

**Department of Social Medicine**



**Research student opportunities 2009**

**University of Bristol**  
**Department of Social Medicine**

The Department of Social Medicine is a leading centre for research and teaching in epidemiology, public health and health services research. It was awarded the top 5\* grade in the 2001 Research Assessment Exercise. The primary characteristic of research in the Department is that it is collaborative and multi-disciplinary. The Department's staff are located either in Canynge Hall (39 Whatley Road, BS8 2PS) or Oakfield House (Oakfield Grove, BS8 2BN).

The Department houses the MRC Centre for Causal Analyses in Translational Epidemiology (CAiTE), Directed by Professor George Davey Smith with Deputy Directors Professor Debbie Lawlor and Professor Ian Day; the premier birth cohort study, ALSPAC (Avon Longitudinal Study of Parents and Children). In addition, the Department hosts the MRC ConDuCT (COllaboration and iNnovation in DifficUlt and complex randomised Controlled Trials) methodology hub, led by Professor Jane Blazeby; and, jointly with Primary Care, the BRTC (Bristol Randomised Trials Collaboration) UKCRC/NCRI-accredited trials unit. From January 2009, the Department will host (with the Universities of Cardiff and Swansea) the UKCRC DECIPHer (Development and Evaluation of Complex Interventions for Public Health Improvement) Centre, led by Professor Rona Campbell.

Undergraduate and postgraduate teaching programmes provide training and career development for undergraduate medical students, public health trainees, clinicians, and research staff. The postgraduate Short Course Programme is a popular source of intensive short courses in epidemiological and HSR research methods and skills. There are weekly seminars in term-time. For the latest news from the Department, please see [here](#). Academic staff in the Department are leaders in their fields and have extensive national and international research collaborations, with several providing health policy advice for government organisations and international bodies. Within its subjects the Department is seen as one of the leading departments in the world. The skills of a variety of health care researchers including statisticians, epidemiologists, geneticists, sociologists, psychologists, anthropologists, health economists, and public health physicians, are all available and contribute to the excellent working environment in which to undertake interdisciplinary research and teaching. The Department is keen to attract graduates from all the above disciplines to carry out postgraduate research.

## **Research**

The Department has an exceptional set of resources for research and a wide range of opportunities exist for students wishing to pursue a higher degree or a career in the fields of epidemiology and health services research. In addition to its high achieving and enthusiastic academic staff, there is also a collection of historical and contemporary cohort studies, many with stored biological samples and linked clinical, sociodemographic and epidemiological data. These cohorts provide opportunities to explore the aetiological aspects of life-threatening diseases across the lifecourse, and their social and health care impact. Our research is strengthened by international collaborations with researchers in Sweden, Norway, Belarus, India, Sri Lanka, USA, Canada, Southern Africa, the Netherlands and Australia. While our research can be classified crudely as “Epidemiology” and “Health Services Research”, much work crosses these artificial boundaries. In particular, large studies such as ALSPAC and ProtecT provide opportunities for intra- and inter-disciplinary projects.

## ***Epidemiology***

The Department’s epidemiology research programme is concerned with life-course, clinical, genetic and molecular epidemiology, bioinformatics, human gene mapping and statistical methods.

1. Lifecourse epidemiology is the investigation of exposures acting at different stages of the life course that may contribute to risk of disease in adulthood. Studies are based in a wide range of historical and contemporary cohort studies and focus on cardiovascular disease, cancer, insulin resistance, suicide and schizophrenia; or on biological mechanisms such as IGFs mediating associations.
2. Clinical epidemiology. Our highly influential work investigating the prognosis of HIV-infected patients starting antiretroviral therapy based on collaborative analyses of HIV cohort studies will expand with newly established collaborations in developing countries, in particular sub-Saharan Africa. Respiratory and nutritional programmes will continue to expand in ALSPAC.
3. Genetic and molecular epidemiology examines ways in which genotypes can aid understanding of the relationships between life-course exposures and adult diseases, to develop methods for dissemination and introduction into practice of genetic epidemiology findings, and to use the principles of Mendelian randomisation to examine the role of environmental exposures for obesity, insulin resistance, diabetes, CHD and common cancers. Funding has already been achieved to allow cell lines to be set up in ALSPAC with the potential to examine imprinting patterns and carry out gene expression studies.
4. Human Gene mapping. A major focus of the department is to map genetic variants that contribute to variation in complex traits and diseases. The department is heavily involved in the genome-wide association analysis of several diseases including prostate cancer, ankylosing spondylitis and osteoporosis as well as analysis of the thousands of phenotypes measured within the ALSPAC cohort.
5. Statistical methods. A major strength of the Department is its internationally renowned statistical expertise and the integration of this expertise with other disciplines in the Department. Our research strategy includes contributions to development and application of statistical methods for dealing with missing data, repeated measures and other hierarchical data structures, and evidence synthesis. We will also extend statistical methods for examining different life course models of relationships between a large number of potentially important exposure variables, and will develop and apply statistical methods for causal inference, including the instrumental variables methods that will underpin analysis of Mendelian randomisation studies.

### ***Health Services Research (HSR)***

This may be defined as research into all aspects of health technologies and the delivery of health care, and is becoming increasingly more important as it becomes necessary to have reliable information on which to base decisions on the allocation of limited resources, decisions which may well have an ethical component. The aim of our HSR programme is to conduct cutting-edge methodological, theoretical and applied research examining the effectiveness and efficiency of health care interventions, health services and healthcare policy, and to assess the impact of these on the experiences of patients and the health of the public. Programmes of work are wide-ranging (see below), and candidates should also look on the departmental website and at the enclosed research interests of staff for further details.

1. Methodological development and implementation of phase III randomised controlled trials - investigating and developing methods to improve the commissioning, design, and conduct of RCTs. This includes, for example, qualitative research to improve recruitment and retention of participants, novel methods to maintain the integrity of ongoing trials, statistical analyses to accommodate missing data and estimate treatment effects amongst compliers, economic evaluations that accommodate missing data and optimise statistical power, methods of meta-analysis, and the use & analysis of patient reported outcomes. This work based in the Departments of Social Medicine and Community Based Medicine is brought together by the MRC-funded ConDuCT trials methodology hub (Collaboration and innovation for difficult-to-door complex randomised controlled trials), and the UKCRC & NCRI-accredited BRTC (Bristol Randomised Trials Collaborative). The ConDuCT hub is a new organisation opening in March 2009 that will bring together a number of potential supervisors in the two departments, will have funding for five four-year Ph.D. studentships (details in the topic guide), and will encourage links with methodologically challenging trials being conducted across the South West region and South Wales as well as elsewhere in the UK. The BRTC has established links with around 20 ongoing or recently completed trials, with the potential for collaboration in using information from these trials as examples for methodological work.
2. Population needs and outcome assessment - investigating inequalities in health and access to and utilisation of health services; finding appropriate ways of capturing the values, preferences and experiences of people in health care contexts so that they can be integrated with other data; modelling disease trends and future healthcare needs; rationing; needs assessment; developing outcome measures; individualisation of care; evaluating health care interventions and outcome
3. Implementation of research findings - understanding the implementation of research findings and best practice through behavioural change; exploring the factors that determine professional and patient behaviour; exploring barriers to and facilitators of the delivery of health care and its utilisation
4. Diagnosis, decision-making and improving health care - investigating and improving the evidence-base for diagnostic procedures; understanding the diagnostic process and clinical decision-making; patient involvement in clinical decision-making
5. Communicable diseases – investigating the epidemiology, prevention and treatment of sexually transmitted diseases: HIV/AIDS, chlamydia; meningococcal disease and hepatitis B and C

## Research students

The Department has a well-established policy for research students, covering training, expectations of students and advisers, reviews, and University and Department facilities. Within the Department, each student has a desk, filing cabinet, a networked PC with email facilities, access to a 'phone, and access to a high performance PC. Each student has two supervisors to give guidance about the research, and there is a Director of Graduate Studies, Professor Richard Martin, who has overall responsibility for postgraduate students. The Department provides research training through a wide range of short courses, for example, several courses in medical statistics, basic and advanced epidemiology, genetic epidemiology, health economics, qualitative research methods, questionnaire design, meta analysis, study design, and data management (see [www.epi.bris.ac.uk/shortc/shortc.htm](http://www.epi.bris.ac.uk/shortc/shortc.htm)).

Seminars are held weekly in term time (see [www.epi.bris.ac.uk/seminar/socmedseminar.htm](http://www.epi.bris.ac.uk/seminar/socmedseminar.htm)).

There is an active research students' training and support group, which meets monthly in term time, with various sessions. There are currently 35 PhD students.

## Research studentships

The following studentships are currently available:

**University of Bristol Scholarships.** For further information please see the website at: [www.bristol.ac.uk/studentfunding/home\\_pg/schols.html](http://www.bristol.ac.uk/studentfunding/home_pg/schols.html)

**Wellcome Trust 4 year PhD studentships:** Students interested in molecular, genetic or lifecourse epidemiology should also view our [Wellcome Trust 4-Year PhD Programme](#). The aim of this 1+3 Programme is to equip epidemiologists with the technical skills and training to be able to exploit rapidly developing new technologies in molecular and genetic sciences within population-based cohorts in ways that could significantly improve our understanding of causal pathways that lead to disease and its progression.

**MRC CAiTE 4 year PhD studentships:** 4 year PhD studentships with a specific focus on causal analyses and translational epidemiology are available through the MRC funded Centre for Causal Analyses in Translational Research (CAiTE), which is also based in the Department of Social Medicine. Please contact Professor Debbie Lawlor for further information ([D.A.Lawlor@bris.ac.uk](mailto:D.A.Lawlor@bris.ac.uk))

Students interested in the application of their area of expertise to the commissioning, design or conduct of phase III randomised controlled trials (RCTs) will be interested in the five four-year Ph.D. studentships we have in this area, funded jointly by the MRC and the University of Bristol as part of the [ConDuCT](#) programme. In the first year students will attend courses from Social Medicine's popular short course programme (<http://www.epi.bris.ac.uk/shortc/shortc.htm>), and elsewhere, to build a comprehensive knowledge of RCT design and conduct. The first year will also provide the opportunity to conduct small studies in different areas of trials methodology, allowing students to confirm their area of interest, build their research skills, formulate a detailed research plan for the subsequent three years, conduct feasibility studies of intended research procedures, and meet and work with theme leads within the Hub.

Students must be defined as "home" for fee purposes to be eligible to apply for many of these award and information and guidance on how this classification is reached can be found at <http://www.bris.ac.uk/academicregistry/fees/class.html>

**University of Bristol Overseas Centenary Postgraduate Research Scholarships:** The Higher Education Funding Council for England (HEFCE) has announced it will no longer fund the Overseas Research Scholarship Award Scheme (ORSAS) for new international students. However, the University will be offering a limited number of scholarships for the 2009/10 academic year only. For further information please see the website at: [http://www.bristol.ac.uk/studentfunding/overseas\\_pg/cent\\_overseas\\_schols.html](http://www.bristol.ac.uk/studentfunding/overseas_pg/cent_overseas_schols.html)

The Department also hosts an MRC doctoral training core (ten four year [1+3] PhD studentships, recruitments 2009 and 2010) applying computational, mathematical and statistical approaches to biomedical questions. Within this programme, which with other research nucleates Bristol Centre for Systems Biomedicine (BCSBmed), students from engineering, computational and mathematical disciplines will be broadly trained in applications to biomedical research emphasizing systems approaches. BCSBmed is interdisciplinary between both of the medical Faculties, Mathematics and Statistics (Science) and Engineering Maths. BCSBmed complements both the EPSRC-funded Bristol Centre for Complexity Sciences (BCCS) and also complements two other doctoral training cores (in Neurosciences and Cell Imaging) in Bristol funded by MRC in 2008.

**Deadlines:** Deadlines and an internal timetable for the above schemes are still under discussion but it is likely that the deadline for submission of Postgraduate Application forms for students wishing to be considered for the various scholarships will be **around 6 February 2009**.

### **Applications**

Applications are invited from graduates with good honours degrees (2i or higher). Research experience or a Masters degree in a related area will be advantageous. Applications should be made using the enclosed application form, an academic CV, a letter indicating your topic of interest and a research proposal. You must include a research proposal comprising at least two A4 pages (see below).

Outlines of topics are provided including the name(s) of the potential supervisors adviser(s). The enclosed outlines offer guidance about the content of the postgraduate research, but these are only brief outlines.

**When you apply, you must include a proposal for the research you would like to undertake, describing the particular methods you would employ and a timetable for the research. We are not expecting a definitive proposal – we are interested in your ideas for transforming these outlines into research for a PhD.**

Your proposal should consist of 2 to 4 sides of A4. You are encouraged to discuss your ideas with the named supervisor(s). The proposal, along with your application form/CV will be used for shortlisting.

Candidates may also apply to study a topic of their own choice. The research areas of members of staff are included with this information and candidates are encouraged to discuss their ideas with a suitable prospective supervisor.

For further information about specified topics, please contact the supervisor named. For general information about research studentships (or if you cannot get through to supervisors), contact Professor Richard Martin, 0117-928 7321, email:Richard.Martin@bristol.ac.uk

Please return applications to Susie Potts, Department of Social Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS (0117 928 7274).

Visit our web page <http://www.epi.bris.ac.uk>

## SUMMARY OF TOPICS

	<b>Supervisor</b>	<b>Title of project</b>
	<b>1</b> David Gunnell	Modeling suicide trends over time and between countries
new	<b>2</b> Chris Metcalfe	Investigating and developing methods to improve the commissioning, design, and conduct of RCTs. (The MRC ConDuCT Trials Methodology Hub).
	<b>3</b> Alison Heawood	Using qualitative methods at the end of randomised controlled trials to evaluate the utility and implementation of trial results
new	<b>4</b> Mona Jeffreys	Determinants and sequelae of sexual and risk behaviours during adolescence
	<b>5</b> Rona Campbell, John Macelod, Ainsley Newson, Catherine Heeney	Ethical aspects of epidemiological research on young people involving linkage to routine individual data.
new	<b>6</b> David Gunnell	Secular trends and intergenerational influences on the suicidal thoughts in the Young Hunt study
	<b>7</b> Dr Matthew Hickman, Professor Rona Campbell	Quantitative and qualitative assessment of injecting risk and drug using networks: developing better behavioural surveillance and effective transmission models of BBV transmission among IDU
	<b>8</b> Kate Tilling and Matthew Hickman	Development and application of methods to estimate prevalence of injecting drug use and other marginal populations
	<b>9</b> Dr John Henderson, Professor Jonathan Sterne	The origins of chronic obstructive pulmonary disease (COPD) in childhood
new	<b>10</b> Yoav Ben Shlomo and Kate Tilling	Modelling the natural history of multiple sclerosis and examining potential prognostic risk factors

new **11** John Macleod

Causal inference in observational studies of substance use

new **12** Matthew Hickman

Modelling the transmission of Hepatitis C and HIV, and the impact of prevention strategies among injecting drug users in U

### **Genetics**

**13** Sarah Lewis and George Davey Smith

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

**14** Dr Santiago Rodriguez and Professor Ian Day

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

**15** Ian Day and George Davey Smith

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

**16** Tom Gaunt and Ian Day

Disentangling instances of causally and pharmacogenetically relevant genomic confounding

**17** Ian Day, Dave Evans and George Davey Smith

A study of genotype influences on reference ranges for clinical analytes ('Range Mendelization')

**18** Jeff Holly and Claire Perks

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

**19** David Evans; George Davey Smith, Matt Brown

Major Histocompatibility Complex (MHC) Genetics of Ankylosing Spondylitis

**20** David Evans, George Davey Smith

Genome-wide association analysis of complex endophenotype measures within the ALSPAC cohort

**21** David Evans, George Davey Smith, Ian Day

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

- 22** Tom Gaunt and Ian Day Which intermediate and disease traits are affected by GH, which by ACE and which by other genes in their region of linkage disequilibrium on human chromosome 17?
- 23** Tom Gaunt and Ian Day Application of a precise accurate quantitative DNA amplification approach in the study of epidemiological consequences of gene copy number polymorphisms
- 24** Tom Gaunt and Ian Day Somatic and population studies of KCNH2 sequence diversity and function: relevance to dysrhythmias and drug side-effects in the population
- 25** Tom Gaunt and Ian Day Genome-wide datamining for non-perfect polyalanine repeats in gene coding sequence and analysis of their role in disease.
- 26** Andy Ness, Ian Day Gene-nutrient interactions in the determination of blood lipid levels and early-stage atherosclerosis in childhood
- 27** Glyn Lewis Genetic predictors of adverse effects of antidepressants
- 28** 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.
- 29** Tom Gaunt and Ian Day Disentangling instances of causally and pharmacogenetically relevant genomic confounding
- 30** 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
- new **31** Dr Santiago Rodriguez and Ian Day Genetic analyses of individuals expressing extremely low levels of plasma protein biomarkers
- new **32** Sarah Lewis Meta-analysis of variants in the vitamin D receptor gene and breast cancer risk
- new **33** Jon Tobias Fine mapping of the association between *osterix* genotypes and bone mineral density in children

- 34** 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
- 35** 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
- 36** David Evans, George Davey Smith Genome-wide association analysis of complex endophenotype measures within the ALSPAC cohort
- 37** David Evans, George Davey Smith, Ian Day “Expression Genetics”: Genome-wide association analysis of gene expression data in the ALSPAC cohort
- 38** David Evans; Matt Brown Major Histocompatibility Complex (MHC) Genetics of Ankylosing Spondylitis
- 39** David Evans; Matt Brown Australian Genome-wide Association Study of Osteoporosis
- 40** Debbie Lawlor and Abigail Fraser Prevalence and determinants of non-alcoholic fatty liver disease (NAFLD) in adolescence

## **Topics 2009**

**Epidemiology and Health Services Research**

## **Title of Project: Modeling suicide trends over time and between countries**

### **Outline of Project**

#### **Background**

Suicide is one of the government's priority areas for reducing deaths<sup>1</sup> and accounts for a significant proportion of potential years of life lost<sup>2</sup>. Two striking features of the epidemiology of suicide are the large variations in its incidence over time and international differences in these temporal trends. Within Europe the main source of recent concern is the rise in youth suicide and a key feature of recent population suicide trends has been the 2-3 fold rise in young male deaths in many, but not all, industrialised countries. Explanations for the differing patterns in suicide trends (by age and gender) between countries are unknown and a greater understanding of factors contributing to observed trends will provide crucial insights into (i) the causes of suicide, (ii) approaches to its prevention and (iii) likely future trends in incidence. An understanding of the influences on time trends in suicide is also important when judging the success or failure of national prevention strategies<sup>3,4</sup>.

#### **Aim**

The aim of this studentship is to use a series of appropriate modeling approaches to identify explanations for differing European trends in suicide over the last 40 years of the 20<sup>th</sup> century. Candidate techniques include time-series analysis, join-point regression and age-period-cohort models.

#### **Data**

Data will be obtained from the WHO mortality database, UN Demographic Yearbooks, OECD Labour Force Survey and Historical Statistics and other relevant sources. All European countries for which age- and gender-specific suicide rates are available from 1960 onwards will be included in the analyses.

#### **Methods**

Time series models describe and explain observed trends in rates in terms of secular trend, seasonal variation, cyclical variation and irregular variation (for example, the seasonal component of suicide trends is well recognised<sup>5</sup>). Such an approach has already been applied to suicide rates in England and Wales<sup>6</sup> and results will be compared across countries. Join-point regression analyses may identify, for each country, the year(s) in which trends in suicide rates change direction. Similar approaches have been applied to cancer mortality rates, providing clues to likely explanations for changes in secular trends<sup>7</sup>. As well as being a useful method of summarizing observed rates, a comparison between countries of join-point years may clarify likely factors precipitating these changes. Both these methods take calendar time as the independent variable, and will be applied separately to data sets for each age/gender/country group. A more complex approach recognizes time as acting in three dimensions – age, period (calendar time) and cohort (time of birth). Although problems arise in trying to estimate all these effects simultaneously, generalized linear models can be used to fit age-period and age-cohort models for each country/gender data set. Joint models can then also be fitted if either cohort or period effects are found to be similar for both males and females. This reduces the number of parameters and enables estimation of age, period and cohort effects in one model. The resulting estimates can be compared across countries. Furthermore, countries exhibiting similar trends may be grouped together in a joint model, further improving the precision of estimates. A Bayesian implementation of these models has been used to make projections of cancer rates<sup>8</sup>, and could equally be applied to suicide rates to predict future trends.

#### **References**

<sup>1</sup> Department of Health (1999). *Saving Lives: Our Healthier Nation*.

<sup>2</sup> Gunnell, D. & Middleton, N (2003). National suicide rates as an indicator of the effect of suicide on premature mortality. *Lancet*; 362: 961-2.

<sup>3</sup> Department of Health (2002). *National suicide prevention strategy for England*. Department of Health, London.

<sup>4</sup> US Department of Health and Human Services (2001). *National strategy for suicide prevention: goal and objectives for action: summary*. Rockville, MD.

<sup>5</sup> Lester, D. (2001). *Suicide prevention: resources for the millennium*. Brunner-Routledge, Hove. p20.

<sup>6</sup> Gunnell, D., Middleton, N., Whitley, E., Frankel, S., Dorling, D. (2003). Why are suicide rates in young men increasing? - a time series analysis of trends in England and Wales 1950-1998. *Soc Sci Med*; 57: 595-611.

<sup>7</sup> Oliver S, May M, Gunnell D. International trends in prostate cancer mortality in the 'PSA era'. *Int J Cancer* 2001;92:893-898

<sup>8</sup> Bray, I. (2002). Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. *Applied Statistics*; 51: 151-63.

**Supervisors**

Professor David Gunnell  
Department of Social Medicine

**Title of project: Investigating and developing methods to improve the commissioning, design, and conduct of RCTs. (The MRC ConDuCT Trials Methodology Hub).**

The MRC ConDuCT Trials Methodology Hub brings together research, into the commissioning, design and conduct of RCTs, being conducted in the University of Bristol Departments of Social Medicine and Community Based Medicine. The Hub has funding for five four-year PhD studentships, joint funded by the MRC and the University of Bristol. In the first year students will attend courses from Social Medicine's popular short course programme, and elsewhere, to build a comprehensive knowledge of RCT design and conduct. The first year will also provide the opportunity to conduct small studies in different areas of trials methodology, allowing students to confirm their area of interest, build their research skills, formulate a detailed research plan for the subsequent three years, conduct feasibility studies of intended research procedures, and meet and work with theme leads within the Hub.

Current and forthcoming areas of work are as follows:

- Developing and integrating qualitative research methods to improve the design and conduct of RCTs
- Statistical methods in the design and analysis of challenging RCTs
- Methods for dealing with missing data, baseline covariates and incorporating productivity costs in economic evaluations conducted alongside RCTs
- Improving trial conduct
- Embedding clinically meaningful patient reported outcomes into RCTs

Many hub members are also involved with the Bristol Randomised Trials Collaboration, which has links with many ongoing and recently completed RCTs.

**Supervisors:** *These include:*

Social Medicine: Jane Blazeby, Sara Brookes, Rona Campbell, Jenny Donovan, Will Hollingworth, Athene Lane, Chris Metcalfe, Sian Noble, Jonathan Sterne, Kate Tilling.  
Community Based Medicine: Tony Ades, Alison Heawood, Sandra Hollinghurst, Alan Montgomery, Tim Peters, Nicky Welton, Nicola Wiles

**Title of project: Using qualitative methods at the end of randomised controlled trials to evaluate the utility and implementation of trial results**

The value of integrating qualitative methods within randomised controlled trials is increasingly recognised, particularly by health services researchers conducting pragmatic community-based trials. To date, qualitative methods (particularly interviews, but also observations and recordings of appointments) have been used at various stages of the trial process, notably during and pre trial. A reasonably common use is for the evaluation of patients' experiences of interventions prior to or during a trial to inform understanding of the acceptability of such interventions (Beattie et al, Emmett et al). Qualitative methods are increasingly used as part of process evaluations during trials, with the aim of improving the quality of trial conduct. Particular attention has been given to how qualitative research can help improve patient recruitment to trials. For example, some qualitative work has investigated clinicians' experiences of recruiting patients to a trial within consultations, identifying barriers and facilitators to patient recruitment (Mason et al 2007). Multiple qualitative methods (including in-depth interviews with RCT participants and recruitment staff, and analysis of information exchanged by recruiters and participants within recruitment appointments) have been used in a trial facing recruitment difficulties with the result of increasing patient acceptance of allocation and randomization rates (Donovan et al 2002, Donovan et al 2008). Early attempts have been made to implement this qualitative package (in the form of a complex intervention) in other trials facing recruitment difficulties, albeit with some challenges, including difficulties establishing collaboration between the qualitative team and RCT staff, poor communication between trial principal investigators and recruiting staff, and recruiters' concerns about having recruitment appointment recorded (de Salis et al 2008, 2008).

Qualitative methods are also increasingly being used within pre-trial feasibility studies, to elucidate and overcome potential process difficulties, and determine appropriate outcome measures, prior to starting the full trial. Such use of qualitative methods was recommended within the Medical Research Council's Framework for evaluating complex interventions (Campbell et al). Yet to date, little attention has been given to the use of qualitative methods at later stages of trials, particularly in relation to the implementation of trial results. This is important as while a trial may produce clinically and statistically significant results, these results may or may not receive recognition or be implemented in the relevant clinical settings. Therefore, the focus of the proposed PhD is to explore how qualitative methods may be incorporated in this post-trial phase to evaluate how trial results are received and implemented, with the ultimate aim of improving the utility of trial results.

**Aim**

To explore how qualitative methods can be used at the end of randomised controlled trials to evaluate the implementation and utility of trial results.

**Methods**

- Qualitative case study approach using a mixture of qualitative methods
- Selection of three ongoing/completed trials to act as 'case studies', drawn from trials within COBM or Social Medicine
- Sample of case studies to vary along certain criteria such as: clinical topic, setting in which trial was conducted, setting in which results may be implemented (may be same as trial setting), audience for the results (may overlap with former), fora/journals where results have been published, how 'controversial' the results are.
- Case studies would be staggered, so start with one or two trials that are in the public domain and have another near completed trial as the third case
- Possibilities for data collection:
  - Interviews or focus groups e.g. with clinicians/staff in settings where results are relevant and could potentially be implemented, the trial staff (such as PI, researchers and data analyst); journal editors or reviewers, others

- Documentary analysis e.g. of responses to publications in journals, trial management group/steering group meeting minutes where dissemination is discussed, written feedback of trial results to various audiences, final reports written by the trial team plus reviewer's comments
- Observation e.g. observation of end-of-trial management group/steering group meetings where dissemination strategies are discussed, observation of trial feedback meetings where results are disseminated to different audiences (e.g. clinical staff, policy/decision-makers), and leading conferences.

## References

Beattie A, Shaw A, Kaur S, Kessler D. Primary care patients' expectations and experiences of online Cognitive Behavioural Therapy (CBT) for depression: a qualitative study. *Health Expectations* (in press).

Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P. Framework for design and evaluation of complex interventions to improve health. *British Medical Journal* 2000, 321(7262):694-696.

De Salis I, Tomlin Z, Toerien M, Donovan J. Qualitative research to improve RCT recruitment: Issues arising in establishing research collaborations. *Contemporary Clinical Trials* 2008, 29:663-670.

De Salis I, Tomlin Z, Toerien M, Donovan J. Using qualitative research methods to improve recruitment to randomized controlled trials: the Quartet study. *J Health Serv Res Policy* 2008; 13 (Suppl 3): 92-96.

Donovan JL, Lane JA, Peters TJ, Brindle L, Salter E, Gillatt D, Powell P, Bollina P, Neal DE, Hamdy FC and the ProtecT study group. Development of a complex intervention improved randomization and informed consent in a randomized controlled trial. *Journal of Clinical Epidemiology* 2008, doi: 10.1016/j.jclinepi.2008.02.010

Donovan J, Mills N, Smith M. *et al.* Improving the design and conduct of randomised trials by embedding them in qualitative research: ProtecT (prostate testing for cancer and treatment) study. *British Medical Journal* 2002;325:766-770.

Emmett C, Shaw A, Montgomery A, Murphy D on behalf of the DiAMOND study group. Women's experience of decision making about mode of delivery after a previous caesarean section: the role of health professionals and information about health risks. *British Journal of Obstetrics and Gynaecology* 2006;113:1438-1445.

Mason VL, Shaw A, Wiles NJ, Mulligan J, Peters TJ, Sharp D, Lewis G. GPs' experiences of primary care mental health research: a qualitative study of the barriers to recruitment. *Family Practice* 2007;24:518-525.

**Supervisors:** *These include:*

**Primary supervisor:** Dr Alison Heawood, Academic Unit of Primary Health Care, Department of Community Based Medicine.

Secondary supervisor: Prof Jenny Donovan, Department of Social Medicine

**Title of Project: Determinants and sequelae of sexual and risk behaviours during adolescence**

**Outline of Project**

**Background**

Recognised sequelae of early romantic and sexual behaviours include teenage pregnancy and sexually transmitted infections (STIs). It has also been reported that early involvement in dating is associated with behavioural problems and lower psychosocial functioning. Most studies of adolescent sexual activity are cross-sectional. The use of data from longitudinal cohort studies will allow us to study the development of romantic and sexual behaviours from early adolescence onwards, and establish whether the trajectory of sexual development can be characterised into broad patterns.

An initial analysis of the ALSPAC data shows that although only a small proportion of adolescents are sexually active at age 12-13, romantic relations begin early, with 24% of 11-12 year olds and 41% of 12-13 year olds reporting having held hands, and 17% of 11-12 year olds and 32% of 12-13 year olds reporting having been kissed on the mouth (Waylen, Journal of Early Adolescence, under review).

There are currently two major public health initiatives in the area of sexual health. A national Chlamydia screening programme was introduced in 2006 for all young people under 25 years ([www.chlamydia-screening.nhs.uk/](http://www.chlamydia-screening.nhs.uk/)). During the 17+ ALSPAC clinics, participants will be offered a Chlamydia screening test, and will be asked to consent to their urine sample to be stored for future STI testing. In 2008, a programme of vaccination against human papillomavirus (HPV), the infection which causes cervical cancer, was launched ([www.immunisation.nhs.uk/Vaccines/HPV](http://www.immunisation.nhs.uk/Vaccines/HPV)). The vaccine is being offered to girls up to age 18 as part of the catch-up programme; the girls in the ALSPAC cohort will be eligible for vaccination.

We hypothesise that there are likely to be early behaviours that are related to the prevalence of Chlamydia infection and the uptake of HPV vaccination.

**Objectives**

1. To describe romantic and sexual behaviour, and their trajectories, from age 11 onwards
2. To investigate associations between adolescent sexual behaviours and other early life factors (such as socioeconomic background, parental attitudes)
3. To investigate determinants of chlamydia screening/ HPV vaccine uptake in the ALSPAC cohort

Students will be encouraged to incorporate elements of interest from the above objectives, and to add their own ideas within the broad subject area of sexual behaviour during adolescence.

**Data**

Prospective data on romantic relations and sexual behaviour have been collected in the Avon Longitudinal Study of Parents And Children (ALSPAC) cohort from age 11 onwards. This is complemented by a comprehensive set of data on pubertal development, socioeconomic background and possible maternal and paternal influences on a child's attitude to early sexual activity. Data on chlamydia screening uptake will be collected at the age 17+ clinics and HPV vaccine uptake will be obtained through linkage to Primary Care Trust records (ethical approval or exemption will be sought within the framework of a broader data linkage project in ALSPAC).

**Methods**

The student will be encouraged to develop a work plan based around their interests. The project may therefore incorporate a range of quantitative and qualitative methods. Statistical methods may include: descriptive statistics, linear, logistic and Cox regression modelling, structural equation modelling.

**Potential supervisors**

Please contact Dr Mona Jeffreys in the first instance. A range of potential supervisors within the Department/ University may be involved depending on the particular direction taken by the student. These include, among others, Dr John Macleod, Dr Andrea Waylen, Dr Caroline Trotter, Dr Matthew Hickman, Dr Paddy Horner and Dr Ardiana Gjini.

**Title of Project: Ethical aspects of epidemiological research on young people involving linkage to routine individual data**

**Outline of Project:**

**Background**

The Department of Social Medicine has recently obtained funding from the Wellcome Trust to undertake a significant research study which will substantially enhance the value of the data currently held by the Avon Longitudinal Study of Parents and Children (ALSPAC). Specifically, the study will involve linking its data with other national and health-care databases (record linkage). An important component of this project is the enriching of the ALSPAC database with data from external sources, including potentially sensitive social data such as criminal records.

This study raises several ethical issues, for example participant consent and issues in data linkage. These will be explored by way of a doctoral project that draws on and feeds into the wider research study. The doctoral project will particularly consider issues around different modes of consent to record linkage and questions of public good versus individual privacy in epidemiological research. Themes that may be drawn on in this project include modes of consent ('opt in' or 'opt out') and whether time limits to the validity of consent are important. The tension between the rights of individuals when compared with the potential benefit to society from record linkage will also be evaluated. Finally, the relevance and value of privacy and confidentiality will be explored.

**Aims of the PhD**

This doctoral project will address some of the ethical implications that may arise in the record linkage project described above. This will involve liaising with the wider research team and ensuring opportunities for empirical research are optimised. The doctoral project will aim to:

1. Scope the ethical issues arising in the record linkage project;
2. Synthesise existing literature on these issues;
3. Develop an appropriate strategy for empirical research on these issues with members of the ALSPAC cohort and possibly also researchers and the wider public;
4. Analyse research data for emergent themes;
5. Synthesise themes together with relevant scholarship in bioethics; and
6. Develop recommendations about appropriate methods of consent and record linkage to social data in cohort studies like ALSPAC. These will then be fed into the wider project, including training modules.

**Methods**

This will be a multi-disciplinary project, combining qualitative research with theoretical bioethics (applied philosophy). Research training will be provided for either or both of these research domains if required. The qualitative research will be undertaken together with the other qualitative streams in the wider research study. The research, as described in the 'Aims' section above, will address ethical themes such as consent and linkage to criminal records. This research can be undertaken with a variety of stakeholders. Data will be analysed for emergent themes but will also then be compared and critically analysed with respect to existing ethical literature and theoretical reasoning.

**References**

1. Huang N, Shih S-F, Chang H-Y, and Chou Y-J. Record linkage research and informed consent: who consents? *BMC Health Serv Res.* 2007; 7: 18.
2. Tate AR, Calderwood L, Dezateux C, Joshi H. Mother's consent to linkage of survey data with her child's birth records in a multi-ethnic national cohort study. *Int J Epidemiol.* 2006 Apr;35(2):294-8.
3. Hogue CJ. Ethical issues in sharing epidemiologic data. *J Clin Epidemiol.* 1991;44 Suppl 1:103S-107S.

4. Neutel CI, Johansen HL, Walop W. 'New data from old': epidemiology and record-linkage. *Prog Food Nutr Sci.* 1991;15(3):85-116.
5. Privacy, epidemiology, and record linkage. *Br Med J.* 1979 Oct 27;2(6197):1018.
6. Goodenough T, Williamson E, Kent J & Ashcroft A. What did you think about that? Researching children's perceptions of participation in a longitudinal genetic epidemiological study. *Children and Society*, 2003; 17: 113-125.

More general ethics literature around issues of privacy, confidentiality and disclosure in research will also be relevant to this work.

**Name of potential supervisors**

Dr Ainsley Newson, Centre for Ethics in Medicine, University of Bristol  
Catherine Heeney, Ethox, University of Oxford

Professor Rona Campbell and Professor John Macleod from the Department of Social Medicine will act as advisers to the project.

**Title: Secular trends and intergenerational influences on the suicidal thoughts in the Young Hunt study**

**Background**

This is an opportunity to undertake a PhD based on a collaborative study between the Department of Social Medicine Bristol and NTNU, Trondheim Norway.

Suicide is one of the main contributors to premature mortality around the world. Despite this, its aetiology remains largely uncertain. In Norway, suicide rates in young (15-24 year old) men and women increased threefold between 1960-2000 (Reseland 2006); the causes for this increase are largely unknown. Whilst recent figures indicate that there has been a decline in young male suicide in Norway between 1988-2004, no such trends have been seen in young females (Gjertsen F, 2007).

It is unclear whether recent changes in the incidence of suicide are due to changing levels of emotional distress amongst young people or due to changes in the incidence of other risk or protective factors for suicide.

The aim of this research is to better understand the aetiology and secular trends in the incidence of suicide in young Norwegians. The research will be based on the Young HUNT studies carried out in the Nord-Trøndelag region in 1995-7 (n=9000); 2000-1 (n=2400) and 2006-08 (n=8800) and involve a linkage between the sub-sample of 16-19 year olds who participated in these studies and their parents who participated in the main (adult) HUNT studies carried out in the same years. Altogether 7,000 participants in the Young HUNT studies have two parents who also participated in the adult studies.

The research group already have considerable experience of collaborative working on HUNT-based studies [Bjerkeset et al 2008] and the proposed PhD project aims to build on this work.

**Aim**

To investigate the following issues:

1. Has the prevalence of suicidal thoughts and other indicators / risk factors for mental illness (e.g. depression measured using the Hopkins 5-item Symptoms Check List (SCL); self esteem; and other measures) increased between 1995-7 and 2006-2008 in Young HUNT?
2. Are maternal characteristics more strongly associated with their daughter's risks of developing suicidal thoughts and mental illness than paternal characteristics. Similarly, are paternal characteristics more strongly associated with their son's risk of developing mental illness than maternal characteristics.
3. Are children of parents who have died by suicide at increased risk of experiencing suicidal thoughts and mental illness? Does any increased risk differ depending on whether it was the mother rather than the father who died and is greater than the risk when a parent died from "natural" causes?

## Methods

Standard statistical methods for cross-sectional and longitudinal studies will be used. Analyses of risk factors will use multivariable approaches to identify specific risk factors for suicidal thoughts and depression. The student will get experience of analysing a large linked dataset. Additional linkage opportunities include the possibility of linking HUNT data to the national prescription register to identify use of antidepressants and tranquilisers in Young HUNT-3.

**Supervisor:** Prof David Gunnell and other supervisors drawn from the Department of Social Medicine and the HUNT Research Group: Turid Holmen; Ottar Bjerkeset; Hans Nordahl; Pal Romundstad

## References

Bjerkeset O, Romundstad P, Evans J, Gunnell D. The association of adult body mass index and height with anxiety, depression, and suicide in the general population: The HUNT study. *Am J Epidemiol* 2008; 167: 193-202.

Gjertsen, Finn [Suicide statistics in Norway, the Nordic and the Baltic countries](http://www.med.uio.no/ipsy/ssff/statistikk/pdf/Gjertsenstatistiskeoppgaverselvmordmai07.pdf)  
<http://www.med.uio.no/ipsy/ssff/statistikk/pdf/Gjertsenstatistiskeoppgaverselvmordmai07.pdf>  
(Internet Publication 2nd May 2007)

Reseland S, Bray I, Gunnell D. Relationship between antidepressant sales and secular trends in suicide rates in the Nordic countries. *Br J Psych* 2006; 188:354-358

**Title of Project: Quantitative and qualitative assessment of injecting risk and drug using networks: developing better behavioural surveillance and effective transmission models of BBV transmission among IDU**

**Outline of Project:**

**Background:** The main thesis is that current measures of injecting risk behaviour are not informative – and if taken at face value may be misleading - and that better measures may be (and must be) obtained through new qualitative and quantitative assessment. Injecting risk behaviour, principally through sharing of used syringes, is a key factor determining the transmission and spread of HCV, HIV and HBV (i.e. blood borne viruses, BBV) among injecting drug users (IDU) [HPA Shooting Up]. Over 80% of diagnosed HCV infection, ~40% of HBV reports, and ~5% of HIV infection is attributed to IDU. The latest evidence also suggests that current or ex-IDU contribute  $\frac{3}{4}$  of the estimated 200,000 HCV cases in England and Wales.

However, current surveillance and research data suggest that HCV incidence recently increased, and since 2001 there has been an ongoing increase in HCV prevalence [Judd 2005a, Sutton, HPA]. Moreover HIV prevalence, after remaining stable at low endemic levels for a decade, is now rising, and there have been marked increases in reports of injecting related bacterial infections [Hope, HPA]. HPA UAPMP data on sharing reported an increase in 1997, but no changes since then, but the analytic value and interpretation of these data is limited [Hope personal communication]. There are 3-fold differences in HCV prevalence among IDU in different geographical settings, for example, from ~20% in rural South Wales and North East England to 60% in London, Manchester and Bristol [Hickman et al, NPHS Wales in prep]. Furthermore recent data from a large study in Wales suggests considerable variation in HCV incidence on a smaller town/city level. Current data on sharing from UAPMP, longitudinal, or enhanced surveillance studies do not predict HCV/BBV infection, and fail to explain the geographical differences in HCV prevalence [e.g. Judd 2005b]. The evidence points to increased risk of HCV infection among homeless and crack IDU, and very recently to substitution treatment as potentially protective.

We believe all three of these factors are mediated through changes in injecting risk i.e. increasing or reducing injecting frequency, size and rate of change of drug sharing group, and syringe sharing events. However, these proximal measures of injecting risk may not be measured with the same degree of accuracy or reliability i.e they are misclassified. Moreover, we believe there maybe a parallel with explaining geographical differences in STI prevalence – where the degree of concurrency between partners rather than average number of sexual partners was the key predictor [Morris]. In contrast, some epidemiological studies emphasise associations between HCV and sharing paraphernalia [Mathei] – suggesting that public health messages also should target paraphernalia as a key transmission risk. Though self-reported behaviours among IDU have been validated [Darke], the research did not extend to sharing, which has been shown to be influenced by social desirability and under-reported in certain study conditions [Crane personal communication]. Initial work on developing a dynamic HCV transmission model for London and UK highlighted key uncertainties in both biological (e.g. viral clearance) and behavioural parameters (sharing frequency), which if one or other was resolved would substantially improve the model projections [Vickerman].

**Objectives and Design:-** The thesis will be explored in three linked parts.

First, the study will test the hypothesis that current measures of injecting risk are not informative. Current systematic reviews of HCV, HIV and HBV prevalence and associations with sharing and injecting risk will be updated. Analyses of UK surveillance data, and other EU data in partnership with EMCDDA, will be conducted to test the strength of association between sharing and HCV infection and whether differences in reported sharing behaviour can explain differences in HCV prevalence. A review of the qualitative literature on reporting problem behaviours also will be conducted to assess reasons for under-reporting socially

undesirable behaviours, and what recommendations have been made to reduce under-reporting.

Second, (and the main part) the study will test the hypothesis that more accurate and better measures of injecting risk can be obtained. A range of qualitative and quantitative surveys of IDU in two settings will be conducted: Bristol and Newport one high and one low HCV prevalence area. This part of the study will allow training in respondent driven sampling (RDS) methods for recruiting IDU. The surveys will explore a series of questions, which will be extended and refined by the qualitative surveys and potential data demands of HCV transmission model. For instance:-

- does self-completion of sharing behaviour under-report sharing frequency
- can qualitative assessments increase reported sharing
- are there reliable proxies for sharing injecting equipment
- what techniques may solicit more accurate responses (e.g. CASI, anonymised response, scenarios)
- what is the ratio of sharing paraphernalia: sharing injecting equipment
- is homelessness associated with greater size and rate of change in drug sharing partners
- what other factors are associated with size and rate of change of sharing partners
- what is the level of concurrency in drug and syringe sharing, and can it predict HCV infection
- what do users recommend for measuring injecting risk
- can users assess lifetime injecting risk
- to what extent does the identity of the interviewer/researcher influence responses

Third, the study will consider whether behavioural surveillance of injecting risk can be improved. It will evaluate the information in relation to CDC and HPA guidelines on evaluating surveillance programmes. In addition, this part of the study will consider:- whether sero-surveillance data are sufficient to inform public health action; whether better information can be obtained for informing and reducing uncertainty within transmission dynamic models; and the utility and recommendations of changing ongoing surveillance.

**Milestones:** The three parts correspond to the three years of the PhD. The first year will be used to conduct literature review, receive analytic training and analyse surveillance data, and set up fieldwork for second year. Second part and fieldwork will be conducted in 2<sup>nd</sup> year. Third part and final write up will be conducted in last year.

#### **References:-**

- Darke S. Self-report among injecting drug users : a review. Drug and Alc Dependence 1998; 51: 253-63*
- Hickman M, Hope V, Brady T et al. Hepatitis C (HCV) prevalence, and injecting risk behaviour in multiple sites in England in 2004. AJE submitted*
- HPA Shooting Up: infections in injecting drug users in the United Kingdom, 2002. (V Hope et al.). Eurosurveillance; 8 (4) 22 January 2004*
- Hope VD, Judd A, Hickman M, Sutton A, Stimson GV, Parry JV, Gill ON. HIV prevalence among Injecting Drug Users in England & Wales 1990 to 2002: Evidence for increased transmission in recent years. AIDS 2005; 19: 1207-14*
- Judd A, Hickman M, Jones S, McDonald T, Parry JV, Stimson GV. Incidence of hepatitis C virus and HIV among new injecting drug users in London – prospective cohort study. BMJ 2005; 300: 24-25 (a)*
- Judd A, Hutchinson S, Wadd S, Hickman M et al. Prevalence of, and risk factors for, hepatitis C virus infection among recent initiates to injecting in London and Glasgow. Journal of Viral Hepatitis 2005; 12: 655-62 (b)*
- Mathei et al. Evidence or a substantial role of sharing of injecting paraphernalia other than syringes /needles to the spread of HCV among IDU. J Viral Hepatitis 2006; doi:10.1111/j.1365-2893.2006.00725.x*

*Morris M, Kretzschmar M. Concurrent partnerships and the spread of HIV. AIDS 1997, 11:641-648.*

*Sutton AJ, Gay NJ, Hope VD, Hickman M. Modelling the Force of Infection for Hepatitis B, Hepatitis C, and HIV in Injecting Drug Users in England and Wales. BMC infectious diseases (in press)*

*Vickerman P., Hickman M., Judd A. Modelling the impact of hepatitis C transmission of reducing syringe sharing: London case study. International Journal of Epidemiology (in press)*

**Research supervision**

Dr Matthew Hickman, Professor Rona Campbell

**Title of project: Development and application of methods to estimate prevalence of injecting drug use and other marginal populations**

Injecting drug use and other hidden or marginal populations are inefficiently and unreliably estimated using population survey techniques. Instead, indirect statistical methods are often proposed, including capture-recapture methods developed originally for animal populations. Capture-recapture is an application of Poisson regression techniques to estimate the size of the unobserved population by fitting a model to a table containing information on the overlap between multiple data sources (ie. on the number and characteristics of subjects who are one or more data sources) Recent advances in capture-recapture have focussed on methods of model selection, internal tests of validity, the use of bootstrapping and simulation methods for generating confidence intervals, and the inclusion of covariates within the analysis. Covariate methods have been applied to problem drug use, but still require some refinement and development. Further, the methods assume a closed population, whereas problem drug use is dynamic, and serial capture-recapture estimates are unlikely to be sensitive enough to detect change in the population over time. Therefore, consideration could be given to applying open capture-recapture methods in order to detect change in the population over time and overcome the violation of the assumption of a closed population.

**Objectives**

1. To develop and extend current methods to estimate the size of hidden populations
2. To examine the assumption that the population remains static in the time period examined (e.g. by developing open models)
3. To apply these methods to the estimation of the size of drug use populations
4. To develop software (e.g. Stata commands) which will enable these methods to be carried out easily and quickly.

**Design**

In addressing objective 1 (above), work is required to:-

- a) test and identify the best method(s) for model selection within a covariate model, determine the implications of selecting the “best fitting model”, deriving a “weighted” estimate, or using the fully saturated model, and recommend the best course of action;
- b) establish limits (or conditions) for CRC in relation to the number of covariates and number and size of data sources and degree of overlap, determine the theoretical implications and limits of CRC with scarce data on the reliability of the prevalence estimates, and assess the implications and best method of dealing with missing categories and data.

In addressing objective 2 (above), “open” methods will be developed. Traditional CRC methods assume a “closed” population (i.e that migration, cessation, and incidence are negligible over the length of the study). Open CRC methods which are the method of choice in ecology have been rarely used for estimating change in injecting drug use or other epidemiological problems and often fail to give credible results. Animal ecology models also have been recommended and developed that combine closed and open models to estimate the population dynamics over time which have yet to be applied to many epidemiological situations.

Objective 3 (above) will be addressed by applying the methods developed to data on injecting drug use populations from data sources in Bristol, Wales, and selected sites in Australia. Objective 4 will involve translating the results from objectives 1-3 into guidelines for practice. This will include development of software/suites of programmes to enable researchers to apply these methods in a consistent and timely fashion.

**Supervisors:** Drs Kate Tilling and Matthew Hickman

**Project title: The origins of chronic obstructive pulmonary disease (COPD) in childhood**

**Outline:** Chronic obstructive pulmonary disease (COPD) is characterized by airflow limitation that is not fully reversible by bronchodilators. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) has defined criteria for diagnosis and staging of COPD and, based on these, a substantial proportion of young adults (20-44 years) has been reported to already have established COPD<sup>1</sup>. Cigarette smoking is recognized as a major risk factor for COPD but only a proportion of smokers develop the condition and there remains much to be learned about other factors that are important in the aetiology of COPD, with increasing interest focusing on early life events and their influence on lung and airway development. There is evidence from longitudinal studies that decrements in pulmonary function that are established in infancy and early childhood persist until adolescence. Failure to achieve maximal pulmonary function in early adult life is likely to be associated with increased respiratory morbidity as pulmonary function declines in later life and possibly with more rapid decrements in pulmonary function through mid-adulthood.

The aims of this project are to investigate the early life antecedents of having low values in adolescence (15-16 years) of FEV<sub>1</sub>, maximal mid-expiratory flow (MMEF) reflecting small airways obstruction, and FEV<sub>1</sub>/FVC ratio that are not bronchodilator-reversible. The research will be based in the Avon Longitudinal Study of Parents and Children (ALSPAC), a longitudinal, prospective birth-cohort study recruited in pregnancy that has followed a population of nearly 14,000 children since birth. The primary outcome of this research will be post-bronchodilator pulmonary function measured in approximately 6000-7000 of this population at age 15+ years (MRC-funded G0401540). We will also apply methodology that we are successfully developing for the classification of wheezing phenotypes based on longitudinal modelling of wheezing symptom data to other respiratory symptoms, including reported cough during childhood. This work includes approaches to modelling data missing at random to maximize the power of the study to detect main effects and interactions between exposures.

Principal research questions of this research will address reports from observational studies of associations between COPD in adults and birth size, particularly focusing on markers of intrauterine growth restraint and subsequent growth during early childhood, and of the relationship between intrauterine and early life exposure to tobacco smoke and subsequent pulmonary function. We will also address novel hypotheses, including the association between maternal and early childhood diet (including antioxidant intake), distance of residence from main roads as a marker of traffic-related pollution exposure, and interactions between these variables and tobacco smoke exposure and irreversible airways obstruction in adolescence. We will also relate the lifetime history of asthma during childhood, including measures of bronchial responsiveness at 8 years, to pulmonary function outcome at 15-16 years to address the potential for some phenotypes of asthma to be associated with remodeling of airways and persistent deficits in pulmonary function. Although the proportion of the ALSPAC population that fulfils GOLD criteria for COPD is likely to be small at the age of 15-16 years, we anticipate that investigation of population traits of pulmonary function measurements to identify those in the lowest deciles of pulmonary function variables without evidence of bronchodilator-reversibility will make a valuable contribution to understanding the associations of early life associations with clinically important pulmonary outcomes. Also, given the richness of the data available in the ALSPAC study, analyses will be adjusted for a number of potential confounding and effects modifying variables, including socioeconomic status, parental history of pulmonary diseases and personal history of smoking validated by measurement of cotinine at 15+ years.

**Reference:**

1. de Marco R, Accordini S, Cerveri I, et al. An international survey of chronic obstructive pulmonary disease in young adults according to GOLD stages *Thorax* 2004;59:120-125.

**Supervisors:**

Dr John Henderson

Professor Jonathan Sterne

**Title of project: Modelling the natural history of multiple sclerosis and examining potential prognostic risk factors**

**Outline of project:**

Multiple sclerosis is a chronic neurodegenerative disorder that starts in young adulthood and has a variable prognosis but in general progresses over time resulting in increased disability, reduced quality of life and excess-related mortality. Clinically patients are usually categorized into three distinct groups based on their presentation. These are (a) relapsing relapsing MS (RRMS) where patients have acute neurological episodes which usually get better and may or may not leave any residual problems, (b) primary progressive MS (PPMS) where patients present usually with a slow insidious decline in functional abilities and (c) secondary progressive MS (SPMS) which usually follows PPMS and where relapses may or may not still occur but there is evidence of clinical decline independent of relapses. Whether these different patterns reflect different disease sub-groups or are merely different manifestations of the same underlying pathology but related to age at onset<sup>1</sup> remains controversial and of great interest. Several prospective clinical cohorts have identified potential prognostic factors that may help predict the clinical course of disease. These include age at onset, gender, and frequency of relapses for RRMS. However these findings are not always consistent across cohorts. More recently the identification of genetic markers of MS risk from GWAS highlight potential candidates for prognostic markers and other biomarkers, such as MRI features at presentation have also been suggested. The need to differentiate patients with slower from more aggressive disease patterns is of major importance as new “disease modifying therapies” (DMT) are emerging<sup>2</sup> which may alter the natural history by enabling repair but are themselves associated with some risk both in the short term and potentially in the longer term if they alter the body’s natural defence systems against neoplasia.

**Objectives:**

1. To undertake a systematic review (with or without a meta-analysis) on the natural history and prognostic markers in MS
2. To use sophisticated statistical methods to model the natural history of MS and examine whether there may be distinct sub-groups or whether age at onset explains variability.
3. To examine for prognostic markers of disease course and derive a prognostic algorithm that may be helpful for clinical trials or patient counselling and advice

**Methods:**

The project will be based on a large MS register based in the University Hospital of Wales (UHW, Cardiff) and led by Dr. Neil Robertson. The University Hospital of Wales (UHW) is the major tertiary referral centre for neurology in Wales serving a local population of 1.2 million. The department of neurology has provided a network of MS clinics across South East Wales since 1999 and additional clinical data is available on patients from 1985.

Approximately 1000 patient contacts are currently documented annually and demographic and clinical data collected routinely at presentation with minimum subsequent current status data sets at each visit including current disease course (classified according to relapsing remitting (RR) secondary progressive (SP) or primary progressive (PP)) and relapse status, expanded disability status score (EDSS) (21), therapeutic interventions together with site and timing of relapses since last review. Initial and subsequent complete datasets are available on 1270 patients with MS comprising more than 90% of the local prevalent patients and may therefore be considered to be a representative sample of the prevalent population. Data on specific features of disease onset are available such as age of first event, age at diagnosis, reported clinical features at first event (split into isolated sensory, longitudinal tract, cerebellar, brainstem, optic neuritis, unifocal or multifocal features) and recovery from first event (either full recovery, incomplete recovery or progressive disease from onset). It has already published epidemiological data on the cohort.<sup>3</sup>

We will examine a variety of statistical methods, including multi-level and latent trait models to explore how well one can explain the heterogeneity of disease progression. Specific clinical

features will be explored to see if they have prognostic value. Findings that look promising will be tested, if possible using independent secondary datasets (with permission of the respective PIs) such as the MS-RSS and the British Vancouver register.

The implications of these findings will be explored in relation to inclusion criteria for RCTs and sample sizes as well as the ability to risk stratify for therapies and counselling.

The PhD student will be supervised by YBS (clinical epidemiologist) KT (senior statistician) and NR (MS specialist neurologist and custodian of dataset)

**References:**

1. Confavreux C et al Age at disability milestones in MS. *Brain* 2006;129:595-605
2. The CAMMS223 Trial Investigators. Alemtuzumab vs. Interferon Beta-1a in Early Multiple Sclerosis. *NEJM* 2008 Volume 359:1786-1801
3. Hirst C et al. Change in disability in patients with multiple sclerosis: a 20-year prospective population-based analysis. *Journal of Neurology, Neurosurgery, and Psychiatry* 2008;79:1137-1143

**Supervisors:** Yoav Ben-Shlomo, Kate Tilling, Neil Robertson

## **Title of Project: Causal inference in observational studies of substance use**

### **Outline of Project**

#### **Background**

Multiple strands of evidence strongly suggest that substance use, of alcohol, tobacco and illicit drugs, is one of the most important environmental influences on health.<sup>1 2 3</sup> Substance use, however, tends to be socially patterned. People who use drugs are often different from people who don't in ways other than the fact of their substance use. These other differences may have profound implications for health that can complicate causal attribution in observational studies. Many types of substance use are associated with social disadvantage.<sup>4 5 6 7</sup> The challenge here is to differentiate between instances where substance use mediates the typical association between disadvantage and poorer health (suggesting one strategy to reduce health inequality) and others where the association between substance use and adverse health or social outcomes mainly reflects the fact that substance use is a marker for disadvantage that damages health through other pathways. Aside from an association with social position, some types of substance use may reflect a tendency to take risks that again may influence health outcomes through multiple pathways, not all involving substance use. Moreover, aside from these issues of confounding, substance use is often subject to strong notions of social desirability that may influence how individuals report substance use to researchers.<sup>7</sup> All these problems mean that observational studies on the causes and consequences of substance use are fraught with methodological difficulties in terms of their usefulness as a basis for causal inference. These difficulties are often not acknowledged and strategies to overcome them are currently underdeveloped. A poor understanding of the causes of substance use is reflected in the limited success of prevention.<sup>9 10 11 12 13</sup> Parental substance use and childhood psychosocial problems, both exposures that are often more common amongst disadvantaged children, are widely held to be key influences on adolescent drug use.<sup>14 15</sup> Incomplete understanding of the consequences of drug use is illustrated by ongoing controversies such as whether cannabis use causes schizophrenia or influences educational attainment.<sup>16</sup> The Avon Longitudinal Study of Parents and Children (ALSPAC) is the UK's premiere resource for the study of the causes of the three commonest types of substance use (alcohol, tobacco and cannabis) and the short-term consequences of these behaviours amongst young people today. Crucially ALSPAC also provides the opportunity to investigate how problems such as confounding and reporting bias may complicate causal inference in this context.

#### **Aim**

The aim of this studentship will be to illustrate how problems of confounding and reporting bias may compromise causal inference in observational studies of adolescent drug use and to develop strategies to overcome these problems.

#### **Data**

Data will be obtained from ALSPAC, up to and including data collected in the age 15+ "Teen Focus 3" clinic and those obtained through linkage to the National Pupil Database. These data will include measures of pre and post-natal parental drug use, multiple measures of parental and family social position up to age 15, measures of childhood psychosocial and educational function, measures of self reported alcohol, tobacco and cannabis use from age 10 onwards and hair-based toxicological measures at age 15. Results from preliminary genome wide association studies on genetic predictors of key substance use phenotypes within an extensively phenotyped subset of the cohort will also be available. Data on educational performance in "Key stage 4" i.e. GCSE examinations will also be used.

#### **Methods**

Building on previous work in ALSPAC at age 10,<sup>17</sup> descriptive analyses will be presented on the prevalence of different substance use phenotypes at ages 13 and 15 and the distribution of these according to measures of social position across the life course. Logistic regression analyses will then examine the association between different measures of parental drug use (for example both maternal and paternal use in the prenatal period and in early childhood) and

measures of psychosocial function (such as conduct problems, bullying involvement, IQ and depression) with these later substance use outcomes. These analyses will be presented before and after measures of life course social position. Subsequent analyses will then examine the association between lifetime substance use up to age 15 and educational outcome at Key Stage 4. These analyses will compare effects of self-reported compared to toxicologically measured substance use and where possible will utilise any potential genetic instrumental variables identified through earlier GWAS studies. Again the influence of adjustment for life course social position on these effect estimates will be examined. It may be possible for students with a particular interest to develop more sophisticated statistical approaches to causal inference in this context such as those involving consideration of latent variables within structural equation or multilevel models.

**Supervisor:** John Macleod, Matthew Hickman.

**Title of Project:** Modelling the transmission of Hepatitis C and HIV, and the impact of prevention strategies among injecting drug users in UK

### **Outline of Project**

#### **Background.**

Hepatitis C (HCV) and HIV cause substantial morbidity. In the UK 150,000 to 300,000 people are infected with HCV – over 80% due to injecting drug use; and nearly 10% of HIV cases are due to injecting. The epidemiology and evidence on the effectiveness of interventions are currently under review (NICE 2009, ACMD 2009).

Key findings are that in different settings in the UK there is variation in the prevalence of HCV/HIV amongst IDUs. Some of the variation in HCV prevalence is associated with homelessness and crack injection, but confusion still surrounds the link with syringe sharing - the main risk factor for spreading these viruses. The differences are likely to be partly due to reporting bias, but also may be due to subtleties in IDU syringe sharing behaviour that have not been recorded in previous surveys. Needle and syringe programmes (NSP) and opiate substitution therapy (OST) are the main intervention strategies for reducing HIV and HCV transmission. However, evidence on their intervention effect is weak, and there is little evidence on the levels of coverage of these and other interventions required to substantially reduce HCV or HIV.

This gap arises partly from limited evidence about what aspects of IDU risk behaviour determine the level of HCV and HIV transmission in different settings in the UK. In addition, there is limited understanding on how increased syringe distribution, or other forms of intervention contact may effect different IDU risk behaviours, such as the rate of syringe sharing, the size and stability of syringe sharing groups, and the degree of concurrent sharing. Without understanding these factors and relationships it is very difficult to evaluate the potential impact of different IDU focused interventions. Opportunities to develop better transmission models are arising because of new data and surveillance in the UK, including HCV Action Plan in Scotland (<http://www.scotland.gov.uk/Publications/2006/09/15093626/0>).

#### **Aim**

To develop novel dynamic mathematical models of HCV and HIV transmission among IDUs in UK.

#### **Data**

UK surveillance and enhanced surveillance data will be available (HPA 2008).

#### **Methods**

Existing models of HCV and HIV transmission will be developed and used as the basis of this PhD (Vickerman 2006,2007). The new models will incorporate network structures, and will be developed in parallel with the collection of enhanced surveillance data in the UK that will attempt to understand better the intervention effect of NSP and OST.

#### **References**

ACMD The prevention of hepatitis C among injecting drug users (Advisory Council on the Misuse of Drugs February 2009)  
 HPA Health Protection Agency, Health Protection Scotland, National Public Health Service for Wales, CDSC Northern Ireland, and the CRDHB. **Shooting Up: Infections among injecting drug users in the United Kingdom 2007**. London: Health Protection Agency, October 2008

NICE Needle and syringe programmes: providing injecting equipment to people who inject drugs. Expected publication February 2009  
<http://www.nice.org.uk/guidance/index.jsp?action=byID&o=11829>

Vickerman P, Hickman M, Judd A Modelling the impact on Hepatitis C transmission of reducing syringe sharing: London case study. Int J Epidemiol. 2007 Apr;36(2):396-405.

Vickerman P, Hickman M, Rodes T, Watts C. Model Projections on the Required Coverage of Syringe Distribution to Prevent HIV Epidemics Among Injecting Drug Users. JAIDS 2006; 42 (3): 355-361

**Supervisors**

Peter Vickerman, Matthew Hickman

**Genetic Epidemiology, Bioinformatics and Molecular Genetics**

**Title of Project: 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 of Project: 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 of Project: 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 of Project: 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*<sup>1</sup> 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 data<sup>2</sup>, confirmed with HapMap data (www.hapmap.org)<sup>3</sup>. 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 disequilibrium<sup>2</sup>. The use of HapMap data<sup>3</sup> 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 project<sup>3</sup> 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 project<sup>3</sup>.
- 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<sup>2,4</sup>.

**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 of Project: 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 of Project: 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 of Project: 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 3<sup>rd</sup>-4<sup>th</sup> 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<sup>-50</sup> 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 of Project: Genome-wide association analysis of complex endophenotype measures within the ALSPAC cohort**

**Outline of Project:**

**Objectives**

1. To identify genetic variants associated with complex endophenotypes that are important measures of growth and development as well as predictors of disease risk in later life using genome-wide association analysis.
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**

The aim of this project is to identify genetic variants responsible for variation in complex endophenotypes via genome-wide association (GWA). The student will select a group of related endophenotypes from the many measures available within the ALSPAC cohort (e.g. all the language related variables; all the bone density and growth phenotypes etc.) and will perform genome-wide association analysis on these traits. Significant associations will be followed up in other cohorts to ensure the robustness and replicability of the findings. Subsequent fine mapping will more precisely identify the functional variants involved.

With the advent of GWA analysis, genetic mapping has now entered an exciting new phase where for the first time it has become possible to robustly identify many of the genetic variants underlying complex traits and diseases. Three recent developments have made this a possibility. First, the availability of genotyping chips containing hundreds of thousands of markers, which provide good coverage of much of the common genetic variation within the genome, has meant that GWA studies are now financially and technically feasible. Second, the publication of the International Haplotype Map, which documents the pattern of linkage disequilibrium across the genome, has facilitated the design and analysis of GWA studies (The International HapMap Consortium, 2005). Finally, the existence of large patient cohorts has been a necessary prerequisite in order to obtain the power necessary to detect loci of small to moderate effect (The Wellcome Trust Case Control Consortium, 2007). These developments have led to a flood of GWA studies that have successfully identified genes conferring risk to a variety of diseases including (but not limited to) coronary heart disease (Samani et al. 2007), breast cancer (Easton et al. 2007), types I and II diabetes (Todd et al. 2007; Zeggini et al. 2007), and inflammatory bowel disease (Parkes et al. 2007).

To date the vast majority of GWA studies have not explored the relationship between genomic variation and detailed assessments of endophenotypes (i.e. intermediate physiological measures that lie between distal determinants and proximal disease states) and how these associations emerge longitudinally. This is important because identification of these genes will allow better characterisation of the biological pathways underpinning growth and development, as well as those responsible for degeneration and disease in later life. Recent results from the Wellcome Trust Case Control Consortium (WTCCC) provide a vivid illustration of the utility of this approach. The WTCCC recently identified an association between SNPs in the *FTO* gene and Type II diabetes (The Wellcome Trust Case Control Consortium 2007; Zeggini et al. 2007). However, because investigators had collected additional detailed endophenotype information, they were able to show that the association between *FTO* and type II diabetes could be completely explained by body mass index (i.e. the diabetics had higher body mass indices than the lighter controls), and that fat mass was the variable primarily driving this association (i.e. not lean body mass). In other words, the *FTO* variant was not responsible for increasing risk of diabetes directly, but rather was likely affecting disease risk by increasing individuals' fat mass. In addition, longitudinal information from the ALSPAC cohort indicated that although the association between *FTO*

and fat mass was not present at birth, it was apparent as early as seven years of age and was subsequently maintained into the pubertal period and beyond (Frayling et al. 2007). It is important to realise that this kind of in depth assessment of the effect of the *FTO* gene would never have been possible without the existence of detailed longitudinal endophenotype information like that provided in the ALSPAC cohort.

Progress identifying loci that affect quantitative endophenotypes is likely to be difficult because of the low power to detect genetic association in unselected individuals (as opposed to case-control studies where we can assume that cases have been selected on the basis of their extreme score on some underlying continuous distribution). In addition, quantitative genetics theory (Lynch & Walsh, 2001), studies of model organisms (Mackay, 2001) and the few loci which have been reliably replicated in human genetic association studies (Frayling et al. 2007; Zeggini et al. 2007) suggest that the majority of genes underlying variation in complex quantitative traits will be of small to moderate effect (i.e. responsible for < 1% of the phenotypic variance). It is crucial therefore that any putative genome-wide association study of quantitative endophenotypes be adequately powered to detect small effects of this magnitude.

The ALSPAC cohort is unique in this regard, being one of the few cohorts in the world with literally hundreds of accurate measures on a variety of phenotypes (including immunological measures, sensory functions, motor functions and coordination, cognitive measures, anthropometry, serum biochemistry etc.), many of them longitudinal, on thousands of children throughout their first fifteen years of life. The sheer size of ALSPAC also means that there is considerable power to detect loci responsible for small proportions of the phenotypic variance. Currently 2000 ALSPAC children have been typed genome-wide on the Illumina 317K SNP chip, and we are currently applying for funding to genotype the remaining ~13,000 ALSPAC children. The successful student will have an exciting opportunity to take a leading role in the design and analysis of what will be one of the world's first genome-wide association studies of quantitative endophenotype measures.

### **Plan of investigation**

1. The project will involve use of several methods of study design and analysis relevant to epidemiology, genetic epidemiology, bioinformatics and biostatistics
2. Cleaning, managing and maintaining a large database of genetic information
3. Employing genome-wide association and other appropriate statistical methodology to identify genetic variants associated with endophenotype variability
4. Prioritizing SNPs for follow up and fine mapping using ALSPAC and additional cohorts
5. Organising follow up genotyping where necessary
6. Using meta-analyses to appropriately pool all data
7. Using appropriate statistical methods to make causal inference from the pooled analyses

### **References**

Easton D. F. et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 447, 1087-93.

Lynch M. & Walsh B. (1997). *Genetics and Analysis of Quantitative Traits*. Sinauer Assoc, Sunderland, MA.

Frayling T.M. et al. (2007). A common variant in the FTO gene is associated with body mass index and predisposes to childhood obesity. *Science*, 316(5826), 889-94.

Mackay T. F. (2001). The genetic architecture of quantitative traits. *Annu Rev Genet*, 35, 303-39.

Parkes M. et al. (2007). Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet*, 39, 830-2.

The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature*, 437, 1299-320.

Samani N. J. et al. (2007). Genomewide association analysis of coronary artery disease. *N Engl J Med* (available online).

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

Todd J. A. (2007). Robust associations of four new chromosome regions from genome-wide analysis of type 1 diabetes. *Nat Genet*, 39, 857-64.

Zeggini E. et al. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, 316(5829), 1336-41.

### **Supervisors**

David Evans, George Davey Smith

**Title of Project: “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 of Project: Which intermediate and disease traits are affected by GH, which by ACE and which by other genes in their region of linkage disequilibrium on human chromosome 17?**

**Outline of project**

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* (many papers) and *GHI*<sup>1</sup> gene region and cardiovascular and metabolic risk phenotypes. We have also demonstrated ~20% linkage disequilibrium between the *ACE* and *GHI* genes using our own data<sup>2</sup>, confirmed with HapMap data (www.hapmap.org). Whilst it is apparent that factors in this region influence disease risk, pathway misinference may arise due to the high levels of linkage disequilibrium<sup>2</sup>. In addition to *ACE* and *GHI* several other genes with the potential to influence cardiovascular and metabolic disease risk exist in this chromosomal region. We propose a comprehensive analysis of the region to identify the important genetic variants that contribute to disease.

**Objectives**

1. Genotype haplotype-tagging SNPs across the *ACE-GH* genomic region on chromosome 17 in British Women's Heart and Health Study
2. Use haplotype analyses to determine which SNPs are genuinely associated with cardiovascular and metabolic phenotypes, and which are associated due to linkage disequilibrium

**Design**

1. SNPs will be selected using data from the HapMap project<sup>3</sup> to select haplotype-tagging SNPs.
2. Genotyping will be carried out by outsourcing
3. Haplotype analyses for association with phenotypes will use various programs as described in previous work<sup>2,4</sup>.

**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 (in press, *Pharmacogenetics and Genomics*)
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.

**Potential supervisors:**

Tom Gaunt, Lecturer in Bioinformatics and Molecular Genetics  
 Ian Day, Professor of Genetic and Molecular Epidemiology

**Title of Project: Application of a precise accurate quantitative DNA amplification approach in the study of epidemiological consequences of gene copy number polymorphisms**

**Outline of Project:**

**Background**

It has become clear in the past couple of years that gene copy number variation (CNVs) is prevalent in the human genome, especially in regions of segmental duplication, and at least 1,000 such loci exist. It is also clear, both from classical examples such as CYP2D6 in drug metabolism and HBA in haemoglobinopathies, and new examples such as CCL3L1 in AIDS susceptibility, that these CNVs have major roles on disease risk traits and phenotypes. Within the Bristol Genetic Epidemiology Laboratories, we have developed a system, Amplification Ratio Control System (ARCS) [work by Dr Philip Guthrie, Research Associate, currently subject to patent review by UoB] which enables pseudo-endpoint multiplex PCR which preserves representation of template ratio very accurately, and allows fine control of that in formats suitable for both liquid phase and gel based assays. This enables the quantitation of gene dose in simple efficient liquid phase assays suitable for high throughput typing, which has previously been problematic for the epidemiological scale study of CNVs. The approach also lends itself to mRNA quantitation, mRNA being the 'proximal phenotypic trait' for most genes. For transcribed haplotype SNPs, we have also developed and applied 'Quantitative Trait Haplotyping' where relative allelic expression of mRNA isoforms can be measured and have applied this to functional studies of the angiotensin receptor type I (AGTR1) [Abdollahi, Human Mutation in press]. This will enable integrated study of selected CNVs for their epidemiological and causal molecular effects.

**Objectives of the PhD**

- 1.- To apply ARCS to selected CNVs and type these markers in relevant phenotyped cohorts
- 2.- To conduct appropriate statistical analyses
- 3.- To develop bioinformatics approaches to mine emergent CNV data for further studies, and possibly to apply ARCS in the more detailed characterisation genomewide, of putative CNVs.

**Design**

Development of ARCS assays for targets of highest clinical epidemiological interest. To type these markers in relevant cohorts, particularly in ALSPAC (10,000 children) for whom lymphoblastoid cell lines for quantitative mRNA studies (where expression occurs in LCLs) will also be undertaken in selected subsets or correlated with emergent HapMap and microarray expression data. Upfront bioinformatic and downstream statistical analyses will be used for target selection. Improved in silico tools for automating both target selection and ARCS assay design will also be developed.

**Techniques and approaches**

Application and extension of the ARCS methodology.  
 Statistical approaches to classification and phenotype correlation studies of CNVs  
 Bioinformatic aspects of CNVs relevant to molecular epidemiological studies  
 Possibly, mRNA quantitative correlation studies (from established lymphoblastoid cell lines in ALSPAC) for CNVs expressed in LCLs.

**References**

1. Buckley PG, Mantripragada KK, Piotrowski A, Diaz de Stahl T, Dumanski JP. Copy-number polymorphisms: mining the tip of an iceberg. *Trends Genet.* 2005 Jun;21(6):315-7. Review.
2. Mohammad R. Abdollahi <sup>1,3</sup>, Rohan M. Lewis <sup>2</sup>, Thomas R. Gaunt <sup>1</sup>, Debbie V. E. Cumming<sup>1</sup>, Santiago Rodriguez<sup>1</sup>, Matthew Rose-Zerilli<sup>1</sup>, Collins A R<sup>1</sup>, Holly E.

Syddall<sup>2</sup>, William M Howell<sup>1</sup>, Iain T. Cameron<sup>2</sup>, Ian N. M. Day<sup>1</sup>. Quantitated Transcript Haplotypes (QTH) of AGTR1, reduced abundance of mRNA haplotypes containing 1166C, and relevance to metabolic syndrome traits(*Hum Mutation*, accepted Aug 2006)

**Potential supervisors**

Tom Gaunt, Lecturer in Bioinformatics and Molecular Genetics  
Ian Day, Professor of Genetic and Molecular Epidemiology

**Title of Project: Somatic and population studies of *KCNH2* sequence diversity and function: relevance to dysrhythmias and drug side-effects in the population**

**Outline of Project:**

**Background**

Sudden and fatal dysrhythmias attributable to ‘long QT syndrome’ have been increasingly recognised in the past decade and familial occurrence has enabled the characterisation of the genes involved. One such gene, *KCNH2* or *HERG* or *LQT2*, which encodes a potassium slow channel crucial in cardiac repolarisation, has become particularly important to the pharmaceutical industry and to medicine in general. This channel has a pore with a rather open outer face and either because of its role or its structure, is prone to being affected by exogenous small molecules such as new drugs, leading to occasional fatal dysrhythmias as a drug side-effect. Compounds for trial are therefore now screened in *in vitro* systems expressing the ‘wild type’ channel. However, it remains unknown (except for common polymorphism), how much diversity of *KCNH2* may exist in the general population. Such diversity might contribute to primary dysrhythmic effects or to dysrhythmias secondary either to enhancement of a particular drug side-effect or to other compromise such as ischaemic heart disease. We have recently developed tools suitable for mutation scanning in available population samples and have also presented theoretical and practical evidence (in the context of cholesterol levels), that paucimorphisms and *formes frustes* mutations may contribute significantly to disease burden without leading to florid familial long QT syndrome.

In the process of setting up mutation assays for population scanning of *KCNH2*, we examined known *LQT2* family mutations and their sequence contexts. We were not surprised to find that quite a lot of mutations were at CpG sites, but we were surprised to note that the protein coding region was not depleted of CpG, whereas CpG depletion is an almost universal phenomenon in coding regions in vertebrate genomes. Further examination has confirmed that this phenomenon affects the exons but not the introns of *KCNH2*, but that it is not necessary for its coding potential since many of these CpG sites bridge a codon wobble position rather than representing the first two bases of arginine codons (CGN). This CpG prevalence has not been reported previously and suggests that *KCNH2* may be particularly prone to point mutations leading to aminoacid substitutions, both in germ line context and population; and at the somatic level. Since methylation at CpG sites predisposes CpG mutation, the frequency of *LQT2* syndrome CpG mutations suggests that as is usual in coding regions, the *KCNH2* coding CpG sites may be methylated in the germline. However, since this is also the reason for vertebrate genome depletion of CpG sites, if true, such methylation raises a paradox. Other systems such as liver cytochrome detoxification genes and HLA and other immune response genes have acquired high degrees of sequence diversity, driven by the population need to ‘survive’ in the face of the many environmental interactions in human history such as eating many different plants and meeting many different infections. Although the immune and detoxification systems might currently be considered exceptional in this respect, there is no reason why other systems such as ion channels susceptible to interactions with diverse small molecules, might not be subject to similar diversification pressures. If there were a high rate of somatic (e.g. in the heart) mutability of *KCNH2*, then it is also possible that heart might be expressing (perhaps advantageously) a spectrum of closely related (one aminoacid different) receptors. CpG site methylation also mark imprinting, a process whereby the allele from one parent is expressed but the allele from the other parent is silenced. While the related gene which is responsible for *LQT1* is known to be imprinted, there is no evidence for this for *KCNH2*.

**Objectives of the PhD**

To determine:

- how much population sequence diversity exists for *KCNH2*?
- what are the molecular, clinical and epidemiological effects of sequence diversity in *KCNH2*?
- what is the germline methylation status of *KCNH2*?

- what is the methylation status in heart of *KCNH2*?
- is there any somatic sequence diversity of *KCNH2*?

### **Design**

1. Mutation scanning for *KCNH2* variants in adult cohorts representing over 5,000 subjects with ECG traces available. Definition of QT dispersion in subjects with identified variants. Relevant epidemiological statistical/genetic analyses. Protein prediction bioinformatics (and collaborative in vitro functional studies with electrophysiologists Harry Witchel and Jules Hancox in Bristol). Extraction of germline DNA and tissue (e.g. heart) DNA for coupled bisulphite modification and sequencing to determine methylation status. Cloning and sequencing of *KCNH2* from heart DNA to look for possible low frequency somatic diversity.

### **Techniques and approaches**

- use of LightScanner, endoVIIMADGE and meltMADGE technology as appropriate for population mutation scanning
- follow up DNA sequencing
- follow up direct assay development for variants found, to scan in other population groups
- bisulphite modification of tissue DNA (already available) for methylation sequencing
- relevant statistical and bioinformatics analyses

### **References**

1. Sanguinetti MC, Tristani-Firouzi M. hERG potassium channels and cardiac arrhythmia. *Nature*. 2006 Mar 23;440(7083):463-9. Review.
2. Thomas D, Karle CA, Kiehn J. The cardiac hERG/IKr potassium channel as pharmacological target: structure, function, regulation, and clinical applications. *Curr Pharm Des*. 2006;12(18):2271-83. Review
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., Rassoulian, 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.* (2006) Prevalence and functionality of paucimorphic and private MC4R mutations in a large, unselected European British population, scanned by meltMADGE. *Hum Mutat.* (*in press*).

### **Potential supervisors**

Tom Gaunt, Lecturer in Bioinformatics and Molecular Genetics  
 Ian Day, Professor of Genetic and Molecular Epidemiology

**Title of Project: Genome-wide datamining for non-perfect polyalanine repeats in gene coding sequence and analysis of their role in disease.**

**Outline of project**

Non-perfect polyalanine repeat expansion in the gene *PHOX2B* (multiple GCN codons) has been implicated in the rare condition *Congenital Central Hypoventilation Syndrome* (CCHS)<sup>1,2</sup>. The mechanism of activity is likely to be through protein misfolding reducing transactivation and DNA-binding activity<sup>3</sup>. The project aims to address two questions: (1) are there undetected repeat expansions within the general population causing minor unexplained phenotypic effects. (2) Does this phenomenon exist in other genes, potentially at a higher frequency with less severe phenotypes.

**Objectives**

1. Develop a liquid-phase assay for the rapid detection of abnormal polyalanine repeat lengths in PHOX2B
2. Analyse the polyalanine repeat in cohort studies (eg ALSPAC) to determine whether it varies in “normal” individuals, and if so determine association with phenotypes, including physical activity and blood pressure.
3. Identify polyalanine repeats genome-wide, focussing on those of around 15-20 non-perfect repeats.
4. Analyse any candidate polyalanine repeats in other genes to determine if they are polymorphic in the general population, and if so whether they associate with relevant phenotypes. A number of cohorts are available depending on the phenotypes, including ALSPAC, BWHHS, Boyd Orr etc.

**Design**

1. Liquid-phase assays will be developed using LightScanner™ (Idaho Technologies) melt-curve analysis to detect differences in number of repeats.
2. New programs will be developed for *in silico* detection of candidate polyalanine repeat tracts using public sequence data utilising Perl or Python programming languages
3. Standard association analyses will be used to determine relationship of genotype with phenotypes.

**Techniques/approaches**

Bioinformatic work will include the development and utilisation of programs to interrogate databases and analyse sequence data for polyalanine repeats. This will require prior programming experience, preferably in Perl, Python or C.

Laboratory molecular work will include the development of melt-curve assays for detection of polyalanine repeat variability, and the high-throughput application of those assays. In addition fragment sizing analysis will be used for verification of genotype and sequencing for detection of internal sequence variability. Some of this work will be in collaboration with Maggie Williams (Clinical Genetics) and Professor Peter Fleming (Infant and Child Health).

## **References**

1. Amiel J, Laudier B, Attie-Bitach T, Trang H, de Pontual L, Gener B, Trochet D, Etchevers H, Ray P, Simonneau M, Vekemans M, Munnich A, Gaultier C, Lyonnet S. Polyalanine expansion and frameshift mutations of the paired-like homeobox gene PHOX2B in congenital central hypoventilation syndrome. *Nat Genet.* 2003 (4):459-61.
2. Matera I, Bachetti T, Puppo F, Di Duca M, Morandi F, Casiraghi GM, Cilio MR, Hennekam R, Hofstra R, Schober JG, Ravazzolo R, Ottonello G, Ceccherini I. PHOX2B mutations and polyalanine expansions correlate with the severity of the respiratory phenotype and associated symptoms in both congenital and late onset Central Hypoventilation syndrome. *J Med Genet.* 2004 (5):373-80.
3. Trochet D, Hong SJ, Lim JK, Brunet JF, Munnich A, Kim KS, Lyonnet S, Goridis C, Amiel J. Molecular consequences of PHOX2B missense, frameshift and alanine expansion mutations leading to autonomic dysfunction. *Hum Mol Genet.* 2005 Dec 1;14(23):3697-708.

## **Potential supervisors:**

Tom Gaunt, Lecturer in Bioinformatics and Molecular Genetics  
Ian Day, Professor of Genetic and Molecular Epidemiology

**Title of Project: 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 metabolism<sup>20</sup>. Assessment of recent diet and physical activity will be made, as these have been shown to affect fat tolerance<sup>21,22</sup>.

**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 of project: 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)<sup>1</sup> 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.<sup>2</sup>

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.<sup>3</sup> 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 of Project: 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 of Project:****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*<sup>1</sup> 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 data<sup>2</sup>, confirmed with HapMap data (www.hapmap.org)<sup>3</sup>. 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 disequilibrium<sup>2</sup>. The use of HapMap data<sup>3</sup> 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 project<sup>3</sup> 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 project<sup>3</sup>.
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<sup>2,4</sup>.

**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 of Project: 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., Rassoulia,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.rodriguez@bristol.ac.uk](mailto:santi.rodriguez@bristol.ac.uk))

Professor Ian Day

## **Title of Project: 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 ( $\beta$ -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](mailto:santi.rodriquez@bristol.ac.uk))

Professor Ian Day ([ian.day@bristol.ac.uk](mailto:ian.day@bristol.ac.uk))

**Title of Project: 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 of Project: 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<sup>1,2</sup>. 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 disorders<sup>3</sup> (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 2007<sup>4-9</sup>) 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 of Project: 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 data<sup>1</sup>. 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 chip<sup>2</sup>. 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 regions<sup>3-11</sup>. 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](http://www.molpage.org)). From the development and subjection 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 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 of Project: Genome-wide association analysis of complex endophenotype measures within the ALSPAC cohort**

**Outline of Project:**

**Objectives**

1. To identify genetic variants associated with complex endophenotypes that are important measures of growth and development as well as predictors of disease risk in later life using genome-wide association analysis.
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** The aim of this project is to identify genetic variants responsible for variation in complex endophenotypes via genome-wide association (GWA). The student will select a group of related endophenotypes from the many measures available within the ALSPAC cohort (e.g. all the language related variables; all the bone density and growth phenotypes etc.) and will perform genome-wide association analysis on these traits. Significant associations will be followed up in other cohorts to ensure the robustness and replicability of the findings. Subsequent fine mapping will more precisely identify the functional variants involved.

With the advent of GWA analysis, genetic mapping has now entered an exciting new phase where for the first time it has become possible to robustly identify many of the genetic variants underlying complex traits and diseases. Three recent developments have made this a possibility. First, the availability of genotyping chips containing hundreds of thousands of markers, which provide good coverage of much of the common genetic variation within the genome, has meant that GWA studies are now financially and technically feasible. Second, the publication of the International Haplotype Map, which documents the pattern of linkage disequilibrium across the genome, has facilitated the design and analysis of GWA studies (The International HapMap Consortium, 2005). Finally, the existence of large patient cohorts has been a necessary prerequisite in order to obtain the power necessary to detect loci of small to moderate effect (The Wellcome Trust Case Control Consortium, 2007). These developments have led to a flood of GWA studies that have successfully identified genes conferring risk to a variety of diseases including (but not limited to) coronary heart disease (Samani et al. 2007), breast cancer (Easton et al. 2007), types I and II diabetes (Todd et al. 2007; Zeggini et al. 2007), and inflammatory bowel disease (Parkes et al. 2007).

To date the vast majority of GWA studies have not explored the relationship between genomic variation and detailed assessments of endophenotypes (i.e. intermediate physiological measures that lie between distal determinants and proximal disease states) and how these associations emerge longitudinally. This is important because identification of these genes will allow better characterisation of the biological pathways underpinning growth and development, as well as those responsible for degeneration and disease in later life. Recent results from the Wellcome Trust Case Control Consortium (WTCCC) provide a vivid illustration of the utility of this approach. The WTCCC recently identified an association between SNPs in the *FTO* gene and Type II diabetes (The Wellcome Trust Case Control Consortium 2007; Zeggini et al. 2007). However, because investigators had collected additional detailed endophenotype information, they were able to show that the association between *FTO* and type II diabetes could be completely explained by body mass index (i.e. the diabetics had higher body mass indices than the lighter controls), and that fat mass was the variable primarily driving this association (i.e. not lean body mass). In other words, the *FTO* variant was not responsible for increasing risk of diabetes directly, but rather was likely affecting disease risk by increasing individuals' fat mass. In addition, longitudinal information from the ALSPAC cohort indicated that although the association between *FTO* and fat mass was not present at birth, it was apparent as early as seven years of age and was

subsequently maintained into the pubertal period and beyond (Frayling et al. 2007). It is important to realise that this kind of in depth assessment of the effect of the *FTO* gene would never have been possible without the existence of detailed longitudinal endophenotype information like that provided in the ALSPAC cohort.

Progress identifying loci that affect quantitative endophenotypes is likely to be difficult because of the low power to detect genetic association in unselected individuals (as opposed to case-control studies where we can assume that cases have been selected on the basis of their extreme score on some underlying continuous distribution). In addition, quantitative genetics theory (Lynch & Walsh, 2001), studies of model organisms (Mackay, 2001) and the few loci which have been reliably replicated in human genetic association studies (Frayling et al. 2007; Zeggini et al. 2007) suggest that the majority of genes underlying variation in complex quantitative traits will be of small to moderate effect (i.e. responsible for < 1% of the phenotypic variance). It is crucial therefore that any putative genome-wide association study of quantitative endophenotypes be adequately powered to detect small effects of this magnitude.

The ALSPAC cohort is unique in this regard, being one of the few cohorts in the world with literally hundreds of accurate measures on a variety of phenotypes (including immunological measures, sensory functions, motor functions and coordination, cognitive measures, anthropometry, serum biochemistry etc.), many of them longitudinal, on thousands of children throughout their first fifteen years of life. The sheer size of ALSPAC also means that there is considerable power to detect loci responsible for small proportions of the phenotypic variance. Currently 2000 ALSPAC children have been typed genome-wide on the Illumina 317K SNP chip, and we are currently applying for funding to genotype the remaining ~13,000 ALSPAC children. The successful student will have an exciting opportunity to take a leading role in the design and analysis of what will be one of the world's first genome-wide association studies of quantitative endophenotype measures.

### **Plan of investigation**

- a. The project will involve use of several methods of study design and analysis relevant to epidemiology, genetic epidemiology, bioinformatics and biostatistics:
- b. Cleaning, managing and maintaining a large database of genetic information
- c. Employing genome-wide association and other appropriate statistical methodology to identify genetic variants associated with endophenotype variability
- d. Prioritizing SNPs for follow up and fine mapping using ALSPAC and additional cohorts
- e. Organising follow up genotyping where necessary
- f. Using meta-analyses to appropriately pool all data
- g. Using appropriate statistical methods to make causal inference from the pooled analyses

### **References**

- Easton D. F. et al. (2007). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 447, 1087-93.
- Lynch M. & Walsh B. (1997). *Genetics and Analysis of Quantitative Traits*. Sinauer Assoc, Sunderland, MA.
- Frayling T.M. et al. (2007). A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood obesity. *Science*, 316(5826), 889-94.
- Mackay T. F. (2001). The genetic architecture of quantitative traits. *Annu Rev Genet*, 35, 303-39.
- Parkes M. et al. (2007). Sequence variants in the autophagy gene *IRGM* and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet*, 39, 830-2.
- The International HapMap Consortium (2005). A haplotype map of the human genome. *Nature*, 437, 1299-320.

Samani N. J. et al. (2007). Genomewide association analysis of coronary artery disease. *N Engl J Med* (available online).

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

Todd J. A. (2007). Robust associations of four new chromosome regions from genome-wide analysis of type 1 diabetes. *Nat Genet*, 39, 857-64.

Zeggini E. et al. (2007). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*, 316(5829), 1336-41.

**Supervisors**

David Evans, George Davey Smith

**Title of Project: “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 of Project: 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.

Ankylosing spondylitis (AS) is a common inflammatory arthritis affecting 0.4% of white European populations (Braun et al. 1998). AS typically develops in the 3<sup>rd</sup>-4<sup>th</sup> 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:

- (g) Cleaning, managing and maintaining a large database of genetic information
- (h) Employing conditional genetic association analysis, principal components analysis and other appropriate statistical methodology to identify genetic variants associated with Ankylosing Spondylitis.
- (i) Prioritizing SNPs for follow up and fine mapping

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; Matt Brown (University of Queensland, Brisbane, Australia).

## **Title of Project: Australian Genome-wide Association Study of Osteoporosis**

### **Outline of Project:**

Osteoporosis is a common disease causing substantial morbidity and mortality worldwide, the risk of which is predominantly genetically determined. The aim of this study is to identify genetic and environmental factors that contribute to the determination of bone mineral density in the general population in order to better understand the etiology of osteoporosis. The study will take advantage of the availability of several major epidemiological cohorts across Australia, and, by combining with the Rotterdam Study, achieve high power to identify even small genetic effects involved in osteoporosis.

### **Objectives**

1. To identify genetic variants that influence bone mineral density.
2. Investigate gene-gene and gene-environment interactions associated with bone mineral density, and to investigate their effect on bone structure and fracture risk.
3. To follow up significant associations in additional samples so as to ensure robustness and replicability of the findings.
4. Fine map the genomic regions of interest so as to precisely identify the functional variants involved.

### **Background**

Osteoporosis is defined as a systemic skeletal disease characterized by low bone mass and microarchitectural deterioration of bone tissue, with a consequent increase in bone fragility and susceptibility to fracture. The burden of osteoporosis to the general community is considerable, with the direct costs to the Australian community of hip fracture alone estimated at \$400 million in 1995 (Randall et al. 1995), and more recent estimates, including indirect costs, are about \$7 billion per annum. In Sweden, osteoporotic fracture is responsible for more hospital bed-days than breast cancer and prostate cancer combined (Johnell et al. 2004). Osteoporotic fracture risk increases with age; the Australian population over 65 years of age is projected to increase from 2.4 million in 2001 to 5 million by 2031 (Australian Bureau of Statistics). Thus, whilst osteoporosis is already a substantial public health problem, the magnitude of the problem will inexorably increase unless more effective treatments and, more importantly, novel approaches to prevention are implemented.

Twin and family studies have demonstrated that osteoporosis is highly familial, and that the tendency of the condition to run in families is predominantly due to genetic factors. This is true of a wide-range of osteoporosis-related phenotypes including bone mineral density (BMD), bone turnover, and skeletal dimensions associated with growth and fracture risk, as well as fracture risk itself (Slemenda et al. 1996; Arden et al. 1996; Koller et al. 2001; Flicker et al. 1996; Deng et al. 2002). The most commonly used screening tool to identify patients with osteoporosis and at increased risk of osteoporotic fracture is BMD measurement (discussed below). The heritability of BMD measured by a variety of methods in twin and intergenerational studies has been shown to be very high. Studies of female twins have shown heritability of BMD to be 0.57-0.92, including studies of post-menopausal twins (Pocock et al. 1987). Estimates from intergenerational family studies have also shown substantial heritability of BMD (0.44-0.67) (Guegen et al. 1995; Krall et al. 1993; Duncan et al. 2003). Several segregation studies, in families drawn from the general population, and ascertained with probands with more severe phenotypes, have demonstrated that the majority of the heritability of BMD is polygenic (Gueguen et al. 1995; Duncan et al. 2003; Deng et al. 2002; Livshits et al. 2004; Cardon et al. 2000; Ginsburg et al. 2001). In specific populations, substantial monogenic effects have been observed, but this has always been on the background of predominantly polygenic effects (Cardon et al. 2000; Cohen et al. 2003; Deng et al. 2004).

The small magnitude of individual genetic effects, likely gender- and site- specificity of those effects, and the small size of most osteoporosis genetic studies have each contributed to the

paucity of success of family studies in this field. Multiple regions have reported to be linked with BMD, but to date only one gene has been identified in the disease from linkage mapping in populations (BMP2). Association has been reported with a large number of candidate genes, but these findings have generally not been of high statistical significance and replication has been difficult. Further, the genes involved were all selected on the basis that there was a strong prior probability that they were involved in bone development or disease. Hypothesis-free genetic studies have the substantial advantage of the possibility of findings that open up new areas of research regarding disease pathogenesis and potential treatments. Major advances in genotyping technology and study design have made genome-wide association studies the gold standard for gene-mapping in common, complex, genetic diseases such as osteoporosis (Hirschhorn et al. 2005; Todd 2007).

We have recently completed a phase 1 genome-wide association study in osteoporosis, studying 135 unrelated white Caucasian postmenopausal women age 55-80 years with total hip bone mineral density (BMD) measurements from the extreme high and low 6.7% of the normal population distribution. This study is equivalent in power to a cohort study of 850-920 cases, and is the first such study done in osteoporosis. We will combine this data with the findings from a genome-wide association study performed in 5000 white Caucasian women from the Rotterdam Study, and select the most significantly associated 7600 SNPs for follow-up in a further Australian cohort of 1500 women recruited with the same BMD criteria as our Australian phase 1 study. Genotypes from phase 1 and 2 will be analysed as a combined cohort. This study has >80% power to identify polymorphisms contributing 0.5% additive heritability of BMD variation, and will give an unprecedented genome-wide analysis of the genetic determinants of osteoporosis.

The high heritability of osteoporosis does not mean that the environment is unimportant. Several non-genetic factors are known to influence osteoporosis phenotypes including diet, drugs (medications, alcohol and cigarettes) and exercise. Although gene-environment interaction has been little studied in osteoporosis, it is likely that genetic determinants interact with environmental factors to determine osteoporosis risk. Controlling for environmental influences may improve the power of studies to identify genetic risks, as well as reducing the risk of false positive findings (Guo 2000; Hung et al. 2004). Standard methods for analyzing gene-environment interaction include logistic or linear regression, and ANCOVA. Data on environmental factors relevant to osteoporosis will be collected from the cases involved, including dietary calcium intake, serum vitamin D levels, cigarette smoking and weight, and a secondary aim of the study will be to investigate gene-environment interaction. Gene-gene interaction is thought to be involved in most complex human genetic diseases (Moore 2003). Systematic methods for investigating it are being developed for both linkage and association studies. These include classification and regression tree and multivariate adaptive regression spline models, multiple dimensionality reduction methods (Hahn, Ritchie & Moore, 2003), and others. These methods do not depend on genes having strong independent effects, allowing the identification of genes whose effect is only moderate.

### **Plan of investigation**

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

- (j) Cleaning, managing and maintaining a large database of genetic information.
- (k) Employing genome-wide association analysis and other appropriate statistical methodology to identify genetic variants associated with osteoporosis.
- (l) Analysis of gene-gene and gene-environment interaction.
- (m) Prioritizing SNPs for follow up and fine mapping.
- (n) Organising follow up genotyping where necessary.

## References

- Arden NK, Baker J, Hogg C, Baan K, Spector TD. The heritability of bone mineral density, ultrasound of the calcaneus and hip axis length: a study of postmenopausal twins. *J Bone Miner Res.* 1996 Apr;11(4):530-4.
- Cardon LR, Garner C, Bennett ST, Mackay IJ, Edwards RM, Cornish J, Hegde M, Murray MA, Reid IR, Cundy T. Evidence for a major gene for bone mineral density in idiopathic osteoporotic families. *J Bone Miner Res.* 2000 Jun;15(6):1132-7.
- Cohen Z, Kalichman L, Kobylansky E, Malkin I, Almog E, Livshits G. Cortical index and size of hand bones: segregation analysis and linkage with the 11q12-13 segment. *Med Sci Monit.* 2003 Mar;9(3):MT13-20.
- Deng HW, Mahaney MC, Williams JT, Li J, Conway T, Davies KM, Li JL, Deng H, Recker RR. Relevance of the genes for bone mass variation to susceptibility to osteoporotic fractures and its implications to gene search for complex human diseases. *Genet Epidemiol.* 2002; 22(1):12-25.
- Duncan EL, Cardon LR, Sinsheimer JS, Wass JA, Brown MA. Site and gender specificity of inheritance of bone mineral density. *J Bone Miner Res.* 2003;18(8):1531-8.
- Flicker L, Faulkner KG, Hopper JL, Green RM, Kaymacki B, Nowson CA, Young D, Wark JD. Determinants of hip axis length in women aged 10-89 years: a twin study. *Bone.* 1996;18(1):41-5.
- Guéguen R, Jouanny P, Guillemin F, Kuntz C, Pourel J, Siest G. Segregation analysis and variance components analysis of bone mineral density in healthy families. *J Bone Miner Res.* 1995;10(12):2017-22.
- Guo SW. Gene-environment interactions and the affected-sib-pair designs. *Hum Hered.* 2000;50(5):271-85.
- Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics.* 2003;19(3):376-82.
- Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet.* 2005;6(2):95-108.
- Hung RJ, Brennan P, Malaveille C, Porru S, Donato F, Boffetta P, Witte JS. Using hierarchical modeling in genetic association studies with multiple markers: application to a case-control study of bladder cancer. *Cancer Epidemiol Biomarkers Prev.* 2004;13(6):1013-21.
- Johnell O, Kanis JA, Jonsson B, Oden A, Johansson H, De Laet C. The burden of hospitalised fractures in Sweden. *Osteoporos Int.* 2005;16(2):222-8.
- Koller DL, Liu G, Econs MJ, Hui SL, Morin PA, Joslyn G, Rodriguez LA, Conneally PM, Christian JC, Johnston CC Jr, Foroud T, Peacock M. Genome screen for quantitative trait loci underlying normal variation in femoral structure. *J Bone Miner Res.* 2001;16(6):985-91.
- Krall EA, Dawson-Hughes B. Heritable and life-style determinants of bone mineral density. *J Bone Miner Res.* 1993;8(1):1-9.
- Livshits G, Deng HW, Nguyen TV, Yakovenko K, Recker RR, Eisman JA. Genetics of bone mineral density: evidence for a major pleiotropic effect from an intercontinental study. *J Bone Miner Res.* 2004;19(6):914-23.
- Moore JH. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered.* 2003;56(1-3):73-82.
- Pocock NA, Eisman JA, Hopper JL, Yeates MG, Sambrook PN, Eberl S. Genetic determinants of bone mass in adults. A twin study. *J Clin Invest.* 1987;80(3):706-10.
- Randell A, Sambrook PN, Nguyen TV, Lapsley H, Jones G, Kelly PJ, Eisman JA. Direct clinical and welfare costs of osteoporotic fractures in elderly men and women. *Osteoporos Int.* 1995;5(6):427-32.
- Slemenda CW, Turner CH, Peacock M, Christian JC, Sorbel J, Hui SL, Johnston CC. The genetics of proximal femur geometry, distribution of bone mass and bone mineral density. *Osteoporos Int.* 1996;6(2):178-82.
- Todd JA. Statistical false positive or true disease pathway? *Nat Genet.* 2006 Jul;38(7):731-3.

## Supervisors

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

**Title of project: Prevalence and determinants of nonalcoholic fatty liver disease (NAFLD) in adolescence**

**Outline of project:**

The specific objectives of this proposal are:

1. To determine the association of growth trajectories (weight, height and BMI) and changes in waist circumference and fat mass from birth to adolescence and with NAFLD ascertained at age ~17.
2. To determine the association of macronutrients and dietary patterns from birth to childhood with ultrasound scan determined NAFLD and biomarkers for NAFLD ascertained at age ~17.
3. To determine the association of physical activity in childhood with ultrasound scan determined NAFLD and biomarkers of NAFLD ascertained at age ~17.
4. To determine the direction of association between insulin resistance and dyslipidaemia with NAFLD by using the repeat measurements of insulin and lipids at 9, 15 and 17 and determining how these relate to biomarkers for NAFLD assessed at 15 and 17.
5. To examine maternal lifestyle characteristics (including smoking and alcohol consumption) during pregnancy in relation to ultrasound scan determined and biomarkers for NAFLD and to compare these associations with paternal characteristics and the same offspring outcomes.
6. To use genetic variants that are associated with putative causal risk factors for NAFLD to test whether their association is truly causal.

The above objectives are likely to cover more than one distinct PhD. We would anticipate students studying in depth 1-3 of the objectives and indeed adding their own areas of interest in relation to the broad topic of NAFLD in adolescence.

**Background**

NAFLD is characterized by the accumulation of fat in the liver with or without inflammation, fibrosis and cirrhosis, in the absence of substantial alcohol consumption and is considered the hepatic manifestation of the metabolic syndrome<sup>1</sup>. There is evidence that NAFLD is increasing in prevalence in adolescents in Europe and other developed countries, and a suggestion that this could have a major impact on future population levels of cirrhosis<sup>2</sup>. Post-mortem studies in the USA suggest a prevalence of NAFLD of 17% in adolescents (aged 15-19) and that the condition is rare before the age of 10<sup>3</sup>. Our recent work in a general population sample from the USA (NHANES) shows a prevalence of elevated ALT of 8% in 'healthy' adolescents<sup>4</sup>. Amongst adults the longer the duration of NAFLD the greater the likelihood of progression to severe liver disease – fibrosis and cirrhosis, it is therefore likely that individuals with NAFLD in adolescence are at a high risk of severe liver pathology in adulthood if steatosis is not reversed (since by definition having the condition from adolescence into adulthood implies a longer duration than steatosis first appearing in adulthood). However, the current state of art regarding the occurrence and determinants of NAFLD in adolescents is based largely on clinical studies, with relatively small sample sizes and that use cross-sectional or retrospective case control study designs. There is an important need for determining the prevalence of NAFLD in general populations of adolescents in Europe and also for large prospective studies that can determine the key risk factors for NAFLD in adolescents.

**Plan of investigations**

Analyses will be based on data from the Avon Longitudinal Study of Parents and Children (ALSPAC), a large population based prospective birth cohort based at the Department of Social Medicine (<http://www.bristol.ac.uk/alspac/>). Data on NAFLD diagnosed by ultrasound (USS) as well as established biomarkers of NAFLD (including ALT, AST, GGT, total bilirubin, fasting glucose, total cholesterol, HDLc, LDLc, triglycerides, apolipoprotein A1 (ApoA1), fasting insulin, alpha 2 macroglobin and haptoglobin), at ages ~15 and ~17 are available as well as all data on potential determinant of NAFLD (anthropometry, diet, etc.). Students will apply appropriate statistical techniques to the specific objectives, including where appropriate multilevel models (e.g. for growth trajectories) and instrumental variables

analyses (e.g. for Mendelian randomization approaches using genetic variants as instrumental variables to establish causality between exposures of interest and NAFLD).

### **References**

1. Marchesini G, Forlani G. NASH: from liver diseases to metabolic disorders and back to clinical hepatology. *Hepatology*. 2002;35:497-499.
2. Day CP. Liver Disease. Non-alcoholic fatty liver disease: current concepts and management strategies. *Clinical Medicine* 2006;6:19-26.
3. Schwimmer JB, Deutsch R, Kahen T, Lavine JE, Stanley C, Behling C. Prevalence of fatty liver in children and adolescents. *Pediatrics* 2006;118:1388-1393

Fraser A, Longnecker MP, Lawlor DA. Prevalence of elevated alanine-aminotransferase (ALT) and factors associated with ALT levels in adolescence: findings from the US National Health and Examination Survey 2001-2004. *Gastroenterology* 2007; 133:1814-1820.

### **Supervisors**

Prof. Debbie Lawlor & Dr. Abigail Fraser