Genetic Epidemiology 21:139–183 (2001)

ABSTRACTS FROM THE

TENTH ANNUAL MEETING OF THE INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY

GARMISCH PARTENKIRCHEN, GERMANY
SEPTEMBER 2–4, 2001

IGES-51

Distribution-Free Maximum Likelihood TDT for Quantitative Traits in Nuclear Families

A. Alcais, L. Abel
Laboratory of Human Genetics of Infectious Diseases, INSERM U.550, Paris, France

The classic TDT is a matched chi-square but the same data can be analyzed using a Maximum-Likelihood statistic (ML-TDT) which was shown to be more powerful. We propose an extension of the ML-TDT to quantitative phenotypes, denoted as ML-QTDT. As already proposed for linkage analysis (Alcais and Abel, 1999), this extension relies on the introduction of a latent binary variable. As usual, only parents heterozygous for the marker allele of interest, A, contribute to the likelihood. The first part of this likelihood depends is expressed using binomial distributions of allele A among children. The second part of the likelihood is a link function which can be either distribution-free (e.g. using deciles) or specified using a cumulative parametric distribution (e.g. normal). This method takes into account in a natural way nuclear families with multiple children and provides a simple likelihood-ratio test. Large simulation studies showed that the ML-QTDT is a valid test for linkage disequilibrium whatever the distribution of the quantitative trait and the study design (in particular, in the context of small sample size and multiple offspring families). These studies also compared the ML-QTDT to recent methods (variance components') implemented in the QTDT software and showed that it was the most powerful approach for the analysis of samples selected through extreme trait values.

IGES-112

Joint Effects of HLA Loci in Causing Rheumatoid Arthritis

C.I. Amos[1], D. Jawaheer[2], T.W. Behrens[3], J.T. Elder[4], W. Chen[1], M.F. Seldin[5], P.K. Gregersen[2]
[1]U.T. M.D. Anderson Cancer Center Houston, TX; [2]North Shore University Hospital, Manhasset, NY; [3]University of Minnesota, Minneapolis; [4]University of Michigan, Ann Arbor; [5]University of California-Davis

The North American Rheumatoid Arthritis Consortium recently completed a genome scan including 301 affected sibling pairs from 257 multiplex families. The HLA region yielded a LOD score of 5.9 under an autosomal dominant inheritance model. To further study the HLA region, we genotyped 50 novel microsatellite markers in the region. Results of these analyses show significant association, using the pedigree disequilibrium test (PDT). Areas showing highly significant results included the region near DRB1 (M6S118, p=3 × 10-4) and near the tumor necrosis factor (TNFβ) locus (M6S125, p=8 × 10-4). To further evaluate effects of HLA region loci, we have also genotyped 383 unaffected individuals for the same 50 microsatellite markers. Results comparing the probands from the linkage study to controls shows a jointly significant associa-

tion between alleles at both loci and case status. The joint analysis showed R-squared values of 0.46, 0.13, and 0.51 for M6S118 alone, M6S125 alone, and M6S118 and M6S125 analyzed jointly. Alleles 4 and 11 of M6S125 showed associations that were significant at the 0.005 and 0.001 levels, while several of the M6S118 alleles were significant at less than the 0.0001 level. These support a weak additional effect of a locus near TNFβ, along with a stronger effect of the HLA DRB1 locus in influencing risk for rheumatoid arthritis.

IGES-32

A Comprehensive Model for Familial Breast Cancer Incorporating BRCA1, BRCA2 and Other Genes

A.C. Antoniou[1], P.D.P. Pharoah[2], B.J. Ponder[2], D.F. Easton[1]
[1]CRC Genetic Epidemiology Unit, Institute of Public Health; [2]CRC Human Cancer Genetics Research Group, University of Cambridge, UK

To provide efficient genetic counselling for breast and ovarian cancer, it is important to have mathematical models that compute mutation carrier probabilities and age specific cancer risks. Existing models do not differentiate adequately between BRCA1 and BRCA2, and fail to take into account that genes other than BRCA1/2 may be involved in genetic susceptibility to breast cancer. We have developed a risk model for familial breast cancer, using segregation analysis of breast and ovarian cancer occurrence in a combined dataset, including a population based series of 1484 breast cancer cases and 156 high-risk families. The model incorporates BRCA1, BRCA2 (allele frequencies: 0.051, 0.068% respectively), the effects of low penetrance genes with multiplicative effects on the risk of breast cancer, and the effect of modifiers on the breast cancer risk in BRCA1/2 mutation carries. 20% of the familial aggregation of breast cancer is estimated to be due to BRCA1/2 and 80% due to the polygenic component. The model discriminates well between BRCA1 and BRCA2 mutation carriers. We demonstrate that the incorporation of modifiers results in a wide variation in the predicted breast cancer risks in unaffected BRCA1/2 mutation carriers according to the family history of breast cancer. The predicted cancer/carrier risks for some families differ markedly from those predicted under other models.

IGES-96

Automated Versus One Model-at-a-Time Segregation analysis

V. Apprey, J. Kwagyan, G.E. Bonney
Statistical Genetics and Bioinformatics Unit, National Human Genome Center at Howard University, Washington, USA

Segregation analysis requires fitting several models to family data. The challenges faced in fitting the necessary models one-at-a-time in each case battling with the very unsettling statement “initial estimates not in the domain of the function.” And the questions, Did it converge? If so, was it at a global maxima? Were there numerical problems? We propose to minimize these problems through automated fitting techniques. In segregation analysis, to infer a major gene, the Elston School of thought requires fitting the following models: I. No familial aggregation, or complete independence model; II. A model with familial aggregation; III. A model including a major type that is not transmitted; IV. A model in which the major type is Mendelian; V. A model in which the major type is environmental; VI. A model in which the major type has arbitrary transmission. To infer Mendelian inheritance, the hypothesis IV. must be accepted against VI, but V must be rejected against VI If the environmental model is also accepted, then one must exercise caution in accepting the Mendelian hypothesis. Now IV and VI are non-nested ypothesis, but each is nested within VI. This means that in fitting these models we are following two seperate lines of increasing likelihoods, which are the same initially but diverge later and come together again at the end. Analysis of type II diabetes is discussed.

IGES-139

Linkage of Multiple Structural and Functional Cardiac Phenotypes to Chromosome 4: The HyperGEN Study

D.K. Arnett[1], R.B. Devereux[2], D. Kitzman[3], A. Oberman[4], P. Hopkins[5], D.C. Rao[6]
[1]University of Minnesota; [2]Cornell University, [3]Wake Forest University' [4]University of Alabama at Birmingham; [5]University of Utah; [6]Washington University

Genome-wide linkage analysis of nine left ventricular (LV) structural and functional phenotypes was conducted in 681 African American (AA) and 481 white hypertensive sibpairs. Sibships were recruited for the Hypertension Genetic Epidemiology Network (HyperGEN, an NHLBI-sponsored study designed to identify genes contributing to hypertension). Echocardiography was performed using a common protocol in four field centers and centrally read. The phenotypes, aortic root diameter (ARD), arterial stiffness (AS), LV internal dimension (LVID), LV mass/height (LVM), early and late diastolic filling (MVE and MVA), contractility (MWS), posterior wall thickness (PWT), and relative wall thickness (RWT), were adjusted for age, age 2, and center within sex-race groups. The mean age was 60 years in whites and 52 years in AAs; 42% and 71%, respectively, were female. Heritability ranged from 0.18 to

0.70 in AA and 0.26 to 0.66 in whites. Microsatellite markers were typed by the NHLBI Mammalian Genotyping Service. Multipoint model-free variance components linkage analysis was implemented using GENEHUNTER. Suggestive linkage was detected in 7 of 9 phenotypes in a region of chromosome 4 spanning 94–150 cM (see table). The consistency of findings across phenotypes and race suggest chromosome 4 contains influential genetic loci contributing tp LV structural and functional measures.

Race Group	94–100 cM	112–129 cM	144–150 cM
AA	PWT 2.6; LVM 1.4	ARD 1.5	
Whites		MVE 3.0; MVA 2.2, ARD 1.3	LVID 2.3; LVM 1.9; MWS 1.4

IGES-110

A Genome-Wide Scan of Finnish Hereditary Prostate Cancer Families Identifies Chromosomes 11, X and Other Regions of Interest.

A. Baffoe-Bonnie[1,4], J. Schleutker[2], E. Gillanders[3], T. Kainu[3], M. Matikainen[2], P. Koivisto[2], T. Tammela[5], J. Trent[2], J. Bailey-Wilson[4], O.P. Kallioniemi[2]
[1]Fox Chase Cancer Center, Philadelphia, USA; [2]Laboratory of Cancer Genetics, Tampere University Hospital, Finland; [3]Cancer Genetics Branch, NHGRI, USA; [4]Inherited Disease Research Branch, NHGRI, USA; [5]Division of Urology, Tampere, Finland

Several predisposition loci for hereditary prostate cancer (HPC) have been suggested. (Smith et al., 1996, Berthon et al., 1998, Xu et al., 1998, Gibbs et al. 1999, Berry et al., 2000). A homogeneous population like Finland, may help to minimize genetic heterogeneity that usually complicates genetic linkage analyses. Here, we report results of a genome-wide linkage scan of 13 multiplex Finnish HPC families, selected from a total of 292 Finnish families identified in our nation wide search (Schleutker et al. 2000). On average, 4 affecteds were genotyped per family (range 2–6), with a mean age of 68.9 years (range 44–99 years) at diagnosis. Altogether, 413 markers were analysed with a ABI3700 capillary sequencer. Two-point and multipoint LOD scores were calculated for all autosomes with FASTLINK, Genehunter and Genehunter-Plus. The results indicated three chromosomal sites with two-point LOD scores greater than 1.5. The most promising area was found at 11q, where the peak two-point LOD score was 2.85 (theta = 0.0) for affected+unaffected analyses. Peak Genehunter HLODs were 3.04 (theta = 0.0) for affected + old unaffected and 3.28 (theta = 0.0) for affected only analyses. One family shows evidence of linkage to the X chromosome. The results suggest several putative novel regions of interest for HPC in Finland. Chromosome 11p linkage has been reported by Ginns et al., 2000.

IGES-115

Breast and Ovarian Cancer Co-Incidence May Be a Marker of Heterogeneity in Hereditary Prostate Cancer

M.D. Badzioch[1], J.L. Stanford[2], K. Markianos[2], S. Kolb[2], M. Janer[3], M.A. Peters[2], E.L. Goode[1], M. Gibbs[2], L. Hood[3], E.A. Ostrander[2], G.P. Jarvik[1]
[1]University of Washington; [2]Fred Hutchinson Cancer Research Center; [3]Institute for Systems Biology, Seattle, Washington

Epidemiological reports suggest that breast cancer (BC) and prostate cancer (PC) share genetic risk factors but few studies have focused on inherited susceptibility to both cancers in high-risk PC families. Using genome-wide non-parametric linkage, we evaluated 30 hereditary PC families with a first-degree family history (FH) of BC in 6 family sets defined by 3 FH groups (1 BC case, 2+ BC, and BC with ovarian cancer (OC)) and 2 median age-of-PC-diagnosis groups (< and >= 66 yrs). (All families were BRCA1/2 negative by SSCP.) The median age of PC diagnosis per family differed among the 3 FH groups (69.1, 67.7, and 63.5 yrs, respectively; p=0.009). Linkage results with an observed p< 0.005 are shown. Empirical significance levels from 100 genome simulations will be presented. The NPL= 6.10, p< 0.0004, was primarily due to a single non-Caucasian family, in which, 3 of 4 founder 4-marker haplotypes were identical. Older-onset PC families with BC and OC provided evidence for linkage to 3 chromosomal regions, the most significant of which, p< 0.002, mapped to chr 7q32-q36, a region showing LOH in PC and possibly harboring an epithelial cell tumor suppressor gene. The earlier age of PC-onset and significant linkage in this group indicate possible heterogeneity in hereditary PC marked by FH of BC and OC that may help locate candidate loci predisposing to these cancers.

Family History	2+ BC	BC and OC	BC and OC	BC and OC	BC and OC
Median PC Age	> 66	< 66	> 66	> 66	> 66
No. Families	4	5	3	3	3
Chr. Location	19p13.3	6p21.3	7q32-q36	7q21-q22	20q11.23-q13.1
NPL	3.17	6.10	3.99	3.07	3.07
p-value	0.00457	0.00039	0.00163	0.00723	0.00723

IGES-161 [INVITED SPEAKER]

Study Designs to Detect Gene Environment and Gene–Gene Interactions in Oral Clefts

T.H. Beaty
Johns Hopkins University, U.S.A.

A number of different study designs are available for testing for gene-environment and even gene–gene interaction, including case-only design, case-control and case-family designs. The strengths and weaknesses of each of these designs will be discussed briefly, and illustrated with findings from studies of oral clefts, a heterogeneous group of birth defects with a complex etiology. Analysis of markers in selected candidate genes were examined on 269 case-parent trios ascertained through a child with an isolated, non-syndromic oral cleft (cleft lip, CL; cleft palate, CP; or cleft lip and palate, CLP). Markers at two candidate genes, transforming growth factor b3 (TGFB3) and MSX1, showed consistent evidence of linkage and disequilibrium due to linkage using several different statistical tests. There was little evidence of heterogeneity in the role of TGFB3 between different types of oral clefts, but MSX1 suggested some heterogeneity between the common types of clefts. MSX1 also showed evidence for interaction between infant's genotype and maternal smoking. Using conditional logistic models to test for gene–gene interaction revealed no evidence of interaction between these genotypes at these two markers, and both genes seem to contribute to the risk of isolated, non-syndromic oral clefts independently.

IGES-145

Family-Based Tests of Association in the Presence of Consanguinity

S. Bennett[1], A. Casbard[1], R.N. Curnow[2]
[1]London School of Hygiene and Tropical Medicine;
[2]University of Reading.

Consanguineous marriages, usually between first cousins or between uncle and niece, are common in certain societies. We have investigated the consequences for the Transmission/disequilibrium Test (TDT), and for two sibling-based tests (the SDT and S-TDT), when parents are related. For the TDT we calculated algebraically the probability that an affected child inherits the given allele from a heterozygous parent. For the SDT and S-TDT, simulation was used to generate discordant sibships. In each case, the Type I error and power were evaluated for three levels of inbreeding and for a range of modes of inheritance and penetrance values.

The Type I error probability of the TDT is unaffected by intermarriage, except for a purely recessive disease allele. Its power is increased for a recessive allele and decreased for a dominant allele, the effects being greater for candidate genes, for uncle-niece marriages, for rare disease alleles and for high genotype relative risk. The Type I error of the sibling-based tests was slightly lower in the presence of consanguinity. Their power was also little affected, being increased for a recessive, multiplicative or additive model, and decreased for a dominant model. Consideration of levels of consanguinity that arise in practice indicates that standard power calculations for each of the tests will usually need only minor modification.

IGES-46

Gene-Environment Interaction between Poly(ADP-ribose)polymerase activity, Smoking and Alcohol: A Case-Control Study on Laryngeal Cancer

H. Becher[1,2], H. Ramroth[2], P. Schmezer[2],
N. Rajaee-Behbahani[2], A. Dietz[1]
[1]University of Heidelberg, Medical School, Germany;
[2]German Cancer Research Center, Heidelberg, Germany

We performed a population-based case-control study in Germany in 1998–2001 with 257 cases and 769 controls frequency matched for age and gender to investigate the effect of activity of Poly(ADP-ribose)-polymerase (PARP), a nuclear enzyme that is catalytically activated by DNA strand breaks on laryngeal cancer risk and its interaction with smoking and alcohol. PARP activity was measured by DNA damage-induced poly(ADP-ribose) formation in human peripheral blood lymphocytes (PBL) with quantitative immunofluorescence analysis. Polymer formation was determined in response of PBL to bleomycin (BLM) which induces DNA strand breaks. BLM-induced polymer formation was not associated with smoking or alcohol drinking. Comparing the highest tertile of PARP activity with the lowest, the odds ratio (OR) was 0.23 (95% CI 0.08–0.67). (linear trend test $p<0.01$). Odds ratio increases with tobacco dose (>40 pack-years OR=18.06, 95% CI 2.72 # 120) and for heavy drinkers (>120 ml ethanol/day OR=4.38, 95% CI 1.14–16.9). The factors acted multiplicatively on laryngeal cancer risk. Thus, individuals in the highest risk groups for all factors have an estimated OR of $18.1\times4.4\times(1/0.23)=344$. Individuals with low PARP-activity and additional high tobacco and alcohol consumption have an extremely high risk which could considerably be reduced by quitting to smoke and to drink alcohol.

IGES-16

Genome-Wide Linkage Analysis of Obesity-Related Traits in the NHLBI Family Heart Study (FHS)

I.B. Borecki[1], M.F. Feitosa[1], S. Hunt[2], D. Arnett[3],
R.H. Myers[4], S.S. Rich[5], M.A. Province [1]

[1]Washington Univ Schl Med, USA; [2]Univ Utah, USA; [3]Univ Minnesota, USA; [4]Boston Univ, USA; [5]Wake Forrest Univ Schl Med, USA;

Obesity is a significant public health problem and identification of underlying genes can suggest novel interventions. In the FHS, 1,328 subjects in 316 sibships ascertained via CHD or a high individual risk score were typed for 243 markers (avg density ~20cM) by the Utah Molecular Genetics Lab. Another 3,027 subjects in 401 of the largest 3-generation pedigrees were also typed for 404 markers (avg density 8.5 cM) by the NHLBI-supported Mammalian Genotyping Service. A genome scan was conducted and data analyzed using a variance components approach. The following phenotypes were examined: BMI, BMI at age 25, highest and lowest BMI since 25 y, largest weight gain and loss, birthweight, current waist circumference, waist to hip ratio, and sum of skinfold thicknesses. The strongest linkage results included: lod=4.9 for BMI on ch 7 at 137 cM near the leptin gene (previously reported on a subset of the data by Feitosa et al. 2000); lod=3.7 for highest BMI on ch 1 at 158 cM (also seen in Pimas and Finns); lod=3.2 for BMI and lod=2.5 for waist circumference on ch 13 at 42 cM (also seen in FUSION study for BMI and 2-h insulin); suggestive signals (lod 2–3.4) on ch 15 at 22 cM for lowest BMI, highest BMI, and waist; and lod 2.1 for sum of skinfolds on ch 20 (seen in previous studies). There appears to be an emerging consensus as to common loci influencing adiposity.

IGES-93

Missing Data in Haplotype Analysis

C. Bourgain, E. Genin, F. Clerget-Darpoux
INSERM U535, Le Kremlin-Bicêtre, France

Given the enormous progress in the knowledge of the human genome, genetic markers are now available throughout the genome. Haplotype analysis, allowing the simultaneous use of information brought by several markers, has thus become increasingly popular. However, we often face the problem of missing data and of haplotype identification. Even when parents are available, phase determination may be ambiguous. This is particularly true for SNP analysis.

We have proposed an haplotype based method for the genetic study of multifactorial diseases in founder populations, the MILC method (Bourgain et al, 2000). MILC is based on the contrast of identity length between haplotypes transmitted to affected offspring and haplotypes non transmitted. Different strategies regarding missing data and haplotype ambiguities are studied for MILC in a context of SNP analysis : discarding ambiguous haplotypes, considering ambiguous loci as missing data or using all possible haplotypes weighted by their frequencies. Power and type I error are evaluated as a function of allele frequency and linkage disequilibrium pattern.

IGES-15

Optimized Group Sequential Study Designs for Tests of Genetic Linkage in Complex Diseases

I.R. Böddeker, H. Schäfer, H.-H. Müller, A. Ziegler
Center for Methodology and Health Research,
Institute of Medical Biometry and Epidemiology,
Philipps-University of Marburg, Germany

Studying genetic linkage in complex traits requires large sample sizes due to small expected effect sizes and the adoption of extremely low significance levels. Using a sequential study design reduces the amount of necessary phenotypings and genotypings on average. Here, interim analyses are conducted with the possibility to stop the investigation early if the result is significant.

We present the application of optimized group sequential designs to the analysis of genetic linkage using the mean-test on affected sib pairs. For designs with two and three stages at an overall significance level of 0.0001 and a power of 0.8, we calculated necessary sample sizes, time points, and critical boundaries for interim and final analyses. Results from Monte Carlo simulation analyses are shown that demonstrate the validity of the asymptotic approximation. Furthermore, average sample sizes required under the null and alternative hypotheses in the different study designs are given.

The application of the sequential designs leads to a maximal increase in sample size of 5% under the null hypothesis compared to the fixed sample design. This is contrasted by maximal savings of 15% in average sample sizes under the alternative hypothesis. As these savings affect the amount of both genotyping and phenotyping required for a study, they lead to a significant decrease in cost and time of the study.

IGES-128

DNA Repair and Survival in Patients with Non-Small Cell Lung Cancer

C. Bosken[1], Q. Wei[1], C. Amos[1], M. Spitz[1]
[1]U.T.M.D. Anderson Cancer Center, Houston, Tx, USA

Genetically determined variability in DNA repair has been associated with the risk of developing lung cancer. We tested the hypothesis that variability in DNA repair would also be associated with survival of patients after diagnosis.

Methods: Patients with histologically documented lung cancer were recruited for a case-control study. Follow-up was obtained from the clinical record and from the U. S. National Death Index. DNA repair was measured by the Host Reactivation Assay. Statistical analy-

ses included Cox proportional hazards models for continuous variables and log-rank test for DNA repair used as a categorical variable with high and low repair patients divided at the median DRC.

Results: Preliminary analysis showed that DNA repair was related to age, sex, and date of entry into the study, but not stage of disease, degree of tumor differentiation or to the presence or absence of weight loss. The final Cox Proportional Hazards models were therefore adjusted for these variables. The relative risk of death associated with efficient DNA repair was 1.13 (CI = 1.03–1.25, p = 0.01) for stage III/IV patients who were treated with chemotherapy alone. There was no effect of DNA repair on patients who were not treated with chemotherapy.

Conclusion: Variability in DNA repair may explain some of the variability in the response to therapy of patients with lung cancer.

IGES-53

Complete Enumeration Methods for Power and Sample-Size Calculations for the Transmission/ Disequilibrium Test

B. Brown
University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

We enumerate all possible offspring for all combinations of parental haplotypes to compute power or requisite sample-size for the TDT test. This allows the extension of the flexible genetic model of McGinnis (1998) to both-parent ascertainment and to the use of all affected offspring. This method is readily understood without mathematical manipulation.

A Fortran 95 computer program to perform these calculations is freely available:

http://odin.mdacc.tmc.edu/anonftp/

IGES-147

Power of Sib Pair Analysis of Diseases with Variable Age-of-Onset in the Presence of Gene–Environment Interaction

L. Briollais
Samuel Lunenfeld Research Institute, Mount Sinai Hospital

Selecting sib pairs with extreme trait values can increase the power to detect quantitative-trait loci but requires the screening of a large number of pairs. We have investigated how accounting for gene–environment (GE) interaction can increase the power of sib pair analysis and allow less extreme trait values to be selected when the trait analyzed is a disease with variable age-of-onset. We assumed that the disease status was determined through an underlying continuous variable (the liability) such that affected individuals have a liability that lies between two time-dependent thresholds, depending on their age-of-onset. We assumed that the liability results from the additive effects of a major gene (codominant or dominant with H=24% and 32% and q=20%), an environmental factor and a GE interaction. The joint probability of belonging to the liability classes (h,l) for a sib-pair was computed from a bivariate normal cdf. Our results showed that the number of sib-pairs required to achieve a power of 80% at a 5% level in each combination of classes (h,l) decreases dramatically in presence of a GE interaction and a residual correlation. This number was in most cases lower than 500 when only one of the two sibs had an extreme age-of-onset value. Selection of sib pairs through a single proband with extreme age-of-onset may be an efficient design in the presence of GE interaction.

IGES-11

Segregation Analysis of Lung Cancer — Results of a Case-Family Study.

K. Bromen[1], H. Pohlabeln[2], W. Ahrens[2], I. Jahn[2], K.-H. Jöckel[1]
[1]Institute of Medical Informatics, Biometry and Epidemiology, University of Essen, Essen, Germany; [2]Bremen Institute for Prevention Research and Social Medicine, Bremen, Germany

Objectives: The contribution of familial factors to lung cancer development has been shown in previous studies. Using segregation analysis we tested how well models of inheritance fit the data of a German lung cancer study when considering environmental factors.

Methods: Each 1004 cases and controls (individually matched by age (+/–5 years), sex and region) were interviewed between 1988 and 1993 using a standardised questionnaire. Among the data assessed was family history of diseases. Information on 945 lung cancer cases and 4674 relatives (parents and siblings) was included in the analysis. Calculations were based on both logistic and time to event models. Parameter estimates were obtained by maximum likelihood methods using program package GAP. Age, sex and smoking (estimated for siblings) were considered as covariates in the analysis.

Results: The analyses based on Cox regression yielded a better data fit than the logistic model. The models without genetic component were rejected when compared to the general model. Of the genetic models the dominant or codominant model provided the best fit depending on the covariates considered.

Conclusions: Our findings support a genetic influence in lung cancer development and confirm the re-

sults of earlier segregation analyses. Population-based family studies with complete covariate data are needed for confirmation.

IGES-121

Linkage Analysis of the Respiratory Disturbance Index in African Americans and Caucasians

S.G. Buxbaum[1], R.C. Elston[1], P.V. Tishler[2], L.J. Palmer[2], S. Redline[1]
[1]Case Western Reserve University, Cleveland, OH; [2]Harvard Medical School, Boston, MA

In a segregation analysis of obstructive sleep apnea (OSA), we have assessed sex-specific transmission patterns in an African American sample of 123 families and a Caucasian sample of 177 families. In each sample, these patterns were assessed using two variables: 1) the respiratory disturbance index (RDI), log transformed and adjusted for age and 2) the RDI, log transformed and adjusted for age and body mass index (BMI). Segregation analysis of the Caucasian sample showed transmission patterns consistent with that of a major gene that were stronger in the age-adjusted variable than in the age- and BMI-adjusted variable. However, in the African American families, adjusting for BMI in addition to age gave stronger evidence for segregation of a recessive gene with an allele frequency of .2 accounting for about 20% of the total variance. These results provide support for an underlying genetic basis for OSA that in African Americans is independent of the contribution of BMI. Subsequently, an informative subset of these families was selected for a genome scan. We now present results of a linkage analysis of the 61 African American families (110 males and 143 females) and the 67 Caucasian families (183 males and 167 females), using model-free and model-based methods.

IGES-119

Analysis Of Accumulating Data In Genomewide Studies of Affected Sib Pairs

S. Bull[1], C. Greenwood[2,3,4], L. Mirea[1], J. Biernacka[1], K. Morgan[3,4]
[1]Samuel Lunenfeld Research Institute, Public Health Sciences, University of Toronto; [2]Epidemiology & Biostatistics, McGill University; [3]Montreal Genome Centre, [4]Human Genetics, Medicine, McGill University

The objective of current genomewide studies of complex diseases is to find chromosomal regions that harbour susceptibility genes for the disease, but the presence of heterogeneity in the study sample will often reduce power to detect linkage. Regression models for allele-sharing assess linkage evidence in a likelihood framework in which overall tests, subgroup tests, and tests for heterogeneity can incorporate covariates suspected to be associated with heterogeneity. When linkage data have accumulated and there is some question about whether early results differ systematically from later ones, analysis of heterogeneity using covariates may help to distinguish between random and systematic effects. We applied regression modelling and permutation methods to investigate variation in allele-sharing associated with the order in which families were typed in a genome scan of inflammatory bowel disease (Rioux et al 2000), taking into account family-level covariates of diagnostic subtype, age at onset, and ethnicity. Evidence for linkage increased initially as the families accumulated but then attenuated as the sample size continued to increase. This pattern could be only partially explained by diagnosis and age at onset. We discuss the implications of heterogeneity and systematic inter-family differences for sequential analysis and replication of linkage results.

IGES-111

Genetic Epidemiology of Hepatitis Infection in the Brazilian Western Amazon Region.

L.M.A. Camargo[1], M.F. Feitosa[2], R.G.M. Ferreira[1], R.C. Pagotto[1], K.R. Oliveira[3], M.M. Moura[4], V. Engracia[3], S.A. Basano[1], B. Beiguelman[1], H. Krieger[1]
[1]Dep. Parasitology, Institute of Biomedical Sciences, University of Sao Paulo, Brazil; [2]Div. Biostatistics, Washington University School of Medicine, St. Louis, USA; [3]CEPEM, Porto Velho, RO, Brazil; [4]Federal University of Rondonia, Brazil

The prevalence of positive serology for hepatitis B (H-B) among the adult population in some localities of Amazon region is high (over 70%). A large population survey aimed to study the causes of the variability of several traits associated with infectious diseases was carried out in Monte Negro (10o15′35″S, 63 o18′06″W). A random sample of 214 families with 922 individuals was typed serologically for H-B (Elisa assay). The prevalence was 59%, not significantly different among genders (p=.09). Regression results showed an increase of H-B prevalence with age. Segregation analysis (Lalouel et al. 1983) indicated a significant familial aggregation (p <.0001). Either the multifactorial or the major effect models were not rejected. The mendelian transmission was not rejected, while the non-transmission of the major gene was rejected (p <.0001). The parsimonious model indicated by Akaike's criterion was the additive mendelian mechanism (q=.36). A search for association between H-B and some genetic markers (ABO, Rh, MNSs, Kell, Duffy, PGM, EsD, CAII and GLO), in a

sub-sample of unrelated adults (parents), did not show any significant association. No significant association was found between H-B and GLO, which is located in chromosome 6, within a measurable linkage distance from the MH complex, where there have been some reports of association (FAPESP, CNPq).

IGES-144

Sufficient Cause Model for Gene–Environment Interaction

S.-I. Cho
Seoul National University School of Public Health, Republic of Korea

Genetic epidemiology often progressed by exploiting the concepts and methods developed independently in the fields of genetics and epidemiology. Gene–environment interaction is one of such topics that may be studied better by utilizing the advances in both fields. The models Ottman proposed in 1990 now became the standard conceptual frameworks to understand gene–environment interaction. On the other hand, the nature of causal interaction has been extensively studied in epidemiology, particularly according to the general model of causation first proposed by Rothman in 1976 and later expanded by Greenland et al. We show that this model, which we call counterfactual sufficient cause model, not only is fully compatible with the models by Ottman, but also provides a broader paradigm for causal interactions in general, including gene–environment and gene–gene interactions, as well as those among environmental factors. The general model of causation offers intuitive explanations within a unified conceptual framework for gene–environment interaction and many other important issues, such as the strength of effect of a genotype, induction period for an occurrence of disease, and confounding by genetic or environmental factors.

IGES-89

On Maximizing Multipoint Heterogeneity Lod Statistics over the Genetic Parameter Space Constrained by Sample Prevalence

C.-H. Chen[1], S.K. Nath[1], S.J. Finch[2], N.R. Mendell[2], D. Gordon[3], G.C. Wedig[1], R.C. Elston[1]
[1]Dept. of Epi & Biostat, Case Western Reserve Univ., USA; [2]Dept. of Applied Math and Stat, SUNY at Stony Brook, USA; [3]Lab. of Statistical Genetics, Rockefeller Univ, USA

We propose a strategy for maximizing multipoint lod scores in the presence of heterogeneity (HLOD) based on Smith's [1961]admixture model. Using the S.A.G.E. program MLOD, the maximization is over the genetic parameter space (allele frequencies, penetrances and admixture proportion) constrained by the prevalence estimated from the corresponding sample. We simulated disease traits under different genetic models on the 30 CEPH families for which complete genomic scan data, 382 markers on 22 autosomes, are available. The trait data are simulated independent of the marker data in order to determine the null empirical distribution of the maximum HLOD scores, and hence cutoffs necessary to control genomewide Type I error.

IGES-71

The Closed Biological Set — A New Approach in Classifying Small Epitopes

J.M. Clark, T.J. McCrary, G.E. Bonney
Statistical Genetics and Bioinformatics Unit, National Human Genome Center at Howard University, Washington D.C.

With the completion of the human genome DNA sequence many consider searching of sequence databases by existing methods already inadequate for a comprehensive analysis. Our approach calls on a characterization of biological sequences from a topological perspective. In this approach, the length of a sequence and its string of symbols is its most important property. It establishes the sequence as belonging to a topological closed set of the symbols for a given sequence length. A version of a new classification system developed by MCRA Applied Technology Inc. is now available at The National Human Genome Center at Howard University.

The major histocompability region of the genome has been extensively investigated using methods for conducting homology searches, such as BLAST. All existing homology searches frequently yield sequences associated with each other in a way that cannot be biologically interpreted. This is particularly true in the HLA region where homology searches are not very accurate on sequences shorter than 500 base pairs. In fact, for sequences shorter than 22 base pairs there exists no significant analytical results. With the new classification system we are capable of analyzing sequences of any length and place them into topological near neighborhoods deterministically. Finally, we have a tool that classifies genes and signal sequences of small length.

IGES-91

Overparametrization of Linkage Tests for Multifactorial Diseases

F. Clerget-Darpoux , H. Selinger-Leneman
INSERM U. 535 — Le Kremlin Bicêtre , France

A major challenge for Genetic Epidemiologists is to identify the genetic factors involved in multifacto-

rial diseases. For these diseases, the number of factors involved, the importance of their individual effect and the level of heterogeneity are unknown.

The information available through the marker segregation in families of affected may be sufficient for detecting linkage and thus for indicating a region in which a susceptibility gene lies. However, it does not generally allow for any simultaneous inference of the susceptibility locus position, of the parameters specifying the correspondence between the disease phenotypes and the genotypes at the susceptibility locus and of the degree of linkage heterogeneity in the data. We show that the statistics involving all these parameters are overparametrized and consequently do not provide more power than less parametrized statistics. In that respect, properties of the HLOD and NPL statistics are compared in the context of a genome scan on a sample of 100 affected sibpairs and a 10cM spaced marker map. The threshold of each statistics corresponding to a global type I error a is first assessed. Then the power — at the same a level — of the two statistics is compared for different situations regarding the model at the disease locus and the level of linkage heterogeneity.

IGES-31

Sample Size Requirements to Control for Stochastic Variation in Magnitude and Location of Allele-Sharing Linkage Statistics in Affected Sibling Pairs

H. Cordell
Department of Medical Genetics, University of Cambridge, UK

Typically, genome scans for complex disease have produced linkage peaks which have proved difficult to replicate in additional independent studies. Here we confirm that this may be due to the large variance in magnitude and position of the linkage statistics when maximized across a region. Simulations suggest that for genes of moderate effect (locus-specific sibling relative risks RR in the range 1.23–1.39), sample sizes of less than 500 affected sib pairs will give unacceptably large standard errors in the magnitudes and locations of significant linkage results. For genes of small effect (RR < 1.13), sample sizes in the region of 1000–2000 pairs may be required to achieve consistency of results between different studies. These figures have important implications for our confidence in location estimates for disease genes obtained from linkage studies of modest size. In particular, collection of larger data sets and/or analysis strategies such as conditioning or narrowing the phenotype definition, in order to increase the relative effect size, may be required before embarking on positional cloning.

IGES-133

Using a Multiple Threshold Liabilty Model in MX to Investigate Linkage of Alcohol Dependence and Related Phenotypes in the COGA Data Set

J. Corbett[1], J.P. Rice[1], N.L. Saccone[1], L. Bierut[1], A. Goate[1], H. Edenberg[2], J. Nurnberger[2], H. Begleiter[3], T. Reich[1]
[1]Washington University School of Medicine, St. Louis, MO; [2]Indiana University Medical Center, Indianapolis, IN; [3]SUNY Health Sciences Center, Brooklyn, NY

Previous linkage analyses of the COGA data set using COGA (DSM-III-R plus Feighner Definite) affectation, DSM-IV affectation, and ICD-10 affectation have found regions of interest for one or more of these diagnoses on chromosomes 1, 2, and 7. For nicotine dependence, suggestive linkage to "habitual smoking" was found on chromosomes 2, 9, and 15 in the COGA data.

The three alcohol dependence diagnoses in COGA are almost completely nested. We propose a multiple threshold liability model with five classes: pure unaffected, unaffected, COGA affected, DSM-IV affected, and ICD-10 affected, with subjects classified by their most severe diagnosis.

We analyze the estimated liability with a Variance Components method in the program Mx, with IBD estimates from all possible sib pairs. The most significant result for alcohol dependence was on chromosome 1 at the marker D1S532, with a lod score equivalent of 2.6. This is in the region of linkage reported for COGA affectation. The other significant result was found on chromosome 4 near the ADH3 locus, with a lod score equivalent of 1.7. While not in a region reported for linkage to alcohol dependence in COGA, it is at the same marker where a signal was reported for the quantitative trait "Maximum Number of Drinks Consumed in a 24 Hour Period," as well as near a possible protective locus found using only pure unaffected subjects.

IGES-38

A Single Nucleotide Polymorphism Genome Scan for a Novel Glomerulocystic Kidney Disease Locus.

J.S. Collins[1], B. Tanriover[2], M.L. Robbin[2], R.C.P. Go[2], L.M. Guay-Woodford[2]
[1]Greenwood Genetic Center, Greenwood, South Carolina, USA; [2]University of Alabama at Birmingham, Birmingham, AL, USA

This is a study of a three-generation African-American family with Glomerulocystic Kidney disease (GCKD) which is unlinked to known PKD1 and PKD2 loci (Sharp et al. 1997). We performed a whole genome scan with 10 family members using the Affy-

metrix GeneChip HuSNP Mapping Assay. Pairwise and multipoint analyses using the GENEHUNTER program identified candidate intervals for this novel GCKD locus on chromosome (chr) 11p15 and chr 22q11. Further targeted analyses using microsatellite markers confirmed the suggestive linkage at chr 11p15. These analyses excluded linkage between this disease and the chr 5q31 PKD2L2 locus, the chr 17 HNF1beta locus, and human homologues of the mouse jck and orpk loci. However, only 980 of the 1494 single nucleotide polymorphism (SNP) markers were assigned, which led to several under-represented chr intervals.

We propose to use information from the Human Genome Project and the SNP Consortium to help integrate these SNP markers with microsatellite and radiation hybrid maps. We are currently typing these SNPs in 5 additional family members, as well as 20 unrelated African-American individuals to help establish allele frequencies in the African-American population. These analyses will be redone using all of this information to follow up the suggestive evidence for linkage of this novel GCKD locus to chr 11p15.

IGES-123

Hierarchical Modeling of Linkage Disequilibrium

D. Conti[1], J. Witte[1]
[1]Case Western Reserve University, Cleveland, OH, USA

In fine-scale gene mapping by linkage disequilibrium (LD), patterns of association may be disrupted by considerable variation in LD estimates due to population specific phenomena such as mutation, genetic drift, mating patterns, and allele frequencies. Spurious results may also arise due to multiple tests across numerous loci. To address these problems, we propose a hierarchical linkage disequilibrium (HLD) model for fine-scale mapping. This approach incorporates higher-level information, such as inter-marker distance on the chromosome, to yield more stable and plausible measures of association. Specifically, reduction of estimation error through the integration of this information refines LD patterns and addresses issues of multiple comparisons. HLD increases power without requiring the estimation of haplotypes and allows for the inclusion of covariates. We present the HLD framework and results from its application to an association study of prostate cancer using multiple SNPs within several androgen pathway candidate genes. Moreover, by including information about the relationships between candidate genes within this biochemical pathway, we show how one can expand HLD to further enhance estimates of LD.

IGES-25

Grade-of-Membership Sibpair Linkage Analysis Maps IDDM11 to Chromosome 14q24.3-q31

E.H. Corder[1], M.A. Woodbury[1], K.G. Manton[1], L.L. Field[2]
[1]The Center for Demographic Studies, 2117 Campus Drive, Duke University, Durham, NC 27701-0408, U.S.A.; [2]Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada

The use of Grade-of-Membership analysis (GoM) (Manton et al. 1994) is demonstrated for sibpair linkage analysis: GoM was used to map the IDDM11 locus to the region of chromosome 14q24.3 identified by Field et al. (1996). Haplotype groups were constructed from sib pair information on the number of shared alleles. The sample consisted of 578 sibling pairs found in 246 multiplex IDDM families. Both siblings were diabetic in 53% of the pairs (AA). Pair members could share 0, 1 or 2 alleles IBS at each of eight linked marker loci spanning IDDM11. Three model-based groups best represented the data on allele sharing: The groups corresponded to 'No', 'One', and 'Two' shared haplotypes for the region. Group 'Two' was larger (37% versus 25%, $p < 0.0001$) and more homogeneous ($p < 0.0001$) than expected by chance consistent with the IDDM11 locus being a determinant of diabetes in multiplex families. Genetic linkage of IDDM to the region was demonstrated by a 19% increase in the proportion of AA pairs over the haplotype groups: 'No': 42%; 'One': 49%; 'Two': 61%, $p=0.0005$ – a 43 % relative increase. The approach can be further adapted for use in multi-locus linkage/association studies and to define phenotype profiles related to specific underlying genetic profiles.

IGES-75

Segregation Analysis of Colorectal Cancer families: Comparison of results from PAP and POINTER.

G.P. Crockford[1], J.H. Barrett[1], D.T. Bishop[1], D.J.B. St John[2], E.A. Debney[2], F.T. McDermott[2], E.S.R. Hughes[2]
[1]Imperial Cancer Research Fund, Leeds, England; [2]Dept. of Gastroenterology, The Royal Hospital, Melbourne, Australia.

Segregation analysis of susceptibility to colorectal cancer (CRC) assuming a mixed model has uniformly identified a rare dominantly inherited susceptibility with penetrance up to age 80 of 0.5 or higher. We investigate this and other competing models using PAP and POINTER. Detailed family histories were obtained and verified for a series of 525 CRC patients of Sir Edward Hughes in Melbourne. Four of these families are known to have

mismatch repair mutations. Segregation analysis was performed on the remaining 521 families. Assuming the mixed model POINTER identified a rare dominantly inherited major gene (allele frequency q=0.0014) with lifetime penetrance of 0.5 and no residual heritability. Repeating the analysis in PAP the best fitting model identified a common dominantly inherited gene (q=0.31) with within genotype polygenic heritability of 0.34. This model predicts a lifetime risk of 0.05 to gene carriers and zero to non-carriers. The heritability estimators in POINTER and PAP are not directly comparable since they use different denominators. Further analysis, using POINTER and fixing heritability identified a model which maximised to a better likelihood than previously identified, with a recessively inherited gene (q=0.715) with lifetime penetrance for carriers of 0.05 and zero otherwise. We investigate the plausibility of the competing models and their implications.

IGES-150

Investigation of Gene–Gene and a Between Gene–Environment Interaction Involving the NAT1 and NAT2 polymorphisms, Associations with Colorectal Cancer/Polyps in the Cleveland Colon Neoplasia Sibling Study:

D. Daley[1], R.C. Elston[1], S. Markowtiz[2,5], A. Wiess[3], J.S. Witte[1], P. Platzer[5], S. Lewis[3], M. Diaz[1], G.L. Wiesner[2,3,4]

[1]Departments of Epidemiology and Biostatics; [2]Medicine and [3]Genetics, Case Western Reserve University; [4]Center for Human Genetics, University Hospitals of Cleveland; [5]Howard Huges Medical Institue Laboratory, at Case Western Reserve University

In response to environmental insults, several enzyme superfamilies have evolved that metabolize and detoxify foreign compounds, such as the N-acetyltransferases (NAT) genes. NAT1 and NAT2 are closely linked and genotypes code for either fast or slow acetylator activity.

We examine the joint genotypic and phenotypic correlations between colon neoplasia and NAT1/NAT2,in siblings who have developed colon cancer or adenomatous polyps before age 65. By genotyping NAT1, NAT2 and D8S261, a dinucleotide marker in close linkage with NAT. Our interim linkage analysis suggests an association between NAT and colon neoplasia, which is consistent with other studies. Mean allele sharing with 186 sib pairs, gives evidence of a potential protective allele at D8S261 and NAT1, the proportions of concordantly unaffected sibs sharing 2 alleles ibd being .5850 and .5867 respectively. Results obtained using the TDT/S-TDT demonstrate evidence for both protective and punitive associations (see table), 6/18 alleles are under represented in the affected population, suggesting linkage disequilibrium. We plan to examine the mean allele sharing, haplotype sharing at NAT1/NAT2, gene/gene/ and gene/enviroment interactions with smoking. Results obtained from using standard linkage analysis with GENEHUNTER and the LINKAGE will be compared with those from using the new Haseman-Elston algorithm in S.A.G.E.

D8S261 (Marker)	Allele 130 protective chi square=3.769 (p=.05)	Allele 142 punitive Z=1.696 (p=.05)
NAT 2 (Marker)	Allele 5B protective Z=1.74 (p=.04), codes for slow pathway*	Allele 7A punitive Z=1.565 (p=.059) codes for fast pathway*

Bartsch et al (1998) reported 64% of patients had a slow pathway at NAT-1, with 36% fast and 64% slow, indicating a protective effect of a fast pathway at NAT-1, we find the same percentages of patients with slow and fast pathways at NAT 1, 36% and 64% re

IGES-138

An Improved Score for MCMC Genome Screening

E.W. Daw[1], S.C. Heath[2]

[1]U.T. M.D. Anderson Cancer Center, Houston, Texas, USA; [2]Memorial Sloan-Kettering Cancer Center, New York, New York, USA

Monte Carlo Markov chain (MCMC) techniques offer a promising approach in dissecting complex traits. Methods implemented in Loki have localized genes for complex traits in real and simulated data. Loki iteratively places quantitative trait loci (QTL) on chromosomes (chr) to estimate posterior probability of linkage (PPL). We estimate PPL over location (l), and QTL variance contribution (c), to examine relative importance of loci. Previously we introduced the Log Of the Posterior placement probability ratio (LOP) to facilitate significance assessment. LOP compares PPL on real chr (T) to PPL on pseudo chr (S) unlinked to the trait. S is generated by random gene drop to match T for the map, allele freq., and missing data patterns. On a grid of l and c: LOP(l,c)=log10(P(l,c|T)/P(l,c|S)), where P(l,c|T) is posterior probability of a QTL on T at l and c, and P(l,c|S) is same for S. Our previous LOP estimate used hits on T and S, and may converge slowly, although non converged estimates are practical for linkage studies. Here, we estimate LOP using Rao Blackwell techniques with probability ratios computed by Loki in each iteration. We find 2-trait locus data with maximum 2-point lod scores of ~0.4 produces significant LOP. We estimate LOP >20 for some traits, so our previous estimate could take billions of years to converge. The new estimate is more efficient, taking hours.

IGES-109

Linkage Simulation Enhances Decision Making Strategies for Study of Lung Cancer (LC) Pedigrees.

M. de Andrade[1], J. Slusser[1], C.I. Amos[2], J. Bailey-Wilson[3], P. Fain[4], M.W. Anderson[5], G.M. Petersen[1]
[1]Health Sciences Research, Mayo Clinic, Rochester, MN; [2]Epidemiology, UT M. D. Anderson Cancer Center, Houston, TX; [3]NGHRI, Bethesda, MD; [4]University of Colorado, CO; [5]University of Cincinnati, OH.

As the Genetic Epidemiology of Lung Cancer Consortium identifies high risk LC families, a challenge has been in deciding which families/individuals to study, given high sporadic rate and difficulty of accrual. Objective: To assess the value of linkage simulation as a way to inform data collection and fine mapping of high risk LC pedigrees. Methods: One thousand replicates of a 40 member multiplex genotyped LC pedigree were simulated using SLINK and GENEHUNTER, compared to the observed NPL score (1.7) with a 13-allele marker. We evaluated maximum information in the pedigree that could be obtained with fine mapping. Results: Bimodal distribution of expected NPL scores were observed with this pedigree structure. Compared to the observed NPL score, the pedigree was found to contain additional power to detect linkage by fine mapping. The inclusion of a newly diagnosed affected LC individual will give sufficient power to detect linkage. Conclusions: Linkage simulation can enhance decision-making in LC mapping studies. Following genome scanning, simulation studies of the power to detect linkage by fine mapping in specific families may direct data collection strategies.

IGES-28

Fine-Scale Mapping with Incomplete Initial Association

J. Dong[1], R. Jiang[1], D. Wang[2], F. Sun[2]
[1]Department of Mathematical Sciences, Michigan Technological University, Houghton, MI 49931, U.S.A.; [2]Department of Mathematics, University of Southern California, Los Angeles, CA 90089,U.S.A.

Our main objective is to introduce a new measure of fine-scale mapping. One of the most widely used measures for fine mapping is p(excess)=(p(i)-q(i))/(1-q(i)) (Bengtsson and Thomson 1981), where p(i), q(i) are the sample allele frequencies of the allele i among the cases and controls, respectively, and the allele i is most closely related to the disease. When a disease gene A was first introduced into the general population, if all the chromosomes carrying disease gene A have the same haplotype, we say that there is a single ancestral haplotype, otherwise, there are multiple ancestral haplotypes. Assuming a single ancestral haplotype, p(excess) is an appropriate measure for fine-scale mapping. We prove that, assuming multiple ancestral haplotypes, p′(excess) is no longer an appropriate measure. Let c(i)=P(i|A) be the conditional probability of a chromosome having allele i given that it has allele A when A was first introduced into the general population. Our new measure is p'(excess)=(p(i)-q(i))/(c(i)-q(i)). We have proved that the magnitude of the new measure is a decreasing function of the genetic distance between the marker locus and the disease locus, and its maximum occurs at the disease locus. We propose a novel method to estimate c(i) and to estimate the location of the disease gene. The new method is successfully applied to several real data sets.

IGES-41

A Multivariate Logistic Model for Computing the Likelihood of a Pedigree Without Loops

Y. Dong[1], M. Snell[1], R. Elston[1]
[1]Dept. of Epi & Biostat, Case Western Reserve Univ., USA

Karunaratne and Elston [American Journal of Medical Genetics 76:428–437] proposed a multivariate logistic model for analyzing a binary trait measured on individuals in a nuclear family. This model has the advantage that the marginal probability of being affected is the same for all members of the pedigree who have the same covariate values. This marginal probability is a logistic function, where the logit depends on both a major gene effect and covariates. We consider here the situation of a pedigree without loops. We show that, when the regressive model assumption holds for pedigree members in different nuclear families, the likelihood function can be expressed as a sum of products of likelihood functions for the various nuclear families contained in the pedigree, divided by the joint likelihood function for those members who belong to more than one nuclear family. We thus establish a recursive algorithm for construction of the likelihood for the pedigree when members who belong to more than one nuclear family do not have missing phenotypes. Finally, we consider ways of dealing with missing phenotypes and introduce a simplifying approximation formula.

IGES-99

Power of Variance Component Linkage Analysis to Detect Quantitative Trait Loci in Inbred Sibships

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

Linkage analysis using the variance component method is now a widely used approach for mapping genes influencing quantitative trait variation in human populations. However, substantial work is still needed on the optimal design of family studies required for this task. In this paper, we consider extensions of the variance component method to inbred sibships. Although performing genetic analysis on inbred pedigrees may be a difficult computational challenge, such pedigrees are rich in linkage information. Even by themselves, inbred individuals are informative for linkage. Williams and Blangero [1999] derived exact expressions for the expected LOD score (ELOD) for various sampling structures, including sib pairs and larger sibships. Extending these results to include inbreeding, we have derived analytical expressions for the ELOD of a single inbred individual and inbred sibships. The accuracy of our derivations was verified using computer simulations. We apply this analytical theory to consider the utility of study designs involving common consanguineous matings, such as cousin marriages. Our results indicate that the recruitment of inbred sibships may represent a powerful alternative strategy for QTL localization in human populations. This work was supported by NIH grants MH59490 and HL28972.

IGES-88

Resampling Methods for Linkage Analysis of Quantitative Traits under Heterogeneity

C.T. Ekstrom
Department of Biostatistics, University of Copenhagen and Steno Diabetes Center

Genetic heterogeneity is one of the major problems in linkage analysis of quantitative traits. The use of variance component models for linkage analysis of quantitative traits have proven to be a powerful tool for detecting and locating genes, but the presence of genetic heterogeneity will decrease the power of a linkage study and may even give biased estimates of the location of the quantitative trait loci.

We consider a mixture of multivariate normals to model locus heterogeneity by allowing only a proportion of the sampled pedigrees to segregate a disease allele at a specific locus. However, for mixtures of normals the classical distribution theory of the maximum likelihood estimates does not hold, and tests of linkage and/or heterogeneity must be evaluated using resampling methods.

Three different test statistics for multipoint linkage analysis in the presence of heterogeneity are examined and compared to traditional linkage analysis in a simulation study. In addition, estimation problems and cases where this mixture approach is feasible are discussed.

IGES-118

The Impact of Heterogeneity on Meta-Analysis of Genetic Linkage

C.J. Etzel
Department of Epidemiology, UT M D Anderson Cancer Center, Houston

Investigations in the development, application and implication of meta-analysis are abundant in current genetic epidemiological literature as the debate over the value of meta-analysis continues. An issue in the forefront of this debate is heterogeneity and its effects on the underlying analysis. In the field of genetic linkage, heterogeneity has many definitions and in itself is a dynamic debate all its own. In this investigation, we simulate four types of between-study heterogeneity that are possible in a meta analysis of genetic linkage of a quantitative trait locus (QTL): population heterogeneity (different marker allele frequencies across studies), marker heterogeneity (different marker maps across studies), ascertainment heterogeneity (different ascertainment schemes across studies), and environment heterogeneity (different environment effects across study populations). Within each study, genetic linkage to a QTL is assessed using the Haseman-Elston (1972) method and meta-analysis is performed using the methods proposed by Li and Rao (1996) and Etzel (1999). The effects of the above mentioned types of heterogeneity are assessed by comparing the estimates of total genetic variance and location of the QTL from the meta-analyses to true parameter values.

IGES-76

Joint Role of HFE Gene Mutations and Environmental Risk Factors in Hepatocellular Carcinoma

S. Fargion[1], M.A. Stazi[2], A.L. Fracanzani[1], M. Mattioli[1], M. Sampietro[1], D. Tavazzi[1], C. Bertelli[1], V. Patriarca[2], C. Mariani[1], G. Fiorelli[1]
[1]Università di Milano, Ospedale Maggiore IRCCS, Milano; [2]Istituto Superiore di Sanità, Roma, Italy

To elucidate the interaction between genetic and environmental factors on cancer occurrence is crucial for disease prevention. The aim of this study was to estimate the joint contribution of the genotypes C282Y and H63D in the HFE gene associated with hereditary hemochromatosis (HH) and other established risk factors to hepatocellular carcinoma (HCC) occurrence in Italian patients.

The case-only approach, specifically designed to estimate departure from multiplicative risk ratios under the assumption of independence between genotype and environmental exposure, was adopted. The HFE gene mutations, performed by PCR, were analysed in 63 HCC male patients consecutively enrolled in our clinic;

none of the patients had a phenotype compatible with homozygous HH. Data on alcohol consumption, HBV and HCV previous infections, iron status were collected at the enrolment.

Iron overload was associated to C282Y and to H63D. At the multivariate analysis the interaction effect between any HFE mutation and exposure to HBV and/or HCV was 2.9 (95%CI 0.9–9.5), while the interaction effect between any HFE mutation and alcohol abuse was 3.0 (95%CI 0.7–14.0). There is a suggestion that heterozygotes for the HFE gene exposed to hepatitis infection or alcohol abuse have a risk of developing HCC three times higher than HFE wild type people exposed to the same risk factors.

IGES-143

Comparison of GxE Interaction Detection Methods

D. Fallin, Y. Yao, L. Kao
Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD

Several recent reports have advocated the use of case-only and case-partial-control (partial G or E information on controls, full information on cases) designs to detect gene-by-environment (GxE) interactions. These methods require an assumption of independence between G and E in healthy individuals. We have designed a Monte Carlo simulation strategy to compare type 1 error rates and power to detect GxE interaction across designs and statistical methods to determine which strategies are most robust to violations of the GxE independence assumption, under various models of G, E, and disease frequencies, effect sizes and interaction scenarios. For example, we find the case-only type 1 error rate for a test of interaction to be .049* when G and E are independent. However, under GxE dependence**, this error rate is .654*, although the error rate for GxE interaction using the full case-control data is .044*. We have explored case-control, case-only, case-partial-control, and case-parent trio designs using measures of interaction including odds ratios (OR), χ^2 and Wald tests, and likelihood ratio tests using logistic, conditional logistic and log-linear regression. We also focus on the relative utility of different methods to maximize power for detection of genetic effects.

*1000 sims; 300cases/controls each; model: ORG=2, ORE=1, ORGE=2(ORI=1) **ORGxE; ctrls=2

IGES-140

Familiality Between Free Fatty Acids and Insulin Sensitivity Before and After 20-Weeks Exercise Training: The HERITAGE Family Study.

M.F. Feitosa[1], Y. Hong[1], S.J. Weisnagel[2], T. Rice[1], I.B. Borecki[1], T. Rankinen[3], A.S. Leon[4], J.S. Skinner[5], J.H. Wilmore[6], C. Bouchard[3], D.C. Rao[1]
[1]Biostatistics, Washington Univ, MO; [2]Kinesiology, Laval Univ, QC, Canada; [3]Pennington Center, Louisiana State Univ, LA; [4]Kinesiology, Univ Minnesota, MN; [5]Kinesiology, Indiana Univ, IN; [6]Health and Kinesiology, Texas A&M Univ, TX

Insulin resistance is associated with increased risk of cardiovascular disease. It has been suggested that free fatty acids (FFA) relate to insulin action. This study assessed the familial resemblance for serum FFA and the concomitant familial effects on FFA and insulin sensitivity (SI) at baseline and in response to 20-weeks of exercise training using data from 502 Whites and 277 Blacks. SI was calculated using minimal model analysis of intravenous glucose tolerance tests. The data were adjusted for age, sex and BMI. SEGPATH correlation analyses provided evidence of familial resemblance for FFA and SI in both races at baseline levels, but not in response to exercise training for FFA. The heritability of FFA and SI at baseline were 16% and 44% in Whites and 30% and 38% in Blacks, respectively. Since FFA response was non-familial, we only performed cross-trait bivariate FFA and SI analysis at baseline levels. The results showed significant cross-trait familial resemblance for Whites but not for Blacks. For Whites, the cross-trait heritability between FFA and SI was 14%, i.e. the amount of covariation that is due to common familial causes. In conclusion, there is a modest familial/genetic influence for FFA at baseline but not for its response to exercise training in both races and there is an indication of common familial/genetic factors influencing baseline FFA and SI in Whites.

IGES-132

Malignant Trigenimal Schwannomas in the Rat: Reevaluation of Strain-Specific Susceptibility using Haplotype Sharing Methods.

C. Fischer[1], L. Beckmann[2], G. te Meerman[3], A. Kindler-Röhrborn[4], J. Chang-Claude[2]
[1]Inst. of Human Genetics, Univ. of Heidelberg, Germany; [2]German Cancer Research Center, Heidelberg, Germany; [3]Dep. of Medical Genetics, Univ. of Groningen, The Netherlands; [4]Inst. of Cell Biology Cancer Research, Univ. of Essen, Germany

Recently haplotype sharing analysis (HSA) methods for detection of disease genes for complex traits in homogenous populations have been proposed (1). HSA has proven to be successful for genes with strong influence on affection status in a sample of moderate size in a simulated isolated population (2). Inbred rat strains with differing sensitivity to experimental tumor induction have been used to search for genes that may be responsible for the development of malignant trigenimal schwannomas. On chromosome 10 a significant result

was found by single point association methods (2). Here, we applied HSA on marker data of rat chromosome 10 in two different ways: 1. comparison of case haplotypes with haplotypes of longterm survivors and 2. comparison of case haplotypes with haplotypes randomly generated under the null hypothesis. Interestingly, comparison with longterm survivors points to a centromeric region which was not significant in the initial evaluation. Because of the simple situation chromosome-wide significance can be estimated by simulation methods. Attempts to evaluate chromosomewide significance of haplotype sharing statistics in general are discussed.

1. Levinson DF et al (2001), Am J Med Genet 105: 65–70
2. Beckmann L et al (2001), Genet Epidemiol to appear
3. Kindler-Röhrborn A et al (1999), Cancer research 59:1109–1114

IGES-45

Molecular and Statistical Analysis of a Novel Susceptibility Gene for Persistent Hepatitis B Virus Infection.

A.J. Frodsham[1], S. Best[1], L. Zhang[1], K. Young[1], S. Lobello[2], C. Venturi-Pasini[2], H.C. Thomas[3], M. Chiaramonte[2], M.R. Thursz[3], A.V.S. Hill[1]
[1]The Wellcome Trust Centre for Human Genetics, Oxford, OX3 7BN, UK; [2]Instituto di Medicina Interna, Universita di Padua, 35126 Padova, Italy; [3]Imperial College School of Medicine, St. Mary's Hospital, London, W2 1NY. UK

Hepatitis B is a disease of global importance, with over 350 million carriers of the virus worldwide. Around 1 million deaths per year can be attributed to the end stage complications of chronic HBV infection. Evidence of a genetic effect on the determination of the outcome of infection has been provided by twin studies and by candidate gene association studies. A full genome wide scan was completed using 21 extended pedigrees from Italy with the peak LOD score located at 6q26-27. Subsequent analysis in a further, smaller cohort of families and fine mapping of the region replicated the previous findings. The combined analysis resulted in a peak LOD score of 2.7 ($p=0.0002$). A region of association at 6q27, away from the peak of linkage, has been identified using TDT ($p<0.05$), in both the above population and a larger cohort from The Gambia. No previously identified genes are located in this region. Novel SNPs have been identified in the contiguous region with the aim of defining the boundaries of this association further. These new SNPs have been identified within, or close to, putative exons predicted by computational analysis, confirmed also by homology to expressed sequence. Statistical analysis of the SNP and microsatellite data will enable us to determine which of these putative exons are contained within the novel hepatitis B susceptibility gene located at 6q27.

IGES-131

Inconsistent Linkage Results for the Chromosomal Region of the ApoAI-CIII-IV Complex and FCHL: An Ascertainment Effect?

F. Gagnon, G.P. Jarvik, A.G. Motulsky, S.S. Deeb, J.B. Brunzell, E.M. Wijsman
University of Washington, Department of Medicine, Seattle, WA, USA

We initiated a genome screen for HDL-C level in 4 large (N=258) familial combined hyperlipidemia (FCHL) families, ascertained through hypertriglyceridemic probands. HDL-C is correlated with the lipid levels that define FCHL. Using joint linkage and segregation analysis based on Bayesian Markov chain Monte Carlo methods, we estimate that 3–4 genes are involved. Our preliminary results, focussing on chromosomes (ch) for which linkage to FCHL or HDL-C has been reported, provide evidence for HDL-C linkage to ch 11q and 13q. Several candidate genes including the ApoAI-CIII-AIV complex, are in the critical region on 11q. From a joint multipoint analysis with all markers on ch 11 and 13, the posterior:prior probability of linkage (in a 4 cM interval) is ~12 and ~15 for ch 11 and 13, respectively. Use of APOCIII as a major-gene covariate reduced evidence for linkage to 11q. Thus APOCIII may be the HDL-C locus on 11q. We are following up on additional signals; preliminary results do not confirm previous reports of linkage to ch 8 and 15. Previous reports of linkage of FCHL to this ch 11 region have been published by others, but results remain controversial and inconclusive. Ascertainment of families among studies varies with respect to procedures that could affect the HDL-C distribution, and thus could explain the controversial results obtained for linkage of FCHL to the 11q23 region.

IGES-77

Detecting Genotyping Problems by Excess Rates of Homozygous Genotypes

F. Geller[1], J. Hampe[2], S. Schreiber[2], A. Ziegler[1]
[1]Institute of Medical Biometry and Epidemiology, Philipps-University, Marburg, Germany; [2]1st Medical Department, Christian-Albrechts-University, Kiel, Germany.

Objective: The aim of our investigation is the identification of genetic markers that distort results because of null alleles. These null alleles may be caused by mutations in the primer binding site or by allele lenghts outside the reading frame. If not identified, this results in excess homozygosity rates and may subsequently lead to spurious findings in linkage analysis. As an example consider an affected sib pair study with incomplete genotype information for the parents. Here, inflated homozygosity rates lead to an overestimation of the IBD status.

Methods: In our presentation we consider affected sib pair studies without parental genotype information. Using a test of homogeneity for genotype frequencies, we test whether the rates of homozygous genotypes are consistent with the observed allele frequencies. We illustrate the method by analysing data for Crohn's disease and ulcerative colitis. Additionally, we investigate the bias that occurs in linkage analysis due to null alleles.

Conclusions: Significant results in linkage analysis should be checked for an excess homozygosity rate. This holds especially when the result is obtained for a single marker and cannot be confirmed in neighbouring markers.

IGES-84

Cancer Risk in Heterozygotes for Ataxia-Telangiectasia (AT).

B. Geoffroy-Perez[1], N. Janin[2], K. Ossian[2], A. Laugé[3], M.F. Croquette[4], C. Griscelli[5], M. Debré[5], B. Bressac-de-Paillerets[2], A. Aurias[3], D. Stoppa-Lyonnet[3], N. Andrieu[1]
[1]U521 Inserm, Villejuif; [2]IGR, Villejuif; [3]Institut Curie, Paris; [4]Centre Hospitalier Feron-Vrau, Lille; [5]Hôpital des Enfants Malades, Paris, France

Predisposition to cancer has been suggested among AT heterozygotes. Now haplotyping can identify heterozygotes for AT mutated gene (HetATM) in AT families allowing the risk of cancer associated with ATM heterozygosity status to be better assessed. A family study of AT patients was performed to estimate the risk of cancer according to different probabilities of being HetATM. We collected the demographic characteristics and occurrence of cancer in 1423 relatives of AT patients. Haplotyping was performed in living relatives. The probability of being HetATM was calculated for deceased relatives. The risk of developing cancer was estimated in the cohort of relatives, and expected numbers of cancer cases calculated from French incidences. No increased risk was observed among relatives for all-site cancers. However, an heterogeneity was found according to ATM heterozygosity status. This was mainly explained by the increased risk of BC observed among HetATM (RR=2.9). Among HetATM, the RR of cancer was non-significantly increased by about 2-fold for sites like stomach, bladder, thyroid, leukaemia and liver. Risks of ovarian, pancreatic, lung, kidney and colorectal cancers were non-significantly increased among relatives with a 50% probability of being HetATM. Except for BC, RR of cancer was not significantly increased at any site.

IGES-33

Integrating Sibship Data For Mapping Quantitative Trait Loci

S. Ghosh, T. Reich
Washington University School of Medicine, St. Louis, USA

Existing methods for linkage analysis divide sibships into all possible sib-pairs using weights to account for non-independence. Simulations show that inferences on linkage are greatly influenced by the choice of weights. We propose a statistical procedure which integrates data on entire sibships into a "contrast function", defined as a linear combination of the quantitative values within a sibship, with the sum of the coefficients being zero. We develop a regression procedure of the squared contrast function on a quadratic function of the matrix of estimated identity-by-descent scores at a marker locus, which is a natural extension of the classical Haseman-Elston approach using sib-pairs (1972). We also combine the mean with the contrast function to provide more information on linkage. The method is extended to multiple, epistatically interacting trait loci. Monte-Carlo simulations are used to assess the efficiency of the proposed method and to evaluate the marginal effects of dominance and heterozygosity. The applicability of our method is illustrated using COGA data on alcohol dependence.

IGES-37

The Sib-TDT Adjusted For Age Of Disease Onset

S. Ghosh, T. Reich
Washington University School of Medicine, St. Louis, USA

The transmission disequilibrium test (TDT) introduced by Spielman et al. (1993) requires data on marker allelic transmissions from heterozygous parents to affected offspring. For diseases with late age of onset, it is not possible to obtain data on parents. Spielman and Ewens (1998) proposed a sib-based TDT in which they compared the proportions of a marker allele among affected and unaffected sibs. However, the Sib-TDT procedure does not take into account age of disease onset, that is, it ignores the fact that some of the currently unaffected sibs may in reality possess the disease mutation and manifest the disease in the future. The purpose of this paper is to adjust the test statistic for age of onset in order to increase its power. We use the Kaplan-Meier survival function (1958) to estimate the conditional affectation probability of unaffected sibs given their marker genotypes, and weight the observed number of affected and unaffected sibs in each sibship. We also propose adjustments for covariates via the Cox Proportional Hazard Model (1972). Monte-Carlo simulations are used to evaluate the increase in power of the Sib-TDT. The performance of the adjusted Sib-TDT is assessed for different levels of sensitivity and specificity of the diagnostic test. The applicability of our method is illustrated using data on Alzheimer's Disease.