

98

A stepwise oligogenic segregation analysis showed the existence of a second major locus underlying serum angiotensin I-converting enzyme levels.

Guo X[1], Cooper R[1], Ward R[2], Luke A[1], Zhu X[1], Rotimi C[1]
[1]Loyola University Medical Center, Maywood, IL;
[2]Oxford University, Oxford, UK

Human serum angiotensin I - converting enzyme (ACE) levels are highly heritable and are influenced by major genetic factors. One quantitative trait locus (QTL) is located within or close to the ACE locus itself. Some studies have demonstrated evidence of association between the I/D polymorphism and plasma ACE levels and/or blood pressure, although the effect of the ACE I/D polymorphism has been reported inconsistently. Rigat et al. (1990) found that the I/D polymorphism in the ACE gene accounts for half the variance of serum enzyme levels and that plasma ACE activity increases in a codominant fashion with the D allele. McKenzie et al. (1995) found that two QTLs jointly influence serum ACE levels in a Jamaican sample. As part of an ongoing study on the genetics of hypertension we examined 1060 African Americans in 267 families from Maywood, IL. As expected, the I/D polymorphism was significantly correlated with plasma ACE levels, with the D allele associated with increased

ACE levels, both in men and women (p=0.0062 and 0.0001 for men and women, respectively). However, only marginal associations were found for SBP in women (p=0.0354) and for DBP in men (0.0498). After controlling for the effect of the I/D polymorphism, a stepwise oligogenic segregation analysis indicated the existence of an additional major locus in the residual phenotype in this African-American sample.

99

Rheumatoid arthritis severity and dosage of the shared epitope: Results of a meta-analysis.

J.J. Chen[1], J.D. Gorman[2], G. Thomson[3], M.E. Suarez-Almazor[4], L.A. Criswell[2].
[1]Saint Louis University, MO, [2]UC San Francisco, CA, [3]UC Berkeley, CA, [4]University of Alberta, Edmonton.

To determine whether the risk of severe RA depends upon the dosage of the HLA-DRB1 shared epitope (SE), we performed an electronic literature search using MEDLINE, followed by hand searches of 6 journals. Studies reporting molecular typing methods and at least 1 of 6 outcomes of interest were selected. Authors were contacted for incomplete data and to provide original data sets. Woolf's weighted odds ratios (OR) were pooled across studies and tested for homogeneity. When possible, ethnic and clinical subgroup analyses were performed. 41 studies met inclusion criteria, with a total of 5433 patients from 17 countries on 5 continents. ORs for the association of 1 SE copy (OR_1) and 2 SE copies (OR_2) vs 0 SE copies with each outcome are shown below. A strong dosage effect of the SE was apparent only for joint surgery [OR for 2 vs. 1 SE copies = 2.07 (1.27–3.37)]. Breslow-Day tests for homogeneity indicated substantial heterogeneity across studies for all outcomes except joint surgery. Subgroup analyses indicated that ethnic variation explains at least part of this heterogeneity.

	# Studies	# Pts	OR_1 (95%CI)	OR_2 (95%CI)
RF+	18	2295	1.8 (1.4–2.4)	2.6 (1.8–3.8)
RA Course	8	832	1.5 (1.0–2.2)	1.7 (1.0–2.8)
Erosion	23	2872	1.8 (1.4–2.2)	2.2 (1.6–2.9)
Joint surgery	7	503	1.8 (1.0–3.3)	3.7 (1.9–7.4)
Nodules	22	2221	1.5 (1.2–2.0)	1.7 (1.3–2.3)
Other EAMs	13	1269	1.2 (0.8–1.8)	1.8 (1.1–2.8)

This meta-analysis demonstrates a dose-dependent increased risk for joint surgery according to # of SE copies. Although not observed for the other outcomes, caution should be exercised given their heterogeneity. Further identification of important sources of heterogeneity will require thorough examination of covariate effects.

100

Penetrances of BRCA1 mutations

Pål Møller and Ketil Heimdal
Unit of Medical Genetics, The Norwegian Radium Hospital, N-0310 Oslo, Norway. pmoller@ulrik.uio.no

Genetic drift is supposed to underly the BRCA1 founder mutations. Additional mechanisms may be instrumental: Lower penetrances in the most frequent mutations and deviating phenotypes in newborn mutation carriers indicate possible fitness variation and distorted segregation. Our previous reports of later disease and excess of carriers in founder mutation kindreds are in keeping with this. Paternal inheritance combined with phenocopies in maternal aunts, may erronously be interpreted as low penetrance. Early reports on penetrance underestimated ascertainment problems. Few have discussed the impact of informative censoring present when considering breast cancer as affected and the mutation carrying women die from ovarian cancer or another mutation-associated phenotype.

We have information on a number of extensive kindreds with the same mutations on the same haplotypes, presumably averaging out putative influence(s) of modifying genes and/or environment and therefore suitable to examine the problems mentioned above.

It will be demonstrated how calculated penetrances in the data sets vary according to assumptions underlying the calculations. Frequent mutations may be frequent because of lower penetrance, and complex patterns of different factors influencing fitness are not excluded. Ascertainment problems are significant if calculating on small sibships and have larger impact the lower the penetrance, while informative censoring remains a problem in large sibships.

101

Genetic factors may influence latent class membership for ADHD and its comorbid traits.

R.J. Neuman, A.C. Heath, K. Bucholz, P. Madden, W. Reich, L. Sun, R.D. Todd
Washington Univ., St. Louis, MO

The goal of this study was to define homogeneous phenotypes of ADHD and its comorbid traits in which there would be increased evidence for genetic influences. ADHD commonly co-occurs with disorders such as anxiety, mood, or ODD. We used latent class analysis (LCA) on a population based sample of young female twins (N=2,904), to simultaneously differentiate among phenotypes with co-occurrence of ADHD, ODD, separation anxiety, or potentially depression phenotypes. To determine whether classes could be considered familial, concordance for class membership was

computed separately for the MZ and DZ twins. LCA identified several clinically important classes with high endorsements of symptoms for just one trait and none other. For example, we observed a class consisting of subjects exhibiting ADHD inattention symptoms without ODD, anxiety or depression symptoms (10%) and another where over 70% of individuals endorsed only the depression symptoms (13%). A substantial proportion of the MZ twin pairs, 68.3%, was assigned to the same latent class versus 38.9% of the DZ twins, consistent with a genetic hypothesis for latent class membership. Of the twins assigned to the depression-only class, almost 33% of the MZ sibs and only 18.3% of the DZ sibs were concordant for class membership. The comparable figures for the inattention class was 65.7% and 28.6% for the MZ and DZs, respectively. These patterns of class membership assignments is the expected pattern for a genetically influenced and highly heritable phenotype and argues for the importance of phenotypic classification.

102

Inference in Genetic Association Studies with highly polymorphic Markers.

J.P. Lewinger, S.B. Bull
Samuel Lunenfeld Research Institute, University of Toronto, ON, Canada.

Genetic association studies with population or family control designs are routinely analyzed using the logistic regression model, based on the unconditional or the conditional likelihood. Small samples and highly polymorphic markers, such as the HLA loci, can result in sparse and unbalanced data. In this situation, the maximum likelihood estimates (MLE's) of the log odds ratios are biased away from the null. In extreme cases, for example in which a specific allele appears in cases but not in controls, the MLE's can be infinite.

Although overall score statistics can still be constructed, when there is an infinite MLE there is no meaningful point estimate available as a measure of association, and the corresponding confidence interval has only one finite bound and can be quite uninformative.

In this study we conducted a series of simulations to evaluate inference methods based on a penalized likelihood function proposed by Firth (1993). This modified likelihood removes first order bias and always yields finite estimates of the log odds ratios and corresponding standard errors.

Application of these methods to sparse data from a study of HLA and inflammatory bowel disease provided conclusions that would have otherwise been unavailable.

103

Phenotypic assortment for height reconsidered

J.D. Grant[1], A.C. Heath[1], K.K. Bucholz[1], P.A.F. Madden[1], & N.G. Martin[2]
[1]Washington University School of Medicine, St. Louis, MO, USA; [2]Queensland Institute of Medical Research, Brisbane, Australia.

Pearson and Lee (1903) reported spousal correlations for height that informed Fisher's (1918) paper on the correlation between relatives. Fisher considered models for assortment that implied either direct phenotypic assortment for a trait or phenotypic assortment for a correlated character, concepts that were revisited by Morton, Rao and others in the 1970s. The present analyses represent an examination of assortative mating for height in a sample of Australians. 4845 husband-wife pairs were identified from respondents in the Australian Twin Study (4162 twin/spouse pairs and 683 parent-of-twin pairs). Modest (but statistically significant) assortative mating was observed for height, with a husband-wife correlation of 0.23. Further analyses were conducted using data from 1388 twin pairs where both twins had spouses and all four individuals had height data (479 MZF; 243 MZM; 243 DZF; 122 DZM; and 301 DZO pairs). Under the hypothesis of direct phenotypic assortment for height, one would predict that the correlation between the twin and the cotwin's spouse would be the product of the twin correlation for height (about .9 for MZ twins and .45 for DZ twins) and the marital correlation, and that the correlation between the spouses of twins would be the product of the twin correlation for height and the square of the marital correlation. However, observed correlations between the spouses of MZ twin pairs (MZF r=0.14; MZM r=0.17) were higher than would be predicted under phenotypic assortment, while the twin-cotwin spouse correlations did not show the expected MZ-DZ difference in magnitude (e.g., MZM r=0.26, DZM r=0.25; MZF r=0.16, DZF r=0.12). The hypothesis of direct or indirect phenotypic assortment for height may need to be reconsidered.

104

Incorporation of covariates in multipoint model-free linkage analysis of binary traits using the MLB method.

Alexandre Alcaïs and Laurent Abel.
INSERM U.436, Paris, France

When the mode of inheritance is unknown, genetic linkage analysis of binary trait is commonly performed using affected-sib-pair (ASP) approaches. When there is evidence that some covariates influence the pheno-

type, incorporation of this information is expected to increase the power of the analysis since it allows 1) a better specification of the phenotype and 2) to take into account unaffected subjects. Here, we show how to account for covariates in the sibship-oriented Maximum-Likelihood-Binomial (MLB) linkage method by means of Pearson's logistic regression residuals which are computed using phenotypic and covariate information on both affected and unaffected sibs. These residuals are subsequently analysed as a quantitative phenotype with the corresponding extension of the MLB approach (Alcaïs and Abel, *Genet Epidemiol*, in press) which can be used without any assumption on the distribution of residuals. Simulation studies compared the power of this method to the standard MLB approach (using affecteds only) in the multipoint analysis of nuclear family data under different generating models according to genetic and covariate (with or without familial correlation) effects. These studies also investigate the interest of collecting families with a single affected sib. Power results will be discussed as a function of the familial ascertainment strategy.

105

A genome-wide linkage analysis of blood pressure and heart rate in the Québec Family Study using a variance components approach (SEGPATH)

T. Rice[1], M. A. Province[1], L Pérusse[2], Y Chagnon[2], T. Rankinen[2], I. B. Borecki[1], C. Bouchard[2], D. C. Rao[1]
[1]Washington University, St. Louis, MO USA; [2]Laval University, Québec City, Québec, Canada

A genome-wide scan for heart rate (HR) and blood pressure (both systolic and diastolic BP) genes was performed using a multi-point variance components model (SEGPATH) in over 200 families participating in the Québec Family Study. A total of 293 microsatellite markers were used with an average between-marker distance of 11.9 cM. Relatively few differences in results were observed depending on whether or not the traits were adjusted for body mass index. Two markers showed good evidence for linkage ($P = 0.0007$) with HR (D1S1653 at 1q22 and D19S247 at 19p13.3). There was suggestive linkage ($P < 0.01$) with one or more of the phenotypes on regions of chromosomes 1 (1p22.1–p21.2, 1p11.2, 1q22–q25.2) and 2 (2p14, 2q21.3–q22.1), and other isolated markers at 3q28, 4q35.1, 6q13, 10p14, 21q22.3, 22q13.1. These chromosomal regions contain a number of candidate genes that are related to HR and/or BP regulation. Together, these results and those from 2 recently published genome-wide studies of BP provide a focus for future studies to identify genes that influence interindividual variation in HR and BP.

106

Precision Mapping of Human QTLs by Combined Linkage/Disequilibrium Analysis

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

The variance components method is a powerful approach for detecting linkage to quantitative trait loci (QTLs). However, the resolution may be compromised and power reduced for complex traits. On the other hand, it is well known that incorporating effect of linkage-disequilibrium (LD) enhances power in standard linkage analysis because of the added information on phase, as in the TDT and its extensions. We consider an extension of the variance components method using SEGPATH (Province et al., 1998), and incorporate both the allelic association and the linkage effects in a combined, comprehensive model. The calculation is conditional on family by decomposing the mean effect of allelic association into an overall among-family effect and deviations of individuals in the family from the family mean. As noted by others, procedures using such a conditional framework avoid detecting spurious associations due to population stratification. Various hypotheses can be tested by likelihood-ratio test of appropriate submodels. For example, if the linkage effect disappears when the LD effect is incorporated in the submodel, we may conclude that the marker locus should be in close proximity to the disease gene. Power analysis using simulated data will be presented along with a careful examination of practical issues. For example, when marker loci have a large number of functional alleles, a randomization procedure is introduced on top of the above modeling to reduce the degrees of freedom and to obtain a closer-to-real p-value.

107

MILC: a new statistic to search for multifactorial diseases susceptibility genes in founder populations

Bourgain Catherine[1], Genin Emmanuelle[1], Quesneville Hadi[2] and Clerget-Darpoux Françoise[1]
[1]Inserm U155, Paris, France; [2]Laboratoire de Dynamique du Génome et Evolution, Institut J.Monod, Paris, France.

A method (Maximum Identity Length Constrast, MILC) is proposed to locate susceptibility genes for multifactorial diseases in founder populations. It uses characteristics of such populations: linkage disequilibrium spanning large regions and significant kinship coefficients. Affected individuals and their parents are typed for linked markers and two groups of parental haplotypes are considered: those transmitted to affected

individuals and those not transmitted, representative of the population. Similarly to the "Haplotype Sharing Statistic" (*HSS*), (Van der Meulen et al, 1997), the lengths of identity are compared between the two groups of haplotypes. In contrast with *HSS* that tests differences of identity at each marker, MILC is based on the maximum of this difference. For a given genetic disease model, the power of this method was evaluated through simulations as a function of population characteristics and marker map. With dense map, the power of this method is generally higher than that of the TDT because it uses information on several markers. For fixed population parameters, results are highly variable among simulations. In fact, the power of our method strongly depends on linkage disequilibrium (LD) among markers. As LD mostly results from random processes it is not predictable and therefore not a simple decreasing function of the genetic distance.

In a context of multifactorial disease, one has to be very cautious in the interpretation of LD mapping results even in founder populations. Our results suggest that one way to increase our power to detect genetic risk factor in these populations can be the use of markers showing strong LD one with the others.

108 [Presidential Address]

Genetic Epidemiology with a Capital E

D. Thomas
Univ of Southern Calif, Los Angeles, CA, USA

Three characteristics of genetic epidemiology which distinguish it from its parent disciplines are a focus on population-based research, a focus on the joint effects of genes and the environment, and the incorporation of the underlying biology of the disease into its conceptual models. These principles are illustrated by a review of the genetic epidemiology of breast and ovarian cancer. Descriptive and mechanistic models for the joint effects of genes and "environmental" risk factors such as hormones and reproductive events are compared to illustrate the need to understand the biology. The contribution of population-based research to the development of the evidence for the involvement of major genes, the discovery of BRCA1/2, and their characterization is reviewed. Interactions of major susceptibility genes, metabolic genes, hormones, and radiation will also be discussed. I conclude with some suggestions for future directions for the field, the journal, and the Society, including recent bioethics initiatives. I believe that the Society should reach out more to the epidemiology community and that the journal should shift its emphasis from pure methodology to more substantive papers that illustrate these principles.

109 [Invited Speaker]

Under what conditions are association studies more efficient than family studies for gene mapping?

E.M. Wijsman
University of Washington, Seattle, WA, USA

A major challenge is how to efficiently and successfully map and identify genes contributing to complex traits. With the advent of automated genotyping methods, discussions have turned from family-based studies to the potential of population-based designs for initial gene localization, both in the context of candidate-gene and whole-genome screens. However, before major investment is made in association studies, it is useful to determine the conditions needed to succeed with such studies, and to compare these to requirements for family-based designs. In this context, it is useful to consider sample sizes needed to achieve similar posterior probabilities of linkage for the two designs.

Power and sample size requirements for a case-control study for markers with arbitrary numbers of alleles can be obtained by considering a measure of genetic distance (G^2) between cases and controls. G^2 is a function of age of disease mutation(s), recombination fraction, number and frequencies of marker alleles, case ascertainment procedures, and fraction of case chromosomes carrying a particular disease allele. General conclusions are that multiallelic markers are usually more efficient in case-control studies than are diallelic markers, and become increasingly more so with increasingly complex traits, with increasing mutation age, and with increasing recombination fractions. With increasing etiologic heterogeneity and ascertainment that is a function of number of cases per family, sample-size requirements also increase more rapidly for case-control than for family-based studies. In general, with the exception of young, monogenetic diseases, population-based studies may be considerably less efficient than are family-based studies, especially when the posterior probability of a correct conclusion is taken into account.

110 [Invited Speaker]

Methods for detecting gene-gene interaction using affected sib pairs.

P. Holmans, Washington Univ. School of Medicine, St Louis.

It is likely that several loci contribute to the genetic susceptibility to complex diseases, and that considerable interactions between these loci exist. If two disease loci interact epistatically, a positive correlation between the number of alleles shared identical by descent (IBD) at each locus is expected, whereas heterogeneity should produce a negative correlation.

We consider three methods for detecting such interactions. Firstly, one can calculate the correlation in estimated IBD sharing for each sib pair at locations corresponding to lod-score peaks obtained under single-locus analysis, and test for a significant departure from zero. Secondly, one can select the pairs for which the estimated IBD is greater (if testing for epistasis) than a pre-determined value (e.g. 1) at the first locus. The test statistic for the second locus is calculated using only these sib pairs. The significance of the result is determined by comparing it to those obtained from randomly selected samples of equivalent size. The third method models the probability that an affected sib pair inherit an allele IBD from a given parent at the second locus as a logistic regression with the number of alleles shared IBD at the first locus included as a covariate.

Expected lod scores were calculated under a variety of two-locus disease models in order to determine which, if any, might yield a significant increase under the methods being tested. The statistical powers of the methods were compared by applying them to datasets containing sibships of varying sizes simulated under these models. The increase in precision in disease locus location given by each method was also investigated.

111 [Invited Speaker]

Gamete Competition Models

Kenneth Lange, Departments of Biomathematics and Human Genetics, UCLA School of Medicine

This talk presents joint work with John Blangero and Janet Sinsheimer on a gamete competition model. The simplest version of the model is an adaptation of the Bradley-Terry model for ranking sports teams. If one assigns to allele i of a marker locus a parameter $\tau_i > 0$, then one can express the probability that a parent with heterozygous genotype i/j transmits allele i as $\Pr(i/j \to i) = \tau i / (\tau i + \tau j)$. Mendelian segregation corresponds to the choice $\tau i = 1$ for all i. To test whether Mendeliain segregation is true, one can estimate the τi from pedigree data and perform a likelihood ratio test under the constraint that one τi equals 1. Although this procedure generates an interesting method for performing segregation analysis with a marker locus, the real promise of the gamete competition model lies in generalizing the transmission / disequilibrium test. For a candidate disease gene, one invokes Mendelian transmission of alleles to normal children and gamete competition transmission of alleles to affected children. The resulting method has the virtues of using full pedigree data and giving an estimate of the strength of transmission distortion to affected children for each allele. The impact of covariates can be incorporated by rewriting $\tau i = \exp(\beta^t x_k)$, where β is a parameter vector and x_k is a covariate vector for the kth transmitted gamete. Typical covariates include disease severity indicators for the child or repeat number for tandem repeat alleles. These ideas will be illustrated by applications to specific genetic diseases.

112 [Invited Speaker]

Whole-genome association studies: which strategy to develop ?

Laurence Tiret, INSERM U525, Paris, France

Whole-genome association studies have recently been proposed as an efficient approach for the identification of susceptibility genes underlying common diseases. There are two different strategies: direct or indirect. The direct strategy is to characterize all common variants of human genes and to test directly their association with disease. The indirect strategy is develop very dense maps of neutral single-nucleotide polymorphisms (SNPs) and to detect susceptibility genes through linkage disequilibrium with unidentified functional variants. The choice of a strategy is dependent on a number of critical elements, including the number of functional polymorphisms within a gene, the combination of their effects, their allele frequencies and the extent of linkage disequilibrium among them.

We performed an extensive molecular screening of the coding and flanking regions of 36 genes for cardiovascular genes. All polymorphisms identified by this screening were further genotyped in 750 subjects of European descent. This study provides new insights into the type and amount of DNA sequence variation that might be expected in human genes, as well as on the extent of linkage disequilibrium within candidate regions.