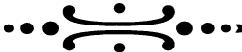


ABSTRACTS FROM THE
SIXTH ANNUAL MEETING
OF THE
INTERNATIONAL GENETIC
EPIDEMIOLOGY SOCIETY



BALTIMORE, MARYLAND

October 27-28, 1997

HOSTED BY:

National Human Genome Research Institute, NIH
Baltimore, MD, USA

University of Maryland, Baltimore
School of Graduate Studies
Baltimore, MD, USA

Johns Hopkins University
Department of Epidemiology
School of Hygiene and Public Health
Baltimore, MD, USA

1

Interest of the binomial maximum likelihood approach in sib-pair linkage methods for analyzing multiplex sibships. Abel L, Alcais A, Mallet A. INSERM U. 436, Paris, France.

Family samples collected for sib-pair linkage studies usually include some sibships with more than two affecteds (multiplex sibships). Several methods have been proposed to take into account these multiplex sibships in the analysis of the sample. Two methods, which are the most widely used, are based on the number of alleles shared by the sib-pairs constitutive of the multiplex sibship, with the first using the total number of these shared alleles ("all possible pairs" method) and the second considering a weighted number of these alleles (Suarez & Van Eerdewegh, Am J Med Genet 1984, 18:135-146.). However, the distributions of the resulting statistics under the null hypothesis of no linkage are not clear, with in particular large inflations of the type I error observed with the all possible pairs method. This problem can be overcome with the use of a likelihood method based on a binomial distribution of parental alleles among affected offspring (Majumder & Pal, Genet Epidemiol 1987, 4:277-287). This approach accounts in a natural way for multiplex sibships and provides a simple likelihood-ratio test (LRT) for linkage involving a single parameter. Simulation studies showed that this likelihood method provides very consistent results in terms of type I errors, whatever the sample

size, and yields power levels similar to those of the other methods. Furthermore, the LRT is theoretically expected to be more powerful than the classical mean test in the analysis of sibships with two affecteds. This binomial likelihood approach appears to be a quite interesting alternative to analyze sib-pair studies.

2

Conditional oligogenic genomic screening. Almasy L, Blangero J. Southwest Foundation for Biomedical Research, San Antonio, TX, USA.

Conditional oligogenic linkage screening, in which a genomic scan for a second quantitative trait locus (QTL) is performed conditional on the location of a known QTL, has the potential to increase power to detect QTLs of moderate effect. Conditioning on the known QTL reduces the residual phenotypic variance, magnifying the effects of any remaining loci. However, significance standards for conditional genomic screening have yet to be developed and the false positive rate is unknown.

We have explored these questions using simulated data from the recent Genetic Analysis Workshop 10. A simulated quantitative trait was influenced by three QTLs, the largest of which accounted for 28% of the phenotypic variance in the trait. An initial multipoint genomic screen was performed for 200 replicates of the data set using

SOLAR, a variance component linkage method for general pedigrees. A second genomic scan was then performed conditional on the location of the strongest QTL, which was detected in the initial screen with a LOD of 3.0 or better in 87% of replicates. In the second phase of the screen, although the location of the first QTL is fixed, its contribution to the genetic variance is re-estimated. A marginal LOD score for a second locus is then computed by comparing the likelihood of a model in which the genetic variance for this QTL is estimated to a model in which it is constrained to zero.

In the initial screen, no false positive LOD scores over 3.0 were observed. However, there were a total of 9 LODs over 2.5 and 16 over 2.0, for false positive rates of 0.045 and 0.08 per replicate. The false positive rate in the conditional screen was nearly identical to that seen in the initial phase. This suggests that the significance standards currently used to interpret genomic scans may also apply to conditional oligogenic genomic screening.

The development of the statistical methods used in this work was supported by NIH grants DK44297, HL45522, HL28972, GM18897, and GM31575.

3

Comparison of variance components and extremely discordant sib pair (EDSP) linkage methods for quantitative traits. Amos CI, de Andrade M. Department of Epidemiology, U.T. M.D. Anderson Cancer Center, Houston, TX.

Variance components methods provide an efficient approach for linkage identification with minimal dependence upon modeling assumptions. However, for random samples, the power to detect a locus having low heritability can be low. Recently, EDSP approaches have been shown to provide a more powerful linkage strategy for traits having low major gene heritability than methods based upon random samples. In the current study we explore properties of variance components methods for ascertained samples (VCAS) and we compare the power of EDSP and VCAS methods. Ascertainment corrections that we considered for the VCAS method include conditioning on the phenotype of the selected sample as well as conditioning on the threshold that led to ascertainment. Simulations show very little difference in terms of power and bias for these two methods of ascertainment. Both methods yield asymptotically unbiased estimates of the environmental and polygenic components of variance. However, the major gene component of variance is biased for both correction procedures. Our current studies include study of sibships of size four for which one sib is above a threshold and another is below a threshold. Comparing VCAS with conditioning on phenotypes to EDSP, we note chi squared values that are 1.35 times larger for the VCAS method compared with the EDSP method. These results show that VCAS can be more powerful than EDSP but that selected samples result in biased parameter estimates.

4

Behavior of the Maximum Likelihood Score when many affected sibpairs are issued from a few untyped parents

Babron MC¹, Truy F¹, Eichenbaum-Voline S², Clerget Darpoux F¹.

¹INSERM U155; ²CNRS URA1323, Paris, France

In the study of multifactorial diseases, model-free analyses such as affected sibpair methods are widely used to detect linkage. The Maximum Likelihood Score (MLS), proposed by Risch (1990) allows estimation of the Identity By Descent (IBD) distribution among sibs given observations on a set of markers. Depending on the available information, the inference on IBD is more or less ambiguous. It is most ambiguous when parental genotypes are missing. This leads in general to multimodal MLS curves.

This ambiguity in IBD becomes crucial when most sibpairs are issued from a few untyped parents. We investigate the behavior of the MLS in such a situation by simulating the segregation of one risk factor linked to a set of markers. The MLS may be artificially increased for one or several markers, leading to even more irregular MLS curves. The maximum of the MLS curve may then be far from its true location.

Furthermore, we show that treating multiple affected sibs in a family as independent sibpairs leads to an even more drastic effect.

This will be illustrated on recently published data on celiac disease. It is well known that there is a risk factor within HLA whereas the MLS was highest at 30cM from HLA and interpreted as evidence for a second locus.

5

Segregation analysis of Attention Deficit Hyperactivity Disorder. Bailey JN, Cantwell D, Asarnow J, Smalley SL. UCLA Neuropsychiatric Institute, LA, CA.

We conducted complex segregation analysis utilizing nuclear family data in a sample of 114 families with a single DSM-IIIR diagnosed ADHD proband. The sample was selected to match the Faraone et al., 1992 sample, and thus the probands were Caucasian, ages 6-18, had IQ>80, and met DSM-IIIR criteria for ADHD. The ratio of male:female probands in the sample was 3.2:1, in contrast to the Faraone analyses which were based on male probands. The frequency of ADHD among first degree relatives was 18% (fathers-21%, mothers -14%, brothers-28%, sisters-13%). Segregation analysis was performed using the RegD program in S.A.G.E.. We assumed a regressive class A model with two affection classes and Hardy-Weinberg equilibrium. Effects of sex were estimated in all models, but age was not included as ADHD is an early-onset disorder. Five modes of transmission were considered: a general model that had non-fixed transmission probabilities, an environmental model with no major gene effect, and three Mendelian models (dominant, recessive, and additive). Model comparisons were made using the likelihood ratio test for models which were nested. The environmental model

was rejected when compared to the general model. None of the Mendelian models were rejected and the dominant model gave a better fit according to Aikaike's criteria. These preliminary findings support those of Faraone et al. (1992) suggesting that an autosomal gene of major effect may contribute to ADHD.

Faraone SV, Biederman J, Chen WJ, Krischer B, Keenan K, Moore C, Sprich S, Tsuang MT (1992): Segregation analysis of attention deficit hyperactivity disorder: Evidence for single gene transmission. *Psych Genet* 2:257-275

6

Segregation analysis of pancreatic cancer. Banks MG, Aston CE, McNamara PJ, Crowley KE, Mulvihill JJ, the Family Study of Pancreatic Cancer group. University of Pittsburgh, Pittsburgh, PA, USA.

Pancreatic cancer (PaCa) is the fifth leading cause of death from malignant disease in Western society. As a public health issue, the devastating effect of PaCa due to its high frequency is aggravated by its extremely poor prognosis (20% one-year survival) and lack of early detection. Determining a means of early detection or better risk assessment is of concern to the relatives of PaCa cases. We applied segregation analysis to data from the Family Study of Pancreatic Cancer at the University of Pittsburgh to examine the evidence for a single major Mendelian gene controlling susceptibility to PaCa. Families were ascertained for having at least two 1⁺ relatives with PaCa or two 2⁺ relatives with PaCa connected by an individual with any other cancer. The families were contacted, questionnaires were administered, and medical records and pathology samples were collected. Seventy families were included in the segregation analysis based on completeness of data. The computer program REGTN of the S.A.G.E. package was used for segregation analysis under a Class 1 model, which specifies distributions of age-of-onset based on genotype. The most likely model for PaCa susceptibility in these families was that of a diallelic major gene with a dominant disease allele controlling the age-of-onset for PaCa. The estimated mean age-of-onset of PaCa in the "high-risk" genotypes, which comprised 11% of our population, was 60.8 years. The estimated age-of-onset for the "low-risk" genotype was 74.6 years, indicating that even those individuals with a "low-risk" genotype are susceptible to PaCa at a reasonable age-of-onset.

7

Genetic analysis of bivariate dichotomous twin data using logistic regression models. Bhadra P¹, Ramakrishnan V^{1,2}, Goldberg J^{1,2}. ¹ University of Illinois-Chicago, USA; ²Vietnam Era Twin Registry, Hines, IL, USA.

This paper presents a new statistical method for analyzing bivariate twin data to assess whether two dichotomous traits, say diabetes and hypertension, are under common genetic influence. Current statistical methods for analyzing bivariate dichotomous twin data use structural equation modeling and assume a unobservable joint liability that is normally distributed. Bivariate logistic regression methods can be used to test for the presence of common genetic influence on two traits that does not require the assumption of a bivariate normal liability distribution. However, the standard bivariate logistic regression methods proposed by Liang, Zeger and Qaqiqah (1992) leads to an overparameterization when using twin pairs. We adapt GEE methods for bivariate twin analyses where the correlation between phenotypes is taken into account. The proposed methods are illustrated using psychiatric data from the Vietnam Era Twin Registry.

8

Evidence for two-locus inheritance of human fatness.

Borecki IB¹, Blangero J², Rice T¹, Bouchard C³, Rao DC¹. ¹Washington University School of Medicine, St. Louis, MO; ²Southwest Foundation for Biomedical Research, San Antonio, TX; and ³Laval University, Ste-Foy, Québec, Canada.

The genetics of human fatness has been the subject of many recent studies, motivated by the increased morbidity and mortality associated with obesity, as well as the increasing prevalence of overweight persons in the USA. The body mass index (BMI) and fat mass (FM), measured by underwater weighing, were assessed in 1630 individuals in approximately 300 families in the Quebec Family Study. The two phenotypes are highly correlated in adults (0.8), and previous segregation analysis revealed evidence for a recessive major gene for each trait. In the present investigation, we utilize bivariate segregation analysis to determine the source(s) of phenotypic correlation: pleiotropic major gene, shared familial factors/polygenes, or shared non-transmitted environment. Analysis was carried out using PAP (Hasstedt 1989) with extensions to the bivariate case provided by Blangero and Konigsberg (1991). Evidence for two pleiotropic recessive loci, together, accounting for 64% and 47% of the variance in BMI and FM, respectively, was obtained. The high degree of covariance between traits due to major locus pleiotropy

is not surprising since they ostensibly measure the same thing, additional covariance could be accounted for by shared polygenes and by shared environment. This study demonstrates that additional power derived from the simultaneous analysis of correlated phenotypes enables the detection of a second major locus that apparently does not affect extreme overweight (as does the primary major locus), but affects variation in the "normal" range.

9

Quantitative symptom measures linked to chromosome 6p in familial schizophrenia.

Brzustowicz LM^{1,2}, Honer WG³, Chow EWC⁴, McAllduff J¹, Hodgkinson K⁴, Bassett AS^{4,5}. Rutgers University, Newark, NJ, USA; ²New Jersey Medical School, Newark, USA; ³University of British Columbia, Vancouver, Canada; ⁴Queen Street Mental Centre, Toronto, Ontario, Canada; ⁵University of Toronto, Ontario, Canada.

A number of recent linkage studies have suggested the presence of a schizophrenia susceptibility locus on chromosome 6p. Evaluating 28 genetic markers spanning chromosome 6 in 10 moderately large Canadian families of Celtic ancestry, we have recently demonstrated significant linkage of positive (psychotic), but not negative (deficit) or general psychopathology symptom scores of the PANSS to chromosome 6p. A multipoint sibpair analysis using symptom scores as quantitative traits produced a nominal p-value of 5.4×10^{-6} at the marker D6S1960. Using simulations studies, we determined this nominal p-value to correspond to empirical p-value of 0.008. This result suggested that the schizophrenia susceptibility locus on chromosome 6p may be related to the severity of psychotic symptoms. We have now further defined the specific symptoms associated with this linkage finding. Analysis of the 14 individual component symptoms of the positive and negative symptom scales suggests that Hostility and Blunted Affect are the most significant individual components. The five-factor categorization of PANSS symptoms (Negative, Positive, Cognitive, Excitement, Depression) was also analyzed, with significant results only for the Excitement factor (formal p-value of 9.1×10^{-8} at D6S1960). Assessment of quantitative behavioral traits, either individually or as composite scores, may provide increased power over categorical phenotypic assignments to detect linkage in complex psychiatric disorders, and may provide further insights into the components of illness most under genetic control.

10

Using Gibbs sampling to fit generalized linear mixed models (GLMMs) to address the correlation of non-Normal phenotypes within nuclear families

Burton P, Palmer L, Gurrin L. TVW Telethon Institute for Child Health Research and University of Western Australia. PO Box 855, West Perth, Western Australia.

Phenotypic assessment can be the most difficult and costly component of research into the genetic epidemiology of common complex diseases. Research is often undertaken by groups with great experience working with the phenotype of interest but little knowledge of genetic epidemiology. Pre-existing cohort studies are often extended to include nuclear families and 'genetic analyses' are undertaken. Unfortunately, residual within-family correlations are likely to be strong and this may cause significant analytical problems for the inexperienced. These problems are compounded by the lack of a unified approach to analysis which easily generalizes to all of the common classes of phenotype: continuous (Normal), binary, censored age at onset etc. Using the GLMM we have developed an approach which models an A.C.C's covariance structure by adding a series of random effects to the linear predictor; this allows most common types of phenotype to be modelled very flexibly. Models are fitted in BUGS, a general purpose Gibbs sampler. By avoiding high dimensional integrals this approach greatly simplifies the extension to non-Normal phenotypes. Using simulated data we will demonstrate that, for a Normally distributed phenotype, the BUGS-based GLMMs generate parameter estimates (and SEs) for fixed effects and random effect variances which are very similar to those of conventional variance component models. We will show that the model extends very naturally to a binary response problem (GLMM: binomial error + logit link) and will report the results of extensive simulation exercises which demonstrate the consistency of parameter estimates in this setting. We will discuss extensions of the model to other types of phenotype and to twin family designs.

11

Localization of a putative gene predisposing to Systemic Lupus Erythematosus (SLE) to a 5 cM region of chromosome 1.

Cantor RM¹, Tsao BP², Badsha H², Grossman J²,

Kalunian KC², Kartung K³, Wallace DJ¹, Hahn BH², Rotter JI¹. Cedars-Sinai Research Institute¹; UCLA School of Medicine², Los Angeles, CA;

Institut für Laboratoriums und Transfusionsmedizin, Bremerhaven, Germany³.

Previously, we identified tentative evidence for linkage of SLE with 5 markers within a 15 cM region of the long arm of human chromosome 1, suggested for analysis by synteny with a murine chromosomal region containing mouse lupus genes (JCI 99:725-731, 1997). The prior sample consisted of 52 affected sibling pairs from Asian, African American, and Caucasian ethnic backgrounds. Seeking confirmation of this finding, we have expanded the sample by 50% to 77 affected sibling pairs, and typed all individuals at 14 markers within a 30 cM region encompassing the 15 cM linked region. Multipoint analysis of this sample with the MAPMAKER/SIBS program yielded a peak LOD score of 3.3, and narrowed the region of interest to 5 cM. Haplotype analysis of the 77 affected sib pairs with 5 markers within the refined 5 cM region resulted in the same proportion of haplotype sharing evidenced in the initial sample (0.62, p=.001), thus

providing additional evidence for linkage to this region. Remarkably, SLE appears to exhibit this linkage across multiple ethnic groups. The estimated λ_s for this locus is 2.14, suggesting that it may account for approximately 20% of the genetic susceptibility to SLE. Currently, this locus shows the only evidence for linkage to SLE of any chromosomal region in man.

12

The study of genes whose phenotype is altered by environmental exposure.

Carpenter CL and Mack TM. University of Southern California School of Medicine, Los Angeles, USA.

Studying genes whose expression depends upon environmental stimuli requires careful consideration. The phenotype of metabolic genes is sometimes permanently altered or induced by environmental exposure.

Differential nicotine metabolism is thought to be one of many factors that contribute to nicotine dependence. While suggestive evidence implicates the role of genetic susceptibility in accounting for differences in nicotine metabolism, no studies have yet been conclusive.

Nicotine metabolism among smokers is different than nicotine metabolism among never-smokers due to the induction of metabolic enzymes by smoking. Because nicotine metabolism differs according to smoking status, the proposed design will compare never-smoking members of monozygotic twin-pairs whose co-twins are smokers with never-smoking members of monozygotic twin-pairs whose co-twins are never-smokers. The never smoking twins carry the same genotype as their co-twin, and therefore comparisons will be between genotypes that vary according to known susceptibility to nicotine dependence, yet are unconfounded by smoking status. Nicotine metabolism will be measured by level of cotinine in urine. Thus we can examine potentially varying levels of nicotine metabolism that correspond to phenotypic expression of underlying genes, without induction by smoking.

13

A Genetic Analysis of Smoking Behavior in 1480 Families.

Cheng LS¹, Swan G¹, Carmelli D¹. SRI International, Menlo Park, CA, USA.

The purpose of this investigation was a genetic study of smoking behavior in 1480 nuclear families. Data was collected by a detailed interview that focused on

smoking and drinking history in members of the Western Collaborative Group Study, a population based cardiovascular study currently in its 36th year of follow-up. Smoking and drinking data were available on all first degree relatives of probands. To evaluate the familial clustering of smoking, we used complex segregation analysis and maximum likelihood statistics to test for the model that best fits the data. Results revealed a lack of fit of the environmental, sporadic, and recessive models, $\chi_{(2)}^2=98.13$, $P<0.0001$, $\chi_{(6)}^2=102.72$, $P<0.0001$, $\chi_1^2=9.72$, $P<0.005$, respectively. Mendelian transmission could not be rejected, $\chi_3^2=7.47$, $P>0.05$. In addition, the effects of spouse and both parents were significant, $\chi_2^2=10.91$, $P<0.005$. The best fitting model for the familial aggregation of smoking was that of a dominant major gene effect with residual familial correlations. Under this model, the penetrance of carriers genotypes was estimated as 65.4%. Forty nine members, distributed in 11 pedigrees, were identified as having a probability of 70% or greater to be gene carriers. Carriers of the susceptibility gene were also 3 times more likely to be alcohol drinkers, 11.04% vs 3.78%, $\chi_1^2=87.7$, $P<0.001$. These results are the first to provide evidence of a major gene for smoking behavior and should aid in future linkage studies.

14

Identity by descent homogeneity test as a simultaneous test of linkage and allelic association.

Clerget-Darpoux F, Babron MC. INSERM U. 155, Château de Longchamp, Bois de Boulogne 75016 Paris, France

The TDT allows testing the two hypotheses H01 (no linkage between marker and susceptibility locus) and/or H02 (no allelic association between marker and susceptibility allele). Similarly, it is also possible to test H01 and/or H02 by testing the homogeneity of the IBD distribution (X_{kl}, Y_{kl}, Z_{kl}) in sib pairs conditioned on the marker genotype (kl) of one of the sibs (the index case). Rejecting the homogeneity of the (X_{kl}, Y_{kl}, Z_{kl}) implies that both H01 and H02 are rejected. We will refer to this test as the identity-by-descent homogeneity test (IBD-HT). Statistical properties of the test will be presented. Note that, under a recessive disease model, homogeneity of the (X_{kl}, Y_{kl}, Z_{kl}) is also expected. Consequently, linkage and gametic disequilibrium cannot be detected in this situation.

This approach will be illustrated on IDDM where the presence of a risk factor in the insulin region can be demonstrated through the heterogeneity of the IBD distributions.

Although the overall IBD distribution in the insulin region does not deviate from (0.25, 0.50, 0.25) in a sample of 95 affected sib pairs, there is significant heterogeneity in the IBD distributions according to the index case's genotype for the 5'FP polymorphism

flanking the insulin gene. When the index cases are homozygous for the class 1 allele, the IBD distribution is (0.29, 0.55, 0.16) whereas for other index cases, the IBD is (0.10, 0.47, 0.43).

This approach as the TDT emphasizes the interest of using both association and linkage information to evidence risk factors in multifactorial diseases.

15

The genetic inheritance of body mass index and waist-to-hip ratio in African-American families. Colilla S¹, Goldberg J¹, Ramakrishnan VR¹, Rotimi C², Cooper R², and Cox N³. ¹University of Illinois at Chicago, USA; ²Loyola University, Maywood, IL, USA; ³University of Chicago, IL, USA.

Most studies of the genetic transmission of obesity have supported the presence of a recessive major gene, but have sampled predominantly white populations. To determine if such observations were consistent in blacks, we used segregation analysis to study the genetic transmission of two obesity measures, BMI and WHR, in a US black population. 315 individuals from 95 black families in Maywood, IL were identified from probands selected from participants in the International Collaborative Study on Hypertension in Blacks (ICSHIB). Class D regressive models (S.A.G.E.) were used to model the transmission of BMI and WHR, which were adjusted for significant predictors. An ascertainment correction was also used due to potential non-random ascertainment of families.

Results of segregation analysis indicate that inheritance of BMI is consistent with the transmission of a major gene. Compared to a general model with unrestricted parameters, both models testing for environmental transmission ($\chi^2=12.93$, 4df) and for polygenic inheritance ($\chi^2=35.66$, 7df) were rejected, providing evidence for the presence of a major effect. A model for a major locus with Mendelian transmission did not differ significantly from the general model ($\chi^2=3.11$, 2df), suggesting that the segregation of BMI in families is consistent with a major gene. An additive model fit the data better than either a dominant or recessive model. WHR, on the other hand, did not appear to be familial.

Our results suggest that the transmission of BMI is consistent with a major gene, but that variation in WHR does not show familial aggregation. The finding that the transmission of BMI in blacks is best explained by an additive rather than a recessive major locus differentiates our results from studies on white populations.

16

Evidence for linkage with markers on chromosome 1p in families with late onset Alzheimer Disease - The NIMH Genetics Initiative AD Study Group. Collins JS¹, Perry RT¹, Watson B¹, Blacker D³, Meyers DA²,

Albert MS³, Tanzi R³, Bassett SS², Rodes L², and Go RCP¹. ¹University of Alabama at Birmingham, USA; ²Johs Hopkins University, Baltimore, MD, USA; ³Massachusetts General Hospital, Boston, USA.

Information was gathered by the NIMH Genetics Initiative AD Study Group as part of a genome wide scan for late onset Alzheimer Disease (AD) genes. In 267 families with AD affected siblings (with no known AD mutations), 29 microsatellite markers spaced at approximately 10 cM intervals on chromosome 1 were genotyped. 145 families had at least two siblings with an age of onset over seventy years. Non-parametric analyses were performed by the programs GENEHUNTER and SIBPAL. Markers considered significant had NPI scores above 1 (GENEHUNTER), and mean IBDs (identical by descent) above 0.54 and p-values below 0.05 (SIBPAL), reported in that order. Significant markers in the full dataset were D1S1646 (1.05, 0.55, 0.01), D1S1597 (1.22, 0.55, 0.005), and D1S1368 (1.28, 0.55, 0.04). In families with an age of onset over seventy D1S1646 (1.21, 0.55, 0.05), D1S1597 (1.56, 0.56, 0.02), D1S552 (1.41, 0.56, 0.01), and D1S1595 (1.27, 0.56, 0.03) were significant. These markers on chromosome 1p are arranged in this order from the p terminus D1S1646<->14cM-->D1S1597<->9cM-->D1S1368 <->10cM-->D1S552, while D1S1595 is more centromeric. These analyses lead us to conclude that the short arm of chromosome 1 may be implicated in late onset AD.

17

Genes shared between leptin and cortisol are not the same as those between leptin and insulin. Comuzzie AG, Mahaney MC, MacCluer JW, Blangero J. Southwest Foundation for Biomedical Research, San Antonio, TX, USA.

In vivo studies suggest that levels of leptin (a protein hypothesized to regulate body fat) are mediated by cortisol, and that this effect may be potentiated by insulin. It has also been proposed that leptin may influence insulin-induced activities.

Using variance component linkage analysis we have recently identified a quantitative trait locus (QTL) at 2p21 linked with serum leptin levels (lod = 4.95). This region of chromosome 2 contains the gene *POMC*, which codes for the precursor for ACTH which in turn regulates the expression of glucocorticoids. Therefore, we have asked to what extent are the levels of these three proteins (i.e., leptin, cortisol, and insulin) influenced by shared genes? To this end, we used data from 769 persons, distributed in 42 families, participating in the San Antonio Family Heart Study to partition the covariance among levels of these three proteins into that attributable to common genetic effects as well as that due to shared environmental factors.

We found significant ($p < 0.05$) genetic correlations between leptin and cortisol, and leptin and insulin ($\rho_G = -0.55 \pm 0.19$, and 0.45 ± 0.12 respectively), but no significant genetic correlation between cortisol and insulin. Therefore, we conclude that the pleiotropy between leptin levels and each of the other two proteins results from unique sets of genes not shared between cortisol and insulin. This work was funded by NIH grants GM15803, HL45522, and DK44297.

18

Haplotypes of the interleukin-1 gene cluster in diabetic nephropathy. Cox A¹, Camp NJ¹, Gonzalez AM¹, El Nahass AM², Shaw J³, Boulton AJM³, Ward JD⁴, Duff GW¹. ¹University of Sheffield, Sheffield, UK; ²Northern General Hospital, Sheffield, UK; ³Manchester Royal Infirmary, Manchester, UK; ⁴Royal Hallamshire Hospital, Sheffield, UK.

The genes for interleukin-1 alpha (IL1A), interleukin-1 beta (IL1B), and interleukin-1 receptor antagonist (IL1RN) are clustered within a 430kb region on chromosome 2q13. The IL-1 cytokines play a central role in the healthy immune response and in the pathogenesis of autoimmune and inflammatory diseases.

We have previously characterised 2 haplotypes of 8 markers of the IL-1 cluster in the healthy caucasian population and we now present data on these haplotypes in patients with non-insulin dependent diabetes and diabetic nephropathy. The carriage rate of haplotype 1 was elevated in patients with nephropathy ($n=79$) compared to NIDDM patients without nephropathy ($n=141$) and healthy controls ($n=198$) ($p=0.03$ and 0.01 respectively). In addition, individual alleles of IL1A and IL1B were elevated in diabetic patients with nephropathy compared to those without, with odds ratios of 3.1 (95% CI 1.7-5.6, $p=0.0002$) and 2.0 (95% CI 1.1-3.6, $p=0.02$) respectively, whilst an IL1RN allele was elevated but did not reach a significance level of $p=0.05$. Composite genotype analysis indicated that the IL1A marker was independently associated with nephropathy, and that there was a possible interaction between IL1B and IL1RN. These data indicate a strong association between the IL-1 gene cluster and nephropathy in NIDDM patients.

19

Clinical usefulness of genotype for predicting long-term RA outcomes. Criswell LA¹, Mu II², Such CL¹, King M-C². ¹UCSF, USA; ²Univ. of Washington, USA.

Background: Rapid advances in molecular genetics raise the possibility that physicians managing RA patients may have the ability to readily obtain detailed genetic information about their patients. Our goal was to determine whether knowledge of DRB1 genotype (shared epitope, SE) aids prediction of long-term RA outcomes for patients whose physicians already have extensive sociodemographic and clinical information.

Methods: Up to 14 years of data were available for 180 Caucasian females. Outcomes included physician's assessment of RA course (0-2), total joint replacement (TJR) surgery, hospitalization for RA, and response to drug treatment. Patients were dichotomized for each drug received in terms of benefit, side effects, and discontinuation due to toxicity on the basis of patient and physician reports. Covariates were age at RA onset, disease duration, family history of RA, education, income, baseline values of function (0-3), painful and swollen joint groups (0-12 and 0-10, respectively), and pain rating (0-100), and rheumatoid factor positivity. Two genotype models were assumed: 1) presence vs. absence of the SE; and 2) # of SE copies. Multivariate logistic regression was used to estimate the clinical usefulness of genotype, given the other covariate data.

Results: Genotypic information contributed clinically and statistically significantly to prediction of RA course, TJR surgery, and RA hospitalization (but not treatment response), even after considering all the covariate data.

Conclusion: Molecular genetic information is useful for predicting clinically important long-term outcomes even for patients with well-established disease for whom a wealth of sociodemographic, clinical, family history, and laboratory data are available.

20

Commingling analysis of phonology measures in pedigrees ascertained through children with severe phonology disorders. Dawson DV¹, Lewis BA¹, Pollak LB Jr¹, Freebairn L¹, O'Donnell B¹, Shriberg LD². ¹Case Western Reserve University, Cleveland, OH, USA; ²University of Wisconsin, Madison, WI, USA.

Commingling analyses were performed on speech and language data collected on 119 individuals, ages 4 through 12 years, from 57 families ascertained through a preschool child with a phonology disorder. Seventy-seven children were classified as affected for a speech/language disorder based on parental report. Significant differences were observed between affected and unaffected children on all measures: phonology, language, receptive language subscores, and speech analysis. Following adjustment for age and/or socioeconomic status, a mixture of two normal distributions was found to fit significantly better than a single normal distribution for the speech analysis

rating; suggestive results were obtained for the total language score, but not the other outcomes. These results suggest that expressive language deficits including phonology may differ from receptive language deficits in their underlying etiology. Further, they suggest that expressive language deficits noted in these families represent a true disorder and are not merely representative of the lower end of the normal continuum of verbal skills.

21

Quantitative trivariate genetic analysis of low-density lipoprotein (LDL) size, triglyceride (TG) and high-density lipoprotein cholesterol (HDL). Edwards KL¹, Mahaney MC², Brunzell JD¹, Motulsky AG¹, Austin MA¹. University of Washington, Seattle, USA; Southwest Foundation for Biomedical Research, San Antonio, TX, USA.

The interrelationships between LDL size, TG and HDL may involve underlying genetic susceptibility. This study evaluated common genetic influences on LDL size, determined by gradient gel electrophoresis, and the natural log of both TG and HDL using data from 86 hyperlipidemic families, including almost 800 individuals, participating in the Genetic Epidemiology of Hypertriglyceridemia Study. Quantitative genetic analysis was used to evaluate genetic (ρ_G) and environmental (ρ_E) correlations between each pair of traits under a polygenic model, adjusting for age and gender.

Each individual trait was found to have a modest and significant ($p < 0.001$) heritability, ranging from 0.33 for TG and LDL size to 0.51 for HDL. The estimated ρ_G between each pair of traits reflects shared genetic influences (pleiotropy). The strong ρ_G between LDL size and TG indicates that the majority of genes influencing LDL size also influence TG. Similarly, the ρ_G between LDL size and HDL and HDL and TG, also suggests shared genetic influences (*all ρ_E and ρ_G $p < 0.001$).

Phenotype Pair	$\rho_G \pm SE^*$	$\rho_E \pm SE^*$
LDL size and TG	-0.88 \pm 0.06	-0.54 \pm 0.05
LDL size and HDL	0.69 \pm 0.09	0.46 \pm 0.06
TG and HDL	-0.56 \pm 0.10	-0.51 \pm 0.07

These results suggest both shared genetic and environmental influences on pairs of lipid and lipoprotein risk factors in these hyperlipidemic families, and may be useful for identifying genes involved in susceptibility to atherosclerosis and diabetes.

22

The Maximum Likelihood Score and the Exclusion Mapping Test.

Eichenbaum-Voline S¹, Babron MC², Prum B¹, Clerget Darpoux F².

¹CNRS URA 1323, ²INSERM U155, Paris, France

The Maximum Likelihood Score Method (MLS) (Risch, 1990) enables the hypothesis of linkage between a disease locus and marker loci to be tested in a sample of affected sib pairs. Let λ be the relative risk conferred by the susceptibility allele(s) of the genetic risk factor. Given the observation, it is also possible to exclude those areas of the genome which are unlikely to contain a genetic risk factor corresponding to a given λ^* by comparing the likelihood of $\lambda = 1$ (no linkage) to the likelihood of λ^* . The computation of the likelihood ratio requires the specification of marker allele frequencies and genetic distance between the markers. The robustness of the exclusion mapping test to misspecification of these values and the reliability of the exclusion criteria are not known. In particular, the assumption of an additive model, which is often done in the analysis, tends to increase the type one error of the exclusion test. Its impact is evaluated. The required sample size to exclude a locus, for a set of different λ values, is analytically computed in case of fully informative sib pairs and estimated for less informative situations.

When presence of a genetic factor with a relative risk greater or equal to λ^* has been excluded for a given type one error, the existence of a risk locus in the region is still possible but its associated relative risk is smaller than λ^* . Then, it is useful to evaluate the minimal sample size which would have been necessary to detect the presence of such a factor with the MLS.

23

Regression analysis of survival data from families with a non-susceptible fraction.

Gagnon DR, Glickman M, Cupples LA. Boston University, MA, USA.

Survival data analysis is frequently applied to family data to estimate lifetime risk and mean onset age of a disease when only a portion of individuals are susceptible. When covariates may affect these parameters, regression models are employed. In this paper, we examine the performance of a model which analyzes correlated survival data where: 1. there is a fraction of the population that is non-susceptible to disease; 2. there is late onset, leading to extensive censoring; 3. there are two censoring mechanisms allowed, i.e., random and disease censoring; and 4. there may be missing information on onset, random or disease censoring ages. This parametric model contains a logistic regression component to predict susceptibility to disease, a Weibull model with a frailty component to predict onset age, and independent Weibull models for the random and disease censoring age distributions. A Gibbs sampling approach is used. This model allows one to examine the association of risk factors for susceptibility to disease and for onset age, as well as estimating the fraction of the population at risk for disease. The performance of the model will be evaluated with

simulated data and an example will be presented for coronary heart disease from the Framingham Heart Study.

24

Arguments for individual susceptibility involved in the outcome of *Trypanosoma brucei gambiense* (Tbg) infection. Garcia A¹, Laveissière C². ORSTOM, IPR / OCCGE, Bouaké, Ivory Coast.

Human African Trypanosomiasis (HAT) or sleeping sickness, transmitted by the bite of an insect vector (*glossina*), remains an important public health problem in Africa. Serological tests such as Card Agglutination Test for Trypanosomiasis (CATT) are useful for mass screening but parasitological techniques are required to confirm diagnosis by detecting trypanosoma in blood or lymph nodes during the first period of the disease (T+). Whatever was the parasitologic technique used, the existence of seropositive but apparently unconfirmed subjects (no parasite in blood or lymph nodes) (CATT+/T-) arise the question of individual susceptibility to infection. To explore the influence of environmental, behavioural and individual risk factors in the outcome of Tbg infection we compared 177 people with confirmed trypanosomiasis (CATT+/T+) to 146 unconfirmed seropositive subjects (CATT+/T-). Measured risk factors were vectorial level of transmission (low or high), living conditions (small hamlets or village) and social way of life (SWL). SWL was defined as *living or not in campement* i.e., temporary dwellings located in coffee or cacao plantations with important mixing of population particularly around wells. These risk factors were described as important determinants for host/vector contact. Individual factors were age, sex and ethnic group (1 native and 4 allogenous groups originating from northern areas or countries with a very low level of trypanosomiasis endemicity).

Although living in small hamlet ($p=0.005$), high entomological risk ($p=0.03$), *campement* way of life ($p<0.0001$) and age ($p=0.008$) were significantly associated with the presence of trypanosoma in blood, ethnic group appeared to be also an important determinant ($p<0.0001$). Ethnic group can be considered as confounding factor since allogenous ethnic groups were significantly exposed to higher environmental and behavioural risk factors than natives ($p<0.0001$). However, multivariate analysis confirmed that ethnic group remains an important relevant factor since the number of T+ subjects was significantly higher within allogenous ethnic groups (mossi, senoufo, baoule or dioula) than within natives ($p<0.001$). Although nutritional habits or use of local medicines by ethnic groups have to be explored, genetically mediated susceptibility to Tbg infection may be involved as shown for other parasitic diseases.

25

Efficient designs for studying measured genes and gene-environment interactions. Gauderman WJ¹, Thomas DC¹, Witte JS². 1. University of Southern California, USA; 2. Case-Western Reserve University, Cleveland Ohio, USA.

Case-control studies that utilize a relative of the case (e.g.

an unaffected sibling) as a control can be used to avoid the problem of bias due to population stratification. However, such relative pairs are overmatched with respect to their genes, which leads to a loss of efficiency compared to using controls selected from the population. We describe study designs using relative controls in which the sample is restricted to families with two or more cases, for example restriction to case-sib pairs that have an affected mother. We review the basic criteria for guaranteeing unbiased estimation, namely that the additional affected relative have the same relationship to the case as to the control. While this is usually automatic for siblings, it is not when using more distant relative controls, such as cousins. We show that these multi-case designs can be much more efficient than unrestricted designs for estimating both main and interaction effects, particularly for the study of rare major genes such as BRCA1. As an example, consider a dominant major gene with allele frequency 0.01 and relative risk 20.0 for carriers compared to noncarriers. In this situation, restricting the sample to case-sib control pairs with an affected parent is 3 times more efficient for estimating genetic relative risk than using unrestricted case/sib-control pairs, and 1.5 times more efficient than using case/unrelated-control pairs. We also discuss methods for exploiting phenotypic information that may be available on family members other than those involved in the case-control comparison.

26

Two hypotheses explaining the contribution of HLA in Rheumatoid Arthritis revisited using the MASC method with data on TNF-LT. Génin E^{1,6}, Babron MC¹, McDermott MF², Mulcahy B³, Waldron-Lynch F³, Adams C³, Clegg DO⁴, Ward RH⁵, Shanahan F³, Molloy M³, O'Gara F³, Clerget-Darpoux F¹, INSERM U155, Paris, France; ²University of London, UK; ³University College, Cork, Ireland; ⁴University of Utah, USA; ⁵University of Oxford, UK; ⁶University of California, Berkeley, USA.

To explain the association between HLA-DRB1 gene and Rheumatoid Arthritis (RA), two main hypotheses have been proposed. The first hypothesis, the shared epitope hypothesis, assumes a direct role of DRB1 in RA susceptibility. The second hypothesis assumes a recessive disease susceptibility gene in linkage disequilibrium with DRB1.

To investigate these two hypotheses, we analyzed data on the HLA-DRB1 and TNF-LT loci in 49 affected sib-pairs. We used the Marker Association Segregation Chi-Square (MASC) method where the genotype distribution of markers among index cases and the haplotype sharing in affected sib-pairs are jointly taken into account. With DRB1 data alone, both hypotheses

were shown to fit but with analysis of TNF data, both hypotheses were strongly rejected. Thus, the TNF data provided additional information for a better understanding of genetic susceptibility to RA than was previously possible using only HLA-DR data.

A theoretical standpoint is addressed here on the advisability of using different linked markers in a candidate region for modeling the contribution of this region in disease susceptibility.

27

THE FUSION (FINLAND-UNITED STATES INVESTIGATION OF NIDDM GENETICS) STUDY

Ghosh S¹, Hauser E², Magnuson V³, Valle T³, Ally D⁴, Watanabe RM², Hagopian W⁵, Bergman R⁶, Tuomilehto J¹, Collins F¹, Boehnke M² for the FUSION Team.
¹NHGRI, NIH, Bethesda, Maryland; ²University of Michigan, Ann Arbor, Michigan; ³National Public Health Institute, Helsinki, Finland; ⁴University of Washington, Seattle, Washington; ⁵University of Southern California School of Medicine, Los Angeles, California.

The FUSION study is a collaborative effort to map genes for NIDDM (non-insulin dependent diabetes mellitus) in the Finnish population. We are performing a genome scan using approximately 400 microsatellite markers in 710 NIDDM-affected sib pairs (ASPs) from 473 families to create a map with an average resolution of 10cM. Using criteria based on serum C-peptide levels and GAD antibodies in diabetics we have excluded 43 families with probable late-onset IDDM. In 183 families we are also typing the spouse and offspring ($n=446$) of one of the affected sibs and measuring insulin resistance and beta cell function in these unaffected family members using a frequently sampled intravenous glucose tolerance test and the minimal model. 220 unaffected controls aged 71 years are also being studied. More than 800,000 first-pass genotypings have been completed so far with an average error rate based on duplicate samples of 0.12%. Our analysis strategy includes non-parametric, multi-point identity-by-descent and single-point identity-by-state methods and association analyses. To date no major susceptibility loci for NIDDM have been identified. We can exclude chromosomes 1, 5, 6, 7, 8, 12, 13, 15 and 16 at a lambda (s) value of 1.55 while chromosome 2 can be excluded at 1.9 and chromosome 20 at 2.1 all under an additive model.

28

Clinical and epidemiologic factors in melanoma-prone families with p16 mutations. AM Goldstein¹,

RT Falk¹, NC Dracopoli², RS Sikorski³, WH Clark⁴, Jr, MA Tucker¹. ¹Natl Cancer Inst, Bethesda, MD, USA; ²Sequana Therapeutics, Inc, La Jolla, CA, USA; ³Natl Human Genome Research Inst, Bethesda, MD; ⁴Beth Israel Deaconess Medical Center, Boston, MA, USA.

The p16 (or CDKN2) gene, located on chromosome 9p21, has been implicated in malignant melanoma (MM) pathogenesis. Although germline p16 mutations confer substantial risk in about 1/4 of melanoma-prone kindreds, other genetic and/or environmental factors likely influence disease expression.

We identified 13 MM kindreds segregating disease-specific germline p16 mutations. Conditional logistic regression was used to estimate relative risks (OR) for p16 and other melanoma risk factors comparing MM cases to non-MM relatives in their family. Univariate (age-adjusted) analyses showed significant odds ratios (OR) for p16, dysplastic nevi (DN), total nevi (TN), complexion (CO), solar injury (SI) and abnormal mole pattern (AMP). Even after adjustment for p16 and age, DN (OR=10, 95% CI 3-67), AMP (OR=6, 2-23), TN (OR=13, 4-59), SI (OR=4, 1-16), and CO (OR=3, 1-13) each showed significant associations with MM. The most parsimonious model revealed that risk for MM increased with p16 mutation (OR=203, 28-4961), TN (OR=18, 4-108), and SI (OR=4, 1-19). We also analyzed the subset of p16 mutation carriers (50 cases, 44 controls) using unconditional approaches. The best fitting model included TN (OR=7, 2-33), SI (OR=5, 1-23), and tanning ability (OR=6, 1-38). Thus, factors (total nevi and skin's response to sun exposure) other than and separate from p16 appear to influence MM expression in these melanoma-prone families.

29

Response Surface Analysis for Lod Score Optimization. Gordon D¹, Finch S², Mendell N², Chen C², McInnis M³, McMahon F³, Koskela R⁴, Botstein D⁴, DePaulo J³, Marr T¹. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA; ²State University of New York at Stony Brook, USA; ³Johns Hopkins School of Medicine, Baltimore, MD, USA; ⁴Stanford University, Stanford, CA, USA.

We apply exploratory data analysis techniques, analysis of variance techniques, and response surface analysis to find the genetic penetrance model that maximizes a transformation of lod scores across markers. The penetrance models we consider follow a $2^5 \times M$ factorial design, with two settings for

each of five categories: fraction of homozygous non-disease individuals affected, fraction of homozygous disease individuals affected, dominant vs. recessive, disease allele frequency, and age of onset (M =number of markers). We perform this analysis using lod scores as reported in GENEHUNTER. We use data from a study of multiplex families ascertained through probands with bipolar disorder. We report results of our analyses on four chromosomes: two for which there is suggested linkage in the literature, and two for which there is no suggested linkage.

30

Family history, ethnicity, and relative risk of breast cancer ((Grabrick DM, Walsh A, Anderson KE, Folsom AR, Sellers TA)) Division of Epidemiology, University of Minnesota, Minneapolis, Minnesota

A family history of breast cancer is considered one of the strongest predictors of the disease. In a large cohort of 24,109 post-menopausal women from Iowa, we examined whether the risk associated with a family history of breast cancer differs by self-reported ethnicity. A total of 907 breast cancer cases occurred over 10 years of follow-up. Using a phylogenetic tree, ethnicities were combined into five groups: Scandinavian; English, Scottish, Welsh, and Dutch; Irish; German; and Other European. The incidence of breast cancer did not significantly differ by ethnicity, although the highest rate appeared to be in Scandinavians and the lowest in Irish. The proportion of women with a family history of breast cancer did not differ by ethnicity when only first degree relatives were considered ($p=0.17$); inclusion of second degree relatives enhanced the differences ($p=0.003$). Differences in mean levels of breast cancer risk factors between ethnicities were generally small, yet statistically significant. Proportional hazards regression was performed to evaluate potential interactions of family history with ethnicity. A family history of breast cancer was associated with increased relative risks among English, Scottish, Welsh, and Dutch, Germans, and Other Europeans, but not among Irish and Scandinavians. The highest rate of breast cancer was seen in women of English, Scottish, Welsh, or Dutch descent, with a family history of breast cancer, while the lowest rate was seen in women of Irish descent without a family history of breast cancer. With the addition of known breast cancer risk factors to the model, a family history of breast cancer no longer appeared to be a risk factor among Other Europeans. Relative risk estimates were attenuated upon addition of known breast cancer risk factors to the model, implying that the distribution of these risk factors by ethnicity may explain some of the differences observed between ethnicities with a family history of breast cancer.

31

A linkage strategy for detection of human QTLs: the effect of IBD imputation on the analysis of extreme sibpairs.

Gu C* and Rao DC

Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA

The use of sibpairs with extreme phenotypes has been shown to be a promising and powerful design for detection of human quantitative trait loci (QTLs). We have dealt with some practical issues such as optimizing trait thresholds for extreme phenotypes elsewhere. Herein, we address another important factor that would influence the efficiency of an extreme sibpair (ESP) design. Namely, how should incomplete linkage information be treated. Since the number of ESPs required for analysis is relatively small, the cost of genotyping their parents can be often justified. However, there are cases where genotypic data on the parents are missing by nature (e.g., deceased). Moreover, unless the parents are completely informative at the marker under consideration or the marker is completely polymorphic, non-zero identical-by-descent (IBD) allele sharing by a sibpair is indeterminable. Imputation of IBD sharing has proved to be an effective and viable way to account for this uncertainty and to utilize information carried in partially informative relatives for 0-1 traits. We show that for quantitative traits the methodology of IBD imputation also enhances the effectiveness of the ESP design. We consider three methods to infer missing IBD information: a simple probability model based on parental mating type and sibpair type, a conditional likelihood model based on the Elston-Stewart algorithm, and a MLR model based on Risch's likelihood ratio test. Their effects on power and cost effectiveness of ESP designs are evaluated using simulations.

32

Linkage Information Content of polymorphic genetic markers. Guo X, Elston RC. Case Western Reserve University, Cleveland, OH, USA.

The Polymorphism Information Content (PIC) value [Botstein et al., 1980] is often used to measure the informativeness of a genetic marker for linkage studies. The PIC value was first derived for the situation of a rare dominant disease, when one of the parents is affected, and is a function of the particular mode of disease inheritance [Chakravarti and Buetow, 1985]. However, the genetic mechanism underlying a disease trait is often unknown, especially for complex diseases. Under these circumstances model-free linkage analysis, which does not require specifying the mode of inheritance of the trait, may be applied. The PIC value calculated for a rare dominant disease is an

528 Abstracts

inappropriate measure of linkage information to design a model-free linkage analysis. In this paper, we develop a Linkage Information Content (LIC) value to measure the informativeness of a marker about the IBD sharing status of a pair of relatives. A marker's LIC value depends on the number of marker alleles and their frequencies, and also on the type of relative pairs to be used in a study. However, it is independent of the mode of inheritance of the disease. Knowing the LIC value and the type of relative pairs to be studied (regardless of whether they are affected pairs or discordant pairs), it is possible to allow for incomplete marker information to determine the effective number of fully informative pairs in a study.

33

Properties of a transmission/disequilibrium test for quantitative traits. Hanson RL, Knowler WC. NIDDK, Phoenix, AZ, USA.

The transmission/disequilibrium test (TDT), which assesses whether an allele is transmitted preferentially to an affected offspring, tests for linkage and association between a marker and a dichotomous trait. One can also test for a linear association of a quantitative phenotype with transmission of a certain parental allele.

Simulations were conducted to evaluate the utility of this approach in analyzing body mass index (BMI) in a study of 328 offspring from 99 nuclear families. The SLINK program was used to simulate biallelic markers linked to a hypothetical biallelic locus that influences BMI. Models used varied in frequency of the high BMI allele, q , (assumed to be the same as the frequency of one allele at the marker locus), the proportion of variance in BMI accounted for by the trait locus, h^2 , and the disequilibrium between the marker and trait loci, δ .

In 10,000 replicates generated under the null hypothesis of no linkage, type I error rates were virtually identical to the nominal values. Power to detect linkage and association depended on q , h^2 and δ . At $\delta=0.25$, $q=0.5$ and $h^2=0.3$, 91% of replicates had $p<0.0001$; equivalent power (89%) at $q=0.2$ required $h^2=0.5$. Power declined rapidly with diminishing values of δ ; at $q=0.5$ and $h^2=0.5$, power was 99%, 75%, 32% and 7% at $\delta=0.25$, 0.20, 0.15 and 0.10, respectively.

Evaluation of a linear association in the TDT may be useful for analyzing linkage and association with genes influencing quantitative traits. However, the presence of strong linkage disequilibrium may be necessary to detect such effects.

34

Cholesterol levels in first degree relatives of patients after myocardial infarction

Hengstenberg C¹, Engel S², Döring A², Tietze M¹, Holmer S¹, Riegger AJG¹, Schunkert H¹, Loewel H². ¹University of Regensburg; ²GSF-Institut of Epidemiology; ³Central Hospital of Augsburg, Germany.

Patients with severe hypercholesterolemia have been shown to benefit from lipid lowering therapy (LLT). In particular, the use of CSE inhibitors may result in primary and secondary prevention of myocardial infarction (MI). Regardless of cholesterol levels, the risk to suffer from MI is increased in first degree relatives. The aim of this study was to analyze lipid levels and the use of preventive strategies in siblings of patients with MI.

Methods: A total of 109 male MI patients who suffered from a heart attack prior to the age of 60 years and 116 brothers of these MI patients were studied by questionnaire as well as anthropometry and laboratory measurements.

Results: The mean age of the MI patients was 56.6 (± 6.5) years, the mean age of their brothers was 54.7 (± 8.4) years. The mean blood levels in patients were total cholesterol (TC) 230 mg/dL, HDL 47 mg/L and LDL 136 mg/L, and in their brothers 236, 48 and 204 mg/L (n.s.), respectively. Levels of Lp(a) were also not significantly different in both groups. In contrast, MI patients took LLT in 41%, whereas their brothers in only 3% ($p<0.001$). The TC/HDL ratio was >5 in 53% of MI patients receiving LLT, in 50% of non-treated patients and in 52% of siblings without LLT.

Conclusion: Treatment of patients after MI includes often lipid lowering drugs, whereas their relatives receive only rarely such drugs despite similar lipid profiles. The risk for MI is increased for first degree relatives of patients with MI and it can be speculated that this risk will even increase with non-treated mild to moderate hypercholesterolemia.

35

Secular trends in mortality in six large families with hereditary breast-ovarian cancer from 1840 to 1994.

Hille ETM¹, Klijn JGM², Meijers-Heijboer EJ³, Rosendaal FR¹, Coebergh JWW^{1,3}, Cornelisse CJ¹, Vandenbroucke JP¹. ¹Leiden University Medical Center; ²Daniël den Hoed Cancer Center, Rotterdam; ³Erasmus University Rotterdam; The Netherlands.

In the present study, we selected six families with hereditary breast-ovarian cancer, irrespective of a known genetic defect. We investigated all-cause mortality from 1840 to 1994 in affected family members, obligate and potential carriers to estimate the impact of a hereditary defect on mortality in women. Data of 153 women, who were followed from 20 years on till death or the end-of-

study date, were analyzed using the family tree mortality ratio (FTMR) method: the observed number of deaths was compared with the expected number, based on the general population and adjusted for age and calendar time. The influence of parity and generation was studied.

In 5162 woman-years 69 deaths occurred, whereas 35 deaths were expected (FTMR 2.0; 95%CI 1.5-2.5). Excess mortality varied strongly between families and was confined to the ages 30 to 60 years (FTMR 4.4; 95%CI 3.1-6.0). When dividing the follow-up time for these women in four calendar periods, there was a secular increase of mortality, partly related to excessive deaths in the youngest generations. In this generation ascertainment bias could have played a role. The FTMR in nulliparous women and women with 1 or 2 children was two times higher than in women with 3+ children.

The impact of the genetic susceptibility on mortality is confined to early ages and might have increased over the century. The unbiased estimate is given by the mortality rates of the oldest generations.

36

The Familial Atypical Multiple Mole-Melanoma (FAMMM) syndrome: mortality and causes of death in six Dutch families from 1830 to 1994. Hille ETM, van Duijn D, Gruis NA, Rosendaal FR, Bergman W, Vandebroucke JP. Leiden University Medical Center, The Netherlands.

To estimate the impact of a hereditary defect of the CDKN2 gene on mortality, we investigated all-cause mortality from 1830 to 1994 and causes of death from 1941 to 1994 in six FAMMM families, showing an autosomal dominant pattern of inheritance for a 19-bp deletion in this gene. Follow-up extended from 20 years after birth till death or the end-of-study date. Mortality data were analyzed with the family tree mortality ratio (FTMR) method: the number of deaths among 224 proven, obligate and potential gene carriers (observed) was compared with that of the general Dutch population (expected), adjusted for sex, age and calendar period.

From 1830 to 1994, 65 deaths took place in 6022 person-years, whereas 42 deaths were expected (FTMR 1.6, 95%CI 1.2-2.0). Excess mortality was confined to the ages 35 to 70 years (FTMR 2.1, 95%CI 1.5-2.9). The FTMR doubled with calendar time from 1870 to the present. Excess mortality could be fully attributed to malignancies, especially to melanoma of the skin and pancreatic carcinoma. There were no significant differences between the families, although the specific cancer pattern within a family varied.

We conclude that, in comparison with the general population, the impact of this CDKN2 gene is rising.

The greatest challenge for the future is, what kind of screening can be offered to these families, especially since, besides deaths from melanoma of the skin, their mortality of pancreatic carcinoma is also heightened.

37

Affected sib-pair analysis on multiplex sibships: Power and Type I error probability

Holmans P.

(Dept. of Psychological Medicine, Univ. of Wales College of Medicine, Cardiff)

Many affected sib-pair studies contain a number of sibships containing three or more affected individuals. The most efficient method for analysing such sibships is not clear. In practice, the most common procedure is to form all possible sib pairs from the sibships, and analyse them as a sib pair sample. Since the resulting sib pairs are not independent, analysing them as such may lead to an inflation of the Type I error (Suarez & Van Eerdewegh 1984, Am J Med Genet 46:242-253). It is usual, therefore, to downweight sib pairs from multiplex sibships in an attempt to counteract the lack of independence. The most commonly-used scheme is to weight pairs formed from a sibship containing N affecteds by $2/N$ (Suarez & Hodge 1979, Clinical Genetics 15:125-136). However, it has been shown (Mounier et al. 1997, Genet Epidemiology, in press) that this weighting is often conservative.

The Type I error probability given by the likelihood-ratio method of sib-pair analysis (Risch 1990 Am J Hum Genet 46:242-253, Holmans 1993 Am J Hum Genet 52:362-374) when applied to multiply-affected sibships was investigated. This was found to depend heavily on the informativity of the marker, the presence of typed parents, and whether the siblings not being counted in the "affected pair" were included in the analysis to give information on missing parental genotypes. In view of these results, and the unknown asymptotic distribution of the likelihood-ratio test statistic when applied to samples of dependent sib pairs, significance levels were obtained by Monte-Carlo methods. The power given by such an approach was compared to that of the mean test (Blackwelder & Elston 1985, Genet Epidemiology 2:85-97) and the method of Green & Grennan (1991, Ann Hum Genet 55:243-249) under a variety of markers and sibship structures.

38

Association studies using DNA pooling

Holmans P, Daniels J, Owen M, McGuiffin P.
(Depts of Psychological Medicine and Medical Genetics,
University of Wales College of Medicine, Cardiff)

Association studies have been proposed as an alternative to linkage studies for locating susceptibility genes for complex traits (Risch and Merikangas 1996, Science 273:1516-1517). The

main advantage of association tests is that they are more sensitive than linkage tests, making it possible to detect relatively small genetic effects, such as may be involved in complex traits, without requiring huge sample sizes. However, since detectable associations with disease genes extend only over relatively short distances, a genome scan for association will require a fine grid of markers (<1cM). This in turn may require a prohibitive amount of genotyping. One way of reducing the amount of genotyping is to "pool" the DNA from various members of the sample. This has been used in QTL studies in animal populations (Darvasi and Soller 1994, Genetics 138:1365-1373), and in studying recessive diseases in inbred human populations (Carmi et al. 1995, Human molecular Genetics 4:9-13). We present a method for performing case-control association studies in out-bred populations.

When a pooled sample of DNA is analysed by GENESCAN software on the ABI fluorescent system, an allele image pattern, consisting of a number of peaks corresponding to the various alleles, is produced. Our method quantifies the difference between the image patterns produced by the case and control samples, using this as a test statistic for association. Test criteria and powers are calculated under a variety of sample sizes and marker allele frequency distributions. The effect of stutter bands and differential amplification is investigated, and various methods of correcting for them compared. The methodology is also applied to real marker data, enabling comparisons with standard methods of analysis to be made.

39

A family history score for breast cancer.

Houwing-Duistermaat JJ, Van Houwelingen JC.
Medical Statistics, Leiden University, Leiden, The Netherlands.

In Dutch family cancer clinics, Claus tables¹ are commonly used for risk assessment. The tables are based on a model assuming a rare autosomal dominant allele leading to increase susceptibility to breast cancer. By using these tables no adjustments are made for other risk factors and the information on healthy relatives is not used. The aim is to derive a family history score (FS) that can be incorporated as a covariate besides other risk factors in a regression model.

One way is to use Claus model to compute the *In of the life time odds* given the family observations and the current age of the woman whose risk has to be modelled. Computation of this score (FS1) involves averaging over the distribution of genotype configurations. By using first order approximations, a straightforward to compute score (FS2) can be derived. FS2 is based on observed healthy and diseased female relatives.

To check the performance of FS2, 2000 dis-

ease and age-at-onset patterns are simulated under Claus model for healthy women of age 15 with mothers of age 35 and grandmothers of age 75. The correlation between FS1 and FS2 for these women is high, namely 0.87. For a second simulated data-set containing women of age 25 with older sisters and mothers of age 55, the correlation is 0.77. The results are promising and FS2 may be an alternative for FS1 in a model for breast cancer.
¹Claus EB et al. Cancer 1994;73: 643-51.

40

Using isolated populations to locate low penetrance disease genes.

Iles MM, Bishop DT.
ICRF Genetic Epidemiology Laboratory, Leeds, UK.

It is widely accepted that the future of genetic studies of complex diseases lies in finding low penetrance genes. Risch and Merikangas (Science 273:1516, 1996) show that for low relative risks sample sizes required for linkage analysis are too large and assert that association studies are feasible. Their analysis assumed that the polymorphism studied within the disease gene is the disease allele. If instead a linked polymorphism is typed then the sample sizes required to detect linkage are far greater (Muller-Myhsok and Abel, Science 275:1328, 1997).

Population structure can play a major part in the relative frequencies of the marker and disease allele. Formulae were derived from those of Risch and Merikangas and various population structures simulated to find the situations under which sufficient power to detect linkage could be achieved with a reasonably-sized sample of individuals. It was assumed in the simulations that a single mutation existed in the first generation, surviving until the present time. It was found that a population which has experienced a period of rapid expansion can be used to locate disease genes using association studies. If the disease allele confers a GRR of 4, the prevalence of the disease in the general population is 5% and a polymorphic marker with ten different alleles is used in every gene, the proportion of simulations with at least 80% power can be seen in the table below. Simulations indicate that while the rate of expansion is unimportant the age of the population has an effect. The results suggest that small isolated populations could be successfully utilised in the future to locate disease genes with lower penetrance.

Final Size	Initial Size		
	10	20	40
640	0.50	0.26	0.08
1280	0.67	0.45	0.21
2560	0.75	0.60	0.37
5120	0.77	0.67	0.52

41

A Markov chain Monte Carlo approach to survival models with detailed family history.
Iversen, Jr ES¹, Parmigiani G¹, Berry DA¹, Schildkraut JM¹. ¹Duke University, Durham, North Carolina, USA.

The genetic status of an individual, as defined by presence or absence of mutation, is often an important predictor of disease risk and/or survival. For example, germline mutations at BRCA1 and BRCA2 are associated with substantially elevated risks of breast and ovarian cancers and there is emerging evidence of an effect on survival. In situations like this one, where genetic testing is expensive but where an accepted inheritance model and population profile of the related disorders are in place, it is possible to estimate the probability that an individual carries a mutation, or combination of mutations, given a family history of the related disorders. We describe a technique for estimating effects associated with genetic status in the Cox proportional hazards model when the genetic status of members of the study group is unknown. The genetic status of an individual is captured in one or more latent categorical variables whose distribution is estimated. Population parameters are sampled according to a literature based distribution, genetic status and parameters in the Cox model from their conditional posterior distributions, with the process repeated in an iterative scheme. The result is an *a posteriori* sample of model parameters that accounts for sampling error, uncertainty in the genetic status of study participants, and uncertainty in estimates of population parameters. The effect associated with each variable in the Cox regression is estimated by a sample average; estimates of uncertainty by sample variances; and functions of the parameters, like survival curves and estimates of their variability, by averages and sample variances of the functions taken over the sample of parameters. Strategies for sampling parameters, issues relating to convergence of the process, and implementation are discussed in context of an example.

42**Family history and early onset colorectal cancer: a population-based case-control-family study.**

Jenkins M¹, Hopper J¹, St.John D², Macrae F², Forrest S³, Giles G⁴. The University of Melbourne; ²Royal Melbourne Hospital; ³Murdoch Institute; ⁴Anti-Cancer Council of Victoria, Melbourne, Victoria, Australia.

Family studies suggest that a proportion of early onset colorectal cancer (CRC) is due to genetic factors whether they be rare high risk mutations or common, risk modifying polymorphisms. In order to estimate the proportion of CRC associated with, and the aetiological fraction of, different genes it is necessary to study CRC and control families in a population-based setting.

Since 1993, we have studied incident CRC cases

diagnosed prior to age 50, living in Melbourne Australia, and identified via the state-wide Victorian Cancer Registry, their spouses (as controls) and the relatives of cases and of controls. We have surveyed 150 families, attempted to verify all reported cancers in relatives, and have collected 520 blood samples and 1,700 questionnaires, making this a valuable resource for studying the genetic epidemiology of early onset CRC.

By case-control analysis, the effect of family history of CRC on risk of CRC under 50 years was an odds ratio of 5.37 (95% C.I. 1.44-34.9) for at least one first degree relative affected, and 1.82 (0.80-4.47) for affected second degree relative(s) only. Family history consistent with the Amsterdam Criteria for Hereditary Non-Polyposis CRC was observed in 8% (3%-15%) of families. Associations between CRC and family history of other cancers will be presented with age-specific risks of cancer in relatives by age of onset and sex.

43**New logistic regressive models for family data.**
Karunaratne PM, Elston RC. Case Western Reserve University, Cleveland, OH, USA.

We consider modeling the familial correlation between two related individuals using a multiple logistic regressive model. It is shown that there is a discrepancy in the marginal probability of the second individual. We investigate the conditions under which this discrepancy can be minimized and show how it has a direct effect on handling missing values and ascertainment. We derive a functional between the parameters in the model that eliminates this discrepancy, hence solving the problems relating to the handling missing values and ascertainment. In order to extend this methodology to a situation of more than two individuals, we present a new model that enables us to separate the effects of the relatives. It is shown that the likelihood for sibship data under this model is independent of the order in which the sibs enter the likelihood calculation.

44**The Future of Genetic Studies of Complex Human Diseases:****An Epidemiologic Perspective**

Khoury M.J., Yang Q.H. Task Force on Genetics and Disease Prevention, Centers for Disease Control and Prevention, Atlanta, GA 30341

With increasing number of genetic markers scattered throughout the human genome, it will become increasingly possible to find genes for complex diseases by studying parental nontransmitted alleles using the transmission disequilibrium test. Risch and Merikanagas (Science 1996;273:1516-1517) have recently argued that this approach has better statistical power for genes with weak to moderate effects than traditional linkage analysis of pedigrees. In this paper, we show that epidemiologic studies that use population-based incident case-control designs will, on the long run, become even more successful not only in quantifying disease risks with specific alleles but in finding disease susceptibility genes as well. For adult onset common diseases, compared with family studies, epidemiologic studies are more practical and less expensive to conduct, have, more often than not, better statistical power than the transmission disequilibrium test, can define the magnitude of association between disease and specific alleles in terms of relative, attributable and absolute risks, can directly assess the impact of gene-environment interaction, and lead to better generalizability of the results to the underlying populations studied. The traditional limitation of epidemiologic studies, confounding due to population stratification will become more correctable as investigators will be able to choose appropriate control subjects, and adjust for population stratification using a combination of interview techniques and genetic markers. We also show that residual population stratification is unlikely to explain strong associations between disease and marker alleles. The other limitation of such studies, linkage disequilibrium will continue to improve as the human genome becomes increasingly dense with known markers and eventually totally sequenced. While family-based studies will continue to have an important role in the search for genes for common diseases, we predict that traditional epidemiologic studies will have an even greater role in finding disease genes and defining the magnitude of risks, an essential prerequisite to disease prevention.

45

Tests for genotype-phenotype association in related persons. Knowler WC¹, Hanson RL¹, Kobes S¹, Long JC². ¹NIDDK, Phoenix, AZ; ²NIAAA, Rockville, MD.

Association between phenotype and genotype provides evidence for a genetic effect. Standard tests, such as χ^2 or ANOVA for discrete or continuous phenotypes, assume independence of subjects. This does not hold in two situations: 1) each person is duplicated, once for each allele, for tests of allelic association; 2) subjects are related, e.g. sampled from a pedigree study or an isolated population. A familial phenotype may then be spuriously associated with a polymorphic marker. A proposed solution is to test for association after simulation of genotypes independent of phenotypes. The observed p-value (unadjusted for relatedness) is ranked with p-values from all simulations to obtain an empirical p-value.

The procedure was used in 1191 genotyped persons of the same ethnicity in 100 extended

pedigrees from a linkage study of type 2 diabetes. Tests of association of diabetes with simulated highly polymorphic unlinked markers yielded $p < 0.05$ in 15-20% of simulations, illustrating the need correct for relatedness. The marker Gm2 was negatively associated with diabetes (observed $p = 0.0005$); corrected for relatedness, the empirical $p = 0.0024$, about 5 times as high.

Whereas this procedure does not directly address multiple comparisons or confounding by ethnic heterogeneity, it has several virtues, including applicability to discrete or continuous phenotypes and to any pedigree structure (including unrelated individuals). All persons with genotypic and phenotypic data are informative.

46

A putative recessive gene for early onset of cancer in families of 257 probands with Ewing's sarcoma family of tumors. Korczak JF, Novakovic B, Wexler LH, Tucker MA, Goldstein AM. National Cancer Institute (NCI), Bethesda, MD, USA.

Ewing's sarcoma family of tumors (ESFT) is a subgroup of pediatric neuroectodermal tumors for which a genetic basis has been suggested. Based on the finding of a significant excess of neuroectodermal tumors (melanoma and brain cancer), bone and stomach cancer in relatives of 257 ESFT probands treated at NCI, we performed segregation analyses using Class A regressive logistic Models 1 (type-dependent age of onset) and 2 (type-dependent susceptibility), implemented in the S.A.G.E. program REGTL. Two definitions of affection were considered: (i) all cancer (ALL) and (ii) ESFT and those cancers in significant excess (SIG). We reported (Am J Hum Genet 59:A182, 1996) a possible recessive gene effect on cancer, either through an earlier onset (for ALL) or increased susceptibility (for SIG).

We report here the effect on the above analyses of allowing the type-dependent parameter (β in Model 1 or γ in Model 2) to be sex-dependent. This approach significantly improved the likelihood of the data only for ALL under Model 1 and, then, for all transmission hypotheses. The best-fitting hypothesis continued to be that of recessive inheritance. However now, for both ALL and SIG, the effect appeared to be manifested through an earlier age of onset.

These results support further study of genetic mechanisms in families with ESFT.

47

Incorporation of age of onset into multipoint linkage analysis using pseudolikelihood.
Li H¹, Huang J². Mayo Clinic, Rochester, MN, USA;²University of Iowa, Iowa City, USA.

For many diseases that have a complex genetic basis, study has suggested that disease genes influence not only the occurrence of disease, but also the age of onset. Current methods in linkage analysis and genome scanning are mainly concentrated on affected relative pairs or 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 fact, affected relatives with different ages of onset may be the result of different genetic etiologies and unaffected relatives are censored at the study time. Therefore, incorporation of age of onset and including contrasts between affected and unaffected pedigree members are important components of effective analysis for the detection of linkage with genetic markers. We propose to use multiple markers to infer the inheritance vector in order to extract information about the inheritance pattern of the disease allele in a pedigree. For a given inheritance vector, we define two neighbor sets for each individual based on allele sharing IBD. We then use the within-set and between-sets conditional hazard ratios to characterize the dependence of age of onset among relatives. A pseudolikelihood ratio test and a score test are proposed for testing linkage. Simulated and real data sets are used to illustrate these new statistical methods.

48

Using genealogic methods to study complex diseases in an isolated population. Lin J-P¹, Hirsch R¹, Jacobsson L², Scott WW³, Ma L³, Pillemer SR¹, Kastner DL¹, Bale SJ¹. NIAMS/NIH, Bethesda, MD, USA;²Lund Univ, Malmö, Sweden; ³Johns Hopkins Univ, Baltimore, MD, USA.

Due to the particular characteristics of many complex traits, they may not be amenable to traditional epidemiologic methods. Here we introduce an approach that defines an isolated population as the "unit" in which to carry out studies of complex disease. The Pima Indians of the Gila River valley are a relatively isolated and homogeneous population; more than 80% of their alleles are Pima-specific. The incidence of type II diabetes, gallbladder disease and rheumatoid arthritis (RA) are all exceedingly high in the population. However, previous studies of RA in the Pima, utilizing traditional methods, failed to detect significant familial aggregation. We constructed a genealogy for this population and used a genealogic index method to con-

duct studies of familial aggregation. We used an algorithm to identify biological relationships among 88 RA cases as well as 20 sets of 88 matched controls. The kinship coefficient was calculated for all possible pairs (3828) of RA cases, and the same method for the sets of controls. The total kinship coefficient among the RA cases, 5.92, was significantly higher than that of the average of the 20 sets controls, 1.99, and was elevated over that of the controls up to second cousin which supports genetic effects in the RA familial aggregation. We also identified all inbred individuals in this population and determined their inbreeding coefficients and the average for the entire population. The mean inbreeding coefficient, 0.00016, was not higher than that of the general population and none of RA cases were inbred. The creation of the Pima genealogy can be anticipated to provide valuable information for the genetic study of other diseases in this population. More importantly, the concept of defining the population, if it is relatively isolated and homogeneous, as the "unit" in which to assess familial aggregation may be a more powerful method for studying the complex diseases than has been available to date.

49

HLA class II in severity of coccidioidomycosis. Louie L, Ng S, Klitz W. University of California, Berkeley, USA.

We are screening HLA class II genes in subjects with coccidioidomycosis or Valley Fever (VF) to determine the role these genes may play in disease severity. We report our preliminary findings for DPB1, DQB1 and DQA1.

Subjects are from Kern County, CA, diagnosed with either mild (controls, n=83) or severe disseminated disease (cases, n=109). HLA genotypes are determined by PCR-SSO and SSPC. The results presented are for the 59 Hispanics and 54 Caucasian who have data available. Allele distributions and specific alleles are compared in contingency tables using the chi-square statistic.

Among Hispanics, no differences are found for DQB1 or DQA1. DPB1 has an overall difference in allele distribution between cases and controls ($p=0.03$). In particular, the *0201 allele is more common among cases (12.5%) than among controls (0%). $p=0.05$. Among Caucasians, no overall disease association with the loci tested is detected, as is true for DPB1*0201 alone.

To our knowledge, no other studies address the role of host genetic factors in severity of VF. Despite our small sample, we observe a difference in DPB1 allele frequencies between cases and controls of Hispanic descent. The *0201 allele is associated with increased risk of severe disease among patients diagnosed with VF. The significance of this risk is not large, possibly due to small sample size, other genes influencing disease, pathogen variation, and/or chance. We continue to collect samples from subjects with VF and screen class II HLA.

In summary, DPB1 is associated with severity of VF. These data support a role for class II HLA in this disease.

50

Genetic Linkage analysis and MCMC techniques: application of Multipoint analyses in complex pedigrees. Macciardi F¹, Morabito A² Institute H San Raffaele, Dept. of Neuroscience, Milan, Italy, ²Institute of Medical Statistics and Biometry, University of Milan, Italy

Genetic Linkage analysis is a powerful statistical tool routinely used in medical application to localize and map disease genes. Typical applications of genetic linkage analyses make use of pedigrees where a given genetic disease/trait is segregating and estimate the best chromosomal location of a putative gene responsible for the disease across a map of marker loci. When linkage analysis is based on likelihood calculations, it is computationally very intensive: exact calculation of the probabilities of the various genetic parameters is very complex and prohibitive, especially when large pedigrees are explored and/or many markers are jointly evaluated. Classical methods for linkage analyses need a large amount of time to carry out estimations for multipoint analyses. Alternatively, Markov Chain Monte Carlo methods have been recently proposed for estimating the probability and computing the likelihood of genetic linkage in pedigrees. This statistical technique is still awaiting for a massive exploitation of its potentialities, even though a consistent amount of work has been already accomplished (Thompson and Guo, 1991; Geyer and Thompson, 1995; Jensen and Kong, 1995; 1995; Lin, 1996). These techniques make feasible using complex models and a complete utilization of all the available information. Thus, it is possible to challenge the linkage problem even on complex pedigrees and to calculate the likelihood for a multipoint map with multiallelic loci. Complex pedigrees where schizophrenia is segregating and several marker loci map to a potential disease locus are evaluated with MCMC techniques and their performance are discussed.

51

Haplotype-based analysis of linkage disequilibrium. Martin RB, MacLean CJ. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, USA.

We propose a new method for extending disequilibrium analysis from single markers to multiple marker haplotypes. Such haplotypes have greater statistical power to disclose the presence of a disease susceptibility locus. Under what we call a 'causal hypothesis', a disease susceptibility mutation originally occurred in some founder, and as the chromosomal region around the mutation was passed from generation to generation, it was trimmed by recombination. If we consider a haplotype of marker loci clustered

around the disease, individuals who inherit the mutation should have similar 'trimmed haplotypes'. The probability of preservation of such trimmed haplotypes is a function of inter-marker distance and time, while random haplotypes which do not contain the disease mutation display similarity only to the extent determined by haplotype frequencies.

We combine this information to formulate a haplotype probability of association with the disease gene. Haplotypes for which this probability is high should be in linkage disequilibrium, measured by transmission disequilibrium or other means, whereas less disequilibrium is expected from haplotypes in which similarity is weak. To test for the presence of a disease susceptibility gene, we use regression analysis and a test of trend in disequilibrium. These methods have been combined into a unix-based software package.

52

Hodgkin's Disease: Family history of cancer and pesticide exposure. Cross Canada Study of Pesticides and Health. University of Saskatchewan, Saskatoon, SK, S7N 0W8, Canada.

Family history of cancer may reflect genetic or common environmental effects. In order to confirm the association between family history of cancer in first degree relatives and increased risk of Hodgkin's disease and to extend those findings by examining environmental exposures: pesticide exposure (\geq 10 hours per year) and cigarette smoking as effect modifiers or confounders, we conducted a Canadian case (n = 316), control (n = 1506) population-based study. Males who were aged between 19 and 84 years at diagnosis of a first incident primary tumour classified as Hodgkin's Disease (ICD-9 201) resident in six Canadian provinces were eligible cases. Controls were frequency matched by age, sex and province to cases of Hodgkin's Disease, Multiple Myeloma, non-Hodgkin's Lymphoma and Soft Tissue Sarcoma. Odds ratios (OR) and 95% confidence intervals (CI) were estimated using conditional logistic regression controlling for the matching variables age and province. We found increased risks: for affected relatives (A): fathers (OR(95%CI) 1.65(1.07, 2.54); (B) mothers 1.58 (1.04, 2.41), (C) at least one sibling 2.14 (1.36, 3.38), (D) one affected generation 1.98 (1.46, 2.69), (E) two or more affected generations 2.13 (0.98, 4.62), (F) family history of leukemia, lymphoma or multiple myeloma 3.61 (1.65, 7.86) and (G) a history of birth defects in case offspring (11.1% vs 6.2%, p= .0015). Neither cigarette smoking or exposure to pesticides \geq 10 hours per year materially modified these risks. We conclude that family history of any cancer and specifically leukemia, lymphoma or multiple myeloma independently increase the risk of developing Hodgkin's Disease. (Health Canada)

53

Linkage of bipolar affective disorder to chromosome 18q: Confirmation in a new pedigree series.

McMahon FJ¹, Shaw S², Xu J³, McInnis MG¹, Hopkins PJ², Cardon L², Simpson SG¹, MacKinnon DF¹, Stine OC¹, Sherrington R², Meyers DA³, and DePaulo JR¹.

¹The Johns Hopkins Univ. School of Medicine, Baltimore, MD; ²Sequana Therapeutics, Inc., La Jolla, CA, ³Univ. of Maryland Medical School, Baltimore, MD

Several groups have reported linkage of bipolar affective disorder (BPAD) to chromosome 18. We reported data from 28 pedigrees that showed linkage peaks at marker loci on 18p and at loci 40 cM distant on 18q. Most of the linkage evidence derived from families with affected phenotypes in only the paternal lineage and from marker alleles transmitted on the paternal chromosome. We now report results from a new series of 30 pedigrees (n=259) genotyped for 13 polymorphic markers spanning chromosome 18. Subjects were interviewed by a psychiatrist and diagnosed with highly reliable methods. Genotypes were generated with automated technology and scored blind to phenotype. Affected sib-pairs showed excess allele sharing at the neighboring 18q markers D18s541 (IBD=0.62, p=0.0003) and D18s38 (IBD=0.54, p=0.0352). Excess sharing was greatest for 18q marker alleles transmitted on the paternal chromosome but allele sharing in this sample was not greater in families with affected phenotypes in the paternal lineage. Little evidence of linkage was detected for markers elsewhere on chromosome 18. Genehunter multipoint non-parametric linkage analysis in the new sample combined with the original sample of families supports linkage on chromosome 18q at the p=0.0019 level, but the susceptibility gene is not well-localized. These results strengthen the evidence that a susceptibility gene for BPAD resides on the long arm of chromosome 18.

54

The influence of marker and disease allele frequencies on genetic association studies. Müller-Myhsok B¹, Abel L².

Department of Molecular Genetics, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany;

²INSERM U 436, Mathematical and Statistical Modelling in Biology and Medicine, Paris, France.

It has recently been argued (Risch and Merikangas, 1996) that the future of genetic studies in complex disorders belongs to association designs, such as the TDT, because of the inherently higher power of such studies. However, this power was computed in the setting that the allele investigated was the disease allele itself. A more common situation is, and could well remain, the analysis of polymorphisms which have a low prior probability to be the disease allele even when within the actual disease gene.

Using the same genetic model as in (Risch and Merikangas, 1996) we computed the number of necessary families for different degrees of disequilibrium δ and marker (m and $1-m$) and disease (p and $1-p$) allele frequencies for given power and type I error values. The situation in Risch and Merikangas corresponds to $\delta=\delta_{\max}$ and $m=p$. The number of necessary families increases dramatically as p differs from m even when $\delta=\delta_{\max}$, and also as δ decreases.

Researchers should thus be aware that the power of association studies can be greatly diminished as soon as the ratio m/p departs from unity and the linkage disequilibrium becomes weaker.

References:

- Risch N and Merikangas M (1996). Science 73, 1516.
Müller-Myhsok B and Abel L in: Scott WK, Pericak-Vance MA, Haines JL, Bell DA, Taylor JA, Long D, Grote MN, Langley CH, Müller-Myhsok B, Abel L (1997). Genetic Analysis of Complex Diseases. Science 275:1327ff.

55

Using Family History Information to Distinguish True and False Positive Model-Free Linkage Results. Jane M. Olson, Robert C. Elston. Department of Epidemiology and Biostatistics, Case Western Reserve University and MetroHealth Medical Center, Cleveland, USA.

Genome scans using tests for increased marker identity-by-descent sharing between pairs of affected siblings have become increasingly common. These analytic methods do not require a priori specification of a genetic model of inheritance for the disease locus and may thus be more robust than model-based methods. However, for a disease with a dominant mode of inheritance, power may be substantially reduced because the methods do not incorporate information about parental disease status and thus cannot consider the parental source of the disease allele. We propose to include family history information in a model-free analysis for the purpose of testing whether a positive linkage result obtained during the course of a genome scan is a true or false positive result. We describe a simple nonparametric test and a likelihood ratio test based on an extension of the Risch affected-sib-pair likelihood. The key to the new test statistics is the interaction between gender-specific marker allele-sharing and gender-specific family history of disease. The method is useful when the disease locus of interest has a dominant mode of inheritance and a sufficient number of parents are genotyped at the marker locus. If these conditions are met, the proposed tests have good power to differentiate between true and false positive linkage results. In general, the models may be used to increase the evidence for linkage (i.e., increase the lod score), although at the expense of estimating additional parameters.

56

Segregation analysis using Gibbs sampling for quantitative traits related to the asthma phenotype.
LJ Palmer, PR Burton, PN LeSouef, AL James, G Ryan, WOCM Cookson, AW Musk. TVW Telethon Institute for Child Health Research and University of Western Australia. PO Box 855, West Perth, Western Australia.

Aim: To investigate the nature and inter-relationships of major genes modulating asthma-associated traits in a population-based sample of Australian Caucasian families.

Methods: A cohort of 1020 individuals comprising 234 nuclear families was comprehensively evaluated. The quantitative traits assessed included serum levels of total and specific IgE, forced expiratory volume in 1 second (FEV₁), forced vital capacity (FVC), and the dose-response slope (DRS) of FEV₁ to methacholine provocation. Bayesian inference using Gibbs Sampling was used to fit ACCSE models with major gene effects. **Results:** Log-total serum IgE levels were consistent with a recessive gene effect (prevalence=56%). Specific IgE levels were consistent with a significant recessive gene effect (prevalence=48%) as was the DRS (prevalence=27%). FVC levels were consistent with a significant dominant gene effect. FEV₁ levels were not consistent with a major gene effect. All estimates were adjusted for age, height, tobacco smoke exposure and gender. Extended modelling suggested that the data were generally consistent with only a small overlap in the additive genetic determinants (i.e., genes) of these quantitative traits. However, there was a substantial overlap in the additive genetic determinants of (a) total and specific serum IgE levels, and (b) FEV₁ and FVC levels. **Conclusions:** These data are consistent with the existence of at least 3 distinct genetic pathways involving major gene effects in the pathogenesis of asthma

58

Etiologic heterogeneity in age of onset of lung cancer.
Petersen GM, Liao CJ, Tsai YY, Hoover D. Johns Hopkins University, Baltimore, USA.

Family studies suggest that early onset lung cancer may be explained in part by major susceptibility genes. We evaluated time trends of lung cancer in the United States by age group in order to test predictions that could be made by proposed genetic and non-genetic etiologies.

Methods: Lung cancer cases aged 20 and older in 1973-92 in the U.S. Surveillance, Epidemiology, and End Results (SEER) program formed the study population. Lung cancer cases were stratified into two age groups: early onset lung cancer (age 20-49 years) and older onset lung cancer (>49 years). Time trend analyses of incidence, proportion of early onset cases relative to all cases, and sex ratios were performed. **Results:** Incidence of early onset lung cancer remained stable over two decades, while a significant trend of increased incidence was found in older onset lung cancer. Over time, there was no change in proportion of early onset lung cancer compared to all lung cancer cases. Compared to a dramatic decrease of the sex ratio in the older onset lung cancer cases, a modest decrease in the sex ratio which stabilized over the last 15 years was found in the early onset cases. **Conclusions:** Most of the increase in lung cancer incidence between 1973 and 1992 and the rise of lung cancer in females is accounted for by cases with onset after age 50. The incidence and sex ratio among early onset cases has remained relatively stable over two decades. These findings suggest that the etiologic basis in these two age groups may differ. The patterns observed for early onset lung cancer are consistent with a hypothesized genetic susceptibility component in this form of lung cancer.

57

Determining Carrier Probabilities for Breast Cancer Susceptibility Genes BRCA1 and BRCA2. Parmigiani G¹, Berry DA¹, Aguilar O¹. ¹Duke University, Durham, NC, USA.

We present methodology, a model and software for evaluating the probabilities that a woman is a carrier of a mutation of BRCA1 and BRCA2, based on her family history of breast and ovarian cancer in first- and second-degree relatives. We discuss how to obtain carrier probabilities from Bayes' rule, using family history as evidence and mutation prevalences as the prior distribution. We present statistical methods for combining data from several published studies into our model, for incorporating uncertainty about the inputs, and for deriving predictive probabilities that asymptomatic individuals will develop breast and ovarian cancer later in life. Our software is being used successfully to provide individualized information to women that consider themselves at high risk of breast and ovarian cancer and are considering genetic testing. In a different arena, the software is serving as a preliminary surrogate for genetic testing in epidemiological investigations of large familial risk studies.

59

Different age of onset of cancer in HNPCC pedigrees with and without identified germ-line mutation.
Presciuttini S, Caroti-Ghelli C, Cama A, Genuardi M, Radice P, Viel A, Anti M, Bertario L, Casale V, Fornasarig M, Messerini L, Ponz de Leon M, Tonelli F, and Pierotti M. AIRC Special Project "Hereditary Colorectal Tumors", Italy

Germ-line mutations in either the hMSH2 or the hMLH1 gene, two key components of DNA mismatch repair, account for about 50% of families with Hereditary Non-Polyposis Colorectal Cancer (HNPCC). This syndrome is usually defined by the presence in a family of at least three cases of histologically proven colorectal cancers in at least two consecutive generations, with at least one case with onset under age 50 yr ("Amsterdam criteria").

As a part of the AIRC Special Project "Hereditary Colorectal Tumors", a collection of all known Italian families with IINPCC was recently published (*Tumori*: **82**: 151, 1996), and a database linking results of mutation analysis with the published pedigrees was established. We show here a comparison of the mean age-of-onset of colon cancer between two groups of families, those with identified pathogenic germline mutation vs. those in which search of the entire coding region of the two above genes was negative. In each family, we selected the three cases on which the ascertainment of HNPCC status was based. Mean age of onset was 42.3 (17 families, 51 cases) in the group with identified mutation, vs. 50.8 (19 families, 56 cases) in the group without mutation ($P < .001$).

We suggest that mutations at MSH2 and MLH1 loci cause a particularly severe phenotype, whereas mutations in other mismatch repair genes (or other genes conferring a high susceptibility to colorectal cancer) cause a significantly higher mean age of onset.

60

Data Management/Analysis Issues in Genome Screens. Rice JP. Washington University Medical School, St. Louis, MO USA.

We are part of the data management center for large-scale genome screens of alcoholism and bipolar disorder. We discuss data management issues in cleaning and processing genotypic data, map construction, and the creation of an interface for genetic analysis. Although, genome screens are in place for many common diseases, there appears to be a lack of available tools to facilitate the transition from laboratory data to analysis using recent programs such as SIBPAL, ASPEX or GENEHUNTER.

We have developed the GENEMASTER system for use both by the molecular genetics labs and have also created an interface to prepare analysis files using NETSCAPE. We present the difficulties and complications from the genetic analytic perspective. In addition, we discuss the problem of pooling data from multiple genome screens to permit meta-analysis from parallel, independent studies.

61

Heritability of Plasma Leptin in a Population Sample of African-American Families. Rotimi C¹, Luke A¹, Li Z¹, Compton J², Bowsher R², Cooper R¹. ¹Loyola University Medical Center, Department of Preventive Medicine and Epidemiology, Maywood, IL, USA and ²Eli Lilly and Company, Department of Drug

Disposition, Lilly Corporate Center, MC 0815, Indianapolis, IN, USA.

The aim of this study was to examine familial patterns of plasma leptin levels and the potential association with cardiovascular risk factors in a population sample of African-American families recruited from metropolitan Chicago. The study included 68 mothers, 31 fathers, 143 daughters and 119 sons, for a total of 361 individuals from 118 families. Leptin levels were adjusted for the effect of age separately for mothers, fathers, daughters and sons. Residuals were then standardized before estimating familial correlation using the maximum likelihood method available in SEGPATH. With the exception of height, plasma leptin level was strongly correlated with all measured anthropometric variables. Familial effect (ie., heritability) of leptin levels was estimated as 39% in this population at high risk for over weight. A significant sex difference was observed and most of the estimated familial effect may be attributed to genetic influences since the spouse correlation was not statistically different from zero. A strong nonshared individual environmental effect is also suggested, however.

62

Further refinements to the broad autism phenotype definition. Santangelo SL, Folstein SE. Tufts/New England Medical Center, Boston, Massachusetts, USA.

Previously, we reported an empirical definition of the *broad autism phenotype* (BAP) developed using data from the Baltimore Family Study of Autism. The BAP is comprised of characteristics, seen in family members of autistic individuals, that are less severe but qualitatively similar to those that define autism. Here, we present further refinements to our model of the BAP.

We developed a multiple logistic regression model of the BAP that included four traits: having few and poor quality friendships; having a high score on the sum of five personality attributes; having a low score on Performance IQ, and having a childhood history of reading or language difficulties. This model correctly classified 69% of the 180 parents of autistic probands and 80 parents of Down syndrome control probands. The distribution of scores from the logistic regression equation was then dichotomized to assign affection status. To maximize specificity, we chose a probability level of 0.85 (specificity=0.96) as the cutpoint by which to distinguish affected from unaffected parents. Anyone with a logistic regression equation score of 0.85 or greater was designated as affected with the BAP. This method identified 25% of autism parents and 4% of DS parents as affected. ROC analysis was then used to determine cutpoints for the individual continuous and ordinal measures comprising the BAP, in order to test whether the traits tended to assort randomly and whether any were

transmitted in Mendelian fashion. The segregation ratios of two of the four traits appear to conform to Mendelian expectations. The implications of these results for genetic studies will be discussed.

63

Prognostic significance of estimated BRCA1 and BRCA2 mutation status in women diagnosed with breast cancer. Schildkraut JM¹, Parmigiani G¹, Berry DA¹, and Iversen, Jr ES¹. ¹Duke University, Durham, North Carolina, USA.

Germline mutations at the BRCA1 and BRCA2 loci are known to confer an elevated risk of both breast and ovarian cancers, however the effect on survival after breast or ovarian cancer of possessing a mutation at one or the other marker is less well understood. Evidence exists of increased survival among ovarian cancer patients possessing mutations at BRCA1; we investigate the relationship between genetic status and survival after breast cancer. Follow-up information was obtained for 4474 breast cancer cases drawn from the Cancer and Steroid Hormone Study via record linkage with the Surveillance, Epidemiology, and End Results registry, with maximum follow-up approaching 157 months. Using a statistical model, we assign carrier probabilities to individuals given their history of breast and ovarian cancer and that of first- and second-degree relatives, as genetic information is unavailable. The effect of a germline mutation at BRCA1 or BRCA2 on survival is estimated treating genetic status as a latent variable; controlling for stage at diagnosis, histology, whether radiation treatment was administered, the individual's smoking history, body mass, and age at diagnosis; and assuming proportional hazards. In a preliminary analysis, mutations at BRCA1 were found to have a mild protective effect, while mutations at BRCA2 were found to have a moderate protective effect. Breast cancer patients with a BRCA1 mutation are estimated to have a hazard rate 0.96 times as large (95% interval estimate from 0.68 to 1.40) and patients with a BRCA2 mutation a hazard rate 0.69 times as large (95% interval estimate from 0.38 to 1.17) as the rate experienced by mutation-free patients. No evidence was found for an interaction between treatment of the cancer by radiation and genetic status.

64

Identification of a novel genetic linkage with bipolar affective disorder. Schofield PR¹, Adams LJ^{1,2}, Kwok JBJ¹, Salmon JA¹, Fielder SL¹, Rosso A¹, Reid A¹, Donald JA³ and Mitchell PB². ¹Garvan Institute of Medical Research, Sydney, Australia; ²School of Psychiatry, University of New South Wales & Mood Disorders Unit, Prince Henry Hospital, Sydney, Australia; ³School of Biological Sciences, Macquarie University, Sydney, Australia.

Bipolar affective disorder is characterised by severe mood swings (mania and depression) and affects 1-2% of the population. No predisposing genes have been identified although many groups are currently undertaking genome screens in an effort to localise susceptibility genes. We have completed an initial genome screen using 214 microsatellite markers on 35 individuals from the most powerful pedigree in our cohort. The data were analysed by two-point linkage analysis under several diagnostic models. Lod scores greater than 2 were obtained for 5 markers. 52 additional members of this pedigree were typed for these markers and a more intensive screen was undertaken in the regions surrounding these markers. Reduced two-point scores were obtained for all markers except those on one chromosome. Two markers in this region had lod scores >2 and multipoint analysis using additional markers gave a maximum lod score of $z=3.18$. These results suggest the presence of a susceptibility locus for bipolar disorder. Concurrently, we are undertaking a 10 cM genome screen of the twelve most informative families (250 individuals, including 59 affected members) in our pedigree cohort. Initially, we have analysed chromosomes for which evidence of linkage has been reported. We do not find any evidence of linkage to the reported chromosome 16 and 18 markers. However, we do find evidence supporting linkage to chromosome 21q22.3 markers D21S198 and PFKL when using the non parametric Affected Pedigree Member method of analysis, supporting a susceptibility locus on chromosome 21q.

65

Methods to identify genetic high- and moderate-risk families and assess gene-environment interaction in a population-based case-control study of breast cancer. Schmidt S^{1,2}, Schaid DJ², Chang-Claude J¹, Becher H¹. ¹Deutsches Krebsforschungszentrum, Heidelberg, Germany; ²Mayo Clinic/ Mayo Foundation, Rochester, MN, USA.

We present a two-stage method to identify genetic moderate- and high-risk families in a population-based case-control study of premenopausal breast cancer. In the first step, we calculate a family risk score for each family to distinguish families with a likely genetic disease component from those consistent with chance aggregation of breast cancer. A subset of the families with evidence for a genetic involvement may be due to one of the known high-penetrance genes for breast and ovarian cancer (BRCA1, BRCA2). Thus, in the second step, we use the MLINK software to compute carrier probabilities for case and control probands under different genetic models for those genes. To investigate the interaction of such high-risk genotypes with environmental risk factors in a logistic regression model, it is necessary to account for misclassification of carrier and noncarrier genotypes. This can be achieved by directly incorporating the carrier probabilities into the

likelihood for a case-control design, although we then fail to differentiate between moderate-risk genotypes with lower penetrance and nonsusceptible genotypes. Evidence for gene-environment interaction in our study data will be assessed, and issues of genotype misclassification will be discussed.

66

Familial aggregation of lung and breast cancer. Schwartz AG¹, Rothrock M¹, Kau T-Y², Weiss L². Allegheny University of the Health Sciences, Pittsburgh, PA, USA; ²Karmanos Cancer Institute, Detroit, MI, USA.

Studies of familial aggregation of cancer have provided important leads in understanding the genetics underlying cancer development. The current study evaluates whether first-degree relatives of young lung cancer cases are at increased risk of lung and other cancers. Families were identified through 187 lung cancer cases from the SEER population-based cancer registry in metropolitan Detroit, under 45 years of age, diagnosed between 1990 and 1997, and 281 controls frequency matched on age, race, and sex. Data were available for 1,293 relatives of cases and 1,715 relatives of controls. Cancer in relatives was evaluated after adjusting for age, race, sex, and smoking status of each family member. A positive family history of lung cancer at an early age increased lung cancer risk among first-degree relatives 2.3-fold (95% CI, 1.1-4.9). Relative risk estimates were 1.6 (95% CI, 1.2-2.1) for all cancers combined and 4.8 (95% CI, 1.9-12.2) for breast cancer. These findings of familial aggregation of lung and breast cancer suggest that common susceptibility genes may act to increase cancer risk in families.

67

Effects of map density, sibship size and computational approach on the power to detect quantitative loci. Shugart YY and Goldgar DE, Genetic Epidemiology, IARC, Lyon, France.

MIM (Goldgar 1990) and MAPMAKER/SIBS (Kruglak and Lander 1995) are two methods of using IBD sharing in a region spanned by multiple markers for mapping quantitative trait loci. Intuitively, MIM has two advantages over MAPMAKER/SIBS. 1) MIM uses information from complete sibships, while the latter breaks up a sibship into sets of sib-pairs, therefore, MIM avoids the problem of the dependence of sib-pairs; 2) MIM estimates IBD sharing over a region instead of a single point and thus performs less tests.

For the purpose of comparison, we simulated quantitative traits under various genetic models, and evaluated the power and false positive rate of both methods in a genome search. In addition, we examined the effects of marker density (5cM, 10cM and 20cM) and sibship size (2, 4, and 6). Preliminary results indicate that a 10cM map provides the optimal tradeoff between power and type I error and that the MIM method, in general, performs better than MAP MAPKER/SIBS.

68

Linkage analysis of diseases with variable age of onset. Siegmund KD, Todorov AA, and Rao DC. Washington University, St. Louis, MO, USA.

We present a multivariate survival model for the linkage analysis of the age of onset of a disease. The approach allows specific modeling of covariates and gene by environment interactions and is applicable to general pedigrees. We compare the power of this approach to that of the t_2 -statistic when only sib pairs are available. For the survival model, the likelihood is expressed as a function of the number of alleles shared ibd at a marker, the censored ages of onset and disease status.

In a preliminary analysis, we considered the case of a biallelic disease locus (disease allele frequency 0.20 and a genetic relative risk 10) under a dominant mode of inheritance. We also assumed that the disease locus lies exactly at a fully informative marker. Two sampling schemes were considered: families ascertained (1) randomly and (2) through affected sib pairs.

For samples of 400 affected sib pairs, the power of the t_2 -statistic and the likelihood ratio test from the correctly specified parametric approach are comparable at the 0.001 significance level (90.6% vs. 90.5%). The likelihood ratio test has slightly higher power when the phenotypes of the parents are included in the analysis.

69

Congenital anal atresias in 225,752 consecutive births. Stoll C, Alembik Y, Dott B, Roth M.P. Centre Hospitalo-Universitaire, Strasbourg, France.

Congenital anal atresias were studied in a small geographical area in 225,752 consecutive births. For each of the 108 new cases studied during the period 1979 to 1995, more than 50 factors were compared in probands and in controls. The prevalence rate of congenital anal atresias was 4.8 per 10,000. Sex ratio was 0.99. Prenatal diagnosis was performed in 14 cases and 7 cases were induced abortions. The more common types of associated malformations in the 86 affected cases (79.6%) with at least one anomaly other than anal atresia were renal agenesis, intestinal atresia, limb reduction defects, genital anomalies, esophageal atresia and ventricular septal defect. At birth infants with anal atresia and other malformations were smaller, weighted less and their head circumference was lower than in controls. Placental weight was also lower than in controls. Pregnancies with anal atresia were more often complicated by threatened abortion, oligoamnios and polyhydramnios. Mothers of children with congenital anal atresia took more often drugs during pregnancy than mothers of controls. Fathers of children with anal atresia were often exposed to occupational hazards than fathers of controls. There was a significant association between anal atresia and consanguinity of parents. The recurrence risk for first degree relatives of probands was 5.2%. First degree relatives of probands had more than twice the prevalence of non-anal atresia than controls. These results are of relevance to genetic counseling.

70

TDT and other association tests: an epidemiological view. Sun F¹, Yang Q², Khoury MJ². ¹Emory University School of Medicine, Atlanta, GA, USA; ²Centers for Disease Control and Prevention, Atlanta, GA, USA.

We give an epidemiological explanation of the transmission disequilibrium test (TDT) and other related association tests for a diallelic candidate locus with alleles "M" and "N" and show that TDT can be analyzed as a case-control design. The odds ratio for TDT can be defined as the ratio for the number of parents who transmitted allele "M" and did not transmit allele "N" to the cases with the number of parents who transmitted allele "N" and did not transmit allele "M". An epidemiological meaning is given for this odds ratio. We show that this odds ratio equals to the ratio of the odds of disease for individuals harboring allele "M" to the odds of disease for

individuals harboring allele "N", given one parent is heterozygous. Under the multiplicative effect of the "M" allele, the odds ratio equals to the risk of "MN" individuals compared to "NN" individuals. A new simple, yet efficient method for estimating the risks of "MN" and "MM" individuals compared to "NN" individuals related to case-parental control design is proposed. Simulations show that the new method gives more accurate estimations for the relative risks compared to Khoury's method (1994) and roughly the same accuracy as the maximum likelihood based method proposed by Schaid and Sommer (1993).

71

Incorporation of missing data into TDT. Szyda J, Bull SB. Samuel Lunenfeld Research Institute, University of Toronto, ON, Canada.

In the classical TDT analysis, a considerable amount of information coming from the affected individuals is abandoned because of incomplete family data (unknown parental genotype at a marker locus). Working with fully informative data only, i.e., with families having both parents and an affected child genotyped, is free of bias which can be introduced by missing data. However, by following this rule it becomes very difficult to collect data sets large enough to guarantee satisfactory power.

In this study a simulation of the null hypothesis distribution and power of three different TDT-based statistics is conducted. These statistics differ in the number of families with missing data that are rescued for the analysis. In addition, a jackknife TDT-based test is evaluated, in which the nuclear family is the resampling unit. For this statistic the jackknife procedure is used to estimate the variance of the difference between the off-diagonal cells in the marker allele transmission table (i.e. the variance of the numerator of TDT statistics). These tests are then applied to the data from a study of inflammatory bowel disease.

72

Restrictions on the components of genetic variance for epistatic models. Tiwari HK, Elston RC. Case Western Reserve University, Cleveland, OH, USA.

In recent years, there has been a focus on methods

for using linkage analysis to understand the etiology of complex disorders which do not follow simple Mendelian, single locus segregation. Some of these methods make use of components of genetic variance. After giving simple general formulations to derive all the components of total genetic variance for multilocus models, we investigate these components for a series of fifteen two-allele two-locus disease models with incomplete penetrance. We discuss the restrictions and limitations implied by the disease prevalence on both these components and the gene frequencies. Finally, we investigate the relative magnitudes of the components of variance for the various models when penetrance is complete. It is found that the epistatic components of variance are non-trivial, relative to the other components of variance, in seven of the models studied. These models may be of special interest for the study of complex diseases in which the recurrence among relatives beyond first degree decreases sharply.

73

Power of the affected sib-pair method in the presence of environmental factors. Todorov AA¹, Siegmund KD¹, Génin E², Rao DC¹. ¹Washington University School of Medicine, St. Louis; ²University of California, Berkeley.

We consider the situation where the penetrances of a locus are modulated by an environmental factor that acts either additively or multiplicatively upon the penetrances. Relatively little consideration has been paid so far to the fact that the population parameters (prevalences, relative risks) place strong constraints on the type of genetic models that are feasible given values of the population prevalence and sibling relative risk. Here, we provide a simple method to generate all possible models given these values and provide the equations for all the inflection points that may be found in the parameter space.

This method has two applications. First, it allows the determination of the range of sample sizes needed to detect linkage under a variety of conditions. Second, it can be used to guide segregation analyses with marker data when the population parameters are known with reasonable accuracy.

Suppose, for example, that the overall prevalence of the disorder is 0.05, the overall sibling relative risk is 3.0, and the environmental correlation is 0.20. 40 to 100 affected sibpairs suffice to give 90% power ($\alpha = 0.001$), if genetic and environmental factors act multiplicatively, the probability of exposure is between 0.10 and 0.20, and the prevalence of the disorder in the exposed population is less than four times that in the population. Generally, larger samples will be required when the prevalence in the exposed group is large.

74

Bivariate familial correlation analysis, by use of Estimating Equations. Application to the Insulin Resistance Syndrome. Tréguoët DA¹, Herbeth B², Juhan-Vague I³, Siest G², Ducimetière P¹, Tiret L¹. ¹INSERM U258, Paris; ²CMP Nancy; ³CHU Timone, Marseille, France

Familial correlation analysis involving more than one trait may give a better insight into the etiology of multifactorial syndromes than univariate analysis. Significant cross-trait correlations between biological relatives but not between spouses may suggest that the two traits share a common genetic basis. We propose an application of the Estimating Equations (EE) technique for estimating intra-trait and cross-trait familial correlations on two quantitative traits. Unlike maximum likelihood methods, the EE method does not require to specify the joint distribution of the traits. Estimation of correlations and of their variance involves an iterative three stage algorithm which converges very quickly. The generalized Wald test can be used to test any specific hypothesis of familial resemblance. This method has a great flexibility for handling covariates and incomplete family data.

This technique has been applied to a family study of metabolic factors involved in the Insulin Resistance Syndrome (IRS) [body mass index (BMI), triglycerides and insulin] carried out in a sample of 228 healthy nuclear families with ≥ 2 offsprings. The three factors exhibited similar patterns of familial resemblance, with correlations between biological relatives higher than between spouses. Despite similar insulin x BMI and insulin x TG intraindividual correlations, cross-trait resemblance between relatives was observed for insulin x BMI but not for insulin x triglycerides. These results indicate that, in the IRS, the clustering of insulin and BMI has a familial basis, whereas the clustering of insulin and triglycerides has another origin.

75

Association between IDDM age of onset and HLA class II genes: family based and cross ethnic study. Valdes AM¹, J Noble², H Erlich² and G Thomson¹. ¹Dept. of Integrative Biology, University of California, Berkeley, CA, USA; ²Roche Molecular Systems, Alameda, CA -Children's Hospital Oakland Research Institute, Oakland, CA, USA.

Clinical onset of insulin dependent diabetes mellitus (IDDM) is not confined to childhood and age-dependent HLA heterogeneity has been observed in Caucasian and Asian IDDM patients. We have analyzed 222 IDDM Caucasian sib pairs typed for DRB1, DQA1, DQB1 and DPB1 (from the Human Biological Data Interchange catalog) and 664 IDDM patients typed for DRB1 and DQB1 from 8 ethnic groups (from the 12th International Histocompatibility Workshop and Conference). The family based study revealed a strong association in age of onset among sib pairs that share at least one haplotype. We compared the observed age of onset distribution among sib pairs sharing both haplotypes and the

expected distribution under the assumption that DR-DQ account for all the age association between sib pairs. We were able to reject that hypothesis in DR3-DR4 and DR4-DR4 sib pairs, indicating that other genetic, environmental and/or stochastic factors have an influence in age of onset association. To this regard, significantly different age of onset distributions were found among DRB1*0301-DQB1*02 / DRB1*0401-DQB1*0302 genotypes depending on the DPB1 genotype present. All the ethnic groups studied showed a decrease in frequency with increasing age of onset of DR3-DR4 genotypes carrying the DRB1*0405 subtype. For other DR4 subtypes this pattern was less clear. On average the relative risk of DRB1*0301/DRB1*0405 genotypes was 4.1 times higher if the age of clinical onset was under 10 years than after age of onset 10. For DRB1*0301-DQB1*02 / DRB1*0401-DQB1*0302 this ratio was 1.1, and 0.98 for DRB1*0301 / DRB1*0404 genotypes. Further, we observed significant differences in several ethnic groups in the age distributions of the various DR4 alleles. Our findings emphasize the complex genetic nature of IDDM and of its clinical onset.

76

Predictive power of individual genetic and environmental factor scores.

VLIETINCK R¹, THOMIS M¹, MAES H², BLIMKIE C.J.³, VAN LEEMPUTTE M¹, CLAESSENS A. L.¹, MARCHAL G.¹, WILLEMS E.¹ and BEUNEN G.¹

¹University of Leuven, Belgium; ²Virginia Commonwealth University Virginia; ³McMaster University, Hamilton, Ontario.

This study validates the use of an individual's genetic (IGFS) and environmental factor score (IEFS), estimated by genetic model fitting on 25 MZ and 16 DZ twin pairs. The question was whether baseline IGFS and IEFS could predict the gains after a 10-week strength training program, and whether IGFS followed genetic rules. As postulated the IGFS correlated highly (.67-.83) with strength before and after 10 weeks training while IEFS did not. Individuals with a high IGFS gained less strength than individuals with a low IGFS, whereas IEFS had less differential power. Individuals with a low IGFS, had a relative risk of 4.1 of having a low baseline strength (CI95: 1.9 to 8.79), and those with a high IGFS of having a high baseline strength (RR=7.0 CI95: 3.1 to 15.76). High or low IGFS at baseline also predicted (RR=6.25) the strength gain after training. The relative risks to predict the baseline strength or strength gains after training based on the IEFS didn't differ from one. There was no IGFS-IEFS interaction within individuals. The predictive value of IGFS could be useful in identifying susceptibility to environmental stress in a variety of multifactorial traits, e.g., diseases and impairments, and in sib-pair selection for QTL.

77

Continuous variable TDTs using regression analytic methods. WALDMAN ID¹, MILLER MB², ROBINSON BF¹, and ROWE DC³. ¹Emory University, Atlanta, GA, USA; ²Washington University School of Medicine, St. Louis, MO, USA; ³University of Arizona, Tucson, AZ, USA.

The Transmission Disequilibrium Test (TDT) is a simple and effective test of linkage disequilibrium between a candidate gene and a disorder. In contrast to other between- and within-family tests of association, the TDT provides an unbiased test of association and linkage in the presence of population stratification, such as that due to ethnic heterogeneity. In comparison with standard linkage analyses, the TDT is considerably more powerful, thus requiring far fewer subjects to detect a true genetic effect. Despite these advantages, a drawback of the TDT has been its application solely to traits that are categorical, such as the presence or absence of a diagnosis. In this study we propose ordinary least squares and logistic regression-based extensions of the TDT to examine the relation between a candidate gene and one or more continuous variables. We use data on disruptive behavior disorder symptoms and the dopamine transporter gene (DAT1) from 90 probands and their parents to illustrate various issues one can address with these methods. These include examining differences in linkage disequilibrium as a function of symptom severity, age of onset, or comorbidity; using association with candidate genes as a way of choosing an optimal diagnostic threshold; and examining gene X environment interactions. In a sample of probands with Attention Deficit Hyperactivity Disorder (ADHD), DAT1 was related curvilinearly to the level of hyperactive-impulsive, but not inattentive, symptoms. The number of Oppositional Defiant Disorder (ODD) symptoms also was related linearly to DAT1, suggesting that the linkage disequilibrium of DAT1 with ADHD is stronger in the presence of ODD. We also discuss limitations of our method, as well as alternative methods that are more appropriate for examining the effects of multiple genes on a single continuous trait.

78

Linkage analysis of smoking behavior and IgG2 levels in early onset periodontitis. Wang Y-F¹, Wang S¹, Sun C¹, Gillanders E¹, Freas-Lutz D¹, Schenkein HA², Diehl SR¹, NIDR, NIH, Bethesda, MD, USA; ²CRCPD, VCU, Richmond, VA, USA.

Segregation analyses suggest that an autosomal codominant major gene may best explain variation in total serum IgG2 with a significant covariate effect attributable to smoking in a collection of early onset periodontitis families. Cigarette smoking is known to relate to periodontitis and IgG2 levels and smoking behavior itself has been shown to be about 50% heritable. Our research is aimed at elucidating the relationships among early onset periodontitis, cigarette smoking and IgG2 levels. We conducted a genome search using 46 African American and 16

Caucasian multiplex families. Smoking behavior was assessed by serum cotinine and total serum IgG2 levels were also measured for our study results. Several linkage analysis approaches such as affected sibpair analysis and regression linkage analysis using SIBPAL program, non-parametric LOD score tests using Genehunter program, and the transmission disequilibrium tests of linkage and association using SIBPAIR program were applied to analyze the data. We used both transformed quantitative measurements and binary classifications of these variables, and sometimes obtained quite different results. Our results suggest that genes on chromosome 1, 9, 10, 13, 14 and 20 may influence smoking behavior and genes on chromosome 3, 5 and 11 may influence levels of total serum IgG2. Our findings should be interpreted with caution until confirmed with flanking DNA markers and additional families. Supported by DE10703.

79

Juvenile Myoclonic Epilepsy with absence linked to chromosome 1p. Weissbecker KA^{1,2}, Westling BW², Serratosa JM³, Jara-Prado A⁴, Alonso ME⁴, Cordova S⁴, Medina MT⁵, Gee M³, Iranmanesh R³, Delgado-Escueta AV³. ¹Tulane Univ. Med. Ctr., New Orleans, LA; ²LSU Med. Ctr., New Orleans, LA; ³Comprehensive Epilepsy Program UCLA & VA Med. Ctrs., Los Angeles, CA; ⁴Nat'l Inst. of Neuro. and Neurosurg., Mexico; and ⁵Nat'l Autonomous Univ. of Honduras.

Sib pair and lod score analyses of 286 microsatellite markers throughout the genome were used to identify loci inferring a susceptibility to Juvenile Myoclonic Epilepsy (JME) with absence in a large Mexican family ascertained through a proband with JME and pyknoleptic absence. Other generalized seizure types seen in six relatives include Grand Mal (GM) only, GM with Childhood Absence (CAE), Juvenile Absence Epilepsy, JME only, and CAE only. There are also two sibling cousins with partial seizures whose affected status varies in the analysis. Sib pair analysis identified 13 markers with significant evidence for linkage ($p < 0.01$), 5 of which are clustered on chromosome 1p. Lod score analysis was done under 6 autosomal dominant models in which the gene frequency, penetrance, and the affected status of the cousins with partial epilepsy were varied. The maximum lod score among all markers was 3.8 ($\theta = 0.0$) for DIS207 under a model with $q = 0.001$ and penetrance of 0.70. (DIS207 gave the highest lod score under all models tested.) The five 1p markers identified in sib pair analysis and 1 additional 1p marker produced lod scores greater than 1.3 under this model. These 6 markers are within a 8.54 Mb region. Multipoint indicated that the most likely location for the susceptibility gene at marker DIS207.

80

Comparison of Variance-Components and Sib-pair Methods for Quantitative Trait Linkage Analysis in Sibships and Nuclear Families. Williams JT, Blangero J. Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, TX, USA.

We compared the performance of variance-component and sibpair-based methods for multipoint linkage analysis of quantitative traits in sibships and in nuclear families. Variance-components linkage analysis was performed using the program SOLAR (Blangero and Almasy 1996), and multipoint sibpair analysis was performed using the program MAPMAKER/SIBS (Kruglyak and Lander 1995). As a benchmark dataset we used the simulated nuclear family data from *Genetics Analysis Workshop 10*. These data comprise 200 replicates of 239 randomly ascertained nuclear families, and each replicate consists of 1164 individuals for whom complete phenotypic and genotypic information is available.

SOLAR and MAPMAKER/SIBS were each applied to the problem of detecting linkage of quantitative trait Q1 to major gene MG1 on GAW chromosome 5. Linkage analysis was undertaken at 2 cM intervals in all 200 replicates of all 10 GAW chromosomes in both the original GAW nuclear families and in the corresponding sibships. With SOLAR, covariate effects due to sex, sex-specific age, and the environmental factor EF were incorporated as fixed effects in the model for the theoretical trait mean. With MAPMAKER/SIBS the effect of these covariates was removed by correcting trait Q1 using the regression coefficients returned by the polygenic model of SOLAR.

The statistical properties of unbiasedness, Type I error rate, power, and efficiency were estimated empirically for each method and used as a basis for comparison. With both nuclear families and sibships, the variance-components and sibpair methods exhibited comparable performance in terms of their Type I error rate and the unbiasedness of the QTL location estimate. However, the variance-components method provided markedly superior performance in terms of the power to detect linkage and in the relative efficiency of the location estimate.

81

Bias and efficiency in case-control studies of measured genes and gene-environment interactions: basic family designs. Witte JS¹, Elston RC¹, Gauderman WJ², Thomas DC². 1. Case Western Reserve University, Cleveland, Ohio USA; 2. University of Southern California, Los Angeles, USA.

We present here a comparison of case-control designs that use population controls versus those that use controls selected from their relatives (i.e., siblings, cousins, or "pseudosibs" based on parental alleles) for estimating the effect of measured genes and gene-environment interactions. Our results indicate that, as expected, when population stratification exists, using population controls can give biased estimates of effect.

We quantify the extent of this bias. Furthermore, using pseudosibs can give biased effect estimates if one uses the conventional conditional likelihood, though the bias is towards the null and disappears with disease rarity. We show how to correct this likelihood to give unbiased effect estimates. Using sibling or cousin controls give unbiased estimates of effect. The designs using population or pseudosib controls, however, are generally the most efficient for estimating the main effect of a measured gene, followed in efficiency by the design using cousins. Nevertheless, when looking at gene-environment interactions, the design using sibling controls can be quite efficient.

82

Genetic component in disability trait. Yashin AI^{1,2}, Iachine IA³, Christansen K³, Holm N³, Vaupel JW^{1,2}. ¹Max Planck Institute for Demographic Research, Rostock, Germany; ²Duke University NC USA; ³Odense University, Denmark.

In this paper the presence of familial and genetic effects in disability is analyzed. For this purpose the data collected in the first wave of 1995 Longitudinal Study of Aging in Danish Twins (LSADT) older than 75 are used. The data are represented by discrete scores in five disability scales describing activity of daily living for male and female Danish twins.

The multi-threshold liability models of disability with age-dependent thresholds are developed and used in the analysis of this data. The presence of familial effects is revealed in all five disability scales for females and in the three scales of such data for males. Genetic effects are found significant in two scales of disability data: one for males and another for females.

Does this result mean that the age at onset of disability measured in appropriate scale is heritable? Not necessarily. In this paper we show that the presence of genetic effects in prevalence data may be induced either by genetic influence on transition from "active" to "disabled" or on transition from "disabled" to "dead".

To identify the "place" of genetic influence the health history model for dependent individuals is suggested and tested on simulated data. The model extends the idea of correlated frailty used in the genetic analysis of bivariate survival data. The analysis of LSADT data with this model shows the presence of substantial genetic influence on transition to disability. The genetic effects on mortality transition conditioned on disability state are found insignificant in these scales.

83

Is self-reported family history of colorectal cancer accurate in adenomatous polyp and colorectal cancer families? Zauber AG⁽¹⁾, Winawer SJ⁽¹⁾, Bishop DT⁽²⁾, Memorial Sloan-Kettering Cancer Center, NY, NY, USA⁽¹⁾, Imperial Cancer Research Fund, Leeds, England⁽²⁾.

Introduction. Family history of colorectal cancer is an important tool to identify families at higher risk for screening and surveillance and to manage surveillance of patients with adenomatous polyps. The accuracy of colorectal cancer diagnosis as reported by a patient for a family member was assessed by comparing the report of colorectal cancer death in a family member with the cause of death as stated on the death certificate.

Methods. Cancer status, site, and age at diagnosis as well as vital status and cause of death were obtained for all first degree relatives of adenomatous polyp patients and of colorectal cancer patients ascertained for the genetic epidemiology project of the National Polyp Study. Death certificates were requested for all patients and relatives with cancer stated as the cause of death. Each cancer death was classified by whether colorectal cancer was stated on the death certificate and by whether the patient (adenomatous polyp patient or colorectal cancer) reported colorectal cancer for the family member.

Results. Death certificates were obtained for 400 first degree relatives of adenomatous polyp patients. Colorectal cancer was listed on the death certificate for 86 first degree relatives; 57 (78%) of these colorectal cancer deaths had been reported as a colorectal cancer death in the relative. No colorectal cancer was reported on the death certificates for 314 first degree relatives; 298 (91%) had not been reported as colorectal cancer deaths by the patient. The patient report for the family member and the death certificate for colorectal cancer were concordant for 298 (89%) first degree relatives.

Conclusion. Colorectal cancer of first degree relatives is reported with good accuracy (89%) by patients. Familial reports of colorectal cancer can be used as a first level of classification of patients by familial risk.

84

Gene-environment interaction and oral clefts. Zeiger J¹, Beaty T¹, Hetmanski J¹, McIntosh, I¹ Hwang S-J². Johns Hopkins University, Baltimore, MD, USA; ²University of Texas, MD Anderson, Houston, Texas, USA.

Oral clefting is a common birth defect with a complex etiology, where both genetic and environmental factors influence risk. This case-control analysis examines the effects on oral clefts of maternal smoking and infant genotype at the *Tag1* site for the transforming growth factor alpha (TGFA) locus using data from two published studies (Hwang et al., 1995; Beaty et al., 1997).

Caucasians born in Maryland were included. Infants in the case group were born from 1984 to 1996 with cleft

palate only (CP, n=111) or cleft lip ± palate (CLP, n=181). Two control groups were used: infants with isolated birth defects other than a cleft born from 1984 to 1992 (n=281) and healthy infants born from 1992 to 1996 (n=85).

Logistic regression analyses were performed to estimate odds ratios (OR) and their 95% confidence intervals (CI) adjusted for maternal age. The risk for CP was significantly increased among infants carrying the C2 allele whose mother reported smoking when compared to both the birth defect controls (OR=3.64; 95%CI=1.50, 8.80) and the healthy infant controls (OR=6.42; 95%CI=1.40, 29.7). In addition, there was an increased risk for CLP when compared to the healthy controls if the mother reported smoking during pregnancy (OR=1.99; 95%CI=0.93, 4.23) regardless of TGFA genotype.

These findings suggest that maternal smoking during pregnancy may be a risk factor for cleft lip ± palate independent of genotype, but it may modify effects of certain genotypes in determining risk for cleft palate only.

85

Sample size calculations for linkage analysis to obesity using extreme sib pairs. Ziegler A¹, Schafer H¹, Hebebrand J². Clinical Research Group, Philipps-University of Marburg, Germany; ¹Institute of Medical Biometry; ²Department of Child and Adolescent Psychiatry.

There is a growing interest in finding genetic loci linked to human obesity. However, the complexity of the phenotype body weight makes this task challenging. Segregation analyses indicate Mendelian inheritance for a recessive major gene with an additional polygenic component and environmental correlation. One approach to establish linkage is based on allele sharing methods for sib pairs. Recently, the use of extreme sib pairs (ESP) has been proposed to increase power for mapping quantitative traits in humans. Sample sizes required to detect linkage for a major gene linked to obesity are calculated based on parameter estimates from published segregation analyses using the ESP approach and its extensions. Furthermore, we calculate the expected proportion of ESP.

Our results indicate that the sample sizes calculated for the models of the different segregation analyses are not difficult to achieve. They suggest that ESP concordant for high trait values should be the design of choice for genetic mapping to human obesity. Upon use of specific ascertainment strategies attempts to localize genes linked to human obesity appear quite realistic.

86

Further indication for linkage of obesity to the *Obese* gene. Ziegler A¹, Roth H^{2,3}, Hinney A², Grzeschik K-H³ and Hebebrand J². Clinical Research Group, Philipps-University of Marburg, Germany;

¹Institute of Medical Biometry; ²Department of Child and Adolescent Psychiatry; ³Institute of Human Genetics.

The human homologue of the mouse *obese* (*OB*) gene has been cloned, but its involvement in human obesity remains unclear. Indication for linkage of the *obese* gene to extreme obesity has been found, but data with no linkage have also been reported. To further understand the participation of the gene in human obesity we investigated linkage and/or association between markers flanking the human *OB* gene (D7S504 and D7S1875) and one marker located within *OB* (an A-G polymorphism at nucleotide 19 of exon 1) and extreme obesity in a sample of German children and adolescents. The analysis of 88 trios (index probands and both parents) for transmission disequilibrium of a haplotype which has previously been determined to be linked to extreme obesity (Reed et al. 1996) revealed a p-value of 0.054 using the one-sided exact binomial TDT. Post hoc analyses revealed p-values of 0.04 for the 214 bp allele of D7S1875 and 0.05 for the haplotype based on the A-allele of the intragenic polymorphism and the 214 bp allele of D7S1875.

Taken together, our findings support the evidence for linkage of extreme obesity to *OB*.

References

Reed DR, et al. (1996): Diabetes 45, 691-694.

87

Using Minimum Distance Estimation to estimate the recombination fraction and allele frequency in robust sib pair analyses for quantitative traits. Ziegler A¹, Kastner C²; ¹Institute of Medical Biometry, Philipps-University of Marburg, Germany; ²Institute of Statistics, Ludwig-Maximilians-University of Munich, Germany.

Olson and Wijsman (1993) have proposed a robust linkage analysis between quantitative traits and marker loci using all relative pairs. Their approach can be used

to test the recombination fraction θ on the squared differences of phenotypes after fitting Generalized Estimating Equations (GEE). However, neither θ nor the allele frequency p can be estimated. Similarly, estimates for dominance and additive variances cannot be obtained.

Fulker and Cardon (1994) proposed a regression approach to estimate θ from at least two flanking markers loci. We show that θ can be estimated by a two-step procedure using only one marker locus.

The first step involves solving GEE. In the second step minimum distance estimation is applied for estimating θ and p . The estimation is illustrated by simulations.

References

- Olson J, Wijsman EM (1993). Genet Epidemiol, 10, 87-102.
Fulker DW, Cardon LR (1994). Am J Hum Gen, 54, 1092-1103; 55, 419.