

School of Social & Community Medicine



Research student opportunities in Genetics 2011

SUMMARY OF TOPICS

	Supervisor	Title of project
1	Sarah Lewis & George Davey Smith	An investigation of intra-uterine nutrition and prenatal development– applying the principle of Mendelian randomization
2	Dr Santiago Rodriguez & Professor Ian Day	Evaluation of the prevalence and functionality of paucimorphic and private mutations in large epidemiological surveys for cardiovascular risk traits
3	Ian Day & George Davey Smith	Hp genotype as a potential predictor for Hb levels, and investigation of possible correlations with selected phenotypes in mother and child.
4	Tom Gaunt & Ian Day	Disentangling instances of causally and pharmacogenetically relevant genomic confounding
5	<u>Ian Day, Dave Evans and George Davey Smith</u>	A study of genotype influences on reference ranges for clinical analytes ('Range Mendelization')
6	Jeff Holly and Claire Perks	An investigation of nutrition-dependent epigenetic modifications of IGFs and PTEN in relation to cancer risk
7	David Evans; George Davey Smith, Matt Brown	Major Histocompatibility Complex (MHC) Genetics of Ankylosing Spondylitis
8	David Evans, George Davey Smith, Ian Day	"Expression Genetics": Genome-wide association analysis of gene expression data in the ALSPAC cohort
9	Andy Ness, Ian Day	Gene-nutrient interactions in the determination of blood lipid levels and early-stage atherosclerosis in childhood
10	Glyn Lewis	Genetic predictors of adverse effects of antidepressants

- | | | |
|-----------|---|--|
| 11 | Ian Day and Philip Guthrie | Hp genotype as a potential predictor for Hb levels, and investigation of possible correlations with selected phenotypes in mother and child |
| 12 | Tom Gaunt and Ian Day | Disentangling instances of causally and pharmacogenetically relevant genomic confounding |
| 13 | Dr Santiago Rodriguez and Ian Day | Evaluation of the prevalence and functionality of paucimorphic and private mutations in large epidemiological surveys for cardiovascular risk traits |
| 14 | Dr Santiago Rodriguez and Ian Day | Genetic analyses of individuals expressing extremely low levels of plasma protein biomarkers |
| 15 | Sarah Lewis | Meta-analysis of variants in the vitamin D receptor gene and breast cancer risk |
| 16 | Jon Tobias | Fine mapping of the association between <i>osterix</i> genotypes and bone mineral density in children |
| 17 | Nic Timpson, David Evans, George Davey Smith | The use of Genomewide data for the design and undertaking of Mendelian randomisation experiments to dissect potentially causal pathways to common disease |
| 18 | Nic Timpson; David Evans; Tom Gaunt; George Davey Smith | The coordinated analysis of genomewide genotype data, dense transcriptomic data and phenotypic data within a sample of the Avon Longitudinal Study of Parents and Children |
| 19 | George Davey Smith, Sarah Lewis & Debbie A Lawlor | Maternal Phenylalanine and offspring development: Mendelian randomisation approach |
| 20 | Nic Timpson, Emma Clark & Hannah Kuper | Analysis of environmental and genetic contributions to scoliosis in an Indian population. |

**PhD Topics in Genetic Epidemiology, Bioinformatics
and Molecular Genetics 2011**

Title: An investigation of intra-uterine nutrition and prenatal development– applying the principle of Mendelian randomization

Outline of project

Heavy alcohol drinking during pregnancy can result in foetal alcohol syndrome, which is characterized in part by growth deficiency and neuro-developmental disorders. However the effects of moderate levels (within the normal range) of drinking on foetal development during are not clear. Several studies have reported reductions in birth weight, whilst others have found no deleterious effects. Similar the effects of low levels of nutrient intake, for example folate and vitamin D, during pregnancy on infant prenatal growth and development are not clear. The problem is that observational studies are often unable to control for confounding by smoking, other nutrients, socioeconomic status and other lifestyle factors. Measurement of duration and amount of intake may be inaccurate, due to wide categories, misreporting of intake and recall bias.

Genetic variants have been identified which may influence exposure to alcohol and other dietary factors through effects on intake propensities and via differences in metabolism. These variants are likely to be distributed randomly with respect to other dietary and lifestyle factors, including smoking, and can be used as surrogates for measuring dietary intake. It would be of interest to determine whether such genotypes in the mother and foetus, are associated with developmental outcomes among infants and hence whether moderate alcohol consumption and low dietary intake of vitamins and other nutrients during pregnancy influences infant phenotypes.

Objectives

- 1) Identify genetic polymorphisms, which metabolise nutrients, or affect exposure propensities and which may be related to development in utero.
- 2) Obtain genotypes for a cohort of mothers and their offspring with respect to the above polymorphisms.
- 3) Analyse associations between the above polymorphisms and prenatal development to determine which nutrients are important.

Techniques/approaches

This is a genetic epidemiology project, which will encompass; genetics, epidemiology, statistics, bioinformatics

References

Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003 Feb;32(1):1-22.

Supervisors

Sarah Lewis and George Davey Smith

Title: Evaluation of the prevalence and functionality of paucimorphic and private mutations in large epidemiological surveys for cardiovascular risk traits

Outline of Project:

Background

There are two classical genetic models for the molecular basis of common diseases. The prevailing model is based on the hypothesis that common polymorphic alleles exert small individual effects but with significant population attributable risk (common disease/ common variant, CD/CV hypothesis) (1). A contrasting model (rare disease/ rare variant, RD/RV hypothesis) assumes that rare but severe single gene mutations can cause a phenotype which shows strong familial clustering (2). In 2004 we proposed an intermediate genetic model: sequence changes at an intermediate frequency [termed “paucimorphisms” (3)] may exist and may have a moderate effect. We have developed a mutation-scanning approach suitable for whole population screening for unknown mutations and have published theoretical and observational evidence supporting the paucimorphism hypothesis (3-5), including the identification and analysis of paucimorphisms in the LDLR and MC4R genes. However, the full population spectrum of rare, paucimorphic, severe, moderate (forme fruste), and silent mutations and effects is largely unknown. The definition of this type of variation in key genes influencing cardiovascular (CV) risk will facilitate translation to public health in the form of disease prediction in CV risk.

Of particular interest are four genes influencing CV risk: *APOB*, *F5*, *PCSK9* and *CYP2A6*. *APOB* (in particular the mutation R3500Q), is involved in familial ligand-defective apoB (FDB) (6). Factor V Leiden occurs due to a single point mutation on the *F5* gene, and is an inherited condition which predisposes affected individuals to thrombosis (7). *PCSK9* has emerged as a potential target for lowering plasma LDL cholesterol levels, with mutations in this gene associated either with hypercholesterolemia or with hypocholesterolemia (8). We have described associations between an allele (160H) of *CYP2A6* and the likelihood of quitting smoking (9), and, in a subsequent study, we found results suggesting that *CYP2A6* haploinsufficiency increases likelihood of continuing smoking in teenagers (10).

Objectives of the PhD

- 1.- To determine the prevalence of paucimorphic and private mutations in candidate genes for cardiovascular risk from large epidemiological surveys available in Bristol.
- 2.- To determine the functionality of paucimorphic and private mutations in order to infer their role in disease causation and their translational value in form of disease prediction.

Design

Four candidate genes for cardiovascular risk, *APOB*, *F5*, *PCSK9* and *CYP2A6*, will be analysed in large cohorts available in Bristol, including ALSPAC.

Techniques and approaches

The scanning of unknown mutations will be performed using a high-throughput mutation scanning technique (meltMADGE) developed in our laboratory. Paucimorphisms and other variants detected by meltMADGE will be confirmed by sequencing. Direct assays of specific variants will be performed using a Light-Typer instrument, a liquid-phase, fluorescence-based, melting-curve analysis instrument.

Genotype-phenotype analyses will include descriptive analyses for rare mutations and statistical association analyses for paucimorphisms. Specific functional assays for particular variants will be designed.

References

1. Lander, E.S. (1996) The new genomics: global views of biology. *Science*, **274**, 536-539.

2. Pritchard, J.K. (2001) Are rare variants responsible for susceptibility to complex diseases? *Am. J Hum Genet*, **69**, 124-137.
3. Day, I.N.M., Al-Harbi, K., Smith, M., Al Dahmesh, M., Zhang, B., Lotery, A., Cross, N.C.P., Ye, S., Eccles, D., Fox, K., Rodríguez, S. (2004) Paucimorphic alleles versus polymorphic alleles in disease causation: theory, observation and detection. *Curr. Genomics*, **5**, 431-438.
4. Alharbi, K.K., Aldahmesh, M.A., Spanakis, E., Haddad, L., Whittall, R.A., Chen, X.H., Rassoulilian, H., Smith, M.J., Sillibourne, J., Ball, N.J., *et al.* (2005) Mutation scanning by meltMADGE: validations using BRCA1 and LDLR, and demonstration of the potential to identify severe, moderate, silent, rare, and paucimorphic mutations in the general population. *Genome Res.*, **15**, 967-977.
5. Alharbi, K.K., Spanakis, E., Tan, K., Smith, M.J., Aldahmesh, M.A., O'Dell, S.D., Sayer, A.A., Lawlor, D.A., Ebrahim, S., Smith, G.D., *et al.* (2007) Prevalence and functionality of paucimorphic and private MC4R mutations in a large, unselected European British population, scanned by meltMADGE. *Hum Mutat.*, **28**, 294-302.
6. Real, J.T., Chaves, F.J., Ejarque, I., Garcia-Garcia, A.B., Valldecabres, C., Ascaso, J.F., Armengod, M.E., Carmena, R. (2003) Influence of LDL receptor gene mutations and the R3500Q mutation of the apoB gene on lipoprotein phenotype of familial hypercholesterolemic patients from a South European population. *Eur J Hum Genet*, **11**, 959-965.
7. Shaw, D.M. (2006) Factor V Leiden: an overview. *Clin Lab Sci*, **19**, 218-221.
8. Lambert, G. (2007) Unravelling the functional significance of PCSK9. *Curr Opin Lipidol*, **18**, 304-309.
9. Gu, D.F., Hinks, L.J., Morton, N.E., Day, I.N. (2000) The use of long PCR to confirm three common alleles at the CYP2A6 locus and the relationship between genotype and smoking habit. *Ann. Hum. Genet.*, **64**, 383-390.
10. Huang, S., Cook, D.G., Hinks, L.J., Chen, X.H., Ye, S., Gilg, J.A., Jarvis, M.J., Whincup, P.H., Day, I.N. (2005) CYP2A6, MAOA, DBH, DRD4, and 5HT2A genotypes, smoking behaviour and cotinine levels in 1518 UK adolescents. *Pharmacogenet. Genomics*, **15**, 839-850.

Name of potential supervisors

Dr Santiago Rodriguez
 Professor Ian Day

Title: Hp genotype as a potential predictor for Hb levels, and investigation of possible correlations with selected phenotypes in mother and child.

Background

Haptoglobin, encoded by the Hp gene, is a protein which scavenges haemoglobin free in blood plasma and thus protects against peroxidative tissue damage. Haptoglobin allele Hp2 comprises a large duplication of exons 3 and 4, relative to allele Hp1, and its protein product forms multimers with inferior scavenging capacity (1). Evidence has recently emerged that Hp genotype is correlated with Hb levels (paper in preparation). This studentship will investigate the utility of Hp genotype as a marker for Hb levels, and therefore as a possible predictor for Hb-associated phenotypes.

Objectives

- 1) To verify that Hp genotype is a robust instrument for predicting Hb levels in mothers and children of the ALSPAC cohort.
- 2) To test the hypothesis that maternal Hp genotype / variations in maternal haemoglobin levels during pregnancy are correlated with the following phenotypes:
 - offspring birth weight (2)
 - offspring SGA (small for gestational age)(3)
 - risk of postpartum depression (4)
 - risk of preterm PROM (premature rupture of membranes)(5)
 - risk of preterm birth (6)
- 3) To test the hypothesis that children's Hp genotype / variations in their haemoglobin levels are correlated with differences in IQ and growth.
- 4) To explore the relevance of Hp genotype to the interpretation of Hb assay in decision cutpoints in clinical situations of anaemia and its management.

Design

- 1) Tagging SNPs for the duplication, if available, will be selected using HapMap(7). If no suitable tagging SNPs can be found, an in-house liquid-phase copy number assay will be refined and used for genotyping.
- 2) Hp genotypes of the ALSPAC cohort will be analysed with Hb data to quantify the association with Hb levels.
- 3) The suitability of Hp genotype as a novel predictor for certain Hb-related phenotypes will be assessed.

Techniques/approaches

Hp genotyping will be outsourced if practicable, or else carried out in our laboratory using either SNP genotyping or a high-throughput liquid-phase assay developed in-house. Hb levels in mothers and children of the ALSPAC cohort have already been measured. The remainder of the studentship will involve acquiring and processing ALSPAC phenotype data and investigating potential mechanisms for any perceived correlations.

References

1. Langlois, M.R. and Delanghe, J.R. (1996) Biological and clinical significance of haptoglobin polymorphism in humans. *Clin Chem*, **42**, 1589-1600.
2. Steer, P.J. (2000) Maternal hemoglobin concentration and birth weight. *American Journal of Clinical Nutrition*, **71**, 1285S-1287S.
3. Scanlon, K.S., Yip, R., Schieve, L.A. and Cogswell, M.E. (2000) High and low hemoglobin levels during pregnancy: Differential risks for preterm birth and small for gestational age. *Obstetrics and Gynecology*, **96**, 741-748.
4. Corwin, E.J., Murray-Kolb, L.E. and Beard, J.L. (2003) Low hemoglobin level is a risk factor for postpartum depression. *Journal of Nutrition*, **133**, 4139-4142.

5. Ferguson,S.E., Smith,G.N., Salenieks,M.E., Windrim,R. and Walker,M.C. (2002) Preterm premature rupture of membranes: Nutritional and socioeconomic factors. *Obstetrics and Gynecology*, **100**, 1250-1256.
6. Scanlon,K.S., Yip,R., Schieve,L.A. and Cogswell,M.E. (2000) High and low hemoglobin levels during pregnancy: Differential risks for preterm birth and small for gestational age. *Obstetrics and Gynecology*, **96**, 741-748.
7. Gibbs,R.A., Belmont,J.W., Hardenbol,P., Willis,T.D., Yu,F.L., Yang,H.M., Ch'ang,L.Y., Huang,W., Liu,B., Shen,Y. *et al.* (2003) The International HapMap Project. *Nature*, **426**, 789-796.

Supervisors

Ian Day, Professor of Genetic and Molecular Epidemiology
George Davey Smith, Professor of Epidemiology

Title: Disentangling instances of causally and pharmacogenetically relevant genomic confounding

Outline of Project:

Background

The genes for growth hormone (*GHI*) and angiotensin converting enzyme (*ACE*) reside within a 450kb region on human chromosome 17. Research by us and others has demonstrated associations between polymorphisms in the *ACE* (>1000 papers) and *GHI*¹ gene region and cardiovascular, metabolic and numerous other risk phenotypes. Since RAS pathway inhibitors are widely used, and recombinant GH is also used, the pathway inferences are potentially of pharmacogenetic significance. We have demonstrated ~20% linkage disequilibrium between the *ACE* and *GHI* genes using our own data², confirmed with HapMap data (www.hapmap.org)³. Whilst it is apparent that factors in this region influence disease risk, we have demonstrate that pathway misinference may arise due to the high levels of linkage disequilibrium². The use of HapMap data³ enables the identification of other causally and pharmacogenetically important genomic regions in which LD may cause genomic confounding.

Objectives

1. Genotype haplotype-tagging SNPs across the *ACE-GH* genomic region on chromosome 17 in British Women's Heart and Health Study and other relevant cohorts
2. Use haplotype analyses to determine which SNPs are genuinely associated with cardiovascular and metabolic phenotypes, and which are associated due to linkage disequilibrium
3. Use HapMap data to identify other regions in which genomic confounding may occur and apply the same approach as to *ACE-GH*

Design

- 1) SNPs will be selected using data from the HapMap project³ to select haplotype-tagging SNPs.
- 2) Pharmacogenetically and causally important genomic regions will be identified from the literature and LD across those regions analysed using data from the HapMap project³.
- 3) Genotyping work will be outsourced using cohort studies based at University of Bristol for which ethical approval is already in place.
- 4) Haplotype analyses for association with phenotypes will use various programs as described in previous work^{2,4}.

Techniques/approaches

The project will be principally bioinformatic and statistical, with the majority of original data being obtained by outsourcing. However, some problem SNPs may require in-house assay development and genotyping, particularly within the *GHI* region due to high sequence homology between *GHI* and four related genes.

References

1. Day IN, Chen XH, Gaunt TR, King TH, Voropanov A, Ye S, Rodriguez S, Syddall HE, Sayer AA, Dennison EM, Tabassum F, Barker DJ, Cooper C, Phillips DI. Late life metabolic syndrome, early growth, and common polymorphism in the growth hormone and placental lactogen gene cluster. *J Clin Endocrinol Metab.* 2004 89(11):5569-76.
2. Shuwen Huang, Xiao-he Chen, John R. Payne, Matt J Smith, Hugh E. Montgomery, Ian Day and Tom Gaunt Haplotype of Growth Hormone and ACE Genes, Serum ACE and Ventricular Growth: Pathway Inference in Pharmacogenetics *Pharmacogenet Genomics.* 2007 Apr;17(4):291-4.
3. The International HapMap Project. *Nature* 2003, 426: 789-796.
4. Rodriguez S, Gaunt TR, O'Dell SD, Chen XH, Gu D, Hawe E, Miller GJ, Humphries SE, Day IN. Haplotypic analyses of the IGF2-INS-TH gene cluster in relation to cardiovascular risk traits. *Hum Mol Genet.* 2004 Apr 1;13(7):715-25.

Supervisors

Tom Gaunt, Lecturer in Bioinformatics and Molecular Genetics

Ian Day, Professor of Genetic and Molecular Epidemiology

Title: A study of genotype influences on reference ranges for clinical analytes ('Range Mendelisation')

Outline of Project:

Background

Measurement of biochemical, haematological and immunological parameters is undertaken in clinical practice for purposes of diagnosis, prognosis, monitoring, screening and sometimes for genetic counselling. Reference ranges and clinical cutpoints are used in decision-making processes for clinical management of patients, and at the populational level for risk identification and prevention strategies in health screened subjects. Interpretation of individual 'levels' usually relies on cross-sectional data available from an appropriate reference group. However, people 'run' at different setpoints, partly on account of genetic influences, but on first test (e.g. presentation with a possible disease) they do not have a baseline previous value for comparison. An example is ACE level, which is assayed inter alia for uncertain diagnosis ?sarcoidosis. ACE level is significantly predicted by a polymorphism in the ACE gene, with opposite homozygotes showing mean twofold differences of level, which can be 'Z scored' for better diagnostic sensitivity and specificity (1,2). Clearly then the reference ranging (which defines the range within which 95% of people fall) should be genotype-specific for improved precision and accuracy. Some widely used indices display very high heritabilities, implying genotypes underpinning their reference range – for example, platelet count heritability is estimated at 80% (3). Of course the distinction needs to be made between the situation where the disease 'causes' the marker change (where genotype information may enable fine tuning of reference ranges – 'Range Mendelization') (e.g. 2); and the situation where the marker is a 'causal' factor in the disease, in which case the genetic data can give insight into causal mechanism but would not be used to fine tune the reference range – 'Mendelian Randomization' (4). Considered across a wide range of clinical decision tools and consequence health management (quality and costs) there is potential high value translation of molecular information. Additionally, in the context of research, the refinement of such investigations will translate back into more refined information about causal pathways and clinical risk.

Objectives

1. Explore a spectrum of clinical laboratory analytes for their associations with specific genotypes (apparent from literature, genome wide studies or in house studies) and in relation to known heritabilities
2. Derive practical approaches to combine genotypic information with quantitative clinical analyte data
3. Examine the clinical consequences and value of deploying the 'Range Mendelization' approach

Design

1. SNPs will be selected using data from literature, patent databases, genome wide studies and local populational/cohort studies
2. Where appropriate, follow up studies relating SNPs to analytes will be undertaken in unselected population samples already available
3. Robust and simple laboratory typing methods will be developed. Z scores or similar indices to handle analytical reference ranges in genotype-specific ways, will be developed
4. Routine hospital and GP based uses of tests will be explored to estimate the overall effects (clinical value and cost implications) of improved decision ranges for specific scenarios of screening, diagnosis, prognosis, monitoring and counselling.

Techniques/approaches

The project is likely focus on a few important clinical analytes and use a combination of approaches drawing from clinical biochemistry (or other laboratory science discipline), from

contemporary genome wide association data (the first emerging in 2007) and on a range of informatic, statistical and epidemiological methods also.

References

1. Shuwen Huang, Xiao-he Chen, John R. Payne, Matt J Smith, Hugh E. Montgomery, Ian Day and Tom Gaunt Haplotype of Growth Hormone and ACE Genes, Serum ACE and Ventricular Growth: Pathway Inference in Pharmacogenetics *Pharmacogenet Genomics*. 2007 Apr;17(4):291-4
2. [Kruit A](#), [Grutters JC](#), [Gerritsen WB](#), [Kos S](#), [Wodzig WK](#), [van den Bosch JM](#), [Ruven HJ](#). ACE I/D-corrected Z-scores to identify normal and elevated ACE activity in sarcoidosis. [Respir Med](#). 2007 Mar;101(3):510-5. Epub 2006 Aug 9.
3. [Evans DM](#), [Zhu G](#), [Duffy DL](#), [Montgomery GW](#), [Frazer IH](#), [Martin NG](#). Multivariate QTL linkage analysis suggests a QTL for platelet count on chromosome 19q. [Eur J Hum Genet](#). 2004 Oct;12(10):835-42
4. [Davey Smith G](#), [Ebrahim S](#) 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? [Int J Epidemiol](#). 2003 Feb;32(1):1-22

Supervisors

Ian Day, Professor of Genetic and Molecular Epidemiology
David Evans, Senior Lecturer in Biostatistical Genetics
George Davey Smith, Professor of Epidemiology

Title: An investigation of nutrition-dependent epigenetic modifications of IGFs and PTEN in relation to cancer risk.

Outline of Project:

Background.

Although there has been considerable progress characterising the molecular profile of cancers the factors determining aetiology and disease progression are still poorly understood. Identified cancer susceptibility genes account for a very small proportion of the common cancers and the predominant contribution to most sporadic cancers is considered to be environmental factors, the most significant of which is undoubtedly nutrition. It has become increasingly apparent that in addition to genes being activated or inactivated due to mutations, genes can be similarly affected without any change in DNA sequence due to epigenetic modifications. Every cell in the body contains exactly the same DNA, but the genes expressed in any cell varies dramatically in a development, tissue and cell specific manner and this pattern of expression is programmed via epigenetic mechanisms. While much of this is set for the lifetime of the organism, the epigenetic regulation of some genes is however modifiable by environmental exposures. Again the most evident exposure being nutrition and many genes amenable to such epigenetic modification are themselves involved in metabolic regulation, enabling adaptation to the nutritional environment. Such epigenetic modification can result in long-term programmed changes in gene expression which can contribute to the development and/or progression of chronic diseases such as cancer. The insulin/insulin-like growth factor (IGF) system comprises key regulators that co-ordinate tissue growth and development with nutrition and genes in this pathway are particularly affected by environmentally modified epigenetic regulation. Very recently groups have started to screen tumours for aberrant DNA methylation, in an analogous manner to much earlier screens for genes that are commonly mutated in tumours, and IGF-pathway genes have consistently been identified with epigenetic modifications. Although there has been much progress in characterising the genetic profile of some common cancers, studies of the epigenome are only just beginning.

Aims of the PhD.

- Examine the methylation and acetylation status of IGF-pathway genes in samples from population cohorts.
- Analyse associations between methylation/acetylation status of genes, circulating levels of IGF-related proteins and components of diet. In cases and controls analyse associations between methylation/acetylation status of genes and risk of common cancers.
- Examine the methylation/acetylation status of IGF-related genes in primary breast and prostate tumours.
- Analyse associations between methylation/acetylation status of genes, nutrition and cancer progression.

Techniques and Methods.

The methylation and acetylation status of IGF-related genes will be analysed using the chromatin immunoprecipitation (ChIP) assay and combined bisulfite restriction analysis (COBRA) and bisulphate DNA sequencing, in addition to methylation-specific polymerase chain reaction (MSP).

The Department of Social Medicine hosts a wide range of cohorts and has strong collaborative links with many international cohorts with appropriate biological samples and extensive rich nutritional data; including a large randomised trial of the treatment for prostate cancer (ProtecT). This project encompasses genetics, nutrition and epidemiology.

References.

1. Waterland RA, Michels KB. Epigenetic Epidemiology of the Developmental Origins Hypothesis. *Annu Rev Nutr.* 2007 Aug 21;27:363-388.
2. Waterland RA, Garza C. 1999 Potential mechanisms of metabolic imprinting that lead to chronic disease. *Am J Clin Nutr* 69:179-97

Supervisors.

Jeff Holly and Claire Perks (with expertise in IGFs and laboratory techniques) together with Richard Martin and David Gunnell (with expertise in epidemiology, cancer and nutrition).

Title: Major Histocompatibility Complex (MHC) Genetics of Ankylosing Spondylitis**Outline of Project:**

Ankylosing spondylitis (AS) is a common inflammatory arthritis, affecting 4/1000 white Caucasians, which causes pain and stiffness predominantly of the spine, and inexorable progressive fusion (ankylosis) of the affected joints. There is a strong genetic component in the risk of developing the condition, with heritability assessed in twins at 97%. Whilst the major gene for the disease is known (*HLA-B27*), there is strong evidence that other MHC genes are involved. The aim of this PhD project is to identify the non-B27 MHC genes which influence susceptibility to and clinical manifestations of AS.

Objectives

1. To identify genetic variants within the MHC that predispose to Ankylosing Spondylitis.
2. To follow up significant associations in additional samples so as to ensure robustness and replicability of the findings.
3. Fine map the genomic regions of interest so as to precisely identify the functional variants involved.

Background Ankylosing spondylitis (AS) is a common inflammatory arthritis affecting 0.4% of white European populations (Braun et al. 1998). AS typically develops in the 3rd-4th decade of life, and occurs more frequently in men than women, with a gender ratio of 2-3:1. Characteristic clinical features are inflammatory back pain, asymmetric peripheral arthritis, enthesitis, and anterior uveitis. The condition primarily affects the spine and sacroiliac joints of the pelvis, causing pain and stiffness and eventual fusion. The characteristic location of the inflammation in AS is in the site of attachment of ligaments and tendons to bones (entheses). Unlike 'seropositive' forms of arthritis like rheumatoid arthritis, in which inflammation leads to bone and joint erosion, in AS initial erosion is followed by relentless new bone formation leading to joint fusion. This process is very poorly understood. Although anti-TNF drugs (e.g. adalimumab, etanercept and infliximab) produce improvements in acute inflammation in AS, there are no treatments which have to date induced remission of AS or retarded progressive joint fusion that inevitably occurs in the disease. Thus there is an urgent need for more effective therapies.

Genetics research has provided important information as to the aetiopathogenesis of AS. There is a strong genetic component in the risk of developing the condition, with heritability assessed in twins at >97% (Brown et al. 1997). Approximately 5% of carriers of the main susceptibility gene (*HLA-B27*) develop AS, and over 95% of AS cases are *B27*-positive, compared with ~ 8% of healthy Europeans (Brown et al. 1996). The most likely genetic model for the condition is that *HLA-B27* is required for the inheritance of the disease, but that other genes are important in modifying its penetrance, explaining why only 1-5% of *B27* carriers develop AS (Brown et al. 2000). The severity of disease is also largely genetically determined, with heritability of disease activity, functional impairment and radiographic disease extent of 51% (Hamersma et al. 2001), 76% (Brown et al. 2003), and 62% (Brophy et al. 2004) respectively.

Amongst immunological diseases, AS is unusual in its strong HLA Class I association. Two Class I genes, *HLA-B27* and *HLA-B60*, have been demonstrated to play independent roles in susceptibility to AS by different research groups in different populations (Brown et al. 1996; Robinson et al. 1987). The association with B27 has been known for over 30 years but remains unexplained. In British Caucasians, *HLA-B27* is associated with disease with an odds ratio of >100 (Brewerton et al. 1973; Schlosstein et al. 1973); the association of *HLA-B60* with AS is weaker with an odds ratio of 3.6 (Brown et al. 1996). Whilst it is generally accepted that *HLA-B27* is involved directly in AS-pathogenesis, it is uncertain as to whether *HLA-B60* is also disease-causing itself, or a marker of an MHC haplotype bearing other disease causing genes. The association of *HLA-B60* with disease is well established in B27-positive cases (Robinson et al. 1989), and there is data suggesting a role in B27-negative AS (Wei et al. 2004). Identifying the other genes involved in AS is likely to further advance our understanding of how B27 itself is involved, and thereby our understanding of the biology and function of the HLA Class I system.

As part of a large Wellcome Trust funded program, the Wellcome Trust Case Control Consortium (WTCCC), we have genotyped 1000 unrelated AS cases and 1500 locale matched controls for 12,000 non-synonymous SNPs spread across the genome, and 2360 MHC SNPs (The Wellcome Trust Case Control Consortium, 2007). These studies showed extremely strong and broad association of the MHC with AS, with association with p-values <10⁻⁵⁰ present from 30.9Mb to 32.5Mb from the p-telomere of chromosome 6. As the controls in this analysis are not matched for *HLA-B27* with the cases, this association probably reflects both linkage disequilibrium with *HLA-B27*, and the presence of non-B27 MHC associated genes.

It has been a major goal of our research to identify the non-B27 genes involved in AS. Our preliminary data indicates that these genes lie both on MHC haplotypes bearing B27, and on non-B27 MHC haplotypes. We aim to identify both sets of genes using pre-existing genotypes from the Wellcome Trust Case-Control Consortium AS study, and from genotyping further cases and controls, and performing analysis controlled for the effects of the AS-associated HLA-B alleles, *-B27* and *B*4001*.

Plan of investigation

The project will involve use of several methods of study design and analysis relevant to epidemiology, genetic epidemiology, bioinformatics and biostatistics:

- (a)Cleaning, managing and maintaining a large database of genetic information
- (b)Employing conditional genetic association analysis, principal components analysis and other appropriate statistical methodology to identify genetic variants associated with Ankylosing Spondylitis.
- (c)Prioritizing SNPs for follow up and fine mapping
- (d)Organising follow up genotyping where necessary

References

- Braun J, Bollow M, Remlinger G, Eggens U, Rudwaleit M, Distler A, et al. (1998) Prevalence of spondylarthropathies in HLA-B27 positive and negative blood donors. *Arthritis Rheum*; **41**(1):58-67.
- Brewerton, D.A., Hart, F.D., Nicholls, A., Caffrey, M., James, D.C. & Sturrock, R.D. (1973) Ankylosing spondylitis and HL-A 27. *Lancet* **1**, 904-7.
- Brophy, S., Hickey, S., Menon, A., Taylor, G., Bradbury, L., Hamersma, J., Brown, M.A. & Calin, A. (2004) Concordance of disease severity among family members with ankylosing spondylitis? *J Rheumatol* **31**, 1775-8.
- Brown, M.A., Brophy, S., Bradbury, L., Hamersma, J., Timms, A., Laval, S., Cardon, L., Calin, A. & Wordsworth, B.P. (2003) Identification of major loci controlling clinical manifestations of ankylosing spondylitis. *Arthritis Rheum* **48**, 2234-9.

Brown, M.A., Kennedy, L.G., MacGregor, A.J., Darke, C., Duncan, E., Shatford, J.L., Taylor, A., Calin, A. & Wordsworth, P. (1997) Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment. *Arthritis Rheum* **40**, 1823-8.

Brown, M.A., Laval, S.H., Brophy, S. & Calin, A. (2000) Recurrence risk modelling of the genetic susceptibility to ankylosing spondylitis. *Ann Rheum Dis* **59**, 883-6.

Brown, M.A., Pile, K.D., Kennedy, L.G., Calin, A., Darke, C., Bell, J., Wordsworth, B.P. & Cornelis, F. (1996) HLA class I associations of ankylosing spondylitis in the white population in the United Kingdom. *Ann Rheum Dis* **55**, 268-70.

Hamersma, J., Cardon, L.R., Bradbury, L., Brophy, S., van der Horst-Bruinsma, I., Calin, A. & Brown, M.A. (2001) Is disease severity in ankylosing spondylitis genetically determined? *Arthritis Rheum* **44**, 1396-400.

Robinson, W.P., van der Linden, S.M., Khan, M.A., Rentsch, H.U., Cats, A., Russell, A. & Thomson, G. (1989) HLA-Bw60 increases susceptibility to ankylosing spondylitis in HLA-B27+ patients. *Arthritis Rheum* **32**, 1135-41.

Satsangi, J., Welsh, K., Bunce, M., Julier, C., Farrant, J., Bell, J. & Jewell, D. (1996) Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease. *Lancet* **347**, 1212-1217.

Schlosstein, L., Terasaki, P.I., Bluestone, R. & Pearson, C.M. (1973) High association of an HL-A antigen, W27, with ankylosing spondylitis. *N Engl J Med* **288**, 704-6.

Wei, J.C., Tsai, W.C., Lin, H.S., Tsai, C.Y. & Chou, C.T. (2004) HLA-B60 and B61 are strongly associated with ankylosing spondylitis in HLA-B27-negative Taiwan Chinese patients. *Rheumatology (Oxford)* **43**, 839-42.

Supervisors

David Evans; George Davey Smith, Matt Brown (University of Queensland, Brisbane, Australia).

Title: “Expression Genetics”: Genome-wide association analysis of gene expression data in the ALSPAC cohort

Outline of Project:

This PhD project offers the exciting prospect of combining the techniques of quantitative genetics and genetic expression analysis to gain important insights into biological networks and guide efforts in gene mapping. There is growing realization that merging classic statistical genetics methods with those involving expression profiling will be crucial for understanding the etiology of complex disease. This project will be the largest such study of gene expression phenotypes in the world and will offer the exciting opportunity to merge these data with results from classical genetic analyses of quantitative traits and disease endpoints. The more mathematically minded student will also have the opportunity to be involved in the development of statistical methods for the analysis of complex multivariate datasets such as these.

Objectives

1. To identify genetic variants associated with gene expression using genome-wide association analysis.
2. Fine map the genomic regions of interest so as to precisely identify the functional variants involved.
3. Characterize the distribution of expression quantitative trait loci across the genome.
4. Use bioinformatics approaches to identify functional relationships among the transcripts affected by common loci and to investigate the structure of the underlying regulatory networks.
5. Investigate the relationship between genetic variants, transcript levels and quantitative phenotypes.

Background

The classic genetic mapping techniques of linkage analysis and positional cloning have been responsible for the identification of thousands of genetic variants that cause hereditary disease (Botstein & Risch, 2003). These variants typically involve insertion/deletions or non-synonymous changes in exons, which subsequently produce major changes in protein structure, and consequently large (often clinical) phenotypic effects. In contrast, most of the genetic variants underlying complex traits and diseases are likely to be of small effect and not involve structural changes in the protein coding sequence. One possibility is that genetic variants that influence the amount of mRNA transcript may be particularly important in the etiology of complex traits and disease. Thus, the genetic study of differential expression within and among populations may yield important insights into the genetic causes of human phenotypic variation.

The idea behind “expression genetics” is to subject levels of gene expression to exactly the same genetic mapping techniques (i.e. linkage and association analysis) that one would use for more “complex” classical quantitative traits (Cheung et al. 2005; Evans & Cardon, 2006; Morley et al. 2004; Stranger et al. 2005). The difference between analyzing transcript levels and traditional phenotypes is that literally thousands of variables are assayed at once. The large-scale nature of the technique has the potential to elucidate many different biological pathways instead of focusing on a handful of outcomes as is the norm in traditional genetic studies. Transcript levels are closely connected with variation at the DNA level and can thus serve as a bridge linking genomic variation with more complicated phenotypes further downstream (Rockman & Kruglyak, 2005). Additionally, many expression quantitative trait loci (eQTLs) are of far larger effect than traditional QTLs and are thus easy to identify with smaller numbers of subjects (Cheung et al. 2005; Stranger et al. 2005).

The aim of this project is to identify genetic variants responsible for variation in gene expression data via genome-wide association (GWA). One thousand children from the Avon Longitudinal Study of Parents and Children (ALSPAC) cohort are currently being genotyped on the Illumina 317K SNP array as well as having thousands of mRNA expression levels measured on a high density micro-array chip. The student will perform genome-wide association analysis on these expression phenotypes and will characterize the distribution of eQTLs throughout the genome. Specifically eQTLs will be divided into *cis* eQTLs (i.e. those eQTLs that are located within/adjacent to the gene whose mRNA levels they influence) and *trans* eQTLs (i.e. eQTLs that are located some distance from the genes whose mRNA levels they influence). For *cis* eQTLs, the student will employ newly developed methods that allow confirmation that the eQTLs genuinely reflect differences in expressed mRNA levels as opposed to differences in hybridization to the probe set (Alberts et al. 2007). In terms of *trans* eQTLs, the focus will be on using bioinformatics approaches to identify functional relationships among the transcripts affected by common loci, and to investigate the structure of the underlying regulatory networks. Finally, and perhaps most importantly, the relationship between genetic variation, variation in transcript levels and endophenotype measures will be investigated.

Expression genetics and the analysis of high dimensional datasets such as these are in their infancy and will involve many statistical challenges. Whilst not necessary for this PhD project, the successful student will have the opportunity to be involved in the development of statistical methods to analyse expression genetics datasets. Possible areas of contribution will include methods to deal with multiple testing, multivariate analysis and Bayesian statistics.

Plan of investigation

The project will involve use of several methods of study design and analysis relevant to epidemiology, genetic epidemiology, bioinformatics and biostatistics:

- (a) Cleaning, managing and maintaining a large database of genetic information.
- (b) Employing genome-wide association to identify genetic variants associated with mRNA transcript levels.
- (c) Describe the distribution of *cis* and *trans* eQTLs in the human genome
- (d) Confirm that *cis* eQTLs are due to actual variation in mRNA levels and do not reflect an artefact of probe hybridization.
- (e) Using bioinformatics tools to identify functional relationships among the transcripts affected by common loci and to investigate the structure of the underlying regulatory networks.
- (f) Investigate the relationship between genetic variants, transcript levels and quantitative phenotypes.

References

Alberts R., Terpstra P., Li Y., Breitling R., Nap J. & Jansen R.C. (2007). Sequence polymorphisms cause many false cis eQTLs. *PLOS one*, 7, e622.

Botstein D. & Risch N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet*, 33 Suppl, 228-37.

Cheung V.G., Spielman R.S., Ewens K.G., Weber T.M., Morley M. & Burdick J.T. (2005). Mapping determinants of human gene expression by regional and genome-wide association. *Nature*, 437, 1365-9.

Evans D.M. & Cardon L.R. (2006). Genome-wide association: a promising start to a long race. *Trends Genet*, 22, 350-4.

Morley M., Molony C.M., Weber T.M., Devlin J.L., Ewens K.G., Spielman R.S. & Cheung V.G. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature*, 430, 743-7.

Rockman M.V. & Kruglyak L. (2006). Genetics of global gene expression. *Nat Rev Genet*, 7, 862-72.

Stranger B.E., Forrest M.S., Clark A.G., Minichielle M.J., Deutsch S., Lyle R., Hunt S., Kahl B., Antonarakis S.E., Tavaré S., Deloukas P. & Dermitzakis E.T. (2005). Genome-wide associations of gene expression variation in humans. *PloS Genet*, 1, e78.

The Wellcome Trust Case Control Consortium (2007). Genome-wide association study of seven common disease and 3,000 shared controls. *Nature*, 447, 661-678.

Supervisors

David Evans, George Davey Smith, Ian Day

Title: Gene-nutrient interactions in the determination of blood lipid levels and early-stage atherosclerosis in childhood

Background

This study will take place within ALSPAC, a population-based prospective study of over 10000 children in South West England. Measures of arterial distensibility and flow-mediated dilatation are available in around 7000 children at age 10y as measures of early-stage arterial disease. Measures of non-fasting cholesterol and triglycerides are available for the whole cohort at age 7y and 11y and repeated measures of diet throughout childhood are available. Funding has already been obtained to genotype the whole cohort and the mothers for ApoE, and we will additionally genotype for SNPS in the APOAI-APOCIII-APOAIV-APOAV gene cluster.

Objectives and Design

1. To investigate cross-sectional associations between blood lipid levels and a number of candidate polymorphisms.
2. To investigate the importance of triglyceride levels in atherosclerosis by determining how candidate genes associated with triglyceride levels are associated with endothelial function in children.
3. To investigate the importance of maternal phenotype in determining childhood atherosclerosis by examining associations with maternal genotype.
4. To investigate associations between candidate genes and postprandial triglyceridemia.
5. To describe how any such associations interact with/are modified by diet.

A genotype-selected group of 400 15-year-old children and their mothers will be invited to a research clinic for collection of fasting blood samples which will be analysed for cholesterol and triglycerides. Cross-sectional associations between genotype and blood lipid and atherosclerosis phenotypes will be produced, and we will investigate the importance of maternal phenotype in determining childhood atherosclerosis by examining associations between maternal genotype and childhood atherosclerosis. We will examine associations between genes affecting triglyceride levels and atherosclerosis to look for evidence that postprandial hyperlipidemia is causally associated with CVD. This study should clarify understanding of how diet and genotype interact to promote atherosclerosis.

Studies on the whole of ALSPAC – This study will cross-sectionally investigate relationships between ApoE and APOAI-APOCIII-APOAIV-APOAV genotypes and phenotypes including endothelial function, serum total, HDL and LDL cholesterol and non-fasting triglycerides. This will take into account interactions with diet, characterised in terms of food and nutrient intakes and eating patterns.

Focussed genotype-selected studies – A group of 400 15-year-old children and their mothers, selected on the basis of genotype, will be invited to a research clinic for collection of fasting blood samples. Rare homozygotes will be oversampled to increase statistical power. These will be analysed for triglycerides, apoB, total cholesterol, HDL cholesterol, LDL particle size and the LDL subfractions LDLI, LDLII and LDLIII. Measurements of these lipoprotein subfractions can be used to give an indication of postprandial triglyceridaemia and metabolism²⁰. Assessment of recent diet and physical activity will be made, as these have been shown to affect fat tolerance^{21,22}.

Techniques/approaches

Nutritional studies in adolescents (clinical research) – wide variety of aspects

Bioinformatics – mainly genomic and genetic

Statistical and statistical genetic analyses – broad range of methodologies

Supervisors

Andy Ness, Ian Day

Title: Genetic predictors of adverse effects of antidepressants

Background:

Antidepressants are widely used and over 30 million prescriptions for an antidepressant were made in 2006. However, many people stop using antidepressants before they have had a chance of being effective. There are a number of reasons for this but include the clinical observation that some individuals seem particularly sensitive to the adverse effects. It is possible that there are genetic predictors of adverse effects.

Most of the currently prescribed antidepressant drugs inhibit the 5HT reuptake mechanism (selective serotonin reuptake inhibitors, SSRIs). The most interesting polymorphism in the gene which encodes the serotonin transporter (SLC6A4)¹ is a 44 base pair insertion/deletion polymorphism within a repetitive unit in the 'promoter' region. Both reporter gene analyses in cell culture and analysis of native receptors in lymphoblastoid cell lines suggest that the long (insertion) form is functionally more active than the short (deletion) form. There is also human data to suggest that healthy volunteers with the long insertion allele have a greater prolactin response to fenfluramine when on SSRIs.²

There has been little work on adverse effects in relation to genotype. There is some evidence that discontinuation rates are lower in people with depression who are receiving SSRIs who have the two long insertion alleles in SLC6A4.³ However, in this study there was no comparison group so the higher rate of discontinuation observed in those with short alleles might not have been a reflection of sensitivity to SSRIs.

METHOD

GENetic Predictors of Outcome in Depression GenPod) is a large randomized trial funded by the MRC. It will complete recruitment at the end of 2007 and has over 570 participants. It randomized between reboxetine 4mg bd and citalopram 20mg. Reboxetine is a highly specific noradrenaline reuptake inhibitor (NaRI) and citalopram is an SSRI. Side effects were measured at 2 6 and 12 weeks using a self-administered questionnaire. The main hypothesis for the funded study was to examine genetic and clinical predictors of outcome according to NaRI and SSRI allocation. There will therefore be opportunities for a PhD project to examine genetic predictors of adverse effects in NaRIs and SSRIs.

Reference List

- (1) Lesch KP, Bengel D, Heils A, Sabol SZ, Greenberg BD, Petri S et al. Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 1996; 274:1527-1530.
- (2) Whale R, Quedstedt DJ, Laver D, Harrison PJ, Cowen PJ. Serotonin transporter (5-HTT) promoter genotype may influence the prolactin response to clomipramine. *Psychopharmacology* 2000; 150:120-122.
- (3) Murphy GM, Jr., Hollander SB, Rodrigues HE, Kremer C, Schatzberg AF. Effects of the serotonin transporter gene promoter polymorphism on mirtazapine and paroxetine efficacy and adverse events in geriatric major depression. *Arch Gen Psychiatry* 2004; 61(11):1163-1169.

Supervisor: Glyn Lewis

Title: Hp genotype as a potential predictor for Hb levels, and investigation of possible correlations with selected phenotypes in mother and child.

Background

Haptoglobin, encoded by the Hp gene, is a protein which scavenges haemoglobin free in blood plasma and thus protects against peroxidative tissue damage. Haptoglobin allele Hp2 comprises a large duplication of exons 3 and 4, relative to allele Hp1, and its protein product forms multimers with inferior scavenging capacity (1). Evidence has recently emerged that Hp genotype is correlated with Hb levels (paper in preparation). This studentship will investigate the utility of Hp genotype as a marker for Hb levels, and therefore as a possible predictor for Hb-associated phenotypes.

Objectives

1. To verify that Hp genotype is a robust instrument for predicting Hb levels in mothers and children of the ALSPAC cohort.
2. To test the hypothesis that maternal Hp genotype / variations in maternal haemoglobin levels during pregnancy are correlated with the following phenotypes:
 - i. offspring birth weight (2)
 - ii. offspring SGA (small for gestational age)(3)
 - iii. risk of postpartum depression (4)
 - iv. risk of preterm PROM (premature rupture of membranes)(5)
 - v. risk of preterm birth (6)
3. To test the hypothesis that children's Hp genotype / variations in their haemoglobin levels are correlated with differences in IQ and growth.
4. To explore the relevance of Hp genotype to the interpretation of Hb assay in decision cutpoints in clinical situations of anaemia and its management.

Design

1. Tagging SNPs for the duplication, if available, will be selected using HapMap(7). If no suitable tagging SNPs can be found, an in-house liquid-phase copy number assay will be refined and used for genotyping.
2. Hp genotypes of the ALSPAC cohort will be analysed with Hb data to quantify the association with Hb levels.
3. The suitability of Hp genotype as a novel predictor for certain Hb-related phenotypes will be assessed.

Techniques/approaches

Hp genotyping will be outsourced if practicable, or else carried out in our laboratory using either SNP genotyping or a high-throughput liquid-phase assay developed in-house. Hb levels in mothers and children of the ALSPAC cohort have already been measured. The remainder of the studentship will involve acquiring and processing ALSPAC phenotype data and investigating potential mechanisms for any perceived correlations.

References

1. Langlois,M.R. and Delanghe,J.R. (1996) Biological and clinical significance of haptoglobin polymorphism in humans. *Clin Chem*, **42**, 1589-1600.
2. Steer,P.J. (2000) Maternal hemoglobin concentration and birth weight. *American Journal of Clinical Nutrition*, **71**, 1285S-1287S.
3. Scanlon,K.S., Yip,R., Schieve,L.A. and Cogswell,M.E. (2000) High and low hemoglobin levels during pregnancy: Differential risks for preterm birth and small for gestational age. *Obstetrics and Gynecology*, **96**, 741-748.
4. Corwin,E.J., Murray-Kolb,L.E. and Beard,J.L. (2003) Low hemoglobin level is a risk factor for postpartum depression. *Journal of Nutrition*, **133**, 4139-4142.

5. Ferguson,S.E., Smith,G.N., Salenieks,M.E., Windrim,R. and Walker,M.C. (2002) Preterm premature rupture of membranes: Nutritional and socioeconomic factors. *Obstetrics and Gynecology*, **100**, 1250-1256.
6. Scanlon,K.S., Yip,R., Schieve,L.A. and Cogswell,M.E. (2000) High and low hemoglobin levels during pregnancy: Differential risks for preterm birth and small for gestational age. *Obstetrics and Gynecology*, **96**, 741-748.
7. Gibbs,R.A., Belmont,J.W., Hardenbol,P., Willis,T.D., Yu,F.L., Yang,H.M., Ch'ang,L.Y., Huang,W., Liu,B., Shen,Y. *et al.* (2003) The International HapMap Project. *Nature*, **426**, 789-796.

Supervisor: Ian Day, Professor of Genetic and Molecular Epidemiology
(Ian.Day@bristol.ac.uk)
Philip Guthrie, Research Fellow Molecular Genetics and Bioinformatics

Title: Disentangling instances of causally and pharmacogenetically relevant genomic confounding

Outline of Project:

Background

The genes for growth hormone (*GHI*) and angiotensin converting enzyme (*ACE*) reside within a 450kb region on human chromosome 17. Research by us and others has demonstrated associations between polymorphisms in the *ACE* (>1000 papers) and *GHI*¹ gene region and cardiovascular, metabolic and numerous other risk phenotypes. Since RAS pathway inhibitors are widely used, and recombinant GH is also used, the pathway inferences are potentially of pharmacogenetic significance. We have demonstrated ~20% linkage disequilibrium between the *ACE* and *GHI* genes using our own data², confirmed with HapMap data (www.hapmap.org)³. Whilst it is apparent that factors in this region influence disease risk, we have demonstrate that pathway misinference may arise due to the high levels of linkage disequilibrium². The use of HapMap data³ enables the identification of other causally and pharmacogenetically important genomic regions in which LD may cause genomic confounding.

Objectives

1. Genotype haplotype-tagging SNPs across the *ACE-GH* genomic region on chromosome 17 in British Women's Heart and Health Study and other relevant cohorts
2. Use haplotype analyses to determine which SNPs are genuinely associated with cardiovascular and metabolic phenotypes, and which are associated due to linkage disequilibrium
3. Use HapMap data to identify other regions in which genomic confounding may occur and apply the same approach as to *ACE-GH*

Design

1. SNPs will be selected using data from the HapMap project³ to select haplotype-tagging SNPs.
2. Pharmacogenetically and causally important genomic regions will be identified from the literature and LD across those regions analysed using data from the HapMap project³.
3. Genotyping work will be outsourced using cohort studies based at University of Bristol for which ethical approval is already in place.
4. Haplotype analyses for association with phenotypes will use various programs as described in previous work^{2,4}.

Techniques/approaches

The project will be principally bioinformatic and statistical, with the majority of original data being obtained by outsourcing. However, some problem SNPs may require in-house assay development and genotyping, particularly within the *GHI* region due to high sequence homology between *GHI* and four related genes.

References

1. Day IN, Chen XH, Gaunt TR, King TH, Voropanov A, Ye S, Rodriguez S, Syddall HE, Sayer AA, Dennison EM, Tabassum F, Barker DJ, Cooper C, Phillips DI. Late life metabolic syndrome, early growth, and common polymorphism in the growth hormone and placental lactogen gene cluster. *J Clin Endocrinol Metab.* 2004 89(11):5569-76.
2. Shuwen Huang, Xiao-he Chen, John R. Payne, Matt J Smith, Hugh E. Montgomery, Ian Day and Tom Gaunt Haplotype of Growth Hormone and ACE Genes, Serum ACE and Ventricular Growth: Pathway Inference in Pharmacogenetics *Pharmacogenet Genomics.* 2007 Apr;17(4):291-4.
3. The International HapMap Project. *Nature* 2003, 426: 789-796.
4. Rodriguez S, Gaunt TR, O'Dell SD, Chen XH, Gu D, Hawe E, Miller GJ, Humphries SE, Day IN. Haplotypic analyses of the IGF2-INS-TH gene cluster in relation to cardiovascular risk traits. *Hum Mol Genet.* 2004 Apr 1;13(7):715-25.

Supervisors

Tom Gaunt, Lecturer in Bioinformatics and Molecular Genetics (tom.gaunt@bristol.ac.uk)
Ian Day, Professor of Genetic and Molecular Epidemiology

Title: Evaluation of the prevalence and functionality of paucimorphic and private mutations in large epidemiological surveys for cardiovascular risk traits

Outline of Project:

Background

There are two classical genetic models for the molecular basis of common diseases. The prevailing model is based on the hypothesis that common polymorphic alleles exert small individual effects but with significant population attributable risk (common disease/ common variant, CD/CV hypothesis) (1). A contrasting model (rare disease/ rare variant, RD/RV hypothesis) assumes that rare but severe single gene mutations can cause a phenotype which shows strong familial clustering (2). In 2004 we proposed an intermediate genetic model: sequence changes at an intermediate frequency [termed “paucimorphisms” (3)] may exist and may have a moderate effect. We have developed a mutation-scanning approach suitable for whole population screening for unknown mutations and have published theoretical and observational evidence supporting the paucimorphism hypothesis (3-5), including the identification and analysis of paucimorphisms in the LDLR and MC4R genes. However, the full population spectrum of rare, paucimorphic, severe, moderate (forme fruste), and silent mutations and effects is largely unknown. The definition of this type of variation in key genes influencing cardiovascular (CV) risk will facilitate translation to public health in the form of disease prediction in CV risk.

Of particular interest are four genes influencing CV risk: *APOB*, *F5*, *PCSK9* and *CYP2A6*. *APOB* (in particular the mutation R3500Q), is involved in familial ligand-defective apoB (FDB) (6). Factor V Leiden occurs due to a single point mutation on the *F5* gene, and is an inherited condition which predisposes affected individuals to thrombosis (7). *PCSK9* has emerged as a potential target for lowering plasma LDL cholesterol levels, with mutations in this gene associated either with hypercholesterolemia or with hypocholesterolemia (8). We have described associations between an allele (160H) of *CYP2A6* and the likelihood of quitting smoking (9), and, in a subsequent study, we found results suggesting that *CYP2A6* haploinsufficiency increases likelihood of continuing smoking in teenagers (10).

Objectives of the PhD

- 1.- To determine the prevalence of paucimorphic and private mutations in candidate genes for cardiovascular risk from large epidemiological surveys available in Bristol.
- 2.- To determine the functionality of paucimorphic and private mutations in order to infer their role in disease causation and their translational value in form of disease prediction.

Design

Four candidate genes for cardiovascular risk, *APOB*, *F5*, *PCSK9* and *CYP2A6*, will be analysed in large cohorts available in Bristol, including ALSPAC.

Techniques and approaches

The scanning of unknown mutations will be performed using a high-throughput mutation scanning technique (meltMADGE) developed in our laboratory. Paucimorphisms and other variants detected by meltMADGE will be confirmed by sequencing. Direct assays of specific variants will be performed using a Light-Typer instrument, a liquid-phase, fluorescence-based, melting-curve analysis instrument.

Genotype-phenotype analyses will include descriptive analyses for rare mutations and statistical association analyses for paucimorphisms. Specific functional assays for particular variants will be designed.

References

1. Lander, E.S. (1996) The new genomics: global views of biology. *Science*, **274**, 536-539.
2. Pritchard, J.K. (2001) Are rare variants responsible for susceptibility to complex diseases? *Am. J Hum Genet*, **69**, 124-137.

3. Day,I.N.M., Al-Harbi,K., Smith,M., Al Dahmesh,M., Zhang,B., Lotery,A., Cross,N.C.P., Ye,S., Eccles,D., Fox,K., Rodríguez,S. (2004) Paucimorphic alleles versus polymorphic alleles in disease causation: theory, observation and detection. *Curr. Genomics*, **5**, 431-438.
4. Alharbi,K.K., Aldahmesh,M.A., Spanakis,E., Haddad,L., Whittall,R.A., Chen,X.H., Rassouljan,H., Smith,M.J., Sillibourne,J., Ball,N.J., *et al.* (2005) Mutation scanning by meltMADGE: validations using BRCA1 and LDLR, and demonstration of the potential to identify severe, moderate, silent, rare, and paucimorphic mutations in the general population. *Genome Res.*, **15**, 967-977.
5. Alharbi,K.K., Spanakis,E., Tan,K., Smith,M.J., Aldahmesh,M.A., O'Dell,S.D., Sayer,A.A., Lawlor,D.A., Ebrahim,S., Smith,G.D., *et al.* (2007) Prevalence and functionality of paucimorphic and private MC4R mutations in a large, unselected European British population, scanned by meltMADGE. *Hum Mutat.*, **28**, 294-302.
6. Real,J.T., Chaves,F.J., Ejarque,I., Garcia-Garcia,A.B., Valldecabres,C., Ascaso,J.F., Armengod,M.E., Carmena,R. (2003) Influence of LDL receptor gene mutations and the R3500Q mutation of the apoB gene on lipoprotein phenotype of familial hypercholesterolemic patients from a South European population. *Eur J Hum Genet*, **11**, 959-965.
7. Shaw,D.M. (2006) Factor V Leiden: an overview. *Clin Lab Sci*, **19**, 218-221.
8. Lambert,G. (2007) Unravelling the functional significance of PCSK9. *Curr Opin Lipidol*, **18**, 304-309.
9. Gu,D.F., Hinks,L.J., Morton,N.E., Day,I.N. (2000) The use of long PCR to confirm three common alleles at the CYP2A6 locus and the relationship between genotype and smoking habit. *Ann. Hum. Genet.*, **64**, 383-390.
10. Huang,S., Cook,D.G., Hinks,L.J., Chen,X.H., Ye,S., Gilg,J.A., Jarvis,M.J., Whincup,P.H., Day,I.N. (2005) CYP2A6, MAOA, DBH, DRD4, and 5HT2A genotypes, smoking behaviour and cotinine levels in 1518 UK adolescents. *Pharmacogenet. Genomics*, **15**, 839-850.

Supervisors

Dr Santiago Rodriguez (santi.rodriquez@bristol.ac.uk)

Professor Ian Day

Title: Genetic analyses of individuals expressing extremely low levels of plasma protein biomarkers

Outline of Project:

Background

The genetic analysis of plasma protein biomarkers is informative for unravelling the genetic causes of human disease and for disease diagnosis. An example is the *KLK3* gene which encodes Prostate Specific Antigen (PSA). PSA is the most effective test currently available for the early detection of prostate cancer and is used for screening, diagnosis and monitoring of prostate cancer after diagnosis [1]. Intra-individual variation in PSA and its implications for early detection of prostate cancer has been previously described [2]. For example, the presence of PSA value less than 4.0 ng/mL does not guarantee the absence of prostate cancer, since up to 25% of men with the disease can have PSA levels less than 4.0 ng/mL. This variation can be due to an intraindividual variation in the measurement of PSA, that has been estimated to be a coefficient of variation of 13.1% [2]. In addition to this source of variation, there is variation in the distribution of PSA in the general population, with some individuals having very low levels. These very low levels could be due to the presence of PSA deficiencies produced by genetic defects causing a lower expression of *KLK3*. Actually, a parallelism exists for the Chorionic Somatomammotropin Hormone (*CSH*). This hormone was much used as a placental biomarker in pregnancy prior to the days of ultrasound. This led to the discovery of instances where there was then shown to be complete genetic *CSH* deficiency. In these instances, the deficiency was due to deletions in the gene [3,4]. Another similar example is analbuminemia and hypoalbuminemia. If some low PSA values in the population are through gene deficiencies, then it would represent an important cause of false negativity or insensitivity for the biomarker.

In this proposal we aim to study the genetic basis of the inter-individual variation of PSA levels at the population level. This will be relevant in understanding whether there may be individuals who will not and could not benefit from testing of PSA for screening or diagnosis. In addition, if we identify inactivating alleles, they effectively represent a natural human gene knock out which would also facilitate future study in vivo or in vitro of *KLK3* gene function. The determination of instances of inactivating alleles in homozygotes or as compound heterozygotes will open the possibility to investigate the possible presence of those alleles in heterozygotes leading to haploinsufficiency (half of usual expression level). We will also be able to relate this genetic variation with phenotypes available in the ProtecT study.

By a similar logic, this principle will be extended to a range of other plasma protein biomarkers in different clinical epidemiological contexts.

Objectives of the PhD

- 1.- To confirm the original PSA measurements in individuals with apparently very low PSA levels, in order to rule out possible measurement or other errors.
- 2.- To characterise in detail *KLK3* for major physical changes (via determination of exonic dose in relation to a reference gene).
- 3.- To perform sequence level studies of *KLK3* in order to look for inactivating mutations responsible for absent or very low expression of *KLK3* (e.g. stop codon mutations, deletions, splicing mutations).
- 4.- To extend this principle to other biomarkers

Techniques and approaches

Exonic dose experiments will be performed with a real-time PCR machine available in our laboratory as previously described [5]. In short, the SYBR Green dye is used as an intercalator in two PCR reactions, one for the target gene and another one for the control gene (b-globin). Then the Ct-value is defined as the number of PCR cycles necessary to achieve a given level of fluorescence in relation to the internal control (β -globin gene). The Ct-value is

then used to determine whether there is normal copy number, heterozygous deletion or duplication.

In order to search for genetic defects in *KLK3* leading to abnormal PSA levels, we will resequence the whole gene in all the individuals with the lowest PSA levels. *KLK3* and promoter region are ~10kb long. Amplification products will be prepared in house and sequencing will be outsourced to the company K-Biosciences

The project will also involve a significant amount of statistical genetic analyses. A number of statistical analyses will be performed in order to test for association between mutations and intermediate traits in order to investigate their functionality and their role in disease diagnosis. These include a) association analyses between genetic variants and continuous phenotypes under the dominant, codominant and recessive models (as appropriate) using standard t test or using multiple regression analyses with and without adjustments for relevant covariates by means of the statistical packages SPSS and STATA as previously described [6–8].

References

1. Lilja H, Ulmert D, Vickers AJ (2008) Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nat Rev Cancer* 8: 268-278.
2. Yan Y (2001) Intraindividual variation of prostate specific antigen measurement and implications for early detection of prostate carcinoma. *Cancer* 92: 776-780.
3. Wurzel JM, Parks JS, Herd JE, Nielsen PV (1982) A gene deletion is responsible for absence of human chorionic somatomammotropin. *DNA* 1: 251-257.
4. Simon P, Decoster C, Brocas H, Schwers J, Vassart G (1986) Absence of human chorionic somatomammotropin during pregnancy associated with two types of gene deletion. *Hum Genet* 74: 235-238.
5. Chaudhary S, Behari M, Dihana M, Swaminath PV, Govindappa ST, Jayaram S, Goyal V, Maitra A, Muthane UB, Juyal RC, Thelma BK (2006) Parkin mutations in familial and sporadic Parkinson's disease among Indians. *Parkinsonism Relat Disord* 12: 239-245.
6. Day IN, Chen XH, Gaunt TR, King TH, Voropanov A, Ye S, Rodriguez S, Syddall HE, Sayer AA, Dennison EM, Tabassum F, Barker DJ, Cooper C, Phillips DI (2004) Late life metabolic syndrome, early growth, and common polymorphism in the growth hormone and placental lactogen gene cluster. *J Clin Endocrinol Metab* 89: 5569-5576.
7. Rodríguez S, Gaunt TR, O'Dell SD, Chen XH, Gu D, Hawe E, Miller GJ, Humphries SE, Day IN (2004) Haplotypic analyses of the IGF2-INS-TH gene cluster in relation to cardiovascular risk traits. *Hum Mol Genet* 13: 715-725.
1. Rodríguez S, Chen XH, Miller GJ, Day IN (2005) Non-recombining chromosome Y haplogroups and centromeric HindIII RFLP in relation to blood pressure in 2,743 middle-aged Caucasian men from the UK. *Hum Genet* 116: 311-318.

Supervisors

Dr Santiago Rodriguez (santi.rodriquez@bristol.ac.uk)

Professor Ian Day (ian.day@bristol.ac.uk)

Title: Meta-analysis of variants in the vitamin D receptor gene and breast cancer risk

Background: Epidemiological evidence suggests that vitamin D from sunlight and diet may be inversely associated with breast cancer incidence. $1,25(\text{OH})_2\text{D}_3$, the physiologically active metabolite of vitamin D binds to the vitamin D receptor (VDR) to modulate the rate of cell proliferation. As with many tissues, the breast, both normal and malignant, expresses VDR. The *VDR* gene is polymorphic at several sites; the *BsmI*, *ApaI*, and *TaqI* polymorphisms are in strong linkage disequilibrium in Caucasians. Association between the VDR gene polymorphisms and cancer development has been suggested by several studies, however, the relationship has not been confirmed by all studies. Meta-analysis can be used as a tool to combine existing evidence and to determine whether a particular polymorphism might be important for a disease of interest. The aim of this project is to carry-out a meta-analysis of the association between VDR receptor variants and breast cancer risk.

Objectives:

- 1) To review the literature on VDR polymorphisms and breast cancer risk.
- 2) To critically appraise the published studies,
- 3) To carry-out a meta-analysis of the existing studies using inclusion criteria,
- 4) To draw conclusions on the role of the VDR polymorphism in breast cancer susceptibility.

Methods:

- 1) Systematic review of the literature to identify all studies, which have looked at the association between breast cancer risk and variants in the VDR gene.
- 2) Extract relevant data from these studies and synthesise tables of evidence.
- 3) Meta-analyse data from existing studies
- 4) Write paper for publication.

Supervisor

Sarah Lewis

Title: Fine mapping of the association between *osterix* genotypes and bone mineral density in children

Background

Osterix is a transcription factor which acts as a 'master switch' in the differentiation of osteoblasts, the cells responsible for making bone (1). Recently, we identified four single nucleotide polymorphisms (SNPs) in the vicinity of this gene which are related to bone mineral density (BMD) in the Avon Longitudinal Study of Parents and Children (ALSPAC), both in a discovery set analysed as part of a genome-wide association study, and in a replication set (paper submitted for publication). The same markers were also associated with BMD in a cohort of older adults from Australia selected for high or low BMD. Identifying the functional variant responsible for this association, by performing fine mapping studies, is of interest since this may yield new insights into the molecular regulation of osteoblast differentiation, which may provide novel drug targets for osteoporosis therapies which act by stimulating bone formation.

Objectives

1. To select and genotype additional candidate SNPs within the *osterix* gene.
2. To identify the SNP or haplotype most likely to be responsible for the association between *osterix* markers and BMD.

Design

1. Additional SNP markers will be selected according to patterns of linkage disequilibrium within the International HapMap.
2. The strength of associations with BMD in ALSPAC will be compared between *osterix* SNPs genotyped previously, and additional markers genotyped as part of this study.

Techniques/approaches

In silico analysis will be performed on human hap map data in order to identify haplotypes and linkage patterns across the *osterix* gene. Additionally, candidate markers will be selected using bioinformatics analyses to evaluate possible functionality. Genotyping of additional *osterix* markers will be outsourced to Kbiosciences. Associations between all available *osterix* markers and BMD will be analysed using both single markers and imputed haplotypes.

References

1. Nakashima K, Zhou X, Kunkel G, Zhang Z, Deng JM, Behringer RR, de Crombrughe B 2002 The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell* 108(1):17-29.

Supervisors

Jon Tobias, Professor of Rheumatology

Ian Day, Professor of Genetics and Molecular Epidemiology

David Evans, Senior Lecturer in Biostatistical Genetics

Title: The use of Genomewide data for the design and undertaking of Mendelian randomisation experiments to dissect potentially causal pathways to common disease.

Outline of Project:

Background

The suite of genomewide data resources that are becoming available from both publicly available sources and from the Avon Longitudinal Study of Parents and Children (ALSPAC) is providing a burgeoning source of information for the guidance and undertaking of Mendelian randomisation experiments^{1,2}. These analyses require robust associations between genetic variation and intermediate phenotypes of interest (or risk factors) in efforts to assess the nature of observational relationships between potentially modifiable environmental risk factors to disease and health outcomes. These Analyses are increasingly being performed at a number of biological levels and include the assessment of biological relationships within potentially causal pathways involved in the aetiology of common disease. Recent examples of this approach have been seen in negative and positive guises, indicating instances where the use of genetic proxy measures for intermediate phenotypes has both confirmed and drawn question around the involvement of observationally highlighted risk factors and metabolic disorders³ (add Freathy 2007). One key advance in these investigations has been the provision of reliable associations between genetic variation and intermediate phenotypes, information that genomewide analyses are increasingly providing.

Objectives of the PhD

From available GWA analyses (a large proportion of which have been published in the early stages of 2007⁴⁻⁹) it is possible to consider phenotypic associates for their suitability for application to Mendelian randomisation (MR) experiments. Such experiments would consider the possibility of re-assessing existing observational associations between potentially environmentally modifiable risk factors and health outcomes of interest. These analyses would follow those underway (concerning the *FTO* locus BMI and both metabolic intermediates and circulating C reactive protein. Such a project would also be able to consider the possibility of expanding on associations found between genomic variation and novel phenotypes in the ALSPAC sample set and the application of this to larger replicate cohorts which may make appropriate the application of MR approaches.

Design

Data from both internal genomewide screening within the ALSPAC cohort and from publicly available resources (for example the Wellcome Trust Case Control Consortium - <http://www.wtccc.org.uk/>) will be assessed for validity and consistency for association with both disease end points and intermediate phenotypes of interest. Robust associations with components of phenotypes available within the ALSPAC cohort will be employed in order to design Mendelian randomisation experiments in order to (i) assess the potentially causal role of these variables in both end point disease and anthropometric features and (ii) dissect the nature and direction of biological pathways involved in the aetiology of both biological disorders and common phenotypic characteristics. These results would be designed in order to feed into further examination of these pathways and future functional examination.

Techniques and approaches

The overall structure of this investigation will follow a logical progression through the application of novel genomewide data (and findings) to Mendelian randomization frameworks. The genotyping and analysis of genetic/phenotypic data in ALSPAC (and other cohort such as the British Women's Heart and Health Study where appropriate) will follow a collection and review of genomewide association analyses results from public and internal resources. This will lead to the application of Mendelian randomization to instances where genetic variants exhibit properties pertaining to proxy markers for potentially modifiable or pathway based risk factors. Both bioinformatics and mendelian randomization and the processes and methodologies that they require, will be essential approaches for the undertaking of this work. Main processing analytical techniques will include the processing and basic analysis of both genotypic and phenotypic data at a population cohort scale. Further methods will require the application of more advanced methodology (with as instrumental variable analysis – reference) in order to formally apply the paradigm of Mendelian

randomization. Other techniques for the interpretation and application of genomewide data results will include meta-analyses and both linear and logistic regression analyses. The use of the analytical software packages Haploview (<http://www.broad.mit.edu/mpg/haploview/>), PLINK (<http://pngu.mgh.harvard.edu/~purcell/plink/>), STATA or equivalents will be essential.

References

1. Davey Smith, G. & Ebrahim, S. 'Mendelian Randomisation': can genetic epidemiology contribute to understanding environmental determinants of disease? *International Journal of Epidemiology* **32**, 1-22 (2003).
2. Davey Smith, G. & Ebrahim, S. Mendelian Randomisation: prospects, pitfalls and limitations. *International Journal of Epidemiology* **33**, 30-42 (2003).
3. Timpson, N.J. et al. C-reactive protein and metabolic syndrome: Mendelian randomisation suggests associations are non-causal. *Lancet* **366**, 1954-1959 (2005).
4. Sladek, R. et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. [see comment]. *Nature* **445**, 881-5 (2007).
5. Steinthorsdotir, V. et al. CDKAL1 influences insulin response and risk of type 2 diabetes. *Nature Genetics* **AOP**(2007).
6. Helgadottir, A. et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491-3 (2007).
7. Consortium, T.W.T.C.C. Genome-wide association study of 14, 000 cases of seven common diseases and 3, 000 shared controls. *Nature* **447**, 661-678 (2007).
8. Salonen, J.T. et al. Type 2 Diabetes Whole-Genome Association Study in Four Populations: the DiaGen Consortium. *American Journal of Human Genetics* **81**, AOP (2007).
9. Diabetes Genetics Initiative of Broad Institute of Harvard and Mit, L.U.a.N.I.o.B.R. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331-6 (2007).

Supervisors: Nic Timpson, David Evans, George Davey Smith

Title: The coordinated analysis of genomewide genotype data, dense transcriptomic data and phenotypic data within a sample of the Avon Longitudinal Study of Parents and Children.

Outline of Project:

Background

The new era of genomic exploration in light of common disease risk has been the result of vast improvements in the scale of available data¹. No one level, genetic chip technology has exponentially increased the availability of genotype data whilst reducing the effective cost per information point respectively. Whilst accurate phenotyping at the molecular and anthropometric level still remains relatively expensive, large-scale data collection approaches are now available for a crucial level of data one step from basal genomic code. This represents transcription data and is now represented by collection strategies such as the Illumina 46k expression chip. The Avon Longitudinal Study of Parents and Children (ALSPAC) represents a unique opportunity in that data is now available at three crucial levels with respect to the elucidation of aetiological relationships with respect to both disorder and normal physiology. ALSPAC is a large scale pregnancy cohort for which 2000 individuals have been scored for (i) extensive end point phenotypes (endophenotypes), (ii) genomewide genetic variation (Illumina) and (iii) the 46k transcript chip². These sources of data offer a series of analytical possibilities which may enhance the understanding of relationships between genetic perturbations and disorder/anthropometric traits, but also in the dissection of possible biological pathways involved in biological systems mediating both homeostasis, development and potential disease predisposition.

Objectives of the PhD

Such investigation should consider the possibility of both basic genotypic variation and phenotypic correlation (at the level of exogenous measurement, or endophenotypes and the transcriptome), but also that concerning the potential influence of regulatory mechanisms which may link genetic variation to differences in ultimate phenotype status, for example siRNA regions and binding regions³⁻¹¹. Such a triangulation effort should aim to comment on the likely impact of and existence of functionally manifest variations in the human genome. Importantly, analyses within such a framework will be able to address both the relationships between genetic variation and phenotypes of interest, but also the co-relation between differential patterns in the transcriptome.

Design

The overall design of this work will involve the collection and preparation of data from three sources. These will be derived from the ALSPAC cohort and will comprise data on (i) genomewide data on 2000 individuals (ii) extensive phenotypic data and (iii) 46-7k transcript chip data on the same 2000 ALSPAC individuals. All of these aspects will require both quality control measures and diagnostic examination before application to further analyses. In particular, the design and implementation of quality control methods for transcriptomic data is an area which will require extensive development, such that will benefit from collaborative activity with groups such as MOLPAGE (www.molpage.org). From the development and submission of this data to quality control measures, the next stage of analysis will involve the bringing together of genetic data, endophenotype data and transcriptomic data.

Techniques and approaches

Other than the development of suitable thresholds and screening criteria for the data employed, one of the main technical aspects for the development of this work will be in the development of suitable approaches for the analysis of large-scale data sets from three sources (phenome/genome/transcriptome). Past basic pairwise investigation of the relationships between these, the application of principle components methods offers potential insight into the assessment of these data sources simultaneously. In this way, the regression of principle components derived separately from each of the analysis strata on outcomes of interest, the regression of collective principal components from all sources on the outcomes of interest and

the analysis of potential interaction terms from derived from the first components from each source may provide three obvious avenues within which to explore this data. These approaches will require the development of analytical capability for basic epidemiological methods and data sets, for genomewide analyses and data sets, for the processing of transcriptomic data and for the further analysis of relationships between these aspects. This is relatively novel use of such resources and would require a certain degree of computational experience.

References

1. Wadman, M. Genome miners rush to stake claims. *Nature* **447**, aop (2007).
2. Stranger, B.E. et al. Genome-wide associations of gene expression variation in humans. *PLoS Genetics* **1**, e78 (2005).
3. Chen, P.Y. & Meister, G. microRNA-guided posttranscriptional gene regulation. *Biological Chemistry* **386**, 1205-18 (2005).
4. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. *Nature* **431**, 343-9 (2004).
5. Obernosterer, G., Meister, G., Poy, M.N. & Kuras, A. The impact of small RNAs. Microsymposium on small RNAs. *EMBO Reports* **8**, 23-7 (2007).
6. Patel, D.J. et al. Structural biology of RNA silencing and its functional implications. *Cold Spring Harbor Symposia on Quantitative Biology* **71**, 81-93 (2006).
7. Ambros, V. The functions of animal microRNAs. *Nature* **431**, 350-5 (2004).
8. Ambros, V. & Chen, X. The regulation of genes and genomes by small RNAs. *Development* **134**, 1635-41 (2007).
9. Hannon, G.J. & Rossi, J.J. Unlocking the potential of the human genome with RNA interference. *Nature* **431**, 371-8 (2004).
10. Mello, C.C. & Conte, D., Jr. Revealing the world of RNA interference. *Nature* **431**, 338-42 (2004).
11. Lippman, Z. & Martienssen, R. The role of RNA interference in heterochromatic silencing. *Nature* **431**, 364-70 (2004).

Supervisors: Nic Timpson; David Evans; Tom Gaunt; George Davey Smith

Title: Maternal Phenylalanine and offspring development: Mendelian randomisation approach.

Phenylketonuria (PKU) is a classic Mendelian trait, that is screened for in new born babies as treatment through phenylalanine restriction prevents the developmental abnormalities – in particular impaired cognitive intellectual development – from appearing. Like most apparently classical Mendelian traits it's genetic background is complex¹, with a large number of mutations causing the syndrome. Independent of offspring PKU, maternal untreated PKU leads to developmental abnormalities in offspring through an in utero influence², in what is probably the first example of what would now be referred to as intergenerational Mendelian randomisation³. Treatment of maternal PKU through phenylalanine restriction prevents the gross offspring abnormalities.

Despite its molecular genetic complexity PKU is now a reasonably well understood condition, both in its within and between generational affects. What is uncertain, however, is whether milder forms of high phenylalanine among the mother – either genetically or environmentally induced – have consequences for the offspring. Milder genetic forms exists (known as “mild PKU” or “non-PKU mild hyperphenylalaninemia”) (<http://www.PAHdb.mcgill.ca>). A small follow-up study is cited as demonstrating little effect of maternal mild PKU on offspring outcomes, but it actually shows a strong correlation of maternal phenylalanine levels with offspring head circumference and adverse effects on all parameters measured (e.g. seven IQ points lower)⁴. Use of genetic variants in the mother related to these milder conditions can be utilised to study the effects in offspring of a modifiable potential cause of suboptimal outcomes. This PhD project will explore this issue.

Projects steps

- 1 Review current evidence on the genetic basis of mild non-PKU and nonphenylalanine hyperphenylalaninemia.
- 2 Review evidence that phenylalanine containing foods and drinks have developmental influences either when consumed by mothers or by offspring.
- 3 In the ALSPAC study investigate the association of maternal consumption of phenylalanine containing foods and drinks and offspring developmental outcomes, from birth weight through to intellectual performance.
- 4 Select paucimorphic or polymorphic SNPs in the phenylalanine gene that may influence phenylalanine levels through gene by dietary intake interaction.
- 5 Utilise Mendelian randomisation techniques adapted to the gene by environment interaction situation⁵ to and investigate this issue.

Supervisors: George Davey Smith, Sarah Lewis & Debbie A Lawlor

¹ Scriver CR and Waters PJ. Monogenic traits are not simple: lessons from phenylketonuria. *Trends Genet.* 1999;15:267–272.

² Levy HL, Waisbren SE. Effects of untreated maternal phenylketonuria and hyperphenylalaninemia on the fetus. *NEJM* 1983; 309:1269-1274.

³ Davey Smith G. Commentary. Capitalizing on Mendelian randomization to assess the effects of treatments. *James Lind Library* 2007.

⁴ [Levy HL](#), [Waisbren SE](#), [Güttler F](#), [Hanley WB](#), [Matalon R](#), [Rouse B](#), [Trefz FK](#), [de la Cruz F](#), [Azen CG](#), [Koch R](#).. Pregnancy experiences in the woman with mild hyperphenylalaninemia. *Pediatrics.* 2003; 112:1548-52.

⁵ Davey Smith G. Mendelian randomization for strengthening causal inference in observational studies: application to gene by environment interaction. *Perspectives on Psychological Science* 2009, in press.

Title: Analysis of environmental and genetic contributions to scoliosis in an Indian population.

Outline of Project:

Objectives

1. To use a novel method to assess the presence of absence of scoliosis from whole body DXA scans for the participants of the Hyderabad Nutrition Trial
2. To assess the prevalence of scoliosis within this study
3. To assess genetic and environmental predictors of scoliosis, and their interaction, in this study

Background

Scoliosis is a condition in where a person's spine develops a side-to-side curve. Although, it is a complex three-dimensional deformity, the spine of an individual with a typical scoliosis may look more like an "S" or a "C" than a straight line. It most commonly starts between aged 10 years and skeletal maturity. It is typically classified as either congenital (caused by vertebral anomalies present at birth), idiopathic (cause unknown, sub-classified as infantile, juvenile, adolescent, or adult according to when onset occurred) or neuromuscular (having developed as a secondary symptom of another condition, such as spina bifida, cerebral palsy, spinal muscular atrophy or physical trauma). Later in life scoliosis can develop due to degenerative change. Scoliosis is not always a benign structural abnormality, as it can cause severe back pain, restrictions on social activity and on participation in work.

Assessment of scoliosis

Currently, gold standard assessment of scoliosis is achieved through x-ray examination, however this is not ideal for multiple examination and is not always available in large numbers of individuals. Scoliometers can be employed to assess the degree of spinal curvature through a non-invasive, nurse administered, physical examination (the Adams forward bending test). Unfortunately, these measurements are difficult to repeat in a standardised manner and often yield unreliable assessment of scoliosis status across large numbers of participants. Novel automated methods of measuring the Cobb angle for scoliosis diagnosis have been developed in adults that use spinal dual energy X-ray absorptiometry (DXA) taken with the patient lying flat, which has a much reduced radiation exposure compared with spinal radiographs, and correlates well with traditional methods of measuring the Cobb angle ($R^2=0.998$). DXA measures also provide an indication of size, site and direction of curve, as well as measures of local bone quality and density without the large radiation exposure of traditional spinal radiograph. In light of this, members of the musculoskeletal centre at Southmead hospital, Bristol have been developing more accurate measures of spinal deformity and applying these to the Avon Longitudinal Study of Parents and Children (<http://www.bristol.ac.uk/alspac/>).

Potential causes of scoliosis

Current knowledge about the causes of the initiation or induction of the scoliotic curve is scarce, particularly outside of high income countries. Few epidemiological studies have identified clear predictors of risk of scoliosis, although girls appear to be more affected than boys. One potential determinant of curve induction that is of great interest at present is a genetic cause, as twin studies indicate that scoliosis may have a genetic component and many potential genetic areas of interest have been identified. However, it is likely, that as for the majority of human disease, scoliosis is due to a complex interplay between environmental and genetic influences. Bone mass, hyperflexibility and neuromotor or growth abnormalities may also contribute towards the development of scoliosis. In addition to improving our knowledge of the epidemiology of scoliosis, the discovery of novel risk factors may lead on to the development of new management techniques. The discovery of accurate prognostic markers

may help decisions about timing of surgery or other interventions.

Hyderabad Nutrition Trial

The prevalence and causes of scoliosis can be assessed within an Indian population using the new methods in the assessment of scoliosis. In collaboration with the London School of Hygiene & Tropical Medicine, we at the MRC CAiTE Centre (University of Bristol) are undertaking a large study in India based on participants of a trial of nutritional supplementation called the Hyderabad Nutrition Trial. In this study, over 2000 women were randomised to receive supplementation during pregnancy and the first five years of their child's life, or to no supplementation. These children were examined at the age of 14-16, and they are now aged 18-21. We are currently undertaking the follow-up of this cohort and all participants are being undergoing whole body DXA scans, as well as being examined for multiple potential correlates (objective and subjective measures of physical activity, work status, quality of life and depression, spirometry, anthropometry, growth, including puberty, bone and muscle parameters, DNA). We anticipate that we will examine 1300 participants from this cohort by the end of 2010, both from supplement and non-supplement arms of this Indian study. The availability of DXA data presents a superb opportunity to assess the presence of scoliosis in the participants and to investigate the influence of environmental and genetic factors on Scoliosis prevalence. It will also be possible to retrace these subjects for further examination.

Plan of investigation

The broad aim of this PhD project is to provide information on the prevalence and causes of scoliosis in a young Indian population.

We anticipate that the student will use data from the Hyderabad Nutrition Trial which will already be available in the department. He/she will supervise the grading of the DXA scans for the presence or absence of scoliosis, and will undertake a validation study of the DXA measures. He/she will then undertake data analyses to assess the prevalence and predictors, both genetic and environmental in this population. The PhD will be housed at the MRC CAiTE Centre, University of Bristol. The supervisors collectively are very familiar with the study, and have the requisite experience in genetic and environmental epidemiology and the assessment of scoliosis.

Supervisors: Nicholas Timpson; Emma Clark; Hannah Kuper