10

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

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

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

11

Excess of early-onset cancers in relatives of neuroblastoma patients

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

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

12

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

RB Martin and CJ MacLean
Virginia Commonwealth University, USA

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

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

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

13

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

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

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

14

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

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

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

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

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

15

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

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

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

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

16

A likelihood-based approach to linkage disequilibrium testing

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

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

17

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

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

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

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

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

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

18

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

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

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

19

Genetic epidemiology of the atherogenic Lipoprotein(a)

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

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

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

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

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

20

Genomic screen for linkage in a family with autosomal dominant chordoma

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

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

21

Bootstrap confidence intervals for relative risks in ASP data

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

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

22

Dissecting complex diseases with FINESSE

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

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

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

23

Familial risks in cancers from a Family-Cancer Database

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

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

24

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

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

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

(supported by convention INSERM/MSD)

25

Risk Models for Familial Breast and Ovarian Cancer

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

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

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

26

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

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

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

27

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

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

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

28

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

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

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

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

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

29

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

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

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

30

Designing a linkage replication study in affected sib pairs

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

The Maximum Likelihood Score (MLS) (Risch, 1990) enables the hypothesis of linkage between a disease locus and marker loci to be tested in a sample of affected sib pairs. Systematic screening of the genome is prone to false positives due to multiple testing. To avoid wrong conclusions, one may choose a stringent threshold for concluding to linkage, but then decreasing the power to detect susceptibility factors. An alternative is to choose a lax threshold in the first sample, then to study a second sample only for the retained regions. In this second study, the threshold Tα corresponding to a given type I error α will depend on the number of regions re-tested.

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

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

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

31

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

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

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

32

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

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

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

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

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

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

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

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

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

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

33

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

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

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

34

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

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

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

35

Increasing the power of family-based association studies

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

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

36

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

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

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

37

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

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

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

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

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

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

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

38

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

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

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

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

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

39

Case-Only Design to Measure Gene-Gene Interactions

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

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

40

A parametric copula model for analyzing familial binary data.

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

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

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

41

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

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

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

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

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

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

References:

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

42

A proposal for a collaborative pedigree database.

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

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

43

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

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

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

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

44

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

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

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

Supported by AISM and by ISS.

45

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

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

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

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

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

46

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

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

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

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

47

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

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

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

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

48

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

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

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

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

49

A genetic epidemiologic study of radiological osteoarthritis

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

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

50

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

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

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