Genetic Epidemiology 15:513–550 (1998)

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

ARCACHON, FRANCE
SEPTEMBER 11-12, 1998

With financial support from:
Institut National de la Santé et de la Recherche Médicale (INSERM)
Commission Européenne
Association Française contre les Myopathies (AFM)
Institut Rhône Poulenc Rorer France

1

Maximum likelihood generalized heritability estimates for blood pressure among Nigerian families.

Rotimi CN[1], Cooper RS[1] Cao G[1], Ogunbiyi O[2], Ladipo M[2], Owoaje E[2] , and Ward R[3]
[1]Loyola University Chicago, [2]University of Ibadan, Nigeria [3]University of Oxford, Oxford, UK

Elevated blood pressure is more common in relatives of hypertensives than in relatives of normotensives indicating familial resemblance of the blood pressure phenotypes. However, most published studies were conducted in families in westernized societies. We examined familial patterns of blood pressure in a population-based sample of 510 nuclear families, including 1552 individuals (320 fathers, 370 mothers, 475 sons and 387 daughters) from Ibadan, Nigeria. The prevalence of obesity is this community is low (body mass index: fathers = 21.6, mothers = 23.6, sons = 19.2 and daughters = 21.0, kg/m^2). The blood pressure phenotype used in all analyses was created from the best regression model by standardizing the age-adjusted systolic blood pressure and diastolic blood pressure to zero mean and unit variance. Heritability was estimated using SEGPATH from the most parsimonious model of no spouse and no sex nor generation difference as 45% for systolic blood pressure and 43% diastolic blood pressure. The lack of a significant spouse correlation suggests little or no effect of common familial environmental influence. However, a strong non-shared environmental effect is also suggested given the heritability estimate of less than 50% for both systolic and diastolic blood pressures.

2

The necessity of individual approach to segregation analysis of some complex diseases.

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

The transmission probability approach is usually used to detect the major gene effect. It has been shown that this approach prevents false inference of major gene. However a false rejection of a major gene effect

may occur due to imperfection of genetic model or misspecification of the disease status. Therefore, to infer a major gene effect one has to pay a special attention to the selection of genetic models and the accuracy of disease status identification.

We performed the segregation analysis of adolescent idiopathic scoliosis using the regressive logistic models with different ways of age-of-onset description and ascertainment correction. We applied two approaches to define the affected status. First, when it has been prescribed to persons with all forms of disease; second, when it has been done for individuals with only pronounced forms of disease. Only in the second case we found the major gene control. The unique method of describing the age dependence of penetrance function allows to detect major-gene effect. This method introduced by Elston and George (1989) corresponds to adolescent nature of disease. It seems that often the segregation analysis of complex diseases requires the special information on the diagnosis and the manifestation of disease. Constructing the genetic models based on this information may be helpful for detecting the major-gene effect.

3

Inflation of Sibling Recurrence Risk Ratio Due to Ascertainment Bias and/or Over-Reporting

Sun-Wei Guo
Institute of Human Genetics and Division of Epidemiology, University of Minnesota, USA

One widely used measure for familial aggregation is the sibling recurrence risk ratio, which is defined as the ratio of risk of disease manifestation given that one's sibling is affected as compared with the disease prevalence in the general population. Known as lambda, now it has been used extensively in the mapping of complex diseases. In this paper, I will show that for a fictitious disease that is strictly non-genetic and non-environmental, lambda can be dramatically inflated due to misunderstanding of the original definition of lambda, ascertainment bias, and over-reporting. Therefore, for a disease of entirely environmental origin, the inflation of lambda due to ascertainment bias and over-reporting is expected to be more prominent if the risk factor also is familially aggregated. This suggests that, like segregation analysis, the estimation of lambda also is prone to ascertainment bias and should be performed with great care. This is particularly important if we use lambda for exclusion mapping, for discriminating between different genetic models, and for association studies, as these practices hinge tightly on an accurate estimation of lambda.

4

Tests and Estimates of Allelic Association in Complex Inheritance

N.E. Morton, A. Collins
Human Genetics, University of Southampton, Princess Anne Hospital, Southampton SO16 5YA, UK

Family-based procedures like the transmission disequilibrium test (TDT) were motivated by concern that sample-based methods to map disease genes by allelic association are not robust to population stratification, migration, and admixture. Other factors to consider in designing a study of allelic association are specification of gene action in a weakly parametric model, efficiency, diagnostic reliability for hypernormal individuals, interest in linkage and imprinting, and sibship composition. Compared with cases and normal controls the TDT is the method of choice for family data, which are not the data of choice for allelic association. The TDT has an efficiency of ½ for parent-offspring pairs, and 2/3 for father-mother-child trios. Against cases and hypernormal controls the efficiency is only 1/6 on the null hypothesis. Although dependent on marker gene frequency and other factors, efficiency for hypernormal controls is never less than for random controls. Efficiency is increased in multiplex families and by inclusion of normal sibs, approaching a case-control design with normal but not hypernormal controls. Isolated cases favour unrelated controls, and only in exceptional populations would avoidance of stratification justify a family-based design to map disease genes by allelic association.

5

Possible departure from Holman's triangle constraints

Dizier MH, Quesneville H, Clerget-Darpoux F
INSERM U155, Paris, France

Holmans (1993) showed that, for any genetic models, the proportions of affected sibs sharing 2, 1 or 0 Identical By Descent parental marker alleles are constrained to belong to a specific " triangle ". Theses constraints are indeed valid as long as in all pairs the phenotypes of the two sibs are underlied by a same model (same probability for a given genotype to be affected). We show in this study that the triangle constraints do not hold anymore in many realistic situations where the sib phenotypes may correspond to different models. It is the case if the two sibs differ for a variable on which the penetrance function depends and which is not taken into account in the phenotype. This variable may be a caracteristic of the trait itself,

eg severe vs. mild form, or the presence/absence of an associated trait or of an environnemental factor. In such situations, relaxing the constraints will improve the power of risk factor detection. In addition, rejection of Holman's constraints will imply that sib phenotypes are underlied by different models and give evidence for an hidden effect of a variable. It is possible to test if the observed heterogeneity is due to a candidate variable by a predivided sample test.

6

Natural Selection for the MTHFR Gene: Mating Type Distortion

L.R. Weitkamp[1], D.C. Tackels[2], A.G.W. Hunter[3], L.B. Holmes[4], and C.E. Schwartz[2].
[1]Univ. of Rochester, Rochester, USA, [2]Greenwood Genetic Center, Greenwood, USA, [3]Childrens Hosp. of E. Ont, Ottawa, Canada, [4]Massachusetts General Hosp., Boston, USA

Heterozygote advantage of the 5, 10-methylenetetrahydro-folate reductase gene (MTHFR) was reported (Weitkamp et al, Lancet, in press) in unaffected first degree male relatives of neural tube defect (NTD) probands (n =258, p = 0.000002) and not in unaffected female relatives (n = 258, p = 0.51). The opposite sex effect on increased MTHFR C677T heterozygosity was observed in spina bifida probands (33 males, p = 0.58; 50 females, p = 0.011). To investigate the mechanism of sex-dependent MTHFR heterozygote advantage in individuals with NTD susceptibility polygenes, we compared MTHFR mating type frequencies in the 167 NTD families and also in 90 random fecund Caucasian couples with expectation based on Hardy-Weinberg equilibrium and observed C and T allele frequencies (0.651 and 0.349; 0.664 and 0.336, respectively). Matings in each cohort were grouped according to whether the probability that an offspring would be heterozygous was 100%, 50%, or 0%. Both cohorts showed mating type distortion (NTD, p = 0.022; random, p = 0.002). However, the distortion was opposite in the two cohorts: e.g., couples that could produce only heterozygous offspring were 7/167 in the NTD families (17.2 expected) and 17/90 in the random couples (9.0 expected). In 69 NTD families of the mating type heterozygote by C/C homozygote the mother was heterozygous in 21 and homozygous in 48 rather than the equal numbers expected (p = 0.001), suggesting an interaction between maternal and fetal genes in fetal survival. Indeed, from these 69 matings, 28 of 36 spina bifida offspring were heterozygous (18 expected, p = 0.0009). These results provide unusual and striking evidence of natural selection.

7

The $(GGN)_n$ polymorphism in the androgen receptor gene is associated with time to relapse and overall survival in men with prostate cancer.

M. Badzioch[1], S. Edwards[2], and R. Eeles[2].
[1]International Agency for Research on Cancer, Lyon, France, [2]Institute of Cancer Research and Royal Marsden NHS Trust, Sutton, UK.

Androgens are required for the growth and function of the prostate gland as well as for the early stages of prostate cancer (PC). US studies suggest that a subset of alleles at one or both of the trinucleotide repeat polymorphisms in the first exon of the androgen receptor gene may be associated with increased risk of PC. Alleles with fewer repeated CAG triplets have been associated with a higher risk of more aggressive disease as well as with increased *in-vitro* activity of the gene product. The other polymorphism, a repeated GGN sequence, has been less studied. We have genotyped 162 British Caucasian PC patients along with geographically matched controls to assess the roles of the $(CAG)_n$ and $(GGN)_n$ polymorphisms in both risk and clinical course of the disease.

We found no association of either polymorphism with PC risk. Significant associations were seen between $(GGN)_n$ alleles with = 17 repeats and shorter disease free survival, DFS, (RR= 1.74, 95% CI= 1.08-2.79) as well as decreased overall survival, OS (RR= 1.98, 1.13-3.45). After adjusting for stage and grade, these effects remained high but became only marginally significant (RR_{DFS}=1.60, p= 0.052; RR_{OS}= 1.60, p= 0.088, respectively). The greatest effects were in early stage, T1-T2 (RR_{DFS}= 3.56, 1.13-11.21) and grade 1, well differentiated (RR_{DFS}= 6.47, 0.57-72.8) tumors.

We conclude that $(CAG)_n$ allele length is not universally associated with PC risk and in this population the $(GGN)_n$ polymorphism is a significant predictor of DFS and OS with longer alleles conferring worse prognosis. Additionally, the $(GGN)_n$ effect is likely to be restricted to the early stages of tumor progression. We hypothesize that although the androgen receptor is universally required for PC development, the nature of its involvement may be variable.

8

Evidence of Linkage Disequilibrium Between POF and FRAXAC1

G. S. Fisch[1], A. Kenneson[2], S.T. Warren[2], D. Zelterman[1]
[1]Division of Epidemiology and Public Health, Yale University School of Medicine, New Haven
[2]Howard Hughes Medical Institute and Depts. of Biochemistry, Pediatrics, and Genetics, Emory University School of Medicine, Atlanta

In Western European populations, mean age of menopause is 51 years. Early menopause or early ovarian failure (EOF) prior to age 47 occurs in 10% of the population, while premature menopause or premature ovarian failure (POF) prior to age 40 occurs in 1%. A genetic influence in menopause has been inferred from studies of women who report a family history of EOF. Cytogenetic analysis of the X chromosome indicates that deletions and translocations co-segregate with POF. Two regions on the X chromosome appear to be critical for normal ovarian function. One is situated between Xq26-28 (POF1). FMR1, the gene responsible for fragile X [fra(X)] syndrome, is also located within POF1 at Xq27.3. Studies have suggested that fra(X) carriers are at increased risk for POF but not EOF. Moreover, the mutation co-segregates with POF in some fra(X) families. Kenneson et al (1997) proposed that the association between fra(X) mutations and POF may result from linkage disequilibrium between FMR1 and a mutation producing POF, and not from the FMR1 mutation itself. Previous studies have shown that FMR1 is in linkage disequilibrium with nearby loci, one of which is labelled FRAXAC1. FRAXAC1 is a polymorphic dinucleotide repeat about 7 kb upstream from the FMR1 repeat region. There are 5 distinct alleles in FRAXAC1, identified simply as A, B, C, D, and E. To test the hypothesis that POF is in linkage disequilibrium with FMR1, 10,600 women, ages 45-54 years, were surveyed. From this sample, 344 cases of EOF were recruited for further analysis, along with 344 age-matched controls who were menopausal from age 47 onward. Of these, 295 women were genotyped for FRAXAC1, producing a total of 590 alleles. Subjects were separated into three distinct groups: A "familial group" of menopausal women with a family history of EOF (N=14); a "sporadic group" of women with no family history of EOF (N=15); two control groups, one of which was composed of women with EOF occuring between ages 40 and 47 (N=164). A second control group consisted of women who were menopausal after age 47 (N=102). To determine if age was a factor in menopause, the frequency of alleles from the two control groups were compared with one another. Results indicated no difference in proportions (χ^2=2.17; ns). Consequently, allele frequencies from controls were pooled. Allele type frequencies were compared among the 3 groups. Given the sparse 3X5 matrix, pseudo-Bayes estimators were computed to smooth the data in the table. Although it appears in only 6% of the population, the B allele in the familial group occurred unusually often. Thus, odds ratios (OR) were computed to compare the frequency of the B allele with A, C, and D, in the familial versus control groups. Similarly, ORs were computed for the B allele in the sporadic versus control groups. Results show large and significant ORs for the B allele in the familial group compared to controls. ORs of the B allele in the sporadic group compared to controls were not significant. These results suggest that POF is in linkage disequilibrium with FMR1 at the FRAXAC1 locus.

9

Radiation hybrid (RH) mapping: selective models and analysis of a small number of genes

G. R. Svischeva
Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, Novosibirsk, Russia

We developed special statistical methods for mapping of combined group of genes including both isozyme and DNA markers. Our models simulating biological process of obtaining RH clones allow stringent and weak retention of a selective gene in the clone. These models have pseudo-Markovian and additive properties and are applicable for analysis of an arbitrary number of genes. In a number of cases (especially in analysis of mammals poorly studied genetically), we collide with a problem of RH mapping of a small number (3, 4 or 5) of genes. At the decision of this problem, localization of structural genes coding for known protein products is of special interest.

The problem of mapping of a small number of genes needs to be considered as a independent statistical problem, as far as there is a number of difficulties in estimating of relative distances between the genes (with which we do not collide at localisation of a large (6 and more) number of genes) and, on the contrary, there is a opportunity of use of criteria for choice of the correct order, not acceptable for a large number of genes. At the analysis of 3, 4 and 5 genes under stringent selection, the number of various types of clones is not enough, that leads to low informativity of an empirical material. In this case, it is impossible to estimate distances between the genes without introducing additional restrictions on the model parameters. However, due to introducing these restrictions, the model may don't reflect the real situation. We have analysed each of possible orders of 3 and 4 genes and solved analytically the problem on ordering genes. Besides, we have obtained analytically estimates of functional dependencies formed by parameters at the description of the model. Thus, the problem of mapping of a small number of genes contracts to ordering genes and to qualitative comparison of distances between genes.

10

The vitamin D receptor FokI start codon polymorphism and bone mineral density is osteoporotic postmenopausal French women

G. Lucotte[1], G. Mercier[2], A. Burckel[2], M. Dougados[3], and C.Roux[3]
[1]Center of Molecular Neurogenetics, Reims, [2]European Laboratory of Prevention and Screening, Paris, [3]Evaluation Center of Bones Diseases, Cochin Hospital, Paris, France

This study examined the association between bone mineral density (BMD) and a T/C polymorphism in the first of two start (ATG) codons in the vitamin D receptor (VDR) gene. The polymorphism was detected using the restriction enzyme *Fok*I, the *F* allele indicating absence of the first ATG and the *f* allele indicating the presence of the first ATG. The *Fok*I genotype was determined in 124 postmenopausal osteoporotic French women who were 45 to 90 years old. The distribution of *Fok*I genotypes in osteoporotics didn't differ significantly from that found in a group of French random blood donors. There were no significant differences by *Fok*I genotype group in the total sample of osteoporotic women for age, years since menopause, height, weight, and BMD at lumbar spine and femoral neck. However, when only those patients under the age of 75 years are analysed (98 subjects, those with the *ff* genotype (10% of the population) had a significant lower BMD at the femoral neck than *FF* and *Ff* subjects. We conclude that the unfavourable *ff* genotype of the VDR gene correlates with decreased BMD at the femoral neck in French postmenopausal women.

11

Excess of early-onset cancers in relatives of neuroblastoma patients

A. Chompret[1,2], L. Brugières[2], F. de Vathaire[1], A. Abel[1], M-A. Raquin[1,2], O. Hartmann[2], J. Feunteun[3], C. Bonaïti-Pellié[1]
[1]Unit of Cancer Epidemiology (U351 INSERM), [2]Deparment of Pediatrics, [3]Unit of Genetic Oncology (UPR1599 CNRS), Institut Gustave Roussy, Villejuif, France

Within the frame of the French study on genetic predisposition to childhood cancer, the family history was investigated among 426 cases of neuroblastoma treated in the Department of Pediatric Oncology of the Institut Gustave Roussy since 1950. Cancer occurrence was registered in their 8558 relatives including 1485 1st-degree, 3654 2nd-degree and 3319 first cousins. Five families (1.2%) display another case of neuroblastoma, one among first-degree relatives, two among second-degree relatives and two among first cousins, which contrasts with the figure (5% in first-degree relatives) usually reported in the literature. Using standardized incidence ratio (SIR), an excess of relatives affected by early-onset cancer (before or at the age of 45) was looked for. The computation of person-years was restricted to the period posterior to 1969, since the information might be unreliable before this date. Besides, parents and grandparents were taken into account only since their age at conception of proband (for parents) or parent (for grandparents), because they had a low probability of being affected before they had their child. The overall SIR was 1.5 (CI 95% 1.1-1.9) and was not different according to the sex of relatives nor to the stage of tumor). Considering the age at diagnosis of proband, the SIR was only slightly higher (1.6) among relatives of cases occurring after 1 year than the SIR among relatives of cases occurring before 1 year (1.3). These results remain unchanged when the only case with a p53 germline mutation and the five cases with a familial neuroblastoma were excluded. This highly suggests the existence of genetic factors, still not identified, predisposing to neuroblastoma and also to other cancer types.

12

Non-parametric methods for localizing genes for complex traits using ancestral haplotypes

RB Martin and CJ MacLean
Virginia Commonwealth University, USA

We present a non-parametric method that systematically scans a chromosomal region for disease mutations embedded in ancestral haplotypes. As the disease mutation is passed down from founder individuals, recombinations and mutations at marker loci alter the nature of the surrounding ancestral haplotypes. We categorize the resulting haplotype fragments using a method previously developed by us—trimmed haplotype analysis—and estimate a haplotype's contribution to disease from its pattern of IBD sharing among a pedigree's affected individuals. The presence of an ancestral haplotype is suggested when ancestral-like haplotypes have high sharing scores. Empirical p-values are bootstrapped from the data using a rapid between-family haplotype permutation scheme, which preserves linkage information but randomizes correlations between haplotypes and their sharing scores.

Our method requires a grid of closely spaced markers and individuals' haplotypes. Map distances and marker order are not required, but are used if known. Our method extracts full information from multiplex

pedigrees, identifies carriers of the disease mutation by the fragments of ancestral haplotype they possess, does not require assumptions regarding mode of inheritance of the disease, allows for the confounding effects of marker mutations, and operates efficiently in the presence of heterogeneity. We have begun implementing these ideas in a FORTRAN-based software package HAL (Haplotype ALgorithm).

13

Interaction between dysplastic nevi and p16 in American melanoma-prone families.

A.M. Goldstein[1], M. Martinez[2], M.A. Tucker[1], F. Demenais[2].
[1]Genetic Epidemiology Branch, National Cancer Institute, Bethesda, Maryland, USA, [2]INSERM U 358, Hôpital St-Louis, Paris, France.

The p16 (or CDKN2A) gene, implicated in malignant melanoma (MM) pathogenesis, negatively regulates cell growth by arresting cells at G1 of the cell cycle. Although p16 mutations confer substantial risk for MM in about 25% of melanoma-prone families, sun-related covariates including dysplastic nevi (DN), total nevi (TN), and solar injury (SI) also influence disease expression. To examine the relationship between p16 and these factors, we conducted combined segregation/linkage analysis using the class D regressive logistic model taking into account variable age at onset of disease, as implemented in the program REGRESS. Genetic and covariate data were collected on 20 American MM-prone families, 13 of which had co-segregating p16 mutations. To deal with missing data, we created two dummy variables: a missing-value indicator set to 1 for unknown and 0 for known, and a second variable set to 1 for exposed and 0 for unexposed plus unknown. We assumed that ascertainment would not alter the interactions and conducted a joint likelihood analysis. Overall, there was a significant improvement in the likelihood when DN (χ^2_2=51.9, p<.001), TN (χ^2_2=24.6, p<.001), or both (χ^2_4=63.8, p<.001) were added to the model. In contrast, inclusion of SI (χ^2_2=4.8, p=.09) was not significant. A significant interaction was detected between DN and p16 when DN was the only covariate in the model (χ^2_2= 4.7, p=.03). Interestingly, the estimate of the regression coefficient (β) was greater in subjects without p16 mutations (3.3) versus those with mutations (2.2). The DN-p16 interaction was no longer significant when TN was added to the model. The latter result raises the problem of adequate power to detect gene-environment interactions, which will be further investigated through simulations.

14

Correlates of obesity display evidence for linkage to chromosomal regions 1p and 8p in familial combined hyperlipidemia (FCHL) pedigrees

CJ.M. Van Der Wallen[1] R.M. Cantor[2,3] M.M.J. Van Greevenbroek[1] C. Lanning[2], J. Geurts[1] C. Pette [1] W.Buurman[1] B. Aouizerat[3] A.J. Lusis[3], J.I. Rotter[2,3] and T.W.A. deBruin[1].
[1]Univ. of Maastricht, Netherlands [2]Cedars-Sinai Med. Ctr, [3]UCLA, Los Angeles, USA.

Murine and human studies of leptin and tumor necrosis factor α (TNFα) activities indicate these traits are increased in the presence of obesity; however, they may be obesity markers rather than causal. While structural genes for leptin and TNFα are known, genes which contribute to their expression in adipose tissue are not. To identify 'obesity' genes, body mass index (bmi), leptin and TNFα activity (soluble p55 serum levels) were measured in 200 siblings (364 pairs) from 49 nuclear families in 35 Dutch FCHL pedigrees with 10% having bmi>30. As expected, bmi correlated with log(leptin) (r=.52, p<.0001) and log (p55) (r=.28, p<.02), which were also correlated (r=.28, p<.03). A multipoint genome scan (markers every 10 cM) analyzed by MAPMAKER/SIBS Haseman-Elston (HE) (LOD scores) and the two point HE method of SAGE (p-values) identified two chromosomal regions with evidence for linkage (set at LOD> 2.0).

Trait	Region = 1p	Region = 8p
leptin	LOD=3.2, p<.00002	LOD = 0.0, p is ns
p55	LOD=0.0, p is ns	LOD = 3.0, p<.005
bmi	LOD=1.2, p<.002	LOD = 0.6, p is ns

Linkage to the same 8p region was reported for leptin in Mexican-Americans. (Nat Gen 3/97, p273). While separate genes contribute to leptin and p55 activities in the FCHL population, pathways to obesity vary among populations at risk for other metabolic disorders.

15

Heterogeneity in the relation of dopamine genes to ADHD: Application of a logistic regression extension of the TDT.

WALDMAN ID[1], MILLER MB[2], ROBINSON BF[1], and ROWE DC[3].
[1]Emory University, Atlanta, GA, USA; [2]University of Missouri, Columbia, MO, USA; [3]University of Arizona, Tucson, AZ, USA.

The Transmission Disequilibrium Test (TDT) is a simple test of linkage disequilibrium between a candidate gene and a disorder that is robust to population

stratification and has considerable statistical power. Despite these advantages, one 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 a logistic regression-based extension of the TDT to examine the relation between a candidate gene and one or more continuous or categorical variables. We illustrate the application of this method to issues of genetic heterogeneity, including differences in linkage disequilibrium by age, sex, or symptom severity and type. We used data on symptoms of childhood Attention Deficit Hyperactivity Disorder (ADHD), Oppositional Defiant Disorder (ODD), and Conduct Disorder (CD), and the dopamine transporter gene (DAT1) and dopamine receptor D4 gene (DRD4) from 122 probands and their parents and siblings. DAT1 was related strongly and linearly to the number of hyperactive-impulsive symptoms and less strongly to inattentive symptoms. The relation of DAT1 with ADHD was much stronger in boys than girls and in older than younger children. The number of ODD and CD symptoms also were related linearly to DAT1, suggesting that DAT1 also influences childhood antisocial behavior. In contrast, DRD4 was related linearly to the number of inattention but not hyperactive-impulsive symptoms. The relation of DRD4 with inattention did not differ by sex but was stronger in younger than older children. The number of ODD and CD symptoms were not related to DRD4. These results demonstrate the utility of the logistic regression extension of the TDT for revealing heterogeneity in the relations between candidate genes and complex traits.

16

A likelihood-based approach to linkage disequilibrium testing

D.E. Goldgar.
Unit of Genetic Epidemiology, International Agency for Research on Cancer, Lyon France.

The Transmission Disequilibrium Test (TDT) was proposed in 1993 by Spielman et al. as a means of testing for association by comparing the frequency of transmitted and non-transmitted alleles. More recently, these and other investigators have extended the method to cases of missing parental data, multiple alleles, multiple affected siblings, and using sibling controls. In a previous paper (Goldgar and Fain; 1984) we demonstrated that incorporation of linkage disequilibrium (LD) into standard linkage analysis provided greatly increased power when both linkage and LD were present and postulated that the approach could be used as a test of association when no population controls were available. Here we demonsrate that all of the TDT-like tests can be performed using PAP assuming a low penetrant rare recessive with no phenocopies. In the simplest case, the test is the likelihood ratio test equivalent to the Mcnemar Chi-square statistic obtained from the TDT. However, any pedigree structure with at least one affected individual can be analyzed in this manner. Within this framework, the test performed is equivalent to testing for a difference in the marker allele frequencies conditional on the two disease alleles. We have examined the utility and power of this approach using four pedigree structures analyzed under 3 genetic models, and assuming complete and partial disequilibrium. Preliminary simulation results show the approach is conservative, that power decreases dramatically under partial LD, and that a design with genotypes on two affected and one unaffected sib provides considerable advantage.

17

Risk of prostate cancer associated with a family history: a population-based case-control study.

Staples M[1], Giles GG[1], Hopper JL[2]
[1]Anti-Cancer Council of Victoria, Australia.
[2]University of Melbourne, Australia.

Having a close relative with prostate cancer (PC) has been associated with an increased risk of this disease. To quantify this risk as a function of degree of relationship and age at onset in the case, we used data on family cancer history from a population-based study of 915 cases and 854 controls conducted in Melbourne, Australia, between 1994 and 1998. Incident cases of a first primary invasive adenocarcinoma of the prostate, diagnosed between 1994 and 1997, in men aged 40 - 74, were identified from the Victorian Cancer Registry. Controls were identified from the Electoral Roll (registering for voting is compulsory in Australia). A self-report of family history of PC was asked for all first-degree relatives and paternal and maternal uncles.

Fifteen percent of cases reported affected first-degree relative(s) only, and 17% at least one affected first-degree relative or uncle, compared to 6% and 7%, respectively, for controls ($p < 0.01$). A total of 21 cases, but no controls, had families fitting the proposed clinical criteria for hereditary PC; namely 1) $\geq$ 3 affected individuals in one nuclear family; 2) affected individuals in three successive generations; or 3) $\geq$ 2 relatives each affected before the age of 55.

Case-control analyses, adjusting for age, showed that having any affected first-degree relative was associated with a 3.2-fold (95% CI: 2.3 – 4.6) increased risk of PC. The risk was similar for an affected father (3.5; 2.2. – 5.4) as it was for any affected brother (3.4; 2.0 – 5.8). The risk associated with an affected uncle only was 2.2 (1.0 – 4.7). Having both a first degree relative and an uncle with PC increased risk to 9.7 (2.2 – 42). The

risk associated with any affected first-degree relative decreased with age at diagnosis (trend p = 0.01), from 5.0-fold (2.7 – 9.2) to 3.0 (1.5 – 6.1) to 2.2 (1.3 – 3.8) for the age groups 40-59, 60-64 and 65-74, respectively.

18

Modeling of HLA Class II susceptibility to type 1 diabetes reveals important role of DPB1.

A.M. Valdes,[1] J.A. Noble,[2] E. Génin,[1] F. Clerget-Darpoux,[3] H.A. Erlich,[2] and G.Thomson[1]
[1]Dept Integrative Biology, 3060 VLSB, Univ. California, Berkeley, CA 94720-3140, USA; [2]Dept Human Genetics, Roche Molecular Systems, Alameda, CA;Children's Hospital Research Institute, Oakland CA, USA; [3]INSERM U. 155, Paris, France.

The association between HLA DR-DQ and type I diabetes (IDDM) has been modeled under the assumption that DR-DQ represent the sole disease loci in the HLA region. Two main classification schemes were used: [1] subdividing index cases into DR3, DR4 and other (X), and [2] subdividing X types into intermediate (I), and protective (P). The Marker Association Segregation Chi-Square (MASC) method was used to test these models in 257 affected sib-pairs from the Human Biological Data Interchange. The test hypothesis that DR-DQ accounts for all of the HLA component to IDDM was rejected because it could not explain the identity by descent (IBD) distribution. To examine the contribution of DPB1 the data were split into two groups: those carrying neither DPB1*0301 nor DPB1*0202 (group A), and those carrying at least one copy of DPB1*0301 or DPB1*0202 (group B). The hypothesis that DR-DQ by itself accounts for the IBD distribution was accepted when the DR3, DR4, I, P scheme was used on group A ($p<0.17$), but rejected on the same set of genotypes with the DR3, DR4, X scheme. The model was strongly rejected among group B cases. Interestingly, group A and group B display very similar frequencies of DR3, DR4, I and P types but have significantly different IBD distributions. We also used a model that incorporates the DPB1 effect. This model was accepted on the complete data set. The existence of evidence from other sources linking DPB1*0301 with increased IDDM susceptiblity speaks in favor of direct involvement of this allele in IDDM predisposition.

19

Genetic epidemiology of the atherogenic Lipoprotein(a)

H.G. Kraft, G. Utermann
Institute for Medical Biology and Human Genetics, Schöpfstr. 41, A-6020 Innsbruck, Austria

Lipoprotein(a) [Lp(a)] is a complex consisting of an LDL particle to which a glycoprotein designated apolipoprotein(a) [apo(a)] is bound via a disulfide bridge. High levels of Lipoprotein(a) are considered as genetic risk factor for premature Coronary Heart Disease. Lp(a) plasma levels vary over 1000-fold between individuals but are extremely constant in a single individual. Sib-pair and twin studies have revealed that the variation of Lp(a) concentrations is almost completely determined by one major gene which is the structural gene for apo(a). We have performed a large genetic epidemiological study relating apo(a) gene variation to Lp(a) levels in several ethnic groups.

The most prominent feature of apo(a) is its size polymorphism which is explained level by a variable number of plasminogen like Kringle-IV repeats in the coding sequence of the gene. In each population this polymorphism excerts a major effect on Lp(a) levels. Alleles with low numbers of K-IV repeats are associated with high Lp(a) levels and vice versa. As a mechanism longer retention times in the ER have been shown for large apo(a) isoforms which make them prone to proteolytic events.The fraction of the variance that is explained by variation in K-IV numbers differs between populations ranging from 28 % in Asian Indians to 76 % in Thai. Although the type of the association is the same in all populations differences in K-IV allele frequencies do not explain differences in Lp(a) concentrations between populations.

A polymorphism with 6 to 11 TTTTA repeats exists in the 5´ untranslated region at -1.3 kb from the transcription start in the apo(a) gene. The effect of this polymorphism on Lp(a) concentrations is present only in some populations and is rather small. Moreover, the type of the association is different between these populations (Chinese, Caucasians, Indians, Japanese). Hence we conclude that this polymorphism has no direct effect on Lp(a) levels and that associations result from linkage disequilibria in certain populations.

Finally, a C/T polymorphism that creates an additional start codon and thereby reduces apo(a) translation *in vitro* was studied. In all populations, mean and median Lp(a) concentrations were lower in CT heterozygotes compared to CC homozygotes. This difference was significant only in African populations (Blacks and KhoiSan). In all but in the KhoiSan strong linkage disequilibria were detected between the C/T polymorphism and the two other polymorphisms which likely explain the lack of a significant effect in some populations. Therefore linkage disequilibria may not only produce artificial associations (e.g. TTTTA polymorphism) but also mask true effects.

20

Genomic screen for linkage in a family with autosomal dominant chordoma

J.F. Korczak[1,2], M.J. Kelley[1], K.A. Allikian[1], A.A. Shah[1], A.M. Goldstein[1], and D.M. Parry[1].
[1]National Cancer Institute, Bethesda, MD, USA,
[2]Georgetown Univ. Medical Center, Washington, DC, USA.

Chordoma, a rare, low-grade, malignant bone tumor derived from remnants of the notochord, is usually sporadic. Only three multiplex families, each with 2-3 affected relatives, have been reported. We evaluated a family with 10 affected individuals in three generations, including two father-son pairs, consistent with autosomal dominant inheritance. Diagnosis was based on histopathology or a mass typical of chordoma on MR scans of the skull base and spine. We genotyped 22 family members for 365 STR autosomal markers spaced about 10 cM apart, using the CHLC version 8 set. Two-point lod score (using LINKAGE) and affected sib-pair and Haseman-Elston regression (using SIBPAL) analyses were performed based on (i) only the 10 affected individuals and (ii) all 22 family members, with disease allele penetrances varying from 50-100%. Initial findings suggested possible linkage of the disease locus to a region on chromosome 7, 17, or 19, based both on parametric (lod scores of 1.0-2.2) and nonparametric (nominal P-values ≤ 0.01) results at each of two adjacent marker loci. Additional markers were typed in the three regions to attain a 2 cM map for further analyses. Chromosome 7 gave the greatest evidence for linkage, where maximum two point and multipoint lod scores of 2.3 and 2.4, respectively, and P-values of about 0.01 were obtained in the affecteds-only analysis. All affected family members shared a common haplotype at markers in a 23.6 cM region. We are sequencing candidate genes in the region of interest and initiating a protocol to accrue additional families in order to identify the chordoma gene.

21

Bootstrap confidence intervals for relative risks in ASP data

Heather J. Cordell[1] and James R. Carpenter[2]
[1]Case Western Reserve University, Cleveland, Ohio, USA
[2]London School of Hygiene and Tropical Medicine, London, UK.

In affected-sib-pair (ASP) studies, locus-specific sibling relative risks (λ_s) are often estimated and used to decide whether to continue the search for susceptibility genes. Typically, a point estimate of λ_s is given, but since this estimate may have substantial variance, it is of interest to obtain confidence limits for the true value of λ_s. Cordell and Olson (1997) proposed two methods for doing this, both of which rely on assumptions which are valid asymptotically as the number of families tends to infinity, but which may not hold in practice. In recent years, many simulation-based techniques for constructing confidence regions have been developed, most being based on a resampling or bootstrap approach. We have conducted simulations to investigate the properties of the most popular bootstrap methods for confidence interval evaluation compared to the asymptotic methods. The aim is to identify from the large pool of methods available, those which yield short intervals with accurate coverage probabilities for ASP data. ASP likelihoods have some unusual features due to the discrete nature of the data and the imposition of genetic possible triangle constraints during the maximization. We find in our simulations that many of the most popular methods of bootstrap confidence interval evaluation perform poorly for ASP data, giving coverage probabilities much lower than claimed. The test-inversion, profile-likelihood and asymptotic methods, however, perform well although some care is needed in choice of nuisance parameter in order to obtain accurate coverage.

22

Dissecting complex diseases with FINESSE

A BIOMED EC-funded project involving these groups:
Oxford: J. R. O'Connell[1,2], N. Margetic[1], M. Farrall[1], G.M. Lathrop[1], D. E. Weeks[1,2]
Paris: F. Demenais[3], M. Martinez[3], G. Durrieu[3]
Berlin: I. Koch[4], K. Rohde[4]
[1]Wellcome Trust Centre for Human Genetics, University of Oxford, UK
[2]University of Pittsburgh, USA
[3]INSERM U. 358, Hôpital St-Louis, Paris, France
[4]MDC, Berlin, Germany

Identifying genes in complex diseases requires models and programs that handle multilocus disease models, environmental effects, multiple markers, and covariates which may interact with the genetic factors. We are developing new software called FINESSE which integrates the VITESSE likelihood algorithm, the REGRESS program, and state-of-the-art optimization algorithms. This software will be organized in a modular way to enable sophisticated users (such as IGES attendees) to modify and extend it as needed. We will also create user-friendly graphical (and textual) interfaces which will facilitate complex and powerful analyses including segregation studies, combined linkage and

segregation analyses, semi-parametric mod score analyses, association studies with candidate genes, and analyses of gene-gene and gene-environment interactions.

23

Familial risks in cancers from a Family-Cancer Database

Kari Hemminki and Pauli Vaittinen
Department of Biosciences at Novum, Karolinska Institute, 141 57 Huddinge, Sweden.

Exact quantification of familial cancer risks is important for clinical, psychological and scientific reasons. The current estimates of familial relative risks (FRRs) of cancer have many uncertainties due to sample size and possible bias in data collection. FRRs often refer to the first-degree relatives of unspecified age and sex, obscuring the risk estimates.We calculated sex-and age-specific FRRs of cancer in offspring of cancer probands at 19 male 20 female cancer sites, based on registered nation-wide data, free from bias. We use the Family-Cancer Database from Sweden, which was constructed from a national cancer registry and a family registry identifying the parents of the offspring born after 1941. The Database contained 550 000 primary cancers. FRRs were calculated from the age-adjusted incidence rates for the offspring.The familial risks at known sites: colon, rectum, breast, ovary, testis, skin (melanoma), nervous system, thyroid and other endocrine glands were confirmed. However, the FRR for breast was lower than that reported in the literature. The FRR of male breast cancer was somewhat higher than that of female breast cancer. The FRR of thyroid cancer, exceeding any cancer, was over two times higher for the male than female offspring, and appeared to constitute an early- and late-onset component. Novel register-based findings were familial risks in cervical and uterine cancer, and in male offspring of male probands kidney and skin (mainly squamous cell) cancer. Familial risks were noted also for lung cancer, lymphoma and leukemia but they may have environmental causes. Because of late onset, analysis of prostate cancer showed no familial effect. The proportion of familial cancers depended on the site, ranging from 11% in prostate to 8.7% in female breast and to well below 1% at many sites.

24

Segregation Analysis of IgE Levels in 335 French Families (EGEA) using different strategies to correct for the ascertainment through a correlated trait (asthma).

Sampogna F[1], Demenais F[2], Hochez J[1], Feingold J[1], Dizier MH[1]
[1]INSERM U155, Paris, France [2]INSERM U358, Paris, France

Genetic factors are known to play a role in asthma and allergy, but the exact nature of this genetic component is still unclear. Study of intermediate phenotypes associated with asthma and allergy, such as serum IgE levels, may help in dissecting the genetic mechanisms underlying these diseases. We conducted segregation analysis of IgE levels in 335 French nuclear families of the EGEA study (Epidemiological study of the Genetics and Environment of Asthma), ascertained through an asthmatic proband (123 parents and 212 children). Different strategies were considered to correct for this mode of ascertainment: A) no correction was applied; B) ln(IgE) levels were adjusted for a family position effect defined as being a proband, a blood-relative or a spouse; C) the asthmatic children-probands were excluded and the likelihood of each family was computed conditionnally on the parents' IgE levels. The class D regressive models, as implemented in the computer program REGRESS, were used to search for a major gene effect while taking into account residual familial correlations and covariates (age, sexe, smoking habits). Whereas a major gene effect could not be detected with strategy A, strategy B and C showed evidence for the transmission of a dominant major gene, which was more significant with strategy B. This gene does not interact with any of the covariates and is responsible for about 15% of IgE variation (the allele frequency is 0.65). A genome wide search, currently in progress, may lead to further identify this genetic component.

(supported by convention INSERM/MSD)

25

Risk Models for Familial Breast and Ovarian Cancer

A.C.Antoniou[1], D.F.Easton[1], S.A.Gayther[2], J.F.Stratton[2], B.J.Ponder[2].
[1]CRC Genetic Epidemiology Unit, Institute of Public Health and [2]CRC Human Cancer Genetics Research Group, Univ Cambridge, UK

We investigated risk models for the inherited susceptibility of breast and ovarian cancer, using data from both high-risk families and a population based series of ovarian cancer. The first data set consisted of 112 families containing 2 or more relatives with epithelial ovarian cancer. BRCA1 and BRCA2 germline mutations were detected in 50% of these families. The second study involved 374 ovarian cancer cases, collected at the Royal Marsden Hospital, London, who had DNA samples analysed for BRCA1 mutations. 12 women were found to be carriers. We constructed genetic mod-

els for ovarian and breast cancer using the computer program MENDEL. In the first study we modelled the effects of BRCA1 and BRCA2 simultaneously and allowed for a third gene predisposing to ovarian cancer. None of the models fitted gave significant evidence for a third gene. Population frequencies of BRCA1 and BRCA2 mutations were estimated to be 0.00128 and 0.00172 respectively. Our results suggest that BRCA1 and BRCA2 may be sufficient to explain the majority of familial ovarian cancer and that families without mutations can be explained by sensitivity of the mutation testing and chance clusters of sporadic cases. Using data on the families of the 12 mutation carriers in the second study, we estimated age specific ovarian and breast cancer risks for BRCA1 mutation carriers. Under the best fitting model the cumulative ovarian cancer risk was 68% by age 70, and the corresponding breast cancer risk was 50%. The high penetrance estimate for ovarian cancer, in comparison with other studies, suggests that modifying genetic or environmental factors may be important determinants of risk.

26

Parametric and non-parametric multipoint linkage analysis for two-locus disease models

K. Strauch, R. Fimmers, and M.P. Baur
Institute for Medical Statistics, University of Bonn, Germany

An extension of the Lander-Green algorithm for genetic linkage analysis is presented which incorporates two-locus models of disease. The original algorithm, base of the linkage analysis software package GENEHUNTER developed by Kruglyak et al., performs parametric and non-parametric linkage analysis for multiple marker loci. In our current extension, parametric analysis allows for exact LOD score calculation given a specific two-locus disease model, i.e., allele frequencies and penetrances. Non-parametric linkage analysis (NPL) carries out allele sharing statistics at two loci for a certain scoring function, without need to specify a mode of inheritance. The approach combines the advantage of modeling genetically complex diseases in an appropriate, more realistic way with the extraction of as much inheritance information of a pedigree as possible by multi-marker analysis.

27

Genetic analysis of antibody responses to specific malaria antigens in Papua New Guinea.

H. A. Stirnadel[1], H.-P. Beck[1,2], M. P. Alpers[2] and T. A. Smith[1,2]
[1]Swiss Tropical Institute, Basel, Switzerland, [2]Papua New Guinea Institute of Medical Research, Madang, Papua New Guinea.

The contribution of genetic factors to immune responses to malaria is increasingly recognized, with implications for disease control strategies such as vaccine development. We therefore explored the familial pattern of inheritance in antibody responses to specific malaria antigens in an area highly endemic for malaria in Papua New Guinea. Previous analysis in this population suggested that both environmental and genetic components affect total IgG responses against RESA (ring-infected erythrocyte antigen) and MSA-2 (merozoite surface antigen). We have now extended the analysis to assess the genetic and environmental contribution to variation in IgG subclass responses to the same antigens. Overall, familial aggregation was found for IgG1, IgG2 responses against RESA, IgG1, IgG3 responses against MSA-2 (3D7) and the IgG2 response against MSA-2 (FC27). Further analysis indicated that allowance in the model for neither sharing of houses nor sharing of HLA haplotypes could explain the genetic variance. Preliminary segregation analysis suggested that genetic regulation was more complex than a single major gene. Our findings demonstrate the presence of familial aggregation of antibody responses to certain malaria antigens, but the underlying mechanisms need further clarification.

28

Analysis of linkage to HLA DRB1 and TNF in rheumatoid arthritis families; a comparison of three TDT methods

S John, J Worthington, JH Barrett
ARC Epidemiology Research Unit, University of Manchester, UK

The transmission/disequilibrium test (TDT) can be used to confirm linkage when evidence of association has been found from case-control data, thus ruling out the alternative explanation of confounding by population stratification. Several different extensions of the TDT for polymorphic loci have been proposed. We analysed data on a set of rheumatoid arthritis (RA) cases, each with at least one parent typed. Two genes known to be associated with RA were investigated: the HLA DRB1 gene (112 cases available), and a microsatellite marker for tumour necrosis factor (TNFα, 58 cases). Three different methods were used: (i) a logistic regression modelling approach, with one parameter for each allele describing its preferential transmission to affected offspring[1]; (ii) another likelihood-based approach, with one free parameter measuring the association between one (unspecified) allele and the disease

locus[2]; (iii) a score test based on the likelihood conditional on parental genotype[3]. The methods differ in their approach to missing parental data. In each case there was evidence for linkage to DRB1, but the evidence was stronger using the first ($p<0.0001$) and third ($p=0.0001$) approaches and surprisingly weak using the second method ($p=0.03$). For TNFα no evidence for linkage was found using the last two methods, but there was some evidence using the logistic regression approach even in this relatively small data set ($p=0.05$ using the chi-squared approximation, $p=0.08$ using Monte-Carlo simulation to estimate the p-value). This analysis suggests that the methods may vary considerably in their power to detect linkage. Simulation studies are required to examine their performance in detail.

[1]Sham & Curtis, 1995. Ann Hum Genet, 59: 323-336.
[2]Terwilliger, 1995. Am J Hum Genet 56: 777-787.
[3]Clayton. Transmit v 2.1. MRC Biostatistics Unit, Cambridge, UK.

29

Parental genotype reconstruction and the sibship test for linkage (S-TDT)

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

The transmission/disequilibrium test (TDT) introduced by Spielman et al. (Am J Hum Genet 52:506-516, 1993) is a simple and powerful method to detect linkage between a marker and a disease susceptibility locus in the presence of linkage disequilibrium between both loci. The TDT requires the affected offspring as well as their parents to be typed at the marker locus. The availability of parental marker genotypes can pose a problem, especially when the disease of interest has a late age of onset. For this reason, Spielman and Ewens (Am J Hum Genet 62:450-458, 1998) recently proposed a method called S-TDT which does not require parental marker genotypes, but uses marker genotypes of unaffected siblings. For some families without parental genotype information, it can be possible to reconstruct parental genotypes from the genotypes of their offspring. It may be tempting to treat these reconstructed families as if parental genotypes had been typed. But Curtis (Ann Hum Genet 61:319-333, 1997) already showed that such a procedure can introduce bias. Curtis indicated that correcting this bias would require knowledge of population marker allele frequencies. Indeed, such reliance on population frequencies is not opportune, since a key benefit of the TDT would be lost in that case. On the other hand, deducing parental genotypes when possible is a quite natural and attractive approach for a geneticist. This paper shows a way to reconstruct parental genotypes, but nevertheless avoids the bias described above without depending on marker allele frequencies. Expressions are presented for the conditional null expectation and variance of the number of alleles A in affected children, given that parental genotypes can be reconstructed. With these expressions, a modified S-TDT is obtained and applied to the same two data sets as used by Spielman and Ewens. Finally, the modified TDT and the S-TDT are compared by means of a simulation study.

30

Designing a linkage replication study in affected sib pairs

Eichenbaum-Voline S[1,2], Babron MC[1], Prum B[2], Clerget-Darpoux F[1].
[1]INSERM U155, Paris, [2]CNRS URA1323, Paris, France.

The Maximum Likelihood Score (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. Systematic screening of the genome is prone to false positives due to multiple testing. To avoid wrong conclusions, one may choose a stringent threshold for concluding to linkage, but then decreasing the power to detect susceptibility factors. An alternative is to choose a lax threshold in the first sample, then to study a second sample only for the retained regions. In this second study, the threshold Tα corresponding to a given type I error α will depend on the number of regions re-tested.

However, even if a risk factor exists, the MLS value may vary greatly from one sample to another. For different genetic models and a given number of affected sib pairs, we calculated the 95% tolerance interval of the MLS and showed it may be very large. For example, for a IBD distribution equal to (0.18, 0.43, 0.39), the possible values for the MLS vary between 0.39 and 6.14 for a sample of 100 affected sib pairs.

To confirm the presence of a risk factor in the second sample with a given power p and a given type I error α, one calculates the minimum sample size such that the probability for the MLS to exceed Tα is greater than or equal to p.

In real situations, the underlying genetic model is unknown. Then it is possible to estimate this minimum sample size by bootstrapping the first sample. This replication study design will be illustrated on celiac disease.

31

Glutathione S-transferase M1, M3, P1 polymorphisms and lung cancer: a case-control study in Caucasian smokers.

N. Jourenkova[1], H. Wikman[2], C. Bouchardy[3], A. Voho[2], P. Dayer[4], S. Benhamou[1], and A. Hirvonen[2]. [1]INSERM U.351, Villejuif, France, [2]Finnish Institute of Occupational Health, Helsinki, Finland, [3]Geneva Cancer Registry, Geneva, Switzerland, [4]University Hospital of Geneva, Geneva, Switzerland.

Glutathione S-transferases (GSTs) are involved in the detoxification of active metabolites of carcinogenic polycyclic aromatic hydrocarbons (PAHs) of tobacco smoke. Subjects from a case-control study of 150 patients with squamous or small cell carcinomas and 172 non-cancer controls, all regular smokers, were analysed for their *GSTM1, GSTM3,* and *GSTP1* genotypes to evaluate if these polymorphisms modulated lung cancer risk. The *GSTM1* gene was deleted in 54.0% of cases and 52.2% of controls. The frequency of *AA*, *AB* and *BB GSTM3* genotypes were 70.7%, 24.0%, 5.3% in cases and 72.7%, 24.4%, 2.9% in controls. The frequency of *AA*, *AG* and *GG GSTP1* genotypes 44.7%, 44.0%, 11.3% in cases and 50.0%, 37.2%, 12.8% in controls. No significant interaction was observed between different combinations of *GSTM1, GSTM3* and *GSTP1* genotypes. The analysis of interaction between *GST* genotypes and exposure to tobacco smoke showed an increase in lung cancer risk associated with *GSTM1*null *& GSTM3(AA) & GSTP1(AG+GG)* genotype in smokers with a history of at least 35 pack-years (OR=2.7, 95% CI=1.2-6.0), probably due to an association between *GST* genotypes and smoking among controls.

32

TNF & prolactin microsatellite polymorphism is not associated with susceptibility to SLE.

L. Zhang[1], H. Li[1], M.F. Seldin, and L.A. Criswell[2]. [1]UC Davis and [2]UCSF, San Francisco, CA, USA.

Objective: To determine whether TNF or prolactin polymorphism is associated with susceptibility to SLE utilizing the transmission disequilibrium test (TDT).

Background: TNFα and prolactin are 2 MHC region genes that are prime candidates for a role in SLE susceptibility. Recent studies utilizing traditional case control designs offer support for association of TNF with SLE, &/or ethnic or clinical subgroups.

Methods: We studied the following 3 microsatellite polymorphisms among 416 members of 148 SLE families utilizing parental and sibling TDTs: TNFα, TNFc, and D6S285 (prolactin). TDT methods retain the sensitivity and power of traditional case control designs, yet avoid spurious associations arising from population admixture.

Results: The results summarized in the table below (2-sided p values shown in cells) provide no evidence of association with SLE susceptibility. Similar results were obtained for Caucasian and nephritis subgroups.

	# families	TNFα	TNFc	D6S285
Parental TDT*	87	0.66	0.64	0.40
Sibling TDT†	61	0.39	0.32	0.58

*Score statistic used; p values calculated based on 1,000 simulations.

†Z_{max} statistic used; p values calculated using normal approximations with Bonferroni correction.

Conclusions: These results, which are based on highly sensitive methods that are free from spurious association arising from population admixture, do not support the hypothesis that polymorphism at the TNF or prolactin loci is associated with susceptibility to SLE.

33

An alternative test for linkage between a marker locus and a quantitative trait

V. George[1], H.K. Tiwari[2] and R.C. Elston[2] [1]Medical College of Wisconsin, Milwaukee, USA, [2]Case Western Reserve Univ., Cleveland, USA.

The transmission/disequilibrium test (TDT) proposed by Spielman, et al for qualitative traits and extended by Allison for quantitative traits is a powerful method for detecting linkage between a marker locus and disease trait in the presence of allelic association. As a test for linkage disequilibrium, however, the TDT makes the assumption that any allelic association present is due to linkage disequilibrium. The test cannot distinguish between linkage disequilibrium and, in the presence of linkage, allelic association due to other causes (such as population admixture or selection). Thus, the interpretation of the TDT, other than as a test for linkage in the presence of allelic association, is questionable. In this presentation we propose a linkage test for quantitative traits that does not make any assumption about the cause of allelic association in the population. We model the allelic association as a nuisance parameter and estimate it along with the other parameters in the model. We further investigate the statistical power of the test as a function of the allelic association and sample size using simulation, taking different causes of allelic association into consideration.

34

Segregation analysis of squamous cell carcinoma of the head and neck: evidence for a major gene determining risk.

M. de Andrade[1], C I Amos[1] and W D Foulkes[2] [1]Department of Epidemiology, UT-MD Anderson Cancer Center, PO Box 189, 1515 Holcombe Blvd, Houston, Texas, USA; [2]Departments of Medicine and Human Genetics, McGill University, Montreal General Hospital, Montreal, Quebec, Canada, H3G 1A4

We and others have shown that a family history of squamous cell carcinoma of the head and neck (SCCHN) is a risk factor for this disease. We performed a segregation analysis on a dataset of 1429 first-degree relatives of 242 unselected cases of SCCHN. Using the SAGE software, we demonstrated that a Mendelian model was favored and a model postulating a purely environmental cause of SCCHN was rejected. The model suggests that 18% of the population for those who smoke and drink are susceptible. The lifetime risk for non-smokers and non-drinkers who were heterozygotes for the susceptible allele was close to zero, but for those heterozygotes who smoked but did not drink the risk approached 60% by age 80. These findings suggest that specific genetic factors account for a significant fraction of the risk of SCCHN associated with a family history of SCCHN.

35

Increasing the power of family-based association studies

J.C. Whittaker[1] and C.M. Lewis[2]
[1]Department of Applied Statistics , University of Reading, U.K.; [2]Division of Medical and Molecular Genetics, UMDS, Guy's Hospital, London, UK

We investigate the power of tests based on the TDT for a number of possible family types, classified by the disease status of family members. We show that parental disease status affects the power greatly. Families containing a single affected parent are preferred over families in which neither parent is affected across a broad range of genetic models appropriate for complex traits. Families with a pair of affected sibs are of great value for all models considered, but extending the TDT to include information from unaffected sibs rarely increases power, provided that parents have been genotyped. The Sib-TDT (Spielman and Ewens, AJHG, 1998) uses genotypes from unaffected siblings in place of the parental genotypes used in the TDT. We show that the power obtained from N TDT trios (single affected offspring and two parents) is approximately equivalent to that from N(k+1)/k sibships with a single affected and k unaffected children in the Sib-TDT. Thus, for example, 2N discordant sib pairs are equivalent to N TDT trios. These results allow evaluation of the optimum number of unaffected sibs to genotype, and allow us to compare the increased power of TDT trios against the wider availability of sibships, particularly in late-onset disorders.

36

A Simple Preliminary Ordering Algorithm for Loci Identified in Overlapping YACs.

C.T. Falk.
The New York Blood Center, New York, USA.

A computer algorithm is presented that assists in the preliminary ordering of loci typed on a set of overlapping YACs. The method is similar to the minimum break ordering method used for radiation hybrid data, but includes steps that incorporate special characteristics of YAC fragments. These additions 1) allow for the incorporation of data on endpoint markers and 2) favor orders that allow for all loci present on a YAC to be part of a contiguous set. Simulated annealing is used to search the locus order space for the set of permutations with the smallest number of obligate breaks (or internal "gaps"). The orders identified can be further refined, based on additional testing or information derived from other sources and then used to align the YACs to form a contig. The method is illustrated with two examples. The first identifies a set of optimal orders for 25 markers on 41 YACs and cosmids containing material from chromosome 12q22. The second identifies optimal orders for 29 markers on 20 YACs and P1 fragments with chromosome 15q26.1 material. In both cases the computer predictions are in agreement with the orders determined in the original studies, except where those studies had access to information from other sources. In one case the computer program predicted an order leading to fewer obligate breaks than the previously determined order.

37

Genome-wide screening by homozygosity mapping: what set of markers to choose?

E. Génin[1], A.A. Todorov[2] and F. Clerget-Darpoux[1]
[1]INSERM U155, Paris, France; [2]Washington University School of Medicine, St Louis, USA

Mapping of genes involved in rare recessive diseases is usually difficult because of the lack of families with more than one affected progeny. This problem may be avoided by using inbred affected individuals and the strategy of homozygosity mapping.

In practice, the use of homozygosity mapping in a genome-wide scan requires that a set of markers regularly spaced and spanning the whole genome be tested. Investigators are then faced with the problem of choosing the spacing of markers.

To help solve this problem, we provide some guidelines by computing (1) the expected length of the region of identity by descent around the disease locus, (2) the distribution, given the spacing of markers, of the number of affected individuals expected not to be homozygous at the marker closest to the disease locus and, (3) the expected type-one error.

We show that, even if the markers are very closely spaced, there is a high probability for at least one affected individual in the sample not to be homozygous at the marker closest to the disease locus. For example,

with markers spaced 1 centimorgan apart, this probability is 14% (in a sample of 10 affected progenies of first-cousins) and reaches 52% with markers spaced 5 centimorgans apart. Excluding a region by the criterion that all affected individuals in the sample are not homozygous may then dramatically increase the rate of false negatives. We thus propose to relax the criterion used to declare a candidate region, based on the sample size and the spacing of markers.

38

Detection of polymorphic sites within genes : how many sites are useful for association studies?

E. Génin, A. J. Schafer
Hexagen, Cambridge, UK

Comparison of polymorphic sites within a gene between cases and controls may be useful for establishing a role for this gene in disease susceptibility. A subset of individuals (called the "detection sample") from a test population can be used to define the polymorphisms, followed by testing of select sites on the remainder of the test population. The choice of these select sites is problematic and will determine the power to detect the effect of the gene in disease predisposition. The power depends on the size of the "detection sample", on the proportion of select sites relative to the total number of polymorphic sites within the gene, on the allelic frequencies at the different sites and on the amount of linkage disequibrium between the sites.

Since little information is available on the amount of linkage disequilibrium that is likely to exist between polymorphic sites within a gene, theoretical predictions are difficult. We therefore adopted an empirical approach to assess the utility of testing a population with a subset of defined polymorphisms, estimating the amount of within-gene linkage disequilibrium using data on polymorphic sites identified within the insulin receptor gene in a sample of 86 unrelated individuals from the UK.

Assuming that one of the identified sites is involved in a trait, we evaluated the power of association tests using one, two, three,..., all polymorphic sites within the gene. Weighting the power against the cost of typing, we were able to determine the most efficient strategy for selecting the polymorphic sites to investigate the effect of this gene in this population. Attempts to generalise the results will require analysis of additional genes and populations.

39

Case-Only Design to Measure Gene-Gene Interactions

Quanhe Yang[1], Muin J. Khoury[1], Fengzhu Sun[2], W. Dana Flanders[2].
[1]Centers for Disease Control and Prevention (CDC), Atlanta, USA, [2]Emory University, Atlanta, USA.

Genetic studies have evolved from simple analyses of single genes to include more sophisticated analyses of complex traits, an evolution that parallel, an increasing recognition of the role of gene-environment interactions in disease etiology. Previous work has shown that the case-only design is an efficient and valid approach to screening for gene-environment interaction under the assumption of the independence between exposure and genotype in the population. In the present study, we show that the case-only design is also a valid and efficient approach to measuring gene-gene interactions under the assumption that the frequencies of genes are independent in the population. Our approach differs from that proposed by Piegorsch et al.(Stat Med 1994;13:153-162) who used a logistic model to measure gene-environment interaction with the case-only design. We show that the cross-product term in a case-only 2-by-2 table measures the departure from the multiplicative joint effects of risk ratios, but not odds ratios. For a rare disease, our results approximate odds ratios. However, our results also show that the cross-product remains a valid measure of departure from multiplicativity of risk ratios even if the disease is not rare. Just as the case-only design requires fewer cases than the case-control design in order to detect gene-environment interaction, it also requires fewer cases to detect gene-gene interactions.

40

A parametric copula model for analyzing familial binary data.

Trégouët DA[1], Bocquet V[1], Visvikis S[2], Soubrier F[3], Ducimetière P[1], Tiret L[1].
[1]INSERM U258, Paris; [2]CMP Nancy; [3]INSERM U358, Paris, France

Whereas the multivariate Gaussian distribution is currently used for analyzing familial quantitative phenotypes, there is no such standard model for familial binary data. Maximum likelihood (ML) models have been proposed to model the joint distribution of binary outcomes within families. However, they either assumed the existence of a normal liability variable to the trait or provide estimates of the aggregation parameters that are dependent on the family size. We propose a ML model, based on the copula theory (Biometrics 1994 50:954-963), that allows to model the joint distribution of a binary trait within families without assuming any liability variable. Besides, familial aggregation is estimated through an association parameter that is independent on the marginal distributions, and therefore on the family size. This model can be extended to segre-

gation and combined segregation-linkage analyses of a binary trait. This model has been applied to a combined segregation-linkage analysis of ACE levels dichotomized in two classes according to the median. The study was carried out in a sample of 95 healthy nuclear families with ≥ 2 offspring. The results confirmed the hypothesis that the I/D polymorphism of the ACE gene is in strong linkage disequilibrium with a major gene influencing ACE levels. When controlling for this major gene, there remained no residual familial aggregation.

41

The Effect of Allelic Heterogeneity on the Power of Transmission/Disequilibrium Tests (TDTs) and Affected Sib-Pair (ASP) linkage tests.

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

It has been pointed out elsewhere[1,2,3] that tests for allelic association can have substantially higher power than tests for linkage, when the gene being tested confers only a small risk of developing disease. Here we investigate whether this remains true under allelic heterogeneity. We calculated the required number of families necessary to achieve 80% power using dominant, recessive, additive and multiplicative multi-allelic disease models for a genomewide study. We demonstrate that for the TDT with single affected offspring, the required sample size increases steeply as the number of susceptibility alleles within the gene increases. The table below reveals this for a single-gene model with multiple, equally frequent, susceptibility alleles (total disease allele frequency of 0.1), with multiplicative allelic effects and an overall genotypic relative risk for individuals with at least one disease allele of 4.

	Number of Susceptibility alleles			
	1	2	5	10
TDT[1]	86	185	481	976
ASP[2]	343	343	343	343

Significance Levels: [1]$\alpha = 5*10^{-8}$; [2]$\alpha = 2.2*10^{-5}$

Researchers should be aware that heterogeneity, marker allele frequency, and the amount of disequilibrium all have an impact on the power of finding linkage for these types of studies.

References:

1.Greenberg DA (1993). Am J Hum Genet 52: 135-143.
2. Risch N and Merikangas M (1996). Science 73, 1516.
3. Camp N (1997). Am J Hum Genet 16: 1424 -1430.

42

A proposal for a collaborative pedigree database.

L. Brun-Samarcq (1), A. Philippi (2), S. Gallina (3), F. Demenais (2), G. Vaysseix (1,4), S. Ayme (5) and E. Barillot (1,4).
(1) INFOBIOGEN, Villejuif, FRANCE (2) INSERM U358, Paris, FRANCE (3) CNRS EP10, Lille, FRANCE (4) Genethon, Evry, FRANCE (5) INSERM SC 11, Villejuif, FRANCE

The genetic dissection of human diseases are often today collaborative programs that involves several distant research groups. We propose to develop an environment that allows researchers to share their pedigree data across the networks. In that respect we have developed a pedigree drawing program, so that collaborators can interact remotely with the family data. Programming environment, based on HTTP services, Java (TM) programming language and AWT graphical library, permits remote access by WEB navigators without any custom installation. Contrary to most pedigree drawing softwares, our program is intended for epidemiologists and statisticians rather than clinicians. It allows automatic drawing of large pedigrees, and inclusion of data such as lodscores. The interface is highly configurable and has a strong power of expression. A demonstration and a distribution package can be found at URL http://www.infobiogen.fr/services/CoPE (CoPE stands for Collaborative Pedigree Environment). We now intend to develop a pedigree database for epidemiologists and clinicians. In this database, a standard set of real pedigree data will be defined to serve as a common reference to test new models of genetic analysis. The repository will also archive the pedigree data released by the community, and provide access to the history of the genetic study of human diseases. This data will be made available under various formats, and also through our Java pedigree drawing program. The pedigree database will also be linked to other databases (e.g. OMIM). A Scientific Advisory Board will define the scientific and ethical framework of this project. The principle of this pedigree database has been proposed and adopted at the 1998 ESHG meeting and is now submitted to the IGES

43

Analysis of familial transmission of the response to DerpI skin-prick test (SPT) in 335 French families of the EGEA study.

F. Meunier[1], M. H. Dizier[2], J. Hochez[2],J. Feindgold[2] and F. Demenais[1].
1 I.N.S.E.R.M. U358, Paris, France, [2] I.N.S.E.R.M. U155, Paris,France.

The presence of a genetic component in allergy is now well established but its exact nature remains still unclear. To better understand the genetic mechanisms controlling the immunological response to allergens, a phenotype associated with asthma, we conducted segregation analysis of skin-prick test (SPT) to *Dermatophagoïdes pteronyssinus I* (Derp I) in a sample of 335 French families ascertained through 335 asthmatic probands (212 children and 123 parents) of the multicenter EGEA (Epidemiological Study of the Genetics and Environment of Asthma) study. SPT response was defined by a difference of the weal with a negative control of at least 3 mm. Whereas SPT was found associated with gender and age in the asthmatic probands and their relatives, no association was significant in the probands'spouses. The effect of age on SPT differed in adults and children, which led us to consider an adulthood indicator (AI) with a cutoff point of 16 years of age and an age*AI interaction. Segregation analysis of SPT response was performed using the class D regressive logistic model, which specifies a regression relationship between each person's phenotype and explanatory variables including a major gene, the phenotypes of older relatives and measured covariates, as implemented in REGRESS. The covariates included in the model were the following : IP for individuals' position in the family (probands, relatives or spouses), gender, age, AI, age*AI and interaction of IP with each other covariate, which may correct for the ascertainment of the families through a correlated trait (asthma).Our results indicate the transmission of a dominant gene with an allele frequency estimated at 0.45. These analyses will be pursued using alternative formulations of the regressive models and other strategies to correct for ascertainment.

This abstract is supported by : convention INSERM-MSD and ACC-SV2/1A028A grant.

44

A linkage analysis of Multiple Sclerosis with candidate region markers in Sardinian and Continental Italian families using three non parametric tests

S. D'Alfonso, L. Nistico', P. Zavattari, D. Bocchio, M. Marrosu, R. Murru, M. Lai, L. Massacesi, C. Ballerini, D. Gestri, M. Salvetti, G. Ristori, R. Bomprezzi, C. Montesperelli, M. Trojano, M. Liguori, D. Gambi, A. Quattrone, P. Momigliano Richiardi, R. Tosi.
The Italian Study Group on the Genetics of Multiple Sclerosis

Multiple Sclerosis (MS) is a complex disease caused by an interplay of environmental and genetic factors. So far, the major histocompatibility complex is the only region whose contribution to MS susceptibility has been established, by association study. However, it accounts for at most 10% of the genetic component. Three whole genome screens were reported. Although no predominat susceptibility gene was detected, some chromosome regions were considered as more likely candidates for presence of MS risk genes because of the clustering of MLS scores and homology with eae loci. In the present study we performed a linkage analysis of markers in these regions as well as of intragenic markers of some individual candidate genes (HLA-DRB1, CTLA4, IL9, CSF1R, ApoE, BCL2, TNFR2). For the first time, the study was targeted on Southern European populations, namely Continental Italians and Sardinians. 69 multiplex families were typed for 67 markers, by a semi-automatic fluorescence-based method. Results were analysed for linkage by three non parametric tests: Genehunter, SimIBD and TDT. The latter being utilised on multiple sibs is a test of linkage in the presence of association. In general, the results did not replicate previous linkage data, confirming the conclusion that no gene is playing a major role in the disease. However three markers, in 2p11.2, 7p15.2 and 17q12, stood out as promising since they showed significant TDT, besides relatively high scores with one of the other two tests. Association analysis with these three markers on 200 simplex families is in progress.

Supported by AISM and by ISS.

45

Risk estimation to select high risk families for identifying germ-line mutations in the breast cancer susceptibility genes

J Chang-Claude[1], H Becher[1], S Schmidt[1], M Caligo[2], D Eccles[3], G Evans[4], N Haites[5], S Hodgson[6], A Schofield[5], D Stoppa-Lyonnet[7]
[1]German Cancer Research Center, Heidelberg, Germany, [2]Univ. Pisa, Italy, [3]CRC, Southampton, U.K., [4]St. Mary's Hospital, Manchester, [5]Univ. Aberdeen, [6]Guy's Hospital, London, [7]Institut Curie, Paris, France

In genetic counselling of a woman on familial breast cancer, an accurate evaluation of the probability that she carries a mutation is needed in making genetic-testing decisions. Different models have addressed the risk of breast cancer for women with a family history of the disease, however, they have not been empirically applied.

We used data from six collaborating centers of a European Union demonstration project including 488 families recruited as research families or counselled for familial breast cancer, representing a broad range of family structures. Families were screened for mutations

in BRCA1 using SSCP, CSGE, DGGE, or FAM analysis followed by direct sequencing. BRCA1 mutations were detected in 83 families. Screening of BRCA2 is ongoing. The probability of being a carrier of a dominant breast cancer gene was calculated for the screenee under the established genetic model for breast cancer. A logistic regression approach was used to investigate the ability of carrier probabilities and additional variables to predict detection of mutation. Preliminary analysis indicates that the estimated probability of a BRCA1 mutation increases with increasing carrier probability and the number of ovarian cancers as well as the presence of a breast and ovarian cancer patient in the family improves the estimation. However, the results were heterogeneous among the different laboratories and the detection of BRCA2 mutations has not yet been accounted for.

46

Study design issues for investigating both genetic and environmental risk factors in a case-control study

G.A. Darlington[1,2] and J. Knight[1,2]
[1]Cancer Care Ontario, Toronto, Ontario, Canada,
[2]University of Toronto, Toronto, Ontario, Canada

In the study of complex diseases, the simultaneous consideration of genetic and environmental factors is of interest. One approach to address questions of this type is the case-control design. While in traditional epidemiologic studies controls are randomly selected from disease-free individuals in the population from which cases are selected, other alternatives for genetic epidemiologic studies have been considered. In particular, advantages and disadvantages of using sibling or cousin controls in studies of candidate genes recently have been addressed in the literature.

If the question of interest in a case-control study is to measure the independent effects of genetic and environmental risk factors, then the use of family controls may, because of overmatching, result in bias in the estimation of environmental effects. The potential extent of this bias will be presented under various assumptions for correlation of environmental factors among relatives. Feasibility of design alternatives also will be presented.

47

Extension of class D regressive model to bivariate linkage analysis: increase of power to detect Quantitative Trait Loci (QTLs)

G. Durrieu and F. Demenais.
I.N.S.E.R.M. U358, Paris, France.

Study of intermediate quantitative traits may help in dissecting the genetics of complex diseases. To improve the detection of QTLs with a pleiotropic effect on correlated traits, we extended the class D regressive model to bivariate analysis. This model can take into account a major gene effect, residual familial correlations, within trait and cross-trait correlations, and regression coefficients on covariates. Power of this approach to map QTLs was assessed by simulating two quantitative traits and a marker in nuclear families. The quantitative data were generated under bivariate class D regressive models including a major gene and residual correlations. We considered either a dominant gene, responsible for 13% of the total variance for trait 1 and 26 % for trait 2, or a recessive gene, accounting for 7.6% of variance for trait 1 and 25% for trait 2. The within trait correlations were 0.2 or 0.4. The intraindividual and interindividual cross-trait correlations were set at 0.15 or 0.30, being either both equal or different. The linked marker was fully informative with recombination fraction, θ, equal to 0.0 or 0.05. We generated 100 replicates of 100 nuclear families (6 children) with at least two sibs with trait values in the upper 5% tail of the distribution. The bivariate and univariate approaches were compared in terms of the ratios of mean maximum lod scores (R1 for trait 1 and R2 for trait 2) by estimating only θ. The ratios of the mean maximum mod scores are computed by estimating all traits parameters with $\theta = 0$. When the gene effect is small, the bivariate analysis increases greatly the power to detect linkage: R1 ranges between 5 and 14 and is highest when the interindividual and intraindividual cross-trait correlations are high and low respectively. When the gene effect is large, R2 varies between 1.09 and 1.5, showing only a slight improvement of power. Further simulations will consider more complex genetic models.

This work was supported by a BIOMED2 grant (BMH4-CT97-2532)

48

Little evidence for a relationship between oral contraceptive use and BRCA1 and BRCA2 mutation status in women diagnosed with breast cancer

J.M. Schildkraut[1], K. Regan[1], L. G. Bluman[1],
A. Miron[1], J.D. Igelhart[1], P.A. Futreal[1]
[1]Duke University Medical Center, Durham, USA.

Using data from subjects participating in a genetic counseling study and the Duke Family Cancer Program, we identified probands diagnosed with breast cancer who had undergone genetic testing for BRCA1 and BRCA2 mutations. Family history information was ob-

tained from a telephone survey. Epidemiologic data, including OC use, reproductive history, and demographic characteristics, was obtained from a self-administered questionnaire. Subjects were restricted to women born since 1940. Of the 65 women included, 15 tested positive for either BRCA1or BRCA2, and 50 were negative for a mutation in both genes. The subjects were 20-55 years of age at diagnosis, 95% were white, and 7 reported being Ashkenazi Jewish. Using logistic regression analysis, we calculated odds ratios (ORs) and 95% confidence intervals (CIs) for the association between duration of oral contraceptive (OC) use and BRCA1/BRCA2 mutation status. Adjusting for potential confounders, ORs for 12-48 months and >48 months of OC use were 1.8 (95% CI=0.4-9.4) and 2.4 (95% CI=0.6-10.2), respectively. There was no evidence for a relationship between duration of OC use before the first full-term birth and BRCA1/BRCA2 mutation status. The overall association between OC use did not change when the analysis was restricted to BRCA1 mutation carriers although there was some suggestion of a relationship between >48 months of OC use before the first full-term birth and BRCA1 mutation status alone (OR = 5.2, 95% CI=0.2-140.3). Although our results are preliminary, we found little evidence of a relationship between OC use and breast cancer in BRCA1/BRCA2 mutation carriers. Our results are not consistent with those of a recently published study, of similar size, in Ashkenazi Jewish women under the age of 40 at diagnosis.

49

A genetic epidemiologic study of radiological osteoarthritis

C.M.van Duijn[1], C.Bijkerk[1,2],J.J.Houwing-Duistermaat[1], I Meulenbelt[2,3], H.A. Valkenburg [1], A.Hofman[1],F.C.Breedveld[3], H.A.P.Pols[1], Th. Stijnen[1] and P. E. Slagboom[2].
[1]Erasmus Univ.Rotterdam, [2]Gaubius Lab., TNO-PG, Leiden, Univ. Hospital Leiden, The Netherlands.

To estimate the genetic influence on the occurrence of radiological osteoarthritis (ROA) in the knees, hips and hands and disk degeneration of the spine, relatives of 118 probands derived from a population-based study, the Rotterdam Study, were studied. Heritability estimates for ROA in the knees, hips, and hands and disk degeneration of the spine were calculated by comparing the data of the siblings with the prevalence of ROA and disk degeneration in the cohort. ROA was defined based on Kellgren's grading system. Hand ROA and disk degeneration of the spine was statistically significantly more frequent in siblings as compared to the cohort. The heritability estimate for hand ROA was 0.56 (95 % confidence interval (CI) 0.34-0.76) and for disk degeneration 0.75 (95 % CI 0.30-1.00). Heritability estimates suggested no evidence for a genetic effect on the occurrence of knee and hip ROA in the general population. The heritability estimate for the sum score of ROA and disk degeneration (a score summating the number of joints affected in the knees, hips, hands and spine) was 0.78 (95 % CI 0.52-0.98), suggesting a generalized susceptibility for ROA. As candidate genes, the collagen genes COL2A1, COL9A1 and COL11A2 were studied. There was evidence for association of ROA to both the COL2A1 and COL9A1.

50

Optimization Procedures for Complex Segregation Analysis: the use of Genetic Algorithms

F. Tores*, G. Durrieu and M. Martinez.
U358 INSERM, Paris, France.

Segregation analysis under the regressive approach is very useful to assess genes involved in complex traits. These models allow to account for a major gene, residual familial correlations, covariates and gene-by-covariate effects. However, finding the global maxima of unrestricted models can be a challenge: the number of parameters is large, and the shape of the likelihood may contain numerous local maxima trapping traditional optimization methods based on gradient. Alternative procedures as Genetic Algorithms can be proposed. This method generates a population of possible solutions which evolves through generations by the mean of some « genetic » laws (selection, reproduction, crossing-over and mutation). Our goal is to compare three optimization strategies: GEMINI alone (GE), GA alone (GA), and GA followed by GEMINI (GAE), for complex segregation analysis. Monte Carlo methods (100 replicates) were used to simulate the segregation of a quantitative trait in samples of 100 nuclear families (2 parents and 6 sibs), under different class D models. Generated modesl varied according to the major gene effects and the patterns of residual correlations. For each replicate, likelihoods of the unrestricted model, defined by 8 parameters (major gene + residual correlations) were maximized under GE, GA and GAE using the REGRESS program. Our results show that the mean (m) and variance (σ^2) of maximum likelihood estimators differ under the three strategies. Accuracy of estimates was not improved with GAE. Indeed, in most cases GA worked better than GE or GAE. With GA, m were close to their true values and σ^2 had the smallest values. The study will be further extended to (1) the general class D models including the three transmission parameters and (2) combined segregation and linkage analysis.

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) ≥ 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/m^2). 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.

89

Haseman and Elston Revisited

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

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

90

Detecting influential families in linkage and association studies

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

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

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

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

91

Susceptibility to celiac disease in Tunisian children and GM immunoglobulin allotypes

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

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

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

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

92

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

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

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

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

93

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

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

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

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

94

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

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

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

95

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

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

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

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

96

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

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

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

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

97

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

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

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

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

98

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

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

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

99

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

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

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

100

The hunt for candidate genes in families with a missing parent

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

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

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

101 [Invited Speaker]

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

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

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

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

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

102 [Invited Speaker]

Genetic epidemiology of infectious diseases.

Laurent Abel
INSERM U.436, Paris, France.

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