Abstracts for the GenomEUtwin Scientific Meeting

Rome, 13–14 December, 2004

The GenomEUtwin-Project is supported by the European Commission under the Quality of Life and Management of the Living Resources program of the 5th Framework Program (no. QLG2–CT–2002–01254) www.genomeutwin.org

VARIATIONS IN THE THROMBOMODULIN GENE PREDISPOSE FOR CARDIOVASCULAR EVENTS SUCH AS STROKE AND CORONARY HEART DISEASE IN A FINNISH CASE–COHORT STUDY FINRISK

K. Ahonen,[1] K. Komulainen,[1] M. Alanne,[1] M. Levander,[1] K. Silander,[1] S. Kulathinal,[2] L. Peltonen, [1] V. Salomaa, [2] and M. Perola[1]

[1] *Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland*

[2] *Department of Epidemiology and Health Promotion, National Public Health Institute, Helsinki, Finland*

Background. Thrombomodulin is an important endothelial anticoagulant protein that decreases thrombin activity and activates protein C. The protein, which is encoded by a single exon, is an endothelial-specific type I membrane receptor that binds thrombin. Mutations in the thrombomodulin gene have previously been associated with coronary heart disease (CHD) and inherited thrombophilia.

Methods and study design. For the case–subcohort study design, we studied 655 cases and 381 members of the subcohort from the Finnish population survey FINRISK 92, which is part of the MORGAM study. Among the 945 individuals studied, 255 were classified as coronary cases, 30 as stroke cases, and 219 as having cardiovascular disease already at the beginning of the follow-up period (baseline CVD). We selected five single nucleotide polymorphisms (SNPs) in the thrombomodulin gene, covering a total of 11,910 bases on chr 20p11.21. These SNPs were selected to capture the major haplotypes and to form a tight locus map throughout the gene. The SNPs were genotyped using the MassARRAY System. We evaluated whether genotypes and haplotypes of the thrombomodulin gene predispose to cardiovascular events, such as CHD and stroke. Analysis was done with Cox proportional hazards regression analysis using SAS v. 8.1. For haplotype estimates, the Phase® program was used.

Results. In univariate analysis, genotype GG of the single nucleotide polymorphism rs6082986 was found to predispose to cardiovascular events in males with a hazard ratio (HR) of 2.3 (p = .0389) compared to the other genotypes. After adjustment for smoking, total cholesterol/HDL-cholesterol ratio, hypertension, age and diabetes, HR of 2.805 (p = .023) was obtained. Genotype TT of SNP rs604851 was found to be associated with cardiovascular events with HR = 2.21 (p = .017) in males when covariates other than baseline cardiovascular disease were included. Estimated haplotype 22222 (for SNPs rs6113909, rs6082986, rs1962, rs3176123 and rs6048519, respectively) was 14.1% more common among females with myocardial infarction either at baseline or during the follow-up period compared to control females. In time-to-event analysis, this haplotype was associated with increased risk for a cardiovascular event with a HR of 3.81 (p = .0034) in univariate analysis and with a HR of 6.36 (p = .0014) when adjusted for multiple covariates, including baseline CVD.

Conclusions. These preliminary results suggest that the thrombomodulin gene has an important role in predisposing to cardiovascular events such as stroke and coronary heart disease. In females, haplotype 22222 appears to predict the risk for cardiovascular events. In males, 2 SNPs (rs6082986, rs6048519) affect cardiovascular disease risk either in univariate analysis or when other phenotype data are included but baseline CVD excluded.

ANALYSIS OF INTERLEUKIN 6 AND FIBRINOGEN α, β AND γ GENES REVEALS A PROTECTIVE HAPLOTYPE FOR CARDIOVASCULAR DISEASE IN FIBRINOGEN GAMMA GENE

M. Alanne,[1] V. Salomaa,[2] K. Komulainen,[1] K. Ahonen,[1] K. Silander,[1] L. Peltonen,[1] J. Saarela,[1] and M. Perola[1]

[1] *Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland*

[2] *Department of Epidemiology and Health Promotion, National Public Health Institute, Helsinki, Finland*

Background. Coronary heart disease (CHD) and atherosclerosis are major causes of death in Western countries. LDL-cholesterol has been considered one of the main risk factors for the disease, but more recent evidence has shown that inflammatory process plays a role in atherosclerosis plaque development. The aim of this study is to analyze the association of CHD and candidate genes related to inflammation in a case–cohort study, FINRISK 92.

Project description. The FINRISK risk factor population surveys are conducted every five years, and are designed to assess the levels of CHD and other chronic disease risk factors in defined geographical areas of Finland. The FINRISK 92 survey is a part of the genetic prospective follow-up study MORGAM, which aims to identify genetic risk factors for CHD by analyzing relevant cascades of potentially interacting candidate genes as well as single functionally important genes. For the case-cohort study, 999 FINRISK 92 individuals were selected from among 8000 participants. We analyzed a biological pathway leading to the production of acute phase reactants by studying single nucleotide polymorphisms (SNP) in interleukin 6 (IL6 gene) and fibrinogen genes α, β and γ (FGA, FGB, FGG).

Main methods. To obtain a set of informative SNPs we selected haplotype-tagging SNPs within the coding and regulatory regions of the genes, having minor allele frequency > 5 %, and also including all functionally significant polymorphisms. The genotyping was done by allele-specific primer-extension, using either an in-house DNA chip or the MassARRAY system. Cox's proportional hazard model was used to estimate the risk of cardiovascular disease (CVD) event during the follow-up period.

Results. For our preliminary analyses we studied the association between CVD and 3 SNPs on IL6, 3 SNPs on FGA, 2 SNPs on FGB and 3 SNPs on FGG genes. Genotypes were available for 570 to 934 individuals. Significant difference in genotype frequencies between CVD cases (coronary or stroke event at the baseline or during follow-up) and subcohort members without an event was found for several SNPs. In a logistic regression model adjusted for age, sex, study area, smoking and IL6 serum concentration, IL6 SNP rs1554606 G/G genotype was independently associated with cardiovascular events (p = .008, n = 413). The risk of an event for an individual with the G/G-genotype was 2.4 times higher than for T/T homozygote individuals (95% confidence interval: 1.32–4.22). In time-to-event analysis we studied the effect of haplotypes on risk for CVD event using Cox's proportional hazard model. For FGG SNPs rs2066860 and rs1800792, CT haplotype carriers had a significantly lower risk for CVD event, with a hazard ratio of 0.129 (95% CI: 0.024–0.686, n = 56), p = .0162, compared to noncarriers. Individuals with aspirin medication for CHD treatment were excluded from both analyses.

Conclusions. These preliminary analyses suggest that two genes regulating acute systemic inflammation may have a role in the pathogenesis of CVD. These genes, IL6 and FGG, also interact at the molecular level. Future analyses will include additional informative SNPs, and assessment of the risk for CVD event when allowing for interactions between SNPs, both within the same gene, and across genes.

REPLICATION OF LINKAGE TO THE PDE4D REGION ON 5Q AND SUGGESTED ALTERNATIVE GENOTYPES ASSOCIATED WITH ISCHEMIC STROKE IN A SWEDISH POPULATION

S. N. Ardnor,[1] P. G. Wiklund,[2] P. Lindgren, I A. K. Nilsson,[1] T. Janunger,[1] S. A. Escher,[1] B. Hallbeck,[1] B. Stegmayr,[2] K. Asplund,[2] and D. Holmberg[1]
[1] *Department of Medical Biosciences, Division of Medical and Clinical Genetics, Umeå University, Sweden*
[2] *Department of Public Health and Clinical Medicine, Umeå University, Sweden*

Scientific background. Genetic components in human stroke have been implicated in several studies, including twin and family studies. Animal models also suggest that susceptibility to ischemic stroke is influenced by genetic factors. In several rare monogenic forms of cerebrovascular disease the genetic components have been identified. For common forms of stroke, recent studies of Icelandic patients have demonstrated linkage to 5q12 and association between phosphodiesterase4D (PDE4D) and ischemic stroke (Gretarsdottir et al., 2002; Gretarsdottir et al., 2003).

Project description. To test the validity of the Icelandic findings in a different population, we studied a family-based sample consisting of 56 nuclear and extended families, including 117 patients with ischemic or hemorrhagic stroke and a nested case control sample, including 275 patients with ischemic or hemorrhagic stroke and 550 matched community controls, from the two northernmost counties of Sweden.

Main methods. Familial cases of stroke were identified from questionnaires sent to all patients affected between 1985 and 1996. One hundred and one families were ascertained. Families were included if there was at least one affected sib-pair and at least one unaffected sibling. Fifty-six families with 117 affected individuals were included. Most were nuclear families but extended pedigrees were also identified. In the association study, subjects had been participants in population-based cardiovascular risk factor surveys. We used a nested case-control design including 275 cases of first-ever stroke. Two controls for each case were matched for sex, age and place of domicile. Forty-three polymorphic microsatellite markers from ABI Prism Linkage Mapping Set v2.5 HD5 were used for genotyping in the linkage study. Multiplex PCRs on GeneAmpPCRSystem 9700 were performed. For association analysis we selected 3 SNP markers based on information available from the public databases, rs1971940 (SNP1), rs716908 (SNP2), and rs294492 (SNP3). Sequences for 2 SNPs, rs12188950 (SNP 4, deCODE SNP 45), rs12153798 (SNP5, deCODE SNP 41) and 1 microsatellite (AC008818–1) were obtained from Dr Solveig Gretarsdottir, deCODE, Iceland. We generated SNP genotypes using the TaqMan allelic 5 discrimination method.

Results and conclusions. Nonparametric multipoint linkage analysis of the family-based data set revealed two peaks with an allele-sharing LOD score > 1.0. Running additional markers with an average intermarker distance of 4.7cM at chromosome 5 yielded an increased maximum allele-sharing LOD score of 2.06 (p = .0010) at marker D5S424 and 1.60 (p = .0033) at marker D5S1969. Conditional logistic regression calculations revealed associations for two of the markers with p values < .05. SNP3 (OR 0.68 [95% CI 0.48–0.96]) and the 'B' allele (–4 bp compared to the shortest 5 allele of CEPH 1347–02) in AC008818–1 (OR 0.69 [95% CI 0.49–0.98]).The apparent protective effect of the 'B' allele in AC008818–1 is in agreement with reports from the deCODE study in which the joint set of alleles in this locus, excluding the at-risk allele, confers a protective effect. Although when correcting for the number of markers and alleles tested, p values did not reach formal significance levels, this observation remains interesting. No significant association to the defined at-risk allele of AC008818–1 in the Icelandic study was obtained in this study, OR: 1.1 (95% CI 0.84–1.45).

THE SNP 'TECHNOLOGY PLATFORM' IN UPPSALA, SWEDEN

T. Axelsson, P. Lundmark, M. Jonssson, M. Lindersson, K. Larsson, U. Liljedahl, A. Mylly, D. Fredriksson, L. Bäckström, C. Pöntinen, and A. Syvänen
Molecular Medicine, Department of Medical Sciences, Uppsala University, Sweden

SNP genotyping in the GenomEUtwin project is shared between the Finnish Genome Centre and the National Public Health Institute in Helsinki, Finland and the Department of Medical Sciences at Uppsala University. The SNP 'technology platform' in Uppsala belongs to the large Swedish Wallenberg Consortium North (WCN) for Functional Genomics. The main task of the WCN SNP platform is to perform SNP genotyping as a service to academic groups within Sweden or, as is the case in GenomEUTwin, in collaboration with international projects. The SNP platform has a staff of five research engineers or laboratory technicians and three biocomputing engineers. During the past years, over 500,000 quality controlled genotypes have been delivered to some 30 research projects. The goal for the year 2005 is to reach an annual capacity of one million delivered genotypes. The budget for SNP genotyping in GenomEUtwin allows production of about three million SNP genotypes, of which half will be produced by the SNP platform in Uppsala.

Two genotyping systems based on 'minisequencing' single nucleotide extension are used at the Uppsala SNP platform. These are a homogeneous minisequencing assay with detection by fluorescence polarization (FP–TDI) using the Analyst AD™ instrument (Molecular Devices), which is optimal for analysis of individual SNPs in a 384-well microtiter plate format, and the GenomeLab SNPstream system (Beckman Coulter) for streamlined 12-plex PCR and fluorescent minisequencing reactions in a 384-well array format. Additional key equipment at the Uppsala SNP platform are three pipetting robots for pre- and post-PCR liquid handling and PCR instruments.

SNP markers for genotyping are commonly retrieved from dbSNP, the International HapMap project and Celera databases. PCR and minisequencing primers are designed using Autoprimer (www.Autoprimer.com). The performance of the SNP assays are evaluated in a set of 192 samples using tests for cluster appearance, allele frequency, Mendelian inheritance, Hardy–Weinberg equilibrium and duplicates as quality criteria before genotyping cohort samples. The quality of the genotype data from the cohort samples is assessed using similar quality criteria prior to delivery.

A major challenge in medium- to high-throughput SNP genotyping is handling of the vast amount of genotype data produced, and to maintain traceability of the steps of the laboratory processing of samples and reagents. For this purpose we are developing our own Laboratory Information Managing System (LIMS). The first module of the LIMS that has been completed is a relational database for storage and handling of genotype data. This module facilitates extraction and comparisons of genotype data as well as the quality control review process and validation of the genotyping results. In our fully developed LIMS, a barcode-assisted system for traceability of samples and reagents will be available.

A quality system according to the requirements of the European ISO/IEC 17025 standard is being introduced at the SNP platform. This standard also includes ISO 9001 and 9002. The goal is to fulfill the requirements of the ISO/IEC 17025 system by the end of the year 2004, and obtain accreditation by the accreditation agency SWEDAC. Standard operation procedures for all significant steps in management, maintenance of instruments and localities, and the laboratory process are being documented and implemented as routine. The quality system of European standard guarantees application of a quality system, technical competence and the ability to generate technically validated results.

ENDOPHENOTYPES IN DEVELOPMENTAL ANXIETY DISORDERS

M. Battaglia,[1,3] A. Ogliari,[1] A. Zanoni,[1] V. Patriarca,[2] C. d'Ippolito,[2] R. Cirrincione,[2] and M. A. Stazi[2]
[1] *The Department of Psychology, 'Vita-Salute' San Raffaele University at the Department of Neuropsychiatric Sciences, Istituto Scientifico San Raffaele, Italy*
[2] *Istituto Superiore di Sanità, Italian Twin Registry, Rome, Italy*
[3] *IRCCS E Medea Bosisio Parini, Italy*

Scientific background. Distinct anxiety disorders (including the commonly observed syndromes of social phobia [SP] and panic disorder [PD]) usually manifest themselves within the first two decades of life, and are moderately heritable, according to twin studies based on structural equation modelling of symptoms' profiles. A deeper understanding of the causal architecture of human anxiety disorders can be promoted by the adoption of endophenotypes. Ideally, the endophenotypes should be etiopathogenetically related to, be more heritable, and have simpler genetic architecture, than the relative clinical syndromes. Recently we studied an endophenotype of PD, namely reactivity to inhaled carbon dioxide (CO2) in adult twins of the Norwegian Twin registry, and found a heritability of around .5 (best-fitting model: AE), compared with heritability figures of around .3 (with best-fitting ACE models) reported by published twin population studies of PD symptoms phenotypes.

Project description. In children at risk for SP/behavioral inhibition we have characterized an endophenotype consisting of a N400 event-related brain potential generated in response to facial expressions of emotions of coetaneous children. In an independent sample of Italian twins we have performed biometrical analyses of different DSM-IV childhood anxiety disorders, including SP. Based on the evidence provided by these two previous independent investigations we present a study of visual ERP reactivity to facial affects, to be performed in twins aged 8 to 11 years. The heritability of the N400 response to facial expressions will be studied in relation to children's degree of SP/behavioural inhibition, and their DNA collected for subsequent genetic studies.

Main methods. In a general population sample of children characterized for their degree of SP/behavioural inhibition and the 5-HTTLPR genotype, we studied the N400 component of information processing in response to happy, neutral, and angry expressions of coetaneous children (study1). A sample of 380 complete pairs of the Italian Twin Registry (aged 8 to 16) completed the SCARED questionnaire for DSM-IV childhood anxiety disorders (study 2). We present here an ERP study of twins belonging in the sample of study 2 as a further characterization of SP. Univariate and bivariate models will be fitted to the phenotype of

https://doi.org/10.1375/twin.7.6.680 Published online by Cambridge University Press

SP/behavioural inhibition and the N400 endophenotype. Molecular genetic analyses will be performed subsequently.

Results. Study 1 showed a significant 'expression by genotype' interaction ($F = 3.57$, $p < .01$), sustained by the difference in N400 amplitude of the 'ss' subjects compared to the 'LL' subjects when they were viewing the anger expression ($p < .017$). The –S allele of the 5-HTTLPR is associated to this N400 endophenotype with an adjusted R^2 of .19 (Battaglia et al., in press). Univariate analyses of twins' responses to the SCARED questionnaire found a heritability of .54 (best fitting model: AE, $\chi^2 = 5.3$ $p = .24$, AIC = –2.61) for DSM-IV childhood SP phenotype.

Conclusions. Data show that (1) the N400 endophenotype is a viable approach to study SP in the developmental years; and (2) DSM-IV SP has considerable heritability.

A bivariate twin study of N400 and SP followed by DNA analyses can help clarify the architecture of social phobia in the developmental years.

PRELIMINARY LINKAGE RESULTS FOR ADULT HEIGHT AND BODY MASS INDEX (BMI)

B. K. Cornes,[1,2] S. E Medland,[1,3] K. I Morely,[1,4] M. A. R. Ferreira,[1,2] and N. G. Martin[1]
[1] *Genetic Epidemiology, Queensland Institute of Medical Research, Australia*
[2] *School of Medicine, Central Clinical Division, University of Queensland, Australia*
[3] *School of Psychology, University of Queensland, Australia*
[4] *Institute for Molecular Bioscience, University of Queensland, Australia*

Reducing the prevalence of obesity is a difficult task, but research into the genetic and environmental influences on obesity may aid in the development of improved treatments and management strategies. We present preliminary results from univariate genome scans for height and BMI, with and without log transformation. Height and BMI phenotypes were available for 1886 and 1883 individuals respectively, from 901 families. Genotypic data were available for between 250–1700 markers depending on individual study participation. Merlin and Minx Variance Components analyses were run at 5cM intervals across the entire genome using age and sex as covariates. Heritability estimates (calculated by Merlin) for height, BMI and log transformed BMI in the genotyped sample were all estimated to be approximately .76. We found areas of suggestive linkage for height on chromosomes 6, 7, 10, 11, 15 and 18, and for BMI on chromosome 3 and 6. These results both replicate and extend existing research. The likely candidate genes in these regions will be discussed.

SHOX, A CANDIDATE GENE FOR BODY HEIGHT

A. Dahlgren,[1] M. Perola,[2] A. Syvänen,[1] K. Christensen,[3] K. O. Kyvik[3]
[1] *Department of Medical Sciences, Molecular Medicine, Uppsala University, Sweden*
[2] *National Public Health Institute, Department of Molecular Medicine, Helsinki, Finland*
[3] *Danish Twin Registry, University of Southern Denmark, Odense, Denmark*

The study of genetics has to date been very successful in identifying monogenic disease genes. However the majority of diseases and common traits are genetically complex involving interaction between many genes and environmental factors. Several complex traits and diseases have been selected for analysis within the GenomEUtwin project. One of these, body height, is the focus of this study. The Short Homeobox Containing Gene (SHOX) was selected as the first candidate gene for analysis after searching literature and databases for genes related to body height. The SHOX gene codes for a transcription factor and is located in the pseudoautosomal region (PAR1) on the X and Y chromosomes. SHOX has been linked in several studies to idiopathic short stature and is known to play a central role in Leri–Weill and Turners syndrome, where short stature is part of the phenotypes of the patients.

The aim of this project is to genotype single nucleotide polymorphisms (SNP) across the SHOX gene in selected sample materials available from the collaborators within GenomEUtwin. Using public and commercial databases and publications, assays for 25 SNPs were set up and validated for analysis on the SNPstream® system (Beckman Coulter) which uses 12-plex minisequencing with fluorescence detection in a 384 well format for high-throughput genotyping. Two groups of samples from the Danish twin registry (Odense, Denmark) have successfully been genotyped, generating over 19,000 genotypes. The sample material consisted of a group of 371 samples from dizygotic (DZ) twins that represent the normal distribution of body height in the Danish population. This control group will be compared to a second group consisting of 396 samples from DZ twins selected to be significantly discordant for body height. The genotyping results will then be used to test for correlation between found variation of genotypes and body height in the test population. The genotyping results are currently being analyzed using the Merlin statistical software package.

This will be the first in-depth SNP analysis of the SHOX gene in relation to body height. Studying body height will not only generate information about the mechanics of body development, but will also give much valuable experience needed to investigate our most common diseases that most likely have a much greater level of genetic complexity.

CAN THE COMORBIDITY OF PHYSICAL INACTIVITY AND CARDIOVASCULAR RISK FACTORS HELP US FIND GENES?

E. J. C. de Geus, G. Willemsen, D. Posthuma, J. J. Hottenga, N. Kupper, J. H. Stubbe, J. M. Vink, and D. I. Boomsma
Department Biological Psychology Vrije Universiteit, Amsterdam, the Netherlands

We measured four independent risk factors for CVD (exercise behavior, HR, RSA, BP) in large samples of Dutch twins. Exercise, HR, RSA and BP are clearly correlated at a phenotypic level. Prospective epidemiology and randomized controlled training studies suggest a causal chain of events, in that exercise reduces cardiovascular disease risk through effects on HR and BP (although the effect on RSA remains debated). In parallel we have shown high heritability of both sports participation and weekly energy expenditure in exercise. This opens up the possibility that genes for blood pressure and heart rate overlap with the genes determining self-chosen levels of exercise in leisure time. In that case, failing to take exercise behavior into account may severely hamper gene finding. Put otherwise, high blood pressure in a physically inactive person is a different phenotype than high blood pressure in a normal or high active person. Reversing the argument, we may argue for a substantial gain in the power of our linkage analyses on these risk factors if exercise behavior is factored in. As a first step we propose to establish the genetic correlation between exercise and cardiovascular risk in our twin sample. Since many subjects do not engage in exercise, exercise participation and quantity of exercise must be dealt with in a dual liability model. If, as expected, a significant genetic correlation is found, this can be grounds for either a multivariate or a multigroup linkage analysis, using exercise as an additional or a stratifier variable. Finally, we can try to proceed by delineating the nature of the genetic correlation between exercise and CVD risk, that is to discriminate pleiotropy from a causal chain of events. The latter may be solvable at the level of twin modeling, but will certainly benefit from access to the actual genes for either exercise or the risk factors.

A GENOME-WIDE SCAN FOR BLOOD PRESSURE SUGGESTS LINKAGE TO CHROMOSOME 11, AND REPLICATION OF LOCI ON CHROMOSOMES 16, 17 AND 22

M. de Lange,[2,1] T. D. Spector,[1] and T. Andrew[1]
[1] *Twin Research & Genetic Epidemiology Unit, St Thomas' Hospital, London, United Kingdom*
[2] *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden*

Hypertension was one of the first complex traits to be studied and is thought to be influenced by polygenic and multiple environmental risk factors. Several genomic studies have found suggestive LOD scores for either blood pressure or essential hypertension, but few loci have been replicated. In this study we performed a genome-wide linkage analysis for systolic (SBP) and diastolic (DBP) blood pressure on 1109 Caucasian female dizygotic twin pairs from the TwinsUK registry in London.

Multipoint linkage analysis 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 chromosome 17 at 70cM (LOD 1.8 SBP) and at 35cM on chromosome 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 $p = .0007$) 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. Future studies should focus on the fine mapping of these replicated regions, which include potential candidate genes.

QUANTITATIVE GENETIC ANALYSIS OF HEIGHT, WEIGHT, AND BMI IN 20-YEAR-OLD ITALIAN TWINS: UNIVARIATE MODELS AND A BIVARIATE CHOLESKY APPROACH

C. Fagnani, R. Cirrincione, R. Cotichini, C. D'Ippolito, L. Nisticò, V. Patriarca, S. Pulciani, S Brescianini, E. Medda, and M. A. Stazi
Istituto Superiore di Sanità, Rome, Italy

Scientific background. Height, weight, and body mass index (BMI) are multifactorial characteristics responding to both genetic and environmental influences. Recently, interest in height as an indicator of childhood living conditions, and in BMI as a measure to document the worldwide growth in the prevalence of obesity, has increased remarkably. However, just a few efforts have been dedicated to quantifying the effects of genes and environment on the variation in height, weight, and BMI, and on the co-variation between height and weight. These issues can easily be investigated by the twin method.

https://doi.org/10.1375/twin.7.6.680 Published online by Cambridge University Press

Project description. We studied a cross-sectional sample of young adult twins in order to estimate the heritability (i.e., the proportion of variance due to genetic factors) of height, weight, and BMI, testing for possible sex differences, and determine whether the covariance between height and weight has a genetic or environmental basis. The same twin sample will be used to select the most informative pairs, which will be included in the GenomEUtwin pooled database for the genetic linkage studies on stature and BMI.

Main methods. The twin method was used in this study. The data set was derived from the Italian Twin Registry, and consisted of 1800 twin pairs. All twins were aged 20 years, and were contacted by mailed questionnaire in two waves: 1100 were born in 1983 and enrolled in 2003, while the remaining 700 were born in 1984 and enrolled in 2004. The effect of genetic and environmental factors was estimated via Structural Equation Modeling. Univariate models for height, weight, and BMI and a bivariate Cholesky decomposition for height and weight were considered.

Results. For BMI, the heritability estimates were .86 and .70 in males and females, respectively. No evidence was found for neither common environment ($\chi^2 = 2.71$, $df = 2$, $p = .26$) nor sex differences ($\chi^2 = 4.14$, $df = 3$, $p = .25$). A bivariate ACE Cholesky model provided heritability estimates of .79 (males) and .69 (females) for height, and of .87 (males) and .74 (females) for weight. It also indicated that genes and environment were simultaneously responsible for the covariance between height and weight, as shown by a substantial genetic (r_g), common environmental (r_c), and unique environmental (r_e) correlation in both sexes (males: $r_g = .45$, $r_c = .90$, $r_e = .33$; females: $r_g = .44$, $r_c = .68$, $r_e = .29$).

IF PARTICIPANT CONSENT IS THE ANSWER, WHAT WAS THE QUESTION?

H. S. Hasan and J. R. Harris
Department of Genes and Environment, Division of Epidemiology, the Norwegian Institute of Public Health, Oslo, Norway

Ongoing longitudinal projects using research biobanks, such as GenomEUtwin require the continual cooperation of participants namely in the form of health information and biological samples. These epidemiological studies are under increasing pressure to put greater time and financial resources in ensuring the continual participation of these participants. In comparison to 20 or 30 years ago, changes in legislation in conjunction with cultural perceptions have resulted in a shift of power from the researcher to the participant. One factor that can be defined as a tool that aids the empowerment of the participant is informed consent.

Presently, not only is it self-evident that obtaining informed consent from research participants is good practice, the process of consent is also a defined legal obligation in many countries. The general information provided by the researcher concerns the aims and purpose and methodology of the study, and the nature and reporting of any results generated. Norwegian legislation additionally specifies that the participant be informed, they can withdraw their consent at any point without giving a reason, and also the confidential nature of data handling and results.

Obtaining informed consent is a continuous organic process — there is both a duty to give the necessary information and also to ensure or facilitate the understanding of the information given. This organic process is in danger of being superseded in favor of the participant consent being used as a defensive contract. The advantages and disadvantages of the informed consent as a process or a 'contract' model will be presented citing the interests of both the researcher and the participant. Ideas and suggestions will be given to help participating centers fulfill obligations to their participants (these obligations will be identified in the presentation). Greater use of the GenomEUtwin website as an additional resource information for participants will be discussed.

MAPPING QTLS FOR HDL-C, LDL-C AND ASSOCIATED PROTEINS AND IDENTIFICATION OF UNDERLYING GENETIC VARIATION: A META-ANALYSIS OF FOUR GENOME SCANS

B. T. Heijmans,[1] H. Putter,[2] M. Beekman,[1] N. Lakenberg,[1] H. J. van der Wijk,[2] J. B. Whitfield,[3] R. R. Frants,[4] U. DeFaire,[5] D. T. Oconnor,[6] N. L. Pedersen,[5] N. G. Martin,[7] D. I. Boomsma,[8] and P. E. Slagboom[1]
[1] *Molecular Epidemiology, Leiden University Medical Centre, the Netherlands*
[2] *Medical Statistics, Leiden University Medical Centre, the Netherlands*
[3] *Clinical Biochemistry, Royal Prince Alfred Hospital, Sydney, Australia*
[4] *Human and Clinical Genetics, Leiden University Medical Centre, the Netherlands*
[5] *Karolinska Institute, Stockholm, Sweden*
[6] *Department of Medicine, University of California at San Diego, California, United States of America*
[7] *Queensland Institute for Medical Research, Brisbane, Australia*
[8] *Biological Psychology, Vrije Universiteit Amsterdam, the Netherlands*

Elevated lipid levels in plasma are key risk factors for cardiovascular diseases. To map quantitative trait loci for lipid levels, genome-wide scans were performed in Dutch, Swedish and Australian twins followed by combined linkage/association analysis to identify responsible genetic variation. To optimize power, the genome scans were analyzed simultaneously with a meta-analysis method that estimates an overall LOD score using the mean of the sample-specific effect sizes weighted by the precision of these estimates. Importantly, this method allows for heterogeneity between studies.

The study focused on plasma levels of HDL-cholesterol and LDL-cholesterol and their main protein constituents, apolipoprotein AI (ApoAI) and apolipoprotein B (ApoB) respectively. Genome-wide scans with an average marker spacing of 5–10 cM were performed in 1416 dizygotic twins (708 pairs) from the Netherlands (207 pairs), Sweden (44 pairs) and Australia (457 pairs). A method that regresses the estimated proportion of alleles shared identical-by-descent on the squared sums and squared differences of trait values of the pairs (Merlin-regress) was used to estimate the QTL effect and standard errors for each of the twin samples separately. These estimates were subsequently used in the meta-analysis method.

Suggestive linkage for HDL-C was observed on 8p23.1 (LOD = 2.0) and 12q21.2 (LOD = 2.2) and for ApoAI on 1q21.3 (LOD = 2.1). In contrast to HDL-C and ApoAI, linkage regions frequently coincided for LDL-C and ApoB (on 2p24.1, LOD score both 2.1; on 2q32.1, 2.0 [LDL-C] and 1.7 [ApoB]; on 19p13.2, 1.9 and 0.7; and on 19q13.31, 1.7 and 0.7). After finemapping, the position of 3 maximum LOD-scores were within 1 cM of major candidate genes, namely *APOB*, *LDLR* and *APOE*. To assess the possible contribution of genetic variation in these genes to the linkages observed, tagging SNPs were measured (6 in *APOB*, 8 in *LDLR* and 5 in *APOE*) and haplotype analysis was performed. Accounting for the effect of *APOB* haplotypes reduced the LOD-score observed for LDL-C on 2p24.1 to 1.0. Accounting for the effect of the *LDLR* and *APOE* haplotypes did not change the LOD-score close to the *LDLR* gene but abolished the linkage signal at the *APOE* gene.

Meta-analysis of multiple genome scans in conjunction with combined linkage/association analysis to test haplotypes of positional candidate genes provides a powerful approach to disentangle complex traits. Using this approach we mapped multiple putative QTLs for lipid phenotypes and identified genes underlying 2 of them.

LINKAGE ANALYSIS GENOME SCANS WITH TENS OF THOUSANDS OF SNPS GIVES SYSTEMATICALLY HIGHER POWER THAN TRADITIONAL MICROSATELLITE-BASED APPROACHES UNDER THE NULL HYPOTHESIS

T. Hiekkalinna,[1] M. Perola,[1] and J. D. Terwilliger[2,3]
[1] *National Public Health Institute, Department of Molecular Medicine, Uppsala University, Sweden*
[2] *University of Helsinki, Finnish Genome Center, Finland*
[3] *Department of Psychiatry and Department of Genetics and Development, Columbia Genome Center, Columbia University, United States of America*

Recent technological advances in SNP genotyping have led to many groups advocating that genome scans for linkage use thousands or even tens of thousands of SNPs instead of the traditional microsatellite based approach. However, there are several potential problems with these markers, which have been well known for several decades, since the older RFLP markers are basically SNPs assayed through a different technology. These problems include reduced ability to detect genotyping errors when they do occur, necessity to properly model linkage disequilibrium among the markers when performing multipoint analysis, and high stochastic variation and corresponding lack of informativeness in single-marker analyses.

It is well known that when performing linkage analysis with markers in LD with one another, one must allow for the LD in the linkage analysis to avoid spurious false positive rates. In the days of RFLPs many false positives were noted from linkage studies failing to correctly model the long-range haplotype frequencies when analyzing tightly spaced markers jointly. While this was noticed and studied decades ago, solutions were also developed then to analyze small numbers of markers in LD with one another, in which linkage analysis was performed conditional on long-range haplotype frequencies estimated from data. However, in the intervening decades, scientists have become reliant on computer programs for massive multipoint linkage analysis which take advantage of Hidden Markov Models, a technique that assumes that genotypes at the next locus are independent of all other loci except the most tightly linked one in a sequence.

In this study, we demonstrate by simulations the tremendous increases in false positive rates when LD exists among linked markers and the markers are analyzed under the assumption of linkage equilibrium. We first simulated a large set of 1000 affected concordant dizygotic (DZ) twins with and without parents genotyped assuming LD among a set of markers estimated from real data from a dense SNP map in Finnish samples, showing the inflation of LOD scores observed and the artificially rapid decline in LOD score around such spurious LD-generated peaks due solely to marker–marker LD. This is repeated for a large sample of twins in which the same haplotype frequency model was simulated in a set of real DZ twins in which the quantitative trait of stature was fixed from real data, again showing inflation of the LOD scores under the null hypothesis. Finally, using modified versions of the

https://doi.org/10.1375/twin.7.6.680 Published online by Cambridge University Press

LINKAGE programs implementing an efficient direct search algorithm for estimating long-range haplotype frequencies in pedigree data, we show that analysis of the same datasets conditional on the observed haplotype frequencies restored the correct null distribution for the statistics considered.

The conclusion is that while some may have claimed the large scale SNP-based genome scanning may be more powerful than that done with microsatellites, the same claim is certainly true under the null hypothesis: that when there is NO gene, large scale SNP genotyping will give systematically higher LOD scores than the microsatellite based approach, both because of LD, as well as the greater inherent stochastic variation such markers carry.

GENETIC INFLUENCE ON CHANGE IN BMI: A LONGITUDINAL STUDY OF FINNISH AND DANISH TWINS

J. Hjelmborg,[1] K. Silventoinen,[2] K. Christensen,[1] and J. Kaprio[2]
[1] *University of Southern Denmark, Odense, Denmark*
[2] *Department of Public Health, the Finnish Twin Cohort Study, University of Helsinki, Finland*

Previous studies of twins and families have shown substantial genetic contribution to the variation in BMI. In this talk we consider genetic influence on change in BMI. The longitudinal cohorts of Danish and Finnish twins allows for estimating genetic influence in measures with certain relations to change in BMI. The Danish (LSADT) cohort consists of up to five observations with two-year intervals for each individual and the Finnish cohort consists of up to seven observations with different time-intervals for each individual. We consider the application of marginal models and subject specific models (bivariate growth curve models) and discuss preliminary results. Furthermore, we present the selection of informative pairs from the GenomEUtwin cohorts for linkage study with respect to level of BMI and change in BMI. This involves the Norwegian, Danish, Swedish, Italian, Finnish and Dutch cohorts.

DEVELOPING A GENOTYPE RESULT MANAGEMENT SYSTEM USING THE GENOMEUTWIN GENOTYPE DATABASE STRUCTURE AS A FOUNDATION

M. Jonsson, D. Fange, P. Lundmark, L. Bäckström, M. Lindersson, K. Larsson, T. Axelsson, and A. Syvänen
Molecular Medicine, Department of Medical Sciences, Uppsala University, Sweden

The rapidly growing amount of data produced at genotyping facilities such as ours puts new automation and scalability demands on the information management performed in the laboratory. A number of software manufacturers can provide commercial laboratory information management systems (LIMS) intended to solve some of these problems. Since many of these are aimed at the laboratory market in general, software systems must either undergo a time-consuming adaptation process to fit the laboratory activities or, still worse, the laboratory process itself must be changed to fit the software.

We chose instead to develop our own database system, allowing more focus on convenient results management rather than the sample tracking and project management associated with most LIMS. The underlying database for this system co-evolved with the GenomEUtwin genotype database,* and the core parts of the two databases are identical. Analogously to the idea of collecting results from different genotyping facilities within GenomEUtwin into one central database, our system collects genotype results from the different genotyping instruments within the laboratory into a single database. Quality controls, statistics calculations and result exports can then be performed independently of the source instrument type.

Two different graphical user interface applications were written on top of the database in close collaboration with the laboratory personnel: one for storing information about samples, primers and markers before the genotyping, and one for viewing the results afterwards. The first has functions for loading data, such as pedigree information and marker information, from text files. It can also load primers directly from the output files of the primer design program used in the laboratory. The second one makes it possible to combine different sets of items (samples, markers and assays) to be able to view and calculate statistics on any desired subset of results. Furthermore, a number of quality controls are automatically carried out when the results are viewed. These quality control steps include checking for consistent results when a sample has been analyzed multiple times, controls of the inheritance pattern if family information is available, a Hardy–Weinberg equilibrium test and success rate calculations for samples. It is also possible to save the viewed data subset and quality conditions in a so called 'session', and the results in such sessions can even be marked for export to a location such as the central GenomEUtwin database to which our database has a connection.

For more than a year, this database system has been used by all personnel in our genotyping service facility with good results. We believe that having structured and well-defined data in our laboratory will greatly facilitate cooperation with other external partners. The database and user interfaces are now being developed further to include more traditional LIMS functionality, such as sample tracking and project progress. The initial focus on tailor-made result management has however proven to be valuable.

**In collaboration with J. Muilu, Finnish Genome Center*

GENOTYPING OF HUMAN MICROSATELLITE LOCI ON CHROMOSOME 12–22 AND X AT UPPSALA GENOME CENTER

J. Jonsson and C. Johansson
Uppsala Genome Center, Department of Genetics and Pathology, Uppsala University, Sweden

Uppsala Genome Center is a service facility providing services for genotyping and sequencing projects mainly to the academic community. Our aim is to provide flexible and customized services for all kinds of genetic studies. UGC can provide DNA-extraction from blood or other sources, genotyping with microsatellite and SNP markers, sequencing service or analysis of custom-prepared products by capillary electrophoresis. The center was established in 1998 as a resource facility for running large genetic mapping projects based on genotyping of microsatellite markers.

The present throughput of UGC is over 1.6 million genotypes per year. The center has an automated procedure for amplification of microsatellite loci and pooling of PCR products. All analyses are performed on the capillary electrophoresis instrument, ABI PRISM® 3700 DNA Analyzer. The genotypes are produced in the software GenoTyper® or GeneMapper™ and are independently double-checked before data delivery.

UGC has been involved in a large number of genotyping projects including mapping of complex disorders in humans, laboratory and domestic animals.

In 2002, UGC became a member of the genomEUtwin providing microsatellite genotypes on chromosome 12–22 and X using the ABI PRISM® Linkage Mapping Set v.2.5 MD10. The genotypes are uploaded to the genomEUtwin genotype database (gtDB) located at the Finnish Genotyping Center in Helsinki, Finland.

About 92,000 genotypes from 650 Danish twins have been produced and uploaded to gtDB. In October 2004 the genotyping of 464 Danish and 550 Finish twins begins, giving a calculated number of 156,000 genotypes.

NONPARAMETRIC MULTIPLE IMPUTATION OF EVENT TIMES FOR SUBJECTS WITH CHD AT BASELINE IN MORGAM GENETIC STUDY

J. Karvanen, S. Kulathinal, and K. Kuulasmaa
International CVD Epidemiology Unit, Department of Epidemiology and Health Promotion, National Public Health Institute, Finland

Scientific background. The objective of the MORGAM study is to explore the association between cardiovascular diseases (CVD) and their classic and genetic risk factors. Contrary to many follow-up studies concentrating only on the CVD mortality, nonfatal events are also registered for most of the MORGAM cohorts. The first occurrence of coronary heart disease (CHD) is one of the main endpoints of the study. Because the first CHD event is not necessarily fatal, the MORGAM cohorts, also contain subjects who had their first nonfatal CHD event before the baseline of the study. For these events, the exact event time is unknown but its upper bound is known. The percentage of subjects with baseline CHD in MORGAM cohorts varies from 0% to 13%, which is a considerable proportion when compared to the percentage of first incidence of CHD during the follow-up that varies from 0.2% to 17%. Hence, the inclusion of the baseline CHD cases in the time-to-event analysis would significantly increase the number of events and provide more information on the relatively young subjects. The use of the baseline CHD cases suits the analysis of genetic risk factors, because genotypes, contrary to many other baseline measurements, cannot be affected by a preceding CHD event.

Project description. In statistical terms, doubly censored time-to-event data is considered. Doubly censored data consist of left-censored observations (subjects with baseline CHD), events observed during the study, and right-censored observations (subjects without events before the end of the follow-up). The goal is to develop statistical methods for doubly censored data in cohort studies and extend these methods to the case-cohort design. The developed methods will be applied to the MORGAM data.

Main methods. A nonparametric multiple imputation approach is proposed for doubly censored data. In this approach, the left-censored observations are imputed recursively using conditional distributions estimated from the data. After imputation, the standard estimation methods for right-censored survival data can be directly applied. The proposed imputation approach is compared with maximum-likelihood estimation under Cox's proportional hazard model. The maximum-likelihood estimation of the model parameters is carried out using the EM algorithm.

Results. The proposed estimation methods are studied in simulation experiments that try to imitate the essential features of the MORGAM data. Simulations allow comparing the models estimated from the doubly censored data to the models estimated from the complete data. Preliminary results suggest good performance of both the imputation

https://doi.org/10.1375/twin.7.6.680 Published online by Cambridge University Press

approach and maximum-likelihood estimation. Besides the simulation results, examples with the MORGAM data will be also presented.

ALLELIC VARIANTS OF UPSTREAM STIMULATORY FACTOR 1 (USF1)-GENE AS RISK FACTORS FOR CARDIOVASCULAR EVENTS: A PROSPECTIVE, POPULATION-BASED STUDY ON A MORGAM COHORT

K. Komulainen,[1] M. Perola,[1] R. Kilpikari,[1] P. Pajukanta,[2] J. Saarela,[1] K. Silander,[1] M. Alanne,[1] P. Ellonen,[1] S. Kulathinal,[3] K. Salminen,[3] K. Kuulasmaa,[3] V. Salomaa,[3] and L. Peltonen[1]

[1] *Department of Molecular Medicine, KTL, Finland, Biomedicum, Helsinki, Finland*
[2] *Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, California, United States of America*
[3] *Department of Epidemiology and Health Promotion, KTL, Helsinki, Finland*

Scientific background. Upstream stimulatory factor I (USF1) is a ubiquitously expressed transcription factor controlling several genes, many of which are involved in lipid and glucose metabolism. We recently associated an allelic variant of USF1 with familial combined hyperlipidemia (FCHL; Pajukanta et al. 2004; Nat Gen 36:371–6), one of the most common dyslipidemias predisposing to premature coronary heart disease. Specific alleles of USF1 seem to also influence features of glucose and lipid homeostasis (Putt et al. 2004; Hum Mol Genet 13:1587–97).

Project description. To assess the general significance of this gene for the risk of cardiovascular disease (CVD), we studied four single nucleotide polymorphisms (SNPs) using a population-based, prospective case–cohort approach. Our population cohort was collected via the FINRISK surveys that are carried out every five years and designed to assess the prevalence and risk factors of cardiovascular diseases in Finland. The FINRISK 92 study represents a cohort of 8000 participants aged 25 to 64 years, randomly chosen from 4 different regions of Finland, collected during 1992 and followed up until 2001. Cases with a CVD event either at the baseline or during the follow-up period as well as an age- and sex-stratified subcohort were selected from the original cohort for this genetic study. The genetic study sample consisted of 758 participants with the total of 6836 person years of follow-up. The FINRISK 92 cohort is one of the cohorts in the ongoing MORGAM Project, a multinational collaborative study to explore the relationships between the development of cardiovascular diseases and lifestyle and genetic risk factors.

Main methods. To define the allelic spectrum of the USF1 gene in the study sample, four USF1 SNPs were genotyped using allele-specific primer extension on microarrays and the MassARRAY System. Genotype distribution in the study subjects for the four SNPs allowed us to predict the haplotypes directly from the genotype combinations. Cox's proportional hazard analysis measuring time-to-event was used to estimate the risk of a CVD event during the follow-up period in relation to genotype groups and haplotypes.

Results. Only 4 different haplotypes with frequencies varying from 13% to 38% were detected among the 758 study subjects. The frequency of the haplotype CCTA was significantly higher in the subcohort females without CVD than in the female CVD cases (24% vs. 32%, respectively, $p = .016$). Our results from the time-to-event analysis indicate that the USF1 allelic variants significantly contribute to the risk of CVD. In females, the presence of a risk haplotype (CTTG) significantly increased the risk for CVD, hazard ratio (HR) being 4.72 ($p = .004$). A protective haplotype (CCTA) conferring 75% smaller risk of CVD compared to carriers of other genotypes (HR 0.25, $p = .008$) was also identified for females. Since the difficulty to assess CVD risk in females is generally recognized, our finding is of a special interest and might have clinical relevance. Our data provide a new candidate gene to be tested in additional GenomEUtwin cohorts for CVD. Future data will show if the analysis of the USF1 risk alleles could be developed to a DNA test for the predication of the CVD risk.

STRATEGY FOR POOLING EVIDENCE FOR LINKAGE FROM DIFFERENT POPULATIONS

J. Lebrec, H. Putter, and H. van Houwelingen

Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, the Netherlands

The 'raison d'être' of the GenomEUtwin project is that pooling of data from the different twin registries across Europe and Australia will provide sufficient numbers for detection of the typically small quantitative trait loci (QTL) to be expected in common disorders. This entails a number of challenges from the statistical point of view. Firstly, since twins to be genotyped are selected on the basis of their trait values (see Putter et al. 2003; Twin Res 6:377–382.), any method of analysis should explicitly accommodate this sampling procedure. Secondly, not all QTLs will be present in all populations (qualitative heterogeneity) and even if they exist in all populations, their respective effects are likely to differ widely (quantitative heterogeneity). Finally, marker data may originate from very different marker maps and this constitutes a further source of heterogeneity to be accounted for. We have already exposed our approach for tackling the first of these challenges in Lebrec et al. 2004 (Genet Epidemiol 27:97–108). We show here how standard techniques for meta-analysis of clinical trials can be adapted and offer a fast and economical solution to the problem of quantitative heterogeneity. An application of this technique to lipid levels is presented by Heijmans et al. in another abstract. Finally, we sketch how the method of Generalized Estimating Equations (in the spirit of Liang et al., 2001; Hum Hered 51:64–78) may be used to increase precision around the estimated location of a putative QTL.

MORGAM: AN INTERNATIONAL PROJECT POOLING COHORT STUDIES OF CARDIOVASCULAR DISEASE

P. McCarron,[1] K. Kuulasmaa,[2] V. Salomaa,[2] S. Kulathinal,[2] K. Asplund,[3] F. Cambien,[4] M. Ferrario,[5] M. Perola,[2] L. Peltonen,[2] D. Shields,[6] H. Tunstall-Pedoe,[7] and A. Evans[1] for the MORGAM Project.

[1] *Queen's University, Belfast, United Kingdom*
[2] *National Public Health Institute, Helsinki, Finland*
[3] *National Board of Health and Welfare, Stockholm. Sweden*
[4] *INSERM, Paris, France*
[5] *Università degli studi dell'Insubria, Varese, Italy*
[6] *Royal College of Surgeons in Ireland, Dublin, Ireland*
[7] *University of Dundee, Dundee, Scotland*

Background. MORGAM (http://www.ktl.fi/morgam) is an acronym derived from MOnica Risk Genetics Archiving and Monograph: the last two parts have been completed, so this abstract concerns the risk and genetic components. The project, based originally on cohorts recruited and followed by the WHO MONICA Project, has grown and now forms part of GenomEUtwin (http://www.genomeutwin.helsinki.fi), a network of excellence for genomics in Europe.

Objective. To investigate the contribution of the classic risk factors to the development of cardiovascular disease endpoints, fatal and nonfatal, to derive a risk factor score, and in a subset of the total pooled cohort, to examine the importance of genetic polymorphism in determining cardiovascular risk. Additionally there are plans to test the inflammatory hypothesis for atherosclerosis, should sufficiently well-stored sera be available.

Subjects and methods. The design is case/cohort in order to establish gene frequency in the different populations and to allow the study of different disease. There are currently 144,447 subjects in the risk cohorts in several European countries and elsewhere, with a subset of these comprising 66,164 subjects for whom DNA is available. Other large cohorts are on the point of joining and more are welcome. Genotyping is mainly through mass spectrometry. We are assessing the possibility of employing a Luminex platform to measure many inflammatory markers.

Results. To date we have a total of 8706 fatal and nonfatal incident CHD and stroke events, with a subset of 3064 with DNA. We have completed the analysis of 180,000 genotypes to date.

Conclusions. It is possible to pool cohorts across Europe provided that adequate quality assessment procedures are in place.

DETAILED DESCRIPTION OF BMI AND RISK OF CORONARY HEART DISEASE AND STROKE: EVIDENCE FROM THE MORGAM PROJECT

P. McElduff,[1] I. Buchan,[1] P. McCarron,[2] S. Sans,[3] J. Dallongeville,[4] A. Pajak,[5] and A. Doring[6] (for the MORGAM Project)

[1] *Evidence for Public Health Unit, School of Epidemiology and Health Sciences, University of Manchester, United Kingdom*
[2] *Department of Epidemiology and Public Health, Queen's University Belfast, Belfast, United Kingdom*
[3] *Programa CRONICAT, Hospital Sant Pau, Barcelona, Spain*
[4] *MONICA-Lille, INSERM U508, Institute Pasteur de Lille, Lille, France*
[5] *Department of Clinical Epidemiology and Population Studies, Institute of Public Health, Jagiellonian University, Krakow, Poland*
[6] *Institute of Epidemiology, GSF Forschungszentrum fur Umwelt und Gesundheit, Germany*

Aims/hypothesis. Obesity and overweight are rising dramatically in developed countries, increasing the possibility of future upturns in coronary heart disease (CHD) and stroke. Few studies have examined the associations between body size and risk of these diseases in both sexes and across geographical regions. We undertook such analyses in the MORGAM (MOnica Risk Genetics Archiving and Monograph) study.

Methods. Height and weight were measured in participants in the WHO MONICA project and several other European cross-sectional studies. Participants of these studies were subsequently followed up for both nonfatal and fatal CHD and stroke events and for all cause mortality. Cox proportional hazards models were used to estimate the risk of a CHD and stroke event for persons classified as underweight, overweight and obese compared to people who were categorised as having a normal level of BMI. In addition we estimated the relative risk per $1 kg/m^2$ increase in BMI after adjusting for age, year of baseline measurement, smoking

https://doi.org/10.1375/twin.7.6.680 Published online by Cambridge University Press

status, marital status, number of years schooling, and history of diabetes and cardiovascular disease. The analysis was repeated to examine the impact of also adjusting for systolic blood pressure and total serum cholesterol, which we considered to be on the causal pathway rather than potential confounders.

Results. Men with BMI less than 20 kg/m^2 had a similar risk of CHD and stroke as men with normal BMI but overweight men were 1.28 (95% confidence interval [CI]: 1.20 to 1.38) times more likely to have a coronary event and 1.19 (95% CI 1.06 to 1.34) times more likely to have a stroke, while in obese men, the respective risks were 1.54 (95% CI 1.42 to 1.68) and 1.42 (95% CI 1.23 to 1.63). Women who were underweight had a higher risk of CHD than women who had normal BMI and women who were overweight had 1.42 (95% CI 1.22 to 1.66) times higher risk of CHD. Obesity conferred a relative risk of 1.77 (95% CI 1.50 to 2.09) for CHD and 1.43 (95% CI 1.17 to 1.74) for stroke. Each 1kg/m^2 increase in BMI resulted in an increased coronary risk of 4.2% (95% CI 3.4 to 5.0) among men and 3.7% (95% CI 2.6 to 4.9) among women; for stroke the estimates are 3.2% (95% CI 2.0 to 4.0) for men and 3.3% (95% CI 1.8 to 4.7) for women. Adjusting for blood pressure and cholesterol attenuated these effects but they still remained highly statistically significant.

Conclusion/interpretation. BMI is an important and independent risk factor for CHD and stroke, and the increased risk does not plateau at higher BMI levels. The increasing levels of BMI among adults and children constitute a major public health problem and are likely to lead to future increases in the incidence of CHD and stroke. Efforts to combat the obesity epidemic are needed, while the graded dose response indicates that even modest declines at all levels of BMI will have beneficial effects.

THE ASSOCIATION BETWEEN GENE POLYMORPHISMS, NOVEL CARDIOVASCULAR RISK FACTORS AND OBESITY

C. McFarlane,[1,2] I. S. Young,[1] F. Cambien,[2] and A. Evans[3]
[1] *Department of Medicine, Institute of Clinical Science, The Queen's University of Belfast, Belfast, United Kingdom*
[2] *INSERM U525, Faculte de Medecine Pitie-Salpetriere, Paris, France*
[3] *Department of Epidemiology and Public Health, The Queen's University of Belfast, Belfast, United Kingdom*

The prevalence of childhood obesity has substantially increased over the last twenty years; however, there is limited information regarding the metabolic consequences and future CVD risk resulting from childhood obesity. We have previously investigated the association of established and novel cardiovascular risk factors with overweight and obesity in a large population-based study in healthy young people in Northern Ireland (Young Hearts Project). We reported significant lipid abnormalities in the obese and overweight children relative to normal children. Furthermore, our findings demonstrated that BMI was a significant independent predictor of the following novel CVD risk factors: soluble ICAM–1, soluble VCAM–1, remnant lipoproteins and C-reactive protein.

Having established a causal relationship between obesity and novel CVD risk factors, we intend to extend these investigations to assess how genetic variability relates to obesity, lipids and inflammatory factors in atherosclerosis. The association between gene polymorphisms and their interaction with the environmental stress imposed by obesity is of particular interest. The project will focus specifically on the investigation of 2–3 currently unexplored candidate genes belonging to the same biological system. We intend to collate previously published data relating to gene variability in association with known phenotypes using specialized databases (GeneBank). Polymorphism screening of 64 DNA samples using a capillary sequencing protocol will assist with the identification of new polymorphisms. Genotype, allele frequency and linkage disequilibrium of newly identified polymorphisms will be assessed in a sample of 300 individuals. The possible functionality of known and newly identified polymorphisms can then be analyzed using bioinformatic tools. Polymorphisms of interest will then be selected for genotyping in the context of a selected MORGAM cohort. We already have extensive clinical and biochemical phenotyping data on these populations; statistical analysis and data mining techniques will be used to investigate the association between known and novel polymorphisms in candidate genes, obesity and biochemical phenotypes.

INITIAL RESULTS FROM LINKAGE ANALYSES OF SMOKING BEHAVIORS IN AUSTRALIAN TWIN FAMILIES

K. I. Morley,[1,2] S. E. Medland,[1,3] M. A. R. Ferreira,[1] Gu Zhu,[1] D. L. Duffy,[1] G. W. Montgomery,[1] M. T. Lynskey,[1,4] S. A. Treloar,[1] W. D. Hall,[2] A. C. Heath,[4] P. A. F. Madden,[4] and N. G. Martin[1]
[1] *Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Australia*
[2] *Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia*
[3] *Department of Psychology, The University of Queensland, Brisbane, Australia*
[4] *Department of Psychiatry, Washington University St Louis, MO, United States of America*

Cigarette consumption increases the occurrence of atherosclerosis, high blood pressure, and increased blood clotting, making cigarette smoking one of the major risk factors for coronary heart disease and stroke. This increased risk applies not just to smokers themselves, but also to passive smokers. Increasing the effectiveness of smoking cessation treatments is one method for reducing cardiovascular disease risk, and this can be aided by better characterizing genetic influences on smoking behavior.

We present results for linkage analyses of smoking behaviors in a large sample of Australian twins and their families. A range of measures of smoking behavior have been collected from these families through questionnaires and interviews, and a subset of these families has also been genotyped for the purposes of linkage analysis. Depending upon study involvement, individuals from 901 families were typed for up to 1700 markers. Smoking behavior phenotypes were available for individuals from 892 of these families.

Using this sample, we performed univariate linkage analyses of smoking behavior, and also investigated the influence of including or excluding nonsmokers from analyses, which has been suggested to have a strong effect on variability in linkage results. Our initial results replicate some of the significant and suggestive linkage peaks identified by other analyses of smoking behavior.

CURRENT STATUS OF THE MIGRAINE RESEARCH IN GENOMEUTWIN

E. J. C. M. Mulder, D. Posthuma, G. Willemsen, E. J. C. De Geus, and D. I. Boomsma.
Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands

The GenomEUtwin project aims to identify the genes and loci that contribute to complex diseases. In collaboration with clinical experts from inside as well as outside GenomEUtwin we designed a new headache questionnaire for the screening and diagnosis of migraine with aura, migraine without aura, and aura without headache. First results of the Dutch validation study of the headache questionnaire will be presented.

We also present results from three migraine studies that have been performed on Dutch twin-family data. In a nine-year follow-up study we selected a migraine-free control group (n = 725), a short migraine history group (without migraine in 1991 but with migraine in 2000; (n = 146), and a group with a long migraine history (with migraine both in 1991 and 2000; n = 85) to study the relation between migraine and cognitive deficiencies, and the relation between migraine, anxious depression, and neuroticism. We conclude that the high scores of self-reported attention problems, cognitive failures, anxious/depression, anxiety, and neuroticism that we found in subjects with migraine cannot be accounted for by normal ageing or to cumulative detrimental effects of the long-term exposure to attacks. Rather, they are probably related to the onset of migraine as such.

Finally, the first genotypic linkage results that we obtained on existing migraine data will be presented.

INITIAL RESULTS FROM LINKAGE ANALYSES OF LCA-DERIVED MIGRAINOUS HEADACHE IN AUSTRALIAN TWIN FAMILIES

D. R. Nyholt,[1] S. E. Medland,[1,2] M. A. R. Ferreira,[1] K. I. Morley,[1,3] G. Zhu,[1] D. L. Duffy,[1] G. W. Montgomery,[1] A. C. Heath,[4] and N. G. Martin[1]
[1] *Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Australia*
[2] *Department of Psychology, The University of Queensland, Brisbane, Australia*
[3] *Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia*
[4] *Department of Psychiatry, Washington University St Louis, MO, United States of America*

Familial typical migraine is a common, complex disorder that shows strong familial aggregation. Studies indicate that migraine affects up to 25% of females compared with 7.5% of males in Western populations. We utilized latent class and genetic analyses to identify subgroups of migraine/severe headache sufferers in a community sample of 12,245 Australian twins (60% female) aged 28 to 90, who had completed an interview based on International Headache Society criteria.

We performed linkage analyses of LCA-derived recurrent migrainous headache (which has a heritability of 41%) in a large sample of Australian twins and their families. Following the previously described approach of Nyholt et al. 2004 (Genet Epidemiol 26:231–244), latent class-0 and class-1 individuals were considered unaffected, while class-2 and class-3 individuals were considered affected. Here we report results from genome-wide linkage analyses in 397 independent sibpairs (111 affected concordant and 286 discordant for migrainous headache). Results will be discussed in relation to published linkage data.

https://doi.org/10.1375/twin.7.6.680 Published online by Cambridge University Press

ACQUIRED OBESITY CHANGES ADIPOSE TISSUE MRNA EXPRESSION, INCREASES ABDOMINAL AND LIVER FAT, AND CAUSES INSULIN RESISTANCE IN MONOZYGOTIC TWINS

K. H. Pietiläinen,[1,2,3] J. Kaprio,[1,6] A. Rissanen,[2] K. Kannisto,[4] E. Korsheninnikova,[3] E. Ehrenborg,[4] A. Hamsten,[4] A. Häkkinen,[5] and H.Yki-Järvinen[3]
[1] *Department of Public Health, the Finnish Twin Cohort Study, University of Helsinki, Finland*
[2] *Obesity Research Unit, Helsinki University Central Hospital, Helsinki, Finland*
[3] *Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland*
[4] *Department of Medicine, Karolinska Institutet, Stockholm, Sweden*
[5] *Department of Oncology, Helsinki University Central Hospital, Helsinki, Finland*
[6] *Department of Mental Health & Alcohol Research, National Public Health Institute, Finland*

Scientific background. Obesity and insulin resistance are genetically regulated and may cluster to the same individuals due to shared genetic background. Cross-sectional studies in individuals do not permit unequivocal distinction between genetic versus environmental effects on obesity-related endophenotypes and comorbidities. In order to enable such distinctions, the study of twins is clearly advantageous.

Project description. We have studied and summarize here the effects of acquired obesity, independent of genetic influences, in monozygotic (MZ) twins discordant and concordant for obesity. The study in discordant and concordant dizygotic (DZ) twins is currently ongoing with comparable detailed and extensive measures of adiposity and insulin resistance. These twins will be included in the GenomEUtwin pooled data base for the genetic studies.

Main methods. Screening five consecutive birth cohorts (1975–1979) of 22- to 27-year-old Finnish twins (the FinnTwin16 study), we found 14 obesity-discordant (BMI difference ≥ 4 kg/m^2) MZ pairs out of 658. Ten pairs participated in clinical studies. Nine concordant pairs (BMI difference ≤ 2 kg/m^2) were examined as controls. Body fat percentage was determined by dual energy x-ray absorptiometry, abdominal subcutaneous (s.c.) and intra-abdominal (i.a.) fat by magnetic resonance imaging, liver fat by proton spectroscopy, and whole-body insulin sensitivity by the euglycemic clamp. The mRNA expression of 17 genes in biopsies of s.c. adipose tissue was measured by real-time RT-PCR.

Results. In the discordant pairs, the heavier co-twins had 22% greater BMI, 64% more abdominal subcutaneous fat, 93% more intra-abdominal fat, 284% more liver fat, and 40% poorer whole-body insulin sensitivity than the leaner co-twins. The adipose tissue mRNA expression showed enhanced cortisol activity (11β–HSD–1), inflammation (TNFα), and coagulation (PAI–1), as well as decreased insulin sensitivity (adiponectin, PPARγ). Intrapair differences (Δ) in MZ twins are corrected for genetic influences and can be ascribed to acquired (in the broadest sense) environmental etiology only. In all pairs, ΔBMI was significantly correlated with Δbody fat (r = .83, p < .001), Δs.c. fat (r = .97, p < .001), Δi.a. fat (r = .82, p < .001), and Δliver fat (r = .57, p = .010). ΔBody fat was associated with insulin sensitivity (r = –.62, p = .009) and with selected ΔmRNA expression as follows: Δ11β–HSD–1 (r = .65, p = .005), ΔTNFα (r = .54, p = .026), ΔPAI–1 (r = .63, p = .007), Δadiponectin (r = –.49, p = .047), ΔPPARγ.

Conclusions. Acquired obesity increases fat accumulation and insulin resistance in several tissues independent of genetic effects.

MULTIVARIATE LINKAGE METHODOLOGY: AN APPLICATION TO ENDOPHENOTYPES OF CARDIOVASCULAR DISEASE

D. Posthuma, G. Willemsen, N. Kupper, D. I. Boomsma, and E. J.C. de Geus
Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands

We explored power to detect linkage in a multivariate design given the following different conditions: a single factor QTL model, a single factor QTL model with equality of path coefficients, and a common pathway model. Multivariate power under these different conditions was compared to power in a univariate design. QTL effects were modeled as random effects in a variance components context. We illustrate these issues of power with an empirical example involving endophenotypes of cardiovascular disease. Prolonged ambulatory recording of Heart Period (HP), Respiratory Sinus Arrhythmia (RSA), and Respiration Rate (RR) has an inherent repeated measure structure and the three variables are highly genetically correlated. We conducted a multivariate linkage scan for HP, RSA, and RR, measured ambulatory at three different measurements during the day, and 1 measurement during the night. We found significant LOD-scores (> 2) on chromosomes 3, 10, 16, 18 and 20. Finding genes involved in the regulation of ambulatory HP and HP variability may provide new angles for preventive therapy in cardiovascular disease. Under certain conditions, the power to detect disease genes increases sharply when repeated measures of the same variable or different genetically correlated variables can be used.

QUALITY ISSUES IN THE CENTRALIZED SAMPLE HANDLING AND STORAGE UNIT IN GENOMEUTWIN

O. Repola, M. Alanne, M. Jussila, and M. Perola
Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland

Background. The National Public Health Institute of Finland, Department of Molecular Medicine has its own centralized DNA-extraction and storage core unit. The core unit has been operating since 1993 and has participated in the collection of large epidemiological DNA sample collections for genetic analysis. During the last decade DNA core has gained its experience by extracting nearly 200,000 blood samples from numerous national and international projects. Expert knowledge is required from the unit, which has a capacity of handling over 13,000 samples a year. The challenge is to meet the needs of scientists by extracting DNA with flexible methods and providing good quality DNA for downstream analysis. Sample logistics and quality control are an inseparable part of the success in completing large genetic studies. Centralized DNA-extraction and storage unit plays a crucial role for the GenomEUtwin Project. Participating centers in different countries collaborate with the local biobanks to collect blood or DNA samples to be shipped to the DNA extraction core in Helsinki. Blood and DNA samples are stored, extracted, quality checked, aliquoted and finally distributed to the different genotyping laboratories in Finland and Sweden.

Main methods. Three different protocols of extraction are utilized in DNA core unit for GenomEutwin based on two different chemical purifying methods; organic extraction (phenol-cloroform-isoamylalcohol) and manual or automated salt precipitation method. Accurate aliquoting and quality control processes play an important role prior to the distribution of DNA samples. For the quality control, bar coding and rigid sample flow are used to avoid sample mix-up. An automated pipeting robot is used for aliquoting and the DNA aliquots are provided concordant with the specifications of the different genotyping laboratories. All of the aliquoted samples are tested for PCR functionality and monitored for possible sample mix-up or contamination. This is achieved by amplifying sex chromosome–specific PCR fragments and determining the sex by separation on an agarose gel or by producing a fingerprint including sex chromosome–specific markers from the DNA samples with an ABI 3730 DNA analyzer. Centralized storage and distribution center for wide international projects like GenomEutwin offers numerous benefits. The sample management is professional and the analytical problems in the genotyping laboratories relating to the DNA quality and quantity is remarkably reduced.

Results. We are constantly reviewing the quality and quantity of the extracted DNA. Yield and purity are monitored together with gender and contamination checks during the aliquoting process. For GenomEUtwin, most DNA extractions are done for the MORGAM project. In total, 2887 twin samples and 4156 Morgam samples have been processed in the core unit to date. The failure rate for the extractions has been very low, < 0.5 % in twin samples (n = 2022) and 0.7 % in Morgam samples (n = 4156). The average yield of DNA from 1 ml of blood was 34.3 µg and 30.2 µg respectively. During quality control, 2 sample mix-ups have been discovered in the twin cohort and six mix-ups and one contaminated sample in the Morgam cohort.

GENE–ENVIRONMENT INTERACTION BETWEEN WEIGHT AND LIFESTYLE FACTORS IN THE NORWEGIAN TWIN PANEL

T. Rønning and J. R. Harris
Department of Genes and Environment, Division of Epidemiology, the Norwegian Institute of Public Health, Oslo, Norway

Introduction. Gene–environment interactions may be an important source of variation for weight development. Different lifestyle factors such as smoking, drinking and exercise affect weight, with some studies suggesting that a genetic susceptibility to obesity modifies the responses to lifestyle factors. For example, a sedentary lifestyle may have an obesity-promoting effect in men with a genetic predisposition (Heitmann et al., 1997; Am J Clin Nutr 66:672–8). Smoking is also related to weight change. Smokers tend to be leaner than nonsmokers and some individuals may use smoking as a way to control weight gain. One recent study which investigated smoking and change in BMI found sex differences in the effect of smoking on weight change (Heitmann et al., submitted Obesity Res). The goal of this project is to analyze how the effects of smoking and exercise modify the genetic and environmental variance structure on weight and to test for interactions between lifestyle and genes.

Main methods. Subjects were drawn from the Norwegian Twin Panel Data which is a population sample of twins identified through the Norwegian Medical Birth Registry. In 1998, 12,700 twins received a postal questionnaire as part of an ongoing longitudinal study of health and development. The response rate was 63% and included 3334 complete pairs and 1377 single responders. Structural equation modeling was used to analyze raw data in Mx. General sex limitation models were fitted

https://doi.org/10.1375/twin.7.6.680 Published online by Cambridge University Press

to raw data. In addition to main effects, lifestyle factors were modeled as moderator effects. Likelihood ratio tests were used to test the significance of the parameters in the model.

Results. Preliminary results, based on the moderator model proposed by Purcell (2002; Twin Res, 5: 554–71.) reveal that smoking significantly moderates the variance components for additive genetic and shared environmental effects, but not for unique environmental effects. We are currently testing for gene by environmental (G γ E) and gene by gene (G γ G) interactions as well as gene by environment correlations (*r*GE). The same analyses will be repeated to include exercise as a moderator variable.

COMBINED ANALYSIS OF GENOME SCANS FROM SIX TWIN COHORTS TO LOCATE QUANTITATIVE TRAIT LOCI FOR BODY MASS INDEX AND STATURE IN THE GENOMEUTWIN PROJECT

S. Sammalisto,[1] T. Hiekkalinna,[1] N. Martin,[2] J. Harris,[3] D. Boomsma,[4] K. Christensen,[5] K. Ohm Kyvik,[5] N. Pedersen,[6] T. Andrew,[7] T. Spector,[7] E. Widen,[8] A. Palotie,[8] L. Peltonen,[1] and M. Perola[1] on behalf of GenomEUtwin

[1] *Department Molecular Medicine, National Public Health Institute, Helsinki, Finland*
[2] *Queensland Institute of Medical Research, Brisbane, Australia*
[3] *Norwegian Institute of Public Health, Oslo, Norway*
[4] *Vrije University, Amsterdam, Netherlands*
[5] *University of Southern Denmark, Odense, Denmark*
[6] *Karolinska Institutet, Stockholm, Sweden*
[7] *St Thomas Hospital, London, United Kingdom*
[8] *University of Helsinki, Finnish Genome Center, Finland*

Scientific background. Many recent studies focusing on gene mapping of common multifactorial traits have proposed that significantly larger sample sizes may be needed for achieving sufficient statistical power in these efforts. Due to the immense resources needed for sample collection and analysis, many investigators have engaged in large multinational collaborative efforts. The participating twin cohorts in the GenomEUtwin project provide an excellent basis for this type of collaboration in unraveling the genetic and environmental factors underlying common traits.

Project description. As a proof-of-principle study for combined analysis of multiple cohorts, we performed quantitative trait loci (QTL) analyses of body-mass index (BMI) and stature (body height) using genotypic data from six genome-wide scans performed in cohorts from the GenomEUtwin participating countries. The study material consisted of 6635 individuals from 2882 families: Australia (n = 2600), Finland (n = 344), Denmark (n = 271), Netherlands (n = 1086), Sweden (n = 102) and United Kingdom (n = 2232).

Main methods. Since the cohorts differed in their choice of genetic markers, our first task was to combine the genotype information along a shared genetic map across the cohorts. The genetic marker maps were integrated using a custom-made program, Cartographer (www.bioinfo.helsinki.fi/cartographer), which utilizes physical location information from the human genome sequence and interpolation of the genetic distances from the deCODE genetic map, using its markers as an anchoring set. The raw marker data was pooled by a program developed by us, MERGESCAN, which uses the location information from Cartographer to facilitate joint analysis of different cohorts for combined genome-wide analyses. We used the linkage analysis package Merlin for variance components linkage analyses with age, sex and cohort as covariates.

Results. The covariate adjusted heritability of BMI was found to be 58% and of stature 79% in the pooled data set. We found evidence for QTLs on chromosomes 8q (multipoint LOD score 3.14) and 15q (multipoint LOD score 1.62) for stature and on chromosomes 7q (multipoint LOD score 2.75) and 20q (multipoint LOD score 1.53) for BMI in the combined sample. Our results show the value of joint analysis of multiple cohorts in identification of human QTLs.

ASSOCIATION BETWEEN HEIGHT AND CHD MORTALITY: A PROSPECTIVE STUDY OF 35,000 TWIN PAIRS

K. Silventoinen,[1] S. Zdravkovic,[2] A. Skytthe,[3] P. McCarron,[4] A. Herskind,[3] M. Koskenvuo,[5] U. de Faire,[2] N. Pedersen,[2] K. Christensen,[3] and J. Kaprio,[1,6] for the GenomEUtwin Project

[1]*University of Helsinki, Department of Public Health, Finland*
[2]*Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institute, Sweden*
[3]*University of Southern Denmark, Danish Twin Registry, Denmark*
[4]*Department of Epidemiology and Public Health, Queen's University Belfast, United Kingdom*
[5]*Department of Public Health, University of Turku, Finland*
[6]*Department of Mental Health and Alcohol Research, National Public Health Institute, Finland*

Background. An inverse association between height and risk for coronary heart disease (CHD) is well demonstrated but it is not known whether this association is because of common genetic factors, social background, or other environmental factors. To study this question, we investigated this association among large European population-based twin cohorts.

Methods. Four twin cohorts from Denmark, Finland, and Sweden with register-based follow-up data on CHD mortality were pooled. The response rates of the baseline surveys varied from 65% to 85% and the length of the follow-up periods from 25 to 38 years. Together, the cohorts included 74,706 twin individuals (34,942 complete twin pairs) with 5946 CHD deaths during nearly two million person years of follow-up. The data were analyzed by Cox and conditional logistic regression models.

Findings. In individual-level analyses, height was inversely associated with CHD mortality in men (HR = 0.93 per 1 *SD* of height, 95% CI 0.89–0.90) and women (HR = 0.95, 95% CI 0.91–0.99). No heterogeneity was found for sex or zygosity. When we analyzed twin pairs discordant for height and CHD mortality, a twin who had died from CHD was on average shorter that the co-twin, both within monozygotic (OR = 1.23, 95% CI 1.04–1.34 in men and OR = 1.32, 95% CI 1.10–1.59 in women) and dizygotic twin pairs (OR = 1.01, 95% CI 0.91–1.13 in men and OR = 1.14, 95% CI 1.01–1.28 in women).

Interpretations. The inverse association between height and CHD found within monozygotic discordant twin pairs strongly suggest that this association is because of environmental factors directly affecting height and CHD risk. Identifying and intervening to address these factors would have beneficial effects on CHD rates in the population.

SOLUBLE ADHESION MOLECULES, VON WILLEBRAND FACTOR AND RISK OF DEVELOPING TYPE 2 DIABETES MELLITUS: RESULTS FROM THE MONICA/KORA CASE–COHORT STUDY

B. Thorand,[1] J. Baumert,[1] L. Chambless,[2] C. Meisinger,[1,3] H. Kolb,[4] T. Illig,[1] and W. Koenig[5]

[1] *GSF-National Research Center for Environment and Health, Neuherberg, Germany*
[2] *University of North Carolina at Chapel Hill, United States of America*
[3] *Central Hospital of Augsburg, Germany*
[4] *German Diabetes Research Institute at the University of Düsseldorf, Germany*
[5] *University of Ulm, Germany*

Background and aims. Endothelial dysfunction is suggested to be involved in the pathogenesis of type 2 diabetes mellitus. Therefore, the aim of the present study was to investigate prospectively the associations between soluble adhesion molecules (sE-selectin, sICAM-1) and von Willebrand factor (vWF) as markers of endothelial dysfunction and incident type 2 diabetes.

Materials and methods. A case–cohort study was conducted in middle-aged men and women based on data from the MONICA/KORA Augsburg studies conducted between 1984 and 2002. Concentrations of adhesion molecules were measured in stored samples of 532 case subjects with incident type 2 diabetes and 1880 noncase subjects. VWF was measured in 199 cases and 661 noncases, respectively. To analyze associations between markers of endothelial dysfunction and incident type 2 diabetes, hazard ratios (HRs) were estimated by Cox proportional hazards models using the SAS macro ROBPHREG developed by Barlow and Ichikawa (1998).

Results. Men and women with elevated levels of sE-selection had a significantly increased risk of type 2 diabetes after adjustment for age, survey, body mass index, smoking status, alcohol intake, physical activity, systolic blood pressure, ratio of total cholesterol/HDL-cholesterol, C-reactive protein and a parental history of diabetes. Hazard ratios (HRs) and 95% confidence intervals (CIs) comparing tertile extremes of sE-selectin were 2.45 (1.74–3.45) and 1.69 (1.10–2.59) for men and women respectively. Elevated levels of sICAM-1 were also associated with an increased risk of type 2 diabetes in men and women, however the association was independent of other diabetes risk factors in men only (HR and 95% CI for tertile 3 vs. tertile 1: Men: 1.50 [1.07–2.10]; Women: 1.14 [0.75–1.74]). VWF was not associated with the risk of type 2 diabetes.

Conclusions. Our data support the hypothesis that endothelial dysfunction is associated with newly developed type 2 diabetes mellitus.

HOW TO QUANTIFY INFORMATION LOSS DUE TO PHASE AMBIGUITY IN HAPLOTYPE STUDIES

H. Uh, H. Putter, J. Houwing-Duistermaat, and H. van Houwelingen

Department of Medical Statistics and Bioinformatics Leiden University Medical Center, Leiden, the Netherlands

Due to current high-throughput genotyping technologies, there is considerable interest in using Single Nucleotide Polymorphism (SNP) markers to conduct association studies for complex disease. Such studies often involve case–control disease-marker association studies, that is, a sample of affected individuals is compared with a control group to test for association between allelic variants and disease status. A haplotype-based analyses require information about which alleles at each genotyped locus correspond to the unique parental chromosomes transmitted to an individual, and how to assign haplotypes from the observed genotypes becomes a challenging problem. This (missing) 'phase' information can be inferred using statistical algorithm, such as EM.

https://doi.org/10.1375/twin.7.6.680 Published online by Cambridge University Press

As Hodge et al. (1999; Nat Genet 21:360–1) showed, the probability of the individual ambiguity increases with the number of the loci, and with the allele frequencies approaching 0.5. This ambiguity can increase the variance of the estimated haplotype frequencies. Consequently accepting the 'best' configuration of haplotypes from EM-algorithm as the 'real' haplotype might lead to misleading results.

Using all possible configurations of haplotype reconstruction we first quantify the information loss per individual and per haplotype due to phase ambiguity. Then we propose to genotype only the parents of the most informative individuals in the sense of Louis (1982; J R Stat Soc B 44:226–33), which could hopefully lead to a more accurate result and would reduce the genotyping costs and efforts. To demonstrate the relative efficiency of our method we finally add more controls to reach the same magnitude of accuracy.

QUALITY CONTROL MANAGEMENT OF MICROSATELLITE GENOTYPING IN THE GENOMEUTWIN PROJECT

E. Widén, A. Leinonen, A. Lehto, and A. Palotie
Finnish Genome Center, University of Helsinki, Finland

The GenomEUtwin project, encompassing six twin cohorts with over 0.6 million twin-pairs, provides a unique opportunity for verifying genetic variants contributing to common complex disorders. During the first four years 2000 twin pairs will be genotyped for 400 multiallelic markers in two genotyping laboratories. Since commercially available genotyping software are only semiautomatic requiring a substantial amount of manual data editing, proper quality control is imperative to ensure good quality data. The most stringent means to control for genotyping errors is to check for violation of Mendelian inheritance in extended pedigrees. In twin samples only sibling-pairs are genotyped, and segregation check becomes impossible. In order to ensure high quality data, we have implemented quality control procedures to detect both errors associated with sample handling and genotyping. To detect sample mix-ups, plate mix-ups, technical problems related to PCR and electrophoresis, allele calling errors, marker mutations and null alleles, each laboratory will implement a local quality-control procedure including manual rescoring of electrophoresis runs, plate control and duplicate samples in asymmetric shifting positions on each plate, evaluation of the proportion of alleles shared between siblings (GRR) and mutation and error detection through multipoint mapping (SIBMED).

Based on data from 52 monozygotic twin pairs (39,500 genotypes) the allele calling inconsistency was 0.5 discrepancies/1000 genotypes. In a data-set of 654 clinically diagnosed dizygotic twin pairs genotyped for 230 markers, 9 pairs (1.4%) had identical genotypes. The discrepancy rate in GenomEUtwin duplicate samples and MZ twins (10,980 genotypes) was 1.2/1000 genotypes. 45% of the discrepancies were allele calling errors, 54% were related to differences in PCR amplification. Using GRR, seven pairs (1.1%) showed allele sharing proportions consistent with sample mix-up. Overall, almost all detected genotyping errors were picked up by manual rescoring of electrophoresis runs. Since genotyping was done at 10 cM resolution, genotyping errors were not detected through multipoint mapping.

THE ESTABLISHMENT OF A BIOBANK: EXPERIENCES OF THE NETHERLANDS TWIN REGISTER

G. Willemsen, J. Klaren, A. van Bruggen, C. Hoekstra, D. Posthuma, E. J. C. M. Mulder, E. J. C. de Geus, and D. I. Boomsma
Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands

In June 2004 the Netherlands Twin Register started its largest biological sample collection yet. Over the next two years, more than 9000 individuals will be approached for blood and urine sampling. Families selected as informative for the phenotypes that form the focus within GenomEUtwin are included in this enterprise. Family selection was based on various criteria, including the informative value for genetic linkage of height, BMI, lipids, blood pressure and migraine. All registered individuals within a selected family are approached. In addition, a number of spouses are asked to participate to provide a control group of unrelated individuals who (usually) come from nontwin families. In all participants fasting blood samples are collected between 7 a.m. and 10 a.m. at the participant's home. Blood is stored for future DNA and RNA analyses. At the moment, blood and urine samples have been obtained in more than 500 participants. Additional phenotyping during the home visit includes assessment of medication use, height, weight and waist/hip ratio. Additional phenotyping in blood includes lipid levels, CRP, glucose, and insulin. The first results for lipids and anthropometrics in 235 participants show that BMI correlates positively with total cholesterol (r = .15, p = .023) and LDL levels (r = .16, p = .013) and negatively with HDL levels (r = –.22, p = .001).

https://doi.org/10.1375/twin.7.6.680 Published online by Cambridge University Press