ISTITUTO SUPERIORE DI SANITÀ

GenomEUtwin Scientific Meeting

European network of twin registries and MORGAM cohorts

Istituto Superiore di Sanità
Rome, 14-15 December 2006

ABSTRACT BOOK

Edited by
Maria Antonietta Stazi and Cristina D'Ippolito
National Centre of Epidemiology, Surveillance and Health Promotion

ISSN 0393-5620
ISTISAN Congressi
06/C11

Istituto Superiore di Sanità
GenomEUtwin Scientific Meeting. European network of twin registries and MORGAM cohorts. Istituto Superiore di Sanità. Rome, 14-15 December 2006. Abstract book.
Edited by Maria Antonietta Stazi and Cristina D'Ippolito
2006, v, 60 p. ISTISAN Congressi 06/C11

This second GenomEUtwin Scientific Meeting organized in Rome provides an up-to-date on research within the GenomEUtwin Project (Genome-wide analysis of European twin and population cohorts to identify genes predisposing to common diseases) supported by the European Commission under the 5th Framework Programme. The project has developed and applied to twin and MORGAM (MOnica Risk Genetics Archiving and Monograph) cohorts new molecular technologies and statistical strategies to define and characterise the genetic, environmental and life-style risk factors of obesity, migraine, coronary heart disease, which represent major health care problems worldwide. The meeting serves as a high level educational platform for students and junior investigators whose work is selected to be presented in this event. The results obtained during this four year program are presented according to the main phenotypes and the methodological issues considered (molecular biology, bioinformatics, statistics). The ethical aspects related to genetic epidemiological studies are also addressed. Poster sessions will facilitate discussions between young and senior investigators of GenomEUtwin partners.

Key words: Twins, Genetic epidemiology, Cardiovascular diseases, Migraine, BMI, Statistical methods, Genotyping, Bioethics

Istituto Superiore di Sanità
GenomEUtwin Scientific Meeting. Network europeo dei registri dei gemelli e coorti MORGAM. Istituto Superiore di Sanità. Roma, 14-15 dicembre 2006. Riassunti.
A cura di Maria Antonietta Stazi e Cristina D'Ippolito
2006, v, 60 p. ISTISAN Congressi 06/C11 (in English)

Questo secondo Convegno di GenomEUtwin vuole fornire un aggiornamento sulla ricerca condotta nell'ambito del progetto GenomEUtwin (Analisi genomica di coorti di gemelli e di popolazioni europee per identificare geni di suscettibilità a malattie comuni) finanziato dalla Commissione Europea nel Quinto Programma Quadro. Il progetto ha sviluppato e applicato nuove strategie di analisi molecolare e statistica sulle coorti dei gemelli e su quelle MORGAM (*MOnica Risk, Genetics, Archiving and Monograph*) per definire e caratterizzare le componenti genetiche, ambientali e relative agli stili di vita in patologie quali l'obesità, l'emicrania, le malattie cardiovascolari, che costituiscono un rilevante problema di sanità pubblica in tutto il mondo. Il Workshop costituisce un'occasione di alto valore formativo per gli studenti ed i giovani ricercatori che sono invitati a presentare il loro lavoro. Vengono presentati i risultati di questo programma quadriennale, relativi ai fenotipi oggetto di indagine e agli ambiti metodologici presi in considerazione (informatica, statistica, biologia molecolare). Vengono inoltre esaminate le implicazioni etiche degli studi di epidemiologia genetica. La sessione dedicata ai Poster faciliterà la discussione e il dibattito tra i ricercatori giovani e senior che hanno lavorato al Progetto.

Parole chiave: Gemelli, Epidemiologia genetica, Malattie cardiovascolari, Emicrania, BMI, Metodi statistici, Genotipizzazione, Bioetica

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).

Per informazioni su questo documento scrivere a: antonia.stazi@iss.it

Il Rapporto è disponibile online sul sito di questo Istituto: www.iss.it

Presidente dell'Istituto Superiore di Sanità e Direttore responsabile: *Enrico Garaci*
Registro della Stampa - Tribunale di Roma n. 131/88 del 1° marzo 1988

Redazione*: Paola De Castro, Egiziana Colletta* e *Patrizia Mochi*
La responsabilità dei dati scientifici e tecnici è dei singoli autori.

TABLE OF CONTENTS

Programme iii

First session
Cardiovascular risk factors 1

Second session
Body mass index/stature 11

Third session
Methods 19

Fourth session
Ethics 27

Fifth session
Longevity & migraine 31

Poster 37

Authors' index 59

PROGRAMME

Thursday 14 December 2006

9.00 Welcome
Maria Antonietta Stazi

9.10 Keynote lecture
The fruits of GenomEUtwin harvested in global and European research
Leena Peltonen

First session
CARDIOVASCULAR RISK FACTORS

Chairperson: **Mervi Alanne**

10.00 *Common variation in four thrombosis genes and the risk of cardiovascular events*
Kirsi Auro

10.15 *Lipid metabolism related genetic variation and cardiovascular events: prospective, population-based study on MORGAM cohort*
Kati Kristiansson

10.30 *Genome-wide search in Swedish DZ twins for linkage to apolipoprotein A1 level: evidence of female-specific linkage at 15q11-15q13*
Patrik Magnusson

10.45 *Evaluation of apolipoprotein E gene variants as risk factors for cardiovascular disease utilizing Finnish prospective population cohorts*
Kaisa Silander

11.00 Coffee break

Second session
BODY MASS INDEX/STATURE

Chairperson: **Usha Chinappen-Horsley**

11.30 *Fine mapping of A QTL for body height on the human X chromosome*
Andreas Dahlgren

11.45 *Genetic influences on change in body mass index. A longitudinal study of Finnish twins*
Corrado Fagnani

12.00 *Consequences of acquired obesity: gene expression in adipose tissue and serum lipidomic profiles in monozygotic twins discordant for obesity*
Jussi Naukkarinen

12.15 *BMI and waist circumference: expressions of the same genes*
Gonneke Willemsen

12.30 Poster session

Lunch

Third session
METHODS

Chairperson: **Mats Jonsson**

14.30 *A mixture model clustering study of cardiovascular disease risk factors*
Jaana Wessman

14.45 *Meta analysis of four genome scans for serum insulin concentrations*
Mario Falchi

15.00 *A new method for analyzing incomplete longitudinal twin data*
Maria Iachina

15.15 *Efficient calculation of empirical p-values for genome-wide linkage and association through weighted mixtures*
Sarah E. Medland

15.30 Coffee Break

16.00 *TwinNet: a user's perspective*
Markus Perola

Fourth session
ETHICS

Chairperson: **Lorenza Nisticò**

16.30 *Informed consent for future research use of biological samples: the genomEUtwin Italian studies experience*
Virgilia Toccaceli

16.45 *Recasting consent in population genomics biobanks*
Homa Hasan

Friday 15 December 2006

09.30 *The history of a very young species: ours*
Luigi Luca Cavalli-Sforza

Fifth session
LONGEVITY & MIGRAINE

Chairperson: **Qihua Tan**

10.30 *Genetic influence on human lifespan and longevity*
Jacob v.B. Hjelmborg

10.45 *Physical activity is associated with longer white cell telomeres and potentially slower biological ageing*
Janice Hunkin

11.00 *Trait component analysis of Finnish and Australian migraine families reveals a clearer picture of migraine linked loci*
Verneri Anttila

11.15 *A genome-wide linkage study of migraine in a Dutch population sample*
Lannie Ligthart

11.30 Coffee Break

Round table
WHAT'S HOT IN MY LAB

Chairperson: **Leena Peltonen**

12.00 **Dorret Boomsma, Kaare Christensen, Alun Evans, Jennifer Harris, Jan-Eric Litton, Jaakko Kaprio, Nick Martin, Nancy Pedersen, Leena Peltonen, Tim Spector, Maria Antonietta Stazi, Anne-Christine Syvanen**

13.00 Steering Committee

Working Lunch

First session

Cardiovascular risk factors

Chairperson
Mervi Alanne

COMMON VARIATION IN FOUR THROMBOSIS GENES AND THE RISK OF CARDIOVASCULAR EVENTS

K. Auro (a), K. Kristiansson (a), M. Alanne (a), J. Wessman (a), K. Silander (a), V. Salomaa (b), L. Peltonen (a), M. Perola (a)
(a) Department of Molecular Medicine, KTL, Biomedicum, Helsinki, Finland
(b) Department of Epidemiology and Health Promotion, KTL, Helsinki, Finland

Genetic base of complex diseases consists of several predisposing, low-penetrance risk factors. Lately, gene-gene interaction studies, combining data from predisposing markers, have attempted to enlighten the polygenic structure of complex traits. Coronary heart disease and ischemic stroke are typical complex diseases, where multiple low risk loci have been reported. Our aim was to investigate gene-gene interactions in a physiological thrombosis cascade of four genes using two large, independent, prospective cohorts of the same population. We hypothesized that common variants of these genes, well known to predispose to venous thrombosis, would also contribute to thromboembolic events of arterial side.

Binding of thrombin to its receptor, thrombomodulin (THBD), activates protein C (PROC), which in turn inactivates factor V (F5). This cascade, when functioning correctly, leads to fibrinolysis. Defects in the cascade are involved in venous thrombosis, the most famous example being factor V Leiden mutation. In arterial thrombosis, however, the role of these genes is less well established. Recent results have indicated that soluble THBD together with soluble intercellular adhesion molecule -1 (ICAM1) would contribute to cardiovascular (CVD) events. In our study, no association of THBD variants with CVD was observed.

We selected 43 single nucleotide polymorphisms (SNPs) to cover all common (>5%) haplotype bins in these four genes. From 14140 FINRISK participants, 2222 (FINRISK-92 n=999 and FINRISK-97 n=1223) were selected for genotyping, including all cardiovascular cases and random subcohorts. To identify potential risk variants and their interactions, we used recursive partitioning techniques: in classification and regression tree analysis we observed altogether 12 SNPs contributing to CVD together with the traditional risk factors. Next, we selected all SNPs identified in the trees grown, and estimated their risks for incident CVD events in Cox's proportional hazards model, adjusted for traditional CVD risk factors. Only variants showing consistent effect in both cohorts and their combination in Cox's model were classified as significant. Interaction of these variants was tested with an interaction term, as deviation from multiplicative model.

This approach identified several variants predisposing for CVD events in our study cohorts. One of the most interesting findings was association of F5 Leiden mutation with ischemic stroke events (HR 2.95, CI 95% 1.17 – 7.43, $p<0.05$ for combined data). Another factor V variant was found to interact with a variant of protein C. Our findings suggest that combining analysis methods reduces the risk of false positive findings and may be useful when studying gene-gene interactions.

LIPID METABOLISM RELATED GENETIC VARIATION AND CARDIOVASCULAR EVENTS: PROSPECTIVE, POPULATION-BASED STUDY ON MORGAM COHORT

K. Kristiansson (a), M. Alanne (a), K. Auro (a), K. Silander (a), V. Salomaa (b), M. Perola (a), L. Peltonen (a)
(a) Department of Molecular Medicine, KTL, Biomedicum, Helsinki, Finland
(b) Department of Epidemiology and Health Promotion, KTL, Helsinki, Finland

Allelic variation in genes from lipid metabolism pathways can result in diverse effects on target cells. Several critical genes in lipid and glucose metabolism are under the control of ubiquitously expressed upstream transcription factor 1 (USF1). Another gene, lipin 1 (LPIN1), encodes a novel molecular protein which is expressed by adipocytes and has marked effects on adipose tissue mass. Our group has earlier shown that LPIN1 also associates with BMI and serum insulin levels in obese individuals and dyslipidemic families (Suviolahti et al.2006). In those dyslipidemic families, another gene, Forkhead box C2 (FOXC2), was shown to associate with high-density lipoprotein cholesterol (HDL-C) levels.

Although evidence exists on their involvement in aetiology of conditions predisposing to early CVD, such as familial combined hyperlipidemia (FCHL) or obesity, the potential effects of these genes on the risk for CVD events at the population level have not been established. Here we report the results from prospective genetic-epidemiological studies on the significance of these genes as risk factors for obesity, variation in serum lipid levels, cardiovascular disease (CVD), and mortality in two large Finnish MORGAM cohorts.

Using a prospective case-cohort design, we genotyped haplotype-tagging single nucleotide polymorphisms (htSNPs) of the genes in two distinct cohorts, followed up during 1992-2001 (FINRISK92) and 1997-2003 (FINRISK97). The total number of follow-up years was 112435 in 14140 individuals, of which 2222 were selected for genotyping based on the case-cohort study strategy. Cox's proportional hazards analysis measuring time-to-event was used to estimate the risk of a CVD event or mortality during the follow-up period in relation to genotype groups and haplotypes.

After adjustment for conventional risk factors, we observed an association of USF1 with CVD and mortality among females: In combined analysis of the two cohorts, female carriers of a USF1 risk haplotype had a two-fold risk of a CVD event (hazard ratio 2.02, 95% confidence interval 1.16-3.53, p=0.01) and an increased risk of all-cause mortality (2.52, 1.46-4.35, p=0.0009). Both in a subcohort (N=786) representative of the original follow-up cohorts of 14140 Finns and in incident CVD cases of the follow-up periods (N=528), we observed an association between LPIN1 and BMI, total cholesterol, triglycerides, HDL-C, and total cholesterol minus HDL-C (P=0.03-0.003). Further, results from Cox proportional hazards analysis suggested that allelic variation of LPIN1 contributes to the risk of CVD (P=0.002-0.05). FOXC2 SNPs associated with BMI, waist-to-hip ratio, and HDL-C and triglycerides in the FINRISK92 subcohort (p=0.04-0.003).

FOXC2 also associated with the prospective risk of CVD and mortality among males (p=0.04-0.008).

Our studies demonstrate that genes identified in exceptional families or even in mice, are affecting morbidity and mortality also at the population level. Our data provides new candidate genes to be tested in additional GenomEUtwin cohorts for CVD and related risk factors.

GENOME-WIDE SEARCH IN SWEDISH DZ TWINS FOR LINKAGE TO APOLIPOPROTEIN A1 LEVEL: EVIDENCE OF FEMALE-SPECIFIC LINKAGE AT 15Q11-15Q13

M. Boman, N.L. Pedersen, P.K.E. Magnusson
Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

The effect of genetic variants underlying atherosclerosis is thought to be mediated and mirrored by intermediate phenotypes such as plasmatic cholesterol levels. Localization of quantitative trait loci (QTLs) influencing levels of plasmatic lipids and (apo)lipoproteins may aid in the search for determinants of susceptibility to complex disorders such as atherosclerotic diseases. Apolipoprotein A1 (Apo A1) is the primary protein constituent of high density lipoprotein (HDL) and is considered to be responsible for the anti-atherogenic effect of HDL. Several previous investigations have provided evidence that both major and minor genetic factors are involved in determining levels of Apo A1. Because of the strong relation between risk of future cardiovascular events and level of Apo A1, identification of chromosomal regions harboring genes accounting for variation in Apo A1 levels between individuals is decidedly valuable. We here describe an effort to map QTLs influencing Apo A1 levels. More than five hundred fraternal twin pairs were recruited from the Swedish Twin Registry. Ascertainment of pairs followed one of two schemes, a lesser part (N=67 pairs) was from a pilot study in which dizygotic (DZ) twins pairs were collected if at least one twin had suffered a myocardial infarction or been subjected to revascularization by the time of the study initiation. Ascertainment of the remaining pairs (N=437) followed age-restricted (>67 years) but otherwise random ascertainment. Apo A1 was measured by Karolinska University Laboratory in blood samples collected at local health care centers. DNA extraction, sample storage and data collection has been handled by KI Biobank. Genome-scanning with over 1000 micro-satellite markers was successful in 1008 individuals of 504 included pairs. All genotyping was performed by deCODE genetics. Variance component linkage analysis of ApoA1 levels adjusted for age, sex and fasting-status was undertaken to map QTLs. In order to minimize effects due to non-normality of the Apo A1 distribution, analyses were repeated after Box-Cox transformation had been employed. Body mass index (BMI) is correlated to level of ApoA1. To increase power to detect loci that influence ApoA1 levels independently of BMI, separate analyses were performed when BMI was additionally included as a covariate into the variance component model. The strongest evidence of linkage in the total material was obtained on chromosome 6p21-12 (lod=2.59). Another locus showing suggestive evidence of linkage were 12q23 (lod=2.38). Recent emphasis on the importance of genotype by sex interaction for QTL mapping encouraged us to investigate sex-specific effects. This was done in analyses stratified by sex (female-female pairs N=144, male-male pairs N=129). Significant female-specific linkage was detected near marker D15S156 on chromosome 15q11-15q13 (lod=4.01). Moreover, suggestive evidence of sex specific linkage was obtained on

chromosome 16p13 (lod=3.09) and on chromosome 13q12-13q13 (lod=2.36) for females and males, respectively. The finding on chromosome 6p21-12 in the present study lends further support to several previous studies reporting QTLs for HDL and other phenotypes related to atherosclerosis at this locus. Although we are aware of the increased risk of false positive signals due to the extra dimension of testing arising from the stratification by sex, we believe the high lod-score observed on chromosome 15q11-15q13 warrants further investigations.

EVALUATION OF APOLIPOPROTEIN E GENE VARIANTS AS RISK FACTORS FOR CARDIOVASCULAR DISEASE UTILIZING FINNISH PROSPECTIVE POPULATION COHORTS

K. Silander (a), K. Kristiansson (a), M. Alanne (a), K. Auro (a), K. Kuulasmaa (b), J. Virtamo (b), M. Perola (a), V. Salomaa (b), L. Peltonen (a)

(a) Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland

(b) Department of Health Promotion and Chronic Disease Prevention, National Public Health Institute, Helsinki, Finland

We estimated the effect of genetic risk factors on cardiovascular disease (CVD) in prospectively followed population cohorts. Our primary study samples were two FINRISK cohorts participating in the MORGAM study. FINRISK surveys are carried out every 5 years to assess the prevalence and risk factors of CVD in Finland. Baseline information on all randomly sampled individuals includes anthropometric measurements, serum lipids, blood pressure and questionnaire on CVD risk factors. Information on fatal and non-fatal coronary and stroke events and all-cause mortality during the follow-up period is obtained from national registers. We utilized the FINRISK-92 cohort (n=5999), which has been followed up for 10 years. From this large cohort, we analyzed a case-cohort sample consisting of a total of 196 incident CHD cases, 68 incident ischemic stroke cases, 282 deaths, and 305 non-case subcohort individuals. As an additional independent case-cohort sample we used 205 incident CHD cases, 80 incident stroke cases, 328 deaths, and 281 non-case subcohort individuals from the FINRISK-97 cohort (n=8141), which has been followed up for 7 years. We selected for study 47 genes likely involved in cardiovascular pathobiology based on their function, previous association with cardiovascular disease, and/or relevant phenotype in animal models. These genes represent a wide array of pathways, including lipid and energy metabolism, inflammation, coagulation, and thrombosis. For each gene, we genotyped a set of haplotype-tagging SNPs, and additional SNPs previously reported to associate with CVD. Time-to-event analysis was used to assess whether allelic variants confer risk for CHD, CVD, or all-cause mortality. Hazard ratios were calculated with a multivariate Cox's proportional hazards model, adjusting for established CVD risk factors. We also tested whether allelic variants were associated with various quantitative variables measured at baseline. In this summary I will highlight results from the apolipoprotein E (ApoE) gene. Results on additional genes will be presented at the meeting. ApoE is a key regulator of plasma lipid levels, mediating the clearance of chylomicron and very low density lipoprotein cholesterol remnants from the circulation. Two SNPs in the ApoE gene yield three common epsilon isoforms, E2, E3 and E4. These variants have been extensively studied in relation to lipid levels, cardiovascular disease and Alzheimer's disease. Our analysis confirmed previous association of epsilon variants with lipid variables, including non-HDL cholesterol (for subcohort p=0.04 for FR92, p=0.004 for FR97), total to HDL cholesterol ratio (for subcohort p=0.04 for FR92, p=0.02 for FR97), and apolipoprotein B (data only for FR97, p=0.0034 for subcohort). However, in

time-to-event analysis no significant association with CVD event was observed for any of the 8 ApoE variants studied. We also analyzed the effect of the epsilon variants in a third Finnish cohort, ATBC, of men aged 50-69 years at time of recruitment (1985-1988) who were daily smokers. The follow up period was from 1992 to 1999. This cohort consisted of 393 CHD and 211 ischemic stroke incident cases, and 646 non-case subcohort individuals. Similarly to the FINRISK samples, the associations to CVD events were not significant, but a strong association was found with lipid variables: for total cholesterol levels p=0.000009 in subcohort (mean values: E2 carriers 5.434, E3 homozygotes 5.872, E4 carriers 6.102 mmol/L); for non-HDL cholesterol p= 0.0000004 in subcohort (mean values: E2 carriers 4.181, E3 homozygotes 4.694, E4 carriers 4.942 mmol/L). These results suggest that genetic variants in ApoE may contribute to cardiovascular disease risk through their effect on lipid levels, but the variants are not independent CVD risk factors.

Second session

Body mass index/stature

Chairperson
Usha Chinappen-Horsley

FINE MAPPING OF A QTL FOR BODY HEIGHT ON THE HUMAN X CHROMOSOME

A. Dahlgren (a), M. Perola (b), U. Liljedahl (a), J. Kaprio (c), T.D. Spector (d), N.G. Martin (e), L. Peltonen (b,f), A.C. Syvänen (a)

(a) Department of Medical Sciences, Uppsala University, Uppsala, Sweden
(b) Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland
(c) Department of Public Health, University of Helsinki, Helsinki, Finland
(d) Twin Research & Genetic Epidemiology Unit, St. Thomas' Hospital, London, United Kingdom
(e) Genetic Epidemiology Unit, Queensland Institute of Medical Research, Brisbane, Australia
(f) Department of Medical Genetics, University of Helsinki, Helsinki, Finland

Standing body height is one of the most basic human phenotypes and complex genetic traits. It has high heritability estimated to around 90% based on twin studies (Silventoinen et al, Twin Res. 5:399, 2003). Data on body height is readily available in many existing sample sets, and could serve as a model for other complex genetic traits. Therefore body height was selected as one of the first phenotypes to be investigated in the GenomEUtwin project. Using combined data from genome scans performed with microsatellites in Finnish, British and Australian twin cohorts, a region on the X chromosome was selected for further fine mapping using single nucleotide polymorphisms (SNP) as markers. About 1500 SNPs covering ~9Mb centered on the marker with the highest combined LOD score were selected for genotyping. The SNPs were selected based on their nucleotide position to be located at an average spacing of 5 kb, validation status and minor allele frequency in dbSNP. A design score which predicts the performance of the SNPs in the Golden Gate assay (Illumina) was also used as a selection criterion. The majority of the selected SNPs are validated in European populations, including 961 SNPs that have been genotyped in the HapMap project. The genotyping was performed using the Golden Gate assay, which utilizes Illumina Bead Array technology in a 96-well format and enables multiplex genotyping of up to 1536 SNPs in each well. This SNP panel was first genotyped in 780 samples from the Finnish twin cohort that were analyzed in the original genome scan. The genotyping was assay conversion rate was about 90%. The genotype data is currently being analyzed using standing adult body height as the quantitative trait looking at both linkage and association.

GENETIC INFLUENCES ON CHANGE IN BODY MASS INDEX. A LONGITUDINAL STUDY OF FINNISH TWINS

J.v.B. Hjelmborg (a), C. Fagnani (b), K. Silventoinen (c), M. McGue (e), M. Korkeila (c), K. Christensen (a), A. Rissanen (c), J. Kaprio (c,d)

(a) Statistics and Epidemiology, Institute of Public Health, University of Southern Denmark, Odense, Denmark

(b) National Centre of Epidemiology, Surveillance and Health Promotion, Italian Institute of Health, Rome, Italy

(c) Department of Public Health, University of Helsinki, Helsinki, Finland

(d) Department of Mental Health and Alcohol Research, National Public Health Institute, Helsinki, Finland

(e) Department of Psychology, University of Minnesota, Minneapolis, MN, USA

Objective: to investigate the interplay between genetic factors influencing individual level and changes in BMI in adulthood.

Design: longitudinal study of twins.

Subjects: the cohort of Finnish twins (N=10556 individuals) aged 20 to 46y at baseline.

Measurements: data on weight and height was obtained from mailed surveys in 1975, 1981 and 1990.

Results: latent growth models revealed a substantial genetic influence on BMI level at baseline in males and females (heritability (h2) is 80% (95% CI 0.79 to 0.80) for men and h2=82% (0.81,0.84) for women) and a moderate to high influence on rate of change in BMI (h2=58% (0.50,0.69) for men, and h2=64% (0.58,0.69) for women). Only a weak indication of genetic pleiotropy was observed; the genetic correlation between baseline and rate of change in BMI was very modest 0.070 (–0.13, –0.068) for males and 0.04 (0.00,0.08) for females).

Conclusion: our population based results provide basis for identifying genetic variants for change in BMI, in particular weight gain. Furthermore, it is demonstrated for the first time that such genetic variants for change in BMI are likely to be different from those affecting level of BMI.

CONSEQUENCES OF ACQUIRED OBESITY: GENE EXPRESSION IN ADIPOSE TISSUE AND SERUM LIPIDOMIC PROFILES IN MONOZYGOTIC TWINS DISCORDANT FOR OBESITY

J. Naukkarinen (a), K.H. Pietiläinen (b,c,d), M. Oresic (e), H. Yki-Järvinen (c), A. Rissanen (b), J. Kaprio (c,f), L. Peltonen (a)

(a) Department of Molecular Medicine, National Public Health Institute, Finland and Department of Medical Genetics, University of Helsinki, Finland

(b) Obesity Research Unit, Helsinki University Central Hospital, Finland

(c) Finnish Twin Cohort Study, Department of Public Health, University of Helsinki, Finland

(d) Department of Medicine, Helsinki University Central Hospital, Finland

(e) VTT Technical Research Centre of Finland, Espoo, Finland

(f) Department of Mental Health and Alcohol Research, National Public Health Institute, Helsinki, Finland

Background: The metabolic consequences of obesity involve complex biochemical signaling and information exchange through genetic and metabolic networks. We have used a unique study sample of monozygotic twin pairs discordant for obesity, with the idea to control for genetic background in the obese and non-obese subjects. We performed genome-wide transcript profiling of adipose tissue from fat biopsies and produced lipidomics profiles of serum to uncover the physiological processes in acquired obesity.

Methods: Fourteen MZ pairs discordant for obesity (10 to 25 kg difference in weight) and ten weight concordant control pairs were identified among 658 MZ twin pairs, aged 24-27 years, from the population-based FinnTwin16 study. Body composition was determined by DEXA, MRI and spectroscopy, and insulin sensitivity by the euglycemic clamp technique. Transcript profiling of abdominal subcutaneous adipose tissue was performed using Affymetrix U133 Plus 2.0 GeneChip arrays in the discordant pairs. Global characterization of lipid molecular species in serum was performed by liquid chromatography coupled to mass spectrometry in all pairs.

Results: In cluster analyses the transcript profiles of adipose tissue were grouped based on the amount of fat tissue, not based on the twinning, underlining the importance of the local regulatory factors in fat even in individuals with identical genetic background. BMI and relative fat content was found to be related to expression of genes that have previously been associated with inflammation/immune response, cardiovascular diseases, and cancer as well as to expression of several genes with no defined function in adipose tissue. In serum lipids, obesity was primarily related to an increase in lysophosphatidylcholines, lipids found in proinflammatory and proatherogenic conditions and to decrease in either phospholipids which are known to have antioxidant properties. These changes were associated with insulin resistance, characteristic of acquired obesity in the young and

presumably healthy twins. We are currently exploring the regulatory pathways in adipose tissue and serum lipid metabolism which link obesity with its metabolic consequences.

Conclusions: Acquired obesity, independent of genetic effects, changes transcript profiles of adipose tissue and lipid profiles in serum. Our findings may help detect novel potential molecular and biochemical links between obesity, diabetes, atherosclerosis and cancer.

BMI AND WAIST CIRCUMFERENCE: EXPRESSIONS OF THE SAME GENES

G. Willemsen, D.I. Boomsma, E.J.C. de Geus, A. van Bruggen, D. Posthuma
Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands

Body mass index (BMI) and waist circumference have both been used in studies of obesity. There is a strong association between these two measures and both measures have been shown to be determined for a large extent by genetic factors, i.e. by a set of common genes. No studies as yet have determined whether linkage analyses indeed deliver the same results for BMI as for waist circumference. We examined this question by using data from two studies of the Netherlands Twin Register in which BMI and waist circumference were measured simultaneously. Body composition data are available for more than 1089 twins and siblings, coming from 491 families and including 75 male monozygotic and 105 female monozygotic twin pairs. In addition, whole genome scan data are available. BMI and waist circumference are highly correlated ($r=.82$, $p<.001$). The pattern of twin correlations suggests equal heritability for BMI and waist circumference, with lower heritability estimates for women than for men. The covariance of BMI and waist circumference is predominantly explained by common genetic factors (76%) and for the remainder by unique environmental factors. Bivariate linkage results indicate common genes on chromosome 9 and chromosome 6, while univariate linkage scans also indicate chromosome 13 for BMI, but not for waist circumference. In conclusion, locating the genes for overweight is not dependent on the operationalisation of body composition as BMI or waist circumference.

Third session

Methods

Chairperson

Mats Jonsson

A MIXTURE MODEL CLUSTERING STUDY OF CARDIOVASCULAR DISEASE RISK FACTORS

J. Wessman (a,b), K. Auro (b), K. Kristiansson (b), M. Alanne (b), K. Silander (b), M. Perola (b), V. Salomaa (c), H. Mannila (a), L. Peltonen (b)

(a) Department of Computer Science and Helsinki Institute of Information Technology, University of Helsinki, Finland

(b) Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland

(c) Department Health Promotion and Chronic Disease Prevention, National Public Health Institute, Helsinki, Finland

In this study we performed a clustering analysis on Finnrisk 1997 cohort sample. The aim was to find out if the individuals in the cohort can be automatically divided into groups which exhibit distinctly characteristic phenotype patterns. We can then use these groups in association analysis, to find out if candidate genes associate to a particular phenotype pattern or "endophenotype". In data mining, clustering in general means automated division of individuals into subgroups ("clusters") such that people in each subgroup are as similar to each other as possible, while being as different from individuals in the other groups as possible. Many methods exist for cluster analysis; some best known examples are the k-means and hierarchical clustering methods often used in gene expression analysis.

Mixture model clustering has several advantages, most importantly 1) the model assumptions are explicit and the model can be interpreted as a description of the underlying population, not only of the individuals analyzed, and 2) missing data can be easily handled by iterative imputation in the context of the model. In this method, we assume that the population clusters are each drawn from some distribution of a particular form, but with different parameters for each group. If we knew the distributions for each cluster, we could easily calculate a probabilistic cluster assignment for each individual based on their data likelihood given the cluster distributions. On the other hand, if we knew the cluster assignment of each person, we could easily calculate the cluster distributions.

Our task, however, is to find out both at the same time; this can be done iteratively via the Expectation-Maximation technique. We have now constructed a first clustering of the full Finnrisk 1997 cohort (n=8141) using mixtures of multivariate normal distributions. Variables included were systolic and diastolic blood pressure, waist-hip-ratio, body mass index, total cholesterole, HDL, triglycerides, apolipoprioteins AI and B, and C-reactive protein. All variables were adjusted for age. In the phenotype analysis phase, we did not use any information about the diagnoses the individuals had, as the aim was to see if the found phenotype patterns can be used to predict disease.

We selected the best number of clusters and analyzed the sensitivity of the clustering solution found to outliers by 10-fold crossvalidation. In the results we see five clusters, varying in number of individuals in each from 231 (smallest) to 4363 (largest). Prevalence of cardiovascular disease (either existing before the recruitment of the cohort, or indident during the followup) in the clusters varies from 12.5 to 8.7 percent (10.6 percent in the sample in general). Preliminary analyses show that most risk factors follow the CVD prevalence figures as you would expect. There are some interesting observations though.

As an example, while the cohort in general has 2.4 percent prevalence of type II diabetes, the cluster with lowest CVD prevalence actually has a diabetes prevalence of over 6 percent. Next, we intend to further characterize the differences between the groups, to analyze what variables best predict future CVD incidents (myocarcial infarction, stroke) inside each cluster, and to perform association analysis of already genotyped candidate genes to the cluster assignments inside the MORGAM case/cohort sample which is a subgroup of this cohort. For validation of the results, we will perform similar analysis in the 1992 and/or 2002 cohorts.

META ANALYSIS OF FOUR GENOME SCANS FOR SERUM INSULIN CONCENTRATIONS

M. Falchi (a), D.I. Boomsma (b), K.O. Kyvik (c), N.G. Martin (d), G. Montgomery (d), D. Posthuma (b), K. Schousboe (c), T.I.A. Sørensen (e), J. Whitfield (d,f), G. Willemsen (b), T.D. Spector (a)

(a) Twin Research & Genetic Epidemiology Unit, St. Thomas' Campus, Kings College London, United Kingdom

(b) Department of Biological Psychology, Free University of Amsterdam, The Netherlands

(c) The Danish Twin Register, Epidemiology, IPH, University of Southern Denmark, Odense, Denmark

(d) Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia

(e) The Danish Epidemiology Centre at the Institute of Preventive Medicine, Copenhagen University Hospital, Denmark

(f) Clinical Biochemistry, Royal Prince Alfred Hospital, Sydney, Australia

Background: elevated serum insulin concentration is a risk factor for complex diseases such as obesity, type-2 diabetes, and cardiovascular disease. Occasionally, single mutations can lead to pathological increase of serum insulin concentrations, but the frequency of these variants in the general population is usually low. The identification of common genetic determinants of insulin levels could provide important insights into strategies for the prevention of hyperinsulinemia. Multicenter studies such as the GenomEUtwin project can help in the identification of genetic factors in European populations and to decrease false positives outcomes.

Methods: we performed genome-wide linkage analyses for insulin concentrations in four study groups from the GenomEUtwin multicenter networks (United Kingdom, Netherlands, Denmark, and Australia). The sample comprised 2,992 genotyped individuals in 1,449 families (the contribution to the overall sample was 60% UK, 20% AU, 10% ND, 10% DK). Since trait model, markers density, and allele frequencies differ among populations, instead of gathering the data from all contributing studies into a singe large dataset we performed multipoint variance components linkage analyses in each sample separately. The results from all countries where then combined together through a rank-based genome-scan meta-analysis (GSMA) method, in order to identify quantitative trait loci (QTLs) influencing serum insulin concentrations among populations. GSMA identifies linkage evidence across chromosomal segments, allowing for the uncertainty in the maximum LOD-score localization expected for complex traits. Empirical p-values were assessed though 10,000 permutations.

Results: the strongest evidence for linkage was found on 17q21-qter (empirical P=0.0009, reaching suggestive genome-wide significance after correction for multiple testing) in a chromosomal segment spanning about 45 cM. Defining narrower segments in the meta-analysis identified evidence for linkage in adjacent bins, diluting the significance of the test in the region. Nominal evidence from GSMA was also observed in 2q (P=0.02).

Conclusion: the multicenter meta-analysis of 2992 subjects provides good evidence of a QTL for insulin on chromosome 17q. Collection of further data from these and other studies will help confirming and refining the localization of this QTL and highlighting the genes involved.

A NEW METHOD FOR ANALYZING INCOMPLETE LONGITUDINAL TWIN DATA

M. Iachina (a), B. Joergensen (b), K. Christensen (c)
(a) Epidemiology Unit, Institute of Public Health, University of Southern Denmark, Odense, Denmark
(b) Department of Statistics, University of Southern Denmark, Odense, Denmark
(c) Epidemiology Unit, Institute of Public Health, University of Southern Denmark, Odense, Denmark

In this work, a new method for analyzing incomplete longitudinal twin data will be presented. This method is based on inverse probability weighted generalized estimating equations and allows having different types of missing data in the sample. A new approach was applied to estimate the effect of genes on the influence of stressful events on the depression symptomatology in the elderly. In the analysis the sample from the Longitudinal Study of Aging Danish Twin (LSADT) and the Danish National Civil Register will be used. Assessments were made every second year and 4,484 persons were followed. Statistical analyses will be performed using both the maximum likelihood procedure implemented in MX, (usually done in twin studies) and the Second Inverse Probability Weighted Generalized Estimating Equations (IPWGEE2) approach. We will evaluate the performance of both approaches in the simulation study. The advantages and disadvantages of the maximum likelihood and the IPWGEE2 approaches will be discussed.

EFFICIENT CALCULATION OF EMPIRICAL P-VALUES FOR GENOME-WIDE LINKAGE AND ASSOCIATION THROUGH WEIGHTED MIXTURES

S.E. Medland (a,b), J.E. Schmitt (a), B.T. Webb (a), P.H. Kuo (a), M.C. Neale (a,c)
(a) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University Medical Center, USA
(b) Genetic Epidemiology Unit, Queensland Institute of Medical Research, Australia
(c) Department of Biological Psychology, Free University of Amsterdam, The Netherlands

Currently the two most popular methods for obtaining empirical p-values for genome-wide linkage and association involve performing 'gene-dropping' simulations or permutation tests (in which the coefficient of genotypic sharing, is randomized across relative pairs) at each locus. However, as the time required to obtain empirical significance is directly proportional to the time taken to complete the initial analyses, the time required to obtain empirical p-values for multivariate analyses using these methods is prohibitive. Utilizing the distributional dependencies between and the permutation test we report a new method of efficient permutation. In summary, the distribution of at locus j is estimated as a weighted mixture of drawn from a pool of 'modal' distributions. This weighting scheme is then applied to the distribution of the permutation tests at the 'modal' loci to obtain an empirical p-value at locus j (which is asymptotically distributed as the permutation test at loci j). This weighted-mixture approach greatly reduces the number of permutation tests performed making it suitable for use in multivariate analyses and other computationally intensive linkage and association analyses. In addition, as the distribution of is a property of the genotypic data for a given sample and is independent of the phenotypic data, the weighting scheme can be applied to any phenotype (or combination of phenotypes) collected from that sample.

Fourth session

ETHICS

Chairperson
Lorenza Nisticò

INFORMED CONSENT FOR FUTURE RESEARCH USE OF BIOLOGICAL SAMPLES: THE GENOMEUTWIN ITALIAN STUDIES EXPERIENCE

V. Toccaceli (a), L. Giannantonio (a,b), L. Nisticò (a)
(a) National Centre of Epidemiology, Surveillance and Health Promotion, Italian Institute of Health, Rome, Italy
(b) University of Tor Vergata, Rome, Italy

Within the ethic perspective, Informed Consent is an effective tool whenever the research is a clinical and experimental one. It has limits when applied in the field of genetic epidemiology with the enrolment of large population section, mostly healthy individuals, and above all when the storage of biological samples is required beyond the aims of the specific study. The required level of specification and explanation of objectives and rationale is impossible to be reached because research to be conducted on stored biological samples is for its nature "open ended". We evaluated the main opposite positions and judgements for what is currently defined a "broadened consent" in the framework of the international bioethical debate on research needs and autonomy and respects of individuals, taking into account three important factors:

- legal requirements recently issued in our Country by the legislative decree n. 196/2003, that makes more strict privacy and confidentiality for personal and sensitive data, giving participants a more active role also in the scientific environment;
- the specific nature of the genetic data, which receive particular attention at national and international level;
- the storage of biological materials for a bio-bank project at the Italian Institute of Health.

Finally, we adopted a "broadened consent" model for GenomEUtwin studies, maintaining, as the baseline, study-specific consent requirements and adding further elements to let participants give or deny their consent for the storage of their biological samples. Euroclot study, in particular, constituted the phase of experimentation of this new model. This new consent has main elements, specific information as well as limits and faults which reflects partly a choice of the researchers' team and partly technical and scientific general lack of knowledge on specific problems (such as time limits for safe storage), all these will be discussed in the presentation. In our experience the adoption of a "broadened consent" was a positive experience for enrolment but administering it make it clear that the tool of Informed Consent, especially for large population studies and future research use of stored biological samples, is effective and responds to its goal only if it is envisaged as a starting point for dialoguing with participants. The human researcher - participant relationship is pre-eminent to make the research go on. Further empirical investigation about the effectiveness of this work, the understanding of the aims of the studies, as well as about participants' very attitude to donation for research is going to be carried out.

RECASTING CONSENT IN POPULATION GENOMICS BIOBANKS

H.S. Hasan, J.R. Harris
Division of Epidemiology, The Norwegian Institute of Public Health, Oslo, Norway

Since GenomEUtwin began there has been growing recognition worldwide of the value of national and international biobanks for studying complex diseases. For example, harmonization of biobank platforms to facilitate sharing of data and samples is a high priority area in the EU's 7th framework programme (FP7). In the wake of these biobank initiatives, and coupled with advances in genomics, ethical considerations for population genomics continue to evolve. The Ethics Core of GenomEUtwin is working closely with other international projects to address the changing nature of these issues in order to build more useful and harmonised platforms from which high quality research can be conducted and long-term participation of research subjects can be secured.

Engaging research participants and relying on their continual support for long-term biobanking projects is vital to the success of biobank research. While this process rests critically upon informed consent, the concepts behind informed consent are moving targets. In this presentation we focus on two key aspects of informed consent that illustrate such shifts. Specifically, we discuss how the meaning of 'informed' consent poses challenges in population biobank research, and we describe emerging ethical issues surrounding the provision of research results to participants. We will present reasons why traditional requirements of 'informed' consent are difficult to implement in population genomics and argue for the need for an alternate consent process. These issues highlight examples where biobank research casts new considerations on traditional approaches to consent. This presentation aims to discuss strategies that reflect the emerging ethical needs of this fast moving area of research.

Fifth session

Longevity & Migraine

Chairperson
Qihua Tan

GENETIC INFLUENCE ON HUMAN LIFESPAN AND LONGEVITY

J.v.B. Hjelmborg (a), I. Iachine (a), A. Skytthe (a), J.W. Vaupel (a), M. McGue (b), M. Koskenvuo (c), J. Kaprio (c), N.L. Pedersen (d), K. Christensen (a)

(a) Institute of Public Health, University of Southern Denmark, Odense, Denmark

(b) Department of Psychology, University of Minnesota, Minneapolis, MN, USA

(c) Department of Public Health, University of Helsinki, Helsinki, Finland

(d) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

There is an intense search for longevity genes in both animal models and humans. Human family studies have indicated that a modest amount of the overall variation in adult lifespan (approximately 20-30\%) is accounted for by genetic factors. But it is not known if genetic factors become increasingly important for survival at the oldest ages. As part of the GenomEUtwin project we are able to study the genetic influence on human lifespan and how it varies with age using the almost extinct cohorts of Danish, Finnish and Swedish twins born between 1870 and 1910 comprising 20,502 individuals followed until 2003-2004. We first estimate mean lifespan of twins by lifespan of co-twin and then turn to the relative recurrence risk of surviving to a given age. Mean lifespan for male monozygotic (MZ) twins increases 0.39 (95\% CI (0.28,0.50)) years for every year his co-twin survives past age 60 years. This rate is significantly greater than the rate of 0.21 (0.11,0.30) for dizygotic (DZ) males. Females and males have similar rates and these are negligible before age 60 for both MZ and DZ pairs. We moreover find that having a co-twin surviving to old ages substantially and significantly increases the chance of reaching the same old age and this chance is higher for MZ than for DZ twins. The relative recurrence risk of reaching age 92 is 4.8 (2.2,7.5) for MZ males, which is significantly greater than the 1.8 (0.10,3.4) for DZ males. The patterns for females and males are very similar, but with a shift of the female pattern with age that corresponds to the better female survival. Similar results arise when considering only those Nordic twins that survived past 75 years of age. The present large population based study shows genetic influence on human lifespan. While the estimated overall strength of genetic influence is compatible with previous studies, we find that genetic influences on lifespan are minimal prior to age 60 but increase thereafter. These findings provide support for the search for genes affecting longevity in humans, especially at advanced ages. If time allows we will discuss the work in progress on the application of extreme value theory methods to the above data on human lifespan for the purpose of studying advanced age mortality. In particularly, the effect of birthcohort, the age of the force of mortality plateaus and genetic influences on extreme lifespan.

PHYSICAL ACTIVITY IS ASSOCIATED WITH LONGER WHITE CELL TELOMERES AND POTENTIALLY SLOWER BIOLOGICAL AGEING

J.L. Hunkin (a), L.F. Cherkas (a), B.S. Kato (a), J.P. Gardner (b), G.L. Surdulescu (a), M. Kimura (b), A. Aviv (b), T.D. Spector (a)

(a) Twin Research & Genetic Epidemiology Unit, King's College London, St. Thomas' Hospital Campus, London, United Kingdom

(b) The Centre of Human Development and Aging, University of Medicine and Dentistry of New Jersey, New Jersey Medical School, Newark, USA

Background: Physical inactivity is an important risk factor for many aging-related diseases. White Blood Cell (WBC) telomere dynamics (telomere length and age-dependent attrition rate) is ostensibly a biological indicator of human aging. We therefore tested the hypothesis that physical activity level in a population based sample is associated with WBC telomere length.

Methods: We studied 2152 Caucasian female twins with a questionnaire on physical activity level, smoking status, height, weight and socio-economic status (SES). The mean terminal restriction fragment length (TRFL), an index of telomere length, was measured in WBC DNA and adjusted for age and other potential confounders.

Results: TRFL decreased with age at a mean rate of 21 base pairs (bp) per year. WBC telomere length was positively associated with increasing physical activity level and this association remained significant after adjustment for age, body mass index (BMI), smoking and SES. Telomeres of the least active subjects were 217 bp shorter than those of the most active subjects ($p = 0.006$). The trend was confirmed in a small group of twin pairs discordant for physical activity level.

Conclusion: Physical activity impacts on WBC telomere dynamics. Sedentary lifestyle (in addition to age, smoking, BMI and low SES) is associated with shorter telomeres. In telomeric year equivalence, the least active women were on average 10 years older than the most active ones. This underscores the importance of health promotion of regular exercise since sedentary life is a risk factor which may accelerate human aging.

TRAIT COMPONENT ANALYSIS OF FINNISH AND AUSTRALIAN MIGRAINE FAMILIES REVEALS A CLEARER PICTURE OF MIGRAINE LINKED LOCI

V. Anttila (a,b,c), M. Kallela (c,d), D.R. Nyholt (e), N.G. Martin (e), M. Wessman (b,c,f), A. Palotie (a,c,g)
(a) Finnish Genome Center, University of Helsinki, Helsinki, Finland
(b) Biomedicum Helsinki, Research Program in Molecular Medicine, Helsinki, Finland
(c) Center of Excellence in Complex Disease Genetics, Helsinki, Finland
(d) Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland
(e) The Queensland Institute of Medical Research, Brisbane, Australia
(f) Folkhälsan Research Center, University of Helsinki, Helsinki, Finland
(g) The Broad Institute of MIT and Harvard, Cambridge, MA, USA

Background and aims Migraine is the most common cause of severe chronic episodic headache. We performed a new type of analysis using the clinical traits associated with migraine as phenotypes in two genome-wide screens of 50 and 45 large Finnish families and one screen of 152 small Australian families. To tackle the complexity of the migraine syndrome, we set out to study if individual clinical aspects of the disease might be inherited independently. Our aim was to analyze whether any of the IHS diagnostic subcomponents provide evidence of linkage in both materials. The analysis was performed on sets of families genotyped in the Finnish Genome Center.

Methods Our Finnish study samples consisted of 952 genotyped individuals within 95 Finnish independent multigenerational migraine families. The same neurologist diagnosed all Finnish subjects. The Australian study sample consisted of 661 genotyped individuals in 152 mainly nuclear families. All genotypes were analyzed by parametric and non-parametric two-point and multipoint linkage analyses, using the diagnostic subcomponents of migraine as the phenotypes.

Results In the analyses, significant evidence of linkage was found to a number of loci. Significant or suggestive evidence of linkage was found to a locus on 10q22 in all three screens and using a number of traits, including the patients' diagnosis. Strongest linkage was found using traits pulsation (multipoint LOD score 3.25) and pain intensity (3.18) as the analyzed trait components in a combined analysis of the larger Finnish and the Australian scan. Another locus, on 17p13, showed significant evidence of linkage (two-point LOD score 4.52) in the first Finnish scan, and was replicated in the other Finnish scan. In addition, several other loci with significant evidence of linkage were found.

Conclusion These findings show the 10q22 locus, detected in two previous studies (one Finnish and one Australian), having an effect in all three study samples, with significant evidence of linkage with several traits. The high reproducibility of this locus in genome-wide studies suggests an important contribution of this locus to the genetic susceptibility of migraine, and provides a promising target for further studies. In addition, our ability to replicate several loci in this study provides us with a clearer picture of the genetic susceptibility to migraine.

A GENOME-WIDE LINKAGE STUDY OF MIGRAINE IN A DUTCH POPULATION SAMPLE

L. Ligthart (a), G. Willemsen (a), J.J. Hottenga (a), D. Posthuma (a), D.I. Boomsma (a), D.R. Nyholt (b)
(a) Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands
(b) Queensland Institute of Medical Research, Brisbane, Australia

Previous studies utilising latent class analysis (LCA) to analyse IHS migraine symptoms in Australian and Dutch samples identified subgroups of migraine sufferers. LCA identified groups of migrainous headache sufferers that differ in headache severity and number of symptoms. The LCA approach has been shown to result in a larger number of individuals being classified as migraineurs, compared to strict adherence to IHS criteria, thus potentially increasing power in genetic studies of migraine. A large Australian study (Nyholt et al. 2005) recently used a classification based on LCA as the primary phenotype in a genome-wide linkage study of migraine. In addition, Nyholt et al. (2005) performed the first linkage analyses of individual IHS symptoms. Using this approach evidence of significant linkage was found on chromosome 5q21. Highly suggestive linkage on chromosome 10 and two additional suggestive linkages were found on chromosomes 8 and 13. Also, previous findings on chromosomes 6p12.2-p21.1 and 1q21-23 were replicated. Interestingly, the chromosome 1 linkage peak was located within 3 cM from the ATP1A2 gene, which has been implicated in familial hemiplegic migraine. Importantly, evidence for linkage to chromosome 10 has since been replicated in three independent samples.

In the present study, a similar approach was used in a Dutch migraine cohort drawn from a population sample of 12210 Dutch twins and their families (age 14-90, 59% female). The data were collected as part of two questionnaires in 2002 and 2004, which included a detailed series of headache questions based on International Headache Society criteria for migraine. We performed genome-wide linkage analysis using LCA migraine and individual IHS symptoms as the phenotypes. Linkage analyses were performed in a sample of 490 twin families, including 741 unaffected-concordant, 501 discordant and 190 affected concordant sibpairs (equivalent to 504, 224 and 144 independent pairs, respectively). Results from these analyses highlighting regions which replicate previous findings will be presented.

Poster

INFLAMMATORY CANDIDATE GENES AND THE RISK OF CARDIOVASCULAR DISEASE IN FINNISH POPULATION

M. Alanne (a), K. Kristiansson (a), K. Auro (a), L. Peltonen (a), K. Silander (a), V. Salomaa (b), M. Perola (a)

(a) Department of Molecular Medicine, KTL, National Public Health Institute, Helsinki, Finland

(b) Department of Epidemiology and Health Promotion, KTL, National Public Health Institute, Helsinki, Finland

Coronary heart disease (CHD) and atherosclerosis are a major cause of death in Western countries. The most significant risk factor for CHD and atherosclerosis is considered to be high level of low-density lipoprotein (LDL) cholesterol, but more recent evidence has shown that an atherosclerotic plaque development could be best described as an inflammatory process. Our approach in this MORGAM study has been to select candidate genes from an inflammatory related pathway and genes that biologically interact with them. The candidate genes studied were Angiotensin II receptor type 1, Angiotensin I-converting enzyme, C-reactive protein, Fibrinogen alpha and gamma chain, Interleukin 6 and Selenoprotein S. We have studied the effect of variation in these genes in two separate, prospectively followed Finnish cardiovascular disease (CVD) case-cohort samples, FINRISK 92 and 97, (n=2222). Altogether 42 SNP polymorphisms have been genotyped in both cohorts and examples of negative and positive findings will be discussed.

C-reactive protein (CRP, chr 1q21-23) is a major acute-phase reactant and an independent predictor of future CVD events. We selected all haplotype tagging SNPs in the CRP gene with frequency >5% in the SeattleSNP database for genotyping. In time-to-event analysis with Cox's proportional hazards model, two haplotypes showed significant association to CVD events in women. In a multivariate model adjusted for relevant CVD risk factors, female carriers of the haplotype tagged by rs3091244 (G>A>T) A-allele had 2.5 times higher risk for a CVD event than non-carriers (95% confidence interval: 1.0-6.4, p=0.049) in the FINRISK 92 cohort. The result in the FINRISK 97 cohort was not significant. In the combined analysis of the two cohorts, carriers of this haplotype had 2.1 times higher risk for a CVD event than non-carriers (1.2-3.9, p= 0.013). This result was evident also after further adjustment with serum CRP levels (HR: 2.2 (1.2-4.3), p= 0.013).

Selenoprotein S (SEPS1, chr 15q26.3) has been associated with increased levels of inflammatory markers IL-1β and TNF-α. SEPS1 has been suggested as candidate gene for traits related to inflammation such as diabetes mellitus and CHD. To capture the variation in the SEPS1 gene region we selected 6 SNPs in an 11 kb region, based on previous association reports, haplotype block status and position. Time-to-event analysis adjusted for relevant CVD risk factors indicated that SNP rs8025174 was associated with higher risk of coronary heart disease events in females in both FINRISK cohorts: In the FINRISK 92 sample female carriers of the minor allele had 6.3 times higher (1.4-28.8, p=0.018) risk for a CHD event than non-carriers. In the FINRISK 97 the estimated hazard ratio was 3.6 (1.1-

11.9, p=0.026) and in joint analysis 3.0 (1.4-6.4, p=0.006). No association was observed in males. SNP rs7178239 located 3 kb downstream of the gene was associated with incident stroke in both females (HR: 3.5 (1.7-6.8, p=0.001)) and in combined analysis of males and females (HR: 1.8 (1.2-2.6, p= 0.007)) when analysing the two cohorts together.

These results suggest that variation in the SEPS1 gene locus contributes to the risk of CVD in both females and males in the Finnish population. Variation in CRP gene does not seem to associate with CVD risk. Our study approach enables us to replicate any findings and draw conclusions of the roles of these genes in CVD risk, but further studies in other populations are, however, required.

MORGAM: AN INTERNATIONAL POOLING OF CARDIOVASCULAR COHORTS

K. Kuulasmaa (a), K. Asplund (b), F. Cambien (c), M. Ferrario (d), S. Kulathinal (e), M. Perola (a), L. Peltonen (a), V. Salomaa (a), D. Shields (e), H. Tunstall-Pedoe (f), A. Evans (g) for the MORGAM Project

(a) National Public Health Institute, Helsinki, Finland
(b) National Board of Health and Welfare, Stockholm, Sweden
(c) INSERM U525, Paris, France
(d) Università degli studi dell'Insubria, Varese, Italy
(e) Centre for Chronic Disease Control, New Delhi, India
(f) University of Dundee, Dundee, United Kingdom
(g) Queen's University, Belfast, United Kingdom

At the last GenomEUtwin Scientific Meeting in Rome we presented a poster giving a general description of the MORGAM Project. This poster has since been modified and re-formatted as a Flyer(see below), which is now being used to promote MORGAM. The updated poster will be presented.

The background and objectives of the MORGAM Project are:

Coronary heart disease (CHD) and stroke are due to the interaction of many genetic and environmental factors. Individually these interactions have small or, at most, moderate effects but may carry considerable public health significance if the genetic variants in question are common in the population. Reliable detection of these effects requires large sample sizes and abundant statistical power, which can only be achieved in a large collaborative study by the use of high throughput genotyping. MORGAM (MONICA Risk, Genetics, Archiving and Monograph), (http://www.ktl.fi/morgam) is such a collaborative study. It includes cohorts from MONICA and also non-MONICA centres. The Project now forms part of GenomEUtwin (http://www.genomeutwin.org), a network of excellence for genomics in Europe. Centres recruited cohorts and organised follow-up locally through their own funding. MORGAM is pooling these cohorts. The funding is devoted to coordination, pooling of samples and data, quality assessment and control, central preparation of DNA and laboratory analysis.

The main objective of the risk factor component is to assess the similarity of risk coefficients for the classic CVD risk factors in different parts of Europe, between men and women and between age groups, using large cohorts with standardized baseline measurements and well-validated outcomes. The objective of the genetic component is to determine statistically significant combinations of SNPs from the multitude of genotypic data which, in combination with environmental factors and possible intermediate quantitative phenotypes, are predictors of incident CHD and stroke events, and total mortality.

GENETICS OF BONE SIZE AND SKELETAL RATIOS

U. Chinappen-Horsley, I. Fogelman, G. Blake, T.D. Spector
King's College London, United Kingdom

Background: Dual-energy X-ray absorptiometry (DXA) provides information on all bone sizes and ratios. Skeletal ratios are widely used in anthropology and forensic pathology and may be useful for investigating the genetics of height. The Twins UK Registry at St. Thomas' Hospital holds combined bone densitometry and genotypic data for 1,300 dizygotic (DZ) pairs of twins scanned on a Hologic QDR-4500W DXA scanner.

Aim: (i) To determine the reproducibility and validity of a novel measurement technique using total body DXA in a pilot study; (ii) To create a dataset of bone lengths and ratios for 6,500 individuals; (iii) To perform linkage analysis on 2,600 individuals with genotypic data and DXA total body information using bone lengths and skeletal ratios as novel phenotypes.

Method: (i) A novel method, linear pixel count (LPC), was used to measure skeletal sizes on DXA images and compared with real clinical measures from 20 subjects and 20 X-rays of the femur and tibia taken in 2003; (ii) The valid and reproducible LPC method was subsequently used to create a dataset of bone lengths and ratios using Microsoft Excel; (iii) Statistical package, STATA, is being used in the analysis to investigate the clinical relevance of skeletal ratios in exploring the genetics of stature. Genetic linkage analysis will combine the 10CM genome-wide marker information existing on 2,600 individual twin subjects with new phenotypic information on height and skeletal proportions.

Results: (i) The LPC method was reproducible with a coefficient of variation (CV%) value of 1.6%. It was also validated against real clinical measures showing positive correlation; (ii) A dataset has been created currently containing bone lengths and ratios for 4,500 male and female twin subjects. Preliminary results suggest the highest heritability estimates (with 95% confidence intervals) are for height (0.73 [0.55-0.91]) and shoulder width (0.75 [0.68-0.82) and the lowest heritability estimates are for humerus (0.24 [0.01-0.47]) and radius length (0.45 [0.36-0.54]). Other estimates include femur length (0.66 [0.46-0.85]) and pelvic width (0.62 [0.41-0.82]); (iii) Genetic linkage analysis is in progress.

Conclusion: Skeletal ratios can be accurately and precisely measured from DXA total body scan images. The new LPC method will enable genome-wide analyses of our population cohort for stature and help to unravel the genes for stature and bone size.

COL11A1-GENE IS BOTH LINKED TO AND ASSOCIATED WITH HUMAN STATURE

J. Kettunen, E. Costiander, S. Sammalisto, L. Peltonen, M. Perola
National Public Health Institute, Helsinki, Finland

Stature (i.e. adult height) is a quantitative trait with high heritability. Various interesting regions in the human genome have been linked to adult stature but only few have been confirmed by later studies (see table Sammalisto et al. 2005). Recently, we localized a quantitative trait locus for stature in chromosome 1p21 (multipoint LOD score 4.25 in sex-stratified males-only sample). The most promising candidate gene in this region of interest was COL11A1, expressed in growth plates and known to be mutated in human syndromes (MIM 154780, MIM 604841) one of which results in short stature. We genotyped 42 single nucleotide polymorphisms (SNPs) in this gene of 230 kb in 92 families totalling 921 genotyped individuals.

Statistical analyses of SNPs and additional multiallelic markers provided consistent evidence of linkage when we analyzed the pedigree with additional family members. The best LOD score was obtained for males only (multipoint LOD score 4.6 in sex-stratified males-only sample) whereas the females did not contribute significantly to the observed linkage.

We monitored the SNPs and SNP haplotypes for the association in the 54 Finnish families contributing to the linkage signal. The associated allele of COL11A1-gene shows association to tall male stature in the linked families. The best associated SNP is a non-synonymous exonic SNP that changes an amino acid 1535 from proline to serine in the carboxyterminal end of the protein's major triple helix ($p=0.0004$). The association remains significant after conservative Bonferroni correction for multiple testing ($p=0.02$).

The associating SNP was genotyped in stature discordant DZ male twins. The A allele frequency in the unrelated sample of the twins was 0.15 in tallest 10 percent and 0.05 in shortest 10 percent. The male A allele carriers are taller than the G allele homozygotes in both family sample and twin sample. Further analyses in the complete twin sample and in the population cohort are ongoing. Thus, we shown association to a genetic variant for human adult stature, implying that genetic variation in the COL11A1-gene contributes to the variation observed in human stature.

A LONGITUDINAL STUDY OF THE EFFECT OF GSTT1 AND GSTM1 GENE COPY NUMBER ON SURVIVAL

Q. Tan, L. Christiansen, K. Christensen
Epidemiology Unit, Institute of Public Health, University of Southern Denmark, Odense, Denmark

Deletions of the glutathione S-transferase superfamily genes GSTT1 and GSTM1 has been associated with oxidative stress related diseases, and thus the GSTT1 and GSTM1 genes serve as excellent candidates for studies of successful ageing and longevity. Accordingly, a few recent studies have investigated the influence of the presence/absence of GST gene deletions on human longevity, however with conflicting results. This may partially be caused by the traditional use of assays unable to discriminate between carriers of one or two functional genes, which probably reduces the power to observe an effect and furthermore precludes distinction between the various genetic models.

Using a quantitative realtime PCR method facilitating quantification of gene copy number we evaluated the influence of GSTT1 and GSTM1 gene deletions on longevity in a longitudinal study of 681 elderly Danish twins. Mortality risk was estimated using the Cox proportional hazard model, adjusting for age. The mean follow-up time was 8.1 years and during this time a total of 337 deaths occurred.

The results demonstrated a non-significant trend for carriage of two copies of the GSTM1 functional gene to be a protective factor, whereas both heterozygosity and homozygosity for the GSTT1 functional gene was associated with a significant increased mortality in women. To our knowledge, this is the first longitudinal study exploring the influence of GST gene polymorphisms on longevity and these data implies that GST gene copy numbers do affect mortality risk in the elderly.

We are currently looking at possible gene-gene interactions, which will be presented at the meeting along with the survival analysis data.

AN ASSOCIATION STUDY OF AN AGING MARKER, LEUKOCYTE TELOMERE LENGTH, AND TWO AGING RELATED CANDIDATE GENES

Z. Feng (a), K. Ahmadi (a), B.S. Kato (a), J.P. Gardner (b), M. Kimura (b), A. Aviv (b), T.D. Spector (a)

(a) Twin Research and Genetic Epidemiology Unit, St. Thomas' Hospital Campus, King's College, London, United Kingdom

(b) The Centre of Human Development and Aging, University of Medicine and Dentistry of New Jersey, USA

Background: Leukocyte telomere length (LTR), a putative marker of aging and longevity, is a highly variable but heritable complex trait. In order to determine the possible underlying genetic variants for LTR variation, we conducted an association study of LTR and two candidate genes for aging-related traits, TGFB1 and KLOTHO, in a large, female Caucasian dizygotic twin population (UK).

Methods and Materials: Terminal restriction fragment (TRF) length, an index of telomere length, was measured using Southern Blotting. Six and four single nucleotide polymorphisms (SNP) were genotyped in TGFB1 and KLOTHO gene respectively and tested for association in 982 healthy female Caucasian dizygotic twins. When there is strong LD between SNPs ($r2>0.5$), haplotypic association was also investigated.

Results: The average age of study subject is 47 years old (range 18-74 years, SD 11.75). TRF length ranges from 5.15 to 9.36 kb with the average of 7.09 kb (SD 0.68). All SNPs are in Hardy-Weinberg equilibrium ($p>0.05$). No significant association was detected for individual SNPs under either codominant or completely dominant models ($p>0.101$). Two SNPs, TGFB1 -509 C-T and codon-10 T-C, showed the strong pair wise LD with $r2=0.705$. Four statistically inferred haplotypes were present between them, but none of them showed significant association ($p=0.7497$) with telomere length.

Conclusion: We failed to find any significant association between LTR and two aging related candidate genes, TGFB1 and KLOTHO. The result suggests that while we couldn't exclude minor effects, neither of these two candidate genes plays an important role in the variation of LTR in our cohort. But as it is unclear whether telomere length dynamics is the cause or the effect of the aging process, it is still possible the genes are associated with aging via alternate mechanisms.

PSORIASIS IN NORWEGIAN TWINS: A POPULATION BASED STUDY

A.M Grjibovski (a), A.O. Olsen (b), P. Magnus (a), J.R. Harris (a)
(a) Department of Dermatology, Ullevaal University Hospital, Oslo, Norway
(b) Department of Genes and Environment, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway

Background: Psoriasis vulgaris is a chronic inflammatory skin disease with a prevalence ranging between 0.1 and 6.5%. Previous studies have demonstrated a greater disease frequency among parents, siblings and other near relatives of the psoriatic patients suggesting familial factors in the aetiology of the disease. However, these studies did not distinguish genetic from shared environmental effects. Twin studies can separate the influence of genetic factors from the effects of the family environment. However, twin studies are often clinically-based and have an inherent risk of overrepresentation of concordant pairs and thus overestimation of the role of genetic factors. Moreover, earlier twin studies on psoriasis were based on small numbers of pairs warranting larger population-based studies.

Aims: i) To study the conditional and relative risk of developing psoriasis in identical and fraternal twins whose co-twin has a positive history of the disease; ii) to estimate the relative contribution of genetic and environmental factors on the liability for psoriasis in Norway; iii) to study sex-specific patterns of heritability of psoriasis in Norwegian twins.

Methods: The data are from a population-based sample of Norwegian twins who have participated in a study on health and development at the Norwegian Institute of Public Health. The sample includes 3,334 complete pairs and 1,377 single responders. Among the complete pairs there were 526 monozygotic male (MZM) pairs, 397 dizygotic male (DZM) pairs, 777 monozygotic female (MZF) pairs, 655 dizygotic female (DZF) pairs, and 979 dizygotic pairs of opposite sex (DOS). The absolute risk of developing psoriasis through 31 years of age conditioned on the presence of psoriasis in the co-twin was calculated using Kaplan-Meier analysis. The relative risk of psoriasis in twins whose co-twins reported the disease was estimated by Cox regression, where co-twins history of psoriasis was introduced as a single covariate. The relative importance of genetic and environmental factors in liability to psoriasis was estimated by tetrachoric correlations and structural equation modeling using the liability-threshold model.

Results: Altogether, 334 (4.2%) endorsed a positive history of psoriasis. No difference in prevalence of the disease across sexes and zygosity groups were found. MZ twins had 40% risk of developing psoriasis from birth through 31 years of age if a co-twin has developed the diseases within the same time frame while in DZ twins the conditional risk was 15%. The risk of developing the disease in a MZ twin whose co-twin has a positive history of psoriasis was 15.8 (95%: 9.0–27.7) times higher compared to those whose co-twins did not have psoriasis. The corresponding relative risk for DZ pairs was estimated to be 2.8 (95% CI: 1.4–5.6). The tetrachoric intraclass correlations for MZ twins were also greater than for DZ twins both in females and males suggesting genetic contribution to the liability to the disease. Homogeneity of thresholds tests performed in saturated models

revealed no differences in liability to psoriasis across sexes and zygosity groups. The best fitting model (ACE) showed that 66% of the variation in liability for psoriasis in this population could be explained by additive genetic effects and the remaining 34% was due to non-shared environmental influences. No sex-specific patterns of heritability of psoriasis were found.

LACK OF AGGREGATION OF ISCHEMIC STROKE SUBTYPES WITHIN AFFECTED SIBLING PAIRS

P.G. Wiklund (a), W.M. Brown (c), T.G. Brott (d), B. Stegmayr (a), R.D.Jr Brown (e), S. Nilsson-Ardnor (b), J.H. Hardy (f), B.M. Kissela (g), A. Singleton (g), D. Holmberg (b), S.S. Rich (c), J.F. Meschia (d)

(a) Department of Medicine, Umeå University, Sweden
(b) Department of Medical Biosciences, Umeå University, Sweden
(c) Department of Biostatistics, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA
(d) Department of Neurology, Mayo Clinic, Jacksonville, Florida, USA
(e) Department of Neurology, Mayo Clinic, Rochester, Minnesota, USA
(f) National Institute on Aging, Bethesda, Maryland, USA
(g) Department of Neurology, the University of Cincinnati, Cincinnati, Ohio, USA

Background: Epidemiological twin and genome-wide linkage studies support inherited stroke risk. We hypothesized that if ischemic stroke subtypes had unique genetic risk factors, subtypes would aggregate within ischemic stroke–affected sibling pairs.

Methods: This family study was based on a pooled analysis of two cohorts of male and female adult sibling pairs with symptomatic ischemic stroke. One cohort of 404 individuals (first proband seen August 30, 1999) was recruited from the United States and Canada, and another cohort of 198 individuals (first proband seen April 17, 1997) was recruited from Umeå, Sweden. Subtype diagnoses were based on Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria.

Results: Agreement for subtype diagnoses within families was poor (mean±asymptotic standard error $\kappa=0.17\pm0.04$). Comparable levels of agreement were seen when restricting the analysis to same-sex sibling pairs ($\kappa=0.22\pm0.05$), to sibling pairs in which the proband's stroke occurred before the age of 65 years ($\kappa=0.16\pm0.05$), or to pairs in which the proband's stroke occurred at or after the age of 65 years ($\kappa=0.19\pm0.05$).

Conclusions: Many genetic risk factors for ischemic stroke may not be specific for one subtype. This finding has implications for genetic association studies and linkage studies.

POWERFUL, CONSERVATIVE AND ROBUST FAMILY-BASED ASSOCIATION TEST: PSEUDOMARKER

T. Hiekkalinna (a,b), J.D. Terwilliger (b,c,d,e,f)
(a) Department of Molecular Medicine, National Public Health Institute, Helsinki, Finland
(b) Finnish Genome Center, University of Helsinki, Helsinki, Finland
(c) Department of Genetics and Development, Columbia University, New York, USA
(d) Department of Psychiatry, Columbia University, New York, USA
(e) Columbia Genome Center, Columbia University, New York, USA
(f) Division of Medical Genetics, New York State Psychiatric Institute, New York, USA

PSEUDOMARKER is joint Linkage and/or Linkage Disequilibrium software for qualitative traits in the presence of errors and unknown model parameters. PSEUDOMARKER can utilize different pedigree sizes jointly such as cases and controls, trios, sib pairs, sib ships and multi generational families, which leads more robust and powerful tests.

PSEUDOMARKER uses "direct search" in ILINK program from the FASTLINK 4.1P package to maximize likelihoods over allele frequencies under null hypothesis of Linkage (HO), presence of Linkage (H1), haplotype frequencies conditional on disease under null hypothesis of no Linkage and LD (H2) and presence of Linkage and LD (H3). Then multiple statistics can be calculated as log-likelihood ratio tests: Linkage = log10 (H1/H0), LD given Linkage = log10 (H1/H3), LD given no Linkage = log10 (H2/H0), Linkage given LD = log10 (H3/H2) and joint test of Linkage and LD.

PSEUDOMARKER statistical approach provides a consistently more conservative, robust, and powerful way to conduct joint linkage and association analysis on heterogeneous data structures than the other statistical approaches in wider use at the present time, especially when there is linkage but no LD, for which many other methods have enormously high type I error rates, and typically lower power than PSEUDOMARKER over a wide variety of models considered.

GENOTYPE INFORMATION HANDLING AT THE SNP TECHNOLOGY PLATFORM IN UPPSALA

M. Jonsson in collaboration with staff members of the SNP Technology Platform in Uppsala
Molecular Medicine, Department of Medical Sciences, Uppsala University, Sweden

Abstract: In the beginning of the GenomEUtwin project, it became clear that our SNP genotyping lab would face the challenge of producing, inspecting and delivering for us yet unseen amounts of genotype information. To start tackling the problem of information handling, a Microsoft SQL Server™ database was set up in the lab which would store data from our different genotyping instruments and allow for easier access and comparisons of data both within and between different projects. As a starting point, the database design was harmonized with the central genotype database in GenomEUtwin, the only difference being that the central GenomEUtwin database would store genotypes from different labs, whereas our local database would store genotypes from different instruments within the same lab. Since the start, a lot of effort has been put into making the information system which evolved around our database into a system which supports every aspect of the genotyping not covered by the software of each instrument.

Here are a few examples of how the system, as a whole, supports different activities throughout the genotyping process: 1) Information about samples and individuals coming to the lab from outside are now accepted only in a special Microsoft Excel™ file with built-in error checking to prevent mistakes in sample names and microtiter plate layouts. 2) The generation of control files for our pipetting robots can be automated for a number of scenarios such as sample dilution and sample rearrangements on microtiter plates. 3) The system is able to keep track of polarities so that results from opposite DNA strands can be compared.

The core of this information system consists of two programs developed in C# for the Microsoft. NET platform. The first, called Chiasma, is used for input of data such as samples names, plate layouts and results into our storage database. Data from ongoing projects are mirrored from the storage database onto another database which is more suitable for data analysis and where the results can be inspected using the second program called SNP Quality Analysis Tool. With the current setup, quality measures for a million genotypes can be calculated in less than five minutes.

At the end of a genotyping project, the approved genotypes are locked in the database and the results are put together in a report. For larger projects we sometimes arrange special methods for data delivery. This has been done for projects within GenomEUtwin by providing an encrypted “tunnel” between one of our servers and the central genotype database server in Helsinki. When certain results in our storage database are flagged for delivery to the GenomEUtwin database, they are automatically copied to our export server where they can be reached via the tunnel from Helsinki. In a co-operation with the

genotyping lab in Helsinki, we also give out allele frequencies from all our genotype results in the same way.

Our information system, which was founded around the GenomEUtwin genotype database design and driven much by the goal to be able to handle information produced for GenomEUtwin, is now a central part of the lab and used in all our genotyping service projects. From its first humble steps it has grown into a full-fledged production system able to handle a high throughput of projects and millions of genotypes. A more detailed description can be found on our home page at www.genotyping.se.

QUALITY CONTROL OF THE DNA SAMPLES IN THE GENOMEUTWIN PROJECT

M. Jussila (a), O. Törnwall (a), J. Kaprio (b), K. Silander (a), P. Laiho (a), M. Laukkanen (a), L. Peltonen (a), M. Perola (a) on behalf of GenomEUtwin
(a) National Public Helath Institute, Helsinki, Finland
(b) Department of Public Health, University of Helsinki, Helsinki, Finland

Background: The National Public Health Institute of Finland (KTL), Department of Molecular Medicine (MLO) has a highly specialized biobank core unit for centralized DNA-extraction, quality control, storage and sample logistics. The core unit has been operating since 1993 and it has been critical for the large epidemiological DNA sample collections of KTL for genetic analysis (www.nationalbiobanks.fi). The unit has also served as a DNA-extraction and sample logistic center for numerous additional national and many international projects, including GenomEUtwin (www.genomeutwin.org). During the years in operation the DNA core has gained extensive experience by extracting DNA from nearly 200 000 blood samples and currently over 15 000 new DNA samples are extracted and aliquoted for multiple projects annually. Large collaborative projects like GenomEUtwin benefit greatly from a professional centralized DNA core unit which produces samples with good and uniform quality. This reduces substancially sample related problems during genotyping.

Methods of quality control: The basis for the quality control lies on the standardized sample flow throughout the process beginning from receiving the blood samples until to the distribution of the DNA samples in desired concentration and quantity for genotyping. The data on extraction, aliquoting, sample logistics and DNA amount follow-up are governed by KTL databases. As part of the DNA Core Unit's quality standards all samples are barcoded to enable automated sample handling processes and to avoid sample mix-ups.

Three different protocols of DNA-extraction have been utilized for the GenomEutwin samples depending on the time and conditions of the previous storage of blood. The protocols are based on two different chemical purifying methods; organic extraction and salt precipitation. An automated pipetting robot (Tecan Genesis) is used for aliquoting the DNA samples according to the specifications provided by each genotyping laboratory. Accurate concentrations for the aliquoted samples are measured by fluorescence based method, PicoGreen. Sample specific corrections are performed after the measurement to meet the criteria in samples where the DNA amount in aliquot is not sufficient. Prior to distribution to genotyping laboratories all aliquoted samples are tested for PCR functionality and monitored for possible sample mix-ups or contamination. This is achieved either by separating amplified sex chromosome specific PCR fragments on an agarose gel or by producing a genetic fingerprint from the samples with an ABI 3730 DNA analyzer using microsatellite markers.

Results: For GenomEUtwin total of 5442 twin samples and 6725 MORGAM samples have been processed in the DNA Core Unit by October 2006. The quality and quantity of the extracted DNA is constantly reviewed throughout the process. The average yield in extraction of DNA from 1 ml of blood has been 35 µg. The extraction failure rate

calculated from the data available has been very low, less than 1%. The need for sample specific corrections has fluctuated significantly between sample cohorts, the average being 24%. However, after corrective actions less than 1% of samples have been dropped out from the genotyping process. The quality control steps have exposed 19 (0.4%) sample mix-ups in the twin cohorts and 18 (0.3%) mix-ups and 17 (0.3%) contaminated samples in the MORGAM cohorts. In the twin cohorts 0.8% of the samples were dropped out due to insufficient quality or quantity of the samples. In the MORGAM project the samples with low yield (<7.5 µg of DNA, 5% of the samples) are amplified by whole genome amplification (WGA). The quality of the amplified samples has been found to be acceptable for genetic studies (Silander et al., Twin Res Hum Genet. 2005 Aug; 8(4): 368-75).

MORGAM CASE-COHORT STUDY: SUBCOHORT SELECTION PROCEDURE AND APPROACHES TO STATISTICAL ANALYSIS

S. Kulathinal (a), J. Karvanen (b), O. Saarela (b), K. Kuulasmaa (b)
(a) IC Health Scientific Secretariat, Centre for Chronic Disease Control, New Delhi, India
(b) Department of Health Promotion and Chronic Disease Prevention, National Public Health Institute, Helsinki, Finland

The MORGAM Project aims to explore the relationships between the development of cardiovascular diseases and their classic and genetic risk factors. MORGAM cohorts are selected from geographically defined populations and are followed up prospectively for all-cause mortality, non-fatal coronary heart disease and stroke events. We describe the case-cohort design planned and adopted by the MORGAM Project. The motivation for using the case-cohort design in the MORGAM genetic study is discussed and issues relevant to its planning and analysis are studied. We propose a procedure for appending the earlier case-cohort selection after an extension of the follow-up period. Similar procedure can be employed to achieve maximum overlap between possible locally planned designs and the MORGAM case-cohort design. Two approaches for statistical analysis are considered: case-cohort analysis that uses only the cases and the subcohort members and full likelihood approach that uses the whole cohort and treats the genotypes of individuals outside the case-cohort set as a missing data problem.

BAYESIAN MODEL SELECTION AND ESTIMATION OF HAPLOTYPE EFFECTS BASED ON PARTIAL ORDERING OF HAPLOTYPES

E. Arjas (a), S. Kulathinal (b), O. Saarela (c)
(a) Department of Mathematics and Statistics, University of Helsinki and National Public Health Institute, Helsinki, Finland
(b) IC Health Scientific Secretariat, Centre for Chronic Disease Control, New Delhi, India
(c) Department of Health Promotion and Chronic Disease Prevention, National Public Health Institute, Helsinki, Finland

Analysis of associations between event endpoints and highly polymorphic candidate genes is complicated by the small number of subjects having a specific genotype and the large number of genotype effect parameters to be estimated. Existing Bayesian solutions to the problem are based on clustering of genotypes according to risk and thus reducing the number of parameters. We approach the same problem from the viewpoint of Bayesian model selection. Considering biallelic SNPs in a candidate region, we order the alleles of each SNP according to risk. If both alleles carry a similar risk, the SNP is considered neutral and does not contribute to the risk model. Such ordering of alleles imposes a partial ordering for the haplotype risks as well as for the haplotype pair risks. The purpose of the partial ordering is to limit the number of independent haplotype effect parameters in the risk model. Our approach serves the dual purpose of selecting the most important SNPs into the model and estimating the haplotype effects in the presence of large number of relatively rare haplotypes.

THE NTR BIOBANK: COLLECTING MORE THAN JUST DNA IN MORE THAN JUST TWINS

G. Willemsen, E.J.C. de Geus, A. van Bruggen, D. Boomsma
Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands

Inherent to its name GenomEutwin focuses mostly on twins, as a valuable source of genetic and environmental information. However, more and sometimes even new information may be obtained by including family members of twins such as siblings and parents, and spouses of twins. The power of genetic epidemiological and linkage analyses increased by adding the siblings to the twin design. In addition, spousal associations, between the fathers and mothers of twins as well as between the twins and their spouses, may shed a light on the process of assortative mating.

In the ongoing biobank project of the Netherlands Twin Register (NTR), data are collected in twin pairs, and also in their siblings, parents and spouses. We now have collected fasting blood and urine samples for 6155 individuals who come from 2350 families. We visit the participants at home in the morning between 7 and 10 a.m. In women, we collect samples in women on a fixed day in the menstrual cycle. In all individuals, we determined lifestyle factors such as BMI and smoking and physiological variables such as lipid profile (total cholesterol, HDL, LDL), glucose metabolism (glucose, insulin) and immune parameters (CRP, blood cell count).

This NTR biobank is unique in its protocol of blood and urine collection, which allows for future genomics, expression, proteomics, and metabolomics studies. Blood is not only stored for DNA isolation, but additional blood samples are treated according to strict protocols to allow for RNA expression studies and the generation of immortalized cell lines. RNA is collected both from LPS challenged and from unchallenged samples. Thereby, this project has become an invaluable source of information for future enterprises.

HUNTING MIGRAINE GENES: AU & EU WORKING TOWARDS A BREAKTHROUGH

D.R. Nyholt
Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Australia

Typical migraine is a frequent, debilitating and painful disorder that normally affects people during their most productive years (25% of females and 7.5% of males). The World Health Organisation recently identified migraine among the world's top 20 leading causes of disability, with an impact that extends far beyond the suffering individual, to the family and community.

Although migraine has a large impact and is highly prevalent in our society, its aetiology remains relatively obscure and there are no laboratory based diagnostic tests that identify those who suffer from the disorder.

However, twin studies indicate that migraine has a significant genetic component, with heritability estimates of 33-65%. Therefore, in an effort to identify the molecular mechanisms underlying the disorder, we have been looking for genomic regions co-inherited (linked) with migraine.

That is, because so little is known about the underlying causes of migraine, a positional cloning approach may be the only feasible way to identify the molecular mechanisms underlying the disorder.

Multiple genome-wide linkage screens within different ancestral populations have reported suggestive and significant linkage to numerous novel chromosomal regions, with minimal overlap between studies.

In contrast, recent separate and combined analyses of AU and EU migraine families have provided consistent evidence for linkage to chromosome 10q and suggest additional genomic regions of interest. A summary of our progress to date in identifying genomic regions likely to harbour genes contributing to migraine susceptibility will be presented, together with suggestions for analytical approaches and future directions.

AUTHORS' INDEX

Ahmadi, K.; 45
Alanne, M.; 3; 4; 8; 21; 39
Anttila, V.; 35
Arjas, E.; 55
Asplund, K.; 41
Auro, K.; 3; 4; 8; 21; 39
Aviv, A.; 34; 45
Blake, G.; 42
Boman, M.; 6
Boomsma, D.I.; 17; 23; 36; 56
Brott, T.G.; 48
Brown, R.D.Jr; 48
Brown, W.M.; 48
Cambien, F.; 41
Cherkas, L.F.; 34
Chinappen-Horsley, U.; 42
Christensen, K.; 14; 24; 33; 44
Christiansen, L.; 44
Costiander, E.; 43
Dahlgren, A.; 13
de Geus, E.J.C.; 17; 56
Evans, A.; 41
Fagnani, C.; 14
Falchi, M.; 23
Feng, Z.; 45
Ferrario, M.; 41
Fogelman, I.; 42
Gardner, J.P.; 34; 45
Giannantonio, L.; 29
Grjibovsk; 46
Hardy, J.H.; 48
Harris, J.R.; 30; 46
Hasan, H.S.; 30
Hiekkalinna, T.; 49
Hjelmborg, J.v.B.; 14; 33
Holmberg, D.; 48
Hottenga, J.J.; 36
Hunkin, J.L.; 34
Iachina, M.; 24
Iachine, I.; 33
Joergensen, B.; 24
Jonsson, M.; 50
Jussila, M.; 52
Kallela, M.; 35
Kaprio, J.; 13; 14; 15; 33; 52
Karvanen, J.; 54
Kato, B.S.; 34; 45
Kettunen, J.; 43
Kimura, M.; 34; 45
Kissela, B.M.; 48
Korkeila, M.; 14
Koskenvuo, M.; 33
Kristiansson, K.; 3; 4; 8; 21; 39
Kulathinal, S.; 41; 54; 55
Kuo, P.H.; 25
Kuulasmaa, K.; 8; 41; 54
Kyvik, K.O.; 23
Laiho, P.; 52
Laukkanen, M.; 52
Ligthart, L.; 36
Liljedahl, U.; 13
Magnus, P.; 46
Magnusson, P.K.E.; 6
Mannila, H.; 21
Martin, N.G.; 13; 23; 35
McGue, M.; 14; 33
Medland, S.E.; 25
Meschia, J.F.; 48
Montgomery, G.; 23
Naukkarinen, J.; 15
Neale, M.C.; 25
Nilsson-Ardnor, S.; 48
Nisticò, L.; 29
Nyholt, D.R.; 35; 36; 57
Olsen, A.O.; 46
Oresic, M.; 15
Palotie, A.; 35
Pedersen, N.L.; 6; 33
Peltonen, L.; 3; 4; 8; 13; 15; 21; 39; 41; 43; 52
Perola, M.; 3; 4; 8; 13; 21; 39; 41; 43; 52
Pietiläinen, K.H.; 15
Posthuma, D.; 17; 23; 36
Rich, S.S.; 48
Rissanen, A.; 14; 15
Saarela, O.; 54; 55
Salomaa, V.; 3; 4; 8; 21; 39; 41
Sammalisto, S.; 43
Schmitt, J.E.; 25
Schousboe, K.; 23
Shields, D.; 41
Silander, K.; 3; 4; 8; 21; 39; 52
Silventoinen, K.; 14
Singleton, A.; 48
Skytthe, A.; 33
Sørensen, T.I.A.; 23
Spector, T.D.; 13; 23; 34; 42; 45

Stegmayr, B.; 48
Surdulescu, G.L.; 34
Syvänen, A.C.; 13
Tan, Q.; 44
Terwilliger, J.D.; 49
Toccaceli, V.; 29
Törnwall, O.; 52
Tunstall-Pedoe, H.; 41
van Bruggen, A.; 17; 56
Vaupel, J.W.; 33
Virtamo, J.; 8
Webb, B.T.; 25
Wessman, J.; 3; 21
Wessman, M.; 35
Whitfield, J.; 23
Wiklund, P.G.; 48
Willemsen, G.; 17; 23; 36; 56
Yki-Järvinen, H.; 15

Stampato da Tipografia Facciotti srl
Vicolo Pian Due Torri 74, 00146 Roma

Roma, settembre 2006 (n.18) 3° Suppl.