Genetic Epidemiology 27: 258–310 (2004)

ABSTRACTS FROM THE

THIRTEENTH ANNUAL MEETING OF THE INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY

Noordwijkerhout, The Netherlands
September 11–12, 2004

1

The mystery of conserved non-genic (CNG) sequences

S.E. Antonarakis
Department of Genetic Medicine and Development, University of Geneva School of Medicine, Geneva, Switzerland.

The comparison of the sequences of human chromosome 21 with that of the syntenic regions of the mouse genome revealed a large number of conserved sequences (>100 nt in length and 70% ungapped identity) that are not transcribed. We called these elements conserved non-genic (CNG) sequences. Most of these map in gene-poor regions of chromosome 21. A large majority of CNGs are also present in several mammalian species, indicating a conservation of more than 120 million years. The patterns of evolutionary conservation allow a sufficient separation of CNGs from both coding regions and non-coding RNAs. Furthermore, the evolutionary characteristics are independent of their position relative to protein-coding sequences. The overall level of conservation of CNGs is higher that exonic sequences and strongly suggests functional importance. We anticipate that mutations in CNGs may contribute to human disorders; a search for those is now in progress.
The function of CNGs is largely unknown and considerable effort is now devoted to the functional analysis of these genomic elements that may account for up to 1–3% of the human genome. Some CNGs may be cis or trans regulatory elements of gene expression, others may be structural elements, and yet others may have a function totally unsuspected todate.
For a complete description see Dermitzakis et al., Nature 420:578, 2002; Science 302:1033, 2003; Genome Res. 14:852, 2004.
I thank the members of the laboratory and the funding agencies for supporting our research.

2

Evidence for a Heritable Component to Death from Influenza in the Utah Population

F.S. Albright(1), PL Orlando(1), LA Cannon Albright(2)
(1) Dept of Pharmacy Practice, University of Utah College of Pharmacy; (2) Dept of Medical Informatics, University of Utah School of Medicine

We have investigated 4,000 deaths from Influenza for evidence of a heritable component. We utilized a population-based resource consisting of a genealogy of Utah, record-linked to death certificates. We estimated significant increased relative risks for death from Influenza in first-degree relatives of individuals who died of Influenza (RR=1.6; 95% CI:1.47, 1.73). Elevated risks in first-degree relatives may represent shared environment, infectious exposure, genetic contribution, or a mixture of these effects. Relative risks in second-degree relatives may be more indicative of a heritable component, since such relatives share more genes than environment. We estimated significant increased relative risk for death from Influenza in second-degree relatives of individuals who died of Influenza (RR=1.2; 95% CI:1.13, 1.29). We estimated an average relatedness statistic (Genealogical Index of Familiality) for Influenza deaths that was significantly greater than expected (mean relatedness cases 4.03, mean relatedness controls 2.76, $p<0.001$). An excess of both close and distant affected relatives was observed for cases. Multiple high risk pedigrees have been identified. Although there are clusters of Influenza death in time in these pedigrees, there are also close relatives in each pedigree whose death from Influenza occurred decades and miles distant from other relatives who died of Influenza. These results support the suggestion of a genetic component to predisposition to death from Influenza.

3

The COL9A1 Gene and Hip Osteoarthritis, a Population-based Linkage and Association Study

B.Z. Alizadeh(1), O.T. Njajou(2), C. Bijkerk(3), I. Meulenbelt(2), F.C. Breedveld(3), S.C. De Wildt(2), L. Sandkuyl(2), J.M. Te Koppele(4), A. Hofman(1), H.A.P. Pols(1,5), P.E. Slagboom(2), C.M. van Duijn(1)
Departments of (1) Epidemiology & Biostatistics, Erasmus MC Rotterdam; (2) Molecular Epidemiology, LUMC Leiden; (3) Rheumatology, LUMC Leiden; (4) Vascular & Connective Tissue Research, LUMC Leiden; (5) Internal Medicine, Erasmus MC Rotterdam, The Netherlands

Collagen IX proteoglycan has been implicated in hip osteoarthritis. We studied two COL9A1 markers (509-8B2 and 509-12B1) in relation to radiographical osteoarthritis (ROA) in the Rotterdam Study, a population-based study of 7983 subjects aged 55 years or over. We used two different designs; First a sibling pair study of 83 probands with multiple joints affected with ROA, and their 222 siblings yielding 445 sibling pairs who participated in the study. Second, an association study in a series of 71

Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/gepi.20029

cases with hip ROA and 269 controls. All subjects were characterised for the two COL9A1 509-8B2 and 509-12B1 markers. The means test was used to assess the proportion of alleles shared in affected sibling pairs. The chi-square test was used to compare the allele distribution in cases and controls. We found that affected sibling pairs with hip ROA shared more often alleles IBD at the 8B2 (mean 0.66 $\pm$ standard error 0.07) and 12B1 (0.65 $\pm$ 0.08) markers than expected ($P < 0.05$). No excess sharing was observed for ROA at other joint sites. When comparing the allele frequency of 8B2 and 12B1 in cases and controls, the frequency of 8B2 alleles in cases differed significantly ($P < 0.01$) from those of controls. We concluded that COL9A1 is involved in hip osteoarthritis.

4

A major gene for lung cancer on 6q23-25

C. Amos(1), J. Bailey-Wilson(2), S. Pinney(3), G. Petersen(4), M. deAndrade(4), J. Wiest(2), P. Fain (6), A. Schwartz(7), M. You(8), W. Franklin(6), C. Klein(3), A. Gazdar(9), H. Rothschild(10), D. Mandal(10), T. Coons(11), J. Slusser(4), C. Gaba(8), E. Ivanenkov(3), A. Perez(3), X. Zhou(1), D. Seminara(5), J. Minna(9), M. Anderson(3)
(1) MDACC; (2) NIH; (3) Univ Cin; (4) Mayo; (6) U Colorado; (7) Karmanos Can Inst; (8) Med Col of Ohio; (9) UTSW Med Cen; (10) Louisiana St Univ HSC; (11) Saccomano Res Inst

Lung cancer risk is greatly increased by cigarette smoking and occupational exposures, but familial factors also play a major role. We conducted a genome-wide linkage analysis of extended pedigrees with multiple lung cancers. Multipoint linkage analysis of 38 pedigrees with 4 or more affecteds (lung or throat), under a dominant model with decreased penetrance, yielded a multipoint heterogeneity LOD (HLOD) score of 3.47 on chromosome 6q at 156cM, near D6S2436. A subset of 23 multigenerational pedigrees with 5 or more affecteds yielded a multipoint homogeneity HLOD score of 4.26. Fourteen additional families with 3 or fewer affecteds yielded negative LOD scores. A predivided samples test for heterogeneity comparing the LODs in the 23 multigenerational families to the remaining families was significant ($p=0.007$). Additional analyses to incorporate smoking behaviors also supported evidence for linkage but with lower LOD scores. The 1-LOD multipoint support interval (from 149 cM, to 165 cM) from the multigenerational families overlaps a genomic region that is deleted in sporadic lung cancers. These results provide strong evidence for a major susceptibility locus influencing lung cancer risk.

5

The BOADICEA model of genetic susceptibility to breast and ovarian cancer: updating and validation.

A.C. Antoniou(1), P.D.P. Pharoah(2), D.F. Easton(1) on behalf of the Boadicea collaborators.
(1) CR-UK Genetic Epidemiology Unit; (2) Human Cancer Genetics Group, Cambridge University, UK

We previously derived a model of breast and ovarian cancer susceptibility using segregation analysis based on a population based series of 1484 breast cancer cases and 156 high-risk families from the UK. We have updated this model using additional data from two UK population based studies of breast cancer and family data from BRCA1/2 carriers identified in 22 population based studies of breast and ovarian cancer. The combined dataset includes 2785 families (301 BRCA1, 236 BRCA2 positive). According to the model, susceptibility to breast cancer is explained by mutations in BRCA1 and BRCA2 plus a polygenic effect (joint multiplicative effects of many genes of small effect). Incidence rates were smoothed using locally weighted regression techniques to avoid large variations between adjacent intervals. A birth cohort effect on the cancer risks is implemented, whereby each individual is assumed to develop cancer according to calendar period-specific incidence rates. The variance of the polygenic component declines with age, from 3.2 at age 30 to 1.3 at age 60. The predicted familial relative risks and BRCA1/2 prevalence among cases are close to those observed in population studies. The model predicts that the average breast cancer risks in BRCA1/2 carriers increase in more recent birth cohorts. For example, the average cumulative breast cancer risk to BRCA1 carriers is 50% for women born in 1920–29 and 58% for women born after 1950.

6

Linkage analysis for AD using Amyloid Beta 42 levels shows evidence for a novel AD gene on chromosome 19

A. Arias(1), Y.S. Aulchenko(1), K. Sleegers(1), G. Roks(2), M. Cruts(3), C. Van Broeckhoven(3), J.C. van Swieten(4), P. Heutink(5), B.A. Oostra(6), C.M. van Duijn(1)
(1) Department of Epidemiology and Biostatistics; and (4) Department of Neurology and (6) Department of Clinical Genetics, Erasmus Medical Centre, Rotterdam, Netherlands; (2) Department of Neurology, Elisabeth Hospital, Tilburg, Netherlands; (3) Dept. of Molecular Genetics (VIB8) Neurogenetics Group, University of Antwerp, Antwerp, Belgium; (5) Section Medical Genomics, Department of Clinical Genetics and Human Genetics, VU Medical Center, Amsterdam, Netherlands

For several families in which Alzheimer Disease (AD) segregates as an autosomal dominant trait, the genetic origin of disease remains unknown. We aim to identify genes implicated in the pathophysiology of AD using Amyloid Beta 42 levels to obtain more phenotypic information. We performed a genome screen of AD in a Dutch family with mean onset of 64 years that included eight patients. We measured Ab42 levels in 15 non-affected individuals still at risk for AD and performed a non-parametric linkage analysis, for which we defined our disease set in two ways.Our first analysis included only AD cases, and yielded NPL scores of 2.15 for marker D1S450 (approximately 210 cM from the PSEN2 gene), 2.12 for marker D7S2465 and 1.92 for marker D19S571. Second, an analysis of AD cases plus non-affected individuals with

high Ab42 levels, yielded lower NPL scores for chromosomes 1 (0.20) and 7 (0.41), but increased the NPL score for chromosome 19 (3.09). The haplotype analyses of the 3 regions showed a 30 cM region of chromosome 1 for which all AD patients were heterozygous. Also, all AD patients were compound heterozygous for haplotypes in chromosome 7 (~20 cM) or chromosome 19 (~40 cM). The chromosome 19 haplotype does not include the APOE gene. Combining the phenotypic data, we see that the strongest evidence for the localisation of the AD gene is in chromosome 19.

7

A method for pooling alleles from different genotyping experiments

Y.S. Aulchenko, A.M. Bertoli-Avella, C.M. van Duijn

Single tandem repeat (STR) polymorphisms are widely used in linkage and association studies. One of the drawbacks of using these markers is that genetic data coming from different experiments cannot be easily pooled together, because both allele length and binning distance may change. As large studies become more and more common, there is an increasing interest in pooling of the genetic data obtained in different experiments. Correct reconstruction of allelic correspondences between genotyping experiments is particularly crucial for association-oriented studies such as candidate gene studies and genome-wide association studies in isolated population.
Here, we suggest maximum-likelihood framework to find the best correspondence between alleles typed in different genotyping experiments. We also address the issue of goodness-of-fit and robustness. We performed a study simulating results obtained in a genome scan using 787 STR markers. The simulations show that the suggested method has good properties with the respect to error rate even if sizes of samples to be pooled is as low as 10 subjects (3% errors), though only 9% of alleles pass our tests. As sample sizes increase to 250 subjects the proportion of alleles pooled reaches 96% with error rate of $<0.1\%$.

8

Genotype-by-sex interaction in the etiology of diabetes: Support for sex-specific QTLS in HyperGEN participants

C.L. Avery(1), B.I. Freedman(2), A. Kraja(3), I.B. Borecki(3), D. Arnett(4), M.B. Miller(4), J.S. Pankow(4), C.E. Lewis(5), R.H. Myers(6), S.C. Hunt(7), K.E. North(1)
(1) Epi Dep, UNC-CH; (2) Int Med Dep, WFU; (3) Bios Div, Wash U Med; (4) Epi Div, U Minn; (5) Div Prev Med, UAB; (6) Neuro Dep, BU; (7) CVD Gen. Div, U Utah

Susceptibility to type 2 diabetes is determined by multiple genetic and environmental factors and likely an interaction between the two. There are established sex differences in both the prevalence of type 2 diabetes and risk factors, and we previously found strong sex effects in linkage for type 2 diabetes in this population. As there is limited research on the sex-specific etiology of type 2 diabetes, we assessed genotype-by-sex interaction using a liability threshold model in an attempt to localize sex-specific diabetes QTLs. Hypertensive siblings and their offspring and/or parents in HyperGEN were recruited at five centers. The diabetic phenotype was defined by the WHO criteria and adjusted for race-center, age, and age^2. In total, 567 diabetic persons were identified in 437 families. Variance component linkage analyses in the combined and sex-specific samples were performed (SOLAR 2.1.2) using race-specific marker allele frequencies estimated in MERLIN. We observed QTL-specific genotype-by-sex interaction on chromosome 17 at 31 cM, with females and males displaying LODs of 3.4 and 0.3 respectively ([P]=0.05 for interaction). These sex-specific signals were considerably higher than linkage signals combining sex groups, underscoring the importance of considering genotype-by-sex interactions in the etiology of type 2 diabetes.

9

Testing for Hardy-Weinberg equilibrium in samples with related individuals

Catherine Bourgain(1), Mark Abney(2), Carole Ober(2) and Mary Sara McPeek(2,3)
(1) INSERM U535, Villejuif, France; (2) Department of Human Genetics and (3) Statistics, The University of Chicago, USA.

The classical Hardy-Weinberg (HW) chi-squared has lately regain attention from human genetics both as an additional strategy to identify genotyping errors and as a tool to detect association. However, classical tests for HW cannot be used on samples with related individuals without dramatically inflating the type I error. In particular they may not be used in isolated populations where the individuals are related through multiple lines of descent. We propose a new Quasi-Likelihood score test for HW (QL-HW test) suitable for any sample with related individuals, including large inbred pedigrees, provided that their genealogy is known. Performed conditional on the pedigree structure, the QL-HW test may detect departure from HW that is not due to the genealogy. Because the computation of the QL-HW test may become intractable for very polymorphic loci in large inbred pedigrees, a simpler alternative, the generalized corrected chi-squared test for HW (GCC-HW test), is also proposed. The statistical properties of the QL-HW and GCC-HW tests are studied through simulations considering a sample of independent nuclear families, a sample of extended outbred genealogies and samples from the Hutterite population, a North American highly inbred isolate. In the nuclear families, the power increase due to the inclusion of all family members instead of only two independent parents, can be as high as 40%.

10

Linkage heterogeneity in 254 hereditary prostate cancer (HPC) families

M.D. Badzioch(1), J.L. Stanford(2), D.M. Friedrichsen(2), S. Kolb(2), E.A. Ostrander(2), M. Janer(3), L. Hood(3), G.P. Jarvik(1)

(1) University of Washington; (2) Fred Hutchinson Cancer Research Center; (3) Institute of Systems Biology, Seattle USA

The 254 Seattle/PROGRESS HPC families (mean cases/family=4.45) have previously undergone a 10 cM genome scan by 2-point lod and multipoint NPL analyses. Additional analyses were undertaken to consider likely locus heterogeneity (het). Model-based 3-point lod methods assessed het by the alpha-test. Using the sex specific Marshfield map, the most significant het was with D7S2212/D7S820 at chromosome (chr) 7q21 (lod < 0, hlod=1.59, p-het=.007) under a recessive model; the location of a proposed HPC locus (Friedrichsen, 2004). Under a dominant model, D11S2371/D11S2002, chr 11q13-14, gave hlod=1.58, p-het=.01. The differences (d) between pairs of affected brothers' ages or Gleason score (GS) were examined for multi-marker ibd allele sharing, with a model free approach (SAGE LODPAL). A > 60 cM span on chr 9q21-33 had age-d effects ($p < .01$), peaking at D9S930/D9S934 (lod=3.07, p=.0002). Chr 11q23-24 gave age-d lod=2.85, p=.0003. GS was known for 266 brother pairs in 94 families and showed -d effects on chrs, 2q37 (lod=2.11, p=.002), 8q21-23 (2.23, .002) and 18q21 (1.81, .004). Small d predicted linkage in most cases. Some markers yielding our results were associated with number of sibs affected (GAAT1A4/D8S1132) by Goddard, et al (2001) and with GS (D9S930/D9S934) by Witte, et al (2000). As studies suggest many HPC loci, methods robust to het and consistent linkage signals among datasets should be pursued. Small d may be a useful covariate in HPC linkage.

11

Genetic and maternal factors in a non-human primate model of obesity. Did your mom make you fat?

J.N. Bailey, M.J. Jorgensen, S.E. Breidenthal, L.A. Fairbanks

Obesity is an expanding worldwide problem, and it increases risk for diabetes, metabolic syndrome, and cardiovascular diseases. Obesity is thought to have both genetic and environmental risk factors. The Vervet Research Colony at UCLA offers a good non-human primate model of obesity and related disorders. This pedigree colony of African Green monkeys has approximately 500 living animals, 7 generations, and almost 30 years of clinical records, including records on yearly weight and diseases. Using body mass index (BMI) as a measure for adiposity, we examined the genetic and environmental contributions using variance component analysis. The analyses were performed using SOLAR. SOLAR is capable of analyzing the entire 1100+ member pedigree without breaking it into smaller pedigrees. We found a significant genetic component h2(+-)=.42 (0.18) p=0.0009. In addition, we tested for a maternal effect observed in the human literature and found a significant maternal component m2=0.18 (0.02) above the genetic component. This indicates that the reported maternal effect from the human literature may not be cultural, but have a biological basis.

12

A Multilocus Multimarker Regression Based Test of Linkage for Affected-Sib-Pair Families with Known Parental Genotypes

M.J. Barber(1), H.J. Cordell(1)

(1) JDRF/WT Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, UK

We address the analytical problem of evaluating the evidence for linkage at a test locus while taking into account the effect of a known disease locus that is linked to the test locus. This has been achieved by extending a single-locus multimarker regression approach that is very similar in concept to a generalized estimating equation approach proposed by Liang et al. (2001) (Hum Hered 51:64–78). In contrast to Liang (2001), which specifically addresses the issue of estimation of disease locus location with a confidence interval within a region of known linkage, we primarily utilize the multimarker regression approach as a test of linkage that can viewed as a multimarker extension of the single marker mean test of Blackwelder and Elston (1986) (Genet Epidemiol 2:85–97). The main advantage of a multimarker regression (as a test of linkage) over multipoint methods is that identity-by-descent (IBD) state uncertainty is not imputed at a test locus given the known parental data but, instead, a multimarker regression models all known marker IBD states simultaneously given the effect at test locus. (This is only possible when parental marker genotypes and thus marker IBD states are known). Simulations have been employed to explore the properties of the proposed Multilocus Multimarker Regression test statistic under both the null and various alternate hypotheses. The method is applied to real data from a Type 1 Diabetes affected-sib-pair study, testing for evidence of an additional disease locus on chromosome 6 while taking into account the strong effect at HLA.

13

Accurate Localization Information from the Two-Point Posterior Probability of Linkage

C.W. Bartlett(1), M.W. Logue(1,2), V.J. Vieland(1,2,3)

(1) Center for Statistical Genetics Research; (2) Program in Public Health Genetics, College of Public Health; (3) Department of Psychiatry, University of Iowa, USA

The posterior probability of linkage (PPL) [Vieland 1998] is a direct measure of the probability of a disease gene linked to a marker or location. The PPL differs in several respects from other linkage statistics; it is likelihood-based but also model-free in that trait parameters are not specified [a priori]. In both simulation studies and real applications we have observed that two-point PPL analyses produce narrow and pronounced peaks, raising the question of whether the location of the peak PPL is an accurate indicator of trait-gene position. We have investigated this by simulating a 20 cM map of polymorphic markers, with a disease gene in the center, and 1 cM intermarker distances. Results thus far indicate a high degree of accuracy in localization. For example, under a recessive

model, based on replicates of 100 affected-sib pair families, there was 58% probability that the peak PPL was within +/−1 cM of the disease gene; 76% probability of being +/−2 cM; 90% of being +/−3 cM, and 100% of being +/−5 cM. Decreasing the proportion of linked families to 50% gave probabilities of (respectively for the same conditions): 33%, 58%, 67% and 100%. Additionally, we have investigated the importance of the size of the peak PPL, both on its own and in relation to flanking PPLs, in predicting localization accuracy. Simulations under additional generating models and for varying sample sizes and pedigree structures will be shown, and compared with LOD scores.

14

MDR and PRP: A comparison of methods for high-order genotype-phenotype associations

L. Bastone(1), M. Reilly(2), D.J. Rader(2), A.S. Foulkes(1)
(1) Division of Biostatistics, University of Pennsylvania; (2) Cardiovascular Division, University of Pennsylvania

Complex diseases such as cardiovascular disease are likely due to the effects of high-order interactions among multiple genes and demographic factors. Therefore, in order to understand their underlying biological mechanisms, we need to consider simultaneously the effects of genotypes across multiple loci. Multifactor dimensionality reduction (MDR) and the combination of patterning and recursive partitioning (PRP) are methods designed to uncover complex relationships between genotype and phenotype without relying on a specific model for the interaction and are, therefore, well-suited to this data setting. We demonstrate that MDR is a special case of PRP in which (1) tree growth is restricted to a single split and (2) misclassification error is used as the measure of node impurity. We illustrate this finding by applying MDR and PRP to characterize the effects of 11 single nucleotide polymorphisms on coronary artery calcification in a nested case-control study of 600 people at risk for cardiovascular disease. As expected, the results of MDR and PRP are consistent for these data. Furthermore, the findings are suggestive of an interaction among markers in Angiotensinogen (AGT), Paraoxonase 2 (PON2), and Plasminogen Activator Inhibitor (PAI-1). The direct parallel between MDR and PRP implies that the extensive theory and literature surrounding recursive partitioning can be applied to either method.

15

Application of Mantel statistics for haplotype sharing analysis to schizophrenia and bipolar affective disorder

L. Beckmann(1), T.G. Schulze(2), C. Fischer(3), S. Ohlraun(2), J. Schumacher(4), S. Cichon(4), P. Propping(4), M.M. Noethen(4), J. Chang-Claude(1), M. Rietschel(2)
(1) German Cancer Research Center DKFZ, Heidelberg, Germany; (2) Div. of Genet Epi, Central Inst. of Mental Health, Mannheim, Germany; (3) Dept. of Hum Genet, Univ. of Heidelberg, Germany; (4) Inst. of Hum Genet, Univ. of Bonn, Germany

Recently, we reported an association between schizophrenia (SZ) and both [DAOA] (encoding D-amino acid oxidase activator) and [DAAO] (encoding D-amino acid oxidase) in German patients. For bipolar disorder (BD), we found a significant association with [DAOA], but only suggestive evidence for an association with [DAAO]. Here, we reanalyzed the data taking a new approach based on Mantel statistics that correlate phenotypic similarity and genetic similarity based on haplotype sharing (Beckmann et al. Genet Epidemiol 2003 25:238). We analyzed 5 SNPs in [DAOA] and 3 SNPs in [DAAO] in 599 patients (299 SZ, 300 BD) and 300 controls. Our results confirmed a significant association with a single SNP in [DAOA] and all three SNPs in [DAAO] for SZ, and the association between a SNP in [DAOA] and BD. The Mantel statistics yielded lower p-values than the previous analysis, which used the Armitage trend test and the standard chi-squared-test. We found a statistical interaction between [DAOA]-SNPs and genotypes in [DAAO] with BD, thus confirming functional molecular data, i.e. the protein-protein interaction between D-amino acid oxidase activator and D-amino acid oxidase. Our results suggest that Mantel statistics using haplotype sharing have more power than conventional methods.

16

A GEE approach for disease gene localization: Using IBD sharing proportions versus mean IBD

J.M. Biernacka(1,2), L. Sun(1,3), S.B. Bull(1,2)
(1) Dept. of Public Health Sciences, Univ of Toronto, Canada; (2) Samuel Lunenfeld Research Institute, Toronto, Canada; (3) Hospital for Sick Children, Toronto, Canada

Allele-sharing models for affected sib pairs (ASPs) can be based on identical-by-descent (IBD) sharing proportions, on mean IBD sharing, or on other parameters. Liang et al. (Hum Hered 51:64–78, 2001) introduced a generalized estimating equations (GEE) approach to estimate two parameters: the location of a trait gene and mean IBD sharing by ASPs at that locus. We recently extended this model to simultaneously localize two linked disease genes in a region (Biernacka et al., Am J Hum Genet 73(5S):193, 2003), and proposed test procedures to evaluate evidence for two versus one disease loci (Biernacka and Bull, Genet Epidemiol, 25(3):239, 2003). To compute empirical p-values of these test statistics, however, we need to specify the IBD proportions at the disease gene under the null one-locus model. Here we present a modification of the one-locus model of Liang et al. to estimate not only the two parameters estimated by their GEE procedure, but also the ASP IBD sharing proportions at a single disease gene. We studied the relative performance of the two methods by simulation and found that, in small samples, the procedure based on mean IBD sharing had better performance. In large samples, however, estimation of IBD sharing proportions at a disease gene yielded more efficient location estimates. Estimation of the location parameter in the one-locus model appears to be robust to the choice of allele-sharing parameters.

17

Influence of phenotype definition and covariates on linkage detection in a genome-wide screen for skin test response to allergens

E. Bouzigon(1), M.H. Dizier(2), J. Maccario(3), C. Krähenbühl(1), C Betard(4), MP Oryszczyn(3), M Lathrop(4), F Kauffmann(3), F Demenais(1)
(1) INSERM EMI 0006; (2) INSERM U535; (3) INSERM U472; (4) CNG, France

A previous genome scan, conducted in 295 French EGEA asthmatic families, led to detect three regions (5p15, 13q33 and 17q23) potentially linked to SPT, a binary phenotype defined as the positive skin test response to at least one of 11 allergens. A new score was constructed by summing the number of positive responses to each allergen (SPTQ). This score, which ranged from 0 to 10, was found to be significantly associated with sex. Since the distribution of SPTQ departed from normality, we considered an ordered categorical variable using either 6 classes (0, 1, 2, 3, 4, $>=5$) or 12 classes by splitting each category according to sex. Linkage was investigated using the Maximum Likelihood Binomial method extended to categorical traits. The most significant results for SPTQ were obtained in 3 regions: 3p22, 9q21 and 21q21 ($0.001 < p < 0.003$) which differed from those found for SPT. While lod score peaks were higher for two regions when sex was taken into account in the analysis: 9q21 (LOD=1.89 with sex vs 1.40 without sex) and 21q21 (LOD=1.90 vs 1.44), the contrary was observed for 3p22 (LOD=1.17 with sex and 1.69 without sex). These results show that considering different definitions of a phenotype (overall allergen response and levels of this response) and incorporating covariates in linkage analysis are of importance to infer the underlying genetic mechanisms.
Supported by ''Cohortes et Collections'' grant (4CH06G)

18

Effect of Residual familial Correlation, Ascertainment Bias and Missing Genotype Information on Penetrance Estimation: Application to a Study of Hereditary Nonpolyposis Colorectal Cancer in Large Newfoundland Kindreds with a Common MSH2 Mutation

L. Briollais(1,2), K.A. Kopciuk(3), W. He(1), E. Parkhomenko(2), J. Green(4), J.R. McLaughlin(1,2)
(1) Dept. of Epi & Biostat, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Canada; (2) Dept. of Public Health Sciences, University of Toronto, Canada; (3) Division of Population Health and Information, Alberta Cancer Board, Canada; (4)Clinical Epidemiology Unit, Memorial University of Newfoundland, Canada

Advances in the identification and treatment of genetically transmitted diseases have led to an increased need for reliable estimates of genetic susceptibility risk. These estimates are used in clinical settings to identify individuals at increased risk of being a disease allele carrier as well as to define the age-specific probability of developing a particular disease given one is a carrier (penetrance). Most current methodologies available to estimate penetrance of genes involved in complex diseases are based on simplifying assumptions such as the absence of residual familial correlation. In a random sample of families, such an assumption can lead to an underestimation of the variability of the penetrance estimate. In a selected sample, this could also lead to a biased estimate of the penetrance. We will illustrate the problem by describing an on-going penetrance study of a rare founder mutation in the MSH2 gene in a sample of HNPCC large Newfoundland families. Then we will quantify the magnitude of bias by simulations. Finally, we will discuss methods to correct for this bias in the context of family study.

19

Exogoneous hormones and risk of breast cancer in BRCA1/2 mutation carriers

R.M. Brohet(1), M.A. Rookus(1), N. Andrieu(2), J. Chang-Claude(3), A. Antoniou, D.F. Easton(6) and D.E. Goldgar(4) for the International BRCA1/2 Carrier Cohort Study (IBCCS)
(1) Netherlands Cancer Institute, Netherlands; (2) Institut Gustave Roussy, France; (3) German Cancer Research Center, Germany; (4) International Agency for Research on Cancer, France; (6) Strangeways Research Lab Cambridge, United Kingdom

The effect of oral contraceptives (OC) and hormone replacement therapy (HRT) on breast cancer risk in BRCA1/2 mutation carriers was examined in an international cohort of 1,601 women with proven BRCA1/2 mutations. To overcome possible survival bias, we restricted the analysis to recent breast cancer cases diagnosed within 5 years before interview. Since affected individuals are more likely to be included in the study, weighted cohort analyses were performed to mimic a true cohort of BRCA1/2 carriers. Women who ever used OC had a slightly increased relative risk (RR) of breast cancer (RR=1.57 95%CI; 1.11–2.21). In contrast with findings in the general population, the increased risk seemed to persist for women who stopped OC use more than 10 years ago (RR=1.69 95%CI; 1.16–2.47). A weaker association was found for current OC use (RR=1.51 95%CI; 0.95–2.39) and no duration-response relation emerged. Among postmenopausal women associated breast cancer risk for HRT was 1.57 (95%CI; 1.06–2.32) and risk seemed to be more profound for current HRT use (RR=2.05 95%CI; 1.03–4.09). In conclusion, use of OC and HRT are associated with a slightly increased relative risk of breast cancer among BRCA1/2 mutation carriers, which should be taken into account in counselling women at high risk.

20

Detecting Gene-Gene Interactions: MDR versus CART

W.S. Bush and M.D. Ritchie
Center for Human Genetics Research, Vanderbilt University Medical School, Nashville, TN

In the quest for disease susceptibility genes, the reality of gene-gene interactions creates difficult challenges for many current statistical approaches. In an attempt to

overcome limitations with current disease gene detection methods, we have previously developed the Multifactor Dimensionality Reduction (MDR) approach. MDR evaluates combinations of genes associated with clinical endpoints (such as cases and controls or drug treatment response and non-response). MDR has demonstrated high power for detecting gene-gene interactions. In the current study, we compare the power of MDR with classification and regression trees (CART). CART is a tree based algorithm that builds classification and regression trees for predicting discrete or continuous outcome variables from discrete or continuous independent variables. CART has been applied in genetic epidemiology in recent years with some success. We simulated data using a variety of epistasis models varying in allele frequency, heritability, and the number of interacting loci. All models represent the extreme case of interactions in the absence of main effects. We estimated power as the number of times that each method identified the correct functional SNPs for each model out of a set of 10 total SNPs. Using simulated data, we show that MDR has significantly higher power that CART for detecting two, three, and four-locus interaction models. This study provides evidence that MDR is a powerful statistical approach for detecting gene-gene interactions. In addition, MDR demonstrates higher power to detect disease susceptibility genes in comparison to more traditional statistical approaches such as CART. MDR will continue to emerge as a valuable tool in the study of the genetics of common, complex disease.

21

Estimating residual effects of unobserved gene loci with respect to a binary trait with applications to the ApoE locus and Alzheimer's disease

S. Böhringer(1), E.R. Martin(2), A. Steland(3)
(1) Institut für Humangenetik, Essen, Germany; (2) Center for Human Genetics, Duke University Medical Center, Durham, NC, United States; (3) Lehrstuhl für Stochastik, Ruhr-Universität Bochum, Bochum, Germany

In complex disorders fine mapping remains a major challange in regions of high linkage disequilibrium (LD). Methods to characterize LD in high risk haplotypes can help to define causative variants. We study a likelihood approach to estimate parameters characterizing disease predisposition in a given region, i.e. haplotype frequencies, penetrance. The model includes parameters to estimate the effect of unobserved loci which are estimated simulataneously. We illustrate the methods on a data set of Alzheimer's for which a causative locus has been demonstrated (Martin et al. 2000). Since the method is sensitive to model assumptions like a certain decomposition of the penetrance function, we perform sensitivity analyses to demonstrate the scope of the method.

22

Selection of a subset of SNPs for association studies: comparison of different strategies

E. Cousin(1), J.F. Deleuze(1) and E. Genin(2)
(1) Evry Genetics Center, Aventis Pharma, Evry, France; (2) INSERM U535, Villejuif, France

The high density of Single-Nucleotide Polymorphisms (SNPs) throughout the genome and the easiness of their genotyping made these markers a widely used tool for association studies in candidate genes. As a matter of fact, within a same gene, the number of identified SNPs may be very important. Typing them all may then result in a large consumption of DNA (which amount is often limited) and may involve multiple-testing problems. In this context, numerous methods have been proposed in order to choose the most appropriate SNP subset to be genotyped in a candidate gene and further tested for association. We compared four different methods of selection based on two distinct strategies: 1) the htSNP method of Johnson et al. (Nature Genetics 2001, 29: 233–237) and 2) the SNP selection method of Stram et al. (Human Heredity 2003, 55: 27–36) which are both based on haplotype information, and two other methods based on pairwise linkage disequilibrium, 3) one developed by Génin (Genetic Epidemiology 2001, 21 suppl 1: S614–S619), and 4) another one by Carlson et al. (American Journal of Human Genetics 2004, 74: 106–120). The four methods were applied on the genotype data of two candidate genes. The subsets proposed by the different methods were compared by estimating the power to detect an association with the studied gene under different genetic models. Our results suggest that method 3 is best at selecting the minimal number of SNPs within the gene while maintaining the power to detect association.

23

Testing for association using tag SNPs in the presence of strong dominance effects

J.M. Chapman, D.G. Clayton
JDRF/WT Diabetes and Inflammation Laboratory, University of Cambridge, UK

Usual tests of association using tag SNPs assume that the alleles of the causal locus act additively and that these alleles are then predicted indirectly via a set of tag SNPs. In the presence of strong dominance effects this model is not correct and an extra term needs to be included which uses the tag SNPs to predict the heterozygosity of the causal locus. Assuming this scenario of a strong dominance effect we present an appropriate test statistic and investigate how much power, if any, we gain by adding this single degree of freedom for dominance.

24

A Comparison of Family-Based Association Tests For Quantitative Traits

I. Chazaro(1,2), J. Dupuis (1), L. Atwood(1), R. D'Agostino(1,2), L.A Cupples(1)
(1) Department of Biostatistics, Boston University, USA; (2) Department of Mathematics and Statistics, Boston University, USA

Genomewide linkage studies have pinpointed regions of the human genome associated with variations in complex phenotypes. The identified segments are often large, making quantitative trait loci (QTL) mapping a challenging process. Family-based association methods have been proposed as a QTL fine mapping strategy to control for population admixture, a major concern in association studies of genes. This study evaluates three family-based association methods for quantitative traits in nuclear families: the measured genotype (MG) method (Boerwinkle et al., 1986. Ann Hum Genet. 50:181–194), the quantitative transmission/disequilibrium test (QTDT) (Abecasis et al., 2000. Am. J. Hum. Genet. 71:1330–1341), and the family-based association test (FBAT) (Laird et al., 2000. Genet. Epi. 19:S36–S42, Rabinowitz et al., 2000. Hum. Her. 50:211–223). These methods are compared, using simulated data, in terms of power and type I error assuming normally distributed traits, a diallelic QTL and a marker locus with different levels of linkage disequilibrium with the QTL. The tests are performed with and without the presence of population admixture. To study the sensitivity of the procedures to the normality assumption, the methods are also assessed in regard to type I error and power using simulated non-normally distributed data. This study also proposes theoretical power calculations for the MG assuming normally distributed traits. Power results from simulations are contrasted with theoretical results for the MG and FBAT (Lange C et al., 2002. Am J Hum Genet 71:1330–1341).

25

Using Biologic Knowledge to Inform Population-Level Inference in the Analysis of Candidate Genes in a Pathway

D.V. Conti, Z. Guan, R.M. Watanabe
Department of Preventive Medicine, University of Southern California, Los Angeles, CA, USA

Many chronic diseases are the result of a complex physiologic process involving various genes and environmental exposures. Often, we use our knowledge of the underlying biological mechanism to prioritize the selection of candidate genes for investigation. However, the analysis frequently ignores this biologic dependency and focuses on one variable at a time or pairwise combinations. Here, we aim to integrate biologic knowledge in our statistical analysis of the joint and interactive effects from gene polymorphisms. In concurrent work investigating Type 2 diabetes, we have developed a compartmental model based on the underlying physiology of glucoregulation. By developing a mathematical model in this manner, the parameters involved correspond to specific biologic processes, and thus, to actual candidate genes. We use this model to simulate glucose and insulin profiles from the oral glucose tolerance test on individuals in a population. This allows us to gauge the impact of genetic effects on population-level statistical analysis. We present a Bayes model-averaging framework that uses structured priors to influence the prior distribution of effect estimates and the space of possible models selected. We demonstrate the potential improvement in inference for effect estimates and model selection when these structured priors are based on knowledge of physiology combined with biologically motivated simulations.

26

Genome Scan for a Trichotomous Blood Pressure Phenotype in the NHLBI Family Heart Study

J. Corbett(1), I.B. Borecki(1), K. North(2), D.K. Arnett(3), M. Miller(3), S.C. Hunt(4), R.H. Myers(5), R.C. Ellison(5), M.A. Province(1)
(1) Washington University School of Medicine, St. Louis, MO; (2) University of North Carolina School of Public Health, Chapel Hill, NC; (3) University of Minnesota School of Public Health, Minneapolis, MN; (4) University of Utah School of Medicine, Salt Lake City, UT; (5) Boston University School of Medicine; Boston, MA

A trichotomous blood pressure phenotype was analyzed on 4,292 sib-pairs taken from a total of 2,882 individuals. The subjects were participants in the National Heart, Lung, and Blood Institute Family Heart Study. The hypertension phenotype was defined by three states: severely hypertensive (systolic blood pressure (SBP) >=160 mm Hg, diastolic blood pressure (DBP) >=90 90 mm Hg, or currently taking >=2 antihypertensive medications), moderately hypertensive (SBP >=140, or DBP >=80, or taking one antihypertensive medication), and normotensive (neither moderately nor severely hypertensive). A genome screen was performed using a variance components method implemented in the software package Mx. Two regions on chromosome 2 yielded suggestive LOD scores (3.30 at 63 cM and 2.83 at 144 cM.) A region at 28cM from the p-telomere on Chromosome 7 yielded a LOD score of 2.36. The region near the larger peak on chromosome 2 is in a region with several previously published linkage findings for blood pressure and hypertension phenotypes, suggesting this region likely harbors one or more genes conferring susceptibility to hypertension.

27

Latent stratification of Alzheimer risk genotypes and age at onset

E.H. Corder(1), A.J. Brookes(2), K. Blennow(3), M.A. Woodbury(1), M.G. Taylor(1), J. Prince(2)
(1) Center for Demographic Studies, Duke Univ., USA; (2) Center for Genomics Research, Karolinska Institute, Stockholm, Sweden; (3) Dept. of Clinical Neuroscience and Transfusion Medicine, Univ. of Goteborg, Sweden

A fuzzy latent class approach to the identification of genetic risk sets was employed to group candidate genes/genotypes for Alzheimer disease (AD). We investigated 508 AD cases and 343 cognitively intact control subjects ranging widely in age. Using grade-of-membership analysis (GoM) five susceptibility groups were identified from information on age, case status, MMSE score, APOE

genotype, and 44 additional gene-based polymorphic markers (38 candidate genes, 19 situated on chromosome 10q). The model-based cognitively intact group lacked risk genotypes and carried a number of protective genotypes (for APOE, SCD, LRP1, APBB1, BCHE, LIPA), about two-thirds the size of the control group. A core set of 17 genes ($H > 0.25$) distinguished the affected groups and replicated associations previously found for the dataset (APOE, ACE, IDE, TNFRSF6, AGER). There was extreme genetic heterogeneity for the younger affected groups (APOE44+ onset <age 65, APOE-unrelated onset <age 70). The two late-onset groups were less complex, one associated with APOE34 and the other having onset after age 75 carried MYST4 alleles not found elsewhere implicating altered transcription. Chromosome 10q may carry multiple risk genes. Classification of individuals was cripser for the simpler older age groups. We conclude that fuzzy latent class analysis (GoM) may be a useful tool for dissection of complex disorders such as AD.

28

Dissecting the Heterogeneity of Rheumatoid Arthritis Through Linkage Analysis of Quantitative Traits

L.A. Criswell(1), W.V. Chen(2), N.B. Ahmad(3), M.H. Wener(3), P.K. Gregersen(4), C.I. Amos(2)
(1) Dept. of Medicine, UCSF, USA; (2) Dept. of Epidem., Univ. of Texas MD Anderson Cancer Center, USA; (3) Dept. of Lab. Medicine, Univ. of Washington, USA; (4) North Shore-LIJ Research Institute, USA

Application of linkage analysis to quantitative components of qualitative traits such as rheumatoid arthritis (RA) may yield valuable information about clinical and genetic heterogeneity. We studied 991 RA patients from 512 sib pair families who were typed for 380 autosomal microsatellite markers. We performed a genome scan using 2 quantitative outcomes: anti-CCP and RF titers. The Merlin statistical package was used to analyze the data by regressing the identity by descent sharing in sib pairs upon quantitative traits, with a prespecified mean and variance estimated from 20 normal subjects (all with nondetectable levels). Results of linkage analysis are summarized below. Overall, there was stronger evidence for linkage to anti-CCP compared to RF titers and results for these traits were strikingly different. There was also little overlap with linkage results for RA defined as a qualitative trait. Analysis of quantitative components of RA may facilitate the identification of genes that contribute to this complex disorder and shed light on distinct underlying mechanisms.

Anti-CCP		RF	
chr (cM)	LOD (p)	chr (cM)	LOD (p)
6 (45)	6.1 (1.2E-07)	5 (23)	1.5 (0.004)
18 (65)	1.6 (0.003)	8 (125)	1.4 (0.005)
3 (20)	1.6 (0.003)	6 (119)	1.1 (0.012)
6 (119)	1.3 (0.007)	1 (233)	1.1 (0.012)
8 (0.7)	1.3 (0.008)		
17 (7)	1.1 (0.011)		

29

Smoking Interacts With Genetic Risk Factors in the Development of RA

L.A. Criswell(1), K.G. Saag(2), T.R. Mikuls(3), J.R. Cerhan(4), L.A. Merlino(5), R.F. Lum(1), K.A. Pfeiffer(1), B. Woehl(6), M.F. Seldin(6)
(1) Dept. Med, UCSF, USA; (2) Dept. Med, U. Alabama, USA; (3) Dept. Med, U. Nebraska, USA; (4) Dept. Epi, Mayo Clinic, USA; (5) Dept. Epi, U. Iowa, USA; (6) Dept. Hum Genet, UC Davis, USA

Exposure to tobacco smoke is an important risk factor for rheumatoid arthritis (RA). We sought to determine whether the impact of this exposure is influenced by the HLA-DRB1 and glutathione S-transferase M1 (GSTM1) loci. Subjects were participants in a prospective, population-based study of elderly women. Incident RA cases were identified and medical records reviewed to confirm RA diagnosis. Cases were matched by age and ethnicity to controls from the same prospective cohort. Conditional logistic regression (CLR) was used to estimate the impact of smoking, HLA-DRB1 [presence of shared epitope (SE)], GSTM1 (homozygosity for null allele) and interactions between these factors on RA risk. 228 individuals were studied, including 76 RA cases and 152 controls. Results of CLR demonstrated striking associations of smoking and both genetic factors with risk of RA, as well as significant gene-environment interaction. Odds ratios (OR) describing the risk of RA for different combinations of exposures ranged from 3 to 16 (reference=SE negative non-smokers with 0 or 1 copies of the GSTM1 null allele). Failure to include either of these genetic factors led to underestimation of the impact of smoking on RA risk (OR=1.5, p=0.2). These results emphasize the importance of considering both genetic and environmental factors, plus their interaction, in studies of complex diseases like RA.

30

Identification of SNP haplotypes in CDKN2A

G.P. Crockford, J.A. Newton Bishop, J.A. Randerson-Moor, M. Harland and D.T. Bishop
Cancer Research UK Clinical Centre in Leeds

Aim: To determine common SNP haplotypes in CDKN2A and the additional variation in haplotypes explained by rare SNPs Mutations in the CDKN2A gene are linked with the onset of melanoma in multiple case melanoma families. We screened 100 samples from multiple case melanoma families and identified 33 SNPs located in the promoter region, intron 1, intron 2 and the 3′ region of the gene; there is only one common exonic SNP. Assays were unavailable for 4 SNPS. Population samples (190 mother-child pairs and 444 twin-parent triplets) and 93 melanoma cases with a family history of melanoma were genotyped for the 29 SNPs. Haplotype analyses was performed having determined a maximally informative set of unrelated individuals using EHPLUS SNPHAP. In the population series there were 7 'common' SNPs (variant allele>=5%), 3 'less common' (>=2%), and 19 rare, of which 4 had no variant allele. The 7 common SNPs

identified 6 haplotypes with frequencies >1%, (36%, 26%, 12%, 9%, 9%, 4%) which account for 96% of the variation in all haplotypes. Sequential inclusion of rare SNPS leads primarily to additional rare haplotypes as well as explaining up to 2.5% of the variation in the common haplotypes when including the 'less common' SNPs. Using the same SNPs in the case series we identified 8 common haplotypes with frequencies >1% (28%, 28%, 14%, 6%, 6%, 4%, 4%, 2%) which account for 92% of the variation. We compare these results with those of Crawford et al (AJHG 74:610–622, 2004) showing a comparable number of haplotypes. These haplotypes are being investigated in a case-control study of melanoma and in a QTL analysis of naevi.

31

Sex specific linkage analysis with Genehunter

J. Dietter(1), T.F. Wienker(1), K. Strauch(1)
(1) Institute for Medical Biometry, Informatics, and Epidemiology, Bonn University, Germany

Most computer programs currently available for both parametric and nonparametric multipoint linkage analysis only allow for the use of sex-averaged marker distances. However, since reliable sex-specific marker maps have become increasingly available during the past years, one should proceed to take advantage of them. We have implemented the usage of sex specific recombination frequencies in the Genehunter-Modscore program previously developed by our group. This will put researchers in the position to use all available marker information. Consequently, this will lead to an increased power to detect linkage, which is especially important for complex traits. At the same time, researchers will have access to the accustomed functionality of the Genehunter program as well as to the extended possibilities of Genehunter-Modscore and Genehunter-twolocus.

32

Detection of linkage of asthma associated with allergic rhinitis to the chromosome 1p31 in the French EGEA study

M.H. Dizier(1), E. Bouzigon(2), M. Guilloud-Bataille(1), E. Genin(1), M.P. Oryszczyn(3), I. Annesi-Maesano(3), F. Demenais(2)
(1) INSERM U535, France; (2) EMI 0006, France; (3) INSERM U472, France

A recent genome scan for two allergic co-morbidities, asthma and allergic rhinitis (AR), conducted in French EGEA families ascertained through asthmatic probands, led to detect linkage of 1p31 to asthma associated with AR (p=0.0002) and, in a lesser extent, to asthma (p=0.005) or AR (p=0.005) when considered separately. Our purpose was to search for heterogeneity of linkage to the 1p region according to the affection status being defined as either presence of these two traits (asthma plus AR) or presence of only one trait (asthma or AR). A strong heterogeneity (p=0.0001) was detected by the Predivided Sample Test (PST) which compared the IBD distributions in 51 sib-pairs with each sib having both asthma and AR and in 48 sib-pairs with each sib having either asthma alone or AR alone. Indeed, the IBD distribution in the latter sib-pairs was very close to the one expected under no linkage (0.25,0.5,0.25). This was also true for the asthma-only sib-pairs and for AR-only sib-pairs. All these results indicate that the 1p31 region is only linked to the phenotype defined by asthma plus AR. Since it has been proposed that asthma and AR may correspond to the same disease and asthma associated to AR to the severe form of this disease, our results suggest that a genetic factor on 1p might be involved in this severe form.

33

Statistical Properties of the Propensity Score as a Single Covariate in Covariate-Based Linkage Analysis

B.Q. Doan(1,3), C.E. Frangakis(2), A.J.M. Sorant(3), J.E. Bailey-Wilson(3), Y Yao(1)
(1) Dept.of Epi & (2) Biostats, JHSPH; (3) IDRB/NHGRI/NIH

To increase the power to detect linkage while minimizing the degrees of freedom, we previously applied the propensity score(PS) to covariate-based linkage analysis (LODPAL). The PS is the logistic regression of the affection status on covariate data, incorporating multiple covariates into 1 variable. Through simulations of genetic models with 2 underlying covariate effects, we showed that in 68% of the models, the PS provided the highest power when compared to the inclusion of 0, 1 or 2 covariates; in the remaining 32%, the power was extremely low (<0.10) in all situations. Type I error rates increased approximately 20% with each increasing covariate analyzed. We examined the independent effects of increasing numbers of underlying covariates and increasing sample sizes. With each additional covariate analyzed in the multiple covariate models, results suggest that the type I error rates continued to increase approximately 20%, with an average inflated type I error rate of 0.129(0.05 nominal level) for 5 covariates analyzed. For 2-covariate models, increasing the sample size to 1000 families was sufficient to eliminate the inflation of the type I error rates, and with the additional power, the superior performance of the propensity score was observed.We are exploring the effect of sample size and statistical properties of PS by verifying that the distribution of the covariates are similar between affecteds and unaffecteds within strata defined by PS values (balancing property)and by comparing different definitions of pair-specific PS covariate values.

34

More efficient permutation corrections for multiple correlated tests

F. Dudbridge(1), B.P.C. Koeleman (2)
(1) Medical Research Council, Cambridge, UK; (2) Dept Medical Genetics, University Medical Centre Utrecht, Utrecht, NL

Although asymptotic distributions are usually available for single tests, one must often resort to a permutation

procedure to correct for multiple testing across correlated tests. These procedures estimate a binomial probability from permutation samples, but do not use all the information from those samples. When single P-values are uniform on the unit interval, the permutation distribution often follows an analytic form. We propose to fit the parameters of the analytic form to the permutation replicates, and calculate the significance level from the fitted distribution rather than by a binomial probability. We show that this approach can achieve a given accuracy using several times fewer replicates than a standard permutation test. In certain situations the fitted distribution can be reused in subsequent tests of the same hypotheses. We give examples for association tests of HapMap SNPs, using beta distributions for Bonferroni corrections, gamma distributions for Fisher products, and extreme value distributions for truncated Fisher products.

35

Comparison of Multilocus Genotypes and Inferred Haplotypes of Single-Nucleotide Polymorphism (SNP) Markers for Linkage Disequilibrium Mapping via Cladistic Analysis

C.V. Durrant, L.R. Cardon, A.P. Morris
Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

We recently described a method for linkage disequilibrium (LD) mapping, using cladistic analysis of SNP haplotypes in a logistic regression framework (1). However haplotype data is often not available, and this approach does not account for uncertainty in the phase assignment for inferred haplotypes. Here, we extend the method to accommodate inferred haplotypes, accounting for uncertainty in phase assignment using posterior probabilities from PHASE (2). An alternative approach is to analyse the phase-unknown multilocus genotypes themselves, so here we also present an analogous method to that described in Durrant et al. (1) but for multilocus SNP genotypes instead of haplotypes. We analysed a sample of 41 cases and 977 controls for a poor drug metaboliser phenotype over unphased genotypes of 32 SNP markers in a 0.89 Mb region of chromosome 22 containing the gene CYP2D6. We compared the overall significance of the region and the results of the sliding window analysis across the region for three methods: (a) the unphased multilocus genotypes, (b) the inferred haplotypes accounting for uncertainty in phase assignment using posterior probabilities from PHASE (2) and (c) the haplotype analysis method described in Durrant et al. (1) using the best-guess phase assignment for each individual. (1) Am J Hum Genet 75 (2004), pp 35–43; (2) Am J Hum Genet 68 (2001), pp 978–989.

36

A novel statistic for analysis of association between disease and a multi-allelic marker

R. el Galta(1), T. Stijnen(2), J.J. Houwing-Duistermaat(1)
(1) Department of Medical Statistics and Bioinformatics, LUMC, Leiden, The Netherlands; (2) Department of Epidemiology and Biostatistics, Erasmus MC, Rotterdam, The Netherlands

Terwilliger [Am. J. Hum. Genet. 56: 777–787, 1995] proposed a powerful model which assumes that only one marker allele is associated with the disease. To test for no association the likelihood ratio statistic (LR) was used. However this statistic appeared to be conservative and estimation of parameters might be slow. We derived the corresponding score statistic which is locally most powerful. Under the null hypothesis the distribution of the score test can be approximated by the standard normal distribution.
To study the type I error of the score test and to compare its performance to Pearson's chi-square χ^2 and LR we carried out a simulation study. The score test appeared to be slightly anticonservative and to have similar power as LR and more power than the χ^2 test. When data were generated under the model that assumes two associated alleles the score test appeared to perform better than LR and χ^2. For small p-values we suggest to use Monte-Carlo permutations, which is time efficient, compared with LR. As illustration we applied the score test to a published case control study on association between COL2A1 gene and radiographic osteoathritis [Ann. Hum. Genet. 63:393–400, 1999]. Both score and χ^2 tests gave p-value of 0.01 whereas LR gave a p-value of 0.03. We concluded that the score test is a powerful tool for detecting disease-marker association.

37

Pedigree selection for QTL linkage analysis under heterogeneity

C.T. Ekstrøm
Statistics, Dept. of Natural Sciences, Royal Veterinary and Agricultural University, Denmark

Pedigrees for linkage studies are often sampled from pedigrees known to segregate a disease. Often, the same pedigrees are also used for linkage analysis of quantitative traits believed to related to the disease etiology. When the disease definition is broad or when the disease etiology is unknown the disease may comprise several sub-diseases (and hence different variants) and locus heterogeneity may be introduced if analysing all pedigrees.
We consider an ordered subset like approach to QTL linkage analysis in the presence of locus heterogeneity where same-structured pedigrees are included in the analysis based on their log-likelihood contribution from a traditional variance component model. The ordered subset like approach is examined and compared through a simulation study to the traditional variance component linkage analysis as well as to a classification maximisation approach, where each pedigrees is included in the analysis based on the posterior probability of segregating disease alleles at a particular locus.

38

Estimating the location of susceptibility genes by incorporating Parent-of-Origin effects into affected sib pair analyses

M.D. Fallin(1), W. Chen(2), V.K. Lasseter(3), P.S. Wolyniec(3), J.A. McGrath(3), D. Valle(4), K-Y. Liang(2), A. E. Pulver(3)
(1) Dept. of Epidemiology, Johns Hopkins Bloomberg School of Public Health (JHBSPH), USA; (2) Dept. of Biostats., JHBSPH; (3) Dept. of Psych. & Behav. Sci, Johns Hopkins School of Medicine(JHSOM); (4) Depts. of Pediatrics and Molec. Biol., JHSOM, & Howard Hughes Medical Institute.

Parent-of-origin (P-O-O) effects may be an important mechanism for several complex diseases and have been noted in both association and linkage findings for several disorders including Alzheimer's disease and bipolar disorder. In such cases, detrimental alleles have a much higher penetrance when inherited from a particular parent. Analyses that do not accommodate this may not have the power to detect a linkage signal or the precision to estimate its location, since high and low penetrant alleles would both be contained in the contributing data. Methods to incorporate parent-of-origin effects in allele-sharing linkage approaches are emerging. We present a new method for taking P-O-O into account when estimating the location of a susceptibility gene as a modification of the GENEFINDER software (Liang et al. 2001) that now considers paternal and maternal chromosomes separately. We apply the method to data from a genome scan for bipolar disorder, which originally suggested four potential linkage regions using NPL statistics (chromosomes 1, 3, 11, and 18). Our GENEFINDER P-O-O analyses suggest the c18 evidence is limited to paternally derived alleles, and provides a different location estimate and narrower 95% CI for that location.

39

Confirmation of Prostate Cancer Predispositon Locus HPCX in Large Utah Pedigrees

J.M. Farnham(1), N.C. Camp(1), J. Swenson(2), S.V. Tavtigian(2), L.A. Cannon-Albright(1)
(1) Genetic Epidemiology, Department of Medical Informatics, University of Utah, Salt Lake City, Utah 84108 USA, (2) Myriad Genetics, Inc., Salt Lake City, Utah 84108 USA

Several prostate cancer genetic predisposition loci have been identified through linkage analysis, and it is now generally recognized that no single gene is responsible for more than a small proportion of prostate cancer. However, published confirmations of these loci have been few and failures to confirm have been frequent. It is clear that the genetic etiology of prostate cancer is complex, including significant genetic heterogeneity, phenocopies and reduced penetrance. Powerful analyses that include both robust statistics and methods to reduce genetic heterogeneity are necessary. We employed a robust multipoint statistic (TLOD), and a novel splitting algorithm to tackle intra-familial heterogeneity by iteratively removing the top generation from the large Utah pedigrees. In a dataset containing pedigrees having no more than five generations, we observed a multipoint TLOD of 2.74 ($p=0.0002$), which is statistically significant after correction for multiple testing. For both the full-structure pedigrees (up to seven generations) and the smaller sub-pedigrees, the linkage evidence was much reduced. In conclusion, this study represents the first significant confirmation of HPCX (Xq27-28) using large Utah pedigrees and argues for the continued utility of large pedigrees in linkage analyses for complex diseases.

40

Pleiotropic QTL on 19p13 for triglycerides and adiposity: The HERITAGE Family Study

M.F. Feitosa(1), T. Rice(1), K.E. North(2), A.T. Kraja(1), T. Rankinen(3), A.S. Leon(4), J.S. Skinner(5), J.H. Wilmore(6), J. Blangero(7), C. Bouchard(3), D.C. Rao(1)
(1) Washington Univ; (2) Univ North Carolina; (3) Pennington Biomedical Res Center; (4) Univ Minnesota; (5) Indiana Univ; (6) Texas Univ; (7) Southwest Foundation Biomedical Res, USA

Motivated by strong correlations between triglycerides (TG) and adiposity traits, we conducted bivariate linkage analysis of TG levels and body mass index (BMI), total fat mass (FAT), percentage of body fat (FATPC), and abdominal subcutaneous fat (ASF). The bivariate linkage approach substantially improved the power to detect linkage on these correlated traits at baseline in White families on chromosome 19q13. Strong evidence of a pleiotropic QTL was found for TG and BMI (LOD=3.3, 48cM), coinciding with the chromosomal location of the LIPE gene, a key enzyme in the mobilization of fatty acids from triglyceride stores in adipocytes. Although, the univariate linkage analyses for TG and obesity traits results were non-significant in this region, there was a significant genotype-by-sex interaction for BMI with linkage detected only in females ($p<0.0001$). A similar genotype-by-sex interaction was noted for TG and for bivariate (TG and BMI), although the magnitude of the effect was not significant. A second linkage peak was observed among TG levels and FAT, AVF and FATPC (LODs=3.0, 2.4 and 2.2, 51cM) near the APOE gene which is involved in lipid-lipoprotein metabolism. Therefore, there is strong evidence that one or more QTLs influence TG levels and obesity traits on 19q13 at or near the LIPE and APOE genes.

41

Modelling germline mosaicism and variable mutation rates in Duchenne muscular dystrophy

Fischer, Christine(1), Krüger Jens(1), Grimm Tiemo(2)
(1) Institute of Human Genetics, University of Heidelberg, Germany; (2) Institute of Human Genetics, University of Würzburg, Germany

For Duchenne muscular dystrophy two biological phenomena have been recognised as important: 1. germline mosaicism and 2. different new mutation rates in male and female depending on mutation type. Both principles have

been investigated separately and their influence on risk estimation in families has been exemplified in the literature.
The aim of this paper is to present a general model which allows to include germline mosaicism an heterogenous mutation rates. Mosaicism is introduced by defining additional alleles at the disease locus in combination with segregation rules. We derived the conditions which have to be fulfilled for a population in mutation selection equilibrium. As a prerequisite for the use of the presented framework for practical applications in genetic counselling model parameters had to be estimated. We use published empirical data and some simplifying assumptions to find admissable solutions for the parameters. Our approach aims to describe the model on the population level and not in individual subjects. This has the great advantage of resulting in tractable algebra. It allows the use of well known algorithms for calculation of likelihoods in pedigrees. Based on this model RISCALW a user-friendly Windows program for risk calculation in families with DMD was developed. It can be used to explore the dependence of results on different model assumptions for example different mutation rates in male and female for point mutations and deletions.

42

Tests of Disease-Haplotype Association when Linkage Phase is Ambiguous, Appropriate for Cohort and Matched Case-Control Studies

WD Flanders(1), MJ Khoury(2), QH Yang(3), H Austin(1)
(1) Department of Epidemiol, Emory Univ, Atlanta, GA; (2) Office of Genomics and Disease Prevention, CDC, Atlanta, GA; (3) National Center on Birth Defects and Developmental Disabilities, CDC Atlanta, GA

The effect of competing risks on tests of disease-haplotype association has been largely ignored for situations in which phase is ambiguous. We show that tests for disease-haplotype association can lead to rejection of the null hypothesis with more than the nominal 5% frequency, even when true, if competing risks exist. This problem tends to occur if a haplotype is associated with overall mortality, even if the haplotype is not associated with disease risk- that is when there are competing risks. In a simulation study, we illustrate the magnitude of this bias (artificially high type I error rate) for cohort studies with a modest number of expected incident cases (about 350). This bias occurs even if the score test is based on a logistic model that includes age as a covariate. The usual survival methods that address competing risks do not apply if linkage phase is ambiguous. Therefore, for cohort studies, we describe a new test based on a modification of the proportional hazards model and for case-control studies, a new test based on a conditional likelihood. In simulation studies, we consider three single nucleotide polymorphisms, assessed in unrelated individuals so that phase is ambiguous. The simulation results illustrate that these proposed tests have the correct size under the null even if there are competing risks.

43

Weighting Haseman-Elston-Regression - Weighting I.B.D. Information

D. Franke(1), R.C. Elston(2), A. Ziegler(1)
(1) Institute of Medical Biometry and Statistics, University Hospital Schleswig-Holstein - Campus Lübeck, Germany; (2) Department of Epi & Biostat, Case Western Reserve Univ., USA

Model free methods to detect linkage are based on the idea that genotypic similarity within sibpairs should result in phenotypic similarity as well (and vice versa). The classical Haseman-Elston (HE) algorithm regresses the squared phenotypic difference on the proportion of alleles identical by descent (i.b.d.). Due to technical reasons, the proportion of alleles i.b.d. can not be known exactly. In the analysis, however, each sibpair is assigned the same weight - regardless of its informativity! We therefore investigated whether a Weighted Least Squares approach is more powerful that defines informativity of i.b.d. values in one of the two ways: as Eucledian distance within the metric space of i.b.d.-distributions or by relative and absolute entropy measurements.
We developed SIBSIM to simulate quantitative traits data in multi-point setups with families of any size (GPL, available at www.imbs.uni-luebeck.de). Especially, we created 75.000 replicates of 300 nuclear families, each with a 13-loci setup and either (co-)dominant or recessive QTL-inheritance. Development versions of sibpal (S.A.G.E.) were used to compute test-statistics and p-values (computational results confirmed independently). Distribution graphs show that, by means of the integral transform theorem, the unweighted HE fits the assumptions of the underlying linear model best - any non-uniform weighting applied invalidates the results.

44

Linkage And Association On Chromosome 4 With Externalizing Symptoms In Alcoholism: Semi-Parametric Regression Approaches

S. Ghosh(1), P.P. Majumder(1), H. Begleiter(2), B. Porjesz(2), H.J. Edenberg(3), T. Foroud(3), A. Goate(4), L.J. Bierut(4), J.P. Rice(4)
(1) ISI, Kolkata; (2) SUNY at Brooklyn; (3) Indiana Univ, Indianapolis; (4) Wash Univ, St. Louis

The Collaborative Study On the Genetics Of Alcoholism (COGA) is a multicenter research program established to detect and map susceptibility genes for alcohol dependence and related phenotypes. Alcohol dependence has well-known externalizing symptoms. An endophenotype defined as the number of externalizing symptoms (ranging from 0 to 24) pertaining to anti-social behavior has been analyzed in this study. The distribution of the phenotype is very skewed and standard methods of linkage analyses that assume normality are not optimal. We have developed a linkage method based on non-parametric regression via kernel smoothing using squared contrast and mean functions of trait values within a sibship and quadratic functions of the sib-pair marker identity-by-descent scores.

A genome-wide scan using the COGA data on 171 sibships has provided significant evidence of linkage on 4q22.3, around 120 cM, a region harboring the alcohol dehydrogenase gene cluster (ADH1–ADH7). Since a biallelic marker in the ADH3 gene was genotyped, we next performed a transmission disequilibrium test for the quantitative trait using a logistic link function, which is a natural statistical extension of the classical TDT. We found that a marker allele was significantly preferentially transmitted to offspring having high values of the endophenotype, indicating a strong allelic association between the ADH3 marker and the putative locus controlling externalizing symptoms.

45

Heterogeneity analysis of diffuse large B-cell lymphoma based on coupled two-way clustering

B.-Sh. Gong, S.-Q. Rao, S.-L. Lv, L. Li, W. Jiang, L. Du, X. Li

It becomes increasingly clear that our current taxonomy of clinical phenotypes is mixed with molecular heterogeneity. Of vital importance for refined clinical practice and improved intervention strategies is to define the hidden molecular distinct diseases using modern large-scale genomic approaches. We have thus developed a novel heterogeneity analysis approach to extracting useful gene signatures and to discovering disease subtypes simultaneously. The basic strategy of the proposed method is to iteratively partition in two ways sample and feature space with super-paramagnetic clustering technique and to seek for hard and robust gene clusters that lead to the best fits of sample partitions and that have the highest conceptual consensus evaluated with Gene Ontology. We have applied the proposed method to a publicly available microarray data set of diffuse large B-cell lymphoma with 4026 genes and 42 samples. A feature subset of 21 genes, most of which are associated with immunoglobulin complex, identifies two categories of patients with very different five-year overall survival rates (55% and 25%).

46

Angiotensin Converting Enzyme Gene Insertion/Deletion Polymorphism and Breast Cancer Risk

A.M. González-Zuloeta Ladd(1), A. Arias Vásquez(1), F. Sayed-Tabatabaei(1), O. Njajou(1), C.M. van Duijn(1)
(1) Genetic Epidemiology Unit, Epidemiology and Biostatistics Department, Erasmus MC, Rotterdam, The Netherlands.

Studies have shown that the Insertion/Deletion polymorphism in the ACE gene accounts for the variability of ACE plasma concentrations. Angiotensin II was shown to act as a growth and angiogenic factor in the breast tissue. Several studies have tried to link both the use of ACE inhibitors and the ACE I/D polymorphism to breast cancer risk. Results from these studies have been inconsistent. In this study we evaluate the relationship of the ACE I/D polymorphism with breast cancer risk in caucasian postmenopausal women.

In our study, the ACE I/D polymorphism was genotyped in 4117 women belonging to the Rotterdam Study. Baseline information on the participants was obtained through a computerized questionnaire. We conducted a logistic regression (LR) and survival analysis to assess the risk of breast cancer by ACE genotype. When taking on account the incident cases only and adjusting for age at entry and age at menopause, the DD carriers showed a significantly increased risk of developing breast cancer when compared to the II carriers (OR=1.86, 95% CI=1.06-3.27, p-value=0.03). This association remained after adjusting for other proposed effect modifying variables. Our survival analysis showed that the cancer free survival was significantly reduced in DD compared to II carriers (OR=1.80; 95% CI: 1.07–3.01, p-value=0.026).
Our results suggest that the ACE I/D polymorphism plays an important role in breast cancer risk and disease free survival in caucasian postmenopausal women.

47

No Association of N-Acetyltransferase Haplotypes with Colorectal Polyp Risk

E.L. Goode, J.D. Potter, J. Bigler
Fred Hutchinson Cancer Research Center, Seattle WA, USA

N-acetyltransferases NAT1 and NAT2 have long been considered candidate genes for colorectal polyps because of their role in carcinogen metabolism. Previous studies of NAT genotypes and phenotypes (often imputed) have failed to find consistent associations with risk or interactions with known carcinogenic exposures; however, heterogeneous polyp groups are often combined. In a clinic-based study of 405 cases with adenomatous polyps, 198 cases with hyperplastic polyps, 122 cases with both types of polyps, and 635 polyp-free controls, we examined six NAT1 polymorphisms (190C>T, 559C>T, 560G>A, del9_1065, −88T>A, and −81C>A) and five NAT2 polymorphisms (191G>A, 341T>C, 590G>A, 803A>G, and 857G>A). Polytomous logistic regression assessed risk for each polyp group with each genotype and imputed phenotypes; no statistically significant associations were found, though interactions with current smoking status were suggested. We also performed haplotype-based logistic regression as implemented in HPlus using the EM algorithm for haplotype estimation and EE methods for standard error estimation. Haplotype-specific risks were estimated for five common NAT1 haplotypes and six common NAT2 haplotypes. An elevated risk of the two-polyp-type phenotype was suggested for NAT1 haplotype 001011 (*14A) with an adjusted OR of 2.1 (95% CI: 0.9–7.7) compared to the 000000 haplotype (*4); however, no statistically significant associations were observed. Although this study may be underpowered, we demonstrate an application of newer methods and find additional evidence against a major role of NATs in colorectal carcinogenesis.

48

Imprinting Detection by Extending a Regression-Based QTL Analysis Method: an Extensive Simulation Study

OY Gorlova(1), L Lei(1), RA Price(2), S Shete(1), CI Amos(1)
(1) The University of Texas M. D. Anderson Cancer Center, Houston, TX; (2) The University of Pennsylvania, Philadelphia, PA

We extended Sham et al.'s regression-based quantitative-trait linkage analysis method to incorporate parent-of-origin effects. We separately regress total, paternal, and maternal IBD sharing on traits' squared sums and differences. We also provide a test for imprinting that indicates whether there is any difference between paternal and maternal regression. We suggest using a panel of statistics to detect imprinting, which includes an overall T statistics (a test for total linkage), both parental T statistics (tests for parental-specific linkages), and the D index (a test for imprinting). When using empirical percentiles the method is very powerful in detecting parent-specific linkage with correct type I error rate for the non-linked parental component, although the test of imprinting is conservative. The method is not sensitive to misspecification of heritability while misspecification of the mean of trait decreases the power dramatically. Missing parental genotypes increase Type I error of both linkage and imprinting test and decrease the power of imprinting test. When the major gene has low heritability, the power of the method decreases. We applied the method to a real data set on 6 body size related measures and 23 loci on chromosome 7 for 255 nuclear families. Maternal imprinting was suggested at 102.31–108.702 cM for 4 out of 6 traits.

49

Use of compartmental modeling to assess genetic effects on regulation of glucose metabolism

Z. Guan, D.V. Conti, R.M. Watanabe
Dept. of Preventive Medicine, University of Southern California, USA

To elucidate the genetic role in complex diseases, such as type 2 diabetes mellitus (T2DM), molecular epidemiologic studies often select candidate genes within a biologic pathway. In this framework, the analysis often evaluates each gene as independent with very little knowledge of how functional changes in the underlying mechanism may impact our ability to find and characterize genetic effects with population-level data. In response, we have developed a physiologically-based compartmental model to simulate the oral glucose tolerance test. The model incorporates biologic knowledge of the glucoregulatory system giving each parameter biologic meaning. We examine how population variation in a single candidate gene impacts glucose concentrations by investigating plausible ranges of parameter difference between genotypes. For example, we simulate genetic effects of peroxisome proliferators-activated receptor-γ gene (PPARG, an accepted T2DM susceptibility gene) on glucose metabolism via a parameter representing transport of insulin into the interstitial compartment. By examining the sensitivity of gene effect estimates, we demonstrate that a small biologic difference in genotypes may be reflected in a sizeable impact in population level risk. In the same way, we investigate the impact of other candidate genes on the pattern of effect estimates. Furthermore, we impose multiple genetic effects in different parts of the model to explore possible gene-gene interactions. We present the model and results over several simulations. We discuss how biologic differences and joint action among polymorphisms may manifest on the population-level. The study highlights the promising advantages of compartmental modeling to dissect the genetic architecture underlying regulation of glucose tolerance.

50

Informative transmission disequilibrium test (ITDT): a test of linkage and association using unaffected Siblings

C. Guo(1), K.L. Lunetta(2), A.L. DeStefano(1), L.A. Cupples(1)
(1) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; (2) Oscient Pharmaceuticals, Waltham, MA, USA

The transmission/disequilibrium test (TDT) was introduced to test linkage disequilibrium between a genetic marker and disease locus, and many extensions and modifications of the TDT have been developed. When the family data contains more than one affected child, the TDT is valid for testing linkage between a marker and a disease locus, but not association. The TDTsp statistic proposed by Martin et al (1997) extended the TDT to allow for multiple affected offspring while remaining a valid test of linkage and linkage disequilibrium. A novel family based test using both affected and unaffected siblings was proposed by Lunetta et al. (2000). Because transmissions from heterozygous parents to both affected and unaffected children are included, this approach is valid for testing linkage but not association. Under some conditions, affected and unaffected siblings can significantly increase the power of a test for linkage and association. Here we propose a new test, the informative transmission disequilibrium test (ITDT), which uses information from all of the affected and unaffected children in nuclear families and provides a valid chi-square test for both linkage and association. Computer simulations show that the power of ITDT is greatly improved compared to the TDTsp, especially under dominant disease models.

51

Adiponectin Structural Gene Accounts for the Major Genome-Wide Linkage Signal for Plasma Adiponectin: The IRAS Family Study

X. Guo(1), M.F. Saad(2), C.D. Langefeld(3), S.R. Beck(3), J. Cui(1), K.D. Taylor(1), J.M. Norris (5), S. Jinagouda(2), C. Darwin (2), R.N. Bergman(6), B.S. Sutton(4), L.E. Wagenknecht(3), D.W. Bowden(4), J.I. Rotter(1)

(1) Medical Genet. Inst., Cedars-Sinai Med. Ctr.; (2) Dept. of Medicine, UCLA, (3) Dept. of Pub. Health Sci., and (4) Biochemistry, Wake Forest University; (5) Dept. of Prev. Medicine, Univ. of Colorado Health Sci. Ctr.; (6) Dept. of Physiology and Biophysics, USC, USA.

Plasma adiponectin (ADP) levels have been shown to be under genetic control and variants in the ADP gene have been associated with insulin resistance, type 2 diabetes, and plasma ADP levels. We assessed the genetic basis of plasma ADP at a genome wide level in Hispanic American (HA) and African American (AA) families. ADP levels were measured by radioimmunoassay (Linco Research, St Charles, MO). A 10 cM genome scan (marker set 11) was conducted on 130 families by the Mammalian Genotyping Center (MGS) in Marshfield in two separate batches: MGS1: 66 families (HA: 45; AA: 21) of 877 individuals (HA: 593; AA: 284), MGS2: 64 families (HA: 43; AA: 21) of 662 individuals (HA: 434; AA: 228). 12 SNPs within the ADP structural gene were genotyped for MGS1 samples using the Sequenom MassArray Genotyping System. Diabetic subjects were excluded, leaving 1539 (1027 HA in 88 families, 512 AA in 42 families) in the analysis. The genome scan linkage analysis mapped a major gene to 3q27 (LOD=6.35 in MGS1, 1.62 in MGS2, and 5.98 in ALL, 218 cM) in the HA sample, but no evidence of linkage in the AA families at this location. 12 SNPs were selected in the promoter, coding sequence and 3-UTR of the ADP gene. Association analysis revealed that ADP level is associated with 2 of the 12 SNPs. Incorporating each SNP in the linkage analysis as a covariate reduced the LOD score dramatically (residual LOD scores range from 1 to 2), suggesting that the ADP structural gene accounts for the majority of the linkage in the HA sample.

52

On the use of familial aggregation stratified on proband exposure to detect gene-environment interaction

E. Génin and C. Bonaïti-Pellié
INSERM U535, Villejuif, France

Gene-environment (GXE) interactions may play important roles in complex disease susceptibility but their detection is often difficult. Here we show how GXE interactions can be detected by investigating the degree of familial aggregation according to the exposure, not only of relatives, but also of the probands. Indeed, in case of gene-environment interaction, the distribution of genotypes of affected individuals, and consequently the risk in relatives, is dependent on their exposure. We derived the formulae for calculating the risks in sibs according to the exposure of probands and sibs for various values of exposure frequency (fE), susceptibility allele A frequency (p), relative risk due to exposure alone RE, relative risks due to genotypes Aa and AA respectively (RG1 and RG2), and interaction coefficients I1 and I2. We found that the ratio R of risks in sibs when the proband is exposed and not exposed was a good indicator of the interaction effect. As expected, this ratio increases with RG1 and RG2, and also with fE. It was usually higher when gene frequency was set to 0.1 than 0.5. It increased also as RE, RG1 and RG2 increased and was higher under recessive than dominant models. We computed the power of the test comparing the risks in sibs according to the proband exposure for samples of 300 and 1000 affected individuals. We concluded that this test was valuable for diseases with moderate familial aggregation, only when the role of the exposure has been clearly evidenced.

53

Estimating proportional ancestry in diverse world populations

I. Halder(1), J. Gelernter(2), M.D. Shriver(1)
(1) Pennsylvania State University; (2) Yale University

In an effort to understand the genetic variation within populations and how this variation might influence disease risk we have analyzed multilocus genotypes for 37 microsatellites from 6 diverse populations with different population histories. The populations studied include European Americans (EA), African Americans, two US Hispanic populations from the east and west coast, Finns, and Thais. The markers include all 13 CODIS STRs and 24 other markers with high African-European allele frequency difference. We present allele frequency estimates for all markers in all populations above and in Spanish (Sp), Sierra Leonese (SL) and Mexican (M) populations. Using a maximum likelihood based method and a separate Bayesian method (implemented in the program Structure); we have estimated individual BioGeographical ancestry (BGA) for all individuals. We observed high correlation between estimates obtained with both methods. All populations studied show wide variations in individual BGA. The two Hispanic groups showed different BGA profiles and were significantly different along all three ancestry axes. Neighbor Joining trees show that the east coast Hispanics are closer to African Americans, while the West Coast Hispanics share more European ancestry. EA cluster with Finns and both populations show some Native American ancestry. Simulations were done to explore the biological basis of this phenomenon. To investigate the influence of East Asian ancestry in Finns, we used Thai's as a presumptive parental population in three way admixture models with Sp and SL populations and observed ancestry proportions similar to those obtained with Mexicans. Replacing SL with M resulted in distributing the proportional non-European ancestry almost equally between the two other groups. Further simulation studies were done to explore the variability observed in our populations and will be detailed.

54

A simple correction of the intrafamilial correlation for association studies

J.B. Harley(1), S.K. Nath(1)
(1) University of Oklahoma & Oklahoma Medical Research Foundation, Oklahoma City, OK, USA.

Mapping genes for complex diseases often involve case-control disease-marker association approaches. However,

modest effect sizes are expected for loci predisposing to complex diseases, hundreds or even thousands of individuals may be required to obtain sufficient power to detect association. Therefore, some power-enhancing strategies are required which can take full advantage of already available multiplex family data. The use of affected relatives from the multiplex pedigrees will increase the sample size. However, the use of multiplex families in genetic association studies is limited by the complications surrounding the correlation that arises between relatives. When left uncorrected the association between relatives who share genetic material identical by descent falsely increases the resulting calculation of significance (type I error).
We propose to recast this problem in a way that provides an exact solution by adjusting the sample size. We reframe the problem by considering the number of independent genomes among the related affecteds instead of using the number of cases. Then the independent genomes are then added across pedigrees. We employ an iterative approach that conditions on the affected relatives sequentially in each pedigree. Using the samples already demonstrating linkage for the association study maximizes the opportunity to detect genetic association. The correction provided herein, which uses all the available biologically related cases from multiplex families, will provide access to increased power for such studies.

55

An Insulin Gene Haplotype: Association with Impaired Glucose Tolerance and Modified Insulin Response

I.M. Heid(1), M. Kolz(1), N. Klopp(1), W. Rathmann(2), W. Koenig(3), C. Meisinger(1), A. Luchner(4), F. Kronenberg(1), H.-E. Wichmann(1), H. Kolb(2), T. Illig(1)

for the KORA study group (1) GSF-Institute of Epidemiology, Neuherberg; Universities of (2) Düsseldorf; (3) Ulm; (4) Regensburg, Germany

To describe the role of the insulin gene (INS) in the pathogenesis of type 2 diabetes (T2D), we analyzed data of 226 subjects with T2D, 233 with impaired glucose tolerance (IGT) and 236 controls from a population-based survey, all characterized regarding glucose tolerance, insulin levels, triglycerides. INS-23, -806, -1128 and -1141 were genotyped and haplotypes reconstructed.
Besides the wildtype ACCC, we observed two haplotypes TCCA (9%) and TGTA (21%). Compared to ACCC/ACCC, frequency of TGTA/TGTA subjects was significantly lower among IGT (2%) as compared to T2D (11%, OR=6.7, p=0.0008). Further, frequency of TGTA/TGTA subjects was lower comparing IGT to controls. We found, for TGTA/TGTA as compared to ACCC/ACCC subjects, (a) twice as high insulin levels among the obese (p=0.03), (b) steeper insulin levels increase per BMI WHO category, (c) higher correlation between BMI and insulin levels (r=0.8 vs. 0.5) and (d) higher triglycerides.
We present for the first time a significantly lower frequency of the INS diplotype TGTA/TGTA among IGT subjects. We hypothesize that the observed increased insulin secretion as reponse to higher body mass in subjects with TGTA/TGTA possibly stresses the beta cells leading to accelerated transition through IGT state into overt T2D and consequently to a lower probability of observing these subjects in IGT state.

56

Segregation analysis of Restless Legs Syndrome (RLS)

W.A. Hening(1), R.A. Mathias(2), M. Washburn(3), R.P. Allen(3), S. Lesage(3), A.F. Wilson(2), C.J. Earley(3)
(1) UMDNJ-RW Johnson Medical School, New Brunswick, USA; (2) NIH/NHRGI/IDRB Baltimore, USA; (3) Johns Hopkins University, Balitmore, USA

RLS is a sensorimotor disorder that has a prevalence of ~7% in the US. A diagnosis of RLS requires establishing an urge to move localized to the legs which is provoked by rest, relieved by activity, and accentuated late in the normal wake period and early sleep period. A segregation analysis done in German families with probands having an age of onset below 30 supported a dominant pattern of inheritance [Winkelmann J et al., Ann Neurol, 2002, 52: 297-302].
Probands were selected from consecutive RLS patients from the Neurology and Sleep clinics of the Johns-Hopkins Bayview Medical Center. Patients willing to have first and second degree relatives contacted were included. An RLS diagnosis was made in those who had the four diagnostic features of RLS [Allen RP et al., Sleep Med, 2003, 4:101–119] and whose symptoms could not be explained by an alternate diagnosis. There are 2060 subjects in 76 pedigrees with 534 sibships. The average pedigree size is 27 (range 6–90) with an average of 4.8 generations. Phenotype data is available on 590 subjects of whom 281 have a confirmed diagnosis of RLS. Initial examination of simple segregation ratios suggests an autosomal dominant mode of inheritance of RLS that appears to be more pronounced in families with an earlier proband age of onset. A formal segregation analysis of these pedigrees is ongoing.

57

Logistic regression and conditional linkage analysis based on NPL scores to detect interaction between two loci: a power comparison

J. Hoegel(1), C. Maier(1), Th. Paiss(2), W. Vogel(1)
(1) Dept Human Genetics, Ulm University, Germany; (2) Dept of Urology, Ulm University, Germany

In genome wide scans for susceptibility genes in complex diseases, locus interaction and heterogeneity may hamper the identification of specific loci. Langefeld and Boehnke developed a conditional logistic regression model based on multipoint family-specific nonparametric linkage (NPL) scores that ''allows for conditional or simultaneous tests of the effects of multiple trait loci and of interactions among sets of loci'' [1]. In the case of two loci, Cox et al. [2] proposed a conditional analysis method where, in each family, NPL scores at one locus A are weighted according to the evidence of linkage at another locus B. The weighted scores at A may then be

compared with the corresponding unweighted scores at A to test for interaction with locus B. Different weighting schemes are in use.
A simulation study will be presented to compare the statistical power of these two approaches in detecting interaction between loci or linkage at a locus after incorporating information from another one. In this context, interaction is reflected by the degree of correlation between the family-specific NPL scores at the loci involved. Using observed data from a genome screen for prostate cancer loci in our family sample, the logistic regression model proved to be superior in many instances. The analyses were based on re-sampling techniques generating samples with predefined properties of NPL score distributions.
1. Genet Epidemiol 21 (Suppl 1), 136–141 (2001)
2. Nature Genet. 21, 213–215 (1999)

58

Genetic contribution to blood pressure: heritability and linkage results from Dutch twin and sibling pairs

J.J. Hottenga(1), J.C.N. de Geus(1), A.H.M. Willemsen(1), H. Snieder(1), H.M. Kupper(1), M. van den Berg(1), L.A. Beem(1), B.T. Heijmans(2), M. Beekman(2), P.E. Slagboom(2), D. Posthuma(1), D.I. Boomsma(1)
(1) Dept. of Biological Psychology, Free University Amsterdam, The Netherlands; (2) Dept. of Molecular Epidemiology, Leiden University Medical Center, The Netherlands.

High blood pressure is a major public health problem of largely unknown cause. We recruited twin families through the Netherlands Twin Registry (NTR) for participation in laboratory and ambulatory studies of blood pressure under rest and stress conditions. A total of 731 twin families participated at least once, consisting of 598 MZ twins, 704 DZ twins, 360 additional siblings, 160 mothers and 160 fathers. Average age of the offspring was 33.8 years at time of participation and there was an equal representation of males and females. The average of a minimum of three blood pressure measurements was used. We explored several strategies to deal with data from subjects with antihypertensive medication. Variation in resting blood pressure levels showed a heritability of around 50% for systolic and diastolic pressure. For a subsample of 491 sibling pairs microsatellite markers were available from two genome scans with an average spacing of 10 cM. Variance components QTL analysis in sib pairs revealed some linkage peaks with suggestive linkage (LOD > 2) and one with significant linkage (LOD > 4).
Acknowledgements: Marshfield Laboratories; NIH (R01 HL55976); NHS (86.083, 88.042, 90.313); USF (VU-USF 96/22); GenomEUtwin (QLRT-2001-01254); NWO (250 4227, 904-61-193); CMSB; CNCR-VU.

59

Modelling excess survival in relatives of pairs of siblings aged 90 years and over

J.J. Houwing-Duistermaat(1), M. Schoenmaker(2), M. Beekman(1), T. de Craen(2), E. Slagboom(1), J.C. van Houwelingen(1)
(1) Dept of Medical Statistics and Bioinformatics; (2) Dept of Internal Medicine, Leiden University Medical Center, Leiden, The Netherlands

In an ongoing study of longevity, Dutch long-lived sibling pairs both older than 90 years are being collected. Current age or age at death are available for the parents, siblings and offspring of these siblings. At this moment 100 sibling pairs are available. Using survival data for the Dutch population, it appeared that excess survival in relatives of long-lived sibling pairs exists (SMR=0.42, 95%CI=0.36–0.48). The aim of this study is to verify whether the excess survival in the families depends on the current ages of the long-lived sib pairs.
A Poisson model is proposed with as covariate a score based on the ages of the long-lived sib-pairs. Denote the long-lived sib-pair by index 1 and 2 and their relatives by index i with $i > 2$. For a specific age interval, let D_i be 1 if subject i died in that interval and 0 otherwise and let H_i be the cumulative hazard for the same age interval. Then the model is given by

$$D_i \sim \text{Poisson}(H_i \exp(2\theta(\mathbf{\Phi}_{i1}(D_1 - H_1))) + \mathbf{\Phi}_{i2}(D_2 - H_2))$$

with $\mathbf{\Phi}_{ik}$ equal to the kinship coefficient between relative i and sibling k (k=1,2). Note that $D_1=D_2=0$. The parameter θ models the excess survival in these families.
The parameter θ is highly significant ($P < 0.001$). Furthermore this model gave a better fit than the model with just a constant. Hence to increase the power of a linkage study the current ages of the long lived sibling pairs should be taken into account.

60

Differential heritability estimates of diabetes related traits from families with lean diabetic members

W.-C. Hsueh(1), A.R. Shuldiner(2), B.D. Mitchell(2)
Dept. of Medicine at (1) UCSF, San Francisco, CA; (2) Univ. of Maryland, Baltimore, MD, USA

Obesity is the most important risk factor of type 2 diabetes (T2DM), which may exert its effects via both genetic and non-genetic mechanisms. We hypothesize that genetic influences on T2DM-related traits, estimated as heritability (h^2), may be greater in the absence of obesity.
Data were available from 424 Amish subjects in 20 families. These families were divided into 2 groups. One included families with a mean BMI of diabetic members at age 25 (before their T2DM onset) on the top 50% (obese-diabetic), the other in the low 50% (lean-diabetic). The BMI cut point was 23.8. H^2 for T2DM traits (fasting glucose and insulin, glucose level 2 hrs and insulin level 30 minutes after a glucose challenge, and HbA1c) were compared between 2 groups.
Our findings (h^2) are shown below. The prevalence rates of T2DM were comparable in 2 groups ($\sim$15%, unadjusted). Overall, h^2 for traits measured in fasting state (fasting

glucose and insulin, HbA1c) were significant and greater in families not enriched with obesity; while h^2 for traits measured after a glucose challenge appeared not to be affected by the presence of obesity. These observations support our previous findings that the genetic effects may be stronger in families not affected by obesity and that genetic influences on "basal" vs. post-challenge states may be different.

Family group	Fasting glu	Glu 120		HbA1c	Fasting ins Ins 30
Obese-diabetic	.02*	.41	.09*	.00*	.21*
Lean-diabetic	.26	.40	.36	.25	.20
Combined	.14	.38	.23	.09*	.23

*$p > .05$.

61

Ascertainment corrected score tests for haplotype-trait association

K. Humphreys(1), J. Palmgren(1,2)
(1) Dept of Medical Epi and Biostat, Karolinska Institutet, Sweden; (2) Dept of Math Stat, Stockholm University, Sweden

We consider the problem of testing for association between a binary trait Y and a genotype G in situations where a lifestyle risk factor E may be necessary to trigger the effect. Both Y and E are subject to non-random ascertainment. E is also assumed to be associated with G. Multiple markers and haplotypes as well as additional environmental and lifestyle factors are considered. Motivation stems from studying the association between candidate genes in hormonal pathways (e.g. an estrogen receptor gene) and breast cancer risk, with the possibility (not known in advance) that an association may only be present in women receiving hormone-replacement-therapy (HRT). It is conceivable that genes in this same pathway affect women's postmenopausal symptoms, and thus also their HRT use. A population based breast cancer case-control study design is used, where women on long-term HRT have been over-sampled. We propose an ascertainment correction to the Schaid et al score test for haplotype-trait association, which induces a modification to the link function in the generalized linear model setting. The procedure is useful for many non-standard generalized linear models, and it is conceptually and computationally simple. We illustrate the method and we discuss validity and power under various ascertainment schemes.

62

Haplotype tagging SNPs selected in one human population are highly informative in additional populations

F.C.L. Hyland, B. Halldórsson, H. Isaac, S. Istrail, C. Scafe, E. Spier, and F. M. De La Vega
Applied Biosystems, 850 Lincoln Centre Dr., Foster City, CA 94404, USA

To confirm disease associations, it is desirable to know whether a set of haplotype tagging SNPs (tSNPs) selected in one population will be informative in other populations. Tagging SNPs were selected separately for each of 4 populations genotyped on >100,000 SNPs spanning the human genome. The number of tSNPs needed to maintain a haplotype r2 above a critical threshold was computed separately for each population: more tSNPs were needed in African Americans than in Caucasians, Chinese or Japanese. The effect of SNP density was examined. When subsets of SNPs of various sizes were sampled, and the degree to which the subset tagged the 'hidden' SNPs was calculated, the hidden SNPs were more completely tagged in Caucasians and Asians than in African Americans. The average haplotype r2 of the Caucasian tSNPs in Caucasians, and of the Caucasian tSNPs in African Americans, Chinese and Japanese was computed, and vice versa. The average haplotype r2 of Caucasian tSNPs used in Asians or African Americans was very close to the haplotype r2 of Caucasian tSNPs used in Caucasians, or Asian tSNPs used in Asians, or African American tSNPs used in African Americans, indicating that tSNPs selected in one population will work well in other populations. The number of tagging SNPs in common across the populations was also examined. About 65% of tSNPs were found to be common across populations. Applying a cost function to maximize common SNPs across populations resulted in a minimal increase in the % of common tSNPs.

63

Frailty models for linkage analysis of survival data

I. Iachine
Dept. of Statistics, Univ. of Southern Denmark, DK

Traditional methods of linkage analysis of sib-pair data for quantitative traits are not always appropriate for the analysis of survival data because of censoring and truncation. Multivariate survival models, on the other hand, are traditionally based on the concept of hazards and relative risks and not on the moment structure of the outcomes, which are key components of many genetic analysis methods, such as the variance component method of linkage analysis. Here the covariance structure of the trait is related to the degree of genetic similarity between family members at a particular genetic locus (described by the conditional distribution of the number of alleles shared identically by descent by a pair of family members given the genetic marker information).

There is therefore a need for integration of methods of linkage analysis with multivariate survival models. The aim of this presentation is to show how frailty models may be used for this purpose. The advantage of the frailty approach is that it allows for flexible modelling of the covariance structure of the frailty variables while allowing for a semiparametric structure of the survival model. The idea is to consider frailty as a new intermediate phenotype mediating the genetic influence on longevity and to apply the methods of genetic linkage analysis (e.g. variance component method) to the frailty variable. The implementation of this idea is particularly simple for the ''correlated frailty'' models – multivariate frailty

models where the pairwise correlation coefficient of frailty is explicitly modelled as a parameter of the survival model. The approach is illustrated by simulation studies.

64

Evidence for a Myopia Susceptibility Locus on Chromosome 22 in an Ashkenazi Jewish Population

G.P. Ibay(2), D. Stambolian(1), L. Reider(1), D. Dana(1), M. Schlifka(1), T. Holmes(2), R. Owens(3), E. Ciner(4), J.E. Bailey-Wilson(2)
(1) Dept. of Ophthalmol., UPenn., PA; (2) IDRB/NHGRI/NIH, MD; (3) Owens Optometrics, Lancaster, PA; (4) Penn. College of Optometry, PA

Mild/moderate (common) myopia is a very common disorder with both genetic and environmental influences. The environmental factors are few, which include near-work and possibly night lights, and are easy to measure. There are no known genetic loci for common myopia. Our goal is to find evidence for a myopia susceptibility gene causing common myopia. Cycloplegic and manifest refraction were performed on 44 large U.S. Ashkenazi Jewish families with at least two affected siblings. Individuals with at least −1.00 D in each meridian of both eyes were classified as myopic. Microsatellite genotyping (387 markers) was conducted by the Center for Inherited Disease Research. Linkage analyses used parametric methods under twelve different penetrance models and nonparametric methods. The family-based association test (FBAT) was used for an association scan. A maximum multipoint parametric heterogeneity HLOD score of 3.54 was observed at marker D22S685 and non-parametric (NPL) analyses gave consistent results, with a p-value of 0.0002 at this same marker. The HLOD's exceeded 3.0 for a 4 cM interval and significant evidence of genetic heterogeneity was observed. This genome wide scan is the first step toward identifying a gene on chromosome 22 with an influence on common myopia. We are following up our linkage result on chromosome 22 with a dense map of over 1500 SNP markers for fine-mapping and association analyses.

65

Unbiased estimation of htSNP efficacy

M.M. Iles
Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Haplotype tagging SNPs (htSNPs) are widely used as a means of capturing the majority of the genetic diversity in a region while minimising the amount of genotyping to be performed. Ideally, selection of htSNPs should be carried out using data on all polymorphisms in a region. However, it is usual to genotype a subset of all the SNPs in a region and use these data to select the htSNPs and judge their efficacy simultaneously. We demonstrate using publicly available real data that such an approach leads to biased estimates of htSNP performance and that this bias increases as the spacing of the observed SNPs increases. At an observed SNP density of 2kb, htSNP analysis suggests that the htSNPs capture on average 95% of the observed variation, when in fact they capture 88% of the unobserved variation. At a density of 10kb htSNP analysis suggests that 93% of the observed variation was captured, when in fact they capture on average only 78%.
We propose a method utilising cross-validation to correct for this problem and show that this approach is both unbiassed and more accurate than existing methods. The method is applicable both to small-scale studies of a few regions and to large-scale studies of whole chromosomes or genomes such as HapMap.

66

Lipids: Heritabilities and ApoE in a Genetically Isolated Population

A. Isaacs(1), F.A. Sayed-Tabatabaei(1), Y. S. Aulchenko(1), A.F.C. Schut(1), C. Zillikens(2), H.A.P. Pols(2), J.C.M. Witteman(3), B.A. Oostra(4), C.M. van Duijn(1)
(1) Gen. Epi. Unit; (2) Dept. of Int. Med.; (3) Dept. of Epi. and Biostat.; (4) Dept. of Clin. Gen.; Erasmus Univ. M.C., Rotterdam, NL

Despite progress in unravelling the genetics of dyslipidemia, most findings are based on families with extreme phenotypes. To further dissect these complex traits, a large extended pedigree ($n > 6000$), derived randomly from the population of a recent genetic isolate in the Netherlands, was ascertained. 951 members of this pedigree took part in comprehensive physical examinations and medical interviews in this ongoing study. Laboratory analysis of these subjects included testing fasting plasma lipids. Heritabilities (h^2) for lipids were estimated, using the SOLAR program. The model included multiple covariates: age, sex, alcohol use, smoking, age$*$sex, age^2, diabetes, lipid therapy, hormone therapy, and inbreeding coefficient (calculated using PEDIG). h^2 for total cholesterol (TC), HDL, LDL, and TG were found to be 0.30, 0.51, 0.20, and 0.22, respectively. In a further analysis, apolipoprotein E (apoE) genotype was included as a covariate. ApoE altered only one estimate, that of TC (h^2=0.17). Allele based analysis demonstrated that apoE was not a significant covariate in the heritability of other lipids. In conclusion, our studies in a randomly drawn pedigree show that genetic factors play an important role in plasma lipid levels, particularly HDL. Our findings also suggest that apoE accounts for a large portion of the additive genetic variance of TC, but not other plasma lipid measures.

67

Lod score and MCMC methods applied to a complex trait: LDL size

G.P Jarvik, M.D. Badzioch, R.P. Igo, F. Gagnon, J.D. Brunzell, R.M. Krauss, A.G. Motulsky, E.M. Wijsman
University of Washington, Seattle, USA.

Low density lipoprotein (LDL) size is associated with vascular disease and with familial combined hyperlipide-

mia (FCHL). We used lod score and Bayesian Markov chain Monte Carlo (MCMC) oligogenic linkage analysis methods for a 10 cM genome scan of LDL peak particle diameter and adjusted for age, sex, body mass index, and triglycerides (PPD_{asbt}), in 4 large families with FCHL (N=185). Complex segregation analyses (CSA) found a codominant PPD_{asbt} model with Vg/Vt=0.6, vs. 0.36 without triglyceride (TG) adjustment. CSA models were used for the lod score analyses. MCMC analyses used joint segregation/linkage analyses with an oligogenic model. We identified significant evidence of PPD_{asbt} linkage to a chromosome (chr) 9p locus (multipoint lod_{max}=3.7, MCMC intensity ratio, IR=21) in family 1, and across all 4 families to chrs 16q23 (lod_{max}=3.0, IR=43) near [CETP] and 11q22 (lod_{max}=3.7, IR=120). Chr 14q24-31, a region with prior suggestive PPD_{asbt} linkage evidence in another sample, yielded an IR=71 and lod_{max}=1.8 in the combined families and IR=24 and lod_{max}=2.4 in family 4. Without TG adjustment, no significant lod scores were obtained. These large FCHL pedigrees demonstrate that LDL size is a trait influenced by multiple loci, and is confounded by TG. This underscores the use of relatively homogeneous families enriched for the trait. Lod score methods performed well, despite locus heterogeneity. However, MCMC methods were required to detect the previously reported 14q locus. These results also demonstrate the applicability of MCMC methods to single large families.

68

The Effect of Ethnicity on the Construction of Human Genetic Linkage Maps

E. Jorgenson(1), H. Tang, M. Gadde(1), M. Province, M. Leppert, S. Kardia, N. Schork, R. Cooper, DC Rao, E Boerwinkle and N Risch(1)
(1) Department of Genetics, Stanford University; (2) This investigation was based on data from the Family Blood Pressure Program (FBPP), which is supported by the National Institute of Heart, Lung, and Blood Institute. All authors are members of the FBPP.

Human genetic linkage maps are based on rates of recombination across the genome. These rates in humans vary by the sex of the parent from which alleles are inherited, chromosomal position and genomic features such as GC content and repeat density. We demonstrate for the first time that ethnicity affects genetic map length in humans. We constructed genetics maps in four racial/ethnic groups: Caucasians, African-Americans, Mexican-Americans, and East Asians (Chinese and Japanese) based on 353 microsatellite markers. These maps were generated from the largest number of human subjects of any map constructed to date. We identified regional and genome-wide differences across ethnic groups. Some, but not all, of this variation was explained by the presence of null alleles at many of the loci and ethnic differences in null allele frequencies. The results of our investigation are instructive both for inferences of possible genetic influences on human recombination as well as for future linkage studies, especially those involving populations of non-Caucasian ethnicity.

69

Polymorphisms in the Innate Immunity Pathway are Associated with Prostate Cancer Aggressiveness

E. Jorgenson(1), S. Plummer(2), G. Casey(2), and J.S. Witte(1)
(1) Department of Epidemiology and Biostatistics, University of California, San Francisco California 94143-0560; (2) Department of Cancer Biology, Cleveland Clinic Foundation, Ohio 44106

The aggressiveness of prostate cancer varies widely, with some tumors progressing rapidly to life-threatening disease, and others remaining relatively latent. Finding genetic variants that help explain this variability would provide valuable information both for screening and determining the best course of clinical treatment. The innate immunity pathway has previously been implicated in the development of prostate cancer, but has not been examined for prostate cancer aggressiveness. Here, we evaluate the relation between polymorphisms in genes involved with innate immunity and prostate cancer aggressiveness as measured by Gleason score in a case-only study (623 cases). We found that carrying at least one copy of the interleukin-1-31 (IL1-31) C→T polymorphism was associated with having a Gleason score greater than 6, a marker for aggressive disease (odds ratio (OR)=1.72, 95% Confidence Interval (CI)=1.22–2.38, p=0.0003). Our findings suggest that the presence of at least one copy of the IL1-31C→T polymorphism increases the risk of being diagnosed with more aggressive prostate cancer in individuals who develop this disease.

70

Confirmation, Localization and Genetic Association of SLEH1 at 11q14 in Pedigrees Multiplex for SLE with Hemolytic Anemia

J.A. Kelly(1), S. Shadfar(1), T.J. Aberle(1), G.R. Bruner(1), K.M. Kaufman(1), J.B. Harley(1)
(1) Oklahoma Medical Research Foundation, OK, USA

The SLE locus, SLEH1, was first identified on 11q14 (LOD=4.7) in 16 African-American (AA) multiplex pedigrees that contained at least one individual affected with both SLE and hemolytic anemia (PNAS 99(18):11766–11771, 2002). In an effort to confirm this finding, we identified an additional 13 AA pedigrees multiplex for SLE with hemolytic anemia in one or more SLE affected. The additional 13 pedigrees confirmed linkage to SLEH1 (LOD=2.0). To evaluate the region for possible genetic association, we typed 112 SNPs spanning 74.3=85 Mb on chromosome 11q14-21 in 40 AA pedigrees linked to D11S2002. One SNP showed a weak association using TDT (p=0.004), PDT (p=0.04) and FBAT (p=0.003). Fine mapping of 19 SNPs around this SNP produced 6 SNPs that showed family based association ($p<0.05$). In addition, two of the six SNPs showed significantly different allele frequencies between 40 cases and 46 controls ($\div 2$=4.66, p=0.03). For one of the two SNPs, a more significant difference was observed in the allele frequencies between cases and controls when the cases

selected had hemolytic anemia ($\div 2$=9.31, p=0.002). More SNPs are currently being tested in the neighborhood to define the LD and to help determine if the association signal detected is robust. These results provide further evidence demonstrating that SLEH1 is a strong and reproducible SLE susceptibility locus. The chromosomal locus originally spanning 10 cM is now confirmed and has been further localized to a 2-3 Mb region on chromosome 11q14-21.

71

Serum Cortisol and DHEAS Levels are Genetically Correlated with Obesity but not Insulin Resistance in Mexican Americans

J.W. Kent Jr., R.A. Bastarrachea, A.G. Comuzzie, J.W. MacCluer, J. Blangero
Dept. of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas USA

Disturbance of the hypothalamus-pituitary-adrenal (HPA) axis is associated with the metabolic syndrome, and has been proposed as a mechanism linking depression and stress to obesity and/or type 2 diabetes. We have examined the relationship between morning serum cortisol levels and selected traits of the metabolic syndrome in 686 Mexican American participants in the San Antonio Family Heart Study. With sex, age, and menopause status included as covariates, cortisol levels are negatively genetically correlated with body mass index (BMI) (rhoG=-0.53 ± 0.25, p=0.03) and serum leptin levels (rhoG=-0.54 ± 0.26, p=0.04), but not with HOMA-IR, a measure of insulin resistance (rhoG=-0.04 ± 0.23, p=0.85) or systolic blood pressure (rhoG=-0.06 ± 0.29, p=0.85). Similarly, the adrenocortical androgen dehydroepiandrosterone sulfate (DHEAS) is negatively genetically correlated with BMI (rhoG=-0.26 ± 0.13, p=0.04) but not with HOMA-IR (rhoG=-0.09 ± 0.14, p=0.54), suggesting a comparable pattern of HPA axis regulation. These results suggest heterogeneity in the genetic pathways relating the HPA axis to various components of the metabolic syndrome.

72

Familial association of autoimmune diseases and lymphoproliferative tumors

K. Kerstann(1), G. Gridley(2), B.J. Stone(2), L. Mellemkjaer(3), J. Olsen(3), K. Hemminki(4), L. Goldin(1)
(1) Genetic Epidemiolgy Branch, DCEG/NCI, Bethesda, MD; (2) Biostatistics Branch, DCEG/NCI, Bethesda, MD; (3) Danish Cancer Society, Copenhagen DK; (4) Karolinska Institute, Stockholm, Sweden

Abnormalities of the immune system, such as immune deficiencies and auto-immunity, are known to be risk factors for mature B-cell tumors. Large cohort studies of patients with systemic autoimmune (AI) diseases (i.e. rheumatoid arthritis and Sjogren's) show that they are at increased risk for developing lymphoma. These associations may reflect some common genetic susceptibility factors or result from chronic immune stimulation. We have analyzed 34 AI disease outcomes (from hospital discharge registries) in first-degree relatives of approximately 52,000 cases diagnosed with 4 lymphoproliferative (LP) tumors: Hodgkin lymphoma (HL), non-Hodgkin lymphoma (NHL), chronic lymphocytic leukemia (CLL) and multiple myeloma (MM) and compared them to relatives of matched control groups from population databases in Sweden and Denmark. The 34 AI conditions were classified into 3 groups – systemic, organ specific, and suspected AI etiology. We used a marginal survival analysis model with a robust variance estimate to take into account dependencies among relatives. Relatives who had an LP tumor diagnosis were not included. There was no increased risk of any of the 3 groups of AI diseases in relatives of HL, CLL and MM cases. Relatives of NHL cases were at significantly increased risk for developing a systemic AI disease. Females had a higher risk for developing an AI disease (as observed in the general population) but familial risk was similar in males and females. Family history of LP tumor and age at diagnosis of case probands were not significant predictors of risk. These results are consistent with the associations seen in cohorts and suggest common genetic susceptibility factors affecting both AI diseases and LP tumors.

73

The Association of Malignancy in Systemic Lupus Erythematosus

X.R. Kim-Howard(1) A.I. Quintero-Del-Rio(2,3), J. Morrisey(2), L.M. Feo(2), J.B. Harley(2,3,4)
(1) Genetic Epidemiology Unit, Oklahoma Med Research Foundation, USA; (2) Arthritis & Immunology, Oklahoma Med. Research Foundation, USA; (3) Univ. of Oklahoma Health Science Center, USA; (4) the US Dept. of Veterans Affairs Med. Center, Oklahoma City, OK, USA.

Purpose: We investigated the frequency and clinical characteristics of malignancies among SLE patients. Method: We reviewed medical records of the 910 lupus patients in the Oklahoma lupus genetic studies, included 171 (75 European-American, 20 African-American and 76 Hispanic) from simplex pedigrees and 739 (430 European-American, 224 African-American and 85 Hispanics) from multiplex pedigrees.
Results: Forty-seven patients developed malignancies, 17 were obtained from the simplex and 30 from the multiplex pedigrees p=0.0017 (c2=9.81, O.R.=2.6, C.I.=1.40–4.85). European-American SLE patients had a higher risk of developing malignancy relative to the African-American and Hispanics (c2=6.63, p=0.036). Fisher's exact test was performed to evaluate the prevalence of the different types of cancers in our population in relation to the US Cancer Prevalence Statistics from 1996–2000. We observed a higher risk of breast (p=0.0092; OR=4.17, C.I.=1.38–12.62) and cervical (p=0.0045; OR=11, C.I.=1.42–86.88) cancer in our population using the median age of our SLE patients (40.3 years). Other cancers observed were lung, lymphoma, thyroid, renal, colon, melanoma, prostate and pancreatic, were not significantly different from the US population.

Conclusion: SLE patients have an increased prevalence of cancers and the simplex pedigrees for SLE have a higher risk of cancer in relation to the multiplex pedigrees. SLE patients may have an increased genetic risk for cancer that could be secondary to immunologic disregulation.

74

A novel implementation of a robust variance component approach exemplified using SOLAR

A. Kleensang, I.R. König
Institute of Medical Biometry and Statistics, University Hospital Schleswig-Holstein - Campus Lübeck, Lübeck, Germany

Variance component (VC) methods are one of the most popular approaches taken for detecting linkage in quantitative traits. One of the main reasons is its power compared with other available methods. A drawback, however, are assumptions about the phenotypic distribution VC methods require. Specifically, multivariate normality is usually assumed.
In case of violating this assumption, it has been shown that the type I error rates can dramatically exceed the nominal levels. A possible solution is to use a robust VC approach to allow for possible deviations from multivariate normality (Amos 1994; Am J Hum Genet 54:535–543). We propose an alternative approach and utilize a resampling approach by a jack-knife estimator of variance that is asymptotically equivalent to the robust variance estimator (Lipsitz et al. 1994; Biometrics 50:842–846). To date, robust estimators are not implemented in most commonly used VC software (e.g. SOLAR, Genehunter).
To fill this gap, we present a program which utilizes a jack-knife estimator of variance for generalized estimating equations applied to the SOLAR program package. We show how this approach can be transferred to every VC program without altering the VC program sources. Moreover, we present results from extensive simulation studies. The validity and robustness of our approach is investigated under different genetic models and phenotype distributions.

75

Efficient two-stage genome-wide association designs based on False Positive Report Probabilities

Peter Kraft
Depts. of Epi. and Biostat., Harvard School of Public Health, USA

Despite recent advances, very-high-throughput (VHT) technologies capable of genotyping hundreds of thousands of SNPs in individual samples remain prohibitively expensive for the large studies necessary to screen substantial sections of the genome for variants with modest effects on disease risk. I present a two-stage strategy, where a portion of available samples are genotyped with VHT technology, and a small number of the most promising variants are genotyped in the remaining samples with standard techniques. The samples sizes in the first and second stages and the corresponding significance levels are chosen to limit the False Positive Report Probability (FPRP), while maximizing the number of Expected True Positives (ETPs). I show that for a fixed budget, the multi-stage strategy has greater power (a larger number of ETPs) than the single-stage strategy (where all subjects are genotyped using expensive VHT technology). Furthermore, concentrating on the FPRP leads to considerable savings relative to strategies designed to control the family-wise error (e.g. Bonferonni correction). The FPRP and number of ETPs can also accommodate researchers' prior beliefs about the number of causal loci and the magnitude of their effects. I show that the expected number of false positives does not change if the true number and effects of causal loci differs from the specified prior, thus limiting the amount of resources spent chasing ''false leads.'' Consideration of the FPRP under different priors suggests it is better to design scans that cover a fraction of the genome well, rather than all of it poorly.

76

SNP and Haplotype Associations in Pharmacogenomics using HapMap CEPH trios

A.T. Kraja, J.W. Watters, R. Lazarus, Q. Zhang, H.L. McLeod, M.A. Province

CEPH cell lines are a powerful and unique resource for phamacogenetic gene discovery. We treated 456 cell lines with 9 different doses of two common chemotherapies (Docetaxel and 5-FU) in an in vitro search for toxicity response genes. Previously (Kraja et al., 2003), we showed that the viability response is highly heritable (50–83%), and identified QTLs (e.g. chrom 9: LOD=3.4) in genome-wide linkage scans using published microsatellite data (Jean Dausset Foundation). We now extend our pharmacogenetic search to the association level, using HapMap published SNPs (300k+ so far) genotyped on 30 CEPH trios. Concentrating on the 4928 SNPs in the QTL region on chromosome 9, we utilized various statistical methods to deal with the daunting multiple comparisons problem (179 SNPs were significant at $p<0.05$, uncorrected), including the False Discovery Rate (FDR), Storey's Q, and Sequential Multiple Decision Procedures (SMDP: Province, 2000). Just three SNPs were chosen as significant gene signals by SMDP at a power of 95% (but requiring 82% of the sample), with R2 (and P) values for the entire sample of R2=0.1372 (p=0.00053), R2=0.1359 (p=0.00056) and R2=0.1359 (p=0.00056), respectively. We can improve the power to detect signals by first reducing the dimensionality of the problem (by nearly half), if we first subselect LD tag SNPs eliminating those in strong LD with one another (Lazarus et al., 2004). Haplotyping further reduces the dimensionality, so that multiple comparisons methods retain greater power to detect signals. Ultimately, we will replicate/refute all identified signals in an independent sample of cell lines.

77

Linkage analyses on chromosomal regions 15q21 and 18p11 in dyslexia - results from the German bi-center study

I.R. König(1), A. Ziegler(1), J. Schumacher(4), M.M. Nöthen(4,5), E. Plume(3), A. Kleensang(1), M. Manthey(4), M. Duell(4), F. Dahdouh(4), A. Warnke(3), P. Propping(4), H. Remschmidt(2), G. Schulte-Körne(2)
(1) Inst. of Med. Biometry & Statistics, University Hospital Schleswig-Holstein Campus Lübeck; (2) Dept. of Child & Adolescent Psychiatry, University Marburg; (3) Dept. of Child & Adolescent Psychiatry and Psychotherapy, University Würzburg; (4) Inst. of Human Genetics, University Bonn; (5) Life & Brain Center, University of Bonn, Germany

Dyslexia is a specific disorder in learning to read and spell in spite of adequate educational resources and normal intelligence. It affects about 5% of school-aged children, making it the most common of childhood learning disorders (Schulte-Körne, J Child Psychol Psychiatry 2001, 42: 985), and especially spelling disorder often persists into adulthood. Linkage studies have previously identified regions likely to harbour genes contributing to dyslexia, including regions on the long arm of chromosome 15 and the short arm of chromosome 18 (Grigorenko et al., Am J Hum Genet 1997, 60:27, Fisher et al., Nat Genet 2001, 30: 86).
We are performing a German bi-center study for studying the molecular genetic background of dyslexia. Employing a single proband sib pair design (Ziegler, Child Adolesc Psychiatry 1999 8: 35), we have collected to date 250 families each characterized by at least one dyslectic child, one sibling and their parents. We have genotyped densely spaced genetic markers in chromosomal regions 15q21 and 18p11. Results from linkage analyses will be presented.

78

An analytic study of the impact of genotyping error in linkage mapping to complex traits using selected sib pairs

J. Lebrec, H. Putter, J. Houwing-Duistermaat, J.C. van Houwelingen
Dept. of Med Stat & Bioinfo. Leiden Univ. Med. Center. The Netherlands

In the search for genetic determinants of complex traits, the use of selective designs appears to be the only way to gain sufficient power to detect typically small gene effects. Simulations have shown that the impact of genotyping error on evidence for linkage could be particularly severe in affected sib pairs designs, virtually masking most of the evidence for linkage. The impact on quantitative traits appears to be less dramatic in random samples, however it is unclear whether the same dramatic power losses hold in selected samples. A method of choice (first proposed by Sham & Purcell [1]) is now emerging for the analysis of quantitative traits arising from selected sib pairs, it boils down to a regression through the origin of excess IBD sharing on a function of the trait value, whose slope is an estimate of the linkage parameter. In this setting, we show analytically what effects some simple error generating processes induce. In analogy with affected sib pairs, the loss of power can be dramatic in extremely concordant sib pairs designs. In extreme discordant designs, the result will tend to be over-optimistic inference. Surprisingly, the effect of genotyping error will be milder when more extreme selection is used. It is argued that the simplistic error generating mechanisms assumed are reasonable models of real-life situations. By simulations, we demonstrate that our findings can be qualitatively extended to non-equally frequent alleles and to multipoint settings.
[1] 2001. Am J Hum Genet 68:1527–1532.

79

A genomewide search for HDL-C in large Quebec families: further evidence for a QTL on chromosome 16q.

M. Lemire(1), J. Faith(1), M. Marcil(2), T.J. Hudson(1,3), D. Gaudet(4), J. Genest(2), J.C. Engert(3)
(1) McGill Univ and Genome Quebec Innov Centre, Montreal, Canada; (2) Cardiovascular Gen Lab, McGill Univ, Montreal, Canada; (3) Depts of Hum Genet and Med, McGill Univ, Montreal, Canada; (4) Univ of Montreal Community Genomic Med Center, Quebec, Canada

To examine the genetic contribution to HDL-C levels in a Quebec patient population, we have recruited large multiplex families around probands with low HDL-C. A genome-wide scan was performed on 362 individuals from the 12 largest recruited families, with sizes averaging 45 individuals per families. Markov chain Monte Carlo multipoint methods were used to estimate the posterior distribution of the number of alleles shared identical by descent between pairs of individuals; exact calculations were performed when computationally possible. Variance components methods were used to assess linkage between chromosomal regions and quantitative trait loci. The results of the genomewide scan identified several regions indicative of linkage, including a locus on chromosome 1q that has a LOD score of 3.06 ($p=0.00009$) and a locus on chromosome 16q that has a LOD score of 2.30 ($p=0.0006$). The linkage on chromosome 16q was replicated in a second sample of 61 families from the Saguenay-Lac-St-Jean region of Quebec, ascertained for NIDDM or CHD. The same analytic methods gave a LOD score of 2.55 ($p=0.0003$). Two studies involving Mexican Americans and Finnish and Dutch cohorts found in the litterature also provide evidence of linkage to this region of chromosome 16. All four linkage peaks are less than 20 cM from one another. Our findings provide strong additional support for the presence of a QTL for HDL-C on chromosome 16q.

80

Exact p-values for Family Based Tests of Association Via Importance Sampling

J.P. Lewinger(1), S.B. Bull(1,2)

(1) Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada; (2) Dept. of Public Health Sciences, Univ of Toronto, Toronto, Canada

Methods to compute p-values for family based tests of association usually rely on first order asymptotic approximations or Monte Carlo simulation. First order asymptotic approximations can be quite poor for smaller sample sizes or when the p-value to be computed is small. Alternatively, many proposals have been made in the literature to estimate p-values via Monte Carlo. These proposals usually rely on the simplest form of Monte Carlo in which the probability of an event is estimated by the proportion of times the event occurs in a number of simulated independent replications of the experiment. This direct method can require a prohibitively large number of replicates for adequate estimation of small probabilities. To estimate a probability close to 0.01 with 1% relative error, for example, over 1,000,000 Monte Carlo replicates are required.
We propose a simple importance sampling method that can efficiently estimate p-values for a very general family of tests of linkage and association based on conditioning on a sufficient statistic for the null hypothesis. The sampling distribution is obtained by exponential tilting of the null distribution of the test statistic under Mendelian inheritance, with the tilting parameter chosen to minimize the variance of the importance sampling estimate. An explicit bound for the Monte Carlo error is given. In particular applications we found up to a 70 fold efficiency gain over the standard Monte Carlo method.

81

Genetic bridges between multiple subtypes of lymphoma

X. Li, L. Du, H.-Y. Wang, S.-Q. Rao, B.-Sh. Gong, Q. Wang

Gene expression microarray technology provides the global information on transcriptional activities of essentially all genes simultaneously, and it thus promotes the new application of traditional feature selection methods in the fields of molecular biology and life sciences. The basic strategy for the traditional feature selection methods is to seek for a single gene subset that leads to the best prediction of biological types, for example tumor versus normal tissues. Because of complexities and genetic heterogeneities of biological phenotypes (e.g. complex diseases), robust computational approaches are desirable to achieve high generalization performance to different classifiers, robustness to perturbations of the data structures, and easy biological interpretations. The purpose of this study is to extend our newly developed ensemble decision approach to analyze multiple heterogeneous phenotypes (for example, the numerous subtypes of lymphoma) and to elucidate the underlying molecular bridges that underpin the subtypes. The results from an application to lymphoma data of five subtypes indicate that the proposed analysis strategy is feasible and powerful to perform biological subtyping and to reversely engineer the molecular connections between the clinical phenotypes. Biological interpretations with Gene Ontology reveal concerted genetic pathways for multiple lymphoma subtypes.

82

No Association between INS C1127T Polymorphism and Prostate Cancer

L. Li(1), M.S. Cicek(2), G. Casey(2), J.S. Witte(3)
(1) Departments of Family Medicine, Epidemiology and Biostatistics, Case Western Reserve University; (2) Department of Cancer Biology, Cleveland Clinic Foundation; (3) Departments of Epidemiology and Biostatistics, Urology, University of California, San Francisco

High levels of insulin and insulin resistance have been associated with increased risk of prostate cancer. Thus it is possible that genetic variants in the insulin gene (INS) may influence the circulating levels of insulin, and an individual's risk of prostate cancer. One recent study reported a significant association between INS C1127T polymorphism and prostate cancer risk. We sought to further clarify the potential relation of this genetic variant with prostate cancer in a sibling-matched case-control study. Included in the analysis were 918 brothers (439 cases and 479 controls) from 413 discordant families. Age-adjusted matched (by family) odds ratio (OR) were: 1.0 for CC (referent), 0.84 (95% CI=0.54–1.31) for CT, and 1.06 for TT (95% CI=0.24–4.77) (p for trend=0.78). Stratified analyses by disease aggressiveness (Gleason score ^{3} 7 or stage ^{3} T2c) or age at diagnosis (<62 or ^{3} 62 years) yielded similar results. Further adjustment for body mass index (BMI) and height did not substantially alter the results. Our data do not support an important role of INS C1127T polymorphism in the etiology of prostate cancer.

83

Assessing the Effect of Age at Onset on Linkage to Bipolar Disorder: Evidence on Chromosomes 18p and 21q

P.I. Lin(1), M.G. McInnis(2), J.B. Potash(2), D.F. MacKinnon(2), J.R. DePaulo(2), P.P. Zandi(1).
(1) Department of Mental Health, the Bloomberg School of Public Health, Johns Hopkins University, USA; (2) Department of Psychiatry and Behavioral Sciences, School of Medicine, Johns Hopkins University, USA

Linkage findings in bipolar disorder to date have been inconclusive because of inconsistent findings. The failure to replicate earlier findings may be attributable to genetic heterogeneity. Previous evidence suggests that the segregation of BP in families may vary depending upon the age at onset. We, therefore, sought to use age at onset as a covariate in a linkage analysis of BP in order to identify linkage signals within more genetically homogeneous subgroups. We used two model-free linkage methods to study the effect of age at onset on linkage evidence. The first method is a regression-based method implemented in the LODPAL program (part of the S.A.G.E. package). The second method is an ordered-subset analysis. We carried out genome scans with age at onset as a covariate for 150

pedigrees with 1,345 individuals collected from the NIMH Bipolar Disorder Genetics Initiative. All families were ascertained through probands affected with bipolar I disorder, and were all multiplex and unilineal families. A total of 512 microsatellite markers had been genotyped across the genome for an average spacing of 7 cM, while only 298 markers were common for all families. Diagnoses of mood disorders were assigned by the best-estimate procedure. Age at onset was defined as the age at the first manic or depressive episode. We also used another independent sample consisting of 65 pedigrees with 573 subjects to replicate positive linkage findings. A total of 245 markers were genotyped for each pedigree across the genome with an average spacing of approximately 5 cM. These 65 pedigrees were also ascertained through probands affected with bipolar disorder. The best findings from the analyses with the regression-based linkage method were observed on chromosomes 21q22.13 and 18p11.2. After inclusion of age at onset as a dichotomous variable, the linkage score near marker D21S1252 increased from a LOD of 0.02 to 3.29 (P-value for the increase in LOD=0.0001; P-value for the overall LOD=0.0003). Near marker D18S1150 the linkage score increased from a LOD of 0.3.

84

Likelihood-Based Inference on Haplotype-Disease Associations

D.Y. Lin and D. Zeng
Department of Biostatistics, University of North Carolina, USA

The population associations between haplotypes and disease phenotypes provide critical information about the genetic basis of complex human diseases. It is highly challenging to make statistical inference about these associations because of unknown gametic phase in genotype data. We present a general likelihood-based approach to this problem for cohort and case-control studies of unrelated individuals. The phenotype can be a disease indicator, a quantitative trait or a potentially censored time to disease variable. The effects of haplotypes on the phenotype are formulated through flexible regression models, which can accommodate a variety of genetic mechanisms and gene-environment interactions. We construct appropriate likelihood functions for various study designs and disease phenotypes, allowing Hardy-Weinberg disequilibrium. The identifiability of the parameters, and the consistency, asymptotic normality and efficiency of the maximum likelihood estimators are established. Efficient and reliable numerical algorithms are developed. Simulation studies demonstrate that the likelihood-based procedures perform well in practical settings. Applications to the Finland-United States Investigation of NIDDM Genetics Study are provided. Areas in need of further development are discussed.

85

Designs for microarray spotting

P.J. Lindsey(1), C. Klaassen(2), H. Wynn(2)
(1) Dept. of Population Genetics, Genomics, and Bioinformatics, University of Maastricht, The Netherlands; (2) EURANDOM, The Netherlands

In order to layout products optimally over the printing array, designs based on lattice integers can be used. Unfortunately in discrete space, solutions are only available for a prime number of products to be placed a prime number of times in grids with prime number of columns and rows.
In practice, these settings are very inconvenient especially when the greatest number of spots must be printed in order to have all currently known genes present once or more on a microarray. One approach to solve this issue, is to subdivide the microarray into smaller grids complying to these criterion and combining these back together in order to get a design where the products are uniformly distributed over the microarray.
These designs scatter the spikes, controls, housekeeping genes, positive and/or negative controls, and empty spots uniformly throughout the arrays. This therefore ensure that, regardless of experimental losses due to unexpected problems, there will remain the necessary information to carry out an analysis taking into account the experimental sources of variation.

86

Association of the polymorphisms of the IGF2-INS-TH gene cluster with susceptibility to Type 1 Diabetes

M. Liu(1), K.-S. Wang(1), M.-Y. Lu(1), B. Bharaj(1), H.-T. Chen(1), J. Curtis(2), L. A. Newhook(2), A. D. Paterson(1,3)
(1) Program in Genetics and Genomic Biol, The Hospital for Sick Children, Canada; (2) Dept. of Pediatrics, Memorial Univ of Newfoundland; (3) Dept. of Public Health Sciences, Univ of Toronto

The insulin variable number of tandem repeats locus (INS VNTR) is a known candidate for the type 1 diabetes (TID) susceptibility locus IDDM2. Its class I alleles increase susceptibility to TID, whereas its class III alleles present dominant protection to TID. We selected six SNPs and one microsatellite marker which could distinguish the VNTR allele classes at the IGF2-INS-TH gene region. Family-based association analysis was performed with FBAT using 538 extended families with at least one T1D probands (3156 individuals) from Newfoundland (NF). Single-marker TDT analysis shows significant association of INS-27 (−23 HphI, allele A. $P<0.000001$) and the TH01 microsatellite marker (allele 183. $P=0.007$) with TID susceptibility. Allele A of INS-27 is in strong linkage disequilibrium (LD) with class I VNTR. Allele 183 of the TH01 microsatellite is in strong LD with class IC VNTR alleles. A haplotype constituted by all the other 5 SNPs is also significantly associated with TID ($P=0.0001$) even though these SNPs are not associated with T1D individually. This haplotype is also in LD with class IC VNTR alleles. When allele 183 of the TH01 marker was added to this haplotype, the association was stronger ($P=0.00003$). Without genotyping

INS VNTR, other polymorphisms at the IGF2-INS-TH gene region provide highly significant association with T1D in NF population.

87

The DRD4 48-bp-Repeat Polymorphism and Mood Disorders: A Meta-analysis

S. López León(1), E.A. Croes, F.A. Sayed-Tabatabae(1), S. Claes(2), C. van Broeckhoven(2), C.M. van Duijn(1)
(1) Genetic Epidemiology Unit, Dept. of Epi & Biostat, Erasmus MC, Rotterdam, The Netherlands; (2) Molecular Genetics Department, University of Antwerp, Antwerpen, Belgium

Background: We conducted a meta-analysis to re-evaluate the role of the dopamine D4 receptor gene 48-base-pair-repeat (DRD4 48 bp-repeat) polymorphism in mood disorders.
Methods: DRD4 48 bp-repeat allele frequencies were compared between 917 patients with unipolar (UP) or bipolar affective disorder (BP), and 1164 controls from twelve samples, using the Cochrane Review Manager.
Results: A significant association was found for the DRD4 two-repeat allele in three groups: UP (OR 1.73, 95%CI 1.29–2.32, $P<0.01$), BP (OR 1.26, 95%CI 1.01–1.57, $P=0.04$), and UP and BP combined (OR 1.41, 95%CI 1.18–1.68, $P<0.01$). There was no evidence for heterogeneity or publication bias.
Conclusions: This meta-analysis showed evidence for a positive association of the DRD4 two-repeat allele with unipolar and bipolar mood disorders. These results show that a meta-analysis can overcome the low power of small sample size studies, leading to a more objective and quantitative summary appraisal of the evidence.

88

Identifying Endometriosis Susceptibility Genes in Puerto Rico

D.M. Mandal(1), I. Flores(2), J.E. Bailey-Wilson(3)
(1) Dept. of Genetics, LSU Health Sciences Center, New Orleans, LA; (2) Dept. of Microbiology, Ponce School of Medicine, Ponce, PR; (3) Inherited Disease Research Branch, NHGRI/NIH, MD

The causes of endometriosis are unknown although the evidence of genetic susceptibility is extensive. The main goal of the present study is to provide the genetic basis of endometriosis in the population in Puerto Rico.
Thirty-two families with two or more affected members were recruited from the island of Puerto Rico. Genotyping data are obtained on 18 markers in chromosome 10 candidate regions. The data have been checked for Mendelian inconsistencies using the program GCONVERT/SIB-PAIR. Significant allelic association was observed at the empiric p-value of 0.0038 at one of the candidate regions. The mode of inheritance of endometriosis is not known. Therefore, the LOD score was calculated assuming two simple models, dominant and recessive, each with an arbitrary 50% penetrance in gene-carriers and 5% penetrance in non-gene carriers. We used Genehunter-Plus to compute multipoint parametric LOD scores using these two models, as well as multipoint non-parametric NPL statistics and the Kong and Cox allele-sharing LOD score. Affected only analysis has been performed by coding the unaffected women as unknown in the analysis. Assuming an autosomal dominant mode of inheritance, the highest LOD score obtained was 1.75 and under the autosomal recessive model, the highest LOD score obtained was 1.26. We will be extending these analyses by using models that incorporate age through multiple liability classes.

89

Genome-wide scan linkage analysis for Parkinson's disease: The European Genetic Study of PD

M. Martinez(1), A. Brice(2), J. Vaughan(3), A. Zimprich(4), M. Breteler(5), G. Meco(6), A. Filla(7), M.J. Farrer(8), C. Bétard(9), J. Hardy(10), G. De Michele(7), V. Bonifati(6), B. Oostra(5), T. Gasser(4), N.W. Wood (3) & A. Dürr(2).
(1) INSERM EMI00-06, France; (2) U289, France; (3) U. College, UK; (4) U. of Tübingen, Germany; (5) Erasmus U., The Netherlands; (6) La Sapienza U., Italy; (7) Federico II U., Italy; (8) Mayo Clinic, USA; (9) CNG, France; (10) NIA, USA.

We undertook a genome-wide linkage study in a total of 227 affected sib pairs from 199 pedigrees with Parkinson's disease (PD). Individuals were genotyped for 391 microsatellite markers at ~10 cM intervals throughout the genome. Multipoint model-free affected sib-pair linkage analyses were carried out using the MLS lod score test. Four new chromosomal regions were identified as positively linked to PD (pointwise significance values $P<0.026$), on chromosomes 11q, 7p, 6p and 19q. Interestingly, our chromosome-19 peak maps at the position of apolipoprotein E, which has recently been suggested to be associated to PD. Two other positive findings, on chromosomes 2p11-q12 and 5q23, show evidence consistent with excess allele sharing in other datasets. The peak on 2p11-q12 (MLS=2.04) falls within a relatively short distance ($\sim$17cM) from the PARK3 region. Although, we observed a stronger support of linkage to this region in the late age at onset subgroup of familiess, these differences were not statistically significant. The peak on 5q23 (MLS=1.05, 130cM at D5S471) encompasses the location of the Synphilin-1 gene, a protein shown to interact in vivo with a-synuclein and parkin, and a component of Lewy bodies, the pathological hallmark of PD.

90

Extension of the Regression of Offspring on Mid-Parent (ROMP) approach to multiple SNP loci

R.M. Mathias(1), M.H. Roy-Gagnon(1,2), A.F. Wilson(1)
(1) IDRB/NHGRI/NIH, Baltimore, USA; (2) Johns Hopkins Bloomberg School of Hygiene and Public Health, Baltimore, USA

The Regression of Offspring on Mid-Parent (ROMP) method is an extension of traditional linear regression of offspring on mid-parent used to estimate trait heritability and heritability attributable to a candidate locus. The change in the regression slope of the mid-parent value when a locus is added to the model is the locus-specific heritability. Previous work has shown that for a bi-allelic quantitative trait locus (QTL)(or a single SNP) ROMP is highly powerful and Type I error is nominal.
Here, we extend ROMP to the multi-allelic case as in the situation of multiple SNPs in complete linkage disequilibrium with the QTL. We considered the cases of 2 and 3 SNPs yielding 4 and 8 equifrequent haplotypes, respectively. Two methods of analysis were used: 1) in a single regression, a coefficient for each haplotype was included (coding 0/1/2 for no copies/one copy/two copies for each allele) and 2) where separate regressions were performed for each allele.
Simulations indicate that ROMP implemented in a multi-allelic scenario has ample power where effects of the alleles at the QTL are additive. In the 2-SNP-case, power was well over 90% where the effects of the alleles was additive, but lower (48%-92%) where the effects were dominant. Results were similar in the 3-SNP-case, although on average power was lower. Power increased as the heritability attributable to the QTL increased. Type I errors were nominal, and both methods of analysis were equally efficient at detecting the association.

91

Empirical P-values for MOD score analysis - an extended version of GENEHUNTER-MODSCORE

M. Mattheisen, K. Strauch
Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany

We have extended GENEHUNTER-MODSCORE for simulations under the null hypothesis of no linkage. When assessing results of MOD score analyses, different criteria have to be supposed than for LOD scores that were calculated under one single model. In order to find the appropriate thresholds, this version of GENEHUNTER-MODSCORE allows the user to perform Monte-Carlo simulations to calculate the P-values for MOD score results. In addition to the P-value corresponding to the MOD score for the real dataset, the P-values are reported for a grid of MOD scores between 1.0 and 5.0. It is also possible to evaluate only this 'P-value grid' for a given pedigree structure and marker map, without taking the actual marker configuration of the real dataset into account. Furthermore the program allows for calculation of P-values in the context of 'simple' LOD score results. Under the assumption of Hardy-Weinberg equilibrium, for each replicate the marker alleles are simulated with respect to the given marker allele frequencies and recombination values. Specification of the fraction of available genotypes is possible, and a sequential sampling method is implemented for the purpose of computational efficiency. Anyhow, such an approach can be computationally very intensive, especially for datasets which include larger pedigrees. However, in case that a simulation with a large number of replicates is infeasible, it is still worthwhile to simulate a smaller amount of replicates, in order to obtain at least a crude estimate of the real P-value.

92

Association Between Haplotypes In The Cholesteryl Ester Transfer Protein (CETP) Gene and High Density Lipoprotein (HDL) Levels: Results from Three Separate Populations

McCaskie P.A.(1,2), Beilby J.P.(3), Chapman C.M.L.(3), McQuillan B.M.(4,5), Thompson P.L.(4,5), Hung J.(4,5), Palmer L.J.(1,2).
(1) Western Australian Institute for Medical Research SCGH Campus, Centre for Medical Research, University of Western Australia, Nedlands, Western Australia; (2) School of Population Health, University of Western Australia, Nedlands, Western Australia; (3) Clinical Biochemistry, PathCentre, Perth, Western Australia; (4) Sir Charles Gairdner Hospital Campus of the Heart Research Institute of Western Australia, Perth, Western Australia; (5) Department of Medicine, University of Western Australia, Nedlands, Western Australia.

Cholesterol ester transfer protein (CETP) plays a major role in lipid metabolism. Three single nucleotide polymorphisms in the CETP gene are known to have functional effects on plasma CETP activity and HDL. It has been suggested by various studies that increased HDL levels are protective against cardiovascular disease (CVD). We hypothesised that these polymorphisms would be associated with variation in cardiovascular disease-associated phenotypes, in particular serum HDL levels. Multivariate haplotypic analysis revealed a common haplotype, TaqIB_B2/-2708_A/-629_A, that was strongly associated with increased HDL levels in three independent populations, independent of known risk factors of CVD such as age, sex, systolic blood pressure and pack years of cigarette smoking. This haplotype was also associated with a decreased risk of CVD in a case-control study. SNP analysis suggested that these polymorphisms were also independently associated with increased serum HDL.

93

Association mapping in an isolated founder population

M.S. McPeek
University of Chicago, Chicago, IL, USA

It has been suggested that founder populations may have certain advantageous properties for complex trait mapping, including (i) that the small number of founders would tend to reduce the genetic heterogeneity in the population and (ii) that linkage disequilibrium would exist over greater distances than in outbred populations, so that a less dense marker map would be required for successful detection of association than would be needed in outbred populations. Founder populations also present special challenges, due to the fact that (i) randomly sampled

individuals may be detectably related and (ii) the pedigree can be extremely large and complex with many inbreeding loops. In association mapping, the relatedness of individuals within the founder population can play a similar role to that of population substructure in more heterogeneous populations. In particular, failure to account for the relatedness of individuals sampled from the population has a tendency to result in false positive detection of association. We are developing quasi-likelihood based methods of inference for association mapping in an isolated founder population, with application to the Hutterites, an isolated founder population in the northern U.S. and western Canada, in which the approximately 800 sampled individuals are virtually all related through multiple lines of descent, with the relationships characterized by a known, 13-generation, 1,623-person pedigree with only 64 founders. I plan to discuss the meanings of the relevant parameters and hypotheses in the founder population context, propose quasi-likelihood-based methods of inference, treat the computational issues, and discuss applications of the methods to genetic studies in the Hutterites. This is joint work with Carole Ober, Catherine Bourgain, Xiaodong Wu, Mark Abney, Tim Thornton, and others.

94

A comparison study of three likelihood haplotype-phenotype association methods applied on a case-control of lung cancer

S. Michiels(1), C. Amand(1), S. Benhamou(2)
(1) Unit of Biostatistics and Epidemiology, Institut Gustave-Roussy, Villejuif, France; (2) INSERM and Evry University EMI 00-06, Evry, France.

Recently, various methods for testing the association between single-nucleotide polymorphism (SNP)-based haplotypes and disease risk have been proposed using case-control data with unknown linkage phase. We aimed to compare the results of three published prospective likelihood-based methods on several genotyped DNA repair genes among 151 lung cancer cases and 171 controls. The retained methods estimate simultaneously haplotype frequencies and haplotype-phenotype effects while allowing adjustment for environmental covariates: Stochastic EM (SEM, Hum Mol Genet 2003; 11:2015–23), Generalised Estimating Equations (GEE, J Hum Genet 2003; 72:1231–50) and Score Test (ST, Am J Hum Genet 2002; 70:425–34). All three methods implement a logistic regression model and assume linkage disequilibrium (LD) between SNPs in the pooled sample of controls and cases. SEM and ST assume HW equilibrium (HWE) in the pooled sample, GEE only in the controls. Both ST and SEM provide a multi-haplotype global test. We selected three genes according to different settings for the hypotheses of HWE and LD. Analyses were adjusted for smoking and age.
Divergent p-values for haplotype effects on lung cancer risk were found for the genes for which the hypotheses of HWE and/or LD did not hold in the controls or in the pooled sample. In conclusion, these differences suggest that the selected methods are very sensitive to the violation of the hypotheses of HWE and/or LD.

95

Entropy and Sequential Floating Forward Selection for selecting functional SNPs in genomewide search

Valentin Milanov, Radoslav Nickolov
Department of Mathematical Sciences, Michigan Technological University, USA

A challenging problem in human genetics is the identification and characterization of susceptibility genes for complex human diseases among thousands of candidate genes genomewide. In particularly gene-gene interaction. The known methods, parametric and non-parametric, are usually limited to small number of functional loci. We utilize suboptimal feature selection method and entropy based criterion for case-only study. Using case and case-control simulated data we demonstrated that the method has a high discovery rate under few simulation models with pure interaction without main effects of the genes. The simulated markers are independent single nucleotide polymorphisms (SNPs).

96

Linkage Conditional on Measured Genotypes Under Conditions of Low Statistical Power

M.B. Miller
Division of Epidemiology, University of Minnesota, USA

Almasy and Blangero (Behavior Genetics, March 2004) demonstrated that when there is a single QTL in a chromosomal region and a significant LOD-score peak ($LOD > 3.0$) is observed in that region, use of the QTL as a covariate in the linkage analysis usually reduces the peak to a non-significant level ($LOD < 0.588$) at the QTL. Thus, one may test a candidate gene as a covariate in quantitative linkage analysis to see if it accounts for an observed peak. The simulation undertaken by Almasy and Blangero was highly powered with a QTL accounting for 28% of the trait variance, a mean LOD-score of 5.57 and 87% of replicates achieving $LOD > 3.0$. It is more realistic to simulate data with a QTL that accounts for much less variance (maybe 10%) and produces much lower statistical power. When we analyze sibling-pair data simulated under such a model, we find that using the QTL as a covariate will rarely reduce a statistically significant peak ($LOD > 3.0$) to nonsignificance ($LOD < 0.588$). We also show through simulation that this finding is due to the fact that much of the contribution to the LOD-score peak comes from lucky associations of extra-QTL genetic effects and random environmental influences with identity-by-descent patterns. Such effects remain when the QTL is used as a covariate. Our findings imply that the method recommended by Almasy and Blangero works differently, and is less definitive, when power is low to detect the effect of the QTL. Alternative methods and interpretations are recommended.

97

Identifying Susceptibility Genes by Accounting for Epistasis in Case-Control Studies of Candidate Genes

Joshua Millstein, David V. Conti, W. James Gauderman
Dept. of Preventive Medicine, Univ. of Southern California

We develop a method applicable to a group of candidate genes that accounts for possible epistasis. Likelihood ratio tests are performed in stages that are determined by the highest order interaction term in each model. In the first stage, the main effect of each candidate gene is tested independently. In the second stage, models with 2-gene interaction terms and their component main effects are tested conditional on the main effects of genes with significant effects in the first stage. Subsequent stages are performed in a similar manner. A reduction in the number of tests performed is achieved by prescreening gene combinations with a goodness-of-fit chi square statistic that depends on association among candidate genes in the pooled case-control group. This statistic is calculated irrespective of disease status, thus the overall Type I error rate is preserved. Methods were compared using 200 replicate simulations of 20 candidate genes from 200 cases and 200 controls with a single 3-gene interaction simulated to increase disease risk. Independent tests of main effects identified only 117 (19.5%) of 600 susceptibility genes across replicates, while our proposed approach identified 274 (45.7%) of the susceptibility genes. Further simulations demonstrated that Type I error rate was controlled. Our method, which can be implemented using standard software, provides a natural approach for using epistasis to increase power in the detection of candidate gene associations.

98

The Relative Efficiency of Penetrance Estimators Using Sib-Pairs

C.G. Minard(1,2), M.D. Swartz(2,3,4), C.I. Amos(1)
(1) Univ of Texas M. D. Anderson Cancer Center, Houston, TX; (2) Univ of Texas Health Science Center School of Public Health, Houston, TX; (3) Rice Univ, Houston, TX; (4) Texas A&M Univ, College Station, TX

We evaluated three different methods of estimating genetic penetrance for sib-pairs: maximum likelihood estimation with both trait and marker information, maximum likelihood estimation with trait information alone, and the MOD score approach. The asymptotic relative efficiency between estimators is used to evaluate penetrance estimation under random sampling and single ascertainment for an autosomal disease in sib-pairs modeled with unknown phase. Joint segregation and linkage nearly always yielded the smallest variance and was used as the standard for calculating relative efficiencies. Under random sampling, the relative efficiencies varied markedly with penetrance and allele frequency for a recessive disorder. However, the relative efficiency of the MOD score approach generally increased with increasing allele frequency. For a dominant disease, segregation analysis alone was generally more efficient than the MOD score approach with respect to joint segregation and linkage. Under single ascertainment, the penetrance estimator for segregation analysis alone was not estimatable. The MOD score approach, however, was fully efficient for a dominant disease with lowly prevalent alleles while the efficiency varied with allele frequency and penetrance for recessive disorders. The MOD score approach yielded relatively more efficient estimates of penetrance for ascertained data compared to its behavior in random sampling.

99

Relation between TAFI gene polymorphisms, the risk of coronary heart disease and plasma TAFI levels in the PRIME Study : use of a new ELISA independent of the Thr325Ile polymorphism

P.E. Morange(1–2), D.A. Tregouet(2), C. Frere(1), D. Arveiler(3), J. Ferrieres(4), P. Amouyel(5), A. Evans(6), P. Ducimetiere(7), F. Cambien(2), on behalf of the PRIME Study group
(1) Hematology Lab., CHU Timone, Marseille - (2) INSERM U525, Faculté de Médecine Pitié-Salpétrière, Paris, France - (3) Dept. of Epidemiology and Public Health, Faculty of Medicine, Strasbourg, France - (4) INSERM U558, Dept. of Epidemiology, Paul Sabatier-Toulouse Purpan University, Toulouse, France - (5) INSERM U508, Institut Pasteur de Lille, France - (6) Dept. of Epidemiology and Public Health, Queen's University of Belfast, Northern Ireland - (7) INSERM U258, Hopital Paul Brousse, Villejuif, France.

Several recent works have shown that plasma levels of thrombin activatable fibrinolysis inhibitor (TAFI), suspected to play a role in atherothrombosis, were highly determined by TAFI gene polymorphisms. However, it has been demonstrated that, in commercially available ELISA, variable antibody reactivity towards TAFI isoforms led to erroneous TAFI levels, reflecting qualitative rather than quantitative variations. The aim of the study was to evaluate the association of TAFI gene polymorphisms with the risk of coronary heart disease (CHD) and with TAFI levels measured by a newly developed 1B1/P4C2 ELISA shown to be a trustworthy method to detect quantitative variations of circulating TAFI. 6 polymorphisms (C-2599G, G-408A, Ala147Thr, Thre325Ile, C+1542G and T+1583A) were genotyped and baseline plasma concentrations of TAFI were measured in a nested case-control study as part of the prospective PRIME Study. 249 participants from France and Northern Ireland who had developed a CHD event during a 5-year follow-up were compared with 492 population- and age-matched control subjects. In France, the Thr147 allele was more frequent in cases than in controls (0.41 vs 0.32; p=0.02) while the inverse relationship was observed in Northern Ireland (0.38 vs 0.33; p=0.19) (p=0.01 for interaction). No other polymorphism was associated with CHD risk Conversely, single-locus and haplotype analyses revealed that 2 polymorphisms, C-2599G and Ala147Thr (or T+1583A that is in nearly complete association) had additive effects on TAFI levels

and explained more than 18% of TAFI variability. This effect was homogeneous in

100

Design of case-controls studies with unscreened controls

V. MOSKVINA(1), P. HOLMANS(1), N. CRADDOCK(2)
(1) Bioinf. & Biostat. Unit; (2) Dept. of Psych. Med., Univ. of Wales College of Medicine, UK

Traditionally in genetic case-control studies controls have been screened to exclude subjects with a history of illness. This control group has the advantage of optimal power to detect loci involved in illness however requires more work and may incur substantial cost in recruitment. An alternative approach to screening is to use unscreened controls sampled from the general population. Such controls are generally plentiful but there is a risk that some may have the same disease as the cases, which will also reduce power to detect associations. The extent of this power loss will depend on the prevalence of the trait in the population, being largest when the prevalence is high. In many situations the power may, be recovered by typing a larger number of the unscreened controls. We have quantified the extent of this power loss, and produced mathematical formulae for the number of unscreened controls necessary to achieve the same power as a fixed sample of screened controls. We have also investigated the cost-benefits of the screened and unscreened approaches according to variation in the relative costs of sampling screened and unscreened controls, together with genotyping costs. We, thus, identify the range of situations in which using unscreened controls is a cost-effective alternative to the screened control method and could be considered when designing a study. Specifically, unscreened controls can be considered for prevalence of disease smaller than 0.2, OR$>$2, and when the cost of genotyping controls in a study is less than 5 times the cost of screening the controls.

101

Comparative analysis of linkage disequilibrium patterns and tagSNP transferability among European populations and CEPH trios

J.C. Mueller(1), E. Lõhmussaar(1,2), R. Mägi(2), M. Remm(2), T. Bettecken(1), P. Lichtner(1), S. Huber(1), A. Pfeufer(1), S. Schreiber(3), T. Illig(4), J. Luedemann(5), P. Pramstaller(6), G. Romeo(7), A. Testa(8), H.E. Wichmann(4), A. Metspalu(2), T. Meitinger(1)
(1) Inst Hum Genet, GSF, BRD; (2) Inst Mol Cell Biol, Univ Tartu, Estonia; (3) Univ Hosp, Kiel, BRD; (4) Inst Epidem, GSF, BRD; (5) Inst Clin Chem Lab Med, Univ Greifswald, BRD; (6) Eurac Res, Bolzano, Italy; (7) Med Genet, Univ Bologna, Italy; (8) Lab Nephrol, CNR, Reggio Calabria, Italy

Linkage disequilibrium (LD) is fundamentally important in association studies. The aim of this project is to compare the LD patterns in European populations with respect to gene mapping approaches. We selected several genomic regions, which contain candidate genes for complex traits. Evenly distributed markers with an average spacing of 2-6 kb were genotyped in 9 population samples including standard control collectives. The CEPH trios of the HapMap project are used as a reference population.
We assessed the similarity of block boundaries among population samples by a novel measure based on bootstrapping. Although main patterns are common to all European populations, small differences in block structure and block content appear between geographical regions. Transferability of LD information was tested by the tagSNP approach. Depending on the analysed gene, tagSNPs selected in the HapMap CEPH trios performed relatively well in other local European samples. We showed that more gain in tagSNP efficiency would be achieved by increasing the current SNP density in the HapMap project relative to defining tagSNPs in each local population.

102

On finding the optimal genotype coding when the mode of disease inheritance is unknown: A new family-based association test with unspecified mode of inheritance

A.J. Murphy(1), C. Lange(1,2)
(1) Dept. of Biostatistics, Harvard School of Public Health, USA; (2) Harvard Medical School, USA

We propose a new family-based association test (FBAT) which does not require the mode of inheritance to be specified. The test can be applied in situations in which the mode of inheritance is unknown. Using either the conditional mean (Lange et al., 2003a,b) or non-parametric density estimation, the optimal coding for the heterozygous genotype is estimated on a continuous scale without biasing the significance level. We discuss estimation strategies for continuous and dichotomous traits. We show the practical relevance of the new test statistic, using simulation studies and an application to an Asthma study. The new approach has been implemented in the PBAT-software package and is available at http://www.biostat.harvard.edu/$\sim$clange/default.htm.

References:

Lange C, Lyon, H, DeMeo D, Raby B, Silverman EK. (2003a). A new powerful non-parametric two-stage approach for testing multiple phenotypes in family-based association studies, Human Heredity, 56:10–17.

Lange C, Silverman EK, Xu X, Weiss, ST, Laird NM (2003b). A multivariate family-based association test using generalized estimating equations: FBAT-GEE. Biostatistics, 4:195–206.

103

Bonferroni Correction for Multiple Testing in Candidate Gene Studies Considering the Underlying Linkage Disequilibrium (LD) Structure

Nicodemus K.K., Kao W.H.L., Fallin M.D.
Johns Hopkins School of Public Health, Baltimore MD, 21205

As the number of SNPs examined increases, so does the type I error rate. Controlling the experiment-wide type I error rate while maintaining power is controversial. The gold-standard method to determine significance levels is permutation testing, but it is computationally intense. Another method is using a correction to the nominal p-value (Bonferroni correction (BC)). BC adjusting for each test may be too conservative, since SNPs in close proximity are often not independent. A modification of the BC that takes into account the underlying LD structure may be an appropriate and easily-implemented approach that does not require extensive computation. A suggestion for such a correction uses principal components (PC) to identify independent subsets of markers to determine appropriate type I error thresholds (Nyholt, 2004). This method does not consider the underlying LD structure. We propose using haplotype-blocking algorithms to define independent sets and use this as the factor to correct for multiple testing. We evaluated several methods for multiple-test correction through simulation studies. Two types of study designs were considered: unrelated case-control and case-parent trios. For each study design, three scenarios of the LD structure were generated: 1. the associated SNP measured and part of a haplotype block, 2. the associated SNP measured but not part of a haplotype block, 3. the associated SNP unmeasured but part of a haplotype block. Also, two levels of effect size for the associated SNPs were simulated, for a total of 6 data sets for each study design. Power and type I error rates were assessed using nominal significance levels, BC by the number of individual tests, BC by PC analysis, BC by using the number of haplotype blocks, and significance levels obtained by permutation methods.

104

Heritability of iron and ferritin in a genetic isolate

O.T. Njajou(1), B.Z. Alizadeh(1), B.A. Oostra(2), D. Swinkels(3), C.M. van Duijn(1)
(1) Departments of Epidemiology & Biostatistics and (2) Clinical Genetics, Erasmus MC Rotterdam; (3) Department of Clinical Chemistry, UMC Nijmegen, the Netherlands

Iron plays a crucial role in the pathogenesis of complex disorders such as Parkinson's disease, Alzheimer's disease, atherosclererosis and cancer. Both iron deficiency and iron overload are common public health problems. From a genetic perspective, iron metabolism is a complex trait, in which both genetic and environmental factors are involved. The purpose of the present study was to estimate the magnitude of genetic influences on iron and ferritin levels in relatives from a genetic isolate in the Netherlands. Estimation of how much of the variation in the levels of iron and ferritin could be attributed to genetic factors was done using the variance component method in sequential oligo-genic linkage analysis routines (SOLAR). The participants analysed in this study included 90 nuclear families with a total of 908 subjects. The proportion of the residual phenotypic variance due to additive genetic effects or heritability estimates were approximately 21.34 (SE=0.064, $P<0.00001$) for iron and 38, 62 (SE=0.066, $P<0.00001$) for ferritin. After simultaneously adjusting for sex and age, the heritability was 23.15 (SE=0.064, $P<0.00001$) for iron and 39.20 (SE=0.066, $P<0.00001$). A substantial proportion of the variance of iron and ferritin can be explained by heredity, independent of sex and age. These results demonstrate the influence of both genetic and environmental factors on iron levels. Identification of genes influencing iron and ferritin levels using a QTL approach is feasible.

105

Application of logistic regression to case-control association studies involving two causative loci

Bernard V North(1), David Curtis(1), Pak C Sham(2)
(1) Academic Department of Psychiatry, Queen Mary's School of Medicine and Dentistry, London E1 1BB, UK; (2) Department of Psychological Medicine, Institute of Psychiatry, De Crespigny Park, London SE5 8AF, UK

Models in which two susceptibility loci jointly influence the risk of developing disease can be explored using logistic regression analysis. Comparison of likelihoods of models incorporating different sets of disease model parameters allows inferences to be drawn regarding the nature of the joint effect of the loci. We have simulated case-control samples generated assuming different two-locus models and then analysed them using logistic regression. We show that this method is practicable and that, for the models we have used, it can be expected to allow useful inferences to be drawn from sample sizes consisting of hundreds of subjects. Interactions between loci can be explored, but interactive effects do not exactly correspond with classical definitions of epistasis. We have particularly examined the issue of the extent to which it is helpful to utilise information from a previously identified locus when investigating a second, unknown locus. We show that for some models conditional analysis can have substantially greater power while for others unconditional analysis can be more powerful. Hence we conclude that in general both conditional and unconditional analyses should be performed when searching for additional loci.

106

Linkage Analysis of Factors Underlying the Insulin Resistance Syndrome

K.E. North(1), L. Almasy, H.H. Göring, S. Cole, V.P. Diego, S. Laston, T. Cantu(2), B.V. Howard(3), E.T. Lee(4), L.G. Best(5), R. Fabsitz(6), J.W. MacCluer(2)
(1) UNC, Chapel Hill, NC; (2) SFBR, San Antonio, TX; (3) MedStar Res, Washington, DC; (4) Univ of OK, OK City, OK; (5) Missouri Breaks, Timber Lake, SD; (6) NHLBI, Bethesda, MD

In previous work in non-diabetics of the Strong Heart Family Study (SHFS), principal components analyses of nine insulin resistance syndrome(IRS)phenotypes identified three factors, namely glucose/insulin/obesity, blood pressure, and dyslipidemia factors, and these factors were

found to be heritable. To localize QTLs potentially influencing these IRS factors, we conducted a genome scan of factor scores in 597 SHFS participants. Using the variance components approach implemented in SOLAR, with age, sex, and study center as covariates, we detected linkage of the glucose/insulin/obesity factor to chromosome 4 (near marker D4S415 at 4q34.3, LOD=2.3). This signal overlaps positive findings reported for BMI and fat, and harbors the NPY5R gene, which has been associated with obesity. We identified linkage of the dyslipidemia factor to chromosome 12 (flanking markers D12S79 and D12S86, 12q24.1-24.23, LOD=2.7). This signal overlaps with positive findings for HDL-C and triglycerides, and overlies the SRB1 gene, which has been associated with dyslipidemia and obesity in several populations. We also found linkage of the blood pressure factor to chromosome 1 (flanking markers D1S413 and D1S249, 1q31.3-32.1, LOD=1.5), which supports several positive findings for blood pressure. The corroboration of existing QTLs will bring us closer to the identification of the functional genes that predispose to IRS.

107

The Vulnerability Score Algorithm: Prediction of illness using the NIMH Bipolar Dataset

J.I. Nurnberger, Jr. and Eric T. Meyer
Indiana University School of Medicine

We have previously described an algorithm which assigns a quantitative index of genetic vulnerability to each member of a family participating in a genomic survey. The algorithm involves the calculation of a Sharing Score for each allele at each locus, based on the proportion of affecteds in the family sharing that allele. Each individual is then assigned the value of his/her two alleles and this is weighted by the Genehunter NPL score for that locus. The weighted values for a set of loci are then summed. We have designed a dataset for testing the algorithm, using a set of Truncated Pedigrees from the NIMH Bipolar Dataset. The Truncated Pedigrees are a subset in which the youngest generation of subjects are removed from the pedigree and used as test subjects.
Linkage was calculated for the whole dataset under narrow and wide diagnostic models. Sets of peak loci were identified from the Genehunter output. Two versions of the algorithm were employed. The first version uses the top loci as selected by the NPL score for the group. The second version uses the top loci selected for each individual. The empirical lifetime morbid risk of illness in relatives in these data is 17% (narrow) 44% (wide). Using version 1 of the algorithm, the model predicts 26% and 61% illness respectively. Using version 2, the model predicts 37% and 54%. Prediction of current affected status is more accurate with version 1 (63% accuracy for both definitions) than with version 2 (42% and 61%). The results support the continued development of the Vulnerability Score method with the use of the overall NPL as a means of selecting loci.

108

Ordered Subset Analysis in Parkinson disease families identifies gene-by-gene interaction between chromosomes 10q and Xq

N. Pankratz(1), W.C. Nichols(2), S.K. Uniacke(2), C.A. Halter(1), A. Rudolph(3), C.W. Shults(4), P.M. Conneally(1), T. Foroud(1), and the Parkinson Study Group
(1) Indiana University; (2) Cincinnati Children's Hospital; (3) University of Rochester; (4) University of California, San Diego and the VA San Diego Healthcare System

Most forms of Parkinson disease (PD) are likely due to complex etiology, making it more challenging to identify underlying susceptibility genes. It is assumed that gene-by-gene interactions are quite common in complex diseases, and one possible method to detect such interactions is to condition on a known locus using Ordered Subset Analysis (OSA). OSA ranks families using a specified covariate and incrementally adds families to the analysis in their ranked order. It is then possible to identify the subset of the data that maximizes the evidence of linkage at another locus. We have analyzed a sample of 362 multiplex PD families. OSA was performed where the covariate used was the family specific non-parametric linkage (NPL) score for a region on chromosome 10q that had previously been linked to PD susceptibility in our sample. Using the 98 families with the highest NPL scores for chromosome 10q ($NPL > 0.70$), the LOD score on chromosome X increased from 1.68, observed in the full sample, to 4.51 in the subsample ($p < 0.01$). This suggests that the putative gene on the X chromosome, which has also been replicated in several independent PD linkage samples, may have a significant interaction with the gene on chromosome 10q. OSA may prove useful in identifying other gene-by-gene interactions in the future.

109

Individual-Specific Family Risk Scores in Extended Pedigrees

V.S. Pankratz
Division of Biostatistics, Mayo Clinic, USA

Understanding an individuals familial risk for a disease can provide important insight into disease etiology and prognosis. For diseases with a known genetic etiology, quantification of this risk can be achieved with a single genetic test. For complex diseases, quantifying familial risk is more difficult. Even so, understanding familial risk can provide valuable information. It can aid individuals in prevention, and it can inform our understanding of disease mechanisms. Current methods of measuring familial risk include counting the numbers of relatives with the disease, or estimating the excess number of cases of disease in the family relative to population expectations. While these methods have proven to be useful, their general applicability is limited. Among the limitations are sensitivity to family size and structure, difficulty in incorporating age at onset information and difficulty in accounting for shared genetic vs. shared environmental

risk. When extensive pedigree and disease data are available, it should be possible to obtain a measure of familial, and genetic, risk that is less sensitive to these limitations. We propose to use Cox proportional hazards regression with random effects to estimate individual-specific familial risks. These estimates are achieved by analyzing age at onset data within large pedigrees while accounting for additive polygenic and shared familial random effects. The performance of this technique will be evaluated, and compared to alterative measures of familial risk, using simulated data and data from the Minnesota Breast Cancer Family Study.
Research supported in part by NIH P01 CA82267.

110

Linkage of biochemical and molecular traits related to obesity in the Old-Order Amish

G.J. Papanicolaou(1), P. Platte(2), E.W. Pugh(3), K. Doheny(3), K.M. Pirke(4), M.-H. Roy-Gagnon(1), A.J. Stunkard(5), C. Francomano(6), A.F. Wilson(1)
(1) Genometrics Sect., IDRB, NHGRI, NIH, USA; (2) Biol & Clin Psych, U. Wuerzburg, DEU; (3) CIDR, JHU SOM, USA; (4) Ctr Psychobiol.Psychosom. Res, U. Trier, DEU; (5) Dept. Psychiatry, U. PA, USA; (6) NIA, NIH, USA

The gradual increase in overweight in the U.S. has increased significantly in the past decade; today nearly two-thirds of the population is overweight and over thirty percent are obese. As part of an ongoing study of traits related to obesity in the Old-Order Amish, seven two- and three- generation families totaling 157 individuals (with a mean sibship of 7.2) were assessed for 21 obesity related traits, including diastolic and systolic blood pressure, and levels of guanine nucleotide-binding protein, beta-3 (GNB3), glucose, insulin, leptin, triiodothyronine (t3), thyroxin (t4), thyroid stimulating hormone (tsh), and uncoupling protein 3 (UCP3). Although families were ascertained on a surrogate measure of obesity, the families were normotensive and did not exhibit signs of diabetes. Genotyping was performed at the Center for Inherited Disease Research (Baltimore, MD) with a modified Marshfield Genetics 8 marker set consisting of 384 short tandem repeat markers with an average distance of 9 cM. Model-independent linkage analysis identified candidate regions for diastolic b.p., glucose, insulin, leptin, t3, t4, and tsh. Traits GNB3 and UCP3 were mapped to respective known locations, providing both an internal control for the analysis as well as an estimation of Type I error.

111

The effect of genetic drift in a young genetically isolated Dutch population

L.M. Pardo(1), C.M. van Duijn(1), Y.S. Aulchenko(1,2)
(1) Genetic Epidemiology Unit, Department of Epidemiology & Biostatistics, Erasmus Medical Center, Rotterdam, The Netherlands; (2) Institute of Cytology and Genetics SD RAS, Novosibirsk, Russia

Genetic drift determines the fluctuation in genetic pool of populations. Isolated populations might genetically deviate from a general population as a result of drift. We assessed the extent of this deviation in a recently genetically isolated population in the southwest of the Netherlands studied as part of the Genetic Research in Isolated Population (GRIP) program. A gene-dropping experiment was performed in a large pedigree from this isolate assuming different founder frequencies. Allelic frequencies in the last generations of this pedigree were estimated. Simulation analysis showed that large fluctuations should be expected for founder frequencies lower than or equal to 1%. At founder frequencies larger than 1%, the fluctuations were small. To compare these findings with empirical data, we analyzed mean heterozygosity and allele diversity of 592 markers in a random sample from this GRIP population and contrasted with a general population (CEPH sample), old large isolate (Icelandic sample) and small-sized population of Talana (Sardinia). GRIP mean heterozygosity and mean number of alleles was lower as compared with CEPH and deCODE families, but higher when compared with Talana population. The average heterozygosity in GRIP population, although significantly reduced (P-value <0.01), does not deviate much from that in general population (CEPH families). This is in agreement with our simulation findings indicating that frequencies of common alleles were not changed much in our isolate. The lower allele diversity may indicate that some rare alleles are lost due to drift. To conclude, the genetic diversity in GRIP is smaller than that in outbred (CEPH) and on isolated (Iceland) populations, and higher then in small populations subject to large drift (Talana).

112

APOE 2 has protective effect on early mortality in normal weight individuals.

M.C. Pardo Silva(1), A.C.J.W. Janssens(2), C. van Broeckhoven(3), A. Hofman(1), J.C.M. Witteman(1) and C.M. van Duijn (1)
(1) Dept of Epi & Biostat, Erasmus MC Rotterdam, the Netherlands; (2) Dept of Public Health, Erasmus MC; and (3) VIB, UIA, Antwerpen, Belgium

The aim of the present study was to investigate the effect of APOE genotype on mortality in interaction with body mass index (BMI), and to determine to what extent this association is driven by the occurrence of coronary heart disease (CHD). We compared survival probabilities and mortality hazard ratios (HR) between APOE genotypes using data from the Rotterdam Study (n=6849). Analyses were stratified by BMI, considering normal weight, overweight and obese subjects. APOE2 carriers showed a better survival from all-cause mortality at early ages (before age 80; p=0.03). This improved survival was found only among normal weight individuals HR=0.61 (95% Confidence Interval [CI] 0.40 to 0.94). After adjustment for gender, smoking status, hypertension, hypercholesterolemia, high-density lipoproteins and diabetes mellitus, the HR for early mortality among APOE2 carriers with normal weight was 0.60 (95% CI 0.38 to 0.94). The HR was higher for early CHD-related

mortality (adjusted HR=0.22, 95% CI 0.05 to 0.92), than for early mortality due to other causes (adjusted HR=0.72, 95% CI 0.44 to 1.18). We conclude that APOE2 confers a genetic advantage for early mortality in normal weight individuals, which is mainly explained by the protective effect of APOE2 from early CHD-related mortality. In overweight and obese subjects, this protective effect is overcome by the adverse effect of increased BMI on health.

113

Gene Variants in Myocardial Potassium Channel Genes KCNQ1 and KCNH2 act Additively to Modify the QT-Interval

Pfeufer, Arne(1), Müller, Jakob(1), Perz, Siegfried(3), Jalilzadeh, Shapour(1), Hinterseer, Martin(2), Goppel, Gertrud(2), Illig, Thomas(3), Löwel, Hannelore(3), Näbauer, Michael(2), Steinbeck, Gerhard(2), Kääb, Stefan(2), Wichmann, H.-Erich(3), Meitinger, Thomas(1)
(1) Institut für Humangenetik, TUM, München, D; (2) Medizinische Klinik I, LMU Klinikum Großhadern, München, D; (3) GSF Forschungszentrum, Neuherberg, D

AIM: Cardiac arrythmias are frequently caused by delayed repolarization under control of voltage gated potassium channels. Extreme phenotypes are the monogenic long QT Syndromes (LQT1-LQT6). We undertook LD mapping of the genes KCNQ1 (LQT1, 11p15.5, 400kb, alpha-subunit of myocardial Iks current) to which linkage of the QT interval has previously been reported (Busjahn, 1999) and KCNH2 (LQT2, 7q36, 35kb, alpha-subunit of myocardial Ikr current) followed by association analysis of frequent gene variants with QT-interval in the general population. METHODS: In n=702 population based probands we recorded 10 second ECGs and calculated a QT-time corrected for age, sex and heart rate (QTCN). Probands were genotyped for 82 SNPs in and around KCNQ1 (550 kb region) and 107 SNPs in and around KCNH2 (250 kb region). LD-structure was characterized by determining LD measures D' and r2. Haplotypes in the individual blocks of LD were determined statistically. Probands with ventricular pacemakers, complete bundle branch blocks or brady- or tachyarrhythmia were excluded from association analysis. In the remaining probands (n=657) the corrected QTCN-interval was associated with genetic variants by linear regression over allele counts. RESULTS: 10 haplotype blocks existed in the genomic region of the KCNQ1 gene and 6 in the KCNH2 region. The strongest association to the QTCN-interval was between two SNP markers in a block in KCNQ1 Intron 1 (AFmin=0.37, dQTCN=+7,0ms, p=0,0072) and in a block in KCNH2 extending from Intron 2 to Exon 15 (AFmin=0.23, dQTCN=+7,2ms, p=0,0049) explaining 1.3% resp. 1.1% of the entire variance of QTCN. Both SNPs were in HWE, as expected in complete linkage equilibrium with each other and were independent with respect to their effects on QTCN (p=0.004), their combined effect explaining 2,4 % of its variance. CONCLUSION: Common gene variants in the genomic region of the KCNQ1 and KCNH2 genes independently and additively influence myocardial repolarization.

114

A full likelihood for sibships with variable age at onset ascertained from population based registry of incident cases: Association of HLA and Type 1 diabetes.

(1)J. Pitkäniemi, (3)J. Corander, (2)J. Partanen, (4)E. Tuomilehto-Wolf, (1)J. Tuomilehto, (3)E. Arjas
(1) Dept. of Public Health, School of Medicine, University of Helsinki; (2) FRC Blood Service, Department of Tissue Typing, Helsinki; (3) Dept. of Mathematics and Statistics, University of Helsinki; (4) Div. of Diabetes and Genetic Epidemiology, National Public Health Institute

In genetic studies of rare complex variable age at onset diseases it is common to ascertain sibships/pedigrees through population based incident case(s). Classical ascertainment correction would lead to substantial loss of data and therefore we propose a full likelihood approach. The full likelihood contains two parts: likelihood of all the ascertained data and likelihood of all subjects in the population that could have become probands but did not. Likelihood of the second part uses information on the population at risk but healthy (not ascertined and no Type 1 diabetes) available at population registry and genotype distribution obtained from the Finnish bone marrow registry (19,836 genotyped subjetcs). Likelihood contributions in the second part of the likelihood can be calculated in resonable time because, given our statistical model for the age at onset and the genotypes, these contributions are similar for all subjects with the same genotype and age. We have implemented this approach with WinBugs program and plan to analyze HLA A, B and DR associations with the risk of Type 1 diabetes.

115

Combining linkage scans from different samples: full genome scans for cognitive ability in Dutch and Australian samples

D. Posthuma(1), M. Luciano(2), E.J.C. de Geus(1), M.J. Wright(2), N.G. Martin(2), D.I. Boomsma(1)
(1) Vrije Universiteit, Amsterdam, The Netherlands; (2) Queensland Institute of Medical Research, Brisbane, Australia

Combining linkage scans using two or more independent samples is a powerful tool in finding evidence for linkage. However, both phenotypic and genotypic sample heterogeneity need to be treated carefully. Here we illustrate these issues using linkage scans for cognitive ability based on two different samples.
Psychometric IQ was assessed in Dutch and Australian extended twin families. The Dutch sample consisted of 793 subjects from 317 families, aged between 17 and 68 years. The Australian sample consisted of 1339 subjects from 603 families, aged between 15 and 22 years. Heritability estimates of psychometric IQ ranged between .80 and .90 in both samples.
On sub-samples of the Australian sample (762 individuals from 345 families) and Dutch samples (245 individuals from 117 families) genotypic marker data has been collected.

Sample heterogeneity with respect to phenotype, subject characteristics, and marker informativeness was evaluated. Both independent and simultaneous scans showed evidence for linkage in several overlapping regions.

116

Should the Monkey Who Types "Hamlet" Win the Pulitzer Prize? Multiple Testing Challenges in the Genomic Era

M.A. Province
Washington University in St. Louis, MO, USA

Cheap and efficient high-throughput genotyping has spawned a golden genomic age. But the very ability to measure hundreds of thousands of polymorphisms on large numbers of subjects presents a major challenge to our understanding of which of them are related to phenotypes through the ''curse of dimensionality.'' Inflation of type I error occurs when conducting a large number of statistical tests of genotype-phenotype relationship (either linkage, association, or both). The classical response is to use more stringent a levels to protect against the high numbers of expected false-positives (Bonferroni, Sidak, Lander-Kruglyak, etc.). But this strategy increases type II error, reduces power, and inflates the number of false-negatives as inevitably as the rain causes rivers to rise. More efficient methods for dealing with the multiple comparisons problem are examined and contrasted, including the False Discovery Rate (FDR) (and Storey's Q), permutation tests, as well as Sequential Multiple Decision Procedures (SMDP Province, 2000). The later comes from the theory of sequential analysis, which can be used even when the data are collected using a fixed sample design. Sequential tests give complete, simultaneous control of both the type I and type II errors of tests (no trading one for another) while using the smallest possible sample size for analysis. For fixed samples, the excess N ''saved'' can be used in a confirmatory, replication phase of the original findings. The SMDP approach replaces the multiple genotype-phenotype tests with a single, simultaneous test of all genotypes, partitioning them into two subsets: the ''signal'' vs. the ''noise,'' with an apriori specifiable global error rate. The method allows efficient screening for the true signals from among a large number uses the smallest possible sample sizes, saves the excess to confirm those findings, controls both types of error, and provides one elegant solution to the debate over the best way to balance between false positives and negatives in the multiple testing paradigm.

117

Phenotypic Dissection of DEXA-based Traits in Osteoporosis

F. Rivadeneira(1–2), M.C. Zillikens(2), J.J. Houwing-Duistermaat(4), T.J. Beck(5), Y. Aoulchenko(1), A.G. Uitterlinden(1–2), B. Oostra (3), C.M. van Duijn(1), H.A.P. Pols(1–2)
(1) Epi & Biostat; (2) Int Med; (3) Clin Genet Erasmus MC Rotterdam, NL; (4) Med Stat & Bioinf LUMC Leiden NL; (5) Radiology Johns Hopkins Baltimore USA

BMD is highly heritable but also an ambigous and dynamic composite trait that limits its use in genetic linkage and association studies. We explored alternative bone phenotypes and hip geometry measurements based on DEXA. Heritabilities (h2) and genetic correlations (GC) were estimated in the Erasmus Rucphen Family (ERF) study, an extended-pedigree study designed to map disease-related QTL. Age and sex adjusted traits included BMD of the hip, spine and total body (including composition) and geometry measurements. Statistically independent phenotypes were identified using principal components analysis (PCA). The h2 of total body, hip and spine BMD ranged from 50–70% and of geometry parameters between 25–50%. Height, fat and lean mass indexes (FMI, LMI) had h2 of 60, 31 and 28%. GC between BMD sites ranged from 70 to 99%(same regions) and was 20%, 15% and 30% with height, FMI, and LMI. PCA of correlated geometry measurements identified two independent components explaining 95% of the variance: one highly correlated with bone stability with h2 of 45% and one highly correlated with dimensions and strength with h2 of 39%. An index mechanosensitivity was estimated as the ratio of bone strength to lean mass had an h2 of 27% and a higher GC with the stability component (42%) than with the strength component(3.5%). Gene identification should be targeted on such bone phenotypes instead of BMD.

118

Phenotypic Dissection of DEXA-based Traits in Osteoporosis

F. Rivadeneira(1–2), M.C. Zillikens(2), J.J. Houwing-Duistermaat(4), T.J. Beck (5), Y. Aoulchenko (1), A.G. Uitterlinden(1–2), B. Oostra(3), C.M. van Duijn(1), H.A.P. Pols(1–2)
(1) Epi & Biostat; (2) Int Med; (3) Clin Genet Erasmus MC Rotterdam, NL; (4) Med Stat & Bioinf LUMC Leiden NL; (5) Radiology Johns Hopkins Baltimore USA

BMD is highly heritable but also an ambigous and dynamic composite trait that limits its use in genetic linkage and association studies. We explored alternative bone phenotypes and hip geometry measurements based on DEXA. Heritabilities (h2) and genetic correlations (GC) were estimated in the Erasmus Rucphen Family (ERF) study, an extended-pedigree study designed to map disease-related QTL. Age and sex adjusted traits included BMD of the hip, spine and total body (including composition)and geometry measurements. Statistically independent phenotypes were identified using principal components analysis (PCA). The h2 of total body, hip and spine BMD ranged from 50–70% and of geometry parameters between 25–50%. Height, fat and lean mass indexes (FMI, LMI) had h2 of 60, 31 and 28%. GC between BMD sites ranged from 70 to 99%(same regions) and was 20%, 15% and 30% with height, FMI, and LMI. PCA of correlated geometry measurements identified two independent components explaining 95% of the variance: one highly correlated with bone stability with h2 of 45% and one highly correlated with dimensions and strength with h2 of 39%. An index of mechanosensitivity estimated as

the ratio of bone strength to lean mass had an h2 of 27% and a higher GC with the stability component(42%) than with the strength component(3.5%). Gene identification should be targeted on such bone phenotypes instead of BMD.

119

Insulin Acutely Regulates Gene Expression in Human Skeletal Muscle in Vivo

D.K. Richardson, J. Finlayson, P.S. Streng, K. Cusi, M. Bajaj, R.A. DeFronzo, L.J. Mandarino, and C.P. Jenkinson
Department of Medicine, Diabetes Division, University of Texas Health Science Center, San Antonio, TX 78229

To determine whether alterations in the regulation of gene expression in skeletal muscle characterize individuals at high risk of developing diabetes, we performed global gene expression analysis in skeletal muscle from Mexican American subjects without (FH−) (43±6 yr, BMI 24±1 kg/m2) and with (FH+) (25±2 yr, BMI 25±1 kg/m2) a family history of diabetes. A euglycemic hyperinsulinemic (80 mU/m2.min) clamp was performed with muscle biopsies taken basally and after 30 and 240 min of insulin infusion. Gene expression profiles of muscle RNA were determined using Affymetrix HG-U133A chips. Insulin-stimulated glucose disposal was significantly lower in FH+ versus FH− (6.4±0.7 vs 9.2±0.6 mg/(kg.min), $P<0.05$). In the FH−, genes that were increased by insulin included, among others, IRS-2 (309±30 to 430±16), insulin receptor (256±43 to 393±50), RAS (588±102 to 787±81) and PI3-K (122±14 to 186±11). In contrast, in FH+, insulin did not alter the expression of genes encoding the insulin receptor and IRS-2 but did increase RAS (864±122 to 1028±99) and PI3-K (191±24 to 251±25) expression. IRS-1 expression was decreased in FH− (494±88 to 376±48) and FH+ (664±82 to 485±93). In addition, a number of genes involved in protein synthesis were increased in expression in the FH− and not in the FH+ following the insulin clamp. This study shows, in vivo, that insulin regulates the expression of genes involved in insulin action and substrate homeostasis in skeletal muscle within minutes.

120

Statistical Properties of Multifactor Dimensionality Reduction (MDR)

M.D. Ritchie, W. S. Bush, L. Fu, and J. H. Moore
Center for Human Genetics Research, Vanderbilt University Medical School, Nashville, TN

Multifactor Dimensionality Reduction (MDR) is a novel statistical approach for detecting gene-gene interactions associated with common, complex diseases. A strength of MDR is its ability to detect interactions in the absence of main effects. MDR has successfully been applied to the study of many common diseases including breast cancer, diabetes, and hypertension. In real data analyses, collecting an adequate sample size is critical for maximizing statistical power, minimizing type I error, and obtaining unbiased estimates of prediction error of MDR models. The purpose of this study is to determine the sample size requirements for MDR analyses to develop unbiased estimates of prediction error. In this study, we generated six, two-locus epistasis models. Each epistasis model was used to simulate a population that consisted of 1000 cases and 1000 controls. We randomly sampled the population with replacement to create 100 data sets of each sample size between 1900 and 100 individuals decreasing by increments of 100. We estimated the average prediction error of the two-locus models using MDR using ten-fold, five-fold, or leave-one-out cross validation. The goal was to determine the bias and variance of the prediction error of the MDR models using various model validation strategies. Based on our simulations, we demonstrate that as the sample size decreases there is an upward bias of the prediction error estimates. In addition, we show that the variance of the estimate increases with decreasing sample size. Accordingly, the cross validation strategy selected impacts the degree of bias and variance for the prediction error estimates. The results of this study provide a framework for study design of future MDR applications.

121

A polymorphism in the regulatory region of PRNP is associated with increased risk and earlier onset of sporadic Creutzfeldt-Jakob disease

P.S.-Juan(1), M.T. Bishop(2), E.A. Croes(1), R.S.G. Knight(2), C.M. van Duijn(1), R.G. Will(2), J.C. Manson(3)
(1) Genetic Epidemiology Unit, Department of Epidemiology & Biostatistics, Erasmus MC, The Netherlands; (2) The National CJD Surveillance Unit, Western General Hospital, UK; (3) Neuropathogenesis Unit, Institute for animal health, UK

Creutzfeldt-Jakob disease (CJD) is a rare transmissible neurodegenerative disorder. Although there are genetic forms, the majority of cases are sporadic. An important determinant for CJD risk and phenotype is the M129V polymorphism of the human prion protein gene (PRNP), but there are also polymorphisms inside the regulatory region (-C101G, G310C and T385C) and further upstream of PRNP (-C1368T). We tested whether these four polymorphisms were associated with the risk of CJD or with disease phenotype in a UK population-based sample of 151 sporadic CJD (spCJD) cases, 89 variant CJD (vCJD) cases and 134 controls. Logistic, linear and Cox regressions were employed in the analysis. Our results confirmed a significant association, independent of PRNP M129V, between G310C, and risk for spCJD (OR=2.8, 95% CI=1.1 to 7.2; p=0.033). Further, patients with spCJD carrying a 310C allele presented at a significantly earlier age (7.6 years earlier (p=0.04)). No significant associations were found with vCJD patients. The associations of the PRNP regulatory region G310C polymorphism with an increased risk and an earlier onset of spCJD suggest that it may act by controlling the expression of PRNP. Our findings support the hypothesis that a PRNP dosage effect may play a role in the pathogenesis of spCJD.

122

Evidence of Association of Autism With A Region On Chromosome 7q

S.L. Santangelo(1,2,3), S.A. Haddad(1), S. Purcell(1,4)
(1) Psych & Neurodevelop Genetics, Mass General Hosp, USA; (2) Dept Psych, Harvard Medical Schl, USA; (3) Dept Epi, Harvard Schl of Public Health, USA; (4) Whitehead Institute for Genomic Research, USA

Autism is a neurodevelopmental disorder of childhood characterized by impairments in social interaction, language development, and patterns of behavior. Models implicate anywhere from 2 to >=15 susceptibility genes, none of which has yet been identified. Several studies have shown evidence for linkage to chromosomes 2q and 7q.
The autism sample was comprised of 30 extremely discordant sib pairs. Single nucleotide polymorphism (SNP) markers at roughly 500 kb intervals were selected to cover a 40cM long segment of chromosome 2q and an 85cM long segment of chromosome 7q. Microsatellite maps were used to estimate the genetic locations of the SNP markers using linear regression. IBD sharing was assessed using Aspex. Family-based association analyses were run using the Whap program developed by Shaun Purcell and Pak Sham.
There was no evidence for linkage to the regions examined on chroms 2q or 7q. Global family-based association tests also returned non-significant results. However, nominally significant allelic associations were found at three markers on chrom 2q and eight markers on chrom 7q. Additional SNPs were typed at 20 kb intervals around two of the significant markers on chrom 7q. Haplotype tests showed evidence of association with a 4-SNP haplotype in this region (empirical p-value 0.010), with 2 of the 4 SNPs in this haplotype showing significant individual allelic associations. All 4 SNPs appear to be in strong LD with each other.

123

Heritability of pulse wave velocity and carotid intima-media thickness in an extended pedigree from an isolated population

F.A. Sayed-Tabatabaei(1), M.J. van Rijn(1), Y. Aulchenko(1), B.A. Oostra(2), J.C.M. Witteman(1), C.M. van Duijn(1)
(1) Department of Epidemiology & Biostatistics; (2) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands

Pulse Wave Velocity (PWV) and carotid intima-media thickness (IMT) represent arterial wall stiffness and thickness. Both traits are associated with cardiovascular diseases but little is known of their heritability. We studied the heritability of PWV and IMT in an extended pedigree from an isolated population in the Netherlands. PWV was measured in 802 and common carotid IMT in 815 individuals belonging to a single pedigree. Natural logarithm transformations of both traits were used in the SOLAR software to estimate heritabilities. Covariates used in the complete multivariable-adjusted model of PWV were sex, age, age^2, heart rate, systolic blood pressure, body mass index (BMI) and smoking. For IMT the covariates were sex, age, age*sex, systolic and diastolic blood pressures, BMI and smoking. The age and sex-adjusted and multivariable-adjusted heritability for PWV were 0.36 (SE=0.09) and 0.30 (0.08). For IMT these statistics were 0.40 (0.08) and 0.33 (0.08), respectively (p for all estimates <0.001). The proportion of PWV variance due to all covariates was 0.55 and for IMT was 0.65. To our knowledge, this is the first report on heritability of PWV. Our findings suggest that a substantial part of variance in vessel wall stiffness and thickness are explained by genetic factors. This opens the possibility to search for genes determining vascular structure using both traits as useful measurements.

124

Robust Multipoint Identical-by-Descent Mapping with Affected Relative Pairs

D.J. Schaid(1), J.P. Sinnwell(1), S.N. Thibodeau(2)
(1) Departments of Health Sciences Research and (2) Laboratory Medicine and Pathology, Mayo Clinic School of Medicine, Rochester, MN, 55905 USA

The genetic mapping of complex traits has been challenging, requiring new statistical methods that are robust to misspecified models. Liang et al. [2001] proposed a robust multipoint method that can be used to simulatneously estimate the trait-locus position and its genetic effect, based on sibpair linkage data. We extend this approach to different types of affected relative pairs by two approachs. One approach is to estimate a single trait-locus position, yet different trait-locus effects, in order to evaluate the consistency of linkage evidence across different types of affected relative pairs. This may be useful for older-onset diseases, where secular changes in diagnostic methods can change the frequency of phenocopies among different types of relative pairs. Our second approach models the trait as a single susceptibility locus, yet with a robust method that requires few assumptions and is computationally efficient. Furthermore, we develop a robust score statistic to test whether there is significant evidence against using a single locus model. These methods are applied to a prostate cancer linkage study, which emphasizes their potential advantages and limitations.

125

Evaluation of the ordered subset analysis (OSA) method for linkage mapping in the presence of gene-environment interaction

S. Schmidt, M. Schmidt, E.R. Martin, E.R. Hauser
Duke University Medical Center, Durham, NC, USA

The ordered subset analysis (OSA) method has recently been proposed as an approach to incorporate environmental covariates into parametric or non-parametric linkage analysis. The goal of the method is to identify more homogeneous subgroups of families on the basis of

disease-associated covariates or phenotypic features, which may provide evidence for linkage that is otherwise obscured by genetic heterogeneity. When linked and unlinked families are distinguished by having different means of a normally distributed covariate, the method was shown to have excellent power to detect linkage in a subgroup of families. In this case, OSA can be superior to the admixture test for parametric linkage analysis implemented in the HOMOG package, particular when families are small and the overall genetic effect is low. Here, we present results of evaluating the performance of OSA when gene-environment (GxE) interaction exists at the individual level. We have incorporated exposure to a binary environmental covariate into the simulation algorithm implemented in the SIMLA program, which generates marker genotypes and binary disease phenotypes for families of fixed size given user-specified ascertainment criteria, parameters for a disease gene model, and a map of marker loci. We have assumed variable proportions of families linked to a disease gene that interacts with an environmental covariate. In the absence of familial correlations of the covariate, we demonstrate that OSA has correct type I error but relatively low power to detect linkage in the presence of GxE interaction. The effect of familial correlations in environmental exposure (ranging from 0.1 to 0.9) and comparisons of using binary vs. continuous covariates will be presented.

126

Maximum-likelihood estimation of type 1 diabetes associations with a genetic factor and independent continuous attribute

J.-H. Shin(1), L. Bekris(2,3), F. Farin(3), T. Kavanagh(3), A. Lernmark(2), B. McNeney(1), J. Graham(1)
(1) Statistics and Actuarial Science, Simon Fraser Univ., Canada, (2) Robert H. Williams Lab., Univ. of Washington, USA, (3) Environmental Health, Univ. of Washington, USA

In case-control studies of rare diseases such as type 1 diabetes, covariate information is often collected on a genetic factor and a continuous attribute such as age. In some instances, it is reasonable to assume that the attribute and genetic factor occur independently in the population. Under this independence assumption, we developed maximum likelihood estimators of parameters in a logistic model of disease risk. Estimates are based on data from both patients and controls and may be obtained by fitting a polychotomous regression model of joint disease and genetic status. Our results extend previous log-linear approaches to imposing independence between a genetic factor and a categorical attribute, thereby avoiding potential loss of information from discretizing a continuous attribute. We apply the method to investigate the effects of age and a variant of the glutamate-cysteine ligase catalytic subunit on type 1 diabetes. The results are compared to those obtained from a standard logistic regression analysis, which does not make use of the independence assumption.

127

The risk of coronary heart disease and the interleukin 6-174 G/C promotor polymorphism

M.P.S. Sie(1), A.G. Uitterlinden(2), C.M. van Duijn(1), J.C.M. Witteman(1)
(1) Dept. of Epidemiology & Biostatistics and (2) Dept. of Internal Medicine, Erasmus Medical Center, Rotterdam

The presence of inflammation in atherosclerosis has been noted from the earliest histological observations and may be related to its pathogenesis. Interleukin 6 (IL-6) is an important inflammatory cytokine. The IL-6 -174 G/C promotor polymorphism is a functional polymorphism and has been the subject of much interest. However, study results on the relation between this polymorphism and the risk of cardiovascular diseases are inconsistent. We therefore investigated this polymorphism in relation to the risk of coronary heart disease (CHD).
6434 persons (2612 male, 3822 female; mean age 69.5 years), participating in the Rotterdam Study, were genotyped. In random subgroups the serum levels of IL-6 and C-reactive protein (CRP) were determined. During 6.3 years of follow-up there were 648 CHD cases and 280 myocardial infarctions. The association between serum levels and genotype was investigated using regression analyses and T-tests. The relation between the genotype and the risk of CHD was investigated using Cox regression analyses. All analyses were corrected for common (risk) factors.
Genotype and allele frequencies were in Hardy Weinberg Equilibrium. The promotor polymorphism was not significantly associated with IL-6 or CRP serum levels. No clear relation between the genotype and the risk of CHD was observed. The -174 G/C IL-6 promotor polymorphism is of no predictive value in the estimation of the risk of coronary heart disease in this elderly Dutch population.

128

Evidence of a recessive gene for Alzheimer's disease at chromosome 3q26

K. Sleegers(1), Y. Aulchenko(1), A. Bertoli(1), A. Arias(1), G. Roks(3), M. Cruts(4), C. van Broeckhoven(4), J. van Swieten(2), P. Heutink(5), B. Oostra(1), C. van Duijn(1)
(1) Gen Epi Unit and (2) Neurology, Erasmus MC Rotterdam, NL; (3) Neurology, St Elisabeth Hospital Tilburg, NL; (4) Mol Genetics, Univ Antwerp, Belgium; (5) Med Genomics, VU Med Center Amsterdam, NL

The genetic model underlying late-onset Alzheimer's disease (AD) is largely unknown. In a recent genetically isolated Dutch population we identified 4 inbred families multiply affected with AD, that could be connected to a common ancestor in 8 generations. We performed a full genome search with 420 markers at an average distance of 10 cM in 8 descendants of this pedigree who were affected with probable late-onset AD. Homozygosity mapping revealed a region with evidence for linkage (Lod score 3.4) on chromosome 3q26. The region could be narrowed

down to a 5.8 cM interval after testing additional markers, with a maximum multipoint Lod score of 3.0. A family segregating AD as a classical recessive trait contributed most to this finding. Although this suggests a recessive mutation at this locus, we cannot exclude a co-dominant effect. In an association analysis in a series of 153 patients and 75 controls from the isolate we replicated the finding for markers D3S1282 and D3S1565. The region of interest contained a plausible candidate gene butyrylcholinesterase (BCHE). Sequencing of the enzyme coding regions of this gene revealed no variants that could explain the occurrence of disease. The locus is likely to encompass a new gene involved in AD.

129

Heritability Estimation and Association Testing Under Non-Random Ascertainment

A.J.M. Sorant, A.F. Wilson
Genometrics Section, Inherited Disease Research Branch, National Human Genome Research Institute, NIH, Baltimore, MD, USA

Statistical properties of estimation and significance testing are often based on assumptions such as random sampling that may be unrealistic for genetic studies. To examine the effects of non-random ascertainment on heritability estimation and association testing of a candidate locus, quantitative traits were simulated using G.A.S.P. for 2000 samples of 100 nuclear families with 4 offspring each. The simulated trait was based on a biallelic locus with minor allele frequencies of 30% and 50%, and heritabilities were 0%, 10% and 30%. In addition to the trait locus, a linked marker and an unlinked marker were generated. For each model, four ascertainment schemes were considered: random sampling (RS), family selected if at least one offspring has a high trait value (H1), selection if at least two offspring have high trait values (H2) and selection if there are extremely discordant offspring (ED).
Each sample was used to estimate overall trait heritability and heritability attributable to the locus and to test for association between the trait and the locus. Several methods were considered: regression on mid-parent (ROMP), the variance components method implemented in SOLAR adapted to use identity by state, and association analysis using the George-Elston model implemented in ASSOC of the S.A.G.E. package. Locus-specific analysis was done for trait and marker loci. For the considered models, the ROMP method was most consistent across different ascertainment schemes, with respect to heritability estimation and power.

130

Dating a recent mutation in MSH2 that increases risk for colorectal cancer

S. Sun(1), I. Thiffault(2), N. Hamel(2), W.D. Foulkes(2), C.M.T. Greenwood(1)
(1) Genetics & Genomic Biology, Hospital for Sick Children, Toronto, and University of Toronto (2) Div. Medical Genetics, Res Inst. of the McGill University Health Centre, McGill University, Montreal

Foulkes et al. (Am J Hum Genet. 2002; 71:1395-412) identified a mutation in MSH2 (MSH2∗1906G→C) that leads to increased risk of colorectal cancer among Ashkenazi Jewish individuals. Cases with the mutation carry a large haplotype in common surrounding the mutation site, suggesting that all cases are descendants of a recent founder. To estimate the age of the mutation, we used both single-locus and multi-locus methods. The software DMLE (Reeve and Rannala, Bioinformatics 2002; 18:894–895) calculates a Bayesian posterior distribution for the mutation age from a coalescence-based model for gene genealogies.
Since the number of cases with disease is small, the single-locus estimates are very sensitive to the number of cases carrying the linked allele. At D2S2391, 1.29 Mb from the mutation, we obtained an estimate of 18 generations based on 14/16 cases carrying the linked allele. When we used a growth rate of 1.5-fold per generation, DMLE estimated the weighted mean number of generations to the founding mutation as approximately 9.9, with 95% coverage from 6.5–14.8 generations, corresponding to approximately 250 years, although the distribution had a long right tail. Estimates of age based on the DMLE posterior distribution are very sensitive to the assumed population growth rate. However, this rapid growth is consistent with estimates of the Ashkenazi Jewish population in 1765 and 1900.

131

Gene mapping meets Bayesian Model Selection in Stochastic Search Gene Suggestion

M.D. Swartz(1,2), P.J. Mueller(3), C.I. Amos(1), M. Kimmel(4)
(1) Dept. of Epi, U. T. M. D. Anderson Cancer Center, USA; (2) Dept. of Stat., Rice U., USA; (3) Dept. of Biostat., U.T. M. D. Anderson Cancer Center, USA; (4) Dept. of Stat. Texas A & M Univ., USA

Mapping the genes for a complex disease, such as Rheumatoid Arthritis (RA), involves finding multiple genetic loci that may contribute to the onset of the disease. Pairwise testing of the loci leads to the problem of multiple testing. To avoid multiple tests, we can look at haplotypes; but this results in a contingency table with sparse counts. Using case-parent triad data, we extend the Bayesian conditional logistic regression model developed by Thomas, et al. [Genetic Epidemiology, 5, 1995, pp 455–466], by defining prior distributions on the allele main effects that model the genetic dependencies present in the HLA region of Chromosome 6. We also added a hierarchical level for model selection that accounts for both locus and allele selection. Thus we cast the problem of identifying genetic loci relevant to the disease into a problem of Bayesian model selection. We evaluate the performance of the procedure with some simulated examples, and then apply our procedure to identifying genetic effects influencing susceptibility to RA. This research is supported by a Genetic

Epidemiology Fellowship supported by NCI grant R25-CA57730, Robert Chamberlain, PhD, Principal Investigator.

132

Classification analysis as a tool for dissection of genetic background underlying complex traits

J. Szyda(1), S. Viitala(2), S. Blott(3), J. Komisarek(4), N. Schulman(2), M. Lidauer(2), A. Mäki-Tanila(2), M. Georges(3), J. Vilkki(2)
(1) Department of Animal Genetics, Wrocaw Agricultural University, Poland; (2) Animal Production Research, MTT Agrifood Research Finland; (3) Department of Genetics, Faculty of Veterinary Medicine, University of Liège, Belgium; (4) Department of Cattle Breeding, Agricultural University of Pozna, Poland

Real and simulated data are used to examine the usefulness of different classification methods (linear- and quadratic discriminant analysis, k-nearest-neighbours) as exploratory tools for the analysis of genetic structure of complex traits. The two real data sets comprise: (i) 810 male individuals of Finnish Ayrshire cattle, characterised by 6 snps and 10 quantitative measurements stemming from 21 large paternal half sib families with additional pedigree information on nongenotyped individuals available for practically whole active population bred in Finland, (ii) 100 female individuals of Jersey cattle, characterised by 6 snps and 7 quantitative measurements with pedigree information available for one generation back. The main goal of the study is the analysis of real data sets comprising (i) estimation of the impact of particular snps on the quantitative measurements, (ii) separation between the oligogenic (i.e. due to snps) and polygenic (i.e. due to cumulated small effects of genes scattered across the genome) components underlying the observed quantitative variation, (iii) evaluate the predictive ability of classification methods across both data sets. Additionally, simulated data are used to describe basic statistical properties of classification method used in this study.

133

Segregation Analysis of Resting Metabolic Rate in a Nigerian Population

B.O. Tayo(1), Y. Liang(2), X. Zhu(3), A. Luke(3), X. Wu(3), A. Adeyemo(4), R.S. Cooper(3)
(1) Dept. of Social and Preventive Medicine, SUNY at Buffalo, USA; (2) Dept. of Biostatistics, SUNY at Buffalo, USA; (3) Dept. of Preventive Medicine and Epidemiology, Loyola Medical School, USA; (4) Dept. of Pediatrics, College of Medicine, Univ. of Ibadan, Nigeria

To investigate the extent to which resting metabolic rate is determined by genetic effects, and its possible mode of transmission between relatives, we performed complex segregation analysis on 1000 individuals from 161 randomly sampled Nigerian families. This study was part of an international collaborative study on the genetics of obesity and hypertension in Blacks. About 65% of the variation in resting metabolic rate was accounted for by combined effects of free fat-mass, fat mass, age, sex, and body mass index. Heritability of covariate-adjusted resting metabolic rate was 36%. We fitted several regressive models to the data and based our selection of the best fit model on likelihood procedures. Results of our segregation analysis indicate evidence of Mendelian transmission with residual familial correlations but no strong support for segregation of a major gene for resting metabolic rate in this population.

134

Jointly characterizing allelic associations and estimating haplotype frequencies from diploid data by graphical modeling

A Thomas
Department of Medical Informatics and Center for High Performance Computing, University of Utah, USA

A method for estimating a graphical model to describe allelic associations between genetic loci (Thomas & Camp 2004) is extended to use diploid genotypes rather than haploid data. It also provides haplotype frequency estimates and estimates of phase for sampled individuals. Resulting haplotype frequencies are shown to be similar to those obtained by the PHASE program. The graphical model approach has some advantages in terms of tractability and can be used to select informative subsets of loci and to map loci influencing phenotypes. We show also that constraining the decomposable graphical models to the class of those whose Markov graph is an interval graph has little effect on estimates of haplotype frequency, but significant computational advantages.

A Thomas & N J Camp (2004) Graphical modeling of the joint distribution of alleles at associated loci. American Journal of Human Genetics 74: 1088–1101.

135

Two-stage sampling designs for gene association studies

D.C. Thomas, R. Xie, M. Gebregziabher

We consider two-stage case-control designs for testing associations between single nucleotide polymorphisms (SNPs) and disease, in which a subsample of subjects is used to select a panel of ''tagging'' SNPs that will be considered in the main study. We propose a pseudolikelihood (Pepe and Flemming, JASA 1991; 86:108–113) that combines the information from both the main study and the substudy to test the association with any polymorphism in the original set. SNP-tagging (Chapman et al, Hum Hered 2003; 56:18–31) and haplotype-tagging (Stram et al, Hum Hered 2003;55:27-36) approaches are compared. We show that the cost-efficiency of such a design for estimating the relative risk associated with the causal polymorphism can be considerably better than for a single-stage design, even if the causal polymorphism is not included in the tag-SNP set. We also consider the optimal selection of cases and controls in such designs and the relative efficiency for estimating the location of a causal variant in linkage disequilibrium mapping. Nevertheless, as the cost of high-volume genotyping plummets and haplotype tagging information from the International

HapMap project (Gibbs et al, Nature 2003; 426:789–96) rapidly accumulates in public databases, such two-stage designs may soon become unnecessary.
Motivation for this approach comes from the design of a study of CHEK2 polymorphisms in breast cancer and their interactions with ionizing radiation and other genes involved in repair of double strand breaks (BRCA1/2, ATM, p53 binding protein, and others). This counter-matched study allows for the possibility of further efficiency gains for testing interaction effects by subsampling on the basis of exposure information (Bernstein et al. Breast Cancer Res 2004; 6:R199–214).

136
Cancer risks and mortality associated with ATM mutations
D. Thompson(1), D. Easton(1), A.M.R. Taylor(2)
(1) CR-UK Genetic Epidemiology Unit, Univ. of Cambridge, UK; (2) Division of Cancer Studies, Univ. of Birmingham, UK

Homozygous mutations in the ATM gene are the principal cause of Ataxia Telangiectasia (AT). Patients with AT are at a grossly elevated risk of cancer, predominantly lymphoid tumours. Various studies have found a higher than expected incidence of breast cancer among female relatives of AT patients, suggesting that heterozygous ATM mutation carriers have an increased breast cancer risk. In contrast, most case-control studies have failed to find an association.
We assembled a large cohort of 169 UK AT patients and their relatives. Information on mortality and cancer incidence was obtained by tracing through the National Health Service Central Registry. Information was available from 1160 relatives, including 247 parents (obligate carriers). Relative risks (RR) of cancer in carriers, allowing for genotype uncertainty, were estimated using a maximum likelihood approach, via the EM algorithm. The RR for all cancers was 2.1 (95% CI 1.3–3.4) in females and 1.2 (0.8–2.0) in males, based on 118 reported cancers. The RR of breast cancer (23 cases) was 2.2 (1.2–4.3), rising to 4.9 (1.9–12.9) below age 50. No other individual site showed a clear excess risk, although the increased female cancer risk could not be explained by breast cancer alone. All-cause mortality was significantly increased in female carriers (RR=1.7, 95% CI 1.0–2.9), but not in males (RR=1.0, 95% CI 0.8–1.4). Mutations that allowed the retention of even a small degree of ATM kinase activity were not associated with the increase in lymphoid tumour risk typically seen in AT patients.

137
MLE's using re-genotyped data
N. Tintle(1), S.J. Finch(1), D. Gordon(2)
(1) Dept. of Applied Math and Stat, SUNY at Stony Brook, USA; (2) Lab. of Statistical Genetics, Rockefeller Univ, USA

Genotyping data is prone to misclassification errors. Often, a random sub-sample of units is re-genotyped in order to understand how frequently misclassification errors occur. We suggest how to find MLE's for misclassification rates and the true genotype distribution when some, or all, of the sample is re-genotyped (reclassified). A simulation study investigates bias and standard errors under different probability distributions and misclassification rates. Current work is using the MLE's to develop a case-control LRT for partially re-genotyped data.

138
Inclusion of a Covariate Yields Major Gains in Power in Linkage Analysis
T.N. Turley-Stoulig(1), A.J.M. Sorant(2), J.E. Bailey-Wilson(2), D.M. Mandal(1)
(1) Dept. of Genetics, LSU Health Sciences Center, LA, USA; (2) NHGRI/NIH, MD, USA
Inclusion of covariates in linkage analysis of complex traits may increase the power to detect linkage. To identify situations in qualitative trait linkage analysis with large power gain, this study evaluates the effect of inclusion of a covariate on power and type I error for qualitative trait linkage analysis under various inheritance models.
A binary trait with penetrance determined by a dominant biallelic locus and modified by a quantitative covariate was simulated with G.A.S.P., using various degrees of penetrance, disease allele frequency and covariate effect. For each model, segregation analysis was performed (REGDHUNT) on a large singly ascertained sample to provide a trait model for use in the model-based analysis. Then the trait and linked (Ө=0.01, 0.05) and unlinked markers were generated for 10,000 samples of 300 nuclear families with 4 offspring. Each linkage sample was analyzed with model-based (LODLINK, S.A.G.E. 3.1) and model-free (revised Haseman-Elston regression SIBPAL, S.A.G.E. 4.5, using the mean-corrected cross-product as dependent variable) methods, with and without the covariate.
Under examined inheritance models, both LODLINK and SIBPAL provided more powerful linkage analysis and conserved type I error with covariate inclusion. Under models with moderate to strong genetic effect and moderate covariate effect, power to detect linkage with LODLINK was increased as much as 6-fold when the covariate was included in analysis.

139
Extent of LD in the East Finland founder population
P. Uimari, O. Kontkanen, M. Pirskanen, R. Fuentes, J.T. Salonen

Linkage disequilibrium (LD) is important in mapping loci predisposing common diseases. The power of an association study is directly related to r^2, the squared correlation coefficient between the two loci. Ardlie et al. (2002) conclude that the extent of LD that is useful in mapping studies range from 10–30 kb for European populations. For African descended populations the extent of LD is smaller. A sample of 118 unrelated

individuals from the East Finland founder population (established some 15 generations ago by some 400–800 individuals) were genotyped for over 80,000 SNPs totalling more than 480,000 pair wise D′ and r^2 measures. From the complete disequilibrium with very close SNPs, r^2 gradually decreased so that the average r^2 was 0.52, 0.41, and 0.29 for distances of 10, 20, and 40kb between SNPs. Similarly, the values for D′ were 0.89, 0.78, and 0.65 for the same distances. A threshold used for minor allele frequency (MAF) had an effect on the mean r^2 so that for a distance of 20kb between the SNPs, r^2 was 0.28, 0.32, 0.41, 0.51, and 0.64, for MAF thresholds of 0.05, 0.1, 0.2, 0.3, and 0.4. For D′ the effect of MAF was smaller. These r^2 values can be used for estimating the sample size required for association studies.

140

Power to Detect Linkage when the Lod Score is Maximized over Mode of Inheritance and Phenotype Definition

A. Ulgen, R. Ottman
G.H. Sergievsky Center, Columbia Univ., USA

We evaluated the power to detect linkage when the lod score is maximized over mode of inheritance (MOI) and phenotype definition. The data set comprised 89 families containing two or more sampled individuals with non-symptomatic epilepsy. Affected individuals had various epilepsy types (generalized 32%, focal 45%, both 5%, unknown 18%), and it was unclear what phenotype should be assumed to result from the susceptibility gene. We simulated genetic marker data in 1000 replicates of the data set, assuming a dominant gene with 75% penetrance and no phenocopies raised risk only for focal nonsymptomatic epilepsy, and was linked to the marker with $\boldsymbol{\theta}$=0.05. In analyses under the generating model, using the conventional lod score threshold of 3.0, power was 99.7% when the phenotype was correctly assumed to be focal epilepsy, but much lower when it was incorrectly assumed to be any type of epilepsy (43%), or generalized epilepsy only (0%). We then estimated power in an analysis maximized over four MOI models (high penetrance dominant, high penetrance recessive, low penetrance dominant and low penetrance recessive), and three phenotype definitions (any type of non-symptomatic epilepsy, generalized only, focal only). In this analysis we used as the critical value the lod score corresponding to p=0.0001 from a previous simulation study of the same dataset (3.94). Power for this analysis was 99.2%. These results suggest that in studies of complex traits with unknown MOI and phenotype definition, analysis maximized over both of these factors can yield good power for linkage detection.

141

Model-free linkage analysis adaptive to genomic imprinting: application to leprosy

Q. Vincent, A. Alcais, L. Abel

Genomic imprinting is a phenomenon by which only one of the two alleles of a gene is expressed, depending on its parental origin. Since paternal and maternal transmissions of marker alleles are weighted equally in classical linkage analysis, imprinting is expected to affect strongly the detection of genes displaying this type of regulation. We propose a method, based on the model-free MLB (Maximum Likelihood Binomial) approach, which takes into account the parental origin of markers by splitting the central probability transmission parameter α into a paternal (αP) and a maternal (αM) specific parameter. This specification allows the natural formulation of (i) a linkage test adaptive to imprinting and (ii) an imprinting test in presence of linkage. In addition, the present method takes advantage of the MLB structure, such as the efficient accounting for large sibships and the handling of multipoint analysis. The distribution of the two tests under their respective null hypothesis of no linkage and non-imprinted linkage are derived and validated through simulations under different genetic models. The linkage test adaptive to imprinting is more powerful than the classical linkage test as soon as modest imprinting is involved. The power of the imprinting test is independent of the intensity of linkage but, as expected, increases with the intensity of imprinting. Further studies show that sex-specific recombination rates do not affect these tests as soon as a relatively dense map ($<$4 cM) is considered. This new method has been used to carry out a genome-wide scan to map potential imprinted susceptibility genes for leprosy, a disease previously shown to be linked to chromosome 6q25. Interestingly, a new locus located on chromosome 7q34 shows significant imprinting and is currently under further investigation.

142

Genomic Screening in Family Based Association Testing

K. Van Steen(1), M.B. McQueen(2), A. Herbert(3), N.M. Laird(1), B. Raby(4), H. Lyon(4,5), J. Su(2), S. Datta(4), C. Rosenow(6), E.K. Silverman(4), S.T. Weiss(4), C. Lange(1)
(1) Department of Biostatistics, Harvard School of Public Health, Boston, USA; (2) Department of Epidemiology, Harvard School of Public Health, Boston, USA; (3) Department of Genetics and Genomics, Boston University School of Medicine, Boston, USA; (4) Channing Laboratory, Harvard Medical School, Boston, USA; (5) Division of Genetics, Children's Hospital, Boston, USA; (6) Genomics Collaboration Genotyping, AFFYMETRIX, INC., Santa Clara, USA

The statistical challenge in analyzing genome-wide association studies for complex traits stems from the severe multiple comparison problem that has to be dealt with. Our proposed methodology offers an alternative way to carry out genome-wide screenings in family-based association studies, while controlling the overall type I error rate. The technique is implemented in the PBAT-software (http://www.biostat.harvard.edu/~fbat/default.html) and uses the unified approach to family-based tests of association, as discussed in, for instance, Laird et al. (2000)

and Lange et al. (2003). We illustrate the proposed screening strategy on SNP data, collected from asthmatic children and further support the applicability of the technique using simulation studies.
Laird N, Horvath S and Xu X (2000) Implementing a unified approach to family based tests of association. Genetic Epi 19 (Suppl 1): S36–S42. Lange C, DeMeo D, Silverman E, Weiss S and Laird NM (2003) Using the noninformative families in family-based association tests: a powerful new testing strategy. Am J Hum Genet 73: 801–11.

143

MTHFR 677 genotype and risk of NTD: distinguishing between embryonal and maternal effects

H.H.M. Vermeulen(1), H. Straatman(2), Y.J.M. van der Linden(3), H.J. Blom(3), M. den Heijer(1)
(1) Dept. of Endocrinology; (2) Dept. of Epidemiology and Biostatistics; (3) Dept. of Pediatrics University Medical Center Nijmegen, The Netherlands

The 677c>t variant in the MTHFR gene, an enzyme in folate metabolism, is associated with the development of neural tube defects (NTD). It remains unclear whether the embryonal or the maternal genotype is of main importance. The relation between embryonal and maternal MTHFR genotype and NTD was evaluated in a database consisting of 76 case parent triads, 18 dyads, and 6 monads. To be able to distinguish between embryonal and maternal effects, several tests and techniques were applied (transmission/disequilibrium test (TDT), transmission asymmetry test (TAT), log-lineair model).
The TDT showed no significant influence of the embryonal 677ct and 677tt genotype compared to the wildtype (X^2_{td}=2.1; p=0.09). The TAT was not significant (p=0.07) but was indicative of an influence of a maternally derived T-allele. The relative risk (95% CI) estimates produced by the log-linear approach for one and two T-alleles were 1.3 (0.7–2.3) and 1.4 (0.6–3.5) for children, and 0.9 (0.5–1.6) and 2.7 (1.0–7.5) for mothers, respectively.
In conclusion, the application of several methods of analysis showed that the relation between MTHFR 677 genotype and NTD seemed to be mediated mainly through the TT genotype of the mother. The TAT was suggestive of a parent-of-origin (PO) effect but could have given invalid results because of the presence of a maternal effect. The results of a valid method of analysis of a PO effect will be available soon.

144

The effect of variation in marker informativeness between families on QTL mapping test statistics

P.M. Visscher(1), N.R. Wray(2)
(1) School of Biological Sciences; (2) Molecular Medicine Centre. Univ. of Edinburgh, UK

Recently it has been suggested that families that are uninformative or partially informative for linkage can cause a 'bias' in the test statistic (i.e., reduce power) when using conventional multipoint non-parametric linkage analysis. Although most conventional non-parametric linkage methods are actually invariant to leaving uninformative pairs in or out of the analysis, QTL mapping methods are not, because measures of identity-by-descent are assumed known without error even when there is uncertainty. Using simulation, we investigated the effect of leaving completely uninformative pairs in or out of the analysis for a sibpair design. Analysis of quantitative traits was either using linear regression or maximum likelihood. The number of sibpairs, the proportion of variance explained by a single QTL and residual family resemblance and the proportion of uninformative pairs were varied. The difference between analyses that included or excluded the uninformative pairs was expressed relative to the mean test statistic and to the sampling variance of the test statistic. It was found that the difference in test statistic between analyses that included or excluded uninformative pairs was very small (0–8% of the mean test statistic for 10–90% uninformative pairs) and unlikely to be of practical significance. Uninformative or partially informative pairs provide information on the estimate of fixed effects, total phenotypic variance and average sibling covariance and should not be left out of the analysis.

145

A Two-Level Genetic Regression Model for Pedigree Analysis

T. Wang and R.C. Elston
Department of Epidemiology and Biostatistics, Case Western Reserve University

Linkage analysis and association analysis are two major approaches to detect disease genes in genetic epidemiology. In the report, we propose a two-level genetic model in which linkage analysis and association analysis are unified in a single regression model: linkage disequilibrium and linkage are detected at the individual-level and the pedigree-level, respectively. When this model is used only for detecting linkage, it results in a new version of Haseman-Elston regression (two-level HE). This version of Haseman-Elston regression can make use of all the trait information in any general pedigree. Under the assumption of normality, it is asymptotically equivalent to the variance component model. For the cases when the sample size is small or the assumption of normality cannot be guaranteed, an unbiased REML estimator or a robust variance estimator is available, respectively. Furthermore, this version of Haseman-Elston regression can simultaneously incorporate individual-level and pedigree-level covariates in a natural manner. Complex genetic mechanisms, such as gene-gene interaction, gene-environmental interac tion and imprinting, can be directly modeled. Computer simulations show that the two-level HE maintains good control of type I error rate and gives satisfactory power.

146

Detection of complex gene-gene interactions in a population-based study of breast cancer in Ontario

Y. Wang(1), I. Rajendram(2), E. Shi(3), V. Onay(4), J. Knight(1), H. Ozcelik(4), L. Briollais (1)
(1) Division of Epidemiology and Biostatistics, Mount Sinai Hospital, Toronto, Canada; (2) Fred A. Litwin Centre for Cancer Genetics, Samuel Lunenfeld Research Institute, Canada; (3) Ontario Cancer Genetics Network, Cancer Care Ontario, Toronto, Canada; (4) Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Canada

The grand challenge to statistical genetics is the development of powerful methods that allows the identification of genes that control biological pathways. Indeed, the genetic determinants of complex human disease are likely to result from the poorly understood interaction of dozens, if not hundreds, of disease genes. Breast cancer provides a useful paradigm for genetically complex diseases. To assess the importance of genetic interaction in breast cancer, we have systematically studied the statistical interactions between specific SNPs (Single Nucleotide Polymorphisms) in 19 genes related to major cancer pathways (cell cycle, estrogen metabolism, DNA repair and immune system) based on a population-based sample of 398 cases and 372 controls collected in Ontario. Three methods were used and compared: logistic regression model (LRM), classification and regression tree (CART), and the multifactor dimensionality reduction method (MDR). The ability to detect complex interactions by these three methods is studied. Model validation and permutation tests are both applied to decrease the false positive findings and they also allow us to assess the reliability of the interactions we found.

147

Family based association study between BCL2 gene and type 1 diabetes in Newfoundland

K.S. Wang(1), X.Q. Liu(1), B. Bharaj(1), M. Lu(1), H.T. Chen(1), J.A. Curtis(3), L.A. Newhook(3), A.D. Paterson(1,2)
(1) Program in Genetics and Genomic Biology, The Hospital for Sick Children, Toronto; (2) Dept. of Public Health Sciences and Psychiatry, Univ. of Toronto; (3) Pediatrics, Memorial Univ. of Newfoundland, Canada

IDDM6 (18q12-q23) has been suggested as a type 1 diabetes (T1D) susceptibility locus. BCL2 (B-cell leukemia 2) is a positional candidate gene for IDDM6. DNA from 625 persons with T1D and 1532 relatives without diabetes from 538 pedigrees were recruited from Newfoundland (NF) where there are documented founder effects and a high incidence of T1D. Two microsatellite markers ((AC)n and D18S51) and 6 SNPs were genotyped. The TDTPHASE program was used to test for association between the markers and T1D. Secondary analysis were performed to compare maternal and paternal transmissions as well as gender specific transmissions in probands because the risk of T1D may differ between males and females and some T1D susceptibility loci are subject to genomic imprinting (i.e. the parental origin of a susceptibility allele affects penetrance). There was no evidence for association of single makers with T1D using the whole sample. However, we observed that the (AC)n 195 allele showed significant transmission disequlibrium from fathers to affected offspring (p=0.00016) but not to unaffected offspring (p=0.13) while the same allele transmitted from mothers to affected offspring did not show significant results (p=0.089). Furthermore, the D18S51 275 allele revealed significant negatively transmission from fathers to male affected offspring (p=0.0072) and there was a significant difference between the transmission of the allele from fathers and mothers (p=0.001) as well as between transmission to male and female affected offspring (p=0.008). Moreover, two paternal haplotypes 1–195 from C8687299-(AC)n and 195-2 from (AC)n-C2855835 demonstrated significant association but the same maternal haplotypes were not associated. Our results for the marker (AC)n that is associated with CD4 T cell count indicates that there are significant association at allele 195 and haplotypes formed by it and two flanking SNPs with T1D in NF, specifically with evidence for gender differences and paternal effects (genomic imprinting).

148

Potential pleiotropic influence of loci on chromosomes 2p, 5p, 10q, and 17q on levels of coagulation factors II and X

D. Warren(1), S. Cole(1), J.M. Soria(2), J.C. Souto(2), J. Hixson(3), J. Fontcuberta(2), J. Blangero(1), L. Almasy(1)
(1) Southwest Foundation for Biomedical Research, San Antonio, TX, USA; (2) Hospital de la Santa Creu i Sant Pau, Barcelona, Spain, (3) University of Texas Health Science Center, USA

The normal balance between the blood coagulation and fibrinolysis cascades suggests shared genetic and/or environmental factors may influence their components. To better understand these influences, levels of factors II (FII) and X (FX) were measured in randomly ascertained Mexican Americans participating in the San Antonio Family Heart Study. During coagulation, FII is converted to thrombin by activated FX; thrombin then converts fibrinogen to fibrin, a clot protein. Data were available for 449 individuals from 21 families. Additional measured covariates include age, sex, cigarette smoking, diabetes status, and diabetes medication or exogenous hormone use. Bivariate variance-components methods were used to examine the correlations between plasma levels of FII and FX. Our results showed significant positive genetic (0.41, p=0.0003) and environmental (0.51, p=0.002) correlations between the two traits. We then used a whole-genome joint multipoint linkage screen using 417 highly informative autosomal short tandem repeat markers spaced at circa 10 cM intervals to localize quantitative trait loci influencing variation in FII and FX. This yielded a maximum multipoint LOD score of 4.15 near marker D17S2193 on chromosome 17q. Significant evidence of linkage was also observed on chromosomes 2p (LOD, 3.90), 5p (LOD, 3.46), and 10q (LOD, 3.24). Whereas the 2p linkage peak overlaps the region of the γ-glutamyl carboxylase gene influencing

vitamin K-dependent proteins including FII and FX, no obvious candidate genes have been described in the other regions exhibiting significant linkage. Our results suggest possible pleiotropic effects of genes influencing FII and FX.

149

Genome-wide scan for loci influencing quantitative immune response traits in the Belem Family Study

E. Wheeler(1), E.N. Miller(1), C.S. Peacock(1), I.J. Donaldson(2), M.-A. Shaw(2), J.M. Blackwell(1), H.J. Cordell(1)
(1) Cambridge Institute for Medical Research (CIMR), Cambridge, UK; (2) University of Leeds, Leeds, UK

Here we report the results from a genome-wide linkage scan to identify genes and chromosomal regions that influence quantitative immune response traits, using multi-case leprosy and tuberculosis families from Belem in Northern Brazil. Specific IgG responses to mycobacterial antigens (PPD, MLSA and MTSA), total serum IgE, and lymphocyte stimulation index (SI) and IFN-gamma production phenotypes in response to MLSA and PPD, were measured in 16 (177 individuals) tuberculosis and 21 (173 individuals) leprosy families. The individuals were genotyped at 405 markers across the genome. The adjusted phenotypes were analysed using a variety of variance-component and regression-based methods. Several regions showed some evidence for linkage to the traits examined. These analyses highlighted a number of practical issues and problems encountered with regard to implementation of the linkage methods. A discussion of these issues and observed differences between the methods will be presented.

150

Linkage and Association Mapping for Late Onset Alzheimer's Disease (LOAD) in a Region of Chromosome 9 Identified by Linkage Analysis

H.W. Wiener(1), R.C.P. Go(1), R. Perry(1), D. Blacker(2), S. Bassett(3)
(1) University of Alabama at Birmingham, Birmingham, AL, USA; (2) MGH, Boston, MA, USA; (3) Johns Hopkins University, Baltimore, MD, USA.

A recent 9cM full genome scan of LOAD in families ascertained by the NIMH Alzheimer's Disease Genetics Initiative indicated suggestive linkage to a region on chromosome 9 (9q22). This peak was enhanced when attention was limited to pedigrees with all affected individuals having onset after age 65. We confirmed this initial result by using five additional flanking STR markers The linkage peak was enhanced in this analysis. We will report here our fine mapping results from typing an additional 18 markers with average spacing of 1 cM. We propose to use family based tests of association to investigate the possible roles of several genes in this narrowed region that may be involved in the pathogenesis of LOAD. Previous work done by our group suggests a possible role of transforming growth factor beta; the receptor for this cytokine is located in this region of chromosome 9, and will be one of our targets. Another candidate is the toll like receptor 4, implicated in infection and inflammatory processes.

151

The effects of environmental factors on myopia in an Ashkenazi Jewish population

Wojciechowski R.(1), Stambolian D.(2), Ciner E.(3), Reider L.(2), Ibay G.P.(1), Bailey-Wilson J.E.(1)
(1) Inherited Dis. Res. Branch, NHGRI, USA; (2) Dept. of Ophthalmology, Univ. of Pennsylvania, USA; (3) PA Coll. of Optometry, USA

Myopia, the most common eye condition in humans, is influenced by heritable factors and by excessive near work during childhood. The myopic phenotype is based on the quantitative measurement of ocular refraction. In order to maximize statistical power, variables that affect refraction should be taken into account in quantitative trait linkage (QTL) analyses. We analyzed the effects of age, sex and self-reported visual activity during childhood on refractive errors in 63 Ashkenazi Jewish families using linear regression models and generalized estimating equations. Refractions ranged from −15.88 to +6.75 diopters (D). The mean refraction was -3.46 ± 3.26 D. Males were, on average, 1.97 D (95% CI: 1.51–2.43) more myopic than females. Greater-than-average near vision activity during childhood was associated with an average of 0.86 D (95% CI: 0.22–1.51) more myopia. Participants who reported more-than-average distance activity during childhood were, on average, 0.99 D (95% CI:0.33–2.31) less myopic than those who reported less frequent distance tasks. After adjusting for age, the mean refraction was 2.28 D (95%CI: 1.77–2.79) more myopic in males and 0.83 D (95% CI: 0.25-1.42) more myopic among participants with greater-than-average near vision activity during childhood. These important variables will be incorporated into future QTL analyses of refractive error.

152

Influences of Lifestyle Habits and P53 Codon 72 and P21 Codon 31 Polymorphisms on Gastric Cancer Risk in Taiwan

M.T. Wu(1), M.C. Chen(1), D.C. Wu(2)
(1) Graduate Institute of Occupational Safety and Health and Department of Occupational Medicine, Kaohsiung, Taiwan; (2) Department of Gastroentology, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan

Influences of lifestyle habits and p53 codon 72 and p21 codon 31 polymorphisms on the risk for developing primary gastric cancer were examined in 89 gastric adenocarcinoma cases (51 males, 38 females) and 192 controls (106 males, 86 females) in a hospital-based, case-control study in Taiwan. In the final regression model, Helicobacter pylori infection and substance use (cigarette

smoking, areca chewing) were significant predictors of risk for developing gastric cancer. Compared with subjects negative for H. pylori infection, positive subjects were 3.65-fold (95% CI=2.07–6.42) more likely to develop gastric cancer. Compared with non-smokers or non-chewers, subjects with more than a 15 pack-year history or more than a 498 betel-year history (about 20 betel quids/day for 25 years) were 2.27- and 4.86-fold more at risk (95% CI=1.06–4.84 and 1.20–19.74), respectively. Frequencies of arg/arg, arg/pro and pro/pro in p53 were 11 (12.4%), 53 (59.5%) and 25 (28.1%) in carcinoma cases and 40 (20.8%), 95 (49.5%) and 57 (29.7%) in control cases, respectively. Frequencies of arg/arg, ser/arg and ser/ser in p21 were 26 (29.2%), 36 (40.5%) and 27 (30.3%) in carcinoma cases and 49 (25.5%), 94 (49.0%) and 49 (25.5%) in control cases, respectively. Neither p53, nor p21 polymorphisms were significantly different in cases and controls (P=0.16 and P=0.41, respectively). Results remained insignificant after dichotomizing with respect to cigarette smoking, areca chewing and H. pylori infection. In summary, our data indicate that in Taiwan, H. pylori infection, smoking and areca chewing are significant risk predictors for developing gastric cancer. p53 codon 72 and p21 codon 31 genotypes did not modify these risks.

153

Robustness of Resampling Methods to Reduce Selection Bias of Genetic Effect Estimates

L.Y. Wu(2), L. Sun(1,3), S.B. Bull(1,2)
(1) Public Health Sciences, University of Toronto; (2) Samuel Lunenfeld Research Institute; (3) Hospital for Sick Children

Locus-specific effect estimates from genomewise scans are subject to upward selection bias because of stringent test criteria adopted in the scan (Goring et al., 2001). Sun and Bull (2003, 2004) proposed three resampling-based estimators that can be applied to the original sample and successfully demonstrated effective bias reduction in analytic and simulation studies of a homogenous population with a single disease gene. The goal of the current study is to examine the robustness of genetic effect estimation in the presence of locus heterogeneity, by extending their simulation scheme to a mixture of two populations with differing genetic loci. We simulate a genomewide linkage scan conducted via allele sharing methods with affected sib pairs. The test statistic is the maximum NPL score exceeding criterion of the genomewide significance and the genetic effect is the expected excess allele sharing. We consider five different population mixtures, in which two disease genes are placed on two different chromosomes with different genetic effects. We examined 10×10 and 20×10-fold cross-validation as well as bootstrap resampling of 100 and 200 samples. Bias and root mean squared error are estimated overall and with stratification by true and false positive localization. Although the three estimators perform slightly differently under true or false positives, the overall results indicate that the three resampling-based estimators effectively reduce the upward selection bias under a two-locus heterogeneity model.

154

A Trend Test for Association in Case-Control Studies with Pedigree Data

X. Wu(1), R.S. Cooper(1) and M.S. McPeek(2)
(1) Department of Preventive Medicine and Epidemiology, Loyola University Chicago Medical Center, Maywood, IL; (2) Department of Statistics, The University of Chicago, Chicago, IL

Association studies are important tools in identifying the genes responsible for human complex diseases. They are often used for evaluating candidate genes obtained through previous linkage studies or biochemical pathways. Compared with family-based association studies, case-control studies are more efficient although care has to be taken to avoid higher false positive rate caused by population stratification. In many cases, only related samples are available for association studies such as samples collected for the initial linkage studies. Neglecting such correlations among individuals could cause seriously spurious associations. Several methods have been developed for case-control studies in related subject. Slager and Schaid (AJHG, 68:1457–1462, 2001) proposed a statistical test based on the Armitage trend test and included a variance that accounts for family relationship. Catherine et al. (AJHG, 73:612–626, 2003) proposed a quasi-likelihood score (QLS) function and constructed a QLS test for allelic association based on the known genealogy. This method is especially useful for complex pedigrees, such as large inbred population. However, it is allele based and requires Hardy-Weinberg equilibrium (HWE) in the founders. Here, we extend this method and propose a trend test based on quasilikehood score function. This new method doesn't require HWE and can be used in complex pedigrees when exact-likelihood calculation is not possible.

155

Defining the lung phenotype for modifier gene studies of cystic fibrosis: longitudinal study of lung function decline

W. Xu, M. Corey

The Hospital for Sick Children and University of Toronto, Canada

Forced expiratory volume in one second (FEV1) is the lung function variable that best describes the long-term course of lung disease in cystic fibrosis, and is a surrogate for ultimate survival. However there is great variation in FEV1 among patients with the same CF genotype. This study aims to redefine the lung function phenotype for most efficient genetic analysis in The Canadian Cystic Fibrosis Modifier Gene Study. Patients in the Toronto CF database were studied if born between 1960 and 1989 and diagnosed under 10 years of age. Mixed model regression analysis of FEV1 % predicted versus age included those with at least 2 pulmonary function tests. Follow-up was truncated to

provide equivalent periods in each of 6 5-year birth cohorts. The course of FEV1 was similar in the first 3 cohorts, with intercept of 92% predicted at age 5 and decline of about 2.5% predicted per year. Rate of decline improved in the last 3 cohorts, to about 1% predicted per year. However the intercept value at age 5 was markedly lower than in the earlier cohorts, perhaps due to better survival to school age of severely affected children in later cohorts. These patterns were not altered when other covariates (sex, age at diagnosis, pancreatic status) were included in regression models. Models incorporating pulmonary infection status showed Burkholderia cepacia was a major factor in the rate of decline of FEV1, whereas the more common Pseudomonas aeruginosa was a significant predictor of intercept at age 5. A composite lung function phenotype will provide more reliable genetic analysis.

156

Evaluation of risk factors for nasopharyngeal cancer in a family study in Taiwan

X. Yang(1), S. Diehl(2), R. Pfeiffer(1), C.J. Chen(3), W.L. Hsu(3), Y.J. Cheng(3), B. Sun(4), A.M. Goldstein(1), A. Hildesheim(1)
(1) DCEG, NCI, NIH, DHHS, USA; (2) Center Pharmacogenomics & Complex Disease Research, UMDNJ, USA; (3) College Public Health, Natl Taiwan Univ, Taiwan; (4) Westat Inc, MD, USA

A study of Nasopharyngeal cancer (NPC) families with Ý2 affected members was conducted in Taiwan (274 families with 2,460 subjects, 502 affected and 1,958 unaffected) to determine the association between familial NPC and potential etiologic factors. Similar to results from a previous case-control study in Taiwan, Guangdong salted fish intake during childhood, exposure to wood, having a CYP2E1 RSA allele, and betel nut use were all associated with higher NPC risk using conditional logistic regression(CLR), although these associations were not as strong as in the case-control study possibly due to the shared environment among family members. Risk associated with wood exposure and salted fish intake was stronger in families with early NPC age-onset(ORwood=3.2, 95% CI=1.7–6.2; ORfish=2.1, 95% CI=1.1–3.9) or Ý3 affected members(ORwood=4.9, 95% CI=1.7–14.0; OR fish=4.4, 95% CI=1.1–19.0). In contrast, betel nut use and RSA allele carrier status appeared to have stronger effects in families with late age-onset(ORbetel=1.9, 95% CI=1.2–2.9; ORrsa=1.6, 95% CI=0.98–2.5) and <3 affected members(ORbetel=1.7, 95% CI=0.9–3.2; ORrsa=2.3, 95% CI=1.2–4.3). To better adjust for degree of relationship among family members, we also calculated ORs using a variance component model. The results from the two methods were similar indicating that the risk estimates from CLR were unbiased.

157

The gene (NOS3) and environment (smoking) interaction affects ACE inhibitor induced cough in Chinese Hypertensives

D. Lee(1), X. Su(2), J. Lu(1), X. Li(2),Y. Hu(1), S. Zhan(2), W. Cao(1), L. Mei(2), Y. Tang(2), D. Wang(2), L. Lee(1), H. Yang(2)
(1) School of Public Health, Peking University, Health Science Center; (2) Genetic Epidemiology Program, Cedars-Sinai Medical Center, Los Angeles, CA.

Cough is a common adverse effect of Angiotensin-converting enzyme inhibitors (ACEI) used in hypertension therapy. The incidence of ACEI induced cough has exhibited ethnic difference, with the highest frequency in Asian populations. To identify genetic factors related to this adverse effect, we tested NOS3 and BKDRB2 genes in a Chinese hypertensive sample.
A case-control sample (cough case=254, control=647) was selected from 1831 hypertensive patients who were ascertained from a Chinese community-based postmarketing surveillance and underwent treatment with the ACEI benazepril for 3 years. Four SNPs were genotyped for each of NOS3 and BDKRB2. Univariate and multivariate analyses were performed for single SNP as well as haplotypes. Interactions between genes and between gene and environmental factors (smoking and drinking) were evaluated by both case-control and case only methods.
Only common haplotypes (>5%) were evaluated (4 from each gene). NOS3-H2 had a protective effect on ACEI induced cough (OR=0.63, 95% CI=0.46∼0.88, p=0.006). When gene-smoking interaction was evaluated, a dose-response relationship between the number of NOS3-H1 and cough was observed in smokers only. The ORs for H2/H2, H2/X as compared with X/X carriers were 5.6 (95% CI=2.1∼15.2, p=0.002) and 3.3 (95% CI=1.3∼8.3, p=0.01), respectively. Such gene-smoking interaction was consistently observed in both case-control and case-only analyses and with or without covariates. BDKRB2-H1 showed an increased risk for cough (p=0.01). Gene-gene interaction between NOS3 and BDKRB2 was observed in case-only analysis, but not in case-control analysis.
NOS3 haplotypes are significantly associated with the risk for ACEI induced cough. There is a significant synergistic effect between NOS3 and smoking. Thus, a haplotype approach and gene-environment interaction evaluation has revealed genetic contributions to adverse effects.

158

The IGF-I genotype is a predictor of survival in subjects with type 2 diabetes and a prior myocardial infarction

M. Yazdanpanah(1), I. Rietveld(1,2), J.A.M.J.L. Janssen(2), O.T. Njajou(1), A. Hofman(1), T. Stijnen(1), H.A.P. Pols(1,2), S.W.J. Lamberts(1), J.C.M. Witteman(1), C.M. van Duijn(1)
(1) Dept. of Epi & Biostat, Erasmus Medical Center Rotterdam, The Netherlands; (2) Dept. of Internal Medicine, Erasmus Medical Center Rotterdam,The Netherlands

The risk of mortality from cardiovascular disease (CVD) is markedly increased in diabetes compared to non-diabetic subjects. We investigated whether this IGF-I gene polymorphism also influences survival. The relation

between this IGF-I gene polymorphism and survival was assessed in a population-based study of 4765 subjects from the Rotterdam study. We investigated survival among 4186 non-diabetic subjects without prior MI, 199 non-diabetic subjects with prior MI, 350 diabetic subjects without prior MI, and 30 diabetic subjects with prior MI during 14 years follow up. We observed no differences in survival time between subjects with the wild type of the IGF-I genotype and the variant carriers in the total population (p=0.3). The survival of the variant carriers resembled that of the whole study population. However, we found a strong reduction in survival in diabetic variant carriers with prior MI. After adjustment for known risk factors as sex, age, cholesterol, HDL-cholesterol, systolic blood pressure, diastolic blood pressure, BMI, WHR and smoking, variant carriers with type 2 diabetes and prior MI had a 4 times higher risk to die than subjects with the wild type (hazard ratio: 4.40; 95% CI=0.96–22.24; p=0.05). In conclusion the IGF-I genotyping of a diabetic subject with a prior MI may help to predict the future risk to die.

159

A More Powerful Transmission/Disequilibrium Test Based on Haplotype Similarity

K. Yu and M. Province
Division of Biostatistics, School of Medicine, Washington University, St. Louis, MO 63110, USA

Taking advantage of increasingly available high-density single nucleotide polymorphism (SNP) markers within genes and across genome, various types of transmission/disequilibrium tests (TDT) using haplotype information have been developed. A practical challenge arising in such studies is the possibility that transmitted haplotypes have inherited disease-causing mutations from different ancestral chromosomes (founder heterogeneity). To reduce the effect of founder heterogeneity, Yu et al. (Genetic Epidemiology, in press) proposed a sequential peeling procedure in the context of population based case-control studies. Here we extent that approach to family based studies. The method is applicable to nuclear families with multiple affected sibs, and with ambiguous phase information. Through simulation studies, we demonstrate that the method has the correct type I error rate in structured population, and higher power than some existing haplotype based TDT.

160

The effect of low frequency genotype error rates on haplotype analysis in collections of cases and controls

D. Zabaneh, A. Green, A. Carey, I.J. Mackay
Oxagen Ltd., Abingdon, Oxon, UK

Case-control designs are being used extensively in candidate gene studies. If blood or DNA samples of cases and controls differ in quality, for example as a result of different collection protocols, or because of variation in post-collection handling regimes, then it is possible that genotype error rates will also differ between cases and controls. For single SNPs, provided errors occur at random with respect to genotype and are of low frequency, the consequences for type I and type II error rates are not grave. However, using simple examples, we demonstrate that the effect of low frequency genotype error on haplotype analysis can be very serious.
By simulating errors on real data sets, we study to what extent this may be a problem in practice. Moreover, we present evidence that undetected genotype errors, occurring at low frequency and differing between cases and controls, are a genuine problem which needs to be addressed in haplotype analysis in case-control studies.

161

High resolution LD mapping of type 2 diabetes susceptibility loci on chromosome 1q

E. Zeggini(1), W. Rayner(1), C. Groves(1), S. Wiltshire(1), G. Hitman(2), M. Walker(3), A. Hattersley(4), L. Cardon(1), P. Whittaker(5), S. Hunt(5), P. Deloukas(5), M. McCarthy(1), for the 1q Consortium
(1) WTCHG and OCDEM, Oxford; (2) Queen Mary, London; (3) University of Newcastle; (4) Peninsula Medical School, Exeter; (5) Wellcome Trust Sanger Institute, UK

Type 2 diabetes (T2D) is a common multifactorial disease attributed to complex interactions between genetic and environmental factors. In recent years, whole genome screens (WGS) have allowed prioritisation of regions implicated in T2D susceptibility. Evidence for linkage across several populations has been obtained for chromosome 1q21-24, replicated in 9 different WGS examining T2D or related traits. The international 1q consortium is undertaking high resolution LD mapping of the implicated 13.5Mb interval across 7 of the linked populations. As part of the 1q consortium, 450 UK Caucasian probands from the Warren 2 WGS and 450 ethnically matched controls have been thus far genotyped for 1536 SNPs using the Illumina platform (median spacing 7kb). Single-point and haplotype-based analyses, using haplotype trend regression, have highlighted among other regions, a notable cluster of T2D-associated SNPs at approximately 153.4Mb. 151 SNPs retained significance after imposing a false discovery rate threshold of 0.05. The average LD correlation coefficient r^2 value for adjacent markers was 0.36. Block-like patterns of LD were observed throughout the region examined, though, as expected, precise boundaries were dependent on the block assignment method used. A further set of 1536 SNPs are being genotyped in the region under scrutiny.

162

Multiple marker information content for admixture mapping

X. Zhu, R.S. Cooper, S. Zhang

Recently several admixture mapping methods have been proposed that use the linkage disequilibrium arising from the recent admixture of two or more genetically distinct

populations. To complement these efforts, a high-density genome admixture map (Smith et al. 2004, AJHG) has also been created. The multiple marker information content of the map panel is crucial to the power of admixture mapping. In this report, we studied the multiple marker information content of this map through simulations for situations where the marker allele frequencies in parental populations are both known and unknown. We found that almost half of the markers could be dropped without substantial loss of information. The multiple marker information content is essentially the same as a comparison of known with unknown marker allele frequencies in parental populations. The effect of multiple marker information on power of admixture mapping will also be discussed.

163

Multivariate genetic analysis of chronic pelvic pain and associated phenotypes

K.T. Zondervan(1), L.R. Cardon(1), S.H. Kennedy(2), N.G. Martin(3), and S.A. Treloar(3)
(1) Wellcome Trust Centre for Human Genetics, Univ. of Oxford, UK; (2) Nuffield Department of Obstetrics and Gynaecology, Univ. of Oxford, UK; (3) Queensland Institute for Medical Research, Brisbane, Australia.

Chronic pelvic pain (CPP), a condition of uncertain aetiology involving long-standing pain localising to the lower abdomen and pelvis, is very common in women of reproductive age. It is difficult to diagnose and treat, and causes a substantial burden of functional disability. CPP has been suggested as a phenotype worth pursuing per se for gene discovery studies. Although heritability estimates have been published for some conditions potentially underlying pelvic pain, the heritability of CPP itself has never been investigated. We analysed data from 623 MZ and 377 DZ female Australian twin pairs aged 29-50 years. Univariate twin pair (tetrachoric) correlations of rMZ=0.43 and rDZ=0.11 plus subsequent univariate model-fitting suggested that genetic influences are important in the etiology of CPP. A heritability estimate of 0.41 (95% CI: 0.25–0.56) was found. Nevertheless, results of multivariate Cholesky decomposition model-fitting incorporating additional significantly correlated phenotypes questioned the value of CPP as a useful independent phenotype on which to conduct genetic studies; contributing conditions such as endometriosis and individual variation in perception of pain are likely to provide more useful phenotypes.

164

A novel screening test for quantitative multivariate linkage analysis

M. de Andrade(1), C. Olswold(1), S.T. Turner(2)
(1) Health Sciences Research and (2) Hypertension, Mayo Clinic, USA

Multivariate quantitative linkage analysis provides an approach to identify genes influencing two or more correlated traits. A major advantage over separate univariate analyses is the greater statistical power to identify loci whose effects are too small to be detected in single-trait analyses. Recently we published the first trivariate genome scan for quantitative linkage analysis using blood pressure measures and body mass index data from the Rochester Family Heart Study. Only one region on chromosome 10 showed strong evidence of linkage (LOD=5.10, $p < 0.000044$). Since the multivariate linkage analysis for quantitative traits can be computationally intensive, it is also important to develop screening tests to help identifying combinations of traits with shared genes for multivariate analysis by using measures of correlation prior to multivariate linkage analysis. Thus, we propose a new test to investigate whether there is a single gene responsible for a particular combination of traits (pleiotropy). This test is simple, fast, and uses the QTL variance component estimates from the univariate quantitative linkage analyses. We applied this test using the univariate linkage analysis results from blood pressure measures and body mass index data from the Rochester Family Heart Study, and we observed a strong agreement between our test and the trivariate quantitative linkage analysis. The same concordance was also observed using simulated data.

165

A whole genome scan for 24-hour respiration rate

E.J.C. de Geus(1), D. Posthuma(1), H.M. Kupper(1), G. Willemsen(1), A.L. Beem(1), P.E. Slagboom(2), D.I. Boomsma(1)
(1) Dept. of Biological Psychology, Free University Amsterdam, Netherlands; (2) Dept. of Molecular Epidemiology, Leiden University Medical Center, Netherlands

Identification of genes causing variation in daytime and nighttime respiration rate could advance our understanding of the basic molecular processes in human respiratory rhythmogenesis. This also serves an important clinical purpose, because such processes have been critically implied in sleep disorders. Using ambulatory monitoring, average respiration rate was obtained during three daytime periods (morning, afternoon, evening) and during sleep. We examined the heritability of the respiration rates using an extended twin design (MZ, DZ, singleton siblings). Ensuing, we performed a sib-pair based linkage analysis using data from 307 sibling pairs. Sibs were genotyped on 374 markers on the autosomes with an average distance of 9.65 cM.
Heritability of respiration rate during the daytime was moderate (41%–50%) whereas heritability at night was very high (81%). Variance components based multipoint model-free linkage analysis yielded evidence for linkage ($LOD > 2$) at chromosomes 3, 7, 10, and 22. Strongest evidence was obtained for respiration rate during sleep. A number of viable positional candidate genes were identified in the regions of linkage that can might influence respiratory control during sleep.

166

A genome-wide scan of blood pressure in 2,218 Caucasian twins suggests novel linkage to chromosome 11, and accurate replication of loci on chromosomes 16, 17 and 22

M. de Lange(1,2), T.D. Spector(1) & T. Andrew(1)
(1) Twin Research & Gen Epi Unit, St Thomas' Hospital, London, UK; (2) Dept of Med Epi & Biostat, Karolinska Institutet, Stockholm, Sweden

Hypertension is a vascular risk factor and is influenced by polygenic and multiple environmental factors. Several genomic studies have found suggestive LOD scores for either blood pressure or essential hypertension, but surprisingly few loci have been replicated in different populations. In this study we performed a genome-wide linkage analysis for systolic (SBP) and diastolic (DBP) blood pressure on 1,109 Caucasian dizygotic twin pairs.
Multipoint linkage analysis accurately replicated the locations of three previously reported linkage peaks; on chromosome 16 at 65cM (LOD 0.8 for SBP and 1.8 for DBP); on chr 17 at 70cM (LOD 1.8 SBP) and at 35cM on chr 22 (LOD 0.97 SBP and 0.99 DBP). Results from multipoint analysis showed one novel suggestive linkage for SBP (multipoint LOD 2.28, two-point p=0.0007) at 35cM on chromosome 11. Results were similar when those on BP medication were excluded.
These are encouraging results for hypertensive research and demonstrate that despite past disappointments, linkage studies can be used to replicate regions from other studies and potentially discover new genetic risk factors of moderate to large effect size. Considering the differences in selection and ascertainment of the previous linkage studies, these results also suggest that some QTLs are likely to influence both the normal range of blood pressure and clinical hypertension, while others will be specific to each trait.

167

The usefulness of genetic testing in predicting risk of complex diseases

C.M. van Duijn(1), Y.S. Aulchenko(1), S. Elefante(1), G.J.J.M. Borsboom(2), E.W. Steyerberg(2), A.C.J.W. Janssens(2)
(1) Department of Epidemiology & Biostatistics, Erasmus MC, Rotterdam, the Netherlands; (2) Department of Public Health, Erasmus MC, Rotterdam, the Netherlands

Complex diseases cannot be predicted with single genes, but there is ongoing debate on the future of multiple genetic tests. We evaluated the discriminative ability of tests including up to 400 susceptibility genes using the area under the ROC curves (AUC). We simulated three populations of 100,000 subjects in which the disease prevalence varied from 1%, 10% and 30% using R software. Allele frequencies of the risk allele varied from 10%–50% and odds ratios (ORs) from 1.25 to 3 (multiplicative effects). The number of genes involved and strength of their association with disease determined the AUC. When the frequency of disease and risk alleles are both 10%, at least 7 genes with strong effects (OR=3.0) are needed in the tool to obtain an AUC of 0.80. When all ORs are 2.0, at least 18 genes are needed to obtain an AUC of 0.80, while 60 genes are needed when ORs are 1.5, and 170 when all ORs are 1.25. Also one can reach an AUC of 0.80 by combining a few strong genes with multiple weak genes, which is a more realistic scenario for most complex diseases. The maximum discriminative ability of a genetic screening tool could be deduced from the heritability and the prevalence of the disease. Based on our simulations studies it is unlikely that we will be able to predict common diseases with a modest heritability such as cardiovascular disease. However, predictions appear to be possible for diseases like Alzheimer's disease.

168

Familial Aggregation of Ischemic Stroke in a Genetically Isolated Population

van Rijn M.J.E.(1), Slooter A.J.C.(1,4), Schut A.F.C.(1), Snijders P.J.L.M.(1), Kappelle L.J.(4), van Swieten J.C.(2), Oostra B.A.(3), van Duijn C.M.(1)

From the dept. of (1) Epidemiology & Biostatistics; (2) Neurology and (3) Clinical Genetics, Erasmus Medical Center, Rotterdam; (4) Dept. of Neurology, University Medical center, Utrecht, The Netherlands

There is increasing interest in the genetics of ischemic stroke. We aimed to study the familial aggregation of ischemic stroke subtypes in a genetically isolated population in the Netherlands. Patients were recruited by the general practitioners. The diagnosis was confirmed by two neurologists. Stroke-free individuals were added as controls. Kinship coefficients and inbreeding coefficients were calculated. We ascertained 91 patients with ischemic stroke and 278 controls. 48 were classified as large-vessel stroke, 40 as small-vessel stroke and 3 as stroke of undetermined etiology. 50 patients could be connected to a common ancestor within 7 to 9 generations. The percentage of unrelated pairs was significantly higher for controls compared with all other groups. The percentage of unrelated pairs was significantly higher for small-vessel compared with large-vessel stroke and for late-onset compared with early-onset stroke. There was significantly more inbreeding in early-onset stroke patients compared with late-onset patients. We found evidence for familial aggregation of ischemic stroke. Familial aggregation was strongest for large-vessel and early-onset stroke patients. Further, we found significantly more inbreeding in early-onset stroke patients compared with late-onset patients, suggesting a recessive mutation might be involved in this subgroup.

169

Evaluation of Nyholt's Multiple Testing Method

D. Salyakina, S.R. Seaman, B. Müller-Myhsok
Max-Planck-Institute of Psychiatry, Munich, Germany

Studies of association between genetic polymorphisms and phenotypic characteristics may require a large number of statistical tests. This leads to a multiple testing problem. Several procedures exist for resolving this problem, but they are desined mostly for independent tests. The best alternative for the analysis of correlated variables such as markers in linkage disequilibrium is a permutation based method, which is computationally intensive. In a recent report, Nyholt (2004) provided a new approach for correcting for multiple testing in association studies. He adapted the method of Cheverud (2001) for linkage studies. Principal-component analysis or, more generally, spectral decomposition of the matrix of pairwise LD between SNPs is used to estimate an effective number of independent tests that is then used in a Sidak correction procedure. We investigated the performance of this method on candidate genes typed for several association studies in our Institute. We considered 31 candidate genes genotyped in 1360 persons, with altogether 291 SNPs and an average distance of 7.8 kilobases between SNPs in the same gene. We simulated independent phenotypes for the actual genetic dataset to compare the Type-I error rates of Nyholt's method with that of a minimum p permutation approach. We shall present the results of this comparison.

170

Rapid Simulation of P-values for Product Methods and Multiple-Testing Adjustment in Association Studies

Seaman S.R. and Mueller-Myhsok, B.
Max-Planck Institute of Psychiatry, Munich, Germany

A major aim of association studies is the identification of polymorhisms, usually single polymorphisms (SNPs) associated with a trait. Tests of association may be based on individual SNPs or on sets of neighbouring SNPs, using, for example, a product method (Zaykin et al., 2002; Dudbridge and Koeleman, 2004) or Hotelling's T test. Linkage disequilibrium, the non-independence of SNPs in physical proximity, causes problems for all these tests. Firstly, multiple-testing correction for individual-SNP tests or for Hotelling's T tests leads either to conservative p-values (if Bonferroni correction is used) or is computationally expensive (if permutation is used). Secondly, calculation of product p-values usually requires permutation. We present the Direct Simulation Approach (DSM), a method that accurately approximates p-values obtained by permutation, but is much faster. It may be used whenever tests are based on score statistics, for example Armitage's trend test and Hotelling's T test. The DSM allows for binary, continuous or count traits and permits adjustment for covariates. We demonstrate the accuracy of the DSM on real and simulated data and illustrate how it might be used in the analysis of a whole genome association study.

171

Regulatory promoter variants of the PTGDR gene are protective for asthma

L.J. Palmer(1), T. Oguma(2), E. Birben(2), L.A. Sonna(2), K. Asano(3), C.M. Lilly(2)
(1) Western Australian Institute for Medical Research, University of Western Australia, Perth, Australia; (2) Brigham and Women's Hospital and Harvard Medical School, Boston, MA; and (3) Department of Medicine, Keio University School of Medicine, Tokyo, Japan.

Background: Mechanistic studies suggest that the prostanoid DP receptor (PTGDR) system is required for the expression of the asthma phenotype.
Methods: Combinations of genetic variants that influence PTGDR transcription were identified and tested for disease association in case-control studies of 518 European-American asthmatics and 175 healthy subjects and in 80 African-American asthmatics and 45 controls. Molecular haplotyping was performed.
Results: We identified 4 novel and 2 previously reported SNPs in the PTGDR gene that defined 5 common 3-SNP haplotypes, which varied in their ability to support transcription of the PTGDR gene and have distinct DNA-binding-protein affinity profiles. Individual SNPs in the PTGDR gene showed significant association with asthma in both populations. Multivariate analysis of the haplotype combinations (diplotypes) present in individuals demonstrated that both European-Americans (OR=0.55, P=0.002) and African-Americans (OR=0.32, P=0.03) bearing at least one copy of the low-transcriptional-efficiency haplotype had lower odds of an asthma diagnosis. Asthma risk was proportional to the transcriptional activity of each haplotype.
Conclusions: Our functional and genetic findings identify the PTGDR gene as a novel asthma susceptibility gene.

172

Genome-wide Meta-analytic Methods Applied to Rheumatoid Arthritis

C.J. Etzel(1), D. Jawaheer(2), M.F. Seldin(3), W.V. Chen(1), P. Gregersen(2), C.I. Amos(1) for NARAC(4)
(1) Department of Epidemiology, UT MD Anderson Cancer Center, Houston, TX; (2) Division of Biology and Human Genetics, North Shore University Hospital, Manhasset, NY; (3) Department of Biological Chemistry, UC Davis, Davis, CA; (4) North American Rheumatoid Arthritis Consortium

In this project, we completed meta-analyses (MA) applied to 4 Rheumatoid Arthritis studies: two studies from NARAC (Jawaheer et al. 2001, 2003) and two studies from the UK (MacKay et al. 2002; Eyre et al. 2004). For each study, we obtained NPL scores by performing interval mapping (2 cM intervals) using GeneHunter2 (Markianos et al 1999; Kruglyak et al. 1996). All affected individuals were included in the analysis. The marker maps differed among the two consortium groups, therefore, the marker maps were aligned after the interval mapping was completed and the NPL scores that were within 1 cM of each other were combined using the method of Loesgen et al. (2001) by calculating the weighted average of the NPL score. This approach avoids some problems in analysis encountered by using GeneHunter when some markers in the sample are not genotyped. This procedure provided

support for a region on chromosome 18, which is being studied further by SNP association analysis. We also present an alternative meta-analytic method that generalizes the method first proposed by Etzel and Guerra (2002). In this method, alignment of the NPL scores is not necessary since the weights of the weighted average MA estimator include an adjustment for location (in cM).

173

Semiparametric estimation of haplotype-disease association with unphased genotype data in family-based studies

G.A. Satten(1), A.S. Allen(2) and A.A. Tsiatis(3)
(1) Centers for Disease Control and Prevention, Atlanta, GA USA; (2) Duke University, Durham NC USA; (3) North Carolina State University, Raleigh NC USA

Using results from coarsened-data semiparametric model theory, we develop tests and estimators of the effect of haplotypes on disease status for family-based association studies. Our approach gives unbiased estimators and tests irrespective of agreement between the assumed and true parental haplotype distribution; if the parental haplotype distribution is properly specified, our approach is optimal (i.e., gives estimators with lowest variance) among methods that do not make restrictive assumptions on parental haplotype distributions. We demonstrate our approach and compare it to alternative approaches using simulated data from a case-parent trio study.

174

Mapping Genes of Complex Diseases in Daghestan Primary and Secondary Genetic Isolates

Kazima B. Bulayeva
Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow, Russia

Genetic isolates that provide outstanding opportunities for identification of susceptibility genes for complex diseases can be classified as primary, with ancient demographic history in a stable environment, and as secondary, with younger demographic history (Neel, 1992). Daghestan contains 26 out of 50 indigenous Caucasus ethnicities that have been in existence for hundreds of generations in same highland region. The ethnics are subdivided into numerous primary isolates. The founder effect and gene drift in these primary isolates caused aggregation of specific haplotype with limited numbers of pathogenic alleles and loci in some isolates compared to others. These expressed as inter-population differences in lifetime prevalence and features of certain complex clinical phenotypes and in patterns of genetic linkage and linkage disequilibrium. Stable highland and ethnic-cultural environments lead to increased penetrance and reduced number of phenocopies, which hamper the identification of any susceptibility genes for complex diseases. Owing to these characteristics of the primary isolates, a comparative linkage study of the same complex disease in the primary isolates allow us to define the number of susceptibility genes for any complex disease and to identify the source of variability and non-replication of linkage analysis results. As part of an ongoing study, seven extended schizophrenia kindreds have been ascertained from Daghestan isolates. Lifetime morbid risk for schizophrenia in the isolates varies from 0 to 5%. An important clinical finding is that certain Daghestan primary genetic isolates tended to have certain types of schizophrenia. For instance, the isolate #6022 is characterized by aggregation of the disorganized type of schizophrenia and a relatively early average age of onset (20.8 + 1.51 years) with no cases of suicide. Another isolate #6007 is characterized by aggregation of paranoid schizophrenia and by a relatively later average age of onset (24.0 + 2.35 years). A genome scan with markers spaced 10 cM apart (Weber/CHLC 9) was carried out on these pedigrees and linkage analysis was performed using descent graph methods, as implemented in Simwalk2. To identify regions containing susceptibility genes within these kindreds, we followed up those regions with NPL statistics and next parametric linkage analysis with the choice of genetic model guided by the results obtained in the NPL. The paper presents results of genome-wide multipoint linkage analysis of schizophrenia in two primary and two secondary genetic isolates of schizophrenia.