51

Power and Sample Size Consideration for detecting epistasis

H.K. Tiwari[1], V. George[2], N.J. Schork[1,3,4]
[1]Case Western Reserve Univ., Cleveland, USA , [2]The Medical College of Wisconsin, Milwaukee, USA, [3]The Jackson Laboratory, Bar Harbor, USA, [4]Harvard University School of Public Health, Boston, USA

In recent years, there has been a focus on to understand the etiology of complex diseases which do not follow simple Mendelian, single-locus segregation. Complex diseases are assumed to be the result of more than one locus and/ or environmental factors and are thought to be influenced by many intermediate quantitative traits. Quantitative traits have been studied extensively by plant and animal geneticists. With the advent of new tools and methods, comprehensive approaches to identify candidate genes underlying quantitative traits for humans are available. Testing the contribution of candidate genes to quantitative trait variation will become commonplace as more and more genes are identified. One of the methods that shows great promise in the analysis of candidate genes involves “Variance Components” models originally introduced by Fisher. Fisher partitioned the total genetic variance into additive part resulting from additive effects or main effects of the genes, a part resulting from dominant effects (allelic interactions) of genes, and a part resulting from epistatic effects (non-allelic interactions) of genes. Cockerham further partitioned epistatic variance into additive x additive, additive x dominant, dominant x dominant components of variance. Tiwari and Elston have shown that epistatic components of variance explain the significant part of the total genetic variance in certain two locus models pertinent to human genetics models. In this paper, we investigate the sample size required to achieve pre-specified power to detect epistatic components of variance for the two-locus models. This investigation includes analytical as well as simulation techniques for the development of the proposed method and illustration through examples.

52

Sibling Controls and a General Score Statistic to Detect Associations of Multi-allelic Genetic Markers with Disease

Daniel J. Schaid and Charles Rowland
Mayo Clinic/Mayo Foundation, Rochester MN

Detecting the association of genetic markers with complex diseases can be a critical first step to identify the genetic basis of disease. Misleading associations can be avoided by choosing as controls the parents of diseased cases, but the availability of parents often limits this design to early-onset disease. Alternatively, sib-controls offer a valid design. For a genetic marker with two alleles, Spielman and Ewens (1998) proposed a statistic that they called the *sib-TDT* (*S-TDT*), which is based on using all affected and unaffected siblings within a sibship. We present an extension of the *sib-TDT* - a general multivariate score statistic - to allow for multiple alleles at the marker locus. An advantage of our approach is that it allows a framework to assess associations using either parents as controls, sibs as controls, or even unrelated controls whose genotype frequencies do not fit Hardy-Weinberg proportions, or pooling any combination of these different designs. Methods to compute sample size and power are presented, allowing for varying sibship sizes, ascertainment criteria, and genetic models of risk. The power of sib-controls can be increased by either increasing the number of affected sibs per sibship, or increasing the number of unaffected control sibs. The sample size results indicate that using sib-controls to test for associations, by either a single marker locus or a genome-wide screen, will be feasible for markers that have a dominant effect, and for common alleles having a recessive effect. The results presented will be useful for investigators planning studies using sibs as controls.

53

Quantitative Characterization of Thought Disorder as a Schizophrenia Related Phenotype.

N.R. Mendell(1), P.S. Holzman(2), S. Matthysse(2), and D.L. Levy(2).
State University of New York at Stony Brook, USA, (2) McLean Hospital, Harvard Medical School, USA.

The low recurrence risk for schizophrenia related disorders in first- degree relatives has limited the power of linkage studies to the clinical phenotype, even one that is broadly defined. Other physiological and cognitive-perceptual traits seem to occur at a higher rate in biological relatives than the clinical symptoms. We report the results of a discriminant analiysis of the measures used in the assignment of the clinical diagnosis of thought disorder. The effect size on comparing the mean discriminant score of a sample of siblings of schizophrenics to a sample of sib- lings of controls is estimated to equal 1.4. This translates into an estimated genotype effect size of 2.8 and an estimated heritability of 0.4.

54

Comorbidity: implications for finding disease genes

Kathryn L. Lunetta[1,2] and Jordan Smoller[2,3]

[1]Dana Farber Cancer Institute, [2]Harvard School of Public Health and [3]Psychiatric Genetics Program, Masachusetts General Hospital, Boston, USA

In the context of complex diseases, an apparant linkage to or association with the study phenotype may in fact be attributable to a correlated phenotype. We examine this issue using an example from psychiatric genetics.

The lifetime comorbidity between social phobia and alcohol abuse ishigh. Could a study of social phobics using a realistic sample sizedetect linkage or association related to a comorbid phenotype such as alcoholism, even if social phobia does not have a genetic basis? Using a genotype relative risk (GRR) model (Risch & Merikangas 1996), we have calculated the number of parent-affected offspring trios forwhich the TDT would detect an association for a number of models. The probability of detecting association depends on the GRR for the genetic trait, the degree of comorbidity between genetic trait and study trait, the frequency of the associated allele, and the rate of disease in non-carriers. We assume two-sided $\alpha = .05$, frequency ofmarker and disease allele = p_D, and maximum linkage disequilibrium. For example, if the GRR=4 for a locus influencing alcoholism, and the risk of alcoholism among social phobics is .5, we would have 80% power to detect an association using the TDT with fewer than 150 social phobic trios if p_D=.1, and with fewer than 120 if p_D=.5, even if social phobia were not influenced by the candidate locus.

We discuss methods for identifying spurious associations due to comorbidity and situations where correlated phenotypes do not functionas confounders to the study phenotype.

55

Comparison of Multivariate Linkage Methods.

CI. Amos[1], M deAndrade, Y Wan.
Department of Epidemiology, UT-MD Anderson Cancer Center, Box 189, 1515 Holcombe Blvd, Houston, Texas, USA

In this study we compare the power of multivariate linkage methods to detect genetic factors. Genetic studies typically collect information for multiple correlated traits and power to detect linkages may be improved if multivariate observations are used in the analysis. We previously developed a multivariate extension of the Haseman-Elston (H-E) test to permit inclusion of data from multiple phenotypes. Although this procedure can be expected to provide a valid test for multivariate linkage, the efficiency of the method is unclear since it entails multivariate regression using markedly skewed distributions. Therefore, we have been comparing this procedure with multivariate variance components (VC) methods. Comparative results in a typical simulation gave power of the univariate and bivariate H-E procedures of 22.7% and 62.4% while the univariate and bivariate VC procedures had powers of 48% and 81% respectively. Although the VC procedure had higher power than H-E tests, the bivariate procedure is much more computationally demanding than the H-E procedure. Therefore, for completing rapid screening of quantitative traits, the multivariate H-E tests shows considerable promise.

56

Association versus linkage studies: a comparison of the sib-TDT and non-parametric linkage analysis in rheumatoid arthritis.

A. Cox[1], N.J. Camp[1], F.S. diGiovine[1], A.J. Silman[2], W.E.R. Ollier[2], J. Worthington[2], and G.W. Duff[1].
[1]University of Sheffield, Sheffield, UK, [2]ARC-ERU, University of Manchester, Manchester, UK.

The cytokine interleukin-1 (IL-1) plays a prominent role in inflammation, and genetic markers of the IL-1 locus have been implicated in a variety of autoimmune and inflammatory diseases. We have investigated the role of IL-1 in rheumatoid arthritis (RA) by genotyping a panel of 5 multiallelic markers of the IL-1 gene cluster in 187 families from the Arthritis and Rheumatism Council for Research National Repository. These data were analysed in 2 ways. First, by use of the non-parametric linkage analysis option of "Genehunter", and second, by combined use of the transmission disequilibrium test (TDT) and recently reported sib-TDT, both of which test for linkage in the presence of association, (the latter in families where parental information is unavailable). There was no evidence of linkage using "Genehunter" in either the overall dataset, or after stratifying the families according to their IBD sharing status at HLA-DRB1. When the data were analysed using the combined TDT and sib-TDT, however, there was some evidence of linkage in the overall dataset and stonger evidence in those families in which affected sibs shared 1 or 0 alleles IBD at HLA-DRB1. The most common alleles at three closely linked markers gave $z_{combined}$ scores of 2.32, 2.33 and 3.07 (p (uncorrected) < 0.02, 0.02 and 0.003 respectively). These results provide preliminary evidence of a weak effect of IL-1 in non-HLA-linked families. This effect is only detected using the more sensitive TDT and sib-TDT, which rely on the presence of association, and demonstrate the utility of the sib-TDT in a late onset disease where parental data are limited.

57

Paraoxonase genotype predicts *in vitro* susceptibility of low density lipoproteins to oxidation.

G.P. Jarvik, A. Chait, T. Hatsukami, and C. Furlong.
University of Washington, Seattle

Traditional lipid-related risk factors do not account for all of the estimated genetic variance in vascular disease. Paraoxonase (PON) is a high density lipoprotein (HDL)-associated enzyme that may have a role in the detoxification of atherogenic oxidized low density lipoproteins (LDL). Inactivation of paraoxonase has been shown to eliminate this capacity of HDL *in vitro*.

A common PON polymorphism has 2 alternative alleles, coding for either a GLN or ARG at amino acid 192. The latter R allele has been associated with an increased risk of vascular disease. Its frequency in Caucasians is 0.31; thus, an estimated 52% of individuals carry the allele (genotypes QR or RR) of which 10% are genotype RR.

We investigated whether variation in LDL oxidation was associated with PON genotype. The time taken to oxidize LDL *in vitro* was measured photometrically, with a shorter lag-time indicating increased susceptibility to oxidation. PON genotype was determined from 2-dimensional enzyme activity plots. The sample of 72 military veterans, aged 52-88 yr. (mean 68), was 94% male and 87.5% Caucasian.

PON genotype was predictive of LDL oxidative susceptibility (p=0.027 by ANOVA). The cardiovascular risk allele R was associated with an increased LDL susceptibility to oxidation as measured by a shorter lag-time. Alternatively, the Q allele appeared to have a dominant effect to increase lag-time. Lag-times for the QQ, QR, and RR genotypes were 54.2, 54.2 and 46.3 minutes, with n= 35, 27, and 10 respectively. A similar result was found in the more homogeneous subsample of 59 Caucasian males (p=0.025).

By measuring the intervening phenotype of LDL susceptibility to oxidation, we are able to demonstrate a possible mechanism of the increased cardiovascular risk associated with the relatively common PON R allele.

58

An approach to the analysis of combined traits.

Yu.S. Aulchenko and T. I. Axenovich
Institute of Cytology and Genetics, Novosibirsk, Russia

Genetic analysis usually operates with individual traits, i.e. traits which characterise particular individual. However, there are traits which characterise a group of individuals. Some reproductive features of a livestock, for example, litter size and sex ratio, represent a simplest type of such "combined traits". In general, these traits might be controlled by genotypes of both parents. In the framework of genetic analysis this will lead to modification of penetrance function. Instead of function $f(x|g)$ for probability of individual's phenotype x given that its genotype is g the function $f(y|g_f g_m)$ should be used. Here y is combined phenotype, g_f and g_m are the genotypes of mating pair.

A variant of this model was used for segregation analysis of litter size in a large pedigree of house musk shrew (*Suncus murinus*). The analysis was performed under a mixed model of inheritance. Single autosomal major-gene was assumed. Both majorgenic and polygenic components were necessary for correct description of the litter size inheritance, inasmuch as the exclusion of any of them led to significant drop of likelihood. The Elston-Stewart's test also provided evidences for a major gene incorporated into the control of litter size inheritance.

The "combined trait" models fit the data significantly better then the models which assumed that litter size is an individual trait controlled by maternal genotype.

59

Non-Parametric Linkage Analysis of Three Candidate Genes to Type 2 Diabetes and Obesity in the Old Order Amish

Hsueh W-C (1), Beaty TH (1), Mitchell BD (2), Shuldiner AR (3).
(1) Johns Hopkins University, Baltimore, MD; (2) Southwest Foundation for Biomedical Research, San Antonio, TX; (3) University of Maryland, Baltimore, MD, USA

The potential involvement of 3 candidate genes in type 2 diabetes and obesity, the Ala54Thr variant of fatty binding acid protein 2 (FABP2), the Gly972Arg variant of insulin receptor substrate-1 (IRS-1), and the Gln27Glu variant of the beta-2 adrenergic receptor (B2AR) was evaluated using affected-sibpair (ASP) analysis and the transmission/ disequilibrium test (TDT) in the Old Order Amish, a well-defined genetically homogeneous Caucasian founder population with very large sibships, a high degree of consanguinity, well-documented genealogies and a shared environment. A total of 915 subjects were included in the analysis; all fit into a single 11-generation pedigree. Using S.A.G.E. 3.0 software, we tested for linkage and/or association with type 2 diabetes (DM), obesity (BMI>=27), central adiposity (defined as having a WHR >= 0.85 in females, WHR >= 0.94 in males), and hypertension (HTN, defined as having a systolic blood pressure >= 140mmHg).

In ASP analysis, the B2AR showed evidence for linkage to BMI with a p-value of 0.004 (384 affected

sib-pairs). In TDT analysis, there was marginal evidence for excess transmission of the Gln/Glu genotype in obese subjects (p = 0.046), although there was no evidence for excess transmission of any particular allele (p = 0.24). The ASP test also provided evidence for linkage of both the B2AR and IRS-1 genes to HTN (p = 0.008 and 0.018, respectively, with 65 affected sib-pairs).

These results indicate suggestive linkage of the B2AR gene to BMI and HTN, and that of the IRS-1 gene to HTN. Further analyses, including evaluation of adjacent DNA markers, affected pedigree-member analysis, and replication of linkages in an independent Amish collection will aid to delineate further the roles of these candidate genes in diabetes and obesity.

60

The delayed lipromin reaction (Mitsuda test) is linked to the human NRAMP1 gene in large Vietnamese sibships

A. Alcaïs[1], F.O. Sanchez[2], N.V. Thuc[3], V.D. Lap[3], P.H. Lagrange[4], E. Schurr[2], and L. Abel[1].
[1] INSERM U.436, Paris, France, [2] Montreal General Hosp. Research Inst., Montreal, Canada, [3] Dermatology Hosp., Ho Chi Minh City, Vietnam, [4] Hop. St Louis, Paris, France

The Mitsuda reaction is considered to have a high prognostic value in leprosy since the immunological responsiveness of the host matches the clinical spectrum of the disease. In particular, a positive reaction indicates resistance to the most contagious lepromatous form. The human NRAMP1 gene is a candidate locus for the control of Mitsuda reaction since 1) its murine homologue Nramp1 controls innate resistance to Mycobacterium lepraemurium 2) it has been linked to leprosy per se.

A sib-pair linkage study was performed in 20 leprosy nuclear families (number of sibs ranging from 2 to 12) from Ho Chi Minh City, Vietnam. Mitsuda reaction was measured by experienced leprologists as classically described. Family subjects were genotyped for several intragenic and flanking NRAMP1 markers, leading to the definition of a fully informative NRAMP1 haplotype.

When considering the Mitsuda reaction as a quantitative trait, significant linkage was obtained with the two sib-pair methods that were used: 1) the classical Haseman Elston ($p<0.02$), 2) the Maximum Likelihood Binomial (MLB) method we developed which considers the sibship as a whole ($p<0.007$). Stronger evidence was obtained by the MLB approach when coding the Mitsuda as a four classes (<3, [3-5],]5-10[, ≥10 mm) categorical trait ($p<0.005$).These findings support the view that NRAMP1 itself could be involved in the control of the delayed response to lepromin injection.

61

Assessing the proportion of HLA-linked and unlinked determinants of Hodgkin's Disease

Yin Yao Shugart
Department of Family Medicine and Clinical Epidemiology, University of Pittsburgh, USA

Through a literature search, we identified a total of 43 sib pairs with Hodgkin's disease (HD) who were HLA haplotyped. A reanalysis of these data revealed a partition of 22 (2 sharing), 15 (1 sharing) and 6 (zero sharing). In an attempt to evaluate the role HLA-linked and unlinked determinants of HD, two non-parametric methods were used. Chakravarti et al (1987) proposed a one-sided test statistic without assuming a genetic mode of inheritance. Using their method, we first tested the null hypothesis of no linkage, where T is asymptotically distributed as a standard normal distribution. We computed T =3.52 with an associated p-value less than 0.005. Thus, the hypothesis of no linkage was rejected and the estimated recombination fraction was 0.192. Furthermore, a heterogeneity test based on the IBD score gave significant results with an estimated linked proportion of 0.70. In addition, we applied Risch's method for relative risk estimation (1991). The relative risk in affected HD siblings (λ) was estimated to be 1.79, which differed significantly not only from what was estimated by Grufferman et al (sevenfold increased risk in siblings of the HD patients, but also from Risch's estimation (λ=5.75 or 2.87) based on 23 HLA-haplotyped affected sib pairs summarized by Hors and Dausset (1983). Taken all together, our results suggest that the association between HLA region and HD continues to hold true, but HLA alone can not account for all the familial aggregation of HD, and that an environmental factor, such as the Esptein-Barr Virus (EBV), may play a significant role. Our current focus is on the interaction between the generic susceptibility due to HLA determinants and the role of EBV infection.

62

Using a generic Gibbs sampler (BUGS) to fit variance components models for Normal and binary traits and for right censored survival times

P.R. Burton, K.J. Tiller, L.J. Palmer
Division of Biostatistics and Genetic Epidemiology, TVWT Institute for Child Health Research, Perth, Australia

Markov chain Monte Carlo (MCMC) methods are set to play an increasing role in the genetic epidemiology of complex diseases. However, the specification of appropriate transition kernels - for example, the set of full conditional distributions required for Gibbs sam-

pling - is time consuming and demands specialist knowledge. BUGS is a generic Gibbs sampler. Once you have specified the basic structure of a model, it generates the full conditional distributions for you; it is therefore easy to use. We have developed a BUGS-based approach to genetic variance components modelling which extends naturally from Normal to binary traits and to phenotypes represented by a (possibly right censored) survival time. We will firstly show how to construct a basic (fixed effects, σ^2_A, σ^2_C, σ^2_{Cs}, σ^2_E) variance components model in BUGS for a Normally distributed phenotype using data from nuclear families. We will then show how the model extends to binary traits, right censored survival times and non-nuclear pedigrees. We will present the results of extensive simulation studies investigating bias and consistency and will illustrate our approach using traits associated with asthma/atopy from the Busselton cohort study. We will also present solutions to a number of fundamental problems that can arise when Gibbs sampling is used in this setting.

63

Major gene focusing: A novel technique for high efficiency linkage analysis and genomic screening for complex traits

L.J. Palmer, K.J. Tiller, W.O.C.M. Cookson, P.R. Burton.
Division of Biostatistics and Genetic Epidemiology, TVWT Institute for Child Health Research, Perth, Australia

Whole genome screening using polymorphic linkage markers is currently being undertaken for a large number of complex human diseases, and many more studies are planned. However, whole genome screens in humans are very costly, and the relative efficiencies of various data selection and analytic strategies are uncertain. In this paper, we describe a novel strategy that uses Gibbs sampling to undertake a segregation analysis prior to any genotyping. This allows identification of the families most likely to be informative for a linkage analysis. Informativeness depends upon the segregation model. For example, for a rare dominant allele, families may be ranked on the posterior probability that at least one parent is heterozygous at the unknown major locus. A genome screen can then be undertaken using standard linkage techniques on the subset of families with the highest ranks. Alternatively, the posterior probabilities may be used as regression weights in the linkage analysis, e.g. in calculating the Haseman-Elston statistic. This produces a high-efficiency linkage statistic for the analysis of previously genotyped samples. Using both simulated data and examples from a whole genome screen for asthma phenotypes, we illustrate the potential of this technique to improve the efficiency of whole genome screens.

64

Haptoglobin Polymorphism Association with Cardiovascular Risk Factors in Healthy Children

M. Bicho[1], C. Monteiro[2], M.J. Laires[2], L. Sardinha[3], S. Llobet[4], P. Marques Vidal[4], M.J. Halpern[4]
[1]Laboratório de Bioquímica, Faculdade de Motricidade Humana, U.T.L.; [2]Laboratório de Genética, Faculdade de Medicina de Lisboa, U.L.; [3]Núcleo de Exercício e Saúde, Faculdade de Motricidade Humana, U.T.L.; [4]Instituto Superior de Ciências da Saúde

Hp1.1 phenotype was described associated with salt sensitivity while Hp2.2 with cardiovascular risk (peripheral ischaemic disease) in hipertensive patients. The aim of the present study was to determine, in healthy children, the relationship between haptoglobin phenotypes and the lipid profile and plasma peroxidation levels. The study included 171 children from Lisbon, 76 boys and 95 girls with a mean age of 13.0 ± 1.7 and 12.9 ± 1.7 years respectively. A blood sample was collected after a 12 h fast to evaluate: the lipid profile: total cholesterol, HDL and LDL cholesterol and triglycerides by commercial kits and apolipoproteins A_1 (Apo A_1) and B (Apo B) by imunonepholometry; and peroxidation indices: plasma TBARS by a thiobarbituric acid assay and *in vitro* LDL peroxidation induced by phenilhydrazine detected by the same assay. Hp phenotypes were determined by PAGE Statistical analysis included variance analysis. We observed significantly different mean values for HDL cholesterol and Apo A_1.

Parameter	Hp1.1	Hp2.1	Hp2.2	p
HDL-C (mg/dl)	52.1 ± 13.7	58.6 ± 13.3	61.2 ± 14.4	0.04
Apo A_1 (g/l)	1.25 ± 0.18	1.39 ± 0.18	1.40 ± 0.18	0.002

According to several authors, Hp is one of the proteins which is bound to HDL. Additionally, to Hp2.2 was attributed higher immune reactivity and capacity to inhibit prostaglandin synthetase. As Hp2.2 individuals show higher HDL-Col and Apo A_1, we might suggest that Hp2.2 proteins may have higher affinity to HDL and be associated with higher levels of this lipoprotein.

65

Characteristics of a Genetic Map for a Cost-effective Genome Screen using Diallelic Markers.

KAB Goddard and EM Wijsman
Univ. Washington, Seattle, USA.

Many new technologies for genotyping diallelic markers have recently been developed which may be lower in cost, and more easily automated than current methods. Studies of complex traits generally require large sample sizes, so a reduction in the cost of genotyping may have a significant impact on the overall cost of a study. Using a combination of analytic and simulation methods, we explored characteristics of uniform vs. clustered diallelic marker maps, and evaluated the cost of a study relative to the cost using microsatellite maps. Issues that were addressed in comparing the map structures include the information content for clustered or single markers, and the map accuracy.

The multi-locus polymorphic information content (MPIC) was derived to evaluate the information content of a cluster of diallelic markers. We found that the diallelic markers should have a common allele frequency between 0.5 and 0.75, there should be at most 5 markers per cluster, and there can be some linkage disequilibrium between markers in a cluster. For an accurate map, a uniformly spaced diallelic marker map is more cost-effective than clustering the markers. However, the genotyping cost per marker for diallelic markers can be at most 60% of the genotyping cost per marker for microsatellite markers. We also show that marker distance misspecification causes a reduction in the lod score under linkage, and inflation of the lod score under certain circumstances when there is no linkage. Overall, when both marker information and map accuracy are taken into account, an optimal solution may be a clustered design with two markers per cluster.

66

Identifying gene carriers for heterogeneous and complex traits

E.W. Daw, E.A. Thompson and E.M. Wijsman
University of Washington, Seattle, WA

The process of identifying genetic traits typically consists of three stages: localization to a chromosomal region; narrowing the region; and, finally, the difficult process of identifying the gene. For simple Mendelian traits, this methodology is well-established. In contrast, a complex trait has multiple genes; without any indication of which genes are segregating in which families, stratification of families by underlying trait loci is difficult, fine-scale mapping is more complicated, and positional cloning is impractical. Recently, Monte Carlo Markov chain (MCMC) methods have been introduced which are able to find simultaneously initial locations for multiple loci contributing to oligogenic quantitative traits (Heath, AJHG 61:748-760 (1997)). With these methods, accuracy of gene localization is greater than has previously been possible for complex traits.

Here we present a method which uses the trait locus genotype simulated in each step of the MCMC analysis to help identify potential "gene carriers" for each of the identified linked loci. So far, the basic method has shown good sensitivity, correctly identifying nearly all the gene carriers in an analysis of Alzheimer's disease. The specificity was not as good, with a number of individuals incorrectly identified as "carriers." We examine several adjustments to improve the specificity without losing sensitivity and contrast these methods with a survival-curve-fitting stratification method. The methods presented here can identify a few families in which to concentrate fine-scale mapping and gene-cloning efforts for complex traits.

67

Comparison of SIBPAL (S.A.G.E.) and MAPMAKER/SIBS for Sib-pair Linkage Analysis of Qualitative Traits

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

Two commonly used software packages for sib-pair linkage analysis of qualitative traits are SIBPAL and MAPMAKER/SIBS. SIBPAL measures the mean proportion of IBD sharing for concordant unaffected, discordant and concordant affected sib pairs and tests whether allele sharing is higher than the 50% expected for concordant sib pairs and lower than expected for discordant sib pairs, using a t statistic for each stratum which is asymptotically normally distributed. MAPMAKER/SIBS maximizes the likelihood described by Risch(1990), restricting the maximization to Holmans possible triangle(1993), and calculates a lod score, which follows a mixture of two χ^2 distributions with 1 and 2 degrees of freedom. In a simulation study we compared the Type I Error rate of these two packages in small samples for two point linkage analysis of qualitative traits using affected sib pairs. For this comparison we used the p value for the t statistic which SIBPAL reports and calculated the corresponding p value from the mixture of χ^2 distributions for the Holmans possible triangle. We generated 18,000 replicates for four sample sizes of 10, 20, 40, and 80 sibpairs. Each replicate consisted of nuclear families with two affected siblings and the genotypes of the parents known. For both programs the larger sample sizes should result in a smaller deviation from the expected type I error rate since both statistics are asymptotically derived. We found that both packages performed well even with small samples and were comparable in their type I error rates. These results suggest that both packages per-

form equally well for pairwise linkage studies of qualitative traits.

68

Population stratification in association studies

John S. Witte,[1] W. James Gauderman,[2] and Duncan C. Thomas[3]
[1]Case Western Reserve University, Cleveland, OH USA
[2]University of Southern California, Los Angeles, CA USA

Historically, conventional association studies of genetic traits have been rather precarious, leading to numerous instances of apparent false positives. In contrast, linkage analysis has served as the gold standard for detecting genetic traits. More rigorous association study designs have improved the ability of this approach to accurately detect genetic effects. A major concern that remains with association studies is population stratification. Here we investigate the bias arising from population stratification under a number of different scenarios. For random population-based association studies (i.e., using non-related controls), the magnitude of population stratification varies considerably—with the extent of this bias depending on the relative allele frequencies and risks between contrasted populations. As an example, assume that at a diallelic locus there is a five-fold ratio of allele frequencies and baseline trait risks (in the same direction) across two populations. Further assume that: the true mode of inheritance is additive; the true genetic relative risk is 20 (for the "high risk" allele); the average allele frequency across the populations is 0.01; and the average population disease rate is 0.1. An association study that ignores the stratification across the two populations will overestimate the true genetic relative risk by 100 percent.

69

Effect of Bilineal Inheritance on the Power of Affected Sib-pair Linkage Analysis

Anita L DeStefano, Michael Nicolaou, L. Adrienne Cupples
Boston University Schools of Medicine and Public Health

Affected sib pairs ascertained for linkage analysis of a common complex disease are likely to contain a substantial proportion of nuclear families in which both parents are affected. Such bilineal families have the potential to reduce the power to detect linkage if, for example, both parents carry deleterious mutations in the same gene but these genes are linked to distinct alleles at nearby markers (e.g. in the absence of linkage disequilibrium). Affected sibs may share no alleles in common if each sib inherited the deleterious gene from a different parent. Simulation studies were carried out to evaluate the effect of including bilineal nuclear families on power of sib pair linkage analysis and to evaluate the efficacy of excluding these pairs. Linkage between an autosomal dominant disease gene and a polymorphic marker was assessed using SIBPAL (S.A.G.E. v3.1) in three populations: 1) 200 sib pairs with one affected parent 2) 250 sibpairs of which 200 had a single affected parent and 50 had two affected parents 3) 200 bilineal sibpairs (both parents affected). Power to detect linkage was 100% at the .0001 significance level for populations 1 and 2. For population 3 (all bilineal families) power was greatly reduced: 1% at the .0001 level and 16.8% at the .05 level. The mean proportion of allele sharing was .65, .62 and .52 for populations 1, 2 and 3, respectively. Although the power to detect linkage was the same in populations 1 and 2, the reduction in the proportion of alleles shared suggests that with a smaller sample size, exclusion of bilineal pairs may increase power.

70

A Monte-Carlo based test for quantitative trait allelic association

D.L. Duffy[1].
[1]Queensland Institute of Medical Research, Brisbane, Australia.

Allelic association for quantitative traits within a pedigree is usually tested for as a fixed effect within a mixed (fixed and random effects) model including a background additive genetic correlation between relatives. In the present approach, a standard fixed-effects model (estimating allelic means) is fitted to the data, but the P-value is generated by gene dropping simulation (assuming the null hypothesis of no association). Multiple replicates of the sample are generated, with the original quantitative trait phenotypes and simulated genotypes, and the model fitting repeated to calculate the distribution of the error mean square under the null hypothesis. It can also be applied as a test for linear trends in allele frequencies between ordered groups (by age cohort or physical location say), with the group score replacing the phenotype, giving very similar results to the alternative approach of testing for trends in allele frequencies by variance-weighted logistic regression. The approach can also be applied on a pedigree by pedigree basis, offering a nonparametric test for linkage.

I will present power calculations. The method is available in the computer package Sib-pair, which can be downloaded via *http://www.qimr.edu.au/davidD/davidd.html*.

71

Genetics of Parkinson's disease: Results from linkage analysis, segregation analysis, and association study.

S. Zareparsi[1], J. Kaye[1], J. Nutt[1], P. Kramer[1], M. Litt[1], E. Harris[2], H. Payami[1]
[1]Oregon Health Sciences Univ., USA, [2]Kaiser Permanente Center for Health Research, USA.

Parkinson's disease (PD) is a prevalent movement disorder of unknown cause. Less than 1% of PD is autosomal dominant, 10-20% are familial with no clear inheritance pattern, 80-90% are non-familial. PD was mapped to chromosome 4q, and a mutation in the a-synuclein gene was subsequently identified in several autosomal dominant families. We screened 65 affected members from 40 PD kindreds, none had the mutation. Using multi-point linkage, we excluded a 16cM region around a-synuclein. Recently, a PD susceptibility locus was mapped to chromosome 2p. Our linkage results from six families indicate that a subset of them show evidence consistent with the reported linkage. Except for the rare autosomal dominant cases, the mode of inheritance of familial PD is not clear. We performed segregation analyses on 136 randomly ascertained patients. The hypotheses of a non-transmissible environmental factor, no major gene or type, and all Mendelian inheritance were rejected. The familial clustering of PD was best explained by non-Mendelian transmission of a rare familial factor that influences age at onset. We performed an association study of onset age of PD with Apolipoprotein E (APOE), which is associated with Alzheimer disease (AD). Clinical and neuropathological overlap is observed between PD and AD. In addition, PD and AD appear to cluster within families. Age at onset of PD varied by APOE alleles in a manner similar to AD. Those with APOE-ε4 allele had the earliest onset, ε3 homozygotes had an intermediate onset, and those with ε2 had delayed onset. Collectively, results suggest genes are involved in the causation, susceptibility to, and expression of PD.

72

Power of Segregation Analysis of Quantitative Traits When Relevant Covariates Are Not Included

Suh-Hang Hank Juo[1, 2], Alexander F. Wilson[1], Terri H. Beaty [3]
[1]National Human Genome Research Institute, National Institutes of Health, Baltimore, USA
[2]Laboratory of Statistical Genetics, Rockefeller University, New York, USA
[3]Department of Epidemiology, Johns Hopkins School of Hygiene & Public Health, Baltimore, USA

When using segregation analysis one needs to adjust for covariates to identify major gene effects on complex traits. However, information on environmental factors is rarely complete in human studies. Therefore the power of segregation analysis, when adequate adjustment for covariates cannot be made, is an important question; and one that can be addressed using the simulated data from the Genetic Analysis Workshop 10 (GAW10). We determined the empirical power of segregation analysis of a quantitative trait under the control of multiple loci using different levels of adjustment for covariate effects, ranging from no adjustment to nearly complete adjustment. One-locus segregation models were used for all analyses, although the generating model for the trait was an oligogenic model. Thus, this study can also assess the utility of one-locus segregation analysis for a quantitative trait controlled by more than one locus.

Results indicate that the ability to detect Mendelian inheritance and to reject environmental transmission in segregation analysis (i.e. power of a "compound" test) is quite low when covariate effects are ignored. However, the power improves to as much as 90% in these data when adequate adjustments for covariates are made. One-locus segregation analysis of oligogenic quantitative traits can show evidence of a major gene in the presence of residual familial correlation.

73

A QTL for obesity maps in the region of the beta-3 adrenergic receptor (*B3AR*).

B.D. Mitchell, A.G. Comuzzie, S.A. Cole, J. Blangero, J.W. MacCluer, J.E. Hixson.
SW Foundation for Biomedical Research, San Antonio, TX, USA.

A missense mutation (Trp64Arg) in the *B3AR* gene has been associated with obesity in some populations, although in several, the effect is observed only in individuals homozygous for the mutation. However, a large number of studies have failed to find evidence for an association, and no studies have found strong evidence for linkage of this gene with obesity. Using 370 microsatellite markers and a 10 cM average map density, we performed a genome-wide scan to detect linkage to body mass index (BMI) in 479 individuals from 10 large Mexican American families. The Trp64Arg variant was typed by PCR. The mean age of study subjects was 38.8 yr. and mean BMI was 29.7 kg/m^2. Linkage analysis was performed using multipoint variance component methods. Lod scores > 1.0 were observed in only two chromosomal regions, including one on chromosome 8p, approximately 65 cM from pter and within 5 cM of the *B3AR*. The multipoint lod at this locus was 2.7. Our sample included 11 subjects who

were homozygous for the Trp64Arg variant. To see if these individuals contributed disproportionately to the linkage, we removed them and repeated the analysis. The lod score was essentially unchanged (lod = 2.8). In conclusion, these results provide evidence for linkage of obesity with the chromosome 8p gene region containing *B3AR*, and the effect of this locus is not confined to those homozygous for the Trp64Arg variant. Obesity in this population may be influenced by the Trp64Arg mutation, by a different mutation in this gene, or by a mutation in a different gene in this region.

74

The power of tests of linkage and association when a threshold is used to classify a continuous trait as discrete.

A.F. Wilson and A.J.M. Sorant
National Human Genome Research Institute, National Institutes of Health, Baltimore, USA.

In studies of traits like hypertension or obesity, information can be lost when arbitrary criteria are used to dichotomize a continuous variable to determine "affection status." The effect of this loss of information on the power of model-independent tests of linkage and association was evaluated using computer simulation. The Genometric Analysis Simulation Program (G.A.S.P.) was used to simulate a continuous variable with a threshold set so that approximately 5% of the population would be affected. One hundred nuclear families, each with four offspring, were ascertained so that at least two offspring were affected in each family. Models considered include: heritabilities from 0 to 0.8, complete and no linkage disequilibrium, and recombination fractions from 0 to 10 cM. The power of four variations of the Haseman-Elston (H-E) sib-pair test for discrete traits, and the TDT test were compared to the H-E test for a continuous trait [S.A.G.E. v3]. The power of the H-E test for concordantly affected pairs (only) was nearly identical to that of the H-E test for the continuous trait. The H-E test requiring that the means for the estimated proportion of alleles i.b.d. be > 0.5 for concordantly affected and unaffected pairs and < 0.5 for discordant pairs, was considerably less powerful. The TDT test of association was more powerful than the H-E test for continuous traits when there was complete linkage disequilibrium; otherwise the power of this test was only marginally better than the type I error rate.

75

A Genome Scan for Autism Provides Additional Evidence for Linkage to Regions on Chromosomes 4, 7, and 16.

Susan L. Santangelo[1,5] for The Collaborative Linkage Study of Autism (CLSA)[1-4].
[1]Tufts University, [2]University of Iowa, [3]Vanderbilt University, and [4]Johns Hopkins University Schools of Medicine, and [5]Harvard School of Public Health.

The CLSA is a collaboration among three clinical data collection sites, two molecular and two analytic laboratories, funded through the NIMH (USA) to carry out a genetic linkage study of autism. The goal is to genotype 150 nuclear families in two replicate samples. We are currently completing our first genome-wide scan (291/344 microsatellite markers completed) of 76 families, including 72 sib pairs, 4 sib trios, and 148/153 parents. There are several instances of apparent duplications and deletions on chromosome 15q11-13. Preliminary results suggest several of the same chromosomal regions found by the International Molecular Genetic Study of Autism Consortium (IMGSAC). Regions on seven chromosomes (2,4,7,11,14,15,16) generated two-point maximum heterogeneity lods (MHL) $\geq$ 1.5. The most significant results so far are for regions on chromosomes 4, 7, and 16, in that order. (IMGSAC's most significant results were for regions on chromosomes 7, 16, and 4, in descending order.) A region on chromosome 4, about 50 cM distal to the region identified by IMGSAC (the map distances may not be identical), generated an MHL of 2.5 and a maximum multipoint lod score (MLS) of 3.0; a region on chromosome 7, the same one identified by IMGSAC, yielded an MHL of 1.9 and an MLS of 2.1; and a region on chromosome 16, approximately 50 cM distal to the area identified by IMGSAC, produced an MHL of 2.3 and an MLS of 1.3.

76

Efficient Monte Carlo Evaluation of the Multivariate Normal Integral

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

The evaluation of pedigree likelihoods for threshold characters under a model of normally distributed disease liability requires evaluation of the integral of the multivariate normal density. The dimensionality of the integral scales with the size of the pedigree, and the recent development of variance components methods for genetic linkage analysis in pedigrees of arbitrary size and complexity have necessitated the evaluation of integrals for which the number of dimensions can reach 100—1000.

Monte Carlo methods of integration are by nature asymptotically slow to converge, so it is not unusual that these methods have not seen widespread use in statistical genetic applications. We describe a series of transformations that renders the multivariate normal

integral amenable to efficient Monte Carlo evaluation, and compare the performance of a Monte Carlo method to a standard approach using repeated conditional univariate integration.

Integral estimates returned by the Monte Carlo method are effectively unbiased, whereas repeated conditional univariate integration tends to overestimate the integral for high-dimensional cases. The conditional integration strategy is exceptionally fast, but the execution time increases rapidly with increasing dimensionality. The Monte Carlo algorithm, however, exhibits nearly constant execution time for the requested accuracy, and for high-dimensional cases can be much faster than repeated conditional univariate integration.

77

Associations between bone density and vitamin D and estrogen receptor polymorphisms in postmenopausal women from the Quebec population

Y. Giguere[1,3], S.Dodin[2], C. Blanchet[2], M. Dumont[2], K. Morgan[4] and F. Rousseau1[3].
[1]URGHM & [2]U. Endo. Reprod, CHUQ-PSFA, [3]Dept. Biol. Medicale, Fac. medecine, U. Laval, Quebec and [4]MGH Res. Instit., McGill U., Quebec, Canada.

We are conducting a population-based study of osteoporosis in 5000 French-Canadian, postmenopausal women. We undertook a preliminary analysis in 575 ambulatory women of associations between polymorphisms of the vitamin D receptor (VDR-BsmI and VDR-FokI) and the estrogen receptor (ESR-PvuII) and three bone density measures adjusted for age and weight. [femoral neck(FN), lumbar(L2L4), and heel(SI)]. The two receptor loci are unlinked.

VDR-FokI and ESR-PvuII RFLPs were in Hardy-Weinberg equilibrium while the VDR-BsmI polymorphism was not (χ^2=6.7, p=0.01). In contrast with previous reports, the pair of VDR polymorphisms were in linkage disequilibrium (two-locus composite disequilibrium analysis, χ^2=5.9, p=0.02). BMD did not differ significantly among VDR-BsmI, VDR-FokI or ESR-PvuII genotypes. Similarly, there was no significant overall interaction between VDR and ESR for any site. However, the ESR-PvuII polymorphism was associated with L2L4 BMD in younger postmenopausal women aged < 60 years [ANOVA p=0.01; (Δ(PP vs. Pp) = 0.7, p=0.005] while the VDR-BsmI polymorphism was associated with L2L4 BMD in older women aged > 70 years [(Δ(BB vs. bb) = 0.8, p=0.03]. In addition, VDR-FokI was associated with FN BMD [ANOVA, p=0.02; (Δ(Ff/ff vs. FF) = 0.4, p=0.006] in women aged 60 to 70 years. Furthermore, women with the VDR-bb/ESR-PP two-locus genotype (9% of women) had the lowest mean Z-score (-0.28 Δ, p=0.02) compared with the rest of their cohort. In conclusion, the French-Canadian population is well suited for population-based genetic analysis, including gene-gene interactions, of common, complex traits such as osteoporosis.

78

Stratified case-control sampling using related controls.

K.D. Siegmund, B. Langholz, D.C. Thomas.
University of Southern California, Los Angeles, USA.

We describe a conditional logistic regression approach to estimate the effects of measured genes, environmental exposures, and gene by environment interactions for the analysis of case-control data arising in families selected through a multi-stage sampling design. We allow cases to be sampled conditional on family history of disease in their parents and consider the use of controls selected from among the case's unaffected relatives. The method weighs the likelihood contributions by the fraction of cases sampled in each stratum of family history. For siblings who are necessarily matched on history of disease in their parents, this likelihood reduces to the traditional conditional logistic likelihood. For cousin controls that do not necessarily share the same family history as the case (e.g. affected mother), the sampling fractions from each stratum of family history are necessary to obtain unbiased parameter estimates. We apply this method to compare the efficiency of study designs for estimating gene and gene by environment interaction effects using sibling and cousin controls. We consider a common metabolic gene with small relative risk (allele frequency = 0.44, Rg=2 under recessive gene action) interacting with a common environmental exposure (prevalence=50%, Re=2, interaction relative risk ratio = 2). A sample of 835 (8.5%) case-control sets is selected from a possible 9795 in a population with a disease prevalence of 2%. We find that under random sampling, sibling controls are less efficient than cousins for estimating the main effect of a single gene (asymptotic relative efficiency (ARE)= 85%) but are more efficient for estimating gene by environment interaction effects (ARE = 115%). The efficiencies in the estimates from the sib control design are slightly improved by over-sampling cases with a positive family history of disease. At the same time, the efficiency of the estimates using cousin controls decline.

79

Localization of Quantitative Trait Loci Using Variance Component Methods in Ascertained Samples

J. Dupuis and P. Van Eerdewegh
Genome Therapeutics Corporation, Waltham, USA

Ascertained samples often yield increased power to detect linkage for complex traits. Combined with a variance component methodology, families ascertained through one extreme offspring provide a powerful design for the localization of quantitative trait loci (QTL). The properties of this approach were examined using a simulation study, controlling for a genome-wide type-I error of 5% using the proper ascertainment correction. Reasonable sample sizes were sufficient to detect linkage to loci of moderate effect and the resulting estimate of the location of the true locus was very accurate in most replicates. For example, 500 sibling pairs and their parents, with one phenotypically extreme offspring in the top 10% of the distribution gave 75% power to detect linkage to loci with heritability of 20% when the total heritability was 70%, assuming a biallelic locus with allele frequency of 10%. For the same model, the median size of the 1-lod support intervals was 5.5 cM with an average coverage probability of 89%, while the average absolute error in the estimate of the true location was 2 cM. More stringent criteria for the selection of a family (i.e. offspring in the top 5%) resulted in smaller lod support regions with improved power, and the ability to detect and localize loci of smaller effects. In addition, we observed that a family with *n* offspring provides more information than *n-1* independent sibling pairs, but not as much as *n(n-1)/2* independent pairs.

We conclude that for complex traits that are quantitative in nature, selecting nuclear families through an extreme offspring is a very efficient way to detect QTLs and narrow the chromosomal region of interest, by making use of all available phenotypic information. Moreover, the power of this method can be improved by using larger sibship sizes.

80

The effect of using incorrect trait genotypic means on the power of model-based linkage analysis in quantitative traits.

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

In model-based lod-score linkage analysis, the mode of inheritance for both the trait and the marker loci should be modeled correctly to achieve the maximum informativeness of the data for linkage analysis. Previous studies have shown that if errors are made in specifying the trait locus but the marker is specified without error, and vice versa, then power is reduced but Type I error rates are not increased. The objective of the present study is to compare the power obtained by the lod-score linkage method when the trait genotypic means are specified incorrectly and when the marker genotypic information on one or both of the parents are missing. Data were generated for quantitative trait and marker loci in nuclear families using G.A.S.P. (V3.3). The trait was due to an additive major locus with a random environmental effect and two equifrequent alleles. The heritabilties of the trait ranged from 0.1 to 0.9. The lod-score test of linkage (implemented in LODLINK) was performed on the trait and the linked marker (heterozygosities ranging from 69% to 80%) in each sample. The results show that in the case of tight linkage, as the heritability increases the amount of decrease in power becomes more substantial with the use of incorrect trait genotypic means. However, when expressed as a percentage of the original power the decrease in power was larger when heritability was lower. For example, in the case with no parental information and 90% heritablility, the power is reduced from 98% to 70% (28.57% reduction) at a lod-score of 2 when the trait model is misspecified, whereas, at 50% heritability, the power decreases from 33% to 19% (42% reduction).

81

Interpreting intralocus disequilibrium in candidate loci for cardiovascular disease.

W. Klitz[1], C. Pallaud[2], M. Grow[3], S. Cheng[3], H.A. Erlich[3], G. Siest[2] and S. Visvikis[2].
[1]Univ. California, Berkeley, USA, [2]Centre de Medicine Preventive, Nancy, France, [3]Roche Molecular Systems, Alameda, USA.

Thanks to the interpretable influence of mutation and recombination, the disequilibrium levels among tightly linked markers contain clues as to the evolutionary history of a region, which may be important in understanding the genetic influence on complex disease phenotypes. Data obtained from a prototype assay for examining variation and interaction among multiple genetic sites occurring in candidate loci for cardiovascular disease were used to examine intralocus disequilibrium in 284 population-based nuclear families. The assay is based upon multiplex PCR amplification and immobilized probe arrays. From a total of 35 sites distributed among 15 loci, endothelial leukocyte adhesion molecule-1 (ELAM), lipoprotein lipase (LPL) and apolipoprotein CIII (ApoCIII), each had multiple sites suitable for analysis. Disequilibrium levels were the maximum possible (i.e. D' = 1.0) in many cases, suggesting an absence of recombination between sites. In some instances, however, all four possible haplotypes were present in the two allele-two site systems, implying recombination, despite the very limited distance between sites. Furthermore, in spite of low sta-

tistical power due to low frequencies of the uncommon alleles at the LPL sites, the configuration of haplotypes suggested the possible influence of selection. The frequency and disequilibrium levels associated with one ApoCIII site also implied a possible evolutionary history of positive selection.

82

Classification Techniques Applied to Allele Sharing Data

Rosalind J. Neuman
Washington Univ., St. Louis, MO

Latent class analysis (LCA) and cluster analysis (CLA) are two classification methods, differing significantly in their underlying assumptions and methodologies, which have traditionally been used to define phenotypic subtypes based upon a set of observed categorical items. The objective in this study was to determine whether LCA and CLA could distinguish susceptibility loci from loci having no effect on a quantitative trait (Q1) based upon the IBD status of pairs of relatives drawn from the GAW10 simulated data sets. To that end two data sets were created. The first contained first cousin pairs from the GAW10 extended pedigrees; the second were sib pairs from the nuclear pedigrees. A pair of relatives was included in the data sets if both were either in the upper 10% of the Q1 distribution or if they were discordant, one in the upper 10%, the other in the lower 10%. Only one pair was chosen from a family, and not all families contributed pairs. 934 pairs of sibs chosen (38.5% concordant for the upper 10% of the Q1 distribution) and 691 first cousin pairs (44.1% concordant high). IBD statuses at 12 markers distributed across 4 chromosomes were used as input for both methods. Both CLA and LCA identified a class in which more than 90% of pairs shared IBD=1 for cousins and IBD=2 for sibs at a marker that was linked to a major susceptibility gene for Q1. In this class, 60-70% of the pairs were concordant high for the Q1 phenotype (depending on the type of relative pair and analyses). Classes were identified in which a large proportion of the pairs shared alleles at markers not linked to a major Q1 gene, however, the proportion of concordant high pairs was not elevated. The structure of the classes derived from the cluster analysis were remarkably similar to those obtained from the latent class analysis. Note that LCA and CLA allow for multiple disease loci situated on different chromosomes. These analyses indicate that LCA and CLA may be useful for detecting chromosomal regions containing susceptibility loci.

83

Power gains in model-independent sib-pair linkage tests for a quantitative trait based on minimum and moving average marker cluster t-statistics.

A. Kingman[1], A.J.M. Sorant[2], E.W. Pugh[3], L.R. Goldin[4], A.F. Wilson[2].
[1]NIDR/NIH, Bethesda, MD, [2]NHGRI/NIH, Baltimore, MD, [3]CIDR, Johns Hopkins U., Baltimore, MD, [4]NIMH/NIH, Bethesda, MD

When using linkage analysis to perform genomic screening for complex traits, regions may be identified where two-point linkage analysis produces modest p-values that, taken individually, would not be considered to be suggestive of linkage. As noted by Goldin and Chase [1997], one approach is to combine information from correlated clusters of markers into "regional" linkage tests. The properties of two classes of regional linkage test statistics, based on the t-statistics from the Haseman-Elston sib-pair linkage test [1972], were investigated over a range of heritabilities and different marker sets. The first class, "min-t", included minimum t values from sets of consecutive markers. The second class, "ma(t)", included moving averages of the t-statistics. The Genometric Analysis Simulation Program (G.A.S.P.) was used to simulate a quantitative trait and marker loci in random samples of nuclear families, based on different assumptions about heritability, marker spacing, etc. The sib-pair linkage t-statistics were obtained using SIBPAL [S.A.G.E.,1997].

Initial results, based on a 10 cM map, suggest that the power for the min-t statistic, requiring that two consecutive t-tests be significant, was increased over two-point tests. Larger gains were observed for moving averages, ma(t), of 2 to 5 consecutive t-test values. The gains in power to detect significant linkage ranged between 5% to 20% for heritabilities between 30% and 60%.

84 [Invited Speaker]

The use of linkage and association for genome scanning and candidate gene strategies

H. Bickeböller
Technische Universität München, Germany

The tool box for the investigation of a susceptibility gene for a complex trait contains linkage and association approaches. Linkage analysis is widely used in genome scans and candidate gene studies. Association, including family-based association studies, is generally used for candidate genes and for fine mapping. With current and upcoming advances in technology (DNA-chips, SNP maps, genomic mismatch scanning) strategies including the respective use of linkage and association need to be further discussed. Both genome

scan and candidate gene strategies have their advantages. They are not exclusive, but should be combined. For genome scans there is no need for defining biological mechanisms. Using non-parametric linkage methods the major problem is to achieve good power even for genes with a moderate effect while using stringent criteria for genome-wide significance levels. Cost, time and map density will play a smaller role in the future. The use of assocation in genome-scans seems at the moment not feasible even with 1cM maps due to the need of a strong to moderately strong linkage disequilibrium. Candidate genes are genes which may be 'functionally related to the disease'. In a narrow sense the gene (or gene region) should relate to demonstrated pathophysiologic abnormality or to an animal model of disease. In a broad sense the gene is part of a biological system hypothesized to play a role in the disorder. Ideally the gene is in a chromosomal region in which (or close to which) some hints to linkage have been observed. The power and the possibility to estimate genetic parameters can be greatly enhanced by simultaneously using linkage and association. If an abundance of candidates is hypothesized (as for most psychiatric diseases) a more stringent significant level, perhaps even close to a genome-wide level, needs to be applied.

85

Evaluating the evidence for linkage of candidate genes with BMI, fat mass, and fat-free mass by using a variance components approach (SEGPATH): The Quebec Family Study

I. Borecki[1], M. Province[1], T. Rice[1], C. Bouchard[2], D.C. Rao[1]
[1] Washington Univ., St. Louis, MO, USA; [2]Laval Univ., Québec City, Québec, Canada.

The Québec Family Study is comprised of a group of randomly-sampled French-Canadian families as well as a set of nuclear families selected for the presence of at least one obese family member. Body mass index (BMI) as well as fat mass and fat-free mass, assessed by underwater weighing, were considered with age, sex, and average energy intake and expenditure as covariates. DNA variation on the following chromosomes with several candidate genes or regions for adiposity was assessed including chromosome 1 – ATPα2, ATPβ1, LEPR; chromosome 2p; chromosome 7 – LEP; and chromosome 8 – LPL. Evidence for linkage was evaluated by using a powerful, newly-developed variance components model (SEGPATH) which includes the effect of the putative QTL, a residual familial component, marital resemblance and excess sibling resemblance. The evidence for linkage at loci along those chromosomes was assessed under a variety of modeling assumptions. Although only modest lod scores were found at any of the specific candidate loci, positive results for linkage with adiposity were found at 1p22-21 and 1q44 (lod > 3) and at 8q24.12 (lod > 2). Variation in results will be discussed as a function of covariate adjustment strategy and the inclusion of other model parameters.

86

Experience with the WPC Statistic in Large Pedigrees

Bertram Müller-Myhsok[1] and Andreas Ziegler[2]
[1]Bernhard-Nocht-Institute for Tropical Medicine, Bernhard-Nocht-Str. 74, D-20374 Hamburg, Germany
[2]Medical Center for Methodology and Health Research, Institute of Medical Biometry & Epidemiology, Philipps-University of Marburg, Bunsenstr. 3, D-35033 Marburg, Germany

The weighted pairwise correlation (WPC) approach first proposed by Commenges (1994) provides a simple and flexible test for genetic linkage. In its current release the WPC program performs single point analyses for qualitative, quantitative or age-dependent traits based on the number of alleles shared identical by state. The WPC statistic is equivalent to a relevant part of the locally most powerful test for homogeneity proposed by Liang (1987). Thus, it is expected to have good power. Corder (1995) found that the WPC statistics was too liberal. Therefore, Commenges and Abel (1996) proposed modifications of the original WPC statistics to improve their robustness. Asymptotic properties of the WPC statistics for independent pairs of observations were derived by Commenges and Jacqmin-Gadda (1997).

We are analyzing pedigree data for Morbus Parkinson in a genome scan. To investigate the behavior of the WPC statistics we explored a single extended pedigree with more than 40 family members. To our amazement, we obtained significant evidence for linkage on 2 of the 22 autosomes studied. However, neither of these two peaks coincide with peaks found by conventional LOD score analysis recently published by Gasser et al. (1998). It is most likely that these two are false-positives.

In this talk we discuss the possible causes for these unreliable results. We give a prospect for further improvement of the WPC statistics to avoid possible false positives.

References

Commenges D (1994). Robust genetic linkage analysis based on a score test of homogeneity: The weighted pairwise correlation statistic. *Genetic Epidemiology* **11**, 189-200.

Commenges D, Abel L (1996). Improving the robustness of the weighted pairwise correlation approach for linkage analysis. *Genetic Epidemiology* **13**, 559-573.

Commenges D, Jacqmin-Gadda H (1997). Generalized score test of homogeneity based on correlated random effects. *Journal of the Royal Statistical Society, Series B* **59**, 157-171.

Corder, E.H. (1995). Weighted pairwise correlation and transmission disequilibrium in a common oligogenic disease. *Genetic Epidemiology* **12**, 747-751.

Gasser T, Müller-Myhsok B, Wszolek ZK, Oehlmann R, Calne DB, Bonifati V, Bereznai B, Fabrizio E, Vieregge P, Horstmann RD (1998). A susceptibility locus for Parkinson's disease maps to chromosome 2p13. *Nature Genetics* **18**, 262-265.

Liang, K.Y. (1987). A locally most powerful test for homogeneity with many strata. *Biometrika* **74**, 259-264.

87

Half sibs in sib pair analysis

Andreas Ziegler and Marcus Kutschmann
Medical Center for Methodology and Health Research, Institute of Medical Biometry & Epidemiology, Philipps-University of Marburg, Bunsenstr. 3, D-35033 Marburg, Germany

One method to establish linkage is based on allele sharing methods for sib pairs. In the last years the use of selected sib pairs to increase power for mapping quantitative traits in humans was discussed intensively. Upon use of this approach, the investigator recruites sib pairs including parents where both sibs have extreme trait values.

Non-paternity is a relevant problem for several phenotypes. For example, Allison (1996) reported non-paternity rates of up to 30% for the quantitative phenotype body mass index (BMI; kg/m²). They were also estimated in several studies being as high as 10%-20% (Le Roux et al., 1992; Macintyre and Sooman, 1991). Undetected non-paternity yields a bias in linkage analysis.

In this paper, we propose an intuitive correction of the usually applied tests so that half sibs may be included in linkage analysis. Additionally, we demonstrate the dependency of non-paternity rates on the chosen extreme sib pair strategy. We furthermore quantify the bias introduced by undetected half sibs in linkage analysis for quantitative traits.

Our results show that linkage analysis with undetected half sibs using extreme concordant sib pairs yield considerably conservative results, while linkage analysis based on extreme discordant sib pairs are markedly to liberal. Our investigation indicates that the relation of all pairs used in a genetic study should be analyzed prior to linkage analysis. For this purpose, techniques as proposed e.g. by Boehnke and Cox (1997) may be applied.

References

Allison, D. B. (1996). The use of discordant sibling pairs for finding genetic loci linked to obesity: practical considerations *International Journal of Obesity* **20**, 553-560.

Boehnke, M. and Cox, N. J. (1997). Accurate inference of relationships in sib-pair linkage studies. *American Journal of Human Genetics* **61**, 423-429.

Le Roux, M. G., Pascal, O., David, A. and Moisan, J. P. (1993). Non-paternity rate and screening in genetic disease analysis. *Lancet* **341**, 57.

Macintyre, S. and Sooman, A. (1991). Non-paternity and prenatal genetic screening. *Lancet* **338**, 869-871.

88

Genetic linkage analysis of parental-origin of shared alleles, and sex of affected offspring in affected sibling pair families with type I diabetes

A.D. Paterson[1], D.M.J.Naimark[2], A. Petronis[1]
[1]Clarke Institute of Psychiatry, [2]Sunnybrook Health Sciences, University of Toronto, Ontario, Canada.

A number of susceptibility loci for type I diabetes have been identified. At two loci, *IDDM1* and *IDDM2*, parent-of-origin effects have been described. At *IDDM7* on chromosome 2q, linkage has been reported in families with affected female offspring. We analysed data from a genome scan from 96 affected sibling pair families, as well as 416 similar families with genotyping data on chromosome 6q (http://www.well.ox.ac.uk/~plyons) to detect parent of origin effects, as well as sex of offspring effects. Parental origin effects were studied using the sib_ibd option of ASPEX v 1.17 and sex of offspring effects using GENEHUNTER v 1.1 were used. Weak positive evidence for linkage (lodscore >1) in paternal meioses was obtained at regions on chromosomes 8, 10 and 17, while similar linkage in maternal meioses was observed on chromosomes 13, 14 and 15. On chromosome 6q evidence for parental origin effects were observed, but were not consistent across two samples, one from the UK, and the other from the US. Regarding sex of offspring-effects, on chromosome 2q we found an opposite effect to that previously reported, with evidence for linkage predominantly from families with male-male affected sibpairs. Additionally, at regions on chromosomes 7 and 8 we found weak positive linkage (NPL>1.5) in families with male-male affected sibpairs. Clearly the results obtained here require extension in addition samples to reach suggested significance criteria. However, linkage analysis of complex traits including parental origin effects, as well as sex-of-offspring effects may assist with the identification of susceptibility loci.