Abstracts for the Sixth Australasian Gene Mapping Meeting
Brisbane, August 29–31, 2007

Publication of these abstracts is sponsored by Sequenom Inc.

IDENTIFYING GENETIC MODIFIERS OF BREAST CANCER RISK IN *BRCA1* AND *BRCA2* MUTATION CARRIERS IN THE CONSORTIUM OF INVESTIGATORS OF MODIFIERS OF *BRCA1* AND *BRCA2* (CIMBA)

Antonis C. Antoniou,[1] Georgia Chenevix-Trench,[2] and Douglas F. Easton,[1] on behalf of CIMBA

[1] *Cancer Research — UK, Genetic Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, United Kingdom*

[2] *Queensland Institute of Medical Research, Brisbane, Australia*

Germline mutations in *BRCA1* and *BRCA2* confer high risks of breast and ovarian cancer. However, epidemiological evidence suggests that these risks are modified by other genetic or environmental factors. Studies of genetic modifiers of *BRCA1/2* have been hampered by small sample size, such that the power to detect even moderate effects on cancer risk has been limited. To address this problem, CIMBA was established to conduct collaborative analyses of genetic polymorphisms, involving many thousands of samples, as modifiers of risk in BRCA1/2 mutation carriers. Here, we evaluate the association of four SNPs with breast cancer risk in *BRCA1/2* mutation carriers. Such analyses are complicated by the fact that mutation carriers are not randomly sampled with respect to their disease phenotype and standard methods of analysis, like Cox regression, do not give valid estimates of the hazard ratios (HR). To correct for this potential bias, we analyzed the data by modeling the retrospective likelihood of the observed SNP genotypes and disease phenotypes conditional on the disease phenotypes. A SNP in the 5'UTR of *RAD51, 135*G > C, has been suggested as a possible modifier of breast cancer risk in *BRCA1/2* carriers. We pooled genotype data on 8512 mutation carriers from 19 studies for this SNP. We found evidence for an increased breast cancer risk in CC homozygotes (HR = 1.92, 95% CI: 1.25–2.94) but not in heterozygotes (HR = 0.95, 95% CI: 0.83–1.07, 2*df* heterogeneity test: $p = .002$). When *BRCA1* and *BRCA2* mutation carriers were analyzed separately, the increased risk was significant only among *BRCA2* mutation carriers in whom we observed HRs of 1.17 (95% CI: 0.91–1.51) among heterozygotes and 3.18 (95% CI: 1.39–7.27) among rare homozygotes (2*df*, $p = .0007$). SNPs in *FGFR2*, *TNRC9* and *MAP3K1* have been recently shown to be associated with breast cancer risk in the general population through a genome-wide association study of breast cancer. To evaluate whether these polymorphisms also modify the breast cancer risk in *BRCA1/2* mutation carriers we genotyped approximately 11000 mutation carriers from 24 studies for these SNPs. The data are currently being analyzed and the results will be discussed. The identification of genetic modifiers of risk may have important implications for the clinical management of *BRCA1/2* mutation carriers.

WASTE OF GENOMIC SPACE OR NOT: FUNCTIONAL ANALYSIS AND EVOLUTION OF A NOVEL 72BP VNTR

D. N. Azmanov,[1] L. Wilson,[2] S. Easteal,[3] and J.A. Cavanaugh[1]

[1] *Medical Genetics Research Unit, ANU Medical School, The Australian National University, Canberra, Australia*

[2] *Infectious Diseases Research, The Canberra Hospital, Canberra, Australia*

[3] *Predictive Medicine Group, JCSMR, College of Medicine and Health Sciences, The Australian National University, Canberra, Australia*

Polymorphic minisatellites (VNTRs) have been used successfully in linkage studies and forensic applications (Armour & Jeffreys, 1992). Investigations on specific VNTRs have shown that they may not be functionally inert sequences, exerting transcriptional and/or translational regulatory functions in conversions of their variable structural information into different expression profiles of neighboring genes (Nakamura et al., 1998). Identification of a novel 72bp VNTR in one of the potential *cis*-regulatory regions in *SLC6A19* (the 'Hartnup gene'), prompted the investigation of the functional role of this VNTR, located 1.6 Kb 3' downstream of the stop codon of *SLC6A19* and 1.7 Kb 5' upstream of the initiation codon of *SLC6A18* (5p15.33). Both genes are differentially expressed in kidney and intestine, with overlapping function as apical membrane broad spectrum neutral amino acid transporters (Romeo et al., 2006). The in vivo expression of the 72bp VNTR sequence was tested in human total kidney cDNA. The in vitro effect of the VNTR expression in primary transcripts was investigated using an EGFP reporter gene transfected into a human embryonic kidney cell line. Depending on the position within the reporter system, either an increase (3'UTR location) or decrease (5'UTR location) of the expression of EGFP was observed. This influence was repeat number mediated with the greatest impact on expression observed with the highest repeat number. Since the 72bp VNTR had regulatory function in vitro, its variability was investigated in a Caucasian population sample and compared to the architecture of the locus in a number of primates and other mammals. Hypervariability, with up to 5 repeat-number alleles and 9 motifs (derived from the internal structure of the repeat) was observed in humans only. The origin of the minisatellite was found to predate the divergence of the boreoeutherian mammals, more than 95 million years ago (Bininda-Emonds et al., 2007). The evolutionary evidence for increased variability in the 72bp VNTR in humans and the functional consequences of the longer repeats (presented only in humans) suggest recent minisatellite expansion and possible adaptive evolution perhaps as a consequence of the impact on expression of the neighboring nutrient transporters SLC6A19 and SLC6A18.

GENOME-WIDE ASSOCIATION SCAN REVEALS 6 NOVEL, REPLICATED RISK LOCI FOR CROHN'S DISEASE

Jeffrey C. Barrett,[1] for the UK IBD Consortium and Wellcome Trust Case Control Consortium

[1] *Wellcome Trust Centre for Human Genetics, Oxford University, United Kingdom*

The Wellcome Trust Case Control Consortium is a UK based genome-wide association study consisting of three parts: (1) 7 complex diseases with 2000 cases each compared against 3000 shared controls, all from the UK and genotyped on 500,000 SNPs, (2) a set of 1500 tuberculosis cases and 1500 controls, all from The Gambia, also on 500,000 markers and (3) 4 complex disease samples (UK) genotyped on 15,000 nonsynonymous SNPs. While numerous genomic regions showed strong statistical association to one or more of these traits, replication of the effects in an independent sample is an important step in confirming these results. We have analyzed initial data from the WTCCC scan for Crohn's disease (a form of inflammatory bowel disease; part of (1) above) and follow-up genotyping in 1182 new cases and 2024 new controls. We now present 6 novel, confirmed risk factors for Crohn's disease, examine the nature of the associated regions, and discuss implications for other genome-wide association studies.

GENOME-WIDE LINKAGE AND ASSOCIATION ANALYSES OF BODY HEIGHT FROM A LARGE SAMPLE OF AUSTRALIAN TWIN FAMILIES

Beben Benyamin,[1,2] Markus Perola,[3] Belinda K. Cornes,[1] Pamela A. F. Madden,[4] Aarno Palotie,[3] Grant W. Montgomery,[1] Leena Peltonen,[3] Nicholas G. Martin,[1] and Peter M. Visscher[1]

[1] *Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia*

[2] *Institute of Evolutionary Biology, University of Edinburgh, Ashworth Laboratories, King's Buildings, United Kingdom*

[3] *National Public Health Institute, Helsinki, Finland*

[4] *Washington University School of Medicine, Department of Psychiatry, St Louis, Missouri, United States of America*

Body height is one of the most important anthropometric measures. It has been suggested as an indicator of childhood living conditions and has been associated with the risk of coronary heart diseases, intelligence, educational attainment, longevity and social mobility. Since height is a highly heritable and normally distributed trait, it has also been considered as an excellent model trait for studying the genetic architecture (the number and position of genes affecting the trait, the magnitude of their effects, and type of gene action) of complex quantitative traits/diseases. To find genes responsible for height difference between individuals, we conducted a genome-wide linkage analysis in a large sample of Australian twin families (8447 individuals in 2861 families, providing a total of 5815 quasi-independent sibling pairs) and a genome-wide association analysis from 800 individuals typed with 100k SNPs. From linkage

analysis, we found several chromosomal regions, which are responsible for individual difference in height, including on chromosomes 3 and 5 (for all individuals); chromosome 1 and 15 (for males only) and chromosome 7 (for females only). We also found that the inclusion of individuals who are too tall or too short compared to their brothers/sisters (within-family outliers) may create false positives or negatives. Therefore, we suggest that the investigation of the effect of within-family outliers, which is usually neglected, should be a standard quality control measure in linkage analysis for complex traits. This measure may reduce the noise for the search of genes of modest effect size as well as help identify rare genes of large effect and of clinical significance (e.g., idiopathic short stature). The results of genome-wide association analysis will be discussed.

ASSOCIATION ANALYSIS OF GENES FROM THE REELIN NETWORK IN SCHIZOPHRENIA

Sonja Bouwer,[1] Kirsten Peters,[1] Steven Wiltshire,[1] David Chandler,[2] Milan Dragovic,[2] Johanna Badcock,[2] Sarah Howell,[2] Assen Jablensky,[2] and Luba Kalaydjieva[1]

[1] *Western Australian Institute for Medical Research, Australia*

[2] *Centre for Clinical Research in Neuropsychiatry, The University of Western Australia, Perth, Australia*

Schizophrenia is a complex disorder with phenotypic and pathogenetic heterogeneity that poses a major challenge to genetic research. A genetically distinct schizophrenia subtype, termed the cognitively deficient subtype (CD), was identified in the Western Australian Family Study of Schizophrenia. The memory and cognitive performance of these patients is suggestive of impaired synaptic function, and neuroimaging data show reduced hippocampal volume. Maternal interviews reveal increased incidence of pregnancy complications, and learning difficulties, extreme shyness and social withdrawal observed in the children, all pointing to the CD subtype as a neurodevelopmental disorder. The concept of schizophrenia as a neurodevelopmental disorder with impaired synaptic development/plasticity is supported by consistent neuropathological findings of mild abnormalities in cortical organization, underdeveloped dendrites and reduced Reelin expression. In this study, we hypothesise that impaired corticogenesis and synaptic development/plasticity are major pathogenetic mechanisms in the CD subtype of schizophrenia. This genetic association study aims to identify single nucleotide polymorphisms (SNPs) within genes in a functional network related to neuronal migration and synaptic development/plasticity that contribute to susceptibility to the CD subtype of schizophrenia. The network of genes is centered around the extracellular glycoprotein, Reelin as it plays a key role during evolution of the cerebral cortex, is consistently down-regulated in postmortem brain tissue from schizophrenia patients and is involved in neuronal migration during corticogenesis and synaptic plasticity in adulthood. Polymorphisms within these genes will be analyzed individually, as predictors of the CD subtype, as well as with more complex multi-SNP interaction and pathway-based regression approaches. In this presentation, the systems biology approach of selecting candidate genes within the Reelin network that encode physically and functionally interacting proteins as well as the SNP selection process for comprehensive coverage of these genes will be discussed. The first genotyping results will also be presented.

A NOVEL ASSOCIATION BETWEEN A *TNFα* HAPLOTYPE AND SERUM FERRITIN IN SYMPTOMATIC HEREDITARY HAEMOCHROMATOSIS

Jennifer M. Brew,[1,2] Helen Rodgers,[2] Dimitar N. Azmanov,[2] Erin J. Walker,[3] Susan R. Wilson,[4] and Juleen A. Cavanaugh[2]

[1] *John Curtin School of Medical Research, Australia*

[2] *Medical Genetics Research Unit, ANU Medical School Canberra, Australia*

[3] *Analytical Genetics Technology Centre, Toronto, Canada*

[4] *Centre for Bioinformation Science, Mathematical Sciences Institute, Australian National University, Canberra, Australia*

Hereditary hemochromatosis (HH) is a late onset inherited disorder of excess iron storage that may lead to severe organ damage from cirrhosis or cancer. In 1996 the HFE gene was identified with two causative mutations, the C282Y mutation and the H63D mutation, described in the development of ca. 85% of HH cases (Feder et al., 1996). The frequency of this disease is high in Caucasians, with estimates of up to 100,000 C282Y homozygotes in the Australian population (Olynjk et al., 1999); however, penetrance is incomplete. Genetic factors in addition to HFE that affect iron homeostasis have been estimated to account for ca. 20–45% of the variation in iron stores in HH (Njajou et al., 2006; Whitfield et al., 2000) but no modifier genes have as yet been identified. The aim of this study was to investigate the impact in Australian HH of a reported association between the –308A promoter polymorphism in the pro-inflammatory cytokine TNFα and serum ferritin (SF) levels (Krayenbuehl et al., 2006). We studied 90 male C282Y homozygotes and 100 C282Y homozygote controls and found no association between that TNFα polymorphism and SF levels in our sample. Three additional TNFα promoter polymorphisms (–1031, –863, –857) and haplotypes across the SNPs were also investigated. No deviation from frequencies predicted under HWE was observed and no single polymorphism was found to significantly affect body iron loading in any group. An association was observed between 1 2-locus haplotype (that does not include the –308 polymorphism) and serum ferritin levels in symptomatic but not nonsymptomatic patients or controls. This association was not affected by age, phlebotomy or blood donation. Since we found that symptomatic C282Y homozygotes had significantly higher mean SF than asymptomatic homozygotes ($p < .05$) we hypothesize that symptoms of HH only develop above a threshold level of body iron stores which might be dictated by the novel haplotype described here.

A GENOMEWIDE NONSYNONYMOUS SNP SCAN FOR SUSCEPTIBILITY TO ANKYLOSING SPONDYLITIS — THE WELLCOME TRUST CASE-CONTROL CONSORTIUM

Matthew A. Brown,[1,2] Linda A. Bradbury,[1] Jennifer J. Pointon,[2] B. Paul Wordsworth,[2] and the Wellcome Trust Case-Control Consortium

[1] *Diamantina Institute of Cancer, Immunology and Metabolic Medicine, University of Queensland, Australia*

[2] *Botnar Research Centre, University of Oxford Institute of Musculoskeletal Sciences, Nuffield Orthopaedic Centre, Oxford, United Kingdom*

Susceptibility to ankylosing spondylitis (AS) is primarily genetically determined, with heritability of > 90%. Although HLA-B27 is known to be the major susceptibility allele for the condition, there is strong evidence that other genes are involved. The aim of this study was to identify other AS-associated genes. 922 British Caucasian AS cases, and 1468 ethnically matched controls were genotyped for 14,436 nsSNPs, and 897 SNPs throughout the major histocompatibility complex (MHC), using Illumina microarray SNP genotyping methods. Association was tested using the Cochrane-Armitage test of trend. Extremely strong association was observed with the MHC, with $p < 10^{-50}$ across > 1.5 Mb. This may be due either to extreme linkage disequilibrium with HLA-B27, or the presence of more than one MHC susceptibility gene operating in this disease. Two SNPs lying in the gene ARTS1 on chromosome 5 were strongly associated with AS (rs27044: $\chi^2 = 23.90$, $p = 1.0 \times 10^{-6}$; rs30187: $\chi^2 = 21.82$, $p = 3.0 \times 10^{-6}$). ARTS-1 (ERAAP, ERAP1) encodes a type II integral transmembrane aminopeptidase with diverse immunological functions. Within the endoplasmic reticulum, ARTS-1 is involved in trimming peptides to the optimal length for MHC Class I presentation. At the cell surface, it is involved in trimming cytokine receptors including receptors for IL-1, IL-6 and TNF. A second SNP in the Calsyntenin-3 gene also reached genome-wide significance (rs7302230, $\chi^2 = 25.15$, $p = 5.3 \times 10^{-7}$). No other nsSNPs met genome-wide-significance levels. A nsSNP in IL23R, rs11209026, which has also been associated with both Crohn's disease and psoriasis in other datasets, was also associated with AS ($p = .001$), even in cases not affected with inflammatory bowel disease (IBD) or psoriasis. This gene may thus at least partially explain the co-occurrence of AS with IBD and psoriasis. The findings of this study require replication before they can be considered definitive. Nonetheless, they indicate the likely involvement of more than one MHC gene in AS, as well as pointing to likely non-MHC genes involved in AS-susceptibility. While we have only characterized a small proportion of the overall diversity of the genome, this study demonstrates the great potential for large hypothesis-free association studies to identify genetic variants involved in common diseases.

RAPID AND ACCURATE HAPLOTYPE PHASING AND MISSING DATA INFERENCE FOR WHOLE GENOME ASSOCIATION STUDIES USING LOCALIZED HAPLOTYPE CLUSTERING

Brian L. Browning[1,2] and Sharon R. Browning[1]

[1] *Department of Statistics, The University of Auckland, Auckland, New Zealand*

[2] *Discipline of Nutrition, The University of Auckland, Auckland, New Zealand*

Whole genome association studies present many new statistical and computational challenges due to the large quantity of data obtained. One of these challenges is haplotype inference: methods designed for small data sets from candidate gene studies do not scale well to the large number of individuals genotyped in whole genome association studies. We present a new method and software for inference of haplotype phase and missing data that can accurately phase data from whole genome association studies. Our method is based on fitting a localized haplotype cluster model (Browning, 2006; Browning & Browning, in press) to initial estimates of haplotype phase. The localized haplotype cluster model is extended to give a hidden Markov model, from which revised estimates of haplotype phase can be sampled. This process is iterated, with most likely haplotype phase inferred at the last iteration. Our method is compared with existing haplotype inference methods, including fastPHASE and HaploRec, on real and simulated data sets with thousands of genotyped individuals. We find that our method outperforms existing methods in both speed and accuracy for large data sets with thousands of individuals and densely spaced genetic markers, and we use our method to phase a real data set of 3002 individuals genotyped for 490,032 markers in 3.1 days computing time, with 99% of masked alleles imputed correctly. Our method is implemented in the Beagle software package (http://www.stat.auckland.ac.nz/~browning/beagle/beagle.html).

DIGENIC INHERITANCE OF APPARENT AUTOSOMAL DOMINANT KERATOCONUS IN A LARGE AUSTRALIAN PEDIGREE

Kathryn P. Burdon,[1] Douglas J. Coster,[1] Jac C. Charlesworth,[2] and Jamie E. Craig[1]

[1] *Department of Ophthalmology, Flinders University, Adelaide, South Australia, Australia*

[2] *Southwest Foundation for Biomedical Research, San Antonio, Texas, United States of America*

Keratoconus (KC) is a debilitating, blinding disease characterized by progressive asymmetrical thinning of the cornea. This incidence is around 1 in 2000, with onset typically in early adulthood. KC is the main indication for corneal graft in the Western world, accounting for just over 30%. A significant proportion of grafts fail, leaving the patients blind. Multiple linked loci have been reported, through sib-pair and extended family studies, although none have been replicated. Only 1 gene (VSX1) has been identified but other studies have not always supported this finding. A large 4-generation pedigree with autosomal dominant keratoconus was identified by the treating physician undertaking corneal grafting in multiple family members. A genome-wide linkage scan was conducted using the Affymetrix 10K SNP array. Multipoint analysis under a fully penetrant dominant model in MERLIN did not reveal any linkage. With the penetrance in heterozygotes reduced to 70%, two peaks were identified on 1p36.23-36.21 and 8q13.1-q21.11, with LOD scores of 1.9 and 2.0 respectively. Haplotype analysis revealed that all affected individuals carried identical haplotypes at both loci, while unaffecteds carried only one or neither of the haploytpes. Digenic linkage analysis was undertaken in GENEHUNTER-TWO LOCUS. The maximum LOD score of 3.40 was observed at 19.1cM on chromosome 1 and from 84.4 to 93.5cM on chromosome 8. Non-parametric analysis revealed a maximum NPL score of 7.8 with a corresponding *p*-value of 0.00024 at the same location. Thus, although the pedigree appears to demonstrate simple autosomal dominant inheritance, the disorder is probably digenic. Favor is shifting away from family studies in the age of full genome association for complex diseases. Here we present a large family demonstrating digenic linkage for keratoconus, providing an intermediate step between a single gene Mendelian disorder and a genetically complex disorder. This illustrates the continuing value of extended family studies. Recent technological advances allow genome-wide linkage scans to be conducted using dense SNP maps, eliminating the need for fine mapping and reducing both time and costs. This makes whole genome linkage scans in extended pedigrees a cheap and fast option for identifying genomic susceptibilty regions.

EPITHELIAL SODIUM CHANNEL γ-SUBUNIT POLYMORPHISMS ARE ASSOCIATED WITH BLOOD PRESSURE IN THE VICTORIAN FAMILY HEART STUDY AND IN UTAH PEDIGREES

Cara J. Büsst,[1] Katrina J. Scurrah,[1] Justine A. Ellis,[1] Yuanpei Xin,[2] Eliot A. Brinton,[2] Paul N. Hopkins,[2] Steven C. Hunt,[2] and Stephen B. Harrap[1]

[1] *Department of Physiology, The University of Melbourne, Victoria, Australia*

[2] *Cardiovascular Genetics Department, University of Utah, Salt Lake City, Utah, United States of America*

The epithelial sodium channel is the rate limiting step of sodium reabsorption in the kidney. The γ-subunit of the epithelial sodium channel (γ-ENaC), encoded by the *SCNN1G* gene, has been implicated in the Mendelian monogenic diseases Liddle's syndrome and pseudohypoaldosteronism type 1. *SCNN1G* is located in a region of chromosome 16 linked to systolic blood pressure (SBP) by a number of independent studies, including the Victorian Family Heart Study (VFHS). We performed a preliminary association analysis of SBP with 25 *SCNN1G* SNPs in an unrelated subset of VFHS subjects selected from the upper and lower 10th percentiles of SBP. Using logistic regression adjusting for age, sex and BMI, we detected association of SBP to 4 SNPs (*p* values between .009 and .01), 3 of which were adjacent and formed a haplotype with a greater global significance ($p = .0001$ from HaploStats). To replicate these findings, we typed 6 of the SNPs previously typed in the VFHS on 1971 relatives in 68 large Utah pedigrees who had available BP and DNA. Of these, 675 have returned so far for a 25-year follow-up exam. FBAT was used to test for association of SNPs and haplotypes while controlling for related observations in families. After adjusting for age, sex and body mass index, significant associations were detected for SNP rs13331086 with DBP at 25-year follow up ($p = .002$) and for change in DBP from baseline to 25-year follow up ($p = .003$). Haplotypes of rs4299163 and rs5740 also revealed association with change in DBP from baseline to 25-year follow up ($p = .013$). Preliminary results for analysis of DBP in the VFHS are consistent with DBP findings in the Utah pedigrees. In conclusion, *SCNN1G* is significantly associated with BP in a subset of the VFHS and in Utah pedigrees at 25-year follow-up.

INVESTIGATION OF MATERNAL POPULATION SUBSTRUCTURE AND ASSORTATIVE MATING IN AN AUSTRALIAN POPULATION SAMPLE

Enda P. Byrne, Allan F. McRae, Zhen Zhen Zhao, Nicholas G. Martin, Grant W. Montgomery, and Peter M. Visscher

Genetic Epidemiology, Queensland Institute for Medical Research, Brisbane, Australia

Population substructure refers to situations where populations that are considered to be homogeneous with respect to allele frequencies actually consist of two or more subpopulations each with their own characteristic set of allele frequencies. An understanding of population substructure is important in the design of case-control association studies, as if unaccounted for it can lead to spurious associations. One way that population structure is manifested is through assortative mating, where mates will tend to choose mates that are of the same geographic and ancestral background or are phenotypically similar to themselves. Due to Australia's relatively recent history of immigration from various regions of Europe, the Australian population is a candidate to contain hidden substructure. The strict maternal inheritance pattern and high mutation rate of mitochondrial DNA make it an ideal tool for population genetics studies. By aligning a large number of European mitochondrial sequences, we selected 69 SNPs that tagged all variants with a frequency > 1% with an r_2 of .8. 3839 Australian individuals from 1037 families were genotyped, yielding 1693 independent mitochondrial haplotypes. We also obtained grandmaternal and great-grandmaternal ancestry data for 1547 of these unique individuals from a previous study. Individuals were grouped according to their maternal ancestry. No significant differences were found in allele or haplogroup frequencies between European groups (UK/Ireland, Northern, Southern, Eastern). We performed a number of tests for maternal population substructure in our sample, including Bayesian (STRUCTURE program) and latent class analysis methods (L-pop). We also tested for assortative mating by comparing the haplotypes of mates to see if they are more similar than random pairs from the population, and to see if mates tend to be of the same haplogroup. We found no significant evidence for maternal population stratification in our sample. The results from mitochondrial association studies on a range of phenotypes will be discussed.

TRANSCRIPTOMIC EPIDEMIOLOGY OF SMOKING

Jac C. Charlesworth,[1] Joanne E. Curran,[1] Matthew P. Johnson,[1] Harald H.H. Göring,[1] Thomas D. Dyer,[1] Vincent Diego,[1] Eric K. Moses,[1] and John Blangero[1,2]

[1] *Southwest Foundation for Biomedical Research, San Antonio, Texas, United States of America*

[2] *Chemgenex Pharmaceuticals, Geelong, Victoria, Australia*

This study used large-scale genome-wide expression profiling to determine the systemic influence of cigarette smoke, as an environmental exposure, on human health. Our hypothesis was that, given a sufficiently large set of individuals, a stable and interpretable pattern of gene expression alterations attributable to cigarette smoke exposure may be obtained. In addition, a significantly large dataset allows for both the investigation of significant results at the individual gene level as well as providing the ability to determine patterns or signatures of gene expression alterations. Studying these patterns of alteration in response to cigarette smoke exposure may provide the key to understanding the pathogenesis of many of the adverse health effects attributable to smoking and the interactions between them. We obtained genome-wide quantitative transcriptional profiles from 1240 individuals from the San Antonio Family Heart Study. The prevalence of smoking in the dataset was 23.6%, with 331 current smokers. Using lymphocyte samples, we identified 20,413 transcripts with significantly detectable expression levels, including both known and predicted genes. With a highly conservative false-discovery rate of 5% we identified 342 transcripts (330 unique genes) whose expression levels were significantly correlated with a discrete measure of smoking behavior, 110 (32.2%) with positive correlations and 232 (67.8%) negatively correlated. The smoking correlated genes had a range of significant functional annotations (including cell death, viral elimination, cancer and immune response) that correspond well with known smoking-related pathologies; however a wide-ranging, negative influence on the immune system was by far the clearest picture to emerge from our transcriptional profiling. Our results indicate that not only individual genes but entire networks of gene interaction are influenced by cigarette smoking. This is the largest in vivo transcriptomic epidemiological study of smoking to date and reveals the significant and comprehensive influence of cigarette smoke, as an environmental variable, on the expression of genes within the lymphocyte transcriptome. The results of this investigation offer insights into the system-wide pathological processes induced by exposure to cigarette smoke by determining its influences at the gene level.

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

LINKAGE AND HERITABILITY ANALYSIS OF MIGRAINE SYMPTOM GROUPINGS: A COMPARISON OF THREE DIFFERENT CLUSTERING METHODS

Carla C.M. Chen,[1,2] Jonathan M. Keith,[1] Kerrie Mengersen,[1] and Dale R. Nyholt[2]

[1] *School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia*

[2] *Queensland Institute of Medical Research, Brisbane, Australia*

Although migraine is highly prevalent in our society (affecting up to 25% of females and 7.5% of males), its etiology remains relatively obscure and there are no laboratory based diagnostic tests that identify those who suffer from the disorder. Moreover, common complex disorders such as migraine are typically an ensemble of various symptoms and/or subtypes with diagnosis based upon satisfaction of clinically accepted criteria. However, the resulting classification of individuals as 'affected' or 'unaffected' for subsequent genetic study may not accurately reflect the phenotypic heterogeneity and/or underlying genetic basis of the disorder. Migraine is a debilitating and painful disorder. The diagnosis of migraine relies on self-reported symptomatology and is based upon the criteria developed by the International Headache Society (IHS). The 2 main subtypes of IHS migraine are migraine with aura (MA) and migraine without aura (MO). However, our preliminary results of principle component analysis (PCA) on Australian twin study data indicated that MA and MO are not two separate entities. These results also indicate the statistical clustering approaches yield more homogenous groupings than the IHS criteria. Compared to the 2 groups (MO and MA) suggested by IHS, latent class analysis (LCA) of migraine symptoms determined the optimum number of subgroups to be 3. We also identified some confusion in the literature regarding terminology and use of another clustering procedure, grade of membership (GoM). In particular, there are 2 types of GoM, one has a form of mixture model (GoM) and the other is a partitioning method (GoM-Fanny). Here we extend our previous findings, by comparing heritability measures between LCA, GoM and GoM-Fanny. We will also contrast the results of genome-wide linkage analyses using phenotypes based on the LCA, GoM and GoM-Fanny symptom groupings.

GENOME-WIDE ASSOCIATION STUDY OF BREAST CANCER: RESULTS FROM THE BREAST CANCER ASSOCIATION CONSORTIUM

Antonis C. Antoniou,[4] Olga M. Sinilnikova, Jacques Simard, Susan L. Neuhausen, Jeffery P. Struewing,[2] Dominique Stoppa Lyonnet, GEMO, Timothy R. Rebbeck, MAGIC, Kate Nathanson, Susan Domchek, Andrew Godwin, Anna Jakubowska, Susan Peock, EMBRACE, Rita K. Schmutzler, kConFaB,[3] Kenneth Offit, Tomas Kirschhoff, Eitan Friedman, Gad Rennert, Irene L. Andrulis, Peter Devilee, Frans Hogervosrt, Matti Rookus, Mark H. Greene, Javier Benítez, Csilla I. Szabo, Heli Nevanlina, Ute Hamann, Adalgeir Arason, Paolo Radice, Paolo Peterlongo, Maria Caligo, Ake Borg, Annika Lindblom, Fergus Couch, Douglas F. Easton[4] and Georgia Chenevix-Trench[1] on behalf of the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA)

[1] *Queensland Institute of Medical Research, Brisbane, Australia*

[2] *Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, United States of America*

[3] *The Kathleen Cuningham Foundation Consortium for Research Into Familial Breast Cancer, Peter MacCallum Cancer Center, Melbourne, Australia*

[4] *Department of Oncology, University of Cambridge, United Kingdom*

Breast cancer exhibits familial aggregation, consistent with variation in genetic susceptibility to the disease. Known susceptibility genes account for less than 20% of the familial risk of breast cancer, and the residual genetic variance is likely to be due to variants conferring more moderate risks. To identify further susceptibility alleles, a 3-stage genome-wide association study was conducted in 390 familial breast cancer cases and 364 controls, followed by analysis of 12,711 SNPs in 3990 unselected breast cancer cases and 3916 controls, followed by a third stage involving 21,668 cases and 20,973 controls from 22 studies (including 3 from Australia) participating in the Breast Cancer Association Consortium (BCAC). SNPs in 5 novel independent loci exhibited strong and consistent evidence of association with breast cancer ($p < 10\text{-}7$), 4 of which contain plausible causative genes (*FGFR2*, *TNRC9*, *MAP3K1* and *LSP1*). In addition, at SNP at 8q24, a region previously found to be associated with risk of prostate cancer, was also identified as a breast cancer susceptibility locus. At the second stage, 1579 SNPs were significant at the $p < .05$ level compared with an estimated 1343 that would be expected by chance, indicating that many further common, susceptibility alleles may be identifiable by this approach. *FGFR2* and *TNRC9* were identified by the Cancer Genetic Markers of Susceptibility (CGEMS) and deCODE GWAS of breast cancer, respectively, along with an additional SNP in a gene desert on 2q. We are now investigating whether the associations between these five SNPs and breast cancer risk vary by clinically important tumor characteristics in up to 20,400 invasive breast cancer cases and 25,550 controls from 18 studies. We are also evaluating their influence on overall survival in 13,527 of these cases. Our results, and those of the deCODE GWAS, show that common genetic variants influence the pathological subtype of breast cancer, and provide further support for the hypothesis that ER-positive and ER-negative disease are biologically distinct. Understanding the etiologic heterogeneity of this disease may ultimately result in improvements in prevention, early detection and treatment.

FINDING FUNCTIONAL VARIANTS WITHIN NONCODING DNA: BALDNESS AND THE ANDROGEN RECEPTOR

Joanna Cobb,[1] Fiona Christensen,[2] Stephen Harrap,[1] Steven Petrou,[2] and Justine Ellis[1]

[1] *Department of Physiology, University of Melbourne, Australia*

[2] *Howard Florey Institute, Melbourne, Victoria, Australia*

The search for coding sequence variants for complex polygenic conditions has been a key focus of recent genetic research. Despite significant effort, few variants have been identified suggesting that coding variants are less common than thought and that noncoding sequence variation may underlie complex conditions. Noncoding DNA harbors regulatory elements capable of making subtle changes to gene expression. However, noncoding DNA is currently poorly annotated, complicating the task of assigning potential functional significance to noncoding SNPs. Efforts to annotate the non-coding genome are commonly based on the theory that highly conserved sequences are functionally important (Pennacchio et al., 2006). We were the first to publish significant evidence of genetic association of Male Pattern Baldness (MPB) with the androgen receptor (*AR*) gene via a synonymous SNP in exon 1 (rs6152) in our Caucasian Victorian Family Heart Study population (Ellis et al., 2001). Since this finding, three independent studies have confirmed this association. Despite extensive analysis of *AR* coding sequence (Ellis et al., 2007) we are yet to identify the causative variant for MPB. We have now begun searching non-coding DNA surrounding *AR*. *AR* is located at Xq12 in a region of strong LD spanning approximately 1Mb. There are ~ 150 validated SNPs in strong LD with rs6152, making the identification of candidate functional MPB SNPs a daunting task. We are establishing an enhancer screening technique, based on the methods of Pennacchio et al. to explore the transcription regulating ability of evolutionarily conserved noncoding sequences surrounding *AR*. We have identified a list of 25 candidate regions with high conservation between several vertebrate species. These sequences will be inserted into *LacZ* reporter constructs. Tissue specific expression of *LacZ* in embryos will identify candidate regions showing transcriptional enhancing ability and permit comparison and analysis of the transcriptional effects of SNPs. Implementation of this method in relation to *AR* and MPB may also inform other research aimed at identifying functional noncoding variants for complex disease.

HERITABILITY AND PREVALENCE OF MIGRAINE IN THE NORFOLK ISLAND POPULATION ISOLATE

H. Cox,1 C. Bellis,[1] S. Quinlan,1 T. Dyer,[2] J. Blangero,[2] and L. Griffiths[1]

[1] *Genomics Research Centre, Griffith University, Gold Coast, Australia*

[2] *Southwest Foundation for Biomedical Research, San Antonio, Texas, United States of America*

Studies have shown that the onset of migraine with and without aura is influenced by both genetic and environmental variables. Presently the type and number of susceptibility genes involved in both of these common forms of migraine are unknown. The objective of this study was to assess the prevalence and heritability of migraine in the population of Norfolk Island, a genetic isolate located several thousand kilometers off the eastern coast of Australia. Migraine was assessed by questionnaires and diagnosed using International Headache Society criterion. A total of 372 individuals with phenotypic data were assessed for migraine prevalence. These individuals comprise a complete pedigree of 6537 individuals dating back 11 generations to 12 maternal Tahitian and 6 paternal European founders. A total migraine prevalence of 23% was observed. Migraine with and without aura were reported to affect 13% and 10% of individuals, respectively. Interestingly 5% of males compared to 18% of females were affected. These results are comparable to estimates in outbred populations, which have been reported to vary from 4% to 9% in males to 11% to 25% in females. Heritability estimates were generated using the statistical program SOLAR (v4.0.7). The heritability of migraine was significant ($p < .05$) and calculated as .42. Heritability estimates for migraine have been reported to range from 34% to 57% in worldwide populations. This data supports the use of the Norfolk Island population as a potentially useful genetic isolate for gene mapping studies aimed at identifying migraine susceptibility genes.

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

LARGE SCALE TRANSCRIPTIONAL PROFILING FOR THE IDENTIFICATION OF GENES INFLUENCING COMMON COMPLEX DISEASES

Joanne E. Curran,[1] Matthew P. Johnson,[1] Harald H. H.Goring,[1] Thomas D. Dyer,[1] Jac C. Charlesworth,[1] Shelley A. Cole,[1] David L. Rainwater,[1] Michael C. Mahaney,[1] Jeremy B. M. Jowett,[2,3] Jean W. MacCluer,[1] Greg R. Collier,[3] Eric K. Moses,[1] and John Blangero[1,3]

[1] *Southwest Foundation for Biomedical Research, San Antonio, Texas, United States of America*

[2] *International Diabetes Institute, Melbourne, Australia*

[3] *ChemGenex Pharmaceuticals, Geelong, Australia*

Classical gene discovery approaches have been slow to identify validated susceptibility genes and are limited by our current knowledge of the genetic pathways underlying disease. In this study we propose a new objective approach to candidate gene discovery that uses large-scale transcriptional profiling to identify novel *cis*-acting genes that correlate with a given quantitative trait. Using RNA extracted from lymphocytes, we obtained genome-wide quantitative transcriptional profiles from 1240 individuals in the San Antonio Family Heart Study. For each sample, 47,289 transcripts were measured using Illumina Human-6 BeadChips. Our goal was to detect heritable quantitative variation in expression and look for QTLs influencing expression levels. In this dataset, we were able to significantly detect ~ 20,000 transcripts. In general, detected transcripts exhibited significant heritable quantitative variation, with less than 1% failing to show evidence for genetic factors influencing expression levels. A significant heritability can be viewed as a validation of the biological utility of such quantitative transcript measures. Using quantitative trait linkage analysis, we identified over 3000 autosomal *cis*-acting QTLs for which we have significant evidence for variation influencing expression levels at the transcript's genomic location. We then tested each transcript for correlation with a number of metabolic syndrome-related phenotypes. As a preliminary test of the underlying importance of *cis*-regulated variation, we chose 10 *cis*-regulated genes whose quantitative transcript levels showed significant correlation with at least one metabolic syndrome-related phenotype for further molecular examination. Resequencing of 2kb of the putative promoter in 96 founders revealed that 70% of these genes exhibited significant evidence for association between promoter SNPs and expression levels. Further investigation of one gene, *VNN1*, in an extended sample dramatically improved the evidence for association between promoter SNPs and expression levels ($p < 1 \times 0^{-77}$). Formal functional analyses using EMSA confirmed the likely regulatory function of one proximal promoter variant. Statistical genetic analysis further revealed that this functional variant was also strongly associated with HDL-C levels confirming the previously unknown role of *VNN1* as a determinant of this important risk-related phenotype. Our results point to the utility of large scale family-based transcriptional data bases for identifying human quantitative trait loci.

SPECTRUM OF *CYP1B1* MUTATIONS IN AUSTRALIAN PRIMARY CONGENITAL GLAUCOMA PATIENTS

David P. Dimasi,[1] Alex W. Hewitt,[1,2] Tania Straga,[1] John Pater,[1] Jane R. MacKinnon,[3] James E. Elder,[4] Theresa Casey,[5] David A. Mackey,[2,4,6] and Jamie E. Craig[1]

[1] *Department of Ophthalmology, Flinders University, Flinders Medical Centre, Adelaide, Australia*

[2] *Clinical Genetics Unit, Centre for Eye Research Australia, University of Melbourne, Royal Victorian Eye and Ear Hospital, Melbourne, Australia*

[3] *Ophthalmology Department, Royal Hospital for Sick Children, Glasgow, Scotland*

[4] *Department of Ophthalmology, Royal Children's Hospital; Melbourne, Australia*

[5] *Department of Ophthalmology, Women's and Children's Hospital, Adelaide, Australia*

[6] *Department of Ophthalmology, University of Tasmania, Royal Hobart Hospital, Hobart, Australia*

In Australia, primary congenital glaucoma (PCG) has an estimated incidence of 1 in 30,000 births. Although 3 putative PCG loci have been identified, mutations in only one gene, *CYP1B1*, have been shown to cause PCG. The prevalence of *CYP1B1* mutations in PCG varies between different populations, approaching 100% in Slovakian Gypsies and Saudi Arabians, while only accounting for 20% in people of Japanese ethnicity. We investigated the prevalence and spectrum of *CYP1B1* mutations in an Australian PCG cohort. DNA was extracted from 35 probands with PCG who were predominately recruited from South-Eastern Australia. Following PCR, the coding exons of *CYP1B1* were directly sequenced. The significance of all identified mutations was assessed through screening 200 chromosomes from examined normal individuals. Of the 35 probands screened, 5 (14.3%) were compound heterozygotic and 1 (2.9%) homozygotic for *CYP1B1* mutations. Two (5.7%) probands were heterozygotic for a *CYP1B1* mutation. A significant degree of allelic heterogeneity also existed, with 10 different mutations identified, 4 of which (D192Y, G329D, P400S and 1208insTCATGCC) were novel. The novel mutations identified were located in codons well-conserved across different species and were not found in control individuals. Our results indicate that mutations in *CYP1B1* are responsible for approximately 20% of PCG cases in Australia. By establishing a *CYP1B1* mutation frequency and spectrum in an Australian cohort, our data will help the implementation of neonatal screening and genetic counseling programs in the future.

LINKAGE ANALYSIS OF A SEVEN-GENERATION PROSTATE CANCER FAMILY IDENTIFIES A PROSTATE CANCER SUSCEPTIBILITY GENE AT CHROMOSOME 5P13Q12

Liesel M. FitzGerald,[1] Briony Patterson,[1] Russell Thomson,[1] Andrea Polanowski,[1] Stephen Quinn,[1] Jesper Brohede,[3,4] Timothy Thornton,[10] David Challis,[5] David Mackey,[1,6,7] Terence Dwyer,[1,8] Simon Foote,[1] Garry N. Hannan,[3,4] James Stankovich,[1,2] James D. McKay,[1,9] and Joanne L. Dickinson[1]

[1] *Menzies Research Institute, University of Tasmania, Hobart, Australia*

[2] *Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia*

[3] *CSIRO Preventative Health National Research Flagship, Australia*

[4] *CSIRO Molecular and Health Technologies, North Ryde, New South Wales, Australia*

[5] *Histopathology Department, Royal Hobart Hospital, Hobart, Australia*

[6] *Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia*

[7] *Department of Ophthalmology, Royal Children's Hospital, Melbourne, Australia*

[8] *Murdoch Children's Research Institute, Melbourne, Australia*

[9] *International Agency for Research on Cancer, Lyon, France*

[10] *Department of Statistics, University of California, Berkeley, United States of America*

Identification of genes conferring susceptibility to prostate cancer has proven particularly challenging, even in the context of gene discovery in complex disease. The heterogeneous nature of prostate cancer is one significant difficulty encountered: to partially address this problem we selected a large prostate cancer pedigree for genome-wide linkage analysis from a population that is relatively homogeneous genetically. In a 7-generation Tasmanian pedigree containing 25 prostate cancer cases, 7 cases and selected first degree relatives were genotyped with Affymetrix 10K SNP arrays. Linkage analysis was performed to identify haplotypes inherited identical-by-descent by multiple cases. Microsatellites were genotyped in the region of interest utilizing additional deceased case DNA samples derived from paraffin embedded tissue specimens. Linkage analysis revealed a 14 megabase haplotype inherited in common by 8 patients ($p = .0017$). Candidate genes in this region were prioritized, and 8 of these were resequenced in 6 individuals to identify polymorphisms segregating with the disease-susceptibility haplotype. Re-sequencing of candidate genes identified 2 polymorphisms within a single gene. These polymorphisms were genotyped in an independent Tasmanian dataset comprising 127 cases with hereditary prostate cancer, 412 sporadic cases and 319 unaffected controls. Polymorphism A was significantly associated with risk both in familial prostate cancer cases ($p = .015$,) and sporadic prostate cancer cases ($p = .0018$) and in all cases combined ($p = .0009$). Polymorphism B was associated with risk in familial cases ($p = .0198$) and in all cases combined ($p = .088$) and but was not significant for sporadic cases alone ($p = .0696$). The gene identified has previously been associated with risk of other cancers.

GENETIC MAPPING OF MYOPIA SUSCEPTIBILITY LOCI

Mohamed Dirani,[1,2] Andrea J. Richardson,[1,2] Kelly K. Pertile,[1,2] Maria Schäche,[1,2] Katrina Scurrah,[3,4] and Paul N. Baird[1,2]

[1] *Centre for Eye Research Australia, University of Melbourne, Australia*

[2] *Vision Cooperative Research Centre, Sydney, Australia*

[3] *Department of Physiology, University of Melbourne, Australia*

[4] *Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Australia*

Myopia (short sightedness) is a complex trait influenced by both genetic and environmental risk factors. To date there have been four myopia susceptibility loci (MYP7-10) identified in 1 previous twin study, but these are yet to be independently verified. Using an independent yet ethnically and phenotypically similar twin cohort, we sought to look for linkage to these chromosomal regions. We recruited 223 dizygotic twin pairs aged between 18 to 80 years from the Australian Twin Registry and assessed for evidence for linkage using Haseman-Elston regression analysis. Each twin pair underwent a comprehensive eye examination, with dilated objective refraction and ocular biometric measurements being obtained. In addition, a general questionnaire was administered and 18ml of blood was collected. The mean refraction measurements were –0.02 ± 2.17 DS, with a range of –14.5 to +6.25 DS. The mean ocular biometric measurements were 23.43 ± 1.08 mm for axial length, 3.47 ± 0.42 mm for anterior chamber depth and 44.13 ± 1.50 for corneal curvature. Using ocular refraction as the quantitative phenotype, the maximum LOD scores for each chromosomal region analyzed were 0.578 for 11p13, 0.255 for 8q23, 1.126 at 3p26 and 0.051 at 4q21. No evidence for linkage of myopia or its underlying biological components such as eye length to the MYP7-10 regions was found in this twin cohort. This is the first study to assess for linkage in a secondary myopia twin cohort and highlights the problems associated with applying linkage results from complex traits to the other populations.

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

A CANDIDATE GENE STUDY OF THE SEROTONIN AND DOPAMINE TRANSPORTER GENES IN RELATION TO DEPRESSION IN PARKINSON'S DISEASE

N. Dissanayaka,[1,2] P. Silburn,[1,2] J. O'Sullivan,[1] and G. Mellick[1,2]

[1] *School of Medicine University of Queensland, Australia*

[2] *Eskitis Institute for Cell and Molecular Therapies Griffith University, Australia*

Parkinson's disease (PD) affects over 40,000 Australians, half of which experience depression. The depression rate is greater than that seen in equivalent disease samples and the general population. Depression contributes more to a poorer quality of life in PD patients than the motor disability caused by the disease. The genetic determinants of depression remain unclear. The etiology of depression is heterogeneous, although its genetic determinants may be more homogenous in the setting of PD. Given that altered neurotransmitter levels of dopamine and serotonin are observed in both PD and depression, the genes that encode for their transporters are candidate genes for depression in PD. This project aims to comprehensively examine the association between the *SLC6A3* (dopamine transporter) and *SLC6A4* (serotonin transporter) genes and depression in PD. We recruited 331 PD patients and 150 controls of European Caucasian origin. A lifetime history of depression was identified using the geriatric depression scale and screening questions. Depressed PD patients were age and gender matched to nondepressed PD and control subjects. The HapMap project database was used to select tagging SNPs (7 in *SLC6A3* and 17 in *SLC6A4)*. Genomic DNA was extracted from whole blood and SNPs were genotyped using a Sequenom system. We also genotyped VNTRs in each gene and a 44bp deletion in the *SLC6A4,* using PCR. Haplotype analyses were performed using the SNPStats program. Univariate and multivariate logistic regression analyses were performed using SPSS. Five common (frequency > 5%) haplotypes defined *SLC6A3* and 4 common haplotypes defined *SLC6A4.* Neither showed association with depression in PD (Chi square probability value: *SLC6A3*: $p = .41$; *SLC6A4*: $p = .69$). There were no significant differences in genotype frequency between PD cases with and without depression for any of the variants studied (all uncorrected p values > .05). This is the first study to comprehensively examine the association between *SLC6A3* and *SLC6A4* genes, and depression in PD. We showed that the common variants of these 2 genes were not associated with depression in PD.

GENETIC ANALYSIS OF ATTENTION-DEFICIT HYPERACTIVITY DISORDER 'SIGNATURE' BRAIN MARKERS

Carol Dobson-Stone,[1,2] Kerrie D. Pierce,[1] Daniel F. Hermens,[3,4] Simon D. Clarke,[3,5] Michael R. Kohn,[3,5] C. Richard Clark,[6] Evian Gordon,[3,4,7] Leanne M. Williams,[3,7] and Peter R. Schofield[1,2]

[1] *Prince of Wales Medical Research Institute, Sydney, Australia*

[2] *University of New South Wales, Sydney, Australia*

[3] *The Brain Dynamics Centre, Westmead Millenium Institute, Westmead Hospital and Western Clinical School, University of Sydney, Sydney, Australia*

[4] *The Brain Resource International Database, The Brain Resource Company, Sydney, Australia*

[5] *Adolescent Medicine, Westmead Hospital, Sydney, Australia*

[6] *Cognitive Neuroscience Unit, Flinders University, Adelaide, Australia*

[7] *Psychological Medicine, University of Sydney, Sydney, Australia*

Attention-deficit hyperactivity disorder (ADHD) is one of the most common childhood psychiatric disorders, with a prevalence of 3–9%. Those affected face social, educational and psychological problems that can last into adulthood. Heritability estimates of ~ 80% indicate that genes play a strong role in the susceptibility to ADHD. However, the genetic contributions are likely to include multiple genes, each of moderate effect. Some well-supported candidate genes have been proposed (e.g., the dopamine receptor D4 [*DRD4*], dopamine transporter [*SLC6A3*] and serotonin transporter [*SLC6A4*] genes), but inconsistent results between studies mean that consensus has not been reached for any gene. One problem may be that most genetic studies use subjective clinical ratings of ADHD. There is also growing support for the view that ADHD, like many complex neurological disorders, is not a discrete entity, but may represent an extreme end of the spectrum of normal brain function. Thus the classification of people into 'ADHD' and 'normal' groups may be too simplistic. The Brain Resource International Database (BRID) is a collection of neurological data (neuropsychological testing and brain imaging) from a large cohort of healthy individuals and clinical groups. BRID test scores from ADHD children and matched controls were used to derive 4 composite cognitive markers of ADHD. These markers provide a quantitative measure of 'ADHD-ness' that can be assessed in healthy and ADHD children, rather than assigning a 'yes/no' diagnosis of ADHD. They are also objective performance measures that may lie closer to the underlying biology of the disorder than diagnosis or clinical symptoms alone, and so may provide stronger statistical associations with predisposing genes. We have genotyped the *DRD4* variable number tandem repeat (VNTR), *SLC6A3* VNTR and *SLC6A4* promoter insertion/deletion polymorphisms in over 300 healthy children (6–17 years old). We are using the ADHD 'signature' markers to investigate the contribution these polymorphisms make to variation in ADHD-like traits in this sample. Identification of ADHD trait–gene associations in the general population will help elucidate the biological pathways involved in ADHD brain function.

SNP ASSOCIATION ANALYSIS OF FILAMIN B (*FLNB*) AND ADULT HEIGHT IN THE VICTORIAN FAMILY HEART STUDY (VFHS)

Anna E. Duncan,[1] Katrina J. Scurrah,[1] Stephen B. Harrap,[1] and Justine A. Ellis[1,2]

[1] *Department of Physiology, The University of Melbourne, Melbourne, Australia*

[2] *Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne, Australia*

Adult height is widely regarded as the complex human trait most amenable to genetic dissection. It shows a high degree of heritability and can be easily and accurately measured. In the Victorian Family Heart Study (VFHS), a healthy population-based Caucasian cohort consisting of adult 2-generation families, we have previously performed a genome-wide linkage analysis for adult height and subsequent fine mapping, identifying a region on chromosome 3p (Ellis et al., 2007). This region contains the physiologically plausible candidate gene, Filamin B (*FLNB*). FLNB is a cytoplasmic actin-binding protein, whose expression pattern includes growth plate chondrocytes. Mutations in *FLNB* are found in various rare skeletal disorders that display short stature. This study investigated whether there was association between variation in *FLNB* and adult height in unrelated individuals from the VFHS that were selected for tall ($n = 96$, range 175–199 cm) or short ($n = 96$, range 141–163 cm) stature. We selected 17 SNPs that tag haplotype blocks throughout the gene based on HapMap data. Genotyping was performed using a variety of methods including sequencing and single nucleotide primer extension. Alleles and genotypes of the tag SNPs were compared between the tall and short groups using χ^2. Eight SNPs demonstrated association at $p < .01$. In particular, alleles of rs1658397, located in intron two of *FLNB*, were highly significantly different between tall and short groups ($p = .00009$). In general, associated SNPs were in low levels of LD with each other, suggesting that more than one functional variant may exist within *FLNB* relevant to stature. We are currently analyzing additional SNPs in an increased sample size to further define these associations.

IS ANDROGEN RECEPTOR COPY NUMBER VARIATION IMPORTANT IN MALE PATTERN BALDNESS?

Justine Ellis,[1,2] Joanna Cobb,[1] Stephen Harrap,[1] and Stefan White[2]

[1] *Department of Physiology, University of Melbourne, Victoria, Australia*

[2] *Murdoch Childrens Research Institute, Parkville, Victoria, Australia*

Copy Number Variation (CNV) is a recently described phenomenon whereby large segments of DNA, from thousands to millions of base pairs, can vary in copy number between individuals. Such variation can lead to dosage imbalance of genes within these segments. Current thinking suggests that CNVs will play an important role in complex disease, and several important examples have been published recently. CNVs can also exist in linkage disequilibrium with SNPs, and thus association of a phenotype with SNPs may be due to the involvement of CNV in determining the phenotype. The important role of the androgen receptor, *AR*, with male pattern baldness (MPB) is now well established. Expression of *AR* is increased in balding hair follicles, and four studies worldwide, including our own original finding (Ellis et al., 2001), have shown strong genetic association of *AR* SNPs with MPB. We have shown that *AR* coding variants are not causative in this condition, and thus we are considering a range of possibilities, including functional noncoding variation, epigenetic differences, and CNV to explain the increased expression levels. Two studies to date have demonstrated that the genomic region containing *AR* is subject to CNV (McCarroll et al., 2006; Redon et al., 2006). We therefore hypothesize that *AR* CNV is associated with MPB. As part of a pilot study to test this hypothesis, we are currently using Multiplex Ligation-dependent Probe Amplification (MLPA) to detect CNV in 6 *AR* exons and a conserved, noncoding region upstream of AR in 24 cases (males $\prime\prime$ 30 years with cosmetically significant baldness) and 24 controls (males ≥ 50 years with no indication of baldness). Results of this analysis will be available at the meeting.

AMACR GENE VARIANTS ARE SIGNIFICANTLY ASSOCIATED WITH PROSTATE CANCER RISK IN THE TASMANIAN POPULATION

Liesel M. FitzGerald,[1] Russell Thomson,[1] Timothy Thornton,[2] Simon Foote,[1] James Stankovich,[1,3] James D. McKay,[1,4] and Joanne L. Dickinson[1]

[1] *Menzies Research Institute, University of Tasmania, Hobart, Australia*

[2] *Department of Statistics, University of California, Berkeley, United States of America*

[3] *Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia*

[4] *International Agency for Research on Cancer, Lyon, France*

Alpha-methylacyl-CoA racemase (*AMACR*) is a well-characterized enzyme that plays a critical role in the metabolism of branched-chain fatty acid molecules. Evidence to suggest an involvement of the AMACR gene in prostate cancer has come from several sources including genome-wide linkage studies, dietary studies and multiple microarray and immunohistochemical studies. More recently, it was shown that AMACR expression is highest in localized prostate cancer and decreases in metastatic prostate

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

cancer and that this decrease in expression is associated with progression and cancer specific death. To date only 1 study has investigated whether sequence variants of AMACR alter the risk for prostate cancer. Zheng and associates (2002) found 4 missense changes (M9V, D175G, S291L, and K277E) that had significantly different genotype frequencies between HPC cases and unaffected controls; haplotype analysis of the M9V and D175G SNPs provided stronger evidence for an association. Multipoint linkage analysis demonstrated that these variants strongly co-segregate with disease in HPC families. The AMACR variants M9V and D175G were screened in a Tasmanian dataset of 359 sporadic and 130 familial Tasmanian prostate cancer cases and allele frequencies were compared to those of a control group consisting of 340 individuals. Genotype-disease associations were estimated by computing odds ratios (OR) using unconditional logistic regression. The MQLS method was used to test for significance of association while adjusting for relatedness between familial cases. The results from these analyses will be presented.

THE ATG16L1 VARIANT (THR300ALA) IS CONFIRMED AS A SUSCEPTIBILITY LOCUS IN CROHN'S DISEASE AND MAY ALSO BE ASSOCIATED WITH ULCERATIVE COLITIS IN AUSTRALIAN IBD PATIENTS

Beth Fowler,[1] Lisa Simms,[1,2] James Doecke,[2] Zhen Zhen Zhao,[2] Penny Webb,[2] Nick Hayward,[2] David Whiteman,[2] Timothy Florin,[4] Grant Montgomery,[2] Juleen Cavanaugh,[5] and Graham Radford-Smith[1,3]

[1] *IBD Laboratory, Royal Brisbane and Women's Research Foundation, Brisbane, Australia*

[2] *Queensland Institute of Medical Research, Brisbane, Australia*

[3] *Department of Gastroenterology, Royal Brisbane and Women's Hospital, Brisbane, Australia*

[4] *Mater Adult Hospital, Brisbane, Australia*

[5] *Australian National University, Canberra, Australia*

Crohn's disease (CD) and ulcerative colitis (UC), are the two most common forms of inflammatory bowel disease, and represent a significant burden to healthcare in developed countries. The Thr300Ala variant in the autophagy-related 16-like 1 (ATG16L1) gene has been newly identified as a susceptibility locus for CD by genome-wide association scan of nonsynonymous DNA polymorphisms. We assessed the contribution of this variant in determining disease susceptibility and phenotype in 2 independent Australian cohorts of CD and UC, and investigated the interaction between CD-associated NOD2 variants and carriage of this variant in patients. The Thr300Ala variant (rs2241880) was genotyped in 682 CD and 588 UC cases, and 1233 controls from South-East Queensland, together with 251 IBD trios. Case-control analysis of genotype and allele frequencies and family-based transmission disequilibrium testing was conducted. A strong association with CD was demonstrated ($p = 1.3 \times 10^{-6}$, OR 1.39 [1.22,1.59]), which was confirmed in family studies ($p = .004$). Ileal patients with or without colonic disease were found to have a significantly highly frequency of the Ala variant ($p = 9.0 \times 10^{-8}$, OR = 1.58 [1.34,1.87]). Analysis of the interaction between ATG16L1 and NOD2 indicated that these two loci independently contribute to Crohn's disease risk. A significant decrease in GG genotype frequency was observed in ulcerative colitis and (UC = 20.4%; AG = 55.8%) compared with controls (GG = 27.2%; AG = 48.5%; $p = .003$). The frequency of the G allele was also moderately decreased in UC cases (48.3%) compared with controls (51.4%), although this was not significant ($p = .077$; OR = 0.88 [0.77, 1.01]). We confirm the strong association between rs2241880 and Crohn's disease, specifically ileal subphenotype, and also suggest a possible association of this variant with ulcerative colitis.

GENOME-WIDE LINKAGE AND TWO-DIMENSIONAL INTERACTION ANALYSIS OF 67 AUSTRALIAN PEDIGREES REVEALS COMPLEX GENETIC ARCHITECTURE FOR BIPOLAR AFFECTIVE DISORDER

Janice M. Fullerton,[1] Jennifer A. Donald,[2] Philip B. Mitchell,[3,4] and Peter R. Schofield[1]

[1] *Prince of Wales Medical Research Institute, Sydney, New South Wales, Australia*

[2] *Department of Biological Sciences, Macquarie University, Sydney, New South Wales, Australia*

[3] *School of Psychiatry, University of New South Wales, Sydney, New South Wales, Australia*

[4] *Black Dog Institute, Prince of Wales Hospital, Sydney, New South Wales, Australia*

As part of an ongoing study into the genetics of bipolar affective disorder (BAD), a series of genome-wide linkage scans were previously conducted on Australian extended pedigrees affected with BAD in 5 waves of parametric analysis, revealing major peaks on 4q35 (Adams et al., 1998); 13q14 (Badenhop et al., 2001); 3q25 and 13q14 (Badenhop et al., 2002); 15q28 and 1p31 (data not published); and 15q26 and 6q26 (McAuley et al., in press). The whole dataset now consists of 67 pedigrees, containing 852 individuals with full genome scan data of whom 286 were affected. A pooled 1-dimensional linkage analysis and 2-dimensional interaction analysis has been conducted in the full Australian cohort. Nonparametric linkage analysis was performed using a broad phenotype definition, where individuals with a clinical diagnosis of bipolar disorder, schizoaffective disorder or recurrent unipolar depression were considered affected. NPL scores from individual families were subject to chromosome-by-chromosome interaction analysis by bivariate correlation. Major NPL linkage peaks were identified on chromosomes 2q31, 4q34, 7q21, 9q31, 13q13, 17q21, 19q13, 20p12 (z scores of 3.12, 2.75, 3.01, 2.61, 2.7, 2.95, 2.69 and 2.35 respectively). Seven main interactions were detected, using a 5% threshold for the most significant inter-chromosomal correlations ($p > .000001$), including two interactions at NPL peaks (4q34 and 1p36; and 17q25 and 6q24), and two interactions involving the same locus at 11q24 (with 2p14 and 5p13 respectively). Nonparametric linkage supports loci previously detected in individual waves using parametric models, and has revealed new loci not previously detected in our cohort. Interaction analysis has revealed a second dimension underlying susceptibility to bipolar affective disorder, and highlights the need for a more powerful approach to detect multiple loci of small genetic effect.

AN INVESTIGATION OF GENE EXPRESSION IN HUMAN BREAST CANCER USING TISSUE MICROARRAY

P. Gabrovska, R. Smith, and L. Griffiths

Genomics Research Centre, Griffith University, Gold Coast, Australia

One in 11 women will be diagnosed with breast cancer, the most common cause of cancer-related death in Australian women. A number of genetic mutations have been identified in human breast cancer, yet the specific mutations required to act in concert to form breast carcinoma cells remain incompletely defined. In this study, a gene expression profiling approach was used to uncover differentially expressed genes related to the disease and the pattern of regulation across the 3 grades of breast cancer development and progression. The investigation was undertaken using 12 breast archival invasive ductal carcinoma tissue samples of progressive grades, 3 of which were benign controls. From these samples, mRNA was extracted and gene expression profiles were determined using microarray hybridisation. Results were analyzed at significance levels of .01 and .05 to detect significantly differentially expressed genes in breast tissue compared to control tissue. In the analysis of the array data, a series of t tests revealed that 184 genes were found to be significantly ($p = .01$) differentially expressed in at least one of the group comparisons, 42 of which were identified as being involved in processes previously implicated or associated with breast cancer. It was also discovered that 8 of these genes were significantly differentially expressed across more than one comparison of groups which included CLDN10, CXCL16, EPSTI1, LOC441259, CDC42EP3, ZAN, TCEA3 and PALMD. Investigating patterns of expression indicated that most of the differentially regulated genes showed up regulation from controls to grade 1 tumors, then a drop in regulation in grade 2 tumors and finally a considerable up-regulation in the final grade 3 tumors. One particular gene, the chemokine CXCL16 was the only gene found to be significantly differentially expressed ($p = .01$) in more than 2 comparisons, thus determining regulation of the gene in a greater number of grade 3 tissues would prove beneficial in potentially identifying this gene as a target for diagnostics and therapeutics. While the results of the study require validation, candidate genes for further investigation have been identified and future studies could now investigate these as potential targets for diagnostic and therapeutic development.

A NOVEL MISSENSE MUTATION IN *NKX2-5* GENE FOUND IN AN AUSTRALIAN FAMILY WITH AUTOSOMAL DOMINANT DILATED CARDIOMYOPATHY AND CONDUCTION-SYSTEM DISEASE

Guanglan Guo,[1] Mauro W. Costa,[1] Robyn Otway,[1] M. Leticia Castro,[1] Christopher Hayward,[2] Anne Keogh,[2] Peter Macdonald,[2] Richard Harvey,[1] and Diane Fatkin[1,2]

[1] *Victor Chang Cardiac Research Institute, Darlinghurst, New South Wales, Australia*

[2] *Cardiology Department, St Vincent's Hospital, Darlinghurst, New South Wales, Australia*

Dilated cardiomyopathy (DCM) is a heart muscle disorder characterized by dilatation and contractile dysfunction of the left ± right ventricles. Recent molecular genetic studies suggest that inherited gene defects underlie DCM in a significant proportion of cases. Familial DCM may present as an isolated disorder, or in association with conduction-system disease (CD). *NKX2-5* is a cardiac homeobox transcription factor that has a critical role in the temporal and spatial patterns of gene expression in the developing heart. Mutations in the *NKX2-5* gene cause congenital heart disease and CD. Studies by our group and others have shown that *Nkx2-5* deficiency in mice is also associated with adult-onset contractile dysfunction. We hypothesized therefore that *NKX2-5* mutations might be a cause of human familial DCM. We have screened the coding region of the *NKX2-5* gene by DNA sequencing in probands of 90 Australian families with DCM ± CD. One novel missense mutation, Ile184Met (C729G), which is located within the highly conserved homeodomain, was identified in one family with DCM + CD. The variant segregated with the clinical affection status in this family and was absent in 208 normal control chromosomes. These results suggest that *NKX2-5* is a novel disease-causing gene for familial DCM + CD but mutations are relatively uncommon. Our findings highlight the potential importance of cardiac transcription factors not only in heart development but also in adult heart function.

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

RESEQUENCING CANDIDATE GENES OF IRON METABOLISM AND THE RELATIONSHIP BETWEEN SNP OR HAPLOTYPE DOSAGE AND IRON PHENOTYPE IN INDIVIDUALS AT HIGH AND LOW RISK OF IRON OVERLOAD

Lyle C. Gurrin,[1] Clare C. Constantine,[1] Chris D.Vulpe,[2] Gregory J. Anderson,[3] Melanie Bahlo,[4] and Christine E. McLaren[5]

[1] *Centre for MEGA Epidemiology, University of Melbourne, Australia*

[2] *Nutritional Sciences and Toxicology, University of California, Berkeley, United States of America*

[3] *Queensland Institute of Medical Research, Brisbane, Australia*

[4] *Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia*

[5] *Department of Epidemiology, University of California, Irvine, United States of America*

Hereditary hemochromatosis [HH] is an inherited disease of dysregulated iron absorption that can lead to total body iron overload with secondary tissue damage in a wide range of organs. Individuals who are homozygous for the C282Y mutation in the *HFE* gene are at increased risk of iron overload and account for 82–90% of clinical cases of HH among those of northern European descent. Many C282Y homozygotes do not, however, develop elevated iron indices suggesting the presence of genetic and environmental modifiers of the HH phenotype. The HealthIron project resequenced exonic regions of eleven candidate genes of iron metabolism in a random sample of 94 C282Y homozygotes and 94 random individuals who are all participants in the Melbourne Collaborative Cohort Study. We investigated the association between SNP genotype and quantitative measures of iron phenotype (log serum ferritin [SF] and transferrin saturation [TS]) using multiple linear regression with individual SNP dosage represented as either a numeric or categorical exposure variable. Haplotypes were inferred statistically using PHASE 2.1, with the haplotype dosage of individuals incorporated into linear regression models. A novel *Cybrd1* SNP was associated with lower serum ferritin in male C282Y homozygotes ($p = .008$) with the same trend in other groups defined by sex and C282Y homozygote status. A *transferrin* SNP (rs1880669) was associated ($p = .004$) with higher transferrin saturation in male non-C282Y homozygotes. A previously reported *transferrin receptor* SNP (rs3817672) associated with increased risk of neoplasm (breast, colorectal and multiple myeloma) when homozygous in association with C282Y homozygosity (Beckman et al., 1999) was associated with increased serum ferritin in male C282Y homozygotes. These associations suggest potential genetic modifiers of iron phenotypes in those genetically predisposed to develop iron overload, and will be the subject of future functional studies.

LINKAGE ANALYSES OF ALCOHOL CONSUMPTION AND DEPENDENCE

Narelle K. Hansell,[1] Katherine I. Morley,[1] John B. Whitfield,[1,3] Penelope A. Lind,[1] Scott D. Gordon,[1] Arpana Agrawal,[2] Michele L. Pergadia,[2] Grant W. Montgomery,[1] Pamela A. F. Madden,[2] Richard D. Todd,[2] Andrew C. Heath,[2] and Nicholas G. Martin[1]

[1] *Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia*

[2] *Departments of Psychiatry, Psychology, and Genetics, School of Medicine, Washington University, United States of America*

[3] *Department of Clinical Biochemistry, Royal Prince Alfred Hospital, Sydney, Australia*

Our aim is to identify quantitative trait loci (QTLs) for alcohol consumption and dependence measures collected in a large community sample. The sample comprises 7268 individuals (4438 females, 2830 males) from 2401 families, with family structures that include twins, non-twin siblings, and parents. Participant age ranges from 18.7 to 91.7 years. Consumption is assessed through measures of frequency and quantity, while dependence is assessed through DSM-IIIR and DSM-IV criteria. Linkage analyses will be run in MERLIN REGRESS. Findings will be discussed in relation to potentially influential genes in regions identified by suggestive and significant linkage peaks and in relation to replication of findings reported in other studies.

GENETIC MODELING AND LINKAGE ANALYSIS OF THE PERCEIVED INTENSITY OF SWEETENERS

Jonathan L. Hansen,[1,2] Danielle R. Reed,[3] Margaret J. Wright,[1] Nicholas G. Martin,[1] and Paul A. S. Breslin[3]

[1] *Queensland Institute of Medical Research, Brisbane, Australia*

[2] *University of Queensland, Brisbane, Australia*

[3] *Monell Chemical Senses Center, Philadelphia, United States of America*

Sweetness is nearly universally liked by most humans, and this liking is unlikely to be learned because it is present at birth. Because studies in mice and cats suggest that the acceptance and intake of sugars is partially determined by genetic factors, we wondered whether similar effects were present in humans. To test this idea, we studied twins and their family members, asking them to rate the perceived taste intensity of four sweeteners, glucose, fructose, neohesperidin dihydrochalcone (NHDC) and aspartame. This genetically informative sample consisted of 392 females and 313 males (mean age of 17.8 ± 3.1 years), including 62 monozygotic and 131 dizygotic twin pairs and 237 sib pairs. Multivariate modeling estimated broad sense heritabilities at 12, 17, 21, and 27% respectively, for perceived intensity measures of glucose, fructose, NHDC and aspartame and indicated that there was only a single common genetic factor, with no genetic factors specific to a particular sweetener. Unique environmental influences for all four compounds were also due to a single factor responsible for 21–34% of phenotypic variation; however the remainder of the phenotypic variance (49-53%) was accounted for by measurement unreliability. A genome-wide linkage scan of 795 microsatellite marker was performed on an average sweet intensity score (heritability = .33, unreliability = .29) in 236 families and suggested a region on chromosome 9 (LOD = 2.74), which is near a region homologous to a sweet preference locus in mice. As we did not reach statistical significance, we sought to tackle the large unreliability in these measures and began testing individuals under supervision in our clinic. Unsurprisingly in this cohort, heritability rose (.21–.39) and unreliability was reduced (.37–.47). The heritability also increased in the average intensity measures to .43 and unreliability reduced to .20. In addition to collecting data from new individuals and retesting some existing participants, 2 subsets have also had either additional microsatellite markers typed or have data from a 100K SNP chip available to build a much stronger linkage marker set for our sample. Linkage analysis is being rerun to hopefully reinforce our suggestive region on chromosome 9.

FUNCTIONAL CONSEQUENCES OF A COMMON AND A NOVEL *CYP17A1* PROMOTER POLYMORPHISM FOR ASSOCIATION STUDIES

Vanessa M. Hayes

Cancer Genetics Group, Cancer Research Program, Garvan Institute of Medical Research, St Vincent's Hospital, Sydney, Australia

Cytochrome P450c17α is a key enzyme in sex hormone production and is encoded by the *CYP17A1* gene. A single nucleotide polymorphism (SNP) located 34-bases upstream of translation initiation, –34T > C (rs743572), has been studied extensively for its potential role, although often controversial, in conferring risk to a large number of hormone (and non-hormone) related diseases and conditions. The majority of these studies are based on the assumption that this polymorphism directly influences gene transcription. With a focus on prostate cancer, numerous studies have reported associations between this SNP and increased/decreased risk of prostate cancer, or no association. In our large European–Australian cases–control study ($n = 1563$) we report a lack of association between this marker and prostate cancer risk (Severi et al., in press) in concordance with a meta-analysis ($n = 5159$) suggesting that this polymorphism is not associated with risk to prostate cancer in European populations (Ntais et al., 2003). The latter study did however suggest a potential influence of this marker in African-based populations. In this study we identify a novel SNP at the –34 position, resulting in a T to A nucleotide change, in 2 African-based populations using denaturing gradient gel electrophoresis. We demonstrate how commonly used genotyping methods, namely restriction fragment length polymorphism analysis and TaqMan allelic discrimination, result in genotype misclassification in the presence of the A-allele. Although the common C-allele and the novel A-allele variants create putative SP-1 and AP-4 transcription factor binding sites, respectively, we show no biological effect of these polymorphic alleles on transcription factor binding or gene expression compared to the wild-type allele. The implication of our findings on association studies of the *CYP17A1* –34T > C SNP are apparent, not only with respect to genotype misclassification in Africans, but also the lack of evidence supporting its role as a functional (direct association) polymorphic marker of disease risk.

GENETIC DETERMINANTS OF MAMMOGRAPHIC DENSITY

Vanessa M. Hayes,[1,2] Jennifer Stone,[3] Lyle C. Gurrin,[3] Graham B. Byrnes,[3] Christopher Schroen,[4] Susan A. Treloar,[5] Emma J. D. Padilla,[1] Gillian S. Dite,[3] Melissa C. Southey,[4] and John L. Hopper[3]

[1] *Garvan Institute of Medical Research, Cancer Research Program, Sydney, Australia*

[2] *Department of Medicine, St Vincent's Hospital Clinical School, University of New South Wales, Sydney, Australia*

[3] *Centre for Molecular, Environmental, Genetic, and Analytic (MEGA) Epidemiology, University of Melbourne, Melbourne, Australia*

[4] *Genetic Epidemiology Laboratory, University of Melbourne, Melbourne, Australia*

[5] *Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Australia*

Mammographic density, the light/white radiographic appearance on a mammogram that represents connective and epithelial tissue, is a strong risk factor for breast cancer which appears to be highly heritable. Little is known about its genetic determinants. We studied 457 women from 207 sisterhoods (104 monozygotic (MZ) twins, 182 dizygotic (DZ) twins, and 171 singletons). Percent mammographic density (PMD), as well as dense area and nondense area, were calculated using a computer-assisted method. We measured 10 single nucleotide polymorphisms (SNPs) from 7 candidate genes (*COMT, HSD3B1, IGFBP3, HER2, VDR, XPD,* and *XRCC3*) and 19 haplotype tagging SNPs from the *IGF-I* gene.

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

Associations between genotypes and mammographic measures were tested: (a) cross-sectionally, using a multivariate normal model fitted using FISHER that allowed separate correlations for MZ, DZ and non-twin pairs; and (b) within sister pairs using paired *t* tests. Weighted regression models examined the association between mammographic measures and the *IGF-I* haplotypes using PROC MIXED in SAS after haplotypes were resolved using tagSNPs. Cross-sectionally, each additional copy of the *HSD3B1* Asn367Thr variant allele was associated with lower PMD (–3.47% per allele, $SE = 1.65$, $p = .035$). Within-pair regression estimates confirmed this association. We also found evidence of associations between 3 polymorphisms in the *VDR* gene and all 3 mammographic measures in within-pair analyses. None of the *IGF-I* tagging SNPs were associated with the mammographic measures but we found evidence of an association between an *IGF-I* haplotype and all 3 measures. We have replicated an association between a variant in the *HSD3B1* gene and PMD which suggests that *HSD3B1* may be an important genetic determinant of mammographic density. We also found that common genetic variation in *IGF-1* is associated with mammographic density. The association between mammographic density and variants in the *VDR* gene merits further investigation.

A REANALYSIS OF 409 EUROPEAN-ANCESTRY AND AFRICAN AMERICAN SCHIZOPHRENIA PEDIGREES REVEALS SIGNIFICANT LINKAGE TO 8P23.3 WITH EVIDENCE OF LOCUS HETEROGENEITY

Elizabeth G. Holliday,[1] Bryan J. Mowry,[1,2] and Dale R. Nyholt[3]

[1] *Queensland Centre for Mental Health Research, The Park, Centre for Mental Health, Wacol, Queensland, Australia*

[2] *Department of Psychiatry, University of Queensland, St Lucia, Queensland, Australia*

[3] *Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Queensland, Australia*

Recent years have highlighted the formidable sample sizes required to detect and replicate complex disease susceptibility loci by linkage analysis. Such evidence has prompted various collaborative efforts for producing enlarged samples. A potential limitation of such pooled samples is that they may consist of subsamples with substantial population genetic differences, including differences in marker and disease allele frequencies. Here, we investigate the impact of population differences in a reanalysis of molecular genetics of schizophrenia 1 (MGS1) affected sibling-pair data. These data comprise 2 samples of distinct ancestral origin: European- (EA; $n = 263$ pedigrees) and African-American (AA; $n = 146$ pedigrees). In an attempt to maximize the linkage information contained within these distinct continental samples, we performed separate linkage analyses of the EA and AA samples, allowing for within-sample heterogeneity, and the combined sample, allowing for both within-sample and between-sample heterogeneity. Significance levels, corrected for the multiple tests, were determined empirically. This approach detected stronger evidence for linkage than any analysis of the combined sample alone, regardless of how heterogeneity was modeled. Most notably, we detected genome-wide significant linkage to 8p23.3. We also detected suggestive linkage to an additional three regions (5p13.3, 7q36.2, and 8p21.3), and observed 6 and 20 regions with nominal $p < .01$ and $p < .05$, respectively (with only 2.4 and 9.1 such peaks expected by chance), suggesting the presence of multiple, small-effect susceptibility loci. Many regions showed pronounced differences in the extent of linkage between the EA and AA samples, which may reflect ancestral differences in the frequency of disease risk alleles. This analysis exemplifies a potentially useful approach for linkage analyses of samples drawn from distinct continental populations.

CASE-CONTROL-FAMILY STUDY OF A COMMON POLYMORPHISM REVEALS THAT 'HIGH-RISK' MUTATIONS MAY PREDISPOSE TO DISEASE: RELEVANCE TO ASSOCIATION STUDIES

John L. Hopper,[1] Vanessa M. Hayes,[2] and Melissa C. Southey[3]

[1] *Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Victoria, Australia*

[2] *Cancer Research Program, Garvan Institute of Medical Research, St Vincent's Hospital, Darlinghurst, Sydney, New South Wales, Sydney, Australia*

[3] *Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Victoria, Australia*

Many studies have assessed whether polymorphisms in steroid hormone metabolism genes are associated with breast cancer risk. We previously measured the *CYP17A1* –34T> C (c.27T > C) promoter polymorphism in a population-based study of 1404 Australian women with breast cancer diagnosed before age 60 years (case probands), 1903 relatives, and 788 controls. Within-family analyses suggested the CC genotype was associated with, on average, a small increased risk. This finding appeared to be influenced by the families of 3 early-onset case probands with multiple affected sisters. *CYP17A1* mutation screening revealed a case proband diagnosed at age 38 years who had a germline protein-truncating mutation (c.775C > T, p.Arg239X), which results in a nonfunctional enzyme and has been reported in a male compound heterozygote with 17 α-hydroxylase deficiency. This mutation was carried by both sisters diagnosed with breast cancer at ages 34 and 42 years, but not by a 57-year-old unaffected sister. It was not found in any of the other tested case probands (48 with multiple-affected relatives and 241 randomly selected) or controls. This study suggested there may be rare mutations in steroid hormone metabolism genes associated with a high dominantly-inherited breast cancer risk. It also demonstrated how 'high-risk susceptibility genes' might be discovered using population-based case-control-family studies. We have since conducted screening for other steroid hormone metabolism genes in a series of 78 early-onset beast cancer cases (i.e., diagnosed before age 40 years) with a strong family history and will present some preliminary findings. This work may have relevance to association studies using unrelated cases and controls that identify multiple highly linked common variants at the one loci associated with risk.

GENOTYPING COPY NUMBER VARIANTS — 1:2 ↑ 2:4

Daryl Irwin,[1,2] Karl Eisler,[2] and David Hawkes[2]

[1] *Sequenom Inc, Brisbane, Australia*

[2] *Australian Genome research Facility, University of Queensland, St Lucia, Brisbane, Australia*

Large-scale copy-number variation (CNV) is an important contributor to conspecific genomic diversity. CNVs have been shown to contribute to phenotype differences in humans and other species. Reliable detection of CNVs remains difficult, being both time-consuming and expensive. Furthermore, detection of CNVs in complex genomes such as *Saccharum officinarum L* (sugar cane $n = 8$–16) is yet to be mastered. We describe techniques to reliably and accurately detect copy number variants using the high-throughput, low cost Sequenom MassARRAY platform. *Saccharum officinarum L* (sugar cane) is a genetically diverse species with a complex genome containing 8-16 copies of the haploid content originating from numerous primordial genomes. Mapping and quantifying copy number of each of the primordial genomes is yet to be mastered. To perform this analysis, polymorphisms must be identified in the genes of interest. These polymorphisms may be in the form of an allele that identifies a single primordial copy or a group of copies that when compared to other groupings the presence of a primordial copy can be delineated and quantitated. Quantitative genotyping techniques have been employed and shown to detect the minimum allele ratio of 1:15 (5.7%) required for gene copy mapping. In addition, variation has been shown to be < 3% providing reliable quantitation and detection of 1:15 Vs 2:15 allelic ratios in a gene duplication event. A further challenge when analyzing CNVs is converting allelic ratios into absolute copy number values. We have overcome this by incorporating an oligonucleotide in the reaction that is the same sequence as the amplicon with an alternate allele at the SNP site. This 'spike' is of known copy number and as it essentially performs the same as the DNA of interest can be compared to determine absolute copy number of the alleles in the reaction. When this is multiplexed with a loci of known copy number, the number of genomic DNA copies can be calculated, facilitating calculation of the absolute numbers of copies of the CNV. Utilizing the quantitative capabilities of the Sequenom MassArray iPlex platform, in conjunction with known oligo spikes, we have developed tools that can efficiently and cost effectively detect and quantitated CNVs in any genome.

ASSOCIATION OF CHROMOSOME 10q26 GENES IN AGE-RELATED MACULAR DEGENERATION (AMD)

F. M. Amirul Islam, Andrea J. Richardson, Robyn H. Guymer, and Paul N. Baird

Centre for Eye Research Australia, University of Melbourne, Australia

Age-related macular degeneration (AMD) is the leading cause of blindness in the developed world. A non-synonymous SNP at amino acid 69 (A69S) in the hypothetical LOC387715 gene on chromosome 10 has been identified as a major susceptibility allele for AMD. Additional SNPs from this and adjacent genes have also been implicated in AMD, although these findings have been inconsistent. We undertook a SNP-tagging analysis of 3 adjacent genes (PLEKHA1, LOC387715 and PRSS11) on chromosome 10q26 to better investigate the involvement of these genes in AMD. 577 individuals with AMD and 173 ethnically matched controls were available for this study. A tag-single nucleotide polymorphism (t-SNP) approach was used to investigate these genes. t-SNPs were selected to encompass the entire PLEKHA1/LOC387715/PRSS11 region, including the promoter region of the PLEKHA1 gene and the 3'UTR region of the PRSS11 gene. 25 t-SNPs with an LD tagging criteria of $r_2 > .8$ and a minor allele frequency (MAF) of at least .1 were selected based on the International HapMap project. UNPHASED software was used for haplotype analysis purposes. All SNPs were in Hardy-Weinberg equilibrium. Four major blocks were determined based on linkage disequilibrium (LD) statistics and *r*2. After Bonferroni correction, 9 individual SNPs were found to be significantly associated with AMD (p value ranging from .001 to 1.510_{-19}). Based on the log likelihood ratio statistics (LRS) and odds ratio

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

(OR), 4 SNPs were selected for genotype and haplotype analysis. The CC genotype of SNP 2 was found to be associated with AMD (OR 17.70, 95% CI 6.91, 45.35) compared with the GG genotype. A two-SNP haplotype (G-G of SNPs 1 and 4) was associated with the highest risk of disease, OR 4.34, compared to the T-C haplotype; a 3 SNP haplotype (G-T-G of SNPs 1, 3 and 4) had an OR 4.16 compared to T-A-C haplotype, while the haplotype G-G-T-G had an OR, of 3.95 Our work confirms previous studies that the LOC387715 gene is a strong susceptibility gene for AMD. It also indicates that multiple SNPs, either individually or in varying haplotype combinations are significantly associated with AMD from both the LOC387715 and PRSS11genes.

A GENOME-WIDE SCAN FOR DIABETES IN MAURITIUS

J. Jowett,[1,2] K. Shields,[1] J. E. Curran,[3] M. P. Johnson,[3] H. H. H. Goring,[3] T. D. Dyer,[3] S. Kotea,[4] S. Kowlessur,[5] P. Chitson,[5] J.W. MacCluer,[3] E. K. Moses,[3] P. Zimmet,[1] and J. Blangero,[2,3]

[1] *International Diabetes Institute, Australia*

[2] *Human Gene Discovery Program, ChemGenex Pharmaceuticals, Australia*

[3] *Southwest Foundation for Biomedical Research, United States of America*

[4] *University of Mauritius, Mauritius*

[5] *Ministry of Health, Mauritius*

The prevalence of Type 2 diabetes and obesity continues to increase in both developed and developing countries presenting major public health issues impacting a wide variety of social and economic measures. Substantial evidence supports a major role of genetics in development of the disease together with modifying effects of the environment. To improve our understanding of the molecular mechanisms underpinning the metabolic abnormalities associated with disease development, we have conducted a genome-wide scan to identify chromosomal regions likely to harbor disease-influencing genes. Using a study design based on large multigenerational pedigrees to improve statistical power of susceptibility gene localisation, we identified a region on chromosome 12q that influences fasting glucose. In addition to ranking candidate disease genes by their known function for further evaluation, we have incorporated differential gene expression data obtained from a large scale genome-wide expression profiling study. As part of the San Antonio Family Heart Study, genome-wide quantitative transcriptional profiles were obtained from 1240 individuals. Using carefully collected lymphocyte samples from the initial examination period approximately 15 years ago, high quality RNA was recovered and profiled for 47,289 transcripts using the Sentrix Human-6 expression BeadChip supplied by Illumina. Expression profiles that correlated with fasting glucose were used as an additional means for ranking candidate genes in the region for ongoing genetic variation analysis.

AN IMPROVED MODEL FOR CALCULATION OF IDENTITY BY DESCENT PROBABILITIES FOR DENSE SNP DATA

Jonathan M. Keith,[1] Kerrie Mengersen,[1] and Peter M.Visscher[2]

[1] *School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia*

[2] *Queensland Institute of Medical Research, Royal Brisbane Hospital, Brisbane, Australia*

Various strategies for mapping disease genes involve estimating the probabilities that a pair of relatives share 0, 1 or 2 alleles Identical By Descent (IBD) at each of a number of marker loci. IBD probability estimates are also required for certain family-based tests of linkage and linkage disequilibrium (LD), such as those implemented in the QTDT package (Abecasis et al., 2000). The state-of-the-art software for evaluating IBD probabilities in small pedigrees is the popular Merlin package (Abecasis et al., 2002). A recently added feature of Merlin is the ability to account for LD between markers in the parental population (Abecasis & Wigginton, 2005). This is an important development, since information about parental haplotypes can be very helpful in resolving uncertainties about the number of IBD alleles at each locus and ignoring LD can create bias in IBD estimation. Moreover, the current move towards denser linkage maps entails that LD between adjacent markers is also increasing. The algorithm implemented in Merlin involves identifying clusters of markers that represent haplotype blocks and estimating the population frequency of each haplotype in each cluster. However, the method has a number of drawbacks: (i) it relies on knowing or estimating a large number of population parameters, (ii) it assumes no recombination within a cluster and no LD between clusters and (iii) it does not scale up to whole-genome, dense-marker data. Here we present evidence showing that accounting for LD can substantially improve the estimation of IBD probabilities. We also describe a new model and method for estimating IBD probabilities that addresses these three drawbacks, by incorporating a Markov model for the parental haplotypes. In particular, the number of parameters scales linearly with the number of markers, and recombination and LD between any pair of adjacent markers is permitted. An additional advantage of the new model is that it allows for and detects likely genotyping errors. We present results using real and simulated data to demonstrate the improved speed and accuracy of the new method compared to Merlin.

A MOLECULAR MODEL FOR INTERACTION BETWEEN GLYCOGEN SYNTHASE KINASE-3β AND TAU GENES IN NEURODEGENERATIVE DISEASES

John B. J. Kwok,[1,2] Clement Loy,[1,2] Marianne Hallupp,[2] Simon Lovestone,[3] and Peter R. Schofield[1,2]

[1] *Prince of Wales Medical Research Institute, Sydney, Australia*

[2] *University of New South Wales, Sydney, Australia*

[3] *Departments of Old Age, Psychiatry, and Neuroscience, Institute of Psychiatry, London, United Kingdom*

Human Glycogen synthase kinase-3beta (GSK-3beta) is a serine/threonine kinase that is involved in numerous physiological responses, and phosphorylates various nuclear and cytoplasmic proteins, including the microtubule-associated protein Tau (MAPT) and beta-catenin. Hyperphosphorylated Tau is deposited in various neurodegenerative diseases including Alzheimer's disease and frontotemporal dementia. Beta-cateinin forms part of the Wnt-signaling pathway. A promoter polymorphism (rs334558) within the GSK-3? gene (*GSK3B*) is associated with transcriptional strength, and an intronic polymorphism (rs6438552) regulates selection of splice sites in vitro. There are significant differences in transcriptional efficiency between the 2 *MAPT* haplotypes (H1 and H2) as measured by luciferase assays and in vivo measurement of brain samples. We have recently shown that the polymorphisms in *GSK3B* interact with the *MAPT* haplotypes to increase the risk of idiopathic Parkinson's disease (Kwok et al., 2005). We observed a similar interaction between *GSK3B* and *MAPT* in 2 late-onset AD cohorts, with the T allele of rs6438552 SNP significantly increasing the risk of AD in individuals with at least one H2 haplotype (odds ratio = 1.5 to 1.8; p = .03–.04). From our functional analyses of *GSK3B* and *MAPT* haplotypes and our association studies, we propose a molecular model of interaction based on discordant levels of *GSK3B* and *MAPT* gene expression. In individuals with different levels of the 2 gene products, that is high levels of GSK-3? and low Tau levels, or *vice versa*, there is increased risk of disease. A corollary of this observation is that excess Tau protein should be able to sequester GSK-3β and shunt it towards cellular pools that have functional consequences on other substrates of GSK-3β. We verified this model by measuring the levels of beta-catenin signaling in cell lines transfected with *MAPT* and *GSK3B* cDNAs. Over-expression of either *MAPT* or *GSK3B* resulted in decreased ?-catenin signaling. Conversely, co-transfection of both of these molecules reversed the decrease in β-catenin signaling, as predicted by our model. Our genetic and biochemical analyses have identified a novel interaction between Tau, GSK-3β and β-catenin, and can have important etiological implications for multiple neurodegenerative disorders.

DCDC2 IS ASSOCIATED WITH NORMAL VARIATION IN READING AND SPELLING ABILITY IN A LARGE AUSTRALIAN POPULATION SAMPLE

Penelope A. Lind,[1] Michelle Luciano,[1] Margaret J.Wright,[1] Grant W. Montgomery,[1] Nicholas G. Martin,[1] and Timothy C. Bates[2]

[1] *Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia*

[2] *Department of Psychology, University of Edinburgh, United Kingdom*

The 6p21-p22 chromosomal region has been consistently and independently identified as a reading disability (developmental dyslexia) locus in a number of linkage studies (Fisher and DeFries, 2002). Candidate genes within this region include *KIAA0319* and *DCDC2*. Recent association studies (Meng et al., 2005; Schumacher et al., 2005) have implicated *DCDC2* as a susceptibility gene for developmental dyslexia. The present study attempted to replicate the association by typing 22 SNPs across *DCDC2* gene in 783 unselected families of twins aged 11 to 25 years old: 144 monozygotic (MZ) twin pairs (including 90 families with additional siblings), 95 MZ twin singletons (42 with siblings), 446 dizygotic (DZ) twins pairs (164 with siblings), 82 DZ singletons (25 with siblings), and 16 families with sibling data only. Twins and siblings (n = 1773) completed 6 tests of reading and spelling (regular, irregular and non-word stimuli) ability, and a single principal component was extracted from the 6 scores for analysis. Single marker analysis was performed within the QTDT program, using regression based estimation while controlling for sex and nonverbal IQ, and assuming an additive model. We found no association with either the 2 *DCDC2* SNPs reported by Meng et al. (2005), or the at-risk rs793862-rs807701 haplotype SNPs reported by Schumacher et al. (2005). We did however, find significant support for association (p < .05) at 3 alternative *DCDC2* SNPs, with the strongest association at rs1419228 (p = .003). The C allele of rs1419228 SNP conferred a lower reading score (size of effect –0.145), as did the G alleles of rs9467076 and rs2262442 (effect size –0.117). Further haplotype analysis will be discussed. Although we have not replicated previous association findings, we have shown that *DCDC2* genetic variation does influence variation in reading and spelling ability in an unselected Australian population.

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

A NOVEL FRAMEWORK FOR 3D VISUALIZATION AND ANALYSIS OF FAMILIAL DATA

A. M. Loh,[1] S. Wiltshire,[1] K.W. Carter,[1] J. Emery,[3] and L. J. Palmer[1,2]

[1] *Centre for Genetic Epidemiology, Western Australian Institute for Medical Research, University of Western Australia, Perth, Australia*

[2] *School of Population Health, University of Western Australia, Perth, Australia*

[3] *School of Primary, Aboriginal and Rural Health Care, Faculty of Medicine, Dentistry and Health Sciences, University of Western Australia, Perth, Australia*

With the ever-growing size of genealogical and family-based datasets around the world, there is an increasing need for software tools to analyze, manage and record the data. Graphical techniques are becoming essential for analyzing large amounts of data, to improve upon the limitations of traditional text-based techniques, particularly in representing complex data interrelationships. A graphical presentation of familial data (e.g. phenotypic, genotypic or demographic) and pedigree structures can show important information in a far less cluttered fashion, enabling rapid analysis and interpretation of results. Current 2D techniques for visualizing pedigrees are limited in the number of individuals that can be displayed simultaneously, particularly for large pedigrees, due to the exponential expansion of generations. 3D pedigree visualization is superior to 2D in many ways. 3D techniques can display many more individuals in a given area. Closely related individuals are more easily displayed near each other, allowing trends within families to be clearly seen. In addition it is possible to display families or individuals of interest in the context of the entire population. To our knowledge, no tool exists to visualize large pedigrees in 3D. However, recent advances in consumer graphics technology allow such a tool to be created and used on most personal computers. In this project, software was created as a proof of concept for 3D pedigree visualization techniques. In its current implementation, the software is capable of importing data in the standard Linkage format, and displaying all pedigrees simultaneously. The software was used to display and navigate through all known families from the Busselton Health Study, along with phenotypic data. This working proof of concept is the basis for a collaborative application freely available to researchers all over the world.

TESTING REPLICATION OF A 5-SNP SET FOR GENERAL COGNITIVE ABILITY IN SIX POPULATION SAMPLES

Michelle Luciano,[1] Penelope A. Lind,[1] Ian J. Deary,[2] Anthony Payton,[3] Danielle Posthuma,[4] Sarah Harris,[2] Lee M. Butcher,[5] Tinca J. C. Polderman,[4] Peter M. Visscher,[1] Timothy C. Bates,[2] Eco J. C. de Geus,[4] Margaret J. Wright,[1] Dorret I. Boomsma,[4] Grant W. Montgomery,[1] Nicholas G. Martin,[1] and Robert Plomin[5]

[1] *Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia*

[2] *Department of Psychology, University of Edinburgh, United Kingdom*

[3] *Centre for Integrated Genomic Medical Research, University of Manchester, United Kingdom*

[4] *Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands*

[5] *Institute of Psychiatry, King's College, London, United Kingdom*

A 5 SNP set has been associated with general cognitive ability in children from the Twins Early Development Study (TEDS;Harlaar et al., 2005). Four of these SNPs were identified through a 10K microarray analysis (Butcher, Meaburn, Knight et al., 2005) while the other SNP was identified through a targeted analysis of brain expressed genes (Butcher, Meaburn, Dale et al., 2005). The present study aimed to replicate the association of this SNP set with general cognitive ability in 6 population samples from Australia (adolescents), the UK (2 Scottish cohorts from Aberdeen and Lothian both measured in childhood and in old age, and an old aged cohort from Manchester), and the Netherlands (a cohort of children and an adult cohort). Allele frequencies for each of the SNPs were comparable across populations and there were no major deviations from Hardy-Weinberg equilibrium for any of the SNPs. Results for the SNP set were significant ($p < .05$) in the 2 Scottish cohorts, although the direction of the effect differed between cohorts. In the Aberdeen cohort (in old age) the effect was positive ($r = .10$) and consistent with the initial finding, but in the Lothian cohort (in childhood) the effect was negative ($r = -.11$). Association tests of the individual SNPs supported replication ($p < .001$) of one SNP (rs99168492) in a single cohort only (i.e., the Aberdeen cohort demonstrating the positive SNP set association). Two other significant associations (rs4128492, rs2382591) found in the Manchester cohort were in the opposite direction to that expected. A test in which the SNP set effect was constrained equal across cohorts was nonsignificant. Our results do not support replication of a 5-SNP set for general cognitive ability in 6 independent samples.

PREDICTION OF QUANTITATIVE PHENOTYPES FROM GENOME-WIDE ASSOCIATION STUDIES

Paula E. Macrossan,[1] Allan F. McRae,[2] Kerrie M. Mengersen,[1] and Peter M. Visscher[2]

[1] *Queensland University of Technology, Brisbane, Australia*

[2] *Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia*

Since the advent of genetic mapping in the 1980s, medical genetics has focused on the genetic mapping of 'Mendelian' diseases, or those controlled by one or a small number of genes. However, most diseases and traits are influenced by multiple genes, each segregating according to Mendel's laws, with the addition of varying degrees of deviation attributable to environmental effects. Such 'quantitative' traits display a continuous distribution of phenotypes, compared with the discrete phenotypes seen in Mendelian traits. Recent advances in affordable high-volume genotyping technology have led to a growing interest in genome-wide association studies (GWAS) for the mapping of such continuous traits. The first success stories using GWAS have been published, and the emerging evidence points to many loci, each with a small effect. The implication is that using individual loci to predict risk to disease or a quantitative phenotype is inaccurate and inefficient. In this study we investigate methods that use multiple loci to give unbiased prediction of a quantitative phenotype. We explore cross-validation methods and iterative regression methods to identify a subset of SNPs that predict phenotypes, and quantify the accuracy and bias of the prediction. The methods are applied to a dataset of 170 families consisting of twins and siblings with a measurement of height who were genotyped with the Affymetrix 100k SNP chip.

HIGHLY COST EFFICIENT GENOME WIDE ASSOCIATION STUDIES USING DNA POOLS AND DENSE SNP ARRAYS

Stuart Macgregor, Zhen Zhen Zhao, Anjali Henders, Nicholas G. Martin, Grant W. Montgomery, and Peter M. Visscher

Genetic Epidemiology and Molecular Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia

Recent advances in large scale genotyping have made genome-wide association (GWA) possible. GWA is one of the primary tools for the identification of loci contributing to susceptibility to complex common human disease, with several applications of GWA published in *Nature* and *Science* recently. However, the major limiting factor in many GWA studies is cost. Individually genotyping GWA samples is often prohibitively expensive, with genome scans of suitable size (hundreds/thousands of cases and controls, hundreds of thousands of markers) typically costing over US$1 million. Alternative approaches that reduce the genotyping cost are therefore highly desirable. We will demonstrate that DNA pooling offers a means of dramatically reducing the cost of GWA studies. Building on previous work on Affymetrix arrays, new methodology will be outlined for statistical analysis of data from the Illumina platform, including a novel quality control metric. The method is based upon contrasting case and control pools, and hence does not require independent estimates of rates of unequal amplification of alleles. Illumina and Affymetrix arrays were applied to the same pools; Illumina arrays were found to offer an order of magnitude decrease in pooling error variance compared with Affymetrix arrays. With Illumina arrays, concordance with individual genotyping data is excellent; in terms of effective sample size it is possible to extract more than 80% of the information available with individual genotyping. Guidance will be given on best study design for pooling based GWA studies. It will be shown that even after taking into account pooling error, one stage scans can be performed for more than 100-fold reduced cost compared with individual genotyping. With appropriately designed 2-stage studies, individual genotyping can provide confirmation of pooling results whilst still providing ~ 20-fold reduction in total cost compared with individual genotyping based alternatives. The large cost savings with Illumina based pooling imply that future studies need only be limited by the availability of samples and not cost of arrays.

GENOME-WIDE LINKAGE AND ASSOCIATION SCANS FOR MELANOMA RISK FACTORS

Nick Martin,[1] David Duffy,[1] Gu Zhu,[1] Grant Montgomery,[1] Nick Hayward,[1] Allan McRae,[1] and Rick Sturm[2]

[1] *Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia*

[2] *Institute of Molecular Bioscience, University of Queensland, Brisbane, Australia*

Moliness, skin, hair and eye color, freckling, and ability to tan are easily measured risk factors for melanoma. We have collected these measures on nearly 800 adolescent twin families for whom we have DNA samples. Microsatellite based genome-wide linkage scans have been performed on 630 families, and Affymetrix 100k SNP scans on 170 families. Linkage scans will be presented for all 800 families, and genome-wide association scans for 170 families. Early analyses of this new expanded dataset support some, but not all, of the previously reported linkages for moliness on 2p, 8q, 9p, 9q and 17p. The GWAS for moliness shows a larger number of hits in the 9p region than expected by chance, and one very large peak on another chromosome that is being followed up. The hits in the 9p region are in the same small area as recently reported associations for T2D and myocardial infarction. Haplotype analysis will tell if the same or different SNPs or haplotypes are implicated. A 3-SNP haplotype in intron 1 of OCA2 has been reported by us as a good predictor of blue eye color, and we are now fine mapping in this region to try and find the causal SNP[s].

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

A SIMPLE MODEL TO EXPLAIN THE AGE-DEPENDENCY OF CANCER SUSCEPTIBILITY GENES AND ENVIRONMENTAL EFFECTS

John D. Mathews, James Dowty, Mark Jenkins, Graham Byrnes, and John L. Hopper

Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, University of Melbourne, Melbourne Australia

For major cancer genes such as BRCA1, the cancer incidence in carriers relative to the incidence in noncarriers (known as the hazard ratio — HR) falls progressively with age. Likewise, the excess relative risk (ERR) for ionising radiation, measuring the same trend as HR, also falls with time since exposure. Both effects can be explained by a simple model for genetic-environmental interactions. The essential assumption is that the rate of disease initiation is determined by multiplicative interactions of measured factors (genes or environmental exposures such as radiation) with unmeasured factors. It then follows that for persons with measured factors, the cases with the earliest ages of onset will be those at highest risk because of the extra effects of unmeasured factors. It also follows that the cohort of unaffected persons still at risk because of a measured factor will become depleted for unmeasured risk factors more quickly than those without the measured gene or environmental exposure, causing the HR or ERR to fall with age/time since exposure. This simple model has now been developed in mathematical detail, and fitted to published data on the age dependency of hazard ratios for BRCA1, BRCA2, FPC. The same model, fitted to published data on the cancer incidence of children exposed in the Japanese A-bomb cohorts, provides strong circumstantial evidence for the role of (unmeasured) genes and environmental effects in modifying the carcinogenic effects of the estimated doses of ionising radiation. From a practical viewpoint, we suggest that in searching for effects of putative susceptibility genes, it will be useful to look for effects which are age-dependent, as well as for effects that are sustained throughout the entire age-span. Because of the competing effects of different genes and environmental factors over the age-span, the average lifetime effect for many genes could be close to zero, even though there might be strong statistical evidence for age-dependence, i.e., for major effects at particular ages. The model also provides a strong rationale for using the mean age of onset, or years of (healthy) life lost, as an ancillary measure of effect in genetic and epidemiological studies.

POWER AND SNP TAGGING IN WHOLE MITOCHONDRIAL GENOME ASSOCIATION STUDIES

Allan F. McRae, Enda M. Byrne, Zhen Zhen Zhao, Grant W. Montgomery, and Peter M. Visscher

Queensland Institute of Medical Research, Brisbane, Australia

The lack of coverage of the mitochondrial genome in both the Illumina 300K and Affymetrix 500K SNP chips has resulted in an omission in the current generation of genome-wide association studies. Given the many implications of mitochondrial variants in disease, a mitochondrial genome-wide association study provides an important addition to nuclear-genome association studies at relatively little cost. Using a collection of mitochondrial sequences of European origin, a panel of 69 SNPs was chosen to tag all SNPs with a minor allele frequency greater than 1% at an $r2$ of .8. The validity of the European sequences as a tagging population was demonstrated through the genotyping of a sample of 3839 individuals from 1037 Australian families of European descent. The allele frequencies and inter-marker linkage disequilibrium within the 1693 independent mitochondrial haplotypes in the Australian sample showed a strong concordance to those in the collection of European mitochondrial sequences. The ability of the tagging set to predict the mitochondrial haplogroup of a sample is also demonstrated. Power estimates are derived for both disease (case-control) and quantitative trait mapping studies. The power when testing all common mitochondrial SNPs is shown to be equivalent to that when testing only tagging SNPs, despite the relatively high ratio of tagging SNPs to total SNPs resulting from the tagging of all SNPs with a minor allele frequency greater than 1%. The sample size requirements of mitochondrial genome association studies are compared to that of nuclear whole genome studies. Remarkably, the trade off between the number of tests being performed, and the proportion of phenotypic variance explained for a fixed additive effect, results in approximately equal sample sizes required for both study types, although the per individual cost for the mitochondrial association study is much less.

UNCOVERING COMMON GENETIC RISK FACTORS FOR PARKINSON'S DISEASE

Greg T. Sutherland and George D. Mellick

Eskitis Institute for Cell and Molecular Therapies, Griffith University, Nathan, Queensland, Australia (on behalf of the Australian Parkinson's Project Team)

Mutations in the so-called PARK genes lead to rare familial forms of Parkinson's disease (PD). However the extent to which common genetic variability around these genes alters risk for common PD remains unclear. The Australian Parkinson's Project is analyzing genetic variability around the PARK loci in a large PD case-control sample recruited from 3 Australian states. The emphasis is on gene-gene and gene-environment interactions between commonly occurring variables. Our aim was to report on a pilot PD association analysis of 87 polymorphisms around13 PARK genes in an initial case-control sample recruited during 2006. PD cases (n = 3356) and unaffected control subjects (n = 298) of white European ancestry were recruited from 3 specialist clinics in Brisbane and the Australian Electoral Roll. Common genetic variables (86 SNPs genotyped on the TaqMan platform, and 1 STR variable genotyped using standard methods) were assessed in all subjects. Haplotypes around the PARK genes were inferred by implementing the expectation-maximization algorithm using the program EH. Other statistical analyses were performed using SPSS and SNPStat. Consistent with previous studies, PD cases were more likely than controls to report a family history of PD, high levels of pesticide exposure, and lower frequency of cigarette smoking. The frequencies of all polymorphisms were comparable to those reported in other samples of European ancestry, although several putative haplotype tagging SNPs were highly correlated (r squared greater than 0.9) in our sample. Weak associations with PD were identified for variables around the alpha-synuclein, parkin, PINK1 and APOE genes at levels insensitive to identification by approaches such as whole-genome association analysis. Our initial analysis supports the validity of the recruitment strategy and experimental design being employed in this project, and confirms that the individual influence of common genetic variability around the PARK genes to PD risk is modest at best. Initial analyses of interactive effects on PD outcome are now underway.

AN IMPORTANCE SAMPLING APPROACH TO SENSITIVITY ANALYSIS OF PRIORS IN BAYESIAN ANALYSIS OF DNA SEQUENCE SEGMENTATION

Kerrie Mengersen,[1] Darfiana Nur,[2] David Allingham,[2] and Judith Rousseau[3]

[1] *School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia*

[2] *School of Mathematical and Physical Sciences, University of Newcastle, Australia*

[3] *CEREMADE, Universite Dauphine, Paris, France*

Driven by the vast amount of DNA sequence data that are currently available for analysis, there is an increasing need to develop corresponding efficient computational and statistical tools. An important part of any statistical analysis is an assessment of the sensitivity of the inferential statements to assumptions made in the modeling process. For example, one aspect of Bayesian analysis to which this applies is the choice of prior distributions for the unknown parameters. However, a thorough investigation of this one issue is daunting if the analysis is complex and time-consuming. For focus, we consider in this paper the problem of identifying homogeneous DNA segments, and examine a Bayesian approach based on a hidden Markov model, which is essentially a mixture model with Markov dependent component indicators. The parameters in this model are the base/nucleotides transition probabilities for the segment types and the transition matrix of segment types. We propose an importance sampling approach to sensitivity analysis of the priors imposed on these parameters, which dramatically improves computational time and therefore increases the ability for comprehensive assessment. The approach is applied to the segmentation of the benchmarking DNA sequences such as intron-7 of the chimpanzee and human alpha-fetoprotein gene. Our results show that inferences based on a Bayesian hidden Markov model for segmentation are indeed sensitive to prior specification. It is therefore recommended that a thorough sensitivity analysis be included routinely in analyses of complex models such as these. Further research is required to formalize this need, and to develop tools that enable this to be undertaken feasibly with respect to time, computation and coverage.

PREDICTION OF FRACTURE RISK BY LRP5 GENE POLYMORPHISMS: A HAPLOTYPE ANALYSIS IN A POPULATION-BASED STUDY

Nguyen D. Nguyen, Bich H. Tran, Jacqueline R. Center, John A. Eisman, and Tuan V. Nguyen

Bone and Mineral Research Program, Garvan Institute of Medical Research, Sydney, Australia

Polymorphisms of the low density lipoprotein receptor related protein 5 (LRP5) gene have been shown to be linked to or associated with bone mineral density (BMD), but their association with fracture risk is not known. The present study sought to assess the association between LRP5 gene polymorphisms and fracture risk in women by using haplotype analysis. Haplotype tagging single nucleotide polymorphims (in order rs314776, rs3736228, rs4988320, rs4988321, rs556442, rs587808 and rs660925) of LRP5 gene was determined in a sample of 1286 (821 women) participants from a population-based cohort. Bone mineral density (BMD) was measured at baseline. During the follow-up period of 1989-2006, 125 men and 338 women sustained at least one fracture. There were no statistically significant associations between any single

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

SNP and BMD or fracture risk. However, haplotype analysis showed that 3 haplotypes (CCAGAAGC [3% in the population], CCGGAGA [3%] and TCGGGAG [2%]) were associated with variation in BMD. Furthermore, carriers of haplotype CTGGGGA (6% in the population) had an increased risk of fracture (RR: 1.4, 95% CI: 1.1–3.0), hip fracture (2.2, 1.0–4.9) and wrist fracture (2.4, 1.1–4.9), after adjusting for age and BMD. There was no significant association between any haplotype and BMD or fracture risk in men. The proportion of fracture cases that is attributable to the haplotype was 2.3%. These results suggest that common genetic variation at the LRP5 gene was modestly associated with fracture risk in women independent of its association with BMD. The present study also suggests that a haplotype-based analysis is more powerful than a SNP-based analysis in the detection of association.

THE UTILITY OF FAMILY-BASED STUDIES IN THE ERA OF (POTENTIAL) GENOME-WIDE ASSOCIATION STUDIES

Dale R. Nyholt

Queensland Institute of Medical Research, Brisbane, Australia

With the availability of high-throughput cost effective genotyping platforms and dense maps of anonymous and frequent single nucleotide polymorphisms (SNPs) spanning the whole human genome, genome-wide association (linkage disequilibrium, LD) studies (GWAS, also termed WGAS) are now feasible. However, GWAS still face major challenges, such as collecting huge samples with minimal inter-population variability, and without compromising on the quality of phenotypes. Furthermore, linkage data contain considerable additional information capable of aiding the discrimination of causal variation from the nearby variation that is merely in linkage disequilibrium (LD). Consequently, linkage studies may still pay dividends even if GWAS are as successful as many of us hope, and efforts to recruit and study high quality family data should continue to take high priority in our efforts to dissect the genetic components of complex human diseases, with linkage and association studies being considered as complementary strategies. Finally, novel approaches utilizing family data that offer potential to increase the efficiency of association studies will be briefly discussed.

COMMON GENETIC INFLUENCES UNDERLIE COMORBIDITY OF MIGRAINE AND ENDOMETRIOSIS

Dale R. Nyholt,[1] Nathan G. Gillespie,[2] Kathleen R. Merikangas,[3] Grant W. Montgomery,[1] Susan A. Treloar,[1] and Nicholas G. Martin[1]

[1] *Queensland Institute of Medical Research, Brisbane, Australia*

[2] *Virginia Institute of Psychiatric and Behavioral Genetics, Richmond, Virginia, United States of America*

[3] *Section on Developmental Genetic Epidemiology, Intramural Research Program, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland, United States of America*

Migraine and endometriosis are highly prevalent, debilitating and painful disorders that affect women during their most productive years. Two recent studies — one within a gynecological population, the other within a migraine population — reported the coexistence of these conditions within the same individual (comorbidity). To further investigate migraine and endometriosis comorbidity, we examined the co-occurrence of migraine and endometriosis within the largest known collection of families containing multiple women with surgically confirmed endometriosis, and in an independent sample of 815 monozygotic and 457 dizygotic female twin pairs, for whom endometriosis status and migraine symptom data were available. Within the endometriosis families, a significantly increased risk of migrainous headache was observed in women with endometriosis compared to women without endometriosis ($p = .009$), with an odds ratio of 1.57. Bivariate heritability analyses indicated no evidence for common environmental factors influencing either migraine or endometriosis, but significant genetic components for both migrainous headache and endometriosis, with heritability estimates of 69% and 49%, respectively. Importantly, a significant additive genetic correlation ($r_G = .27$) and bivariate heritability (17%) were observed between migraine and endometriosis. Controlling for the personality trait Neuroticism made little impact on this association. These results confirm the previously reported comorbidity between migraine and endometriosis, and indicate common genetic influences completely explain their cooccurrence within individuals. Therefore, to enhance the physical health and emotional well-being of many women — via improved treatment and pain relief — the presence of migraine should always be investigated in women with endometriosis and vice versa.

ANALYSIS OF GENOME REGIONS LINKED TO AGE-AT-APPENDECTOMY IN AUSTRALIAN TWINS

Chris Oldmeadow,[1] Ian Wood,[1] Kerrie Mengersen,[1] Peter Visscher,[2] and David Duffy[2]

[1] *Queensland University of Technology, Brisbane, Australia*

[2] *Queensland Institute of Medical Research, Brisbane, Australia*

It has been shown that there are significant common environmental and genetic contributions to the onset of appendectomy, although no account for variation in the age of onset were made. Utilizing questionnaire data, estimates of genetic and common environmental contributions to the age at appendectomy in 3808 Australian twin pairs were calculated. Microsatellite data has been incorporated with the survey data to facilitate a model-free linkage analysis. One method of analyzing dependent censored age-at-onset data arising from a twin study involves the introduction of a bivariate Gaussian distributed random effect on the log scale to capture the dependencies between related individuals. When interest rests in calculating the components of variance of this random effect, such as in linkage analysis, estimation procedures involve finding approximations to the marginal likelihood using methods, such as the penalized partial likelihood (PPL), or simulating the marginal likelihood (or posterior distribution for a Bayesian analysis) using Markov Chain Monte Carlo (MCMC). Unfortunately, approximation methods may be unreliable and MCMC simulation can be computationally demanding for analysis over an entire genome. Here, results from a typical PPL analysis were used as a guide to direct attention to particular areas of interest in the genome. MCMC simulation was then performed on these areas of interest.

THE FIRST LINKAGE STUDY TO IDENTIFY GENES UNDERLYING INTRAHEPATIC CHOLESTASIS OF PREGNANCY IDENTIFIES A NOVEL SUSCEPTIBILITY LOCUS

Jodie N. Painter,[1] Anne Ropponen,[2] Miia Savander,[1] and Kristiina Aittomäki[1,3]

[1] *Folkhalsän Institute of Genetics, University of Helsinki, Helsinki, Finland*

[2] *Department of Obstetrics and Gynaecology, Helsinki University Central Hospital, Helsinki, Finland*

[3] *Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland*

Intrahepatic cholestasis of pregnancy (ICP) is the most common pregnancy-related liver disease worldwide. Symptoms generally appear in the third trimester, and include generalized mild to intense itching and increased levels of serum bile acids (≥ 10 μmol/L). The condition is usually relatively benign for the mother, as the cholestasis typically resolves following delivery, leaving no apparent liver damage. However, ICP entails a small but substantial risk to the fetus, with increased rates of premature birth, intrauterine distress, and in rare cases, fetal death. The precise etiology is currently unknown in the majority of cases, but it appears to be a genetically heterogeneous complex disease resulting from the inability of the liver to metabolise increasing levels of placentally derived hormones, particularly estrogen and progesterone, as pregnancy progresses. We have performed 2 genome-wide scans for linkage for ICP in Finland, in the expectation that families from a small, relatively genetically homogeneous population would share the same underlying genetic susceptibility to cholestasis. For the first scan, 8 families were genotyped at over 360 microsatellites, with an average inter-marker distance of 10 cM, resulting in 8 peaks, with NPL scores indicative of suggestive linkage. All but one of these peaks were due to linkage signal from only one of the 4 largest families. As no mutations were found in potential candidate genes, we performed a second scan on the 4 largest families with Affymetrix 50K *XbaI* SNP chips. This scan resulted in the identification of 29 loci not detected in the microsatellite scan, most of which were again family-specific. Three families contributed to a novel ICP susceptibility locus on chromosome 6 with the highest genome-wide single-point NPL score of 5.09 ($p < .000001$). This peak remained significant in subsequent analyses with reduced numbers of SNPs, considering both linkage disequilibrium between SNPs, and including only SNPs informative in all 4 families. Refining this locus in the search for candidate genes is currently underway.

LINKAGE ANALYSIS OF A LARGE TASMANIAN HEMATOLOGICAL CANCER FAMILY

Devindri Perera,[1] Elizabeth Tegg,[1] James Stankovich,[1,2] Russell Thomson,[1] Jeremy Silver,[2] Katherine Marsden,[3] Ray Lowenthal,[3] Ceri Flowers,[1] Annette Banks,[1] Anne Piaszczyk,[1] Melanie Bahlo,[2] Simon Foote,[1] and Jo Dickinson[1]

[1] *Menzies Research Institute, Hobart, Australia*

[2] *Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia*

[3] *Royal Hobart Hospital, Hobart, Australia*

We are performing linkage analysis in a 6-generation Tasmanian family with 10 cases of hematological cancer. The family contains one first cousin marriage resulting in 7 offspring, of whom 4 have been diagnosed with chronic lymphocytic leukemia (CLL) and one with diffuse large B-cell

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

lymphoma. The DNA samples from one deceased CLL case, one unaffected sibling, and 8 offspring of cases have been genotyped with Affymetrix 250k arrays. We are testing the hypothesis that all 5 affected siblings have inherited 2 copies of a rare susceptibility allele from one of their common great-grandparents. To investigate this, we are looking for chromosomal segments inherited homozygous–by-descent in the genotyped case. We are also looking for offspring of the deceased cases to be heterozygous for the allele. We will also investigate the possibility of a dominant susceptibility allele being transmitted through the family. For the CLL case a blood sample was donated at a late stage of disease, and isolated lymphocytes were predominantly leukemic. SNP genotyping revealed a high number of Mendelian errors between the genotypes of the case and his daughter. The leukemic cell population genotyped was shown to carry the common CLL deletion on chromosome 13q14.

ASSESSMENT OF *TGIF* AS A CANDIDATE GENE FOR MYOPIA

Kelly K Pertile,[1,2] Maria Schache,[1,2] F. M. Amirul Islam,[1,2] Christine Yi-Chin Chen,[1,2] Mohamed Dirani,[1,2] Paul Mitchell,[2,3] and Paul N Baird[1,2]

[1] *Centre for Eye Research Australia, University of Melbourne, Melbourne, Australia*

[2] *Vision Cooperative Research Centre, Sydney, Australia*

[3] *Centre for Vision Research, Department of Ophthalmology, Westmead Millennium Institute, University of Sydney, Sydney, Australia*

Myopia is a complex eye disorder with the likelihood of multiple genes being involved in its development. Transforming growth β-Induced factor (*TGIF*) has been identified as a candidate gene for high myopia, based on its location within a linked myopia region, as well as its biological role. However, current data from SNP association studies have been inconclusive. Given that the *TGIF* gene has a role in ocular growth, we wished to undertake a case-control association study in a Caucasian population, to investigate the association of variants within this gene with refraction, as well as ocular biometric measures (axial length, anterior chamber depth, and corneal curvature). A SNP tagging approach was used to evaluate the association between *TGIF* variants and refraction and ocular biometric traits. Nineteen SNPs, including 12 tagging SNPs (tSNPs), were identified within the *TGIF* gene, it's promoter, and 2kb upstream of the promoter region, and genotyped in 257 cases with myopia (high and moderate/low myopes) and 294 controls (no myopia consisting of emmetropes and hypermetropes). Alleles were determined by a MALDI-TOF based approach, and genotype frequencies were analyzed using Chi-square tests and one way ANOVAs. All but one of the SNPs was in Hardy-Weinberg equilibrium (HWE), and this was removed from further analysis. After adjusting for multiple testing using a Bonferroni test, none of the remaining 18 SNPs showed significant association with refractive measurements or ocular biometric measurements in the study population. This case-control study evaluated the association of the *TGIF* gene in individuals with differing levels of refractive errors, including myopia as well as hypermetropia. Based on this study, we can conclude that the *TGIF* gene is unlikely to play a role in either refractive error or ocular biometric measures. Future investigations on identifying myopia candidate genes should therefore focus on other myopia linked regions and genes.

INVESTIGATION OF GENES INVOLVED IN MEMORY PERFORMANCE IN A GENETICALLY DISTINCT SUBTYPE OF SCHIZOPHRENIA

Kirsten Peters,[1] Steven Wiltshire,[1] David Chandler,[2] Assen Jablensky,[2] and Luba Kalaydjieva[1]

[1] *Western Australian Institute for Medical Research, University of Western Australia, Australia*

[2] *Centre for Clinical Research in Neuropsychiatry, University of Western Australia, Australia*

Schizophrenia is a debilitating mental illness with a devastating impact on affected individuals and their families. It is a complex disorder, resulting from the interaction of multiple susceptibility genes and poorly defined environmental factors. The list of putative disease genes is growing, with replication attempts having variable success largely due to the phenotypic and genetic heterogeneity of the disease. A comprehensive phenotype strategy has been employed in the Western Australian Family Study of Schizophrenia in an effort to reduce this phenotypic heterogeneity. Measurements of general cognitive ability, attention, executive function, memory and speed of information processing identified 2 cognitively contrasting schizophrenia subtypes: (a) general cognitive deficit (CD), and (b) cognitively spared (CS). The CD subtype is a genetically distinct, homogenous, highly heritable, biologically relevant trait that co-segregates with clinical illness. The core feature of the CD subtype is impairment in the encoding in verbal memory, particularly after short delay (immediate recall). This study aimed to investigate the involvement of a number of genes recently shown to influence normal memory performance in humans, including NMDA and metabotrobic glutamate receptors, adenylyl cyclase, and protein kinases CAMKII, PKA and PKC (de Quervain & Papassotiropoulos, 2006). Our results suggest association with the CD subtype for 3of the 7 reported genes including *GRIN2A*, *PKA* and *PKC*. Further analyses of these genes will be presented in an extended case-control sample.

INTERLEUKIN-23 RECEPTOR POLYMORPHISMS ARE A MAJOR DETERMINANT OF SUSCEPTIBILITY TO ANKYLOSING SPONDYLITIS

John D. Reveille,[1] Xiaodong Zhou,[1] Linda A. Bradbury,[2] Lon R. Cardon,[3] John C. Davis,[4] Panos Deloukas,[5] David M. Evans,[3] Andrew Keniry,[5] Rhian Gwilliam,[5] Ralph McGinnis,[5] Jenny J. Pointon,[6] Laurie Savage,[7] Wellcome Trust Case-Control Consortium, Thomas L. Learch,[8] Michael M. Ward,[9] Michael H. Weisman,[8] Paul Wordsworth,[6] and Matthew A. Brown[2]

[1] *University of Texas Health Science Center at Houston, Houston, United States of America*

[2] *University of Queensland, Brisbane, Australia*

[3] *Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom*

[4] *University of California-San Francisco, California, United States of America*

[5] *Wellcome Trust, Sanger Institute, Cambridge, United Kingdom*

[6] *University of Oxford, Oxford, United Kingdom*

[7] *Spondylitis Association of America, Sherman Oaks, California, United States of America*

[8] *Cedars-Sinai Medical Center, Los Angeles, United States of America*

[9] *National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, Maryland, United States of America*

IL23R polymorphisms are associated with inflammatory bowel disease and psoriasis. We sought to define this potential association with ankylosing spondylitis (AS). Eight IL23R SNPs were genotyped in the 1000 British AS cases and 1500 controls and in a further cohort of 634 white North American AS cases and 672 controls. The North American cases included Caucasian patients from 2 cohorts: 1) the Prospective Study of Outcomes in Ankylosing Spondylitis (PSOAS; n = 390), and 2) the North American Spondylitis Consortium, with 244 AS probands from multiplex AS families, both meeting modified 1984 New York criteria. The PSOAS patients were further examined for functional impairment using the BASFI and the HAQs, and radiographic severity using the BASRI. Genotyping was performed with the iPLEX assay (MassArray, Sequenom) in the British samples, and by ABI Taqman assays in the North American samples. Association was tested in each dataset independently, and in the combined dataset with p-values determined by simulation with clustering within each dataset, using the program 'PLINK'. In the UK dataset, strong association was seen in 7 of 8 genotyped SNPs ($p < .002$), with peak association seen at rs11209032 ($p = 6 \times 10–6$). In the North American dataset, association was observed with all genotyped SNPs ($p < .04$), with peak association observed with marker rs1343151 ($p = 3.8 \times 10–5$). In the combined dataset, the strongest association observed was with SNP rs11209032 (odds ratio 1.3, 95% CI 1.2–1.4, $p = 3.5 \times 10–8$). The attributable risk fraction for this marker in the North American confirmation cohort is 12%. No association with seen with either functional or radiographic severity. Considering only cases known not to have IBD (n = 1066), the association remained strong, with peak association remaining at rs11209032 ($p = 6.9 \times 10–7$), indicating that there is a primary association with AS, and that the observed association was not due to coexistent clinical IBD. We demonstrate that polymorphisms of IL23R are a major determinant of AS susceptibility but not severity, at least partially explaining the clinical association between AS, psoriasis and inflammatory bowel disease.

TAG-SNP ANALYSIS OF THE VEGF GENE IN AGE-RELATED MACULAR DEGENERATION (AMD)

Andrea J. Richardson, F. M. Amirul Islam, Robyn H. Guymer, Melinda Cain, and Paul N. Baird

Centre for Eye Research Australia, University of Melbourne, Melbourne, Australia

Age-related macular degeneration (AMD) is the leading cause of poor vision in the developed world and its pathogenesis remains unknown. The most common form of end stage disease is neovascular or wet AMD, in which there is underlying growth of new blood vessels. We undertook an association study to investigate the previously associated vascular endothelial growth factor gene, a known angiogenic gene, in an AMD case-control cohort. Five hundred and seventy seven individuals with AMD (early, atrophic and neovascular AMD), and 173 ethnically matched controls were available for this study. We used a tag-single nucleotide polymorphism (t-SNP) approach to investigate this gene, using t-SNPs that encompassed the VEGF promoter, as well as its coding region. t-SNPs were selected using the International HapMap project, with an LD tagging criteria of $r2 > .8$, and with all SNPs having a minor allele frequency (MAF) of at least 0.1. The t-SNPs used in this study included rs833061 (promoter), rs25648, rs3024997, rs2146323, rs3025030, rs3025035 and rs10434 (3' UTR). Alleles were determined by a MALDI-TOF based approach followed by statistical analysis. The 577 affected Caucasian individuals had a mean age of 73.3 years. The 173 control individuals had a mean age of ascertainment of 71.0 years. All but one of the t-SNPs was in Hardy-Weinberg equilibrium (HWE). The SNP found not to be in HWE (rs3024997) was removed from further analysis. No association was found between any of the VEGF t-SNPs analyzed in this study and any sub-type of AMD. Despite a previous association study identifying statistical significance between VEGF SNPs and AMD, we found no evidence to support the involvement of this being a susceptibility gene in our AMD population. The complex nature of AMD indicates that a combination of other genes and/or environmental factors may be playing a role in angiogenesis, and that the VEGF gene is unlikely to be involved in AMD in our study population.

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

PHARMACOGENOMICS, CLOZAPINE AND MYOCARDITIS

Kathlyn Ronaldson,[1] John McNeil,[1] and Justin Rubio[2]

[1] *Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia*

[2] *Howard Florey Institute, University of Melbourne, Parkville, Melbourne, Australia*

Myocarditis is a drug hypersensitivity reaction occurring in perhaps 1% of patients with treatment-resistant schizophrenia in response to clozapine. We are conducting a case-control study to identify genotypic, phenotypic and environmental risk factors for myocarditis with clozapine. Genetic variations predisposing to drug hypersensitivity may be involved in the following functions: drug bioactivation, drug detoxification, immune responsiveness, tissue injury, or tissue repair. Considerable success in identification of predisposing polymorphisms for drug hypersensitivity has been achieved following searches of the human leukocyte antigen (HLA) region of the DNA, which controls immune function. Investigation of genes for other functionalities has been less useful, to date. One group found an HLA-B variant to be present in 94% of patients with hypersensitivity for the antiretroviral abacavir. This same group linked an HLA-DRB variant to hypersensitivity to nevirapine, but only in the presence of patient immunocompetence. Another group identified an interdependent association between a TNFα gene variant and 2 HLA variants in the Class II region and serious carbamazepine hypersensitivity. For our pharmacogenetic analysis of genetic factors predisposing to myocarditis, we plan to conduct exploratory mapping of variation in the HLA complex by typing the HLA-A, B and DRB1 genes at low resolution. Cases and controls will be matched for ethnicity. We will also type SNPs in the coding and non-coding gene of the enzyme metabolizing clozapine, CYP1A2. DNA samples will be saved for investigation of other regions and/or for whole genome scan.

GENETIC MODELING AND FAMILY-BASED ASSOCIATION STUDY INCLUDING HAPLOTYPE ANALYSIS OF ADHD AND ITS COMORBIDITY WITH READING DISABILITY IN AUSTRALIAN TWINS

Abdullah Sheikhi, Jan Piek, Neilson Martin, and David Hay

School of Psychology, Curtin University of Technology, Perth, Western Australia, Australia

Attention Deficit Hyperactivity Disorder (ADHD) and Reading Disability (RD) are common childhood neuro-behavioral disorders that frequently overlap together. However, this comorbidity is still not well understood. This study aimed to investigate the genetic components for ADHD subtypes, and RD separately, and ADHD RD comorbidty, and explore if there is a unique gene pattern for ADHD RD comorbidity, or whether each disorder has it is own genes. Three approaches were applied to data from 2611 Australian twin families. In the first, genetic modelling was used to explore whether ADHD/RD comorbidity originated from genetic or environmental components or both. In the second, Latent Class Analysis (LCA) was used to produce genetically independent classes of ADHD and RD. Finally, family-based genetic association, including haplotype analysis, was also included, for 190 individuals, testing 21 SNPs from 5 ADHD candidate genes (DRD4, DAT1, SNAP25, COMT, and HTR1B), and 4 RD candidate genes (MRS2L, KIAA0319, TTRAP, and THEM2) from the 6p22.2 region. Univariate and bivariate results indicated the presence of genetic components on each ADHD subtype and RD category, and also showed the existence of genetic factors in the ADHD/RD comorbidity. LCA produced 9 ADHD/RD latent classes. Haplotype analysis detected one significant haplotype block containing 2 htSNPs (rs4680 and rs165599), of the COMT gene, with 3 risk alleles associated with some phenotypic RD components. The study concluded that the use of LCA is more appropriate than DSM-IV for genotyping analysis, as the 9 ADHD/RD latent classes were more symptomatically homogenous, to the extent that the identification of such distinctive clusters of symptoms represented more etiologically pure genetic forms of disorders. The study also concluded that COMT, SNAP-25, and KIAA0319 acted differentially on RD alone, ADHD alone, and comorbid ADHD/RD latent classes.

INVESTIGATION OF OBJECTIVE VERSUS SUBJECTIVE MEASURES IN A STUDY OF MELANOMA RISK FACTORS

Sri N. Shekar,[1,4] David L. Duffy,[1] Tony Frudakis,[2] Richard A. Sturm,[3] Grant W. Montgomery,[1] and Nicholas G. Martin[1]

[1] *Genetic Epidemiology Unit, Queensland Institute of Medical Research, Brisbane, Australia*

[2] *DNAPrint, Genomics Inc., Sarasota, Florida, United States of America*

[3] *Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia*

[4] *School of Medicine, University of Queensland, Brisbane, Australia*

A large amount of time and resources are required in collecting phenotypes and genotypes for linkage analyses in gene mapping studies. This usually includes selecting, contacting, collecting phenotypes from, and bleeding, participants followed by blood processing and genotyping. Here we examine the difference in results between having objective, continuous measures compared to subjective, ordered ratings. In a study of the genetics of melanoma risk factors, phenotypes for 1276 families have been collected since the early 1990s. Here we perform linkage analyses on a subset of these individuals for a variety of melanoma risk factors. Both objective and subjective measures were collected for some measures. For example, hair color was quantified using a spectrophotometer as well as being rated on a 5-point scale. In both analyses a lod of approximately 1.6 was found near OCA2, the eye color gene. In contrast, the linkage peaks in the analysis of objectively measured hair color appeared to coincide with genomic regions housing known pigmentation genes, with one marker on 7p significantly linked to hair color (LOD_{obj} ~ 3.3; LOD_{subj} ~ 0). The peaks in regions housing pigmentation genes were less numerous in the analysis using subjective ratings. We expect that, with a reduction in genotyping cost, and subsequent increase in the number of genome wide association studies, the accuracy with which traits are estimated may influence the success of a genetic study. The results are discussed in the context of the relative differences in time, resources, and cost between studies.

DETECTING COPY NUMBER VARIATION IN RELATED INDIVIDUALS

Jeremy Silver,[1,2] Henrik Bengtsson,[3,4] Jim Stankovich,[1,5] Sam Berkovic,[2,6] Richard Huggins,[2] and Melanie Bahlo[1]

[1] *Walter and Eliza Hall Institute, Melbourne, Australia*

[2] *University of Melbourne, Melbourne, Australia*

[3] *Lund University, Lund, Sweden*

[4] *University of California — Berkeley, California, United States of America*

[5] *Menzies Research Institute, Hobart, Australia*

[6] *Austin Health, Melbourne, Australia*

The term copy number (CN) variation describes large (> 1kb) insertions and deletions. Despite some disagreement about the average number and size of CN variants, recent studies show that most people carry several such variants, and that they are widespread in the human genome (Sebat et al., 2004; Iafrate et al., 2004). Furthermore, the vast majority of CN variants (> 90%) are inherited rather than due to somatic rearrangements (Redon et al., 2006). SNP chips can be used to detect both linkage and CN variation, and statistical methods have been developed to find CN aberrations from SNP chip data. If a CN variant causes a heritable phenotypic disorder, the variant itself may mask detection by linkage analysis. Samples used in linkage analysis are not independent due to their relatedness. However none of the available CN algorithms take this into account. We present a method for detecting CN variants in related individuals. It is a modification of the widely used Lander-Green algorithm, based on hidden Markov models. This algorithm can be used to detect sporadic and inherited CN variants. We have implemented this method in the popular statistical package R. The SNP chips required for detecting smaller CN variants (i.e. 1kb-100kb) include hundreds of thousands of markers, thereby generating large amounts of data. We make use of software that handles the large datasets from Affymetrix SNP chips in memory-efficient manner. We apply this method to Affymetrix 500K chips of HapMap trios. We will present some preliminary results, and compare with other available methods for detecting CN variants that do not incorporate the relatedness of samples.

GENETIC ANALYSES IN A NATIONWIDE SAMPLE OF HIGH VERSUS LOW BONE DENSITY INDIVIDUALS DEMONSTRATES ASSOCIATION WITH MULTIPLE WNT PATHWAY GENES

Anne-Marie Sims,[1] Neil Shephard,[2] Alison Dowling,[1] Tracy Doan,[1] Emma L. Duncan,[1] John Eisman,[3] Graeme Jones,[4] Geoffrey Nicholson,[5] Richard Prince,[6] Ego Seeman,[7] Gethin Thomas,[1] John A. Wass,[8] and Matthew A. Brown[1,9]

[1] *Diamantina Institute for Cancer, Immunology and Metabolic Medicine, Brisbane, Australia*

[2] *Western Australian Institute of Medical Research, Perth, Australia*

[3] *Garvan Institute of Medical Research, Sydney, Australia*

[4] *Menzies Research Institute, Hobart, Australia*

[5] *Barwon Health, University of Melbourne, Geelong, Australia*

[6] *School of Medicine and Pharmacology, University of Western Australia, Perth, Australia*

[7] *Departments of Medicine and Endocrinology, Austin Health, University of Melbourne, Melbourne, Australia*

[8] *Nuffield Orthopaedic Centre, Headington, Oxford, United Kingdom*

[9] *Botnar Research Centre, University of Oxford, Oxford, United Kingdom*

Although the high heritability of bone mineral density (BMD) variation has long been established, few genes have been conclusively demonstrated to affect the variation of BMD in the general population. Extreme-truncate selection has been proposed as a more powerful alternative to unselected cohort designs in quantitative trait association studies. We sought to test whether these theoretical predictions resulted in real power gains in studies of the bone densitometry measures BMD, bone mineral content (BMC) and femoral neck area, by investigating their association with Wnt pathway members in an extreme-truncate selected cohort (absolute value BMD z scores 1.5 – 4.0). Using a relatively small cohort (n = 344), we demonstrated strong association with *LRP5*, polymorphisms of which have previously been demonstrated to influence BMD (minimum p = .0006). In

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

addition, we demonstrated strong association of a Wnt antagonist, *SFRP1*, with BMD and BMC (minimum $p = .00042$), and confirmed previously reported associations of *LRP1*, *LRP6* and *SOST* with BMD. Two other Wnt pathway genes, *Wnt3a* and *DKK2*, also showed nominal association with BMD. This study demonstrates that polymorphisms of multiple members of the Wnt pathway influence bone density variation. Further, we have confirmed in a practical trial that study designs involving extreme-truncate selection and moderate sample sizes can robustly identify genes of relevant effect sizes involved in bone density variation in the general population. This has implications for the design of future genome-wide studies of quantitative bone phenotypes relevant to osteoporosis.

A PHASE 1 GENOME-WIDE ASSOCIATION STUDY IN OSTEOPOROSIS

Anne-Marie Sims,[1] Emma L. Duncan,[1] Alison Dowling,[1] Tracy Doan,[1] David Evans,[2] John Eisman,[3] Graeme Jones,[4] Geoffrey Nicholson,[5] Richard Prince,[6] Ego Seeman,[7] John A Wass,[8] and Matthew A. Brown[1,9]

[1] *Diamantina Institute for Cancer, Immunology and Metabolic Medicine, Brisbane, Australia*

[2] *Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom*

[3] *Garvan Institute of Medical Research, Sydney, Australia*

[4] *Menzies Research Institute, Hobart, Australia*

[5] *Barwon Health, University of Melbourne, Geelong, Australia*

[6] *School of Medicine and Pharmacology, University of Western Australia, Perth, Australia*

[7] *Departments of Medicine and Endocrinology, Austin Health, University of Melbourne, Melbourne, Australia*

[8] *Nuffield Orthopaedic Centre, Headington, Oxford, United Kingdom*

[9] *Botnar Research Centre, University of Oxford, Oxford, United Kingdom*

Genomewide-association studies (GWAS) are a powerful method for identifying genes of small to moderate effect involved in common heritable diseases. Apart from some genes of large effect, such as *LRP5*, most genes involved in BMD variation have yet to be identified. We sought to identify those genes by performing a GWAS study in individuals selected with extreme BMD values. A phase 1 GWAS was completed in 69 high ($+1.5 < z$ score $< +4$) and 66 low ($-4 < z$ score < -1.5) BMD cases, genotyped for 317,000 SNPs using the Illumina HumanHap300 chip. A phase 2 study investigating positive findings is underway. For a range of different genetic models this study has equivalent power to a cohort study of 850 to 920 unselected individuals. The overall genotyping success rate for this project was 99.6%; a duplicate sample was identical at all 315,953 successful genotypes. Many genes previously associated with osteoporosis were identified in this screen at significance levels that would carry them forward into phase 2, including *Alox12*, *ANKH*, *COL1A1*, *ESR1*, *IL-6*, *Klotho*, *LRP5* and *LRP6*. Nine SNPs in *NELL-1* achieved $p < .024$, strongly suggesting that *NELL-1* is associated with BMD. Transgenic mice over-expressing *NELL-1* develop craniosynostosis, and have accelerated osteoblast differentiation and mineralizing activity. All 3 SNPs in *Sp7* encoding osterix were BMD-associated ($p = .02–0.0009$). No difference could be identified in comparing British and Australian samples, indicating that British historic controls can be used safely in comparison with Australian cases. GWAS using efficient cohort selection designs are a powerful, cost-effective method for identifying BMD-associated genes.

MAKING FULL USE OF HIGH DENSITY SNP DATA COLLECTED ON RELATED INDIVIDUALS

Jim Stankovich[1,2]

[1] *Menzies Research Institute, University of Tasmania, Hobart, Australia*

[2] *Walter & Eliza Hall Institute of Medical Research, Melbourne, Australia*

While genome-wide association studies are a powerful way to detect common susceptibility variants, linkage analysis with large, multigenerational pedigrees remains a more powerful method to map rare variants. Unlike association analysis, the power of linkage analysis is not diminished by allelic heterogeneity. For linkage analysis in 2- or 3-generation families, lower density SNP sets (of around 10,000 markers) extract most information on patterns of identity-by-descent (IBD) sharing. However for more distantly-related individuals in large pedigrees, extra information on IBD-sharing can be gained from higher density SNP sets. (Thomson et al., 2007). Higher density SNP sets also enable the identification of IBD sharing between pairs of people not known to be related to one another. This opens the possibility of recycling data from genome-wide association studies to perform 'population-based linkage analysis' (Purcell et al., in press). Some questions remain about this approach. Firstly, how much more powerful is this method when applied in founder populations, with their increased amount of recent relatedness? And would it be useful to estimate the age of various IBD haplotypes when assessing the significance of sharing patterns? If so, the advent of high-throughput resequencing may be useful, by counting new mutations that have arisen on IBD haplotypes since the common ancestor (not to mention the obvious benefit of resequencing for identifying rare causative variants). Another issue is how to adjust results from association analysis to allow for the nonindependence of related cases and controls. Recently several methods have been developed to correct for this non-independence, given family structures of arbitrary complexity. One of these methods (Thornton & McPeek, in press) also takes advantage of the fact that cases with many affected relatives are more likely to carry susceptibility alleles than cases with no affected relatives, increasing power under many disease models.

β2-ADRENERGIC RECEPTOR, α-ADDUCIN AND G-PROTEIN β3 SUBUNIT GENES POLYMORPHISMS AND RETINAL VASCULAR CALIBRE IN OLDER PEOPLE: THE CARDIOVASCULAR HEALTH STUDY

Cong Sun, F. M. Amirul Islam, and Tien Yin Wong

Centre for Eye Research Australia, University of Melbourne, Melbourne, Australia

Retinal arteriolar narrowing reflects chronic blood pressure levels, and may predict cardiovascular mortality. Recent genome-wide linkage study revealed that the linkage regions for retinal vascular caliber overlap with regions that have been previously associated with hypertension and coronary heart disease. The aims of the study were to examine the associations of different single nucleotide polymorphisms (SNPs) and haplotype of α-adducin (ADD1) genes, β2-adrenergic receptor (ADRB2) genes, and G-protein β3 subunit (GNB3) genes with retinal vascular caliber. Polymorphisms of these genes have been linked with hypertension and cardiovascular diseases in prior studies. We hypothesize that retinal vascular caliber may behave as intermediate phenotype and the results from this study may provide insights into the pathogenesis of hypertension. A total of 1842 persons aged 69 to 95 years participating in a population-based study from 4 US communities were included in the analysis. A computer-assisted method was used to measure retinal vascular calibre. We analyzed DNA extracted from blood samples of participants for common allelic variants of the 4 genes. In total, 4 SNPs at these loci were analyzed for their association with retinal vascular calibre in 1554 Whites and 228 African–Americans individuals, respectively. UNPHASED software was used for haplotype analysis. All SNPs were in Hardy-Weinberg equilibrium in Whites and African–Americans. There SNPs were used for single genotype and haplotype analysis after excluding Gln27Glu which had high linkage equilibrium (LD = 1) and correlation ($r = .83$) with SNP Arg16Gly. No significant association with retinal vascular caliber was found in single gene or the haplotype analysis. We further tested 16 multilocus comparisons and found that there were no significant deviations (retinal arteriolar caliber: 4 comparisons for Whites and 4 for African–Americans; retinal venular caliber: 4 comparisons for Whites and 4 for African–Americans). The results provide no evidence that variants at the 4 loci are related to retinal vascular caliber in either Whites or African–Americans.

USING LINKAGE DISEQUILIBRIUM TO ESTIMATE EFFECTIVE SEPARATION TIMES FOR HUMAN POPULATIONS

John A. Sved,[1] Allan F. McRae,[2] and Peter M. Visscher[2]

[1] *School of Biological Sciences, Sydney University, Sydney, Australia*

[2] *Queensland Institute of Medical Research, Brisbane, Australia*

The separation time between populations that are derived from a common ancestral population can be estimated from the amount of linkage disequilibrium within and between the populations. If there is a level of linkage disequilibrium $r20$ in the parent population before separation, it is expected that, for loci separated by recombination fraction c, $ri\ rj$ will be reduced from this value by an amount $(1\text{-}c)2$ in each generation, where ri and rj are correlations of alleles for a pair of loci in populations i and j. The advantage of such a 2-locus calculation over single locus FST calculations is that population size should not affect the estimate of divergence. We have used the Hapmap Phase II data to calculate values of $ri\ rj$ for different pairs of populations. Over 250,000,000 pairs of loci (SNPs) within 1 cM were available for the linkage disequilibrium calculation. Cumulated over all such pairs, we find that there is a positive signal in r values up to 0.3 cMs for the separation between African and non-African populations. We estimate the effective separation time of African and non-African populations as less than 1000 generations. Such a time can only be consistent with current estimates of historical separation if there has been substantial migration between African and non-African populations, or if assumptions of our models are otherwise violated. We investigated by computer simulations under a neutral evolutionary model how our estimates of separation time are influenced by migration and changes in effective population size.

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

THE ROLE OF APOPTOTIC REGULATORS IN PLATELET DISORDERS

Rachel B. Tan,[1] Andrew W. Roberts,[1] Warren Alexander,[1] Chris Barnes,[2] Paul Monagle,[2] Annabel Tuckfield,[3] Harshal Nandurkar,[4] Shaun Jackson,[5] Adrienne A. Hilton,[1] Douglas Hacking,[1] Douglas J. Hilton,[1] and Benjamin T. Kile[1]

[1] *The Walter and Eliza Hall Institute of Medical Research, Parkville, Australia*

[2] *The Royal Children's Hospital, Parkville, Australia*

[3] *The Royal Melbourne Hospital, Parkville, Australia*

[4] *St. Vincent's Hospital, Fitzroy, Australia*

[5] *Australian Centre for Blood Diseases, Melbourne, Australia*

Recent mapping and characterisation of a number of ENU mutations in the mouse has implicated members of the Bcl-2-like family of apoptotic regulators in the control of platelet numbers (Mason et al., 2007). It is clear that in mice, mutations in such genes can lead to either a marked deficiency of platelets (thrombocytopenia) or a gross excess of platelets (thrombocytosis). Platelet disorders are commonly diagnosed in humans. However, the underlying etiologies of the conditions are often unknown. Patients with thrombocytopenia are at increased risk of prolonged bleeding due to a decreased ability to form blood clots. Patients with thrombocytosis may have an increased tendency for their blood to clot, and so can be at an increased risk of stroke, heart attack, or blood vessel blockage. Our aim was to translate our findings from the mapping of mouse mutations to the human condition, and investigate the role of apoptotic regulators in human platelet disorders. We hypothesised that polymorphisms in apoptotic regulators such as *BCL2L1* and *BAK1* lead to some cases of inherited or acquired human thrombocytopenia and thrombocytosis. In collaboration with hematologists at the Royal Children's Hospital, St Vincent's Hospital, the Australian Centre for Blood Diseases and the Royal Melbourne Hospital, we are recruiting patients with persistent thrombocytosis or thrombocytopenia of unknown cause. We are collecting DNA from each of these participants and are in the process of sequencing exons and exon/intron boundaries for a number of genes encoding apoptotic regulators. We have preliminary sequence data from an initial cohort of 40 participants with thrombocytosis. There is a pressing need for the development of better diagnostic tools to distinguish different categories of platelet disorders, and more effective therapies tailored to the different subgroups. Through applying our understanding of the genetics of platelet regulation in the mouse to investigations of human platelet disorders we hope to begin addressing these issues.

THE TASMANIAN FAMILIAL LEUKEMIA AND LYMPHOMA RESEARCH STUDY

Elizabeth Teggz,[1] Jim Stankovich,[1,2] Russell Thomson,[1] Devindri Perera,[1] Katherine Marsden,[3] Ray Lowenthal,[1,3] Ceri Flowers,[1] Annette Banks,[1] Anne Piasczyk,[1] Melanie Bahlo,[2] Simon Foote,[1] and Jo Dickinson[1]

[1] *University of Tasmania, Menzies Research Institute, Hobart, Australia*

[2] *Walter & Eliza Hall Institute of Medical Research, Melbourne, Australia*

[3] *Royal Hobart Hospital, Hobart, Australia*

The Tasmanian familial leukemia and lymphoma research group was formed in early 2006. We propose to investigate the genetics of hematological malignancies (HMs) using both genome wide association and familial linkage analyses. A study conducted in the 1970s and 1980s identified a large number of HM families with multiple cases of disease. From these records and those of the Tasmania Cancer Registry (TCR), we have performed extensive genealogical research and expanded these families, identifying a number of new cases. To date, 15 priority pedigrees have been selected for immediate recruitment, comprising numerous affected people. The initial part of the study has involved confirming and reclassifying cases according to the WHO classification of hematological malignancies. A significant difficulty encountered has been that historically, most hospitals and the TCR have shredded many of the deceased patients notes. However of the 135 cases in our 15 priority families, the diagnosis was able to be confirmed/reclassified and some clinical information obtained for 105 cases to date. This success has been due to the excellent records available from the study conducted in 1970s/80s. The LK0016 pedigree has ~ 970 descendants from the founder pair, spread over 8 generations, with 11 (10 confirmed) affected individuals, of which 5 are all siblings. The LK0124 family contains ~ 3948 descendants over 10 generations, with 18 (17 confirmed) affected individuals. Within this family, 3 of the affected had a rare site of involvement of their diffuse large B cell lymphoma (DLBCL). They all presented with primary central nervous system (CNS) lymphoma which comprise 1% of DLBCL. Examination of data available from the Royal Hobart Hospital pathology database shows there have only been 13 other patients diagnosed since October 1993 primary CNS lymphoma. Therefore, to find 3 in this one pedigree is interesting, and genome wide linkage analysis using the Affymetrix 250K arrays are about to be performed for this pedigree. This, and other interesting clinical facts about the affected patients in these pedigrees, will be presented.

THE GENETIC CAUSES OF RENAL DISEASE IN AN AUSTRALIAN ABORIGINAL POPULATION

Russell Thomson,[1] Simon Foote,[1] Jim Stankovich,[1,2] Matthew Jose,[3] John Matthews,[4] and Tim Thornton[5]

[1] *Menzies Research Institute, Hobart, Australia*

[2] *Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia*

[3] *University of Melbourne, Melbourne, Australia*

[4] *Department of Health and Human Services, Tasmania, Australia*

[5] *University of California — Berkeley, California, United States of America*

In the Northern Territory, rates of end-stage renal disease (ESRD) amongst Aboriginals are more than 20 times that of non-Indigenous Australians. This disease occurs in the context of other diseases known to predispose to ESRD. The Tiwi Islanders living on Bathurst and Melville Islands north of Darwin have the highest measured prevalence of renal disease among Australian Aboriginals, despite a profile for either hypertension or diabetes which is similar to other communities. Given the extremely high prevalence and heritability of renal failure in the Tiwi population, we hypothesize that a causative allele that is rare in Western populations has become common on the Tiwi Islands (either through drift or positive selection). We examine the population genetics of the Tiwi population through use of models such as the coalescent process. As a result, we will hypothesize on the level of linkage disequilibrium and the disease model for renal failure amongst the Tiwi population. We are about to commence genotyping 357 Tiwi Islanders using the recently released 500K Affymetrix SNP chip, that contains a further 500,000 STSs, to assess genome stability. With recent advances in statistical genetic analysis techniques, it is now possible to carry out association in the presence of linkage and gain more power from linkage analysis, using case-control data. An overview of these advances will be given, with reference to the Tiwi Island Project.

ESTABLISHING A SUCCESSFUL MAPPING STRATEGY FOR IDENTIFICATION OF THE CAUSAL MUTATION IN ENU GENE-VARIANT MOUSE STRAINS

Belinda Whittle,[1] Edward Bertram,[1,2] and Chris Goodnow[1,2]

[1] *Australian Phenomics Facility, Australian National University, Canberra, Australia*

[2] *John Curtin School of Medical Research, Australian National University, Canberra, Australia*

Many of the health problems we face today, autoimmunity, allergy, obesity, cancer, and cardiovascular disease, stem from discordance between genetics and the environment. The Australian Phenomics Facility (APF) provides a unique approach to study human and animal health by producing pedigrees of mice with informative gene alterations for understanding causal relationships between genes and disease. Genome-wide chemical mutagenesis of mouse spermatogonia with Ethylnitrosourea (ENU) allows a large percentage of mammalian genes to be altered without any prior knowledge about their physiological roles. Pedigrees of these gene-variant mice can be scanned for phenotypes of interest in a highly parallel fashion. The point-mutations produced by ENU yield a large range of gene variant alleles that can reveal different functional roles that are not revealed by knockout studies. The APF produces pedigrees of ENU-gene variant mice on inbred, intercross, or genetically sensitized backgrounds. These libraries are structured for detection of new recessive or dominant alterations in mammalian physiological or pathological processes. The pedigrees are designed so the offspring at the G3 stage from any one G2 breeder pair will carry approximately 20 homozygous recessive random gene mutations. The APF mapping team has many years experience establishing specialized strategies to successfully locate the causative mutations in ENU-gene variant mice. This presentation will describe the current strategies and technologies used to map the causal mutation when breeding ENU-gene variant mice to a closely related mapping strain (C57BL/6 v C57BL/10) or a distantly related mapping strain (C57BL/6 x CBA/Ca), and the effect of the number of ancestral meiosis in these pedigrees. Chromosomal linkage is established through a genome wide scan, followed by linkage confirmation and fine mapping using the Affymetrix 5K mouse SNP chip and Amplifluor SNP technologies.

IDENTIFICATION OF MICE CARRYING ENU-INDUCED SUPPRESSOR MUTATIONS TO EXPERIMENTALLY INDUCED EPILEPSY

Lucy Whittock, Terence O'Brien, Ben Kile, and Simon Foote

Menzies Research Institute, Hobart, Australia

The epilepsies are a common group of chronic neurological conditions characterized by recurrent spontaneous unprovoked epileptic seizures. The cumulative lifetime incidence of epilepsy is likely to be at least 2-4% of the population. Up to 30% of patients with epilepsy will fail to have their seizures controlled by medication, despite trials of multiple different drugs (i.e., pharmacoresistant epilepsy). The redundancy in the efficacy profile of current anti-epileptic drugs is most likely due to the fact that most have been developed by similar targeted approaches. Consequently,

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

most anti-epileptic drugs have similar principal mechanisms of action. Seizures induced in mice by intraperitoneal injection of the potent glutamate agonist kainic acid is a well-characterized model of epilepsy. The goal of this study is to identify ENU-induced mutations in mice that suppress the induction of seizures and epileptogenesis caused by kainic acid. ENU mutagenesis is a non-targeted strategy that potentially queries the involvement of every gene in the genome, and has the potential to identify novel epilepsy genes. The study plans to screen 10,000 progeny (G1) of ENU-treated male mice. This dominant screen on a C57B/6 background involves treating mice up to 3 times, once per hour, and recording the injection number at which each mouse seizes (1, 2, 3 or no seizure). Preliminary experiments showed that control mice do not survive 3 injections. Mutant mice that survived (and therefore carry suppressor mutations) were backcrossed to normal C57B/6 for progeny testing. So far 3,979 G1 mice have been screened and 19 candidate suppressor mutants identified. Of these, 2 have been further screened for heritability of the suppressor mutation. Both lines have shown consistent increased resistance to kainic acid compared to the control population. Analysis of the G1 mice showed that there is no effect of sex on seizure frequency (Pearson's Chi-squared test p value = 0.3449). The genes responsible for the suppressor mutations will be identified using classical positional cloning techniques.

USE OF HOMOZYGOSITY TESTING TO DETECT DISEASE GENES IN DOGS BREEDS

Alan Wilton and Jeremy Shearman

School of Biotechnology and Biomolecular Sciences, University of New South Wales, New South Wales, Australia

The breeding structure of many domestic animals, including dogs, results in isolated populations, each segregating for different disease genes. Because of the extensive use of popular sires and subsequent inbreeding in the closed populations, recessive genetic defects can become common in the population and become a significant medical problem. The various breeds are a good source of models for human genetic disease, and humans are good models for the animal genetic diseases. Since affected animals are identical-by-descent for the disease gene, they will be homozygous for the DNA in a large region around the locus. This is supported by the large regions of linkage disequilibrium in dog breeds. Our research in dogs has shown this method to be effective, having used it to identify 2 disease genes in Border collies, one for a nerve degenerative disease (CL) and the other for an immune dysfunction (TNS). We identified candidate loci based on comparison of disease symptoms with human disease, and used comparative genomics to identify the homologous genes in the dog. Screening of the dog genome sequence in the region identified possible microsatellites within the gene, or flanking it, that were used for testing for homozygosity. As few as 7 affected dogs were sufficient to exclude all except one locus as the TNS gene from a group of 15 candidates. Linkage analysis on large highly inbred pedigrees with SuperLink Online confirmed that we were in the right region with lod scores of 15. Analysis of the populations showed that the carrier rate is high (3–5%) and that linkage disequilibrium is strong in the region. No recombinants were identified between the disease gene and markers 0.5 Mb on either side of the gene, even though the mutation is relatively old. The same mutation occurs in several distantly related populations of show dogs and working dogs. Identification of the gene has allowed the development of a genetic test for carriers that breeders can use to help slowly eliminate the disease from the breed.

FACTORS AFFECTING THE VALUE OF GENOMIC PROFILING FOR DISEASE RISK BASED ON GENOME-WIDE ASSOCIATION STUDIES

Naomi R. Wray,[1] Michael E. Goddard,[2] and Peter M. Visscher[1]

[1] *Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia*

[2] *Faculty of Land and Food Resources, University of Melbourne, and Department of Primary Industries, Victoria, Australia*

Empirical studies suggest that the effect sizes of individual causal risk alleles for complex diseases are small, with most genotype relative risks in the range of 1.1 to 2.0. By implication, many such loci are needed to explain the observed genetic variance, but how many is many? Genome-wide association studies (GWAS) are currently underway worldwide, using large numbers of cases and controls (~ 2000) in order to identify variants of small effect from the mire of multiple testing. For any single locus, the increased risk of disease for a carrier is small. However, to what extent is it possible to identify individuals who are at increased risk of disease because they have received multiple risk variants? We quantified the value of prediction of genetic risk to disease, or 'genomic profiling', based on a range of realistic combinations of the number, size and distribution of risk effects that underlie complex diseases. We simulated a case-control study, estimated a prediction equation from this study, and then tested the accuracy with which the equation predicts disease risk in an independent sample of individuals. When the number of loci contributing to the disease is greater than 50, a large (~ 10,000) case-control study is needed to identify a set of risk loci for using in profiling, larger than the studies than are currently underway. For diseases controlled by 1000 loci of mean relative risk only 1.04, a case-control study with 10,000 cases and controls can lead to selection of ~ 75 loci that explain more than 50% of the genetic variance. We show that an individual's genetic risk can be predicted accurately so that the 5% of people with the highest predicted risk are 3 to 7 times more likely to suffer the disease than the population average. The value of genomic profiling depends on the heritability and prevalence of the disease, suggesting greater success of genomic profiling for some diseases than others.

ANALYSIS OF AROMATASE GENE (*CYP19A1*) POLYMORPHISMS IN FEMALE PATTERN HAIR LOSS

L. Yip,[1,2] S. Harrison,[2] R. Sinclair,[2] G. Severi,[3] D. Muller,[3] S. Harrap,[1] and J. Ellis[1]

[1] *Department of Physiology, University of Melbourne, Melbourne, Australia*

[2] *Department of Dermatology, St Vincent's Hospital, Melbourne, Australia*

[3] *Cancer Epidemiology Centre, The Cancer Council, Victoria, Australia*

The *CYP19A1* gene is located on chromosome 15q21 and codes for the aromatase enzyme, which functions to convert androgens into estrogens. In scalp hair follicles, this conversion step is important to reduce the concentrations of androgens available to drive androgenetic alopecia (AGA). However, the role of androgens in female AGA has not been well established; hence the term female pattern hair loss (FPHL) is preferred. It has been shown that aromatase enzyme concentrations in frontal hair follicles of female balding scalps are significantly greater than that in males (Sawaya et al., 1997), and this may partially account for the varying degrees of hair loss and clinical presentations seen in both sexes. This case-control association study aims to define a relationship between polymorphisms in and around the *CYP19A1* gene and the phenotypes of FPHL, for 484 cases and 471 controls, using tag SNPs identified through the International HapMap Project database. Allele and genotype frequencies of tag SNPs were compared between the case and control groups using χ^2-analysis. So far, the analyses of 2 tag SNPs that collectively capture 22 SNPs did not show any significant differences in allele or genotype frequency between cases and controls. However, the association between *CYP19A1* and FPHL cannot yet be discounted. We are currently examining the remaining 33 tag SNPs necessary to capture all the known 141 SNPs within the gene region. A genetic epidemiological study on FPHL of such a large scale has not been previously described, and a positive association between polymorphisms on CYP19A1 and FPHL would confirm the contribution of this candidate gene in FPHL. It would also strengthen the hypothesis that androgen and estrogen-mediated pathways are involved in FPHL.

ANALYSIS OF THE ESTROGEN RECEPTOR BETA GENE, *esr2,* IN FEMALE PATTERN HAIR LOSS

L. Yip,[1,2] S. Harrap,[2] R. Sinclair,[1] and J. Ellis[2]

[1] *Department of Dermatology, St Vincent's Hospital, Melbourne, Australia*

[2] *Department of Physiology, University of Melbourne, Melbourne, Australia*

Female pattern hair loss (FPHL) has a polygenic mode of inheritance with an unproven relationship with androgens. The genes involved in its pathogenesis remain unknown, but are likely to include genes related to the androgen and estrogen pathways. The *ESR2* gene is located on chromosome 14q22-24 and codes for the estrogen receptor beta (ER-β). In contrast to ER-α, ER-β is the predominant subtype within the outer root sheath and epithelial matrix, making it likely that the *ESR2* gene directly modulates the hair growth cycle. This study aims to evaluate the relationship between *ESR2* gene variants and FPHL. Allele and genotype frequencies of tag single nucleotide polymorphisms (tag SNPs) in the *ESR2* gene were compared between 100 cases with FPHL (stage 3 or greater), and 90 controls (aged > 50 years with no hair loss conditions), using χ^2-tests. Tag SNPs are representative variants in a region of linkage disequilibrium, whereby their examination would be sufficient to capture the genotypic information of all known SNPs in that region. Three *ESR2* tag SNPs that collectively capture 33 SNPs in the gene have been analyzed and there were no significant differences in allele or genotype frequency between cases and controls ($p > .1$). By essentially examining 33 SNPs located in the *ESR2* gene region, we have been unable to identify genetic association between *ESR2* and FPHL. However, the association of this gene with FPHL cannot yet be discounted. We are currently analyzing the remaining tag SNPs necessary to capture all known SNPs in the gene region, and have recruited additional cases and controls through the Melbourne Collaborative Cohort Study to increase our cohort size to 484 cases and 471 controls — making this the first large scale genetic epidemiological study on FPHL.

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press

GENETIC ASSOCIATION ANALYSIS OF ORDERED CATEGORICAL OUTCOMES IN NUCLEAR FAMILIES USING MARKOV CHAIN MONTE CARLO

Sophie G. Zaloumis,[1] Katrina J. Scurrah,[1,2] Lyle C. Gurrin,[2] and Stephen B. Harrap[1]

[1] *Department of Physiology, University of Melbourne, Melbourne, Australia*

[2] *Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, University of Melbourne, Melbourne, Australia*

Statistical methods for the analysis of data from family-based genetic association studies are well developed for continuously valued and binary outcome measures. Ordinal categorical outcomes have received much less attention, and are often analyzed as unordered categorical responses or simply reduced to binary measures. Ordinal data from related individuals presents special statistical challenges because of the possibility that outcome measures are correlated between family members. Current methods allow only simple correlation structures between observations, assuming (unrealistically) that the outcomes are equally correlated among family members, or are restricted to data from twin pairs. These difficulties can be overcome by taking a simulation based approach to model fitting, using Markov Chain Monte Carlo (MCMC) methods in Bayesian framework. This approach has been used to successfully analyze SNP genotypes and ordered categorical phenotypes in nuclear families, and will be illustrated with an application on understanding associations between Male Pattern Baldness (MPB) and the androgen receptor gene in data from the Victorian Family Heart Study. An additional focus will be on investigating the differential effect of SNPs across the categories of an ordinal phenotype — for example, does a particular polymorphism in the androgen receptor gene have an increasing effect on the probability of being in a higher baldness category? The statistical methods presented can be generalized to arbitrary categorical phenotypes.

THE EVOLUTION AND MAINTENANCE OF HOMOSEXUALITY IN THE POPULATION

Brendan P. Zietsch,[1] Katherine I. Morley,[1] Sri N. Shekar,[1] Karin J. H. Verweij,[1,2] and Nicholas G. Martin[1]

[1] *Genetic Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Australia*

[2] *Department of Biological Psychology, VU University, Amsterdam, the Netherlands*

Homosexuality has been shown to have a genetic basis, and on the assumption that it reduces reproductive success, scientists have been unable to explain how it could have evolved. One hypothesis proposes that while genes predisposing to homosexuality may detract from homosexuals' adaptive fitness, they may confer some advantage for nonhomosexuals that carry them. Until now, evidence for what such an advantage may be has not been forthcoming. Here we show in a large twin sample (N = 4904) that heterosexuals with a homosexual twin have more opposite-sex sexual partners in their lifetime than do heterosexual twin pairs. We also show that psychologically masculine females and feminine men are more likely to be homosexual, but when they are nonhomosexual, have more opposite-sex partners. Further, homosexuals have more sex partners in total than nonhomosexuals. Moreover, structural equation modeling on the twins suggests all these variables are partly influenced by common genes. Taken together, these results suggest that genes predisposing to homosexuality may confer a mating advantage in nonhomosexuals, which provides a plausible explanation for the evolution and maintenance of homosexuality in the population.

https://doi.org/10.1375/twin.10.4.663 Published online by Cambridge University Press