MRC Integrative Epidemiology Unit
At

Research student opportunities in Integrative Epidemiology 2014/15

SUMMARY OF TOPICS

	Supervisor(s)	Title of project
1	Prof Debbie Lawlor, Prof Caroline Relton, Dr Hannah Elliott	Epigenetic variation, maternal health and pregnancy outcomes in the Born in Bradford study
2	Prof George Davey Smith, Prof John Crosby, Dr Chris Arthur, Dr Evie Stergiakouli	Combining untargeted metabolomics with genetic data to identify causal influences on disease
3	Prof Caroline Relton, Prof David Evans, Dr Lavinia Paternoster, Dr Tom Gaunt	Genome-wide association studies of DNA methylation
4	Prof John Crosby, Dr Chris Arthur, Prof Caroline Relton, Prof George Davey Smith, Dr Tom Gaunt	Mapping variation in histone post-translational modifications and implications for population health
5	Prof George Davey Smith, Prof Caroline Relton	Comparison of maternal and paternal body mass index and their relationship with offspring epigenetic patterns, health and development
6	Prof Marcus Munafò, Dr Nic Timpson, Dr Jennifer Ware	Using recall methods to understand genetic influences on addictive behaviours
7	Prof Marcus Munafò, Dr David Coyle	Active and passive monitoring of health behaviour and mood
8	Prof Marcus Munafò, Prof George Davey Smith	Investigating the causal relationship between mental health and health behaviour
9	Dr Dima Damen, Prof Majid Mirmehdi, Prof Ian Craddock	Long-term video monitoring of eating behaviour
10	Prof Kate Tilling, Dr Laura Howe, Dr Neil Davies	Using Mendelian randomization to infer causality in life course epidemiology
11	Prof Kate Tilling, Dr Nic Timpson, Dr Kate Northstone	Genome-wide association studies of dietary patterns and Mendelian randomization applied to diet

12	Prof Kate Tilling, Prof Caroline Relton, Prof Debbie Lawlor	Modelling epigenetic changes over time
13	Dr Tom Gaunt, Dr Santi Rodriguez, Prof Caroline Relton, Prof Dave Evans	Genome-wide analysis of methylation and selection
14	Prof George Davey Smith, Prof Caroline Relton, Prof Frank Windemeijer	Two step epigenetic Mendelian randomization
15	Dr Nic Timpson, Dr Oliver Stegle, Prof George Davey Smith	The coordinated analysis of multi-omic data: "Synergomics".
16	Dr Nic Timpson, Dr Nicole Soranzo, Prof George Davey Smith	Analysis of common complex traits and genome-wide sequence data
17	Dr Nic Timpson, Prof George Davey Smith, Prof Marcus Munafò	The use of genome-wide data for the design and undertaking 'recall by genotype' experiments
18	Prof Debbie Lawlor, Dr Laura Howe, Prof Mika Ala-Korpela	Metabolomic changes across the menopausal transition and their relationship with health in women
19	Dr Laura Howe, Dr Abigail Fraser	Age at puberty and cardiometabolic health in adolescence
20	Prof Debbie Lawlor, Prof Caroline Relton	Cross generational epigenetics in a multi-generational cohort
21	Prof Caroline Relton, Prof Debbie Lawlor	Epigenetic changes across the menopausal transition and their relationship with health in women
22	Prof Debbie Lawlor, Dr Corrie-Macdonald-Wallis, Dr Abigail Fraser	Lifestyle changes during pregnancy and their impact on long term cardiovascular health in women
23	Prof Debbie Lawlor, Prof Caroline Relton, Prof Mika Ala-Korpela	Maternal pregnancy glycaemia, insulin sensitivity, metabolomics, epigenetics and later offspring adiposity
24	Dr Abigail Fraser, Dr Corrie Macdonald-Wallis, Prof Kate Tilling	Maternal haemoglobin in pregnancy: trajectories, risk factors and consequences for offspring. The Avon Longitudinal Study of Parents and Children (ALSPAC)

25	Dr Luisa Zuccolo, Prof Caroline Relton, Marcus Munafo	Neurodevelopmental consequences of prenatal alcohol exposure: a population based MRI study
26	Dr. Abigail Fraser, Dr. Laura Howe, Prof. Gene Feder	Childhood abuse and intimate partner violence and their long term health sequelae: the Avon Longitudinal Study of Parents and Children
27	Dr Tom Gaunt, Dr Colin Campbell, Jose Soeane	Approaches to data integration and visualisation in multi-omics and epidemiological data
28	Dr Tom Gaunt	Patterns of correlation in DNA methylation: inferring functional relationships from temporal and spatial epigenetic changes
29	Dr Tom Gaunt, Dr Colin Campbell, Jose Soeane	Aggregating methods for analysis of association between methylation and phenotypes or outcomes
30	Dr Nic Timpson, Dr Matt Jones and Dr Ullrich Bartsch Research Assistant, School of Physiology & Pharmacology	Linking sleep phenotypes to psychiatric risk genotypes
31	Dr Nic Timpson, Prof Alastair Poole	Methylation of platelet F2RL3 gene provides a major mechanism for smoking-induced coronary thrombosis and cardiovascular disease risk
32	Dr Nic Timpson, Prof Jeremy Tavare	Investigating cellular, transcriptomic and physiological consequences of a functional adenylate cyclase 3 variant: reverse translation of epidemiological findings.
33	Lavinia Paternoster, Prof Caroline Relton, Sara Brown (University of Dundee)	An epigenome-wide association study of atopic dermatitis
34	Professor Peter Vickerman, Professor Matthew Hickman and Dr Natasha Martin	Mathematical modeling of the impact of HCV vaccination among people who inject drugs.
35	Professor Peter Vickerman, Professor Matthew Hickman and Dr Natasha Martin	Mathematical modeling of an emerging epidemic: sexual hepatitis C transmission among HIV positive men who have sex with men.
36	Dr Sarah Lewis and Prof Caroline Relton	An investigation of intra-uterine nutrition and prenatal development– applying the principle of Mendelian randomization

37	Dr Nic Timpson, Dr Colin Steward, Professor Davey Smith, Dr John Harvey	Systematic analysis of the genetic contribution to hemopoietic transplant failure aimed at improving future donor selection
38	Dr Nicholas Timpson; Emma Clark and Hannah Kuper	Analysis of environmental and genetic contributions to scoliosis in an Indian population.
39	Dr Beate St Pourcain, Prof George Davey Smith	Systematic screen of *omic* resources for variation with autism symptomatology
40	Dr Beate St Pourcain, Prof George Davey Smith	Utilising genome-wide data as an instrument for studying environmental risk factors in complex behaviour
41	Dr Beate St Pourcain, Prof George Davey Smith and Prof Jonathan Sandy	Robust evaluation of the relationship between non-syndromic forms of cleft, biological intermediates as well as developmental outcomes in later life

1 Title: Epigenetic variation, maternal health and pregnancy outcomes in the Born in Bradford study

Background

The Born in Bradford study (or BiB) was established in 2007 to examine how genetic, nutritional, environmental, behavioural and social factors impact on health and development during childhood, and subsequently adulthood in a deprived multi-ethnic population[1]. Between 2007 and 2011 detailed information on socio-economic characteristics, ethnicity and family trees, lifestyle, environmental exposures, physical and mental health on 13,776 pregnant women recruited at around 28 weeks of gestation has been collected. 40% of the women are White British and 45% Pakistani, with the remaining women being other White European (3%), other South Asian (7%) and other ethnicity (6%). Extensive clinical and biochemical measures have been collected during pregnancy, including measures of glucose tolerance, as well as measures of fetal and postnatal growth.

Epigenetic processes are thought to be involved in mediating prenatal and early life influences on long term health. The analysis of DNA methylation in infants at birth can help to understand these potential links, identify biological consequences of prenatal exposure and potentially identify early predictors of developmental trajectories and adverse health consequences. Furthermore, variation in DNA methylation is known to differ markedly between ethnic groups; some of this may be attributable to genetic differences and some to marked differences in environmental and lifestyle factors. Nothing is known about how ethnic variation in epigenetic patterns impacts upon pregnancy outcome.

BiB provides a unique cohort in which to explore ethnic differences in adult DNA methylation and the extent to which it is influenced by genetic / lifestyle characteristics and the relationships of these characteristics, as well as pregnancy characteristics to fetal/offspring DNA methylation. There are marked differences in lifestyle characteristics between the two main ethnic groups in BiB: none of the Pakistani women smoked in pregnancy whereas 47% of the White women did; all of the Pakistani women are life-long tee-totallers whereas over 70% of the White women consumed some alcohol in pregnancy; dietary intake is markedly different between the two. This provides a valuable opportunity to consider the prenatal influences on offspring outcomes and whether epigenetic variation may play a role in mediating the relationship between such exposures and outcomes. It also allows a direct comparison of these pathways between ethnic groups.

Objectives

1. To identify ethnic differences in DNA methylation patterns between Pakistani and white British women the BiB cohort
2. To interrogate the relationship between maternal DNA methylation variation and (a) differing lifestyle patterns (e.g. smoking, alcohol, diet) in pregnant women from the two ethnic groups, and (b) pregnancy characteristics (e.g. variation in fasting and postload glucose, metabolomic and inflammatory profiles)
3. Assess the relationship between maternal lifestyle, glucose, metabolomic and inflammatory profiles in pregnancy and offspring cord blood DNA methylation patterns in the two ethnic groups and determine whether any associations differ between the two groups.
4. Apply causal analysis methods to epigenetic data

Methods

This studentship will consider a range of prenatal exposures and their relationship with offspring DNA methylation.

Epigenome-wide analysis will be undertaken using Illumina HumanMethylation450 BeadChip to identify exposure-specific signatures and methylation-phenotype associations. The studentship will include training in causal analysis methods, including application of the Mendelian randomization approach to epigenetic data[2]. The opportunity to undertake laboratory work will be encouraged as part of a programme of interdisciplinary training. This will involve the validation of differentially methylated regions through quantitative DNA methylation analysis using Pyrosequencing. The candidate will also gain exposure to and develop expertise in bioinformatics in the interpretation of data arising from their studies to explore the functional relevance of the observations made. In summary this studentship will involve epigenetics, epidemiology, statistics and bioinformatics and will be placed within the context of maternal health and offspring epigenome in a multi-ethnic study population.

Disciplines and training

Epigenetic epidemiology

Causal analysis methods
The candidate will have the opportunity to make an extended exchange visit to a collaborating group working in this area.

References

1. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, Fairley L, Lawlor DA, Parslow R, Petherick ES, Pickett KE, Waiblinger D, West J; on behalf of the Born in Bradford Scientific Collaborators Group. Cohort profile: The Born in Bradford multi-ethnic family cohort study. Int J Epidemiol. 2012 Oct 12. [Epub ahead of print]
2. Relton CL, Davey Smith G. Two stage epigenetic Mendelian randomization: a strategy for establishing a causal role for epigenetic factors in pathways to disease Int J Epidemiol 2012; 41(1): 161-76.

Supervisors/Collaborators

Prof Debbie Lawlor, Prof Caroline Relton, Dr Hannah Elliott (Supervisors) and Prof John Wright, Dr Donald Whitelaw and Prof Mika Ala-Korpela (Collaborators)

2 Title: Combining untargeted metabolomics with genetic data to identify causal influences on disease

Background

There is considerable excitement regarding the incorporation of metabolomic data into epidemiological studies, usually focussed around a limited number of labelled metabolites. A key challenge is that as yet unidentified metabolites are likely to play important roles in many disease processes. Datasets from liquid chromatography/mass spectrometry (LC/MS) metabolomic experiments may contain many thousands of peaks (sometimes referred to as metabolite features). Metabolite identification remains a major bottleneck in these investigations [1], being both expensive and highly time consuming. Genetic variants have recently been used to filter unknown metabolites for follow-up [2]. This project will use formal Mendelian randomization (MR) approaches [3] with untargeted LC/MS metabolomic data to provide a highly efficient way of identifying which peaks should be further interrogated. Peaks that reliably associate with genetic variants that in turn associate with important health related outcomes will be selected for follow-up metabolite identification [4], as these will be strong candidates for being on causal pathways to the health outcomes. This work will relate to several disease outcomes and disease-related traits in the Avon Longitudinal Study of Parents and Children (ALPSAC), replicated in TwinsUK and KORA with collaborators Prof Tim Spector and Prof Thomas Illig, and in case-control studies of several diseases. The findings are directly translational as they will identify causal factors related to disease that may be modified to reduce risk.

Objectives

1. To combine Mendelian randomization approaches and LC/MS metabolomics to identify metabolic intermediates clearly associated with genotype groups
2. To investigate the relationship between identified LC/MS 'peaks' and relevant health outcomes.
3. To identify LC/MS 'peaks where their constituents are unknown.

Methods

This project will employ conventional MR and metabolomics. Supervision is by groups with expertise in MR (Davey Smith, Stergiakouli) and mass spectrometry (Crosby and Arthur).

Disciplines and training

Causal and translational epidemiology
Bioinformatics & Computational Medicine
Statistical methodology
The student will have the opportunity to make an extended visit to a collaborating centre.

References

1. Nordstrom A, Lewensohn R. Metabolomics: moving to the clinic. Journal of Neuroimmune Pharmacology 2010; 5: 4-17.1 Jonsson P et al. Extraction, interpretation and validation of information for comparing samples in metabolic LC/MS data sets. Analyst 2005; 130: 701-707:
2. Krumsiek J, Suhre K, Evans AM et al. Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. PLoS Genetics 2012;8:e1003005
3. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiology 2003;32:1-22.
4. Rasche F, Scheubert K, Hufsky F et al. Identifying the unknowns by aligning fragmentation trees. Analytical Chemistry 2012;84:3417-26.

Supervisors/Collaborators

Prof George Davey Smith, Dr Evie Stergiakouli, Prof John Crosby, Dr Chris Arthur (Supervisors), Prof Richard Evershed, Prof Tim Spector, Prof Thomas Illig (Collaborators).

3 Title: Genome-wide association studies of DNA methylation

Background
DNA methylation is a key regulator of gene expression. The genetic basis of gene expression has been investigated in depth across tissues and populations and the results have identified many genetic variants associated with gene expression variation, or expression quantitative trait loci (e-QTLs)[1]. However much less is known about the genetic basis of variation in DNA methylation and whether this explains genetic variation in gene expression. Recent studies have suggested that genetic variants can regulate the mean level of quantitative DNA methylation, as well as variability in this phenotype[2-4]. A better understanding of the genetic contribution to variation in DNA methylation (meth-QTLs) will have many benefits. It will (a) improve understanding of how genetic variation influences gene expression and consequently phenotype; (b) allow partitioning of the genetic contribution to DNA methylation to allow a purer consideration of the environmental influences on DNA methylation and (c) it will potentially identify genetic instruments (proxies for DNA methylation variation) that can be applied in causal analysis methods.

Objectives
The proposed studentship will use existing genome-wide SNP data and genome-wide Illumina HumanMethylation450 BeadChip (HM450k) data available on the Avon Longitudinal Study of Parents and Children (ALSPAC) to undertake GWAS analysis of DNA methylation.

Methods
Genome-wide association analysis of genome-wide methylation will involve association analysis of both common and low frequency variants derived from genome-wide SNP chip platforms (Illumina 550K or 660K) and low density genome-wide sequencing, as well as imputed variants from the thousand genomes and UK10K projects. HM450K data is available at birth, age 7, age 15-17 in the same children and during pregnancy and 17 years later in their mothers. This provides multiple options for defining a methylation 'phenotype'. Dimension reduction approaches will also be considered. Because of computational constraints associated with the extremely large number of phenotypes modelled, statistical analysis will proceed using SNPTEST on a cleaned set of 8365 unrelated individuals of confirmed British ancestry. Replication will be sought through collaboration with cohorts including TwinsUK, KORA and Generation R, depending on the age group and phenotypic definition of the primary ALSPAC analysis undertaken. Analyses will be enhanced through participation in the recently established international Genetics of DNA Methylation Consortium. Bioinformatic tools will be utilised to explore the functional significance of associations observed between genetic variation and DNA methylation in relation to publicly available gene expression data, ENCODE and other sources. This studentship will include training in genetic epidemiology, epigenetics, statistics and bioinformatics

Disciplines and training
Epigenetics
Statistical genetics
Extended visit hosted by a collaborator from the Genetics of DNA Methylation Consortium (listed below as collaborators).

References

1. Grundeberg E et al. mapping cis- and trans-regulatory effects across multiple tissues in twins. Nature Genet 2012; 44(10): 1084-89.
2. Zhang D et al. genetic control of individual differences in gene specific methylation in human brain. Am J Hum Genet 2010; 86: 411-19.
3. Bell JD et al. DNA methylation associates with genetic and gene-expression variation in HapMap samples. Genome Biol 2011; 12(1): R10.
4. Quon G et al. Patterns of methylation heritability in a genome-wide analysis of four brain regions. Nucleic Acids Res 2013; 41(4):2095-2104.

Supervisors/Collaborators
Prof Caroline Relton, Prof David Evans, Dr Lavinia Paternoster, Dr Tom Gaunt (Supervisors), Dr Christoph Bock, Dr Jordana Bell, Dr Bas Heijmans, Prof Jonathan Mill (Collaborators)

4 Title: Mapping variation in histone post-translational modifications and implications for population health

Background

Epigenetic processes play an integral role in the regulation of gene expression in development and disease. Population-based studies of epigenetic variation and its impact on health and development have largely been limited to the analysis of DNA methylation as this is a stable measure, quantifiable in biobanked DNA. It is however only part of the epigenetic landscape; histone modifications provide another very important component, as highlighted in recent ENCODE data[1]. The requirement for native chromatin, expense and labour intensity of analysis of histone modifications has meant that their analysis is not amenable to large sample sets. Building on very recent developments in this field[2,3], this project will involve development of mass spectrometry based approaches to profile multiple histone modifications in collaboration with and drawing upon the extensive expertise within the University of Bristol Department of Chemistry. Histone analysis will be undertaken on samples from the Avon Longitudinal Study of Parents and Children (ALSPAC) which has extensive social, behavioural, lifestyle and clinical data collected prospectively across the life course[4]. This study also has detailed genetic and DNA methylation data and collectively these data sources will underpin analysis of the causes and consequences of variation in histone modifications at a population level. Data generated from these analyses will be incorporated into our openly accessible epigenome browser (ARIES-explorer).

Objectives

1. To develop and apply a novel method of assaying histone post-translation modifications using LC MS/MS
2. To apply this approach to population health sciences
3. To interrogate the data generated to establish the determinants (genetic and environmental) of inter-individual variation in histone modification signatures

Methods

Histone proteins will be extracted from lymphoblastoid cell lines from individuals from the Avon Longitudinal Study of Parents and Children (ALSPAC). These will undergo LC-MS/MS analysis using an OrbiTrap system; the optimization of this protocol will form part of the studentship. The methodology will be optimized for high throughput. Samples will then be selected for histone analysis to test specific hypotheses using a 'recall-by-genotype' study design e.g. individuals with high genetic load for psychiatric disorders compared to those with low genetic load.

Disciplines and training

In summary this studentship will include molecular genetics, chemistry (mass spectrometry), genetic epidemiology, statistics and bioinformatics.

References

1. Ecker JR, Bickmore WA, Barroso I, Pritchard JK, Gilad Y, Segal E. Genomics: ENCODE explained. Nature. 2012 Sep 6;489(7414):52-5.
2. Tian Z, Tolić N, Zhao R, Moore RJ, Hengel SM, Robinson EW, Stenoien DL, Wu S, Smith RD, Paša-Tolić L.Enhanced top-down characterization of histone post-translational modifications. Genome Biol. 2012 Oct 3;13(10):R86. [Epub ahead of print]
3. Lin S, Garcia BA. Examining histone posttranslational modification patterns by high-resolution mass spectrometry. Methods Enzymol. 2012;512:3-28.
4. Boyd A, Golding J, Macleod J, Lawlor D, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort Profile: The 'Children of the 90s' – the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol 2013;42:111-127

Supervisors/Collaborators

Prof John Crosby, Dr Chris Arthur, Prof Caroline Relton, Prof George Davey Smith, Dr Tom Gaunt (Supervisors)

5 Title: Comparison of maternal and paternal body mass index and their relationship with offspring epigenetic patterns, health and development

Background

Family-based comparisons are a very useful tool to improve causal inference in observational epidemiology. Parental comparisons can be used for exploring the extent to which associations of intrauterine exposures (such as maternal pregnancy smoking, alcohol and BMI) might be causally related to offspring outcomes in later life. The underlying assumption is that if the associations represent causal intrauterine effects, one would expect stronger maternal-offspring associations than paternal-offspring associations, for the same exposure in the father assessed at the time of their partner's pregnancy. Where associations are similar for both parents we assume that they are driven by genetic or postnatal environmental characteristics, such as family socioeconomic position, that would result in an association in both parents. As a proof-of-concept we demonstrated that maternal smoking in pregnancy was strongly associated with lower offspring birth weight[1], whereas paternal smoking is not associated with birth weight[2]. The same principle has been applied to childhood health outcomes including ADHD[3]. Epigenetic marks, including DNA methylation, are integrally involved in the regulation of gene expression, are influenced by a wide range of environmental exposures and impact upon health and development in children. Maternal and paternal comparisons will help to understand whether epigenetic signatures are influenced *in utero* or via paternal influences which may be genetic in origin or post-natally conveyed.

Objectives

1. To identify DNA methylation patterns in offspring associated with prenatal exposures.
2. To assess the relationship between maternal and paternal prenatal exposures and offspring DNA methylation patterns
3. To assess the relationship between differential methylation in offspring and subsequent health outcomes in childhood and adolescence.

Methods

This project will apply the approach of maternal and paternal comparison to consider the influence of BMI (and other prenatal exposures) on offspring epigenetic patterns, and in turn on health and development of offspring. Maternal and paternal BMI will be related to genome-wide methylation patterns in offspring cord blood, age 7 and age15-17 blood. Mendelian randomization utilizing maternal and paternal genetic variants related to body mass index will also be carried out in the studies that have these data.

Disciplines and training

Extended visit to Danish National Birth Cohort and Norwegian Mother and Child Cohort;
Short course in Epigenetic Epidemiology

References

1. Tyrrell J, et al. Genetic variation in the 15q25 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) interacts with maternal self-reported smoking status during pregnancy to influence birth weight. Hum Mol Genet. 2012 Dec 15;21(24):5344-5358.
2. Davey Smith G. Assessing intrauterine influences on offspring health outcomes: can epidemiological findings yield robust results? Basic and Clinical Pharmacology and Toxicology 2008;102:245-256.
3. Langley K, Heron J, Davey Smith G, Thapar A. Maternal and paternal smoking during pregnancy and risk of ADHD symptoms in offspring: testing for intrauterine effects. Am J Epidemiol 2012 176(3): 261-8.

Supervisors/Collaborators

Prof George Davey Smith, Prof Caroline Relton (Supervisors), Thorkild Sorenson (Copenhagen), Camilla Stoltenberg (Oslo) (Collaborators)

6 Title: Using recall methods to understand genetic influences on addictive behaviours

Background
Genome-wide association studies are beginning to identify variants reliability associated with addictive behaviours (e.g., tobacco and alcohol use). However, the mediating pathways between genetic variation and behavioural variation remain unclear. Recall methods select informative individuals on the basis of either genotype or behaviour, to enable more intensive, laboratory-based phenotyping, and the exploration of mechanistic pathways. For example, there is now clear evidence that the *CHRNA5-A3-B4* gene cluster influences heaviness of smoking via the ability to tolerate high doses of nicotine that are usually aversive. This can in principle be explored in human participants by taking detailed and precise measures of smoking behaviour (e.g., depth of smoking inhalation, number of puffs per cigarette) or assessing response to nicotine challenge. Neither of these is feasible in large-scale population studies, but can be achieved in smaller-scale studies while retaining acceptable statistical power if participants are recruited on the basis of genotype (i.e., extreme homozygotes) to maximize differences in the biological substrate of interest. As variants emerge for other addictive behaviours (e.g., alcohol consumption), it will be possible to apply these methods more widely to extend our understanding of the neurobiological pathways through which genetic variation gives rise to behavioural variation.

Objectives
To explore the neurobiological mechanisms underlying established associations between genetics variants and addictive behaviours.

Methods
This project will use a range of intensive laboratory phenotyping methods, including structural and functional MRI, pharmacological challenge, and behavioural testing. It will also enable the use of novel data capture techniques (e.g., ambulatory monitoring of behaviour in naturalistic settings) to be explored (e.g., portable smoking topography monitors to measure smoking behaviour over the course of a whole day).

Disciplines and training
Causal analysis methods
Intensive phenotyping of behavioural traits
Experimental psychology
Health behaviours

References

1. Ware JJ, van den Bree M, Munafò MR. (2012). From men to mice: CHRNA5/CHRNA3, smoking behavior and disease. Nicotine Tob Res; 14(11): 1291-9.
2. Zuccolo L, Fitz-Simon N, Gray R, Ring SM, Sayal K, Smith GD, Lewis SJ. (2009). A non-synonymous variant in ADH1B is strongly associated with prenatal alcohol use in a European sample of pregnant women. Hum Mol Genet; 18(22): 4457-66.

Supervisors/Collaborators
Prof Marcus Munafò, Dr Nicholas Timpson, Dr Jennifer Ware (Supervisors)

7 Title: Active and passive monitoring of health behaviour and mood

Background
Traditional epidemiological methods have relied on either detailed assessments collected in-person at clinics by trained staff, or remotely via postal questionnaire. The former enables detailed, rich data to be collected, but in an artificial environment and at considerable expense. The latter enables data to be collected at much lower expense, but the data collected are typically less detailed and reliant on accurate self-report by the participant. However, the rapid development of smartphone and sensor technologies means that it is now possible to capture detailed information on health-related behaviour and mood in naturalistic settings, without the need for research participants to attend laboratory or clinic assessments. This has the potential to dramatically increase the reach of epidemiological studies and clinical trials, and enhance the quantity, quality and validity of data collected. This can take the form of self-report data collected at multiple time points over the course of a day, so that rapid fluctuations in, for example, mood or behaviour can be measured. Alternatively, computerised tasks designed to measure aspects of cognitive function can be delivered via smartphone or the Internet. Both of these methods require active engagement of the participant. Passive methods also exist which exploit the functionality of modern smartphones, such as GPS to collect location information. Data collected using active and passive methods can then be integrated, so that for example relationships between mood, behaviour and location (or other environmental variables) can be determined. These methods have the potential to transform data collection in epidemiology.

Objectives
To develop the use of smartphone technology and other sensing methods to capture data on health-related behaviours and cognitions, using active and passive methods.

Methods
This project will develop and validate a range of active and passive sensing methods to capture data on health-related behaviours and cognitions. These will be validated against existing data-collection methods, including clinic-based assessments, postal questionnaires and Internet surveys, before being deployed in larger studies to address specific research questions.

Disciplines and training
Data capture methods
Informatics and electronics
Experimental psychology
Heath behaviours

References

1. Ferguson SG, Shiffman S. (2011). Using the methods of ecological momentary assessment in substance dependence research--smoking cessation as a case study. Subst Use Misuse; 46(1): 87-95.
2. Worringham C, Rojek A, Stewart I. (2011). Development and feasibility of a smartphone, ECG and GPS based system for remotely monitoring exercise in cardiac rehabilitation. PLoS One; 6(2): e14669.

Supervisors/Collaborators
Prof Marcus Munafò, Dr David Coyle, Dr Andy Skinner (Supervisors)

8 Title: Investigating the causal relationship between mental health and health behaviour

Background
A number of mental health problems, such as anxiety, depression and schizophrenia, are strongly associated with health-related behaviours, such as tobacco and alcohol use. However, the causal nature of these relationships remains unclear. Do the psychiatric disorders predispose to tobacco and alcohol use, or do tobacco and alcohol use contribute to disease pathogenesis? As genetic variants associated with these phenotypes (both mental health problems and health-related behaviours) emerge, it is possible to use these as instrumental variables in Mendelian randomisation analyses to explore causality. This approach has been used successfully to investigate the relationship between cigarette smoking and both body mass index and depression, using single genetic variants associated with tobacco use. However, as yet, the relationship between mental health disorders and health behaviours has not been explored, due to a lack of suitable genetic instruments for common mental health problems. One solution to this problem is to use polygene scores to capture a greater proportion of variance in, for example, depression symptoms, and thereby explored the causal role of depressed mood in tobacco and alcohol use. In principle, this approach enables the causal relationship between mental health and health behaviour to be fully explored, with both directions of causality amenable to this approach.

Objectives
To use individual genetic variants and genetic risk scores to explore the causal nature of the effects of mental health problems on health behaviours (specifically, tobacco and alcohol use).

Methods
This project will explore the use of genetic risk scores as instrumental variables in Mendelian randomisation analyses, including bidirectional Mendelian randomisation, of mental health and health-related behaviours. This could include a number of relationships where the causal nature of the relationship is unclear, such as tobacco use and depression, alcohol use and depression, and cannabis use and psychosis.

Disciplines and training
Causal analysis methods
Experimental psychology
Health behaviours
Mental health disorders

References

1. Freathy RM, Kazeem GR, Morris RW, Johnson PC, Paternoster L, Ebrahim S, Hattersley AT, Hill A, Hingorani AD, Holst C, Jefferis BJ, Kring SI, Mooser V, Padmanabhan S, Preisig M, Ring SM, Sattar N, Upton MN, Vollenweider P, Waeber G, Sørensen TI, Frayling TM, Watt G, Lawlor DA, Whincup PH, Tozzi F, Davey Smith G, Munafò M. (2011). Genetic variation at CHRNA5-CHRNA3-CHRNB4 interacts with smoking status to influence body mass index. Int J Epidemiol; 40(6): 1617-28.
2. Lewis SJ, Araya R, Smith GD, Freathy R, Gunnell D, Palmer T, Munafò M. (2011). Smoking is associated with, but does not cause, depressed mood in pregnancy--a mendelian randomization study. PLoS One; 6(7): e21689.

Supervisors/Collaborators
Prof Marcus Munafò, Prof George Davey Smith, Dr Amy Taylor (Supervisors)

9 Title: Long-term video monitoring of eating behaviour

Background
Obesity, depression and cardiovascular disease are some of the biggest health issues and fastest rising categories of healthcare costs. By 2015, obesity is forecast to cost the NHS £27bn each year, not least because it accounts for over 80% of the risk of developing type 2 diabetes.

Objectives
Tracking the body movement along with detecting objects and active zones within the surrounding environment will allow constant monitoring of activities, such as eating. This is important in management of obesity (where sedentary behaviour is of great interest) but also in malnutrition and dehydration (these being especially important for patients recuperating from surgery), adherence to rehabilitation regimes (e.g. after stroke), self-administration of medicines (e.g. after cardiovascular surgery) and injections (e.g. in the 300,000 people in the UK with type 1 diabetes).

In this project, we aim to investigate novel methods for long-term monitoring and analysis of eating behaviour. By long term, we refer to days, weeks, and months, in order to allow the investigation of behaviour degradation (or improvement). We plan to develop models and techniques that can be generically applied to different users in different settings, be it in their kitchen, their sitting room while watching TV, or while interacting with others. The various aspects of eating behaviour will be monitored, such as time of day, speed, quantities and within-meal breaks. Analysis should focus on intra-person variations (i.e. changes in behaviour across time), as well as inter-person differences (i.e. compared to averages across age and health conditions).

Methods
Bristol has a long-established record in tracking objects, humans and animals[1,2]. For fast, real-time processing, state-of-the-art Active Shape Models with invariant features and Unscented Kalman Filter tracking will be used to model and track body and limbs – as a precursor to behaviour analysis and activity recognition by trajectory modeling and matching, i.e. for analysis of eating behaviour. Alternative approaches that employ bottom-up and top-down automatic model selection methods to model behaviour and activities by reasoning about the temporal and causal correlations among discretely measured events in the scene, will also be explored, for example by using Bayesian Networks[3].

Disciplines and training
Nutrition
Health behaviours
Statistical methods

References

1. Daubney et al., "Real Time Pose Estimation of Articulated Objects using Low-Level Motion," IEEE International Conference on Computer Vision and Pattern Recognition, 2008.
2. J. Chiverton, M. Mirmehdi, X. Xie, "Automatic Bootstrapping and Tracking of Object Contours. IEEE Transactions on Image Processing, pp. 1231–1245, 2012.
3. D. Damen et al., "Explaining Activities as Consistent Groups of Events - A Bayesian Framework using Attribute Multiset Grammars," International Journal of Computer Vision (IJCV), pp 83-102, 2012.

Supervisors/Collaborators
Dr Dima Damen, Prof Majid Mirmehdi, Prof Ian Craddock (Supervisors), Prof Julian Hamilton-Shields, Prof Andy Ness, Prof Ashley Cooper (Collaborators).

10 Title: Using Mendelian randomization to infer causality in life course epidemiology

Background

Observational associations in epidemiology may suffer from residual confounding, even after adjustment for measured variables.[1] Mendelian Randomization (MR) is a tool used to improve the strength of causal inference from observational data.[2-7] In MR studies, genetic variants related to the exposure of interest are used to give a causal, unconfounded[8] estimate of the exposure-outcome association. MR has successfully been applied to many research questions in epidemiology (e.g. [9-11]). However, to date, applications have focused on the situation where there is a single measure of the exposure and a single measure of the outcome. In life course epidemiology, the focus is on assessing the influence of exposures throughout the life course on later health and disease.[12,13]. Answering questions in life course epidemiology often requires the analysis of longitudinal data[14,15]. For example, it has been hypothesized that early infancy could be a 'sensitive period' during which rapid growth is particularly detrimental for later obesity and cardiovascular health. Using appropriate longitudinal analysis methods can help to understand these relationships, but applying MR to such questions would also strengthen causal inference[16,17].

Objectives

The aim of this PhD is to extend the framework of MR, developing methods to apply to situations where either the exposure of interest, the outcome, or both are measured repeatedly.

Methods

Depending on the interests of the student, the PhD could focus on various subject areas, but some potential example applications include:

1. *A single measure of the exposure and repeated measures of an outcome*: Observational data suggests that maternal smoking is associated with maternal gestational weight gain and maternal blood pressure during pregnancy. The causality of these associations is unclear, and MR using longitudinal models of maternal weight gain/blood pressure over pregnancy as the outcome could help to understand this. In particular, the association between smoking and weight gain/BP during different periods of pregnancy may differ.
2. *Repeated measures of the exposure and a single measure of the outcome*: Evidence from MR studies suggests that height has a causal effect on cognitive and behavioural outcomes. This analysis could potentially be extended to determine whether height growth in specific periods represents a sensitive period in the development of the outcomes.
3. *Repeated measures of both the exposure and the outcome*: MR studies have attempted to untangle the causality of observed associations between obesity and depression. It is possible that any causal relationship may differ across the life course, and using trajectories of both fatness and mental health could help to untangle this.

The PhD will use data from the Avon Longitudinal Study of Parents and Children[18,19] and other cohorts with whom the School of Social and Community Medicine have existing links. Depending on the interests of the student, this PhD could involve the application of MR to problems of interest and/or statistical methodology development.

Disciplines and training

Causal and translational epidemiology
Bioinformatics & Computational Medicine

References

1. Lawlor DA, Davey Smith G, Kundu D, Bruckdorfer KR, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? *Lancet.* 2004;363(9422):1724-1727.
2. Davey Smith G, McCarron P, Okasha M, McEwen J. Social circumstances in childhood and cardiovascular disease mortality: prospective observational study of Glasgow University students. *Journal of Epidemioliology & Community Health.* 2001;55:340-341.
3. Davey Smith G, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32(1):1-22.
4. Davey Smith G, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? *BMJ.* 2005;330(7499):1076-1079.

5. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med.* 2008;27(8):1133-1163.
6. Burgess S, Butterworth A, Malarstig A, Thompson SG. Use of Mendelian randomisation to assess potential benefit of clinical intervention. *BMJ.* 2012;345:e7325.
7. Davies NM, Davey Smith G, Windmeijer F, Martin RM. Issues in the reporting and conduct of instrumental variable studies: a systematic review. *Epidemiology.* 2013;24(3):363-369.
8. Davey Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med.* 2007;4(12):e352.
9. Lawlor DA, Nordestgaard BG, Benn M, Zuccolo L, Tybjaerg-Hansen A, Davey Smith G. Exploring causal associations between alcohol and coronary heart disease risk factors: findings from a Mendelian randomization study in the Copenhagen General Population Study. *Eur Heart J.* 2013.
10. Bonilla C, Gilbert R, Kemp JP et al. Using genetic proxies for lifecourse sun exposure to assess the causal relationship of sun exposure with circulating vitamin d and prostate cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2013;22(4):597-606.
11. Vimaleswaran KS, Berry DJ, Lu C et al. Causal relationship between obesity and vitamin D status: bi-directional Mendelian randomization analysis of multiple cohorts. *PLoS Med.* 2013;10(2):e1001383.
12. Nordestgaard BG, Palmer TM, Benn M et al. The effect of elevated body mass index on ischemic heart disease risk: causal estimates from a Mendelian randomisation approach. *PLoS Med.* 2012;9(5):e1001212.
13. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol.* 2002;31:285-293.
14. Ben-Shlomo Y. Rising to the challenges and opportunities of life course epidemiology. *Int J Epidemiol.* 2007;36:481-483.
15. Tilling K, Davies NM, Nicoli E et al. Associations of growth trajectories in infancy and early childhood with later childhood outcomes. *Am J Clin Nutr.* 2011.
16. Tilling K, Howe LD, Ben-Shlomo Y. Commentary: methods for analysing life course influences on health--untangling complex exposures. *Int J Epidemiol.* 2011;40(1):250-252.
17. Howe LD, Tilling K, Benfield L et al. Changes in ponderal index and body mass index across childhood and their associations with fat mass and cardiovascular risk factors at age 15. *PLoS One.* 2010;5(12):e15186.
18. Boyd A, Golding J, Macleod J et al. Cohort Profile: The 'Children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol.* 2012.
19. Fraser A, Macdonald-Wallis C, Tilling K et al. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol.* 2012.

Supervisors/Collaborators

Prof Kate Tilling, Dr Laura Howe, Dr Neil Davies (Supervisors)

11 Title: Genome-wide association studies of dietary patterns and Mendelian randomization applied to diet

Background

Diet is a potentially modifiable exposure which has been linked to many outcomes, both in childhood and throughout adult life. However, study of diet is complicated by three factors which limit the degree to which causality can be inferred from observational studies of diet: measurement error, residual confounding and reverse causality. Some of these limiting factors could be explored using information from genetic associations with dietary patterns[1]. For example, there is some evidence that variations in the *TAS2R38* taste receptor gene is associated with a nutrient intake pattern indicative of healthy eating in some studies[2] but not others[3]. A UK twin study has also provided evidence that dietary patterns have heritability components between 39 and 49%[4]. This PhD offers the chance to examine genotypes related to dietary patterns within Avon Longitudinal Study of Parents and Children (ALSPAC) by using existing data to undertake genome-wdie association studies (GWAS) of dietary patterns and behaviours, and use this information in a Mendelian randomization (MR) framework[5,6]. The results of these analyses will be used to explore dietary patterns in childhood in relation to outcomes in adolescence, including IQ, child behaviour and educational attainment[7-9].

Objectives

1. Carry out a systematic review (and meta-analysis if appropriate) of studies relating genotypes to dietary patterns in both childhood and adulthood.
2. Relate dietary patterns in ALSPAC to metabolomic data used to give fine scale parameters of dietary intake.
3. Conduct a GWAS of the metabolites found to be related to dietary patterns, within ALSPAC (and possibly other studies).
4. Use the results of the GWAS, and previous work on genotypes related to taste[19] to investigate the use of MR to determine possible causality in the relationships between diet and outcomes, both in examples where causality is expected and where the association is thought to be due to confounding/measurement error.

Methods

Associations between childhood diet and later childhood/adolescent outcomes will be assessed in the Avon Longitudinal Study of Parents and Children (11;12), a prospective birth cohort in which over 14,000 pregnant women were recruited in the early 1990s. Extensive data are available on childhood diet; including 3-day diet diaries at 7, 10 and 13 years of age and food frequency questionnaires at 3, 4, 7, 9 and 13 years (http://www.bristol.ac.uk/alspac/researchers/resources-available/data-details/clinic/documents/nutrition-2.pdf). Food group and estimated nutrient intakes are available together with dietary patterns derived using both principal components analysis and cluster analysis. A wide range of later outcomes, both physical and cognitive, and factors that may mediate the relationships between diet and later outcome measures could be investigated (e.g. DXA-assessed body fat and lean mass). To establish the causality of the role of diet, genotype information will be used to examine confounding and reverse causality. For example, where specific SNPs are related to dietary patterns but to mediating factors (e.g. fat mass) only through those dietary patterns, the potential for these as instrumental variables in MR analyses will be explored. Where SNPs are related to the mediating factor but not the dietary patterns, then MR could be used to examine reverse causality and/or residual confounding.

Disciplines and training

Nutritional epidemiology
Causal and translational epidemiology
Statistical methodology

References

1. El-Sohemy A, Stewart L, Khataan N, Fontaine-Bisson B, Kwong P, Ozsungur S, et al. Nutrigenomics of taste - impact on food preferences and food production. Forum Nutr 2007;60:176-82.
2. Feeney E, O'Brien S, Scannell A, Markey A, Gibney ER. Genetic variation in taste perception: does it have a role in healthy eating? Proc Nutr Soc 2011;70(1):135-43.

3. Timpson NJ, Christensen M, Lawlor DA, Gaunt TR, Day IN, Ebrahim S, et al. TAS2R38 (phenylthiocarbamide) haplotypes, coronary heart disease traits, and eating behavior in the British Women's Heart and Health Study. Am J Clin Nutr 2005;81(5):1005-11.
4. Teucher B, Skinner J, Skidmore PM, Cassidy A, Fairweather-Tait SJ, Hooper L, et al. Dietary patterns and heritability of food choice in a UK female twin cohort. Twin Res Hum Genet 2007;10(5):734-48.
5. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey SG. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 2008;27(8):1133-63.
6. Palmer TM, Sterne JA, Harbord RM, Lawlor DA, Sheehan NA, Meng S, et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. Am J Epidemiol 2011;173(12):1392-403.
7. Feinstein L, Sabates R, Sorhaindo A, Rogers I, Herrick D, Northstone K, et al. Dietary patterns related to attainment in school: the importance of early eating patterns. J Epidemiol Community Health 2008;62(8):734-9.
8. Northstone K, Joinson C, Emmett P, Ness A, Paus T. Are dietary patterns in childhood associated with IQ at 8 years of age? A population-based cohort study. J Epidemiol Community Health 2012;66(7):624-8.
9. Wiles NJ, Northstone K, Emmett P, Lewis G. 'Junk food' diet and childhood behavioural problems: results from the ALSPAC cohort. Eur J Clin Nutr 2009;63(4):491-8.
10. Timpson NJ, Heron J, Day IN, Ring SM, Bartoshuk LM, Horwood J, et al. Refining associations between TAS2R38 diplotypes and the 6-n-propylthiouracil (PROP) taste test: findings from the Avon Longitudinal Study of Parents and Children. BMC Genet 2007;8:51.
11. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: The 'Children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol 2012.
12. 12.Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey SG, et al. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol 2012.

Supervisors/Collaborators

Prof Kate Tilling, Dr Nic Timpson, Dr Kate Northstone (Supervisors)

12 Title: Modelling epigenetic changes over time

Background
Epigenetic markings on the genome (which govern gene expression) are dynamic and believed to show plasticity, especially in early life. Changes in epigenetic patterns over time may have important implications for development and health in childhood, and in later life in ageing and age-related disease. Population-based studies of epigenetic variation have largely been limited to the analysis of DNA methylation as this is a stable measure, quantifiable in biobanked DNA but few studies have attempted to examine how DNA methylation changes with age, partly because of the relatively few longitudinal studies to have collected and banked biological samples at multiple time points across the life course, and partly because of the complexity of the methods involved[1].

Objectives
To further develop methods for modelling epigenetic changes over time, and/or relating these to changes in phenotypic variables (e.g. growth).

Methods
This project will use existing data available on the Avon Longitudinal Study of Parents and Children (ALSPAC). Genome-wide DNA methylation data together with a wide array of phenotypic data are available. Methods to be used or developed will depend on the interest of the student. Methods to be used could include: correlating methylation between mothers and offspring, and examining how this correlation changes between birth and 7 and 15 years; identifying sites with high correlation, and also those with the lowest correlation, across generations; identifying those 100 methylation sites that change the most between each pair of sequential time points, and identifying methylation sites most highly correlated with a given phenotype, and examining their change and that in the phenotype.
Methodological development could include: extending linear mixed modelling approaches; developing novel methods such as independent component analysis (in collaboration with Dr J Marchini, Oxford University); developing novel data reduction techniques, including derivatives of principal components and partial least squares; functional data analysis methods such as those which have been used to analyse repeated metabolomic data[1]; and "bump hunting" methods such as those which have been used to identify regions of methylation associated with gestational age at birth[2]. Methylation (and changes in methylation) could be related to changes in traits such as BMI, using similar approaches to those used to examine longitudinal trajectories in continuous outcomes[3] (e.g. growth[3]), and non-normally distributed outcomes (e.g. alcohol use[4]), and methods developed for relating changes in two outcomes simultaneously[5].

Disciplines and training
Causal and translational epidemiology
Bioinformatics
Metabolomics and epigenetics with a laboratory attachment

References

1. Ng J, Barrett LM, Wong A, Kuh D, Davey Smith G, Relton CL. The role of longitudinal studies in epigenetic epidemiology: Challenges and opportunities. Genome Biol. 2012 Jun 29;13(6):246
2. Berk M, EbbelsT and Montana G. A statistical framework for biomarker discovery in metabolomic time course data Bioinformatics 2011; 27 (14): 1979–1985
3. Jaffe AE, Murakami P, Lee H, Leek JT, Fallin MD, Feinberg AP, Irizarry RA. Bump hunting to identify differentially methylated regions in epigenetic epidemiology. In J Epidemiol 2012 41(1): 200-9.
4. Paternoster L, Howe LD, Tilling K, Weedon MN, Freathy RM, Frayling TM, Kemp JP, Smith GD, Timpson NJ, Ring SM, Evans DM, Lawlor DA. Adult height variants affect birth length and growth rate in children. Hum Mol Genet. 2011 Oct 15;20(20):4069-75.
5. Heron J, Macleod J, Munafò MR, Melotti R, Lewis G, Tilling K, Hickman M. Patterns of alcohol use in early adolescene predict problem use at age 16. Alcohol Alcohol. 2012 Mar-Apr;47(2):169-77.
6. Macdonald-Wallis C, Lawlor DA, Palmer T, Tilling K. Multivariate multilevel spline models for parallel growth processes: application to weight and mean arterial pressure in pregnancy. Stat Med. 2012 Nov 20;31(26):3147-64.

Supervisors/Collaborators
Prof Kate Tilling, Prof Caroline Relton, Prof Debbie Lawlor (Supervisors)

13 Title: Genome-wide analysis of methylation and selection

Background
Environmental factors can influence the role of genes in human traits via both natural selection (a long-term, multi-generational response) and epigenetic changes (shorter-term intra-individual or trans-generational response). Extensive literature provides examples of both of these mechanisms (reviewed in 1&2). However, for the first time, the availability of high-density genetic data in multiple populations and high-density genome-wide methylation in large-scale cohorts enables detailed analysis of both mechanisms in parallel to determine the relationship between these two environmental response mechanisms.

Available Data
The Accessible Resource for Integrated Epigenomics Studies (ARIES) is a two year project to acquire genome-wide methylation data on 1000 children and mothers of the Avon Longitudinal Study of Parents and Children (ALSPAC), and combine this with existing rich data resources. The ALSPAC resource of over 10,000 mothers and their children also includes extensive clinic and questionnaire data, genome-wide SNP arrays, expression data, whole genome sequencing and metabolomics (in progress). This unique collection of data provides a unique opportunity to examine the relationship between environmental/lifestyle factors, molecular mechanisms and health.

Objectives and Plan
- Systematically review the available evidence for selection genome-wide, and perform additional analyses using 1000 genomes, HapMap and other datasets to identify loci with strong evidence of natural selection
- Analyse relationship between environmental/lifestyle exposures, differentially methylated loci and loci under selection to determine in what contexts methylation interacts with or provides an alternative to natural selection in facilitating adaptation to different exposures.
- Analyse trans-generational persistence of differential methylation in relation to exposures to assess influence of short-term selective pressures on methylation.

Methods
- Methods to assess natural selection – eg Cross Population Extended Haplotype Homozogysity
- Bioinformatics/biostatistical approaches to integrate and analyse datasets
- Methylome-wide association study (MWAS) using, for example, CpGassoc
- Statistical analyses of natural selection and methylation changes

Disciplines and training
Genetics
Epigenetics
Bioinformatics

References

1. Alegría-Torres JA, Baccarelli A, Bollati V. Epigenetics and lifestyle. Epigenomics. 2011 Jun;3(3):267-77. doi: 10.2217/epi.11.22.
2. Bamshad M, Wooding SP. Signatures of natural selection in the human genome. Nature Reviews Genetics 4, 99-111

Supervisors/Collaborators
Dr Tom Gaunt, Dr Santi Rodriguez, Prof Caroline Relton, Prof Dave Evans (Supervisors)

14 Title: Two step epigenetic Mendelian randomization

Background
Epigenetic processes are thought to contribute to the risk of a variety of common complex diseases. However, observed associations between epigenetic variation and disease can be difficult to interpret due to the possible contribution of confounding or the disease itself causing changes in epigenetic patterns (reverse causation). Innovative strategies are needed to overcome these problems.
We recently presented a strategy, based on the well-established framework of Mendelian randomization[1], to interrogate the causal relationship between exposure, DNA methylation and outcome[2]. The two step approach firstly uses a genetic proxy for the exposure of interest to assess the causal relationship between exposure and methylation. A second step then utilises a genetic proxy for DNA methylation to interrogate the causal relationship between DNA methylation and outcome. This PhD project will seek to apply this novel methodology, testing its assumptions and limitations in a range of different settings using available data.

Objectives
The project will use existing data, available on the Avon Longitudinal Study of Parents and their Children (ALSPAC) as well as other complementary cohorts with data available to address specific hypotheses. Analysis will initially be focused upon established Mendelian randomization tools, such as genetic proxies used for smoking and alcohol use, but will expand beyond this during the course of the studentship. The development of methods for the efficient identification and evaluation of genetic proxies for DNA methylation will form a major part of the project. This is a novel area of methodological development with the potential for wide application.

Methods
Two-step epigenetic MR will apply established instrumental variables analysis methods[3] and will incorporate new aspects such as the use of a 'two step' approach, where two instruments are utilized, the first for the exposure-mediator association and the second for the mediator-outcome association.

Disciplines and training
Epigenetics
Statistical methodology
Bioinformatics
Opportunity for an extended visit to Prof Thomas Illig's (Munich) group to apply two-step epigenetic MR approaches to the KORA Cohort.

References

1. Relton CL, Davey Smith G. Epigenetic Epidemiology of Common Complex Disease: prospects for prediction, prevention and treatment. PLoS Medicine 2010;7:e1000356
2. Relton CL, Davey Smith G. Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int. J. Epidemiol. 2012;41: 161-176.
3. Angrist D, Pischke JS. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press 2009

Supervisors/Collaborators
Prof George Davey Smith, Prof Caroline Relton, Prof Frank Windemeijer (Supervisors), Prof Thomas Illig (Collaborator)

15 Title: The coordinated analysis of multi-omic data: "Synergomics".

Background

Genomic and next generation sequencing technologies have exponentially increased the availability of genotype data whilst reducing the effective cost per information point respectively. Whilst accurate phenotyping at the molecular and whole body physiology level still remains relatively expensive, large-scale data collection approaches are now available and across multiple omic spectra including the metabolome, the transcriptome, the methylome and the wider phenome. Within the new MRC Integrative Epidemiology Unit at the University of Bristol (IEU, http://www.bristol.ac.uk/integrative-epidemiology/), the Avon Longitudinal Study of Parents and Children (ALSPAC, www.bris.ac.uk/alspac) represents a unique opportunity to jointly analyze multiple omic data sets, their characteristics and cross-talk. ALSPAC is a large-scale pregnancy cohort (n~8000 mother/child pairs at the core) where data is now available on extremely detailed, longitudinal, phenotypes, the metabolome (using NMR based technology), the methylome (Illumina 450k methylation array data), the transcriptome (Illumina 48k transcript data) and the genome (from both next generation sequence and array based platforms). These sources of data offer a series of analytical possibilities which may enhance the understanding of relationships between genetic perturbations and health related traits, but also in the dissection of possible biological pathways involved in systems meditating both homeostasis, development and disease predisposition.

Objectives

1. To integrate data from multiple omic sources and interrogate their inter-relationship and association with a range of phenotypes or traits.
2. To address specific hypotheses relating to how certain exposures act upon these molecular phenotypes independently or in a co-ordinated manner.

Methods

The overall design of this work will involve the collection and preparation of data from multi-omic sources. These will be derived from the ALSPAC cohort and will comprise data on (i) whole genome sequence data on 1800 individuals (ii) genome-wide common variant array data on ~8000 mother child pairs (iii) extensive phenotypic data and (iv) extended phenotypic data from metabolomic, methylomic and transcriptomic data sources. All of these will require both quality control measures and diagnostic examination before application to further analyses.

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 multiple sources and potentially multiple time points. 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 (for example undertaking bivariate and higher order analysis of phenotypes, such as genetic variation, DNA methylation and gene expression for the same gene or using dimension reduction techniques to deduce regulatory pathways that are shared or specific to some molecular layers. These approaches will require the development of analytical capability for basic epidemiological methods and data sets, for genome-wide analyses and data sets, for the processing of omic 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.

Disciplines and training

Genomics
Bioinformatics
Statistical analysis methods
Relevant training as part of the Wellcome Trust Advanced Courses
Extended visit to the European Bioinformatics Institute under the supervision of Dr Oliver Stegle

References

1. Wadman, M. Genome miners rush to stake claims. *Nature* **447**, aop (2007).
2. Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genetics* **44**, 1084-9 (2012).

3. Wurtz, P. *et al.* Metabolic signatures of insulin resistance in 7,098 young adults. *Diabetes* **61**, 1372-80 (2012).
4. Tukiainen, T. *et al.* Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci. *Human Molecular Genetics* **21**, 1444-55 (2012).
5. Li, J. & Burmeister, M. Genetical genomics: combining genetics with gene expression analysis. *Human Molecular Genetics. 14 Spec No* **2**, R163-9 (2005).
6. Boerjan, W. & Vuylsteke, M. Integrative genetical genomics in Arabidopsis. *Nature Genetics* **41**, 144-5 (2009).

Supervisors/Collaborators
Dr Nic Timpson, Oliver Stegle, Prof George Davey Smith (Supervisors)

16 Title: Analysis of common complex traits and genome-wide sequence data

Background
A new wave of increasingly detailed genetic information is being produced by next generation sequencing technologies. This opens up new possibilities in understanding the contribution of genetic variation to many traits and diseases. This project focuses on the analysis of a comprehensive spectrum of genetic variation, evidence for association with complex phenotypes and following up on the activity of the UK10K sequencing project (www.uk10k.org). Work will employ genome-wide whole sequence data from the Avon Longitudinal Study of Parents and Children (ALSPAC, www.bris.ac.uk/alspac) along with data from TwinsUK to examine the nature of the genomic architecture of complex traits (across a frequency and variation type spectrum), to explore potentially *de novo* association signals and to follow up confirmed signals using multi-omic data under collection and using genotype based recall analyses.

Objectives
Genome-wide sequence data available on ALSPAC will be used to perform analysis of the genetic contribution to common complex traits at a new level of detail. With existing genetic data (common genetic variation across the genome) and further collections at the level of the phenome, methylome and transcriptome, the potential application of these data is broad and will allow for a number of hypotheses and possible methodological developments to be explored.

Methods
This proposal aims to explore the potential in analyzing the spectrum of genetic variation existing in low read depth genome-wide sequence data alongside a rich phenotypic resource. Work will use these data and imputed data from the entire ALSPAC cohort to assess the impact of intermediate frequency genetic variants on specific phenotypes of interest and to assess the potential for this type of genetic data in the assessment of rare exomic coding variation. There is further potential to develop the use of these data along lines of population genetic inference, the analysis of coding variation and functional changes in the pursuit of recall experiments or Mendelian randomization exercises. In addition, the parallel analysis of data from numerous sources including the methylome, the transcriptome and the environment will be possible.

Disciplines and training
Genomics
Statistical analysis methods
Relevant training as part of the Wellcome Trust Advanced Courses
Extended visit to the Wellcome Trust Sanger Institute under the supervision of Dr Nicole Soranzo

References

1. Morrison, A.C. *et al.* Whole-genome sequence–based analysis of high-density lipoprotein cholesterol. *Nature Genetics* AOP(2013).
2. Wu, M.C. *et al.* Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. *American Journal of Human Genetics* 89, 82-93 (2011).

Supervisors/Collaborators
Dr Nic Timpson, Dr Nicole Soranzo, Prof George Davey Smith (Supervisors)

17 Title: The use of genome-wide data for the design and undertaking of 'recall by genotype' experiments

Background
A programme of research is currently being developed focused on the development and undertaking of recall by genotype (RBG) studies. These involve highly detailed phenotypic measurement in subjects stratified by genetic variation. There are numerous benefits to this approach which will be explored during the studentship including the ability to harness genetic variation to make causal inferences about the role of environmental and behavioural influences on disease. The MRC Integrative Epidemiology Unit (IEU) provides a wide range of opportunities to apply the RBG approach which will capitalize on local expertise and resources in conjunction with external collaborators. Currently highlighted areas of focus for initial RBG experiments are neuropsychological outcomes (such as patterns of neuro-oscillatory variation and associations with sleep and schizophrenia risk), the causal impact of differential lipid levels on patterns of gene regulation in the genome, nicotine reception/smoking behaviour and disease risk and the causal impact of variation in the human microbiome.

Objectives
The aim of this PhD will be to trial various aspects of the RBG design and to develop, execute and evaluate a specific RBG study. Thesis work will be developed by undertaking one such experiment with the ultimate aim of testing this approach whilst providing a contribution to the scientific area of interest.

Methods
Work will explore the properties of RBG methods as a means of applying Mendelian randomization (MR) and undertaking causal analyses. Using *genome-wide genetic data* available from the imputation of complete genetic variant collections down to a minor allele frequency of ~1% alongside comprehensive phenotype databases, the student will test the assumptions made concerning the properties of recall groups. The student will extend previous work on the use of multiple genetic variants for the construction of predictive scores[3], evaluating their ability to increase variance explained whilst retaining the integrity of MR *i.e.* is grouping individuals by multiple genetic variants more informative than grouping by a single genetic variant?
The project will also aim to define the conditions most conducive to RBG designs. Conventional statistical techniques and simulation can be used to assess the power gains from assessing balanced groups of genetic variants of differing frequencies (versus random sampling) according to a series of models of genetic effect working with others within the unit to explore differing approaches to the construction of aggregate gene scores which aim to maximise the variance explained in factors of interest.
RBG designs allow refined phenotypic measurements in informative groups assigned by genotype. One example of work in this area is work on established associations between smoking and DNA methylation at the coagulation related gene locus *F2RL3*. It is postulated that smoking induces differential methylation of this locus which in turn causes cardiovascular disease[4-6]. In the absence of large collections with smoking, methylation and cardiovascular disease data the hypothesis is difficult to test. To overcome this, a two sample approach can be used to identify genetic variants that can act as a surrogate for DNA methylation in a first sample, and these can be used in a second sample to assess the causal relationship between methylation and cardiovascular disease in an MR framework. This two sample approach will be adopted in an appropriate RBG as part of the studentship.

Disciplines and training
Genomics
Causal analysis methods
Statistical methods
Relevant training as part of the programme of Wellcome Trust Advanced Courses

References

1. Greenland, S. An introduction to instrumental variables for epidemiologists. *International Jounal of Epidemiology* **29**, 722-729 (2000).
2. Lawlor, D.A., Harbord, R.M., Sterne, J.A.C., Timpson, N.J. & Davey Smith, G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Statistics in Medicine* **27**, 1133-1163 (2008).
3. Palmer, T.M. *et al.* Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research* (2011).

4. McGuire, S.E. & McGuire, A.L. Don't throw the baby out with the bathwater: enabling a bottom-up approach in genome-wide association studies. *Genome Research* **18**, 1683-5 (2008).
5. Beskow, L.M., Linney, K.N., Radtke, R.A., Heinzen, E.L. & Goldstein, D.B. Ethical challenges in genotype-driven research recruitment. *Genome Research* **20**, 705-9 (2010).
6. Grant, R.W. *et al.* Design of a randomized trial of diabetes genetic risk testing to motivate behavior change: the Genetic Counseling/lifestyle Change (GC/LC) Study for Diabetes Prevention. *Clinical Trials* **8**, 609-15 (2011).
7. Marteau, T.M. *et al.* Effect on adherence to nicotine replacement therapy of informing smokers their dose is determined by their genotype: a randomised controlled trial. *PLoS ONE [Electronic Resource]* **7**, e35249 (2012).

Supervisors/Collaborators
Dr Nic Timpson, Prof George Davey Smith, Prof Marcus Munafo (Supervisors)

18 Title: Metabolomic changes across the menopausal transition and their relationship with health in women

Background

Longitudinal modelling of data from several countries suggests no clear change in CHD risk in women around the time of the menopause, unlike that seen for breast cancer,[1,2] but it is possible that any causal effect of the perimenopause on CHD risk factors, such as lipids, do not result in an effect on CHD events until many years after the timing of the menopause. The large Women's Health Initiative RCT suggested that oestrogen replacement in postmenopausal women did not confer CHD benefit, and was overall detrimental to women's health via increased risk of breast cancer and thromboembolic disease. However, the mean age of women in that trial was 60 years and they were a mean 10 years postmenopausal at randomisation; recent evidence suggests that oestrogen replacement around the time of the menopause may be cardioprotective with no increased risk for other outcomes.[3-5] It has therefore been suggested that the hormonal changes occurring around the time of the menopause could result in adverse cardiometabolic risk factor changes that have long-term effects in terms of increased CHD risk.[3-5] Most studies that have examined changes in cardiometabolic risk factors with repeat measurements have included small numbers (N ~ 50-1000).[6-12] These studies suggest that the menopausal transition is associated with specific adverse changes in lipids and BMI, but that BP changes are primarily driven by ageing. Limited evidence suggests that during the transition women lose lean and gain fat mass and become more insulin resistant.[13-15] To date no studies have examined changes in liver enzymes, though it is increasingly recognised that these are influenced by body fat and are related to future diabetes and, potentially, to CHD.[16-18] Only one study (N~1000) has been able to model reproductive hormonal changes alongside cardiometabolic changes. It found that adverse menopausal related lipid changes were related to changes in endogenous estrogen and gonadotrophins.[10-12] Thus, the most robust evidence is for an effect of perimenopausal changes in reproductive hormones on adverse lipid profiles, though these changes are complex and require further exploration.[19]

Objectives

1. To explore change in 216 labelled metabolites (measured by NMR spectroscopy) longitudinally in women going through perimenopausal changes
2. To examine the association of changes in sex hormones with changes in metabolites over the perimenopause
3. To determine whether any changes in women's metabolome over the perimenopause are related to pre and post menopausal change in carotid intima media thickness
4. To apply causal methods, such as Mendelian randomization, to these analyses

Methods

This studentship will use largely use data from the ALSPAC mother's study in whom ~ 3,000 women will have had repeat (four occasions) assessment of fasting metabolites, as well as fasting glucose, insulin, inflammatory markers, sex-hormones, DXA determined fat and lean mass, BMI and waist circumference. A wide range of life-style, medical, obstetric and socioeconomic data has been repeated collected on these women from 20 years prior to the current perimenopause study and across this study. Genome-wide data are also available. The genetic and Mendelian randomization work will also make use of collaborative studies involved in a large genome-wide consortium of menopausal age and changes.

Integrated PhD training plans

Metabolomics, some chemistry / laboratory experience will be supported
Epidemiology
Causal analyses
Genetic epidemiology
Opportunities for extended national / international visit

References

1. Lawlor DA, Ebrahim S, Davey SG. Role of endogenous oestrogen in aetiology of coronary heart disease: analysis of age related trends in coronary heart disease and breast cancer in England and Wales and Japan. *BMJ* 2002;**325**:311-312.
2. Vaidya D, Becker DM, Bittner V, Mathias RA, Ouyang P. Ageing, menopause, and ischaemic heart disease mortality in England, Wales, and the United States: modelling study of national mortality data. *BMJ* 2011;**343**:d5170

3. Salpeter SR, Walsh JM, Greyber E, et al. Coronary heart disease events associated with hormone therapy in younger and older women. A meta-analysis. *Journal of General Internal Medicine* 2006;**21**:363-366
4. Salpeter SR, Walsh JM, Greyber E, et al. Mortality associated with hormone therapy in younger and older women: a meta-analysis. *Journal of General Internal Medicine* 2004;**19**:791-804.
5. Schierbeck LL, Rejnmark L, Landbo C, et al. Effect of hormone replacement therapy on cardiovascular events in recently postmenopausal women:randomised trial. *BMJ* 2012:**345**:e6409.
6. Do KA, Green A, Guthrie JR, Dudley EC, Burger HG, Dennerstein L. Longitudinal study of risk factors for coronary heart disease across the menopausal transition. *American Journal of Epidemiology* 2000;**151**:584-593.
7. Matthews KA, Meilahn E, Kuller LH, Kelsey SF, Caggiula AW, Wing RR. Menopause and risk factors for coronary heart disease. *New England Journal of Medicine* 1989;**321**:641-646.
8. Jensen J, Nilas L, Christiansen C. Influence of menopause on serum lipids and lipoproteins. *Maturitas* 1990;**12**:321-331.
9. Lindquist O, Bengtsson C, Lapidus L. Relationships between the menopause and risk factors for ischaemic heart disease. *Acta Obstetricia et Gynecologica Scandinavica - Supplement* 1985;**130**:43-47.
10. Matthews KA, Crawford SL, Chae CU, et al. Are changes in cardiovascular disease risk factors in midlife women due to chronological aging or to the menopausal transition? *J Am.Coll.Cardiol.* 2009;**54**:2366-2373.
11. Kuller LH, Meilahn EN, Gutai J, et al. Lipoproteins, estrogens and the menopause. 2004;179-197.
12. Derby CA, Crawford SL, Pasternak RC, Sowers M, Sternfeld B, Matthews KA. Lipid changes during the menopause transition in relation to age and weight: the Study of Women's Health Across the Nation. *Am.J.Epidemiol.* 2009;**169**:1352-1361.
13. Sowers MF, Crutchfield M, Jannausch ML, Russell-Aulet M. Longitudinal changes in body composition in women approaching the midlife. *Ann.Hum.Biol.* 1996;**23**:253-265.
14. Wing RR, Matthews KA, Kuller LH, Meilahn EN, Plantinga PL. Weight gain at the time of menopause. *Archives of Internal Medicine* 1991;**151**:97-102.
15. Macdonald HM, New SA, Campbell MK, Reid DM. Longitudinal changes in weight in perimenopausal and early postmenopausal women: effects of dietary energy intake, energy expenditure, dietary calcium intake and hormone replacement therapy. *Int.J.Obes.Relat Metab Disord.* 2003;**27**:669-676.
16. Fraser A, Harris R, Sattar N, Ebrahim S, Davey Smith G, Lawlor DA. Gamma-glutamyltransferase is associated with incident vascular events independently of alcohol intake. Analysis of the British Women's Heart and Health Study, a systematic review and meta-analysis. *Arteriosclerosis, Thrombosis & Vascular Biology.* 2007; **92**:1058-1061:
17. Fraser A, Thinggaard M, Christensen K, Lawlor DA. Alanine aminotransferase, gamma-glutamyltransferase (GGT) and all-cause mortality: results from a population-based Danish twins study alanine aminotransferase, GGT and mortality in elderly twins. *Liver Int* 2009;**29**:1494-1499.
18. Fraser A, Harris R, Sattar N, Ebrahim S, Davey SG, Lawlor DA. Alanine aminotransferase, gamma-glutamyltransferase, and incident diabetes: the British Women's Heart and Health Study and meta-analysis. *Diabetes Care* 2009;**32**:741-750.
19. Bittner V. Menopause, age, and cardiovascular risk: a complex relationship. *J Am.Coll.Cardiol.* 2009;**54**:2374-2375.

Supervisors/Collaborators

Prof Debbie Lawlor, Prof Mika Ala-Korpela, Dr Laura Howe (Supervisors), Dr Anna Murray, Prof Scott Nelson, Prof Naveed Sattar, Prof Di Kuh (Collaborators)

19 Title: Age at puberty and cardiometabolic health in adolescence

Background

Cardiovascular diseases (CVD) are the leading cause of death worldwide. Whilst the vast majority of cardiovascular events occur from middle age onwards, cardiovascular risk is known to start in early life. Early puberty has been postulated as a risk factor for CVD; studies have demonstrated associations between earlier puberty and greater CVD risk. However, the causality of these associations is not established. This proposal explores the association between age at puberty and cardiometabolic health in late adolescence, using data from the Avon Longitudinal Study of Parents and Children (ALSPAC) primarily.

This project is an opportunity to work on a longitudinal study with repeat measures. Below are a number of potential objectives that could be the basis for more than one distinct PhD. We would anticipate students studying in depth 2-3 of the objectives and adding their own areas of interest in relation to the broad topic of age at puberty and cardiometabolic health later in life. Students may choose to focus on statistical methodology (modelling repeat measures, causal inference), on the more applied objectives, or a mixture of the two.

Objectives

1. To examine associations of age at puberty (age at menarche in females and peak height velocity in females and males) with atherosclerosis (carotid intima media thickness) and measures of cardiac structure and function (from conventional and tissue Doppler echocardiography) at age 17 years.
2. To examine associations of age at puberty with trajectories (changes over time) of blood pressure, lipids, glucose and insulin in adolescence using repeated measures of blood pressure, lipids, glucose and insulin throughout adolescence.

To examine potential pathways linking age at puberty and cardiometabolic health:

3. To assess whether i. declines in physical activity, greater likelihood of smoking initiation, and/or ii. greater subsequent increases in adiposity (measured by BMI, whole-body and central fat mass from age at puberty to age 17) mediate any association between age at puberty and cardiovascular structure and function at age 17.
4. To use a Mendelian Randomization approach and/or a sibling study design to improve causal inference regarding the relationship between age at puberty and cardiometabolic health later in life.

Supervisors/Collaborators

Dr Laura Howe and Dr Abigail Fraser (Supervisors)

20 Title: Cross generational epigenetics in a multi-generational cohort

Background
The inheritance of epigenetic signatures across generations is an area that attracts considerable interest but little evidence exists that this phenomenon occurs in humans[1]. Epigenetic patterns are largely erased post fertilization but it is plausible that some areas of the genome escape this erasure and these may carry with them important information influencing development and ultimately health.
Epidemiological approaches can be applied to study the commonality of phenotypic traits (including molecular phenotypes) between generations and can help to establish the causal pathways which lead to apparent cross generational transmission – whether it be 'real' or arising due to shared genetic, environment or behavioural factors.

Objectives
1. To determine associations of maternal (G0) pregnancy and postnatal (assessed at ~ 16, 19 and 21 years postnatal) genome wide DNA methylation patterns with their offspring (G1) cord blood and later life (assessed at ages 7 and 16 years) and establish whether there are similar patterns of change in G0 and G1 between pregnancy/birth and later life and the extent to which any similar patterns are related to shared household environmental characteristics
2. To determine the association of G0 pregnancy DNA methylation profiles with their G1 daughters pregnancy DNA methylation profiles
3. To determine the association of G1 cord blood DNA methylation profiles with their children's (G2) cord blood methylation profiles
4. To use multivariable regression analyses. bioinformatics information, cross-cohort comparisons and Mendelian randomization, as appropriate, to understand mechanisms that might produce cross generational relationships in DNA methylation patterns

Methods
Analyses will primarily be conducted in the ALSPAC cohort, including ALSPAC-G2. Currently genome-wide DNA methylation data are available at repeat time points for 1000 mother (G0)-offspring (G1) pairs, generated as part of the ARIES project (www.ariesepigenomics.org). To date (July 2013) over 100 G2 (children of the index children in ALSPAC) participants from 90+ families have been recruited, but recruitment has been at different times in the life course, including between fetal life and up to age 4years for the G2 children; currently only a minority of these have been recruited in pregnancy and have cord blood on the G2 infant. Data suitable for DNA extraction is being collected on parents (G1) and offspring G2 and we anticipate that by the end of 2016 (i.e. within the mid/final years of anyone starting a 4-year PhD programme in 2013 or later) we will have at least 150 ALSPAC participants on whom there is DNA methylation data on G0 (pregnancy), G1 (pregnancy and cord blood) and G2 (cord blood) linked generations. There will also be the possibility of undertaking similar analyses in the Pelotas (Brazil) multigenerational cohort or working with PhD students undertaking similar research in those cohorts.

Integrated PhD training plans
Epigenetics
Causal analysis methods
Epidemiology
Some laboratory work and field work will be encouraged

References
1. Davey Smith G. Epigenesis for epidemiologists: does evo-devo have implications for population health research and practice? Int J Epidemiol 2012 41(1): 236-47.

Supervisors/Collaborators
Prof Debbie Lawlor, Prof Caroline Relton (Supervisors), Prof Cesar Victora (collaborators)

21 Title: Epigenetic changes across the menopausal transition and their relationship with health in women

Background

It has been suggested that the hormonal changes occurring around the time of the menopause could result in adverse cardiometabolic risk factor changes that have long-term effects in terms of increased CHD risk.[1-3] Most studies that have examined changes in cardiometabolic risk factors with repeat measurements have included small numbers (N ~ 50-1000).[4-10] These studies suggest that the menopausal transition is associated with specific adverse changes in lipids and BMI, but that BP changes are primarily driven by ageing. Limited evidence suggests that during the transition women lose lean and gain fat mass and become more insulin resistant.[11-13] Only one study (N~1000) has been able to model reproductive hormonal changes alongside cardiometabolic changes. It found that adverse menopausal related lipid changes were related to changes in endogenous estrogen and gonadotrophins.[8-10] As well as hormonal changes, and potential lipid and other metabolic changes occurring at the time of the perimenopause, this time in a woman's life is associated with lifestyle changes, either as a result of trying to manage symptoms or in relation to changing family roles or ageing processes. Epigenetic changes may underlie any sex-hormonal or lifestyle change effects on cardiometabolic outcomes. Nutrition, smoking and other lifestyle factors, as well as common genetic variation, have all been shown to alter DNA methylation (DNAm) patterns.[14] Although DNA methylation patterns are known to change with advancing age (region-specific bi-directional hypo- and hypermethylation[15]) very little is known about how these changes are influenced by the menopausal transition or whether any menopausal sex-hormone related changes in DNA methylation affect menopause related changes in cardiometabolic health. In pilot using ALSPAC data we have observed differences in gene-specific DNA methylation between pre- and post-menopausal women; Illumina 27K DNA methylation array analysis has identified 199 CpG sites that differ by ≥5% between pre- and post-menopausal women (n=24). We are currently extending this work using the Illumina 450K array, validating and replicating these initial findings and exploring their associations with cardiometabolic outcomes.

Objectives

1. To explore change in genome wide DNA methylation patterns longitudinally in women going through perimenopausal changes
2. To examine the association of changes in sex hormones with changes in DNA methylation over the perimenopause
3. To determine the association of lifestyle characteristics such as smoking, physical activity, diet and alcohol consumption with DNA methylation change over the perimenopause
4. To determine whether any changes in women's DNA methylation over the perimenopause are related to changes in lipid, glucose and insulin profiles and with pre and post menopausal change in carotid intima media thickness
5. To apply causal methods, such as Mendelian randomization, to these analyses

Methods

This studentship will use largely use data from the ALSPAC mother's study[16] in whom ~ 3,000 women will have had repeat (four occasions) assessment of fasting metabolites, as well as fasting glucose, insulin, inflammatory markers, sex-hormones, DXA determined fat and lean mass, BMI and waist circumference. DNA methylation is being assessed repeatedly over the same clinics in these women using the Illumina 27K DNA methylation array. A wide range of life-style, medical, obstetric and socioeconomic data has been repeated collected on these women from 20 years prior to the current perimenopause study and across this study. Genome-wide data are also available. The genetic and Mendelian randomization work will also make use of collaborative studies involved in a large genome-wide consortium of menopausal age and changes.

Integrated PhD training plans

Epigenetics including some laboratory experience will be supported
Epidemiology
Causal analyses
Genetic epidemiology

References

1. Salpeter SR, Walsh JM, Greyber E, et al. Coronary heart disease events associated with hormone therapy in younger and older women. A meta-analysis. *Journal of General Internal Medicine* 2006;**21**:363-366
2. Salpeter SR, Walsh JM, Greyber E, et al. Mortality associated with hormone therapy in younger and older women: a meta-analysis. *Journal of General Internal Medicine* 2004;**19**:791-804.
3. Schierbeck LL, Rejnmark L, Landbo C, et al. Effect of hormone replacement therapy on cardiovascular events in recently postmenopausal women:randomised trial. *BMJ* 2012:**345**:e6409.
4. Do KA, Green A, Guthrie JR, Dudley EC, Burger HG, Dennerstein L. Longitudinal study of risk factors for coronary heart disease across the menopausal transition. *American Journal of Epidemiology* 2000;**151**:584-593.
5. Matthews KA, Meilahn E, Kuller LH, Kelsey SF, Caggiula AW, Wing RR. Menopause and risk factors for coronary heart disease. *New England Journal of Medicine* 1989;**321**:641-646.
6. Jensen J, Nilas L, Christiansen C. Influence of menopause on serum lipids and lipoproteins. *Maturitas* 1990;**12**:321-331.
7. Lindquist O, Bengtsson C, Lapidus L. Relationships between the menopause and risk factors for ischaemic heart disease. *Acta Obstetricia et Gynecologica Scandinavica - Supplement* 1985;**130**:43-47.
8. Matthews KA, Crawford SL, Chae CU, et al. Are changes in cardiovascular disease risk factors in midlife women due to chronological aging or to the menopausal transition? *J Am.Coll.Cardiol.* 2009;**54**:2366-2373.
9. Kuller LH, Meilahn EN, Gutai J, et al. Lipoproteins, estrogens and the menopause. 2004;179-197.
10. Derby CA, Crawford SL, Pasternak RC, Sowers M, Sternfeld B, Matthews KA. Lipid changes during the menopause transition in relation to age and weight: the Study of Women's Health Across the Nation. *Am.J.Epidemiol.* 2009;**169**:1352-1361.
11. Sowers MF, Crutchfield M, Jannausch ML, Russell-Aulet M. Longitudinal changes in body composition in women approaching the midlife. *Ann.Hum.Biol.* 1996;**23**:253-265.
12. Wing RR, Matthews KA, Kuller LH, Meilahn EN, Plantinga PL. Weight gain at the time of menopause. *Archives of Internal Medicine* 1991;**151**:97-102.
13. Macdonald HM, New SA, Campbell MK, Reid DM. Longitudinal changes in weight in perimenopausal and early postmenopausal women: effects of dietary energy intake, energy expenditure, dietary calcium intake and hormone replacement therapy. *Int.J.Obes.Relat Metab Disord.* 2003;**27**:669-676.
14. Relton CL, Davey Smith G. Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. *PLoS Med* 2010;**7**:e1000356
15. Bjornsson HT, Sigurdsson MI, Fallin MD, et al. Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 2008;**299**:2877-2883.
16. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J, Macleod J, Molloy L, Ness A, Ring S, Nelson SM, Lawlor DA. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology* 2012; doi:10.1093/ije/dys066

Supervisors/Collaborators

Prof Debbie Lawlor, Dr Caroline Relton (Supervisors), Dr Anna Murray, Prof Scott Nelson, Prof Naveed Sattar, Prof Di Kuh (Collaborators)

22 Title: Lifestyle changes during pregnancy and their impact on long term cardiovascular health in women

Background
Large well conducted systematic reviews and meta-analyses have shown that women who experience hypertensive disorder of pregnancy (HDP) are at marked increased risk of cardiovascular disease (CVD) in later life,[1,2] and that those who experience gestational diabetes (GDM) are at increased risk of later diagnoses of type 2 diabetes.[3] GDM and type 2 diabetes appear to be manifestations of the same underlying condition; during pregnancy women become temporarily insulin resistant via hormonal changes instigated to appropriately feed the developing fetus and this 'carbohydrate stress test' of pregnancy reveals those at risk of glucose intolerance.[4] Consequently, like type 2 diabetes, GDM is a risk factor for future coronary heart disease (CHD).[5] We have shown that 17-18 years post the affected pregnancy, women who experienced HDP had elevated BP, dyslipidaemia, increased glucose and insulin and greater Framingham CVD prediction score compared to those who remained normotensive.[6] Women who experienced GDM had higher fasting glucose and insulin 17-18 years after pregnancy, but other risk factors were not elevated comparing these women to those who were not diagnosed with GDM.[6] The American Heart Association recommends that experience of pregnancy complications form an integral part of screening for, and prevention of, CVD in women.[7] This is based on the assumption that these risk factors increase the predictive ability of existing prediction tools and that pregnancy is a sensitive period for persuading women to make beneficial lifestyle changes. However, it is unknown whether either is true. We have shown that a substantial proportion of women who are smokers quit in early pregnancy,[8] and that in women who quit in the first trimester, BP becomes similar to life-long non-smokers shortly after quitting.[9] However, it is important to note that smoking is related to lower BP in pregnancy, but higher CHD risk in general.[9] Lastly, we have shown that women reduce levels and intensity of physical activity in pregnancy.[10] To our knowledge, no previous study has examined changes in behavioural risk factors across pregnancy and then postnatally to explore the extent to which any pregnancy related changes (beneficial or detrimental) are sustained postnatally and to examine their association with future atherosclerosis.

Objectives
1. To use appropriate statistical methods to develop trajectories of change in lifestyle risk factors – smoking, alcohol, diet and physical activity – from the index pregnancy through to 20+ years postnatal, taking account of subsequent pregnancies, in women.
2. To use the trajectories in objective 1 to determine whether pregnancy specific lifestyle changes are sustained and/or influence later risk factor trajectories post-natally
3. To determine associations of genetic variants with different patterns of lifestyle trajectory through pregnancy and postnatally
4. To examine associations of pregnancy and postnatal risk factor trajectories with mid-life carotid intima media thickness
5. To use Mendelian randomization and other causal methods to determine whether any associations in objective 4 are causal
6. To determine whether any diet and physical activity changes that result from a pregnancy lifestyle advice intervention in overweight and obese women are sustained postnatally

Methods
Objectives 1-5 will largely be completed in the ALSPAC birth cohort,[11,12] with additional contributions of relevant cohorts from the EAGLE consortia for objective 6. For objective 1 a range of statistical methods will be explored to determine the best method for quantifying different patterns of change in order to address objectives 2-5. Objective 6 will use data from the UPBEAT (UK based) and LIMIT (Adelaide, Australia) randomized controlled trials that have each used similar protocols to randomize obese (UPBEAT) or LIMIT (overweight/obese) to a lifestyle intervention aimed at promoting the adoption of pregnancy safe diets and levels of physical activity that limit gestational weight gain and reduce the risk of gestational diabetes and associated adverse perinatal outcomes. Long-term follow-up of mothers and offspring to 3+ years post natal is begin undertaken in these trials.

Integrated PhD training plans
Advanced statistical methods
Epidemiology
Genetic Epidemiology

Causal analyses, including intention to treat analyses

References

1. Bellamy L, Casas JP, Hingorani AD, Williams DJ. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. *BMJ* 2007;**335**:974
2. McDonald SD, Malinowski A, Zhou Q, Yusuf S, Devereaux PJ. Cardiovascular sequelae of preeclampsia/eclampsia: a systematic review and meta-analyses. *Am Heart J* 2008;**156**:918-930.
3. Bellamy L, Casas JP, Hingorani AD, Williams D. Type 2 diabetes mellitus after gestational diabetes: a systematic review and meta-analysis. *Lancet* 2009;**373**:1773-1779.
4. Sattar N, Greer IA. Pregnancy complications and maternal cardiovascular risk: opportunities for intervention and screening? *BMJ* 2002;**325**:157-160.
5. Shah BR, Retnakaran R, Booth GL. Increased risk of cardiovascular disease in young women following gestational diabetes mellitus. *Diabetes Care* 2008;**31**:1668-1669.
6. Fraser A, Nelson SM, Macdonald-Wallis C, et al. Associations of pregnancy complications with calculated cardiovascular disease risk and cardiovascular risk factors in middle age: the Avon Longitudinal Study of Parents and Children. *Circulation* 2012;**125**:1367-1380.
7. Mosca L, Benjamin EJ, Berra K, et al. Effectiveness-based guidelines for the prevention of cardiovascular disease in women--2011 update: a guideline from the american heart association. *Circulation* 2011;**123**:1243-1262.
8. Freathy RM, Ring SM, Shields B, et al. A common genetic variant in the 15q24 nicotinic acetylcholine receptor gene cluster (CHRNA5-CHRNA3-CHRNB4) is associated with a reduced ability of women to quit smoking in pregnancy. *Hum Mol Genet* 2009;**18**:2922-2927.
9. Macdonald-Wallis C, Tilling K, Fraser A, Nelson SM, Lawlor DA. Established preeclampsia risk factors are related to patterns of blood pressure change in normal term pregnancy: findings from the Avon Longitudinal Study of Parents and Children. *J Hypertens* 2011;**29**:1703-1711.
10. Liu J, Blair SN, Teng Y, Ness AR, Lawlor DA, Riddoch C. Physical activity during pregnancy in a prospective cohort of British women: results from the Avon longitudinal study of parents and children. *Eur J Epidemiol* 2011;**26**:237-247.
11. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J, Macleod J, Molloy L, Ness A, Ring S, Nelson SM, Lawlor DA. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology* 2012; doi:10.1093/ije/dys066.
12. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort Profile: The 'Children of the 90s'; the index offspring of The Avon Longitudinal Study of Parents and Children (ALSPAC). *International Journal of Epidemiology* 2012; doi:10.1093/ije/dys064.

Supervisors/Collaborators

Prof Debbie Lawlor, Dr Corrie Macdonald-Wallis, Dr Abigail Fraser (Supervisors)

23 Title: Maternal pregnancy glycaemia, insulin sensitivity, metabolomics, epigenetics and later offspring adiposity

Note: this topic includes at least two PhD and appointed students will be expected to define their specific area of research within this broad area and work with any other student(s) working in the same broad area

Background

The developmental overnutrition hypothesis suggests that the increased glucose supply to the fetus resulting from greater maternal fatness or fat gain in pregnancy leads to permanent changes in offspring metabolism, behaviour and appetite regulation with resultant obesity and cardiometabolic problems in later life.[1] However, the underlying mechanisms for these effects in humans are largely unknown.

During the first two trimesters of normal pregnancy maternal fat accumulation increases, fuelled by an increase in lipogenesis. There is an associated, physiologically appropriate, increase in insulin resistance, which causes an increase in circulating maternal glucose which is necessary for normal fetal growth and development. Maternal glucose freely crosses the placenta and stimulates fetal pancreatic insulin secretion (maternal insulin does not cross the placenta) which is a growth hormone and key to normal fetal lipid and glucose metabolism. In the third trimester, there is a decline in maternal fat accumulation, coinciding with an increase in adipose tissue lipolysis in association with the increase in insulin resistance. This leads to an acceleration in the breakdown of maternal fat and results in maternal hyperlipidaemia during the period of maximal fetal growth. In normal pregnancy this is characterised by a threefold increases in plasma triglyceride and lesser increases in plasma total cholesterol, high-density lipoprotein (HDLc) and low-density lipoprotein (LDLc).[2] The increase in maternal circulating lipids allows for the accumulation of a pool of fatty acids in the placenta through hydrolysis of maternal lipoproteins by placental lipases. These fatty acids are transported to the fetal liver, where they are re-esterified into triglycerides, providing a necessary source of lipids for the fetus. The normal physiological effects of pregnancy on glucose, insulin and lipid metabolism appear to be enhanced in women with higher BMI. This results in marked increases in the delivery of fuels to the fetus during the main period of fetal growth. Studies in humans have shown that high triglyceride and free fatty acids (FFA) levels in late pregnancy are associated with greater birthweight and infant percentage body fat.[2] However, it remains unclear whether this developmental overnutrition in late pregnancy results in greater offspring fatness throughout life.

The greater delivery of glucose and FFAs to the developing fetus results in enhanced fetal hepatic and pancreatic responses during the development of these organs. It is feasible that this will have effects on the fetal metabolome which could mediate the impact of developmental overnutrition on later offspring outcomes, though this has not been previously assessed. Nor has there been any assessment of the relationship between the fetal metabolome and later offspring adiposity and cardiometabolic risk factors.

Epigenetic mechanisms are increasingly proposed as key mediating mechanisms in the developmental origins field.[1,2] Whilst it is possible that environmentally responsive, mitotically stable epigenetic phenomenon, such as DNA methylation, could provide a plausible mediating mechanism for developmental overnutrition, and there is some support for this from animal models, as yet, there are few studies that empirically provide evidence for this in humans. Though a recent within sibling study of extremely obese women who had children before and after bariatric surgery provides some evidence that differential DNA methylation at candidate gene sites might mediate the effect of weight loss in these women and subsequent offspring outcomes.[3]

Objectives

1. To determine whether maternal fatness in early pregnancy and fat gain during pregnancy are associated with maternal circulating levels of insulin, glucose and labelled metabolites.
2. To determine whether maternal fatness in early pregnancy, fat gain during pregnancy, circulating insulin, glucose and labelled metabolites are associated with cord blood insulin/c-peptide and labelled metabolites.
3. To determine whether maternal fatness in early pregnancy, fat gain during pregnancy, circulating insulin, glucose and labelled metabolites are associated with cord white cell DNA methylation.
4. To determine whether cord blood insulin/c-peptide or metabolites are associated with indicators of fetal growth and adiposity (assessed by fetal USS measures, birth anthropometry and cord leptin) and also with offspring adiposity and cardiometabolic risk through childhood adolescence and early adulthood.

5. To determine whether cord DNA methylation is associated with fetal growth and adiposity, and also with offspring adiposity and cardiometabolic risk through childhood adolescence and early adulthood.
6. To examine potential mediating pathways (including of maternal and fetal glucose/insulin axes, maternal and fetal metabolome, fetal growth and fetal DNA methylation) between maternal gestational fatness and fat gain and later offspring fatness and cardiometabolic risk factors
7. To use appropriate methods to determine whether these associations are likely to be causal

Methods

Data from the ALSPAC[4,5] and BiB[6] pregnancy/birth cohorts and the UPBEAT and LIMIT[7] pregnancy RCTs will be used to address the objectives, with the most appropriate of these four studies being used for the specific objectives each student focuses their PhD on. The extent to which the associations are due to causal intrauterine processes, as opposed to being due to confounding by shared familial genetic and lifestyle characteristics, will be assessed by (a) multivariable adjustment for a wide-range of potential confounding factors; (b) replication across different studies; (c) cross-population comparisons between the two main ethnic groups within BiB; (d) negative exposure control comparisons for gestational weight gain associations; (e) use of Mendelian randomization, including two-stage Mendelian randomization for determining the extent to which metabolomic and epigenomic mechanisms mediate maternal pregnancy exposure-offspring outcome associations; (f) use of intention to treat analyses in two RCTs.

Integrated PhD training plans

Epigenetics, including the possibility of some laboratory experience
Metabolomics, including the possibility of some laboratory/chemistry experience
Bioinformatics
Epidemiology
Causal analyses
Genetic epidemiology

References

1. Lawlor DA. The Society for Social Medicine John Pemberton Lecture 2011: Developmental overnutrition – an old hypothesis with new importance? *International Journal of Epidemiology* 2013;42:7-29. doi: 10.1093/ije/dys209
2. Lawlor DA, Relton C, Sattar N, Nelson SM. Maternal adiposity—a determinant of perinatal and offspring outcomes. *Nature Endocrine Reviews* 2012; 8:679-688
3. Guénarda F, Deshaiesb Y, Cianflone K, Kral JG, Marceauc P, Vohla M-C. Differential methylation in glucoregulatory genes of offspring born before vs. after maternal gastrointestinal bypass surgery. PNAS 2013; 110: 11439–11444
4. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, Henderson J, Macleod J, Molloy L, Ness A, Ring S, Nelson SM, Lawlor DA. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology* 2012; doi:10.1093/ije/dys066.
5. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Molloy L, Ness A, Ring S, Davey Smith G. Cohort Profile: The 'Children of the 90s'; the index offspring of The Avon Longitudinal Study of Parents and Children (ALSPAC). *International Journal of Epidemiology* 2012; doi:10.1093/ije/dys064.
6. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, Fairley L, Lawlor DA, Parslow R, Petherick ES, Pickett KE, Waiblinger D, West J, on behalf of the Born in Bradford Scientific Collaborators Group. Cohort profile: The born in Bradford multi-ethnic family cohort study. International Journal of Epidemiology 2012; doi:10.1093/ije/dys112.

Supervisors/Collaborators

Prof Debbie Lawlor, Prof Caroline Relton, Prof Mika Ala-Korpela (Supervisors), Prof John Wright, Prof Lucilla Poston, Prof Jodie Dodds, Prof Julie Owen, Prof Scott Nelson, Prof Naveed Sattar, Dr Abi Fraser, Dr Laura Howe, Dr Corrie Macdonald-Wallis, Dr Hannah Elliot (Collaborators)

24 Title: Maternal haemoglobin in pregnancy: trajectories, risk factors and consequences for offspring. The Avon Longitudinal Study of Parents and Children (ALSPAC)

Background

Maternal haemoglobin concentrations are known to decrease in early pregnancy, with a slight increase towards the end of gestation.(1) Women who have haemoglobin concentrations at either the bottom (maternal anaemia) or the top end of the distribution in each trimester have been found to have a higher risk of delivering a low birth weight offspring and delivering preterm compared with those with mid-range haemoglobin levels.(2-3) However, a greater decrease in haemoglobin between trimesters is associated with a reduced risk of preterm birth.(4) Risk factors such as older maternal age, being overweight pre-pregnancy, lower maternal education, multiparity and twin pregnancy have been reported to be associated with maternal anaemia (low haemoglobin) during pregnancy,(5) but it is not clear whether these are also associated with the pattern of change in haemoglobin across pregnancy.

Iron is essential for proper cell differentiation and cell growth and plays an important role in the functioning of the immune system(6) Whilst it has generally been assumed that infants born at term and with an adequate birth weight have adequate iron stores for the first 4–6 months of life, several studies(e.g. 7,8) have shown an association between maternal pregnancy haemoglobin concentrations and iron status in offspring in infancy. Hence it is possible that maternal haemoglobin concentrations are associated with offspring outcomes both in infancy and later on in life.

Objectives

To describe the average pattern of change in haemoglobin concentrations during pregnancy.

To examine associations of potential risk factors with maternal haemoglobin concentrations and/or patterns of changes in haemoglobin across the course of pregnancy.

To investigate the associations of maternal haemoglobin concentrations and/or patterns of changes in haemoglobin during pregnancy with the following outcomes in offspring:

patterns of growth in childhood and adolescence

childhood infections

cardiovascular risk factors (e.g. blood pressure, lipids, insulin, glucose) during childhood and adolescence

cognitive and behavioural outcomes in childhood and adolescence

respiratory health in childhood and adolescence

To assess whether there is a causal relationship between maternal haemoglobin concentrations in pregnancy and offspring outcomes using a Mendelian randomization approach.

To determine whether differential methylation patterns in offspring cord blood mediate any of the observed associations.

The Avon Longitudinal Study of Parents and Children (ALSPAC)

ALSPAC is a longitudinal, population-based pregnancy cohort based in Bristol that recruited 14,541 pregnancies with expected dates of delivery 1st April 1991 to 31st December 1992 (http://www.alspac.bris.ac.uk.)(9)

In ALSPAC all antenatal haemoglobin measurements have been abstracted from obstetric records (median 3 and interquartile range of 2 to 3 measurements per woman). Early and late pregnancy haemoglobin concentrations have been derived from these routinely collected measurements and it is possible to longitudinally model trajectories of change in haemoglobin across pregnancy using multilevel models. In addition, there is genome wide data on both mothers and offspring and DNA methylation is available for a subgroup.

There are a number of potential PhD projects that could be undertaken based on this proposal, from the very statistical to the much more applied. The student will be encouraged to develop a work plan based around their interests. Students may choose to combine a classic epidemiologic approach with a focus on one or more of statistical modeling of repeat measures (in Stata and/or MLwiN) using multilevel models, strengthening causal inference using a Mendelian randomization approach, investigating the role of epigenetics in explaining observed associations.

References:

1. Meng Lu ZG, RL; Cliver, SP; Cutter, G, Blankson M. The relationship between maternal hematocrit and pregnancy outcome. Obs & Gynecol. 1991;77(2):190-4.
2. Steer P, Alam MA, Wadsworth J, Welch A. Relation between Maternal Hemoglobin Concentration and Birth-Weight in Different Ethnic-Groups. BMJ. 1995;310(6978):489-91.
3. Zhou LM, Yang WW, Hua JZ, Deng CQ, Tao XG, Stoltzfus RJ. Relation of hemoglobin measured at different times in pregnancy to preterm birth and low birth weight in Shanghai, China. Am J Epidemiol. 1998;148(10):998-1006.
4. Zhang QY, Ananth CV, Li Z, Smulian JC. Maternal anaemia and preterm birth: a prospective cohort study. Int J Epidemiol. 2009;38(5):1380-9.
5. Zhang QY, Li Z, Ananth CV. Prevalence and risk factors for anaemia in pregnant women: a population-based prospective cohort study in China. Paediatric and Perinatal Epidemiol. 2009;23(4):282-91.
6. Beard JL. Iron biology in immune function, muscle metabolism and neuronal functioning. J Nutr 2001;131:568S-579S.
7. de Pee S, Bloem MW, Sari M, Kiess L, Yip R, Kosen S. The High Prevalence of Low Hemoglobin Concentration among Indonesian Infants Aged 3-5 Months Is Related to Maternal Anemia. *J Nutr* 2002;132:2215-2221.
8. Chaparro CM. Setting the stage for child health and development: prevention of iron deficiency in early infancy. *J Nutr* 2008;138:2529-2533.
9. Fraser A, MacDonal-Wallis C, Tilling K, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. Int J Epidemiol. 2013;42:97-110.

Supervisors/Collaborators

Please contact Abigail Fraser (Abigail.Fraser@bristol.ac.uk) in the first instance.

25 Title: Neurodevelopmental consequences of prenatal alcohol exposure: a population based MRI study

(Note: this topic includes at least two PhD and appointed students will be expected to define their specific area of research within this broad area)

Background

There is clear evidence that heavy alcohol intake during pregnancy disrupts normal brain development with long-term adverse cognitive and behavioural consequences in offspring [1,2]. While gestational alcohol *abuse* remains a problem, occasional use of *small-to-moderate amounts* of alcohol is a more prevalent behaviour during pregnancy, and therefore potentially associated with greater population-level harms. However, it is unknown whether this drinking pattern is harmful for the developing brain, since most recent epidemiological studies fail to find adverse effects [3]. Nevertheless, evidence is emerging from different study designs more robust to biases including confounding, suggesting that even small amounts of alcohol during pregnancy could potentially affect neurodevelopment and result in lower cognitive and academic scores [4,5], as well as from animal models [6]. Epigenetic marks (e.g. DNA methylation changes) resulting in altered regulation of gene expression provide one likely mechanism for the neurodevelopmental effects of alcohol, with key developmental stages of intrauterine and early life thought to be sensitive periods when the epigenome is particularly responsive to external stimuli such as alcohol [7], and preliminary evidence linking epigenetic marks to future disease risk. To our knowledge, the subtle effects of alcohol use on the brain have not been studied yet in large-scale population samples, nor has the potential mediating role of DNA methylation on such effects.

Objectives

The overall aim of this project is to explore the associations of alcohol consumption during pregnancy with offspring structural brain morphology (structural properties of white and grey matter), with particular emphasis on their causal nature and on the role played by DNA methylation. Specific objectives are:

1. To study the effects of prenatal exposure to moderate levels of alcohol on structural brain morphology in school age children and adolescents.

2. To explore the causal nature of the observed associations using causal analysis methods such as cross-cohort comparisons, negative control and Mendelian randomization methods.

3. To investigate whether DNA methylation is a causal mediator of the above associations, by studying whether prenatal alcohol exposure is causally associated with DNA methylation changes detected at birth and persisting into adolescence, and whether such alcohol-induced methylation changes are causally associated with brain MRI and DTI outcomes.

Methods

There are a number of potential PhD projects that could be undertaken based on this proposal. The student will be encouraged to develop a work plan based around their interests.

All potential studentships will involve the use of multiple data sources including three ongoing cohort studies: Generation R [9], the Avon Longitudinal Study of Parents and Children (ALSPAC) [10] and the Saguenay youth study [11], and a number of different methodologies. Exposure and outcome measures can be readily harmonized across the 3 studies, and a wealth of data exists already to allow investigation of possible confounding and mediating effects (e.g. prenatal and postnatal environmental and psycho-cognitive

assessments, genome-wide genetic scans and DNA methylation data). Together, these 3 studies provide enough power to detect or exclude even subtle effects (e.g. [8]), and enough design differences and a wealth of harmonised data to rule out major biases and confounding as explanation of the results. Epigenome-wide DNA methylation data available in all 3 cohorts will allow the interrogation of the mediating role of DNA methylation by using two-step Mendelian randomization, a newly proposed method for establishing whether DNA methylation causally explains (at least part of) a known exposure-outcome association [14]. Uncovering a potential causal pathway will add strength to the accumulating evidence for long-term neurodevelopmental consequences of prenatal alcohol exposure.

The student(s) will have the opportunity to familiarise with different research methodologies. These will include epigenomics and neuroimaging data processing and analyses (e.g. see http://www.godmc.org.uk/, http://www.ariesepigenomics.org.uk and [12]), and causal analysis methods such as cross-cohort comparisons, negative control methods, and Mendelian randomization [13,14].

Disciplines, research skills and training

Epidemiology
Causal analysis methods
Epigenetics
Neuroimaging
The candidate will be provided with opportunities for international collaborations (Rotterdam or Toronto)

References

1. Irner T.B., Substance exposure in utero and developmental consequences in adolescence: a systematic review. Child Neuropsychol, 2012. 18(6): p. 521-49.
2. Lebel C., et al., A Longitudinal Study of the Long-Term Consequences of Drinking during Pregnancy: Heavy In Utero Alcohol Exposure Disrupts the Normal Processes of Brain Development. The Journal of Neuroscience, 2012, 32(44): 15243-15251.
3. Flak A.L., et al., The Association of Mild, Moderate, and Binge Prenatal Alcohol Exposure and Child Neuropsychological Outcomes: A Meta-Analysis. Alcohol Clin Exp Res, 2013 Aug 1. [Epub ahead of print].
4. Zuccolo L., et al., Prenatal alcohol exposure and offspring cognition and school performance. A 'Mendelian randomization' natural experiment. Int J Epidemiol, 2013. 42(5):1358-70.
5. Lewis S.J., et al., Fetal alcohol exposure and IQ at age 8: evidence from a population-based birth-cohort study. Plos One, 2012 2012;7(11):e49407.
6. Valenzuela C.F., et al., Does moderate drinking harm the fetal brain? Insights from animal models. Trends in neurosciences, 2012;35(5):284-292.
7. Haycock P.C., Fetal alcohol spectrum disorders: the epigenetic perspective. Biol Reprod, 2009 Oct;81(4):607-17.
8. El Marroun H., et al., Prenatal tobacco exposure and Brain morphology: A prospective study in young children. Neuropsychopharmacology, 2013 Oct 7. [Epub ahead of print].
9. Jaddoe V.W., et al., The Generation R Study: Design and cohort update 2012. Eur J Epidemiol, 2012 Sep;27(9):739-56.
10. Boyd A., et al., Cohort Profile: The 'Children of the 90s' - the index offspring of the Avon Longitudinal Study of Parents and Children. Int J Epidemiol, 2013. 42(1): p. 111-27.
11. Pausova Z., et al., Genes, maternal smoking and the offspring brain and body during adolescence: design of the Saguenay youth study. Hum Brain Mapp, 2007;28:502-18.

12. White T., et al., Pediatric population-based neuroimaging and the Generation R Study: the intersection of developmental neuroscience and epidemiology. Eur J Epidemiol, 2013. 28(1): p. 99-111.
13. Gage, S.H., et al., Using Mendelian randomisation to infer causality in depression and anxiety research. Depress Anxiety, 2013 Jul 11. [Epub ahead of print].
14. Relton C.L. and Davey Smith G., Two-step epigenetic Mendelian randomization: a strategy for establishing the causal role of epigenetic processes in pathways to disease. Int J Epidemiol, 2012;41(1):161-176.

Supervisors/Collaborators
Dr Luisa Zuccolo and Prof Caroline Relton (Supervisors)
3rd supervisor:
Prof Marcus Munafò, UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology
Role: phd supervision, expertise in genetic and cognitive influences on addictive behaviour.

External advisors:
Prof Tomas Paus, The Rotman Research Institute, University of Toronto
Prof Ingmar Franken, Dept. Clinical Psychology, Erasmus University Rotterdam
Role: expertise in MRI image processing and analysis.

26 Title: Childhood abuse and intimate partner violence and their long term health sequelae: the Avon Longitudinal Study of Parents and Children

Background:

In recent years, the potential role of abuse victimization in both childhood and adulthood in the aetiology of obesity and cardiovascular diseases has attracted increasing interest. This reflects the growing recognition of the role of psychosocial risk factors in cardiovascular risk[1], as well as the co-occurrence of mental and cardiovascular ill-health[2]. Long term health sequelae of childhood abuse are severe and numerous and include depression, anxiety, eating disorders, illicit drug use, alcohol abuse, and risky sexual behaviour. An emerging body of evidence suggests that a history of child abuse is also associated with growth trajectories[3], greater adiposity[4,5], food addiction[6], hypertension[7], diabetes mellitus[8], smoking[9,10], inflammation markers[11], and cardiovascular disease (CVD) events in later life[12-17]. Less is known about the associations of both childhood abuse and IPV with cardiometabolic and mental health in men, though some studies of child abuse have included both men and women[13,15,16], with one study reporting an association of childhood sexual abuse with an increased risk of myocardial infarction in men but not in women[18]. In women, associations of IPV with smoking, high cholesterol, heart disease and stroke have also been found and these were stronger than equivalent associations in men[19]. Interestingly, in the US Nurses Health II Study, physical and sexual IPV in adulthood were not strongly associated with an increased risk of hypertension[20] or diabetes[21], though severe psychological abuse (indicated by only 2% of cohort members) was associated with both.

Evidence in support of a role of abuse in the aetiology of cardiometabolic ill-health stems from several sources. Preclinical and clinical studies demonstrate profound and lasting effects of early stress on the hypothalamic-pituitary-adrenal and noradrenergic stress systems, including heightened glucocorticoid, norepinephrine, and autonomic stress reactivity, as well as altered dopaminergic and serotonergic function[22-25]. Survivors of childhood sexual abuse have increased heart rate and blood pressure response to cognitive and social challenges[26]. Moreover, posttraumatic stress disorder is prevalent among abused women[27] and is associated with increased risks of coronary heart disease, smoking[9], overweight[28], dyslipidemia[29], diabetes mellitus[30], hypertension[31], and possibly inflammation[32]. Finally, epigenetic processes affecting gene expression may be mediators of these postulated pathways[33].

The Avon Longitudinal Study of Parents and Children (ALSPAC) offers a unique opportunity to study patterns of experience and perpetration of abuse across the lifecourse and their association with cardiometabolic health in both men and women, and potentially across two generations in the same family.

Aims:

The following are potential aims for this project and not all have to be addressed.
The student will be encouraged to develop a work plan based around their own interests, results of a literature review (see aim 1) and the data available in ALSPAC and other cohorts.

1. To systematically review the evidence on the association between child and adult abuse with cardiometabolic health and other outcomes of interest.
2. To estimate the concordance between different family members' (mothers, fathers and offspring trios) responses to questions regarding abuse taking into account different time points and wording of the questions
3. To study associations of childhood abuse with cardiometabolic and other outcomes of interest.
4. To describe longitudinal trajectories of IPV and assess their association with cardiometabolic and other outcomes of interest.
5. To investigate the co-occurrence of abuse in childhood and IPV in adulthood and to investigate whether IPV modifies associations of abuse in childhood with cardiometabolic health and other outcomes of interest.
6. To investigate the role of established cardiometabolic risk factors (adiposity, smoking, alcohol intake) in explaining associations between abuse and cardiometabolic health.
7. To investigate the role of mental health in explaining associations of interest.
8. To investigate the potential role of differential DNA methylation in associations of interest.

Data:

The Avon Longitudinal Study of Parent and Children (http://www.bristol.ac.uk/alspac/) is a prospective birth cohort with over 20 years of active follow up of both parents and their offspring[34,35]. Participants have been asked about effects of parental emotional and physical cruelty in childhood, sexual abuse in childhood, and IPV, the latter on more than one occasion. Detailed information is available on socio-economic indicators, mental health, health related behaviours (e.g smoking, alcohol intake, drug use), and various other potential sources of stress such as financial difficulties. Adiposity (BMI and DXA assessed fat mass) and measures of cardiometabolic health (e.g. blood pressure, lipids, glucose, insulin, inflammation markers, measures of cardiac structure and function (offspring), carotid intima media thickness (mums and offspring) and pulse wave velocity (fathers) have also been repeatedly assessed.

Reference List:

1. Neylon A, Canniffe C, Anand S, Kreatsoulas C, Blake GJ, et al. (2013) A global perspective on psychosocial risk factors for cardiovascular disease. Prog Cardiovasc Dis 55: 574-581.
2. Elderon L, Whooley MA (2013) Depression and Cardiovascular Disease. Progress in Cardiovascular Diseases 55: 511-523.
3. Denholm R, Power C, Li L (2013) Adverse childhood experiences and child-to-adult height trajectories in the 1958 British birth cohort. International Journal of Epidemiology 42: 1399-1409.
4. Noll JG, Zeller MH, Trickett PK, Putnam FW (2007) Obesity risk for female victims of childhood sexual abuse: a prospective study. Pediatrics 120: e61-67.
5. Williamson DF, Thompson TJ, Anda RF, Dietz WH, Felitti V (2002) Body weight and obesity in adults and self-reported abuse in childhood. Int J Obes Relat Metab Disord 26: 1075-1082.
6. Mason SM, Flint AJ, Field AE, Austin SB, Rich-Edwards JW (2013) Abuse victimization in childhood or adolescence and risk of food addiction in adult women. Obesity (Silver Spring).
7. Riley EH, Wright RJ, Jun HJ, Hibert EN, Rich-Edwards JW (2010) Hypertension in adult survivors of child abuse: observations from the Nurses' Health Study II. J Epidemiol Community Health 64: 413-418.
8. Rich-Edwards JW, Spiegelman D, Lividoti Hibert EN, Jun HJ, Todd TJ, et al. (2010) Abuse in childhood and adolescence as a predictor of type 2 diabetes in adult women. Am J Prev Med 39: 529-536.
9. Anda RF, Croft JB, Felitti VJ, Nordenberg D, Giles WH, et al. (1999) Adverse childhood experiences and smoking during adolescence and adulthood. JAMA 282: 1652-1658.
10. Nichols HB, Harlow BL (2004) Childhood abuse and risk of smoking onset. J Epidemiol Community Health 58: 402-406.
11. Bertone-Johnson ER, Whitcomb BW, Missmer SA, Karlson EW, Rich-Edwards JW (2012) Inflammation and early-life abuse in women. Am J Prev Med 43: 611-620.
12. Rich-Edwards JW, Mason S, Rexrode K, Spiegelman D, Hibert E, et al. (2012) Physical and sexual abuse in childhood as predictors of early-onset cardiovascular events in women. Circulation 126: 920-927.
13. Dong M, Giles WH, Felitti VJ, Dube SR, Williams JE, et al. (2004) Insights into causal pathways for ischemic heart disease: adverse childhood experiences study. Circulation 110: 1761-1766.
14. Goodwin RD, Stein MB (2004) Association between childhood trauma and physical disorders among adults in the United States. Psychol Med 34: 509-520.
15. Scott KM, Von Korff M, Angermeyer MC, Benjet C, Bruffaerts R, et al. (2011) Association of childhood adversities and early-onset mental disorders with adult-onset chronic physical conditions. Arch Gen Psychiatry 68: 838-844.
16. Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, et al. (1998) Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. Am J Prev Med 14: 245-258.
17. Batten SV, Aslan M, Maciejewski PK, Mazure CM (2004) Childhood maltreatment as a risk factor for adult cardiovascular disease and depression. J Clin Psychiatry 65: 249-254.
18. Fuller-Thomson E, Bejan R, Hunter JT, Grundland T, Brennenstuhl S (2012) The link between childhood sexual abuse and myocardial infarction in a population-based study. Child Abuse Negl 36: 656-665.

19. Breiding MJ, Black MC, Ryan GW (2008) Chronic Disease and Health Risk Behaviors Associated with Intimate Partner Violence—18 U.S. States/Territories, 2005. Annals of Epidemiology 18: 538-544.
20. Mason SM, Wright RJ, Hibert EN, Spiegelman D, Forman JP, et al. (2012) Intimate partner violence and incidence of hypertension in women. Ann Epidemiol 22: 562-567.
21. Mason SM, Wright RJ, Hibert EN, Spiegelman D, Jun HJ, et al. (2013) Intimate partner violence and incidence of type 2 diabetes in women. Diabetes Care 36: 1159-1165.
22. Bremner JD, Vermetten E (2001) Stress and development: behavioral and biological consequences. Dev Psychopathol 13: 473-489.
23. Bremner JD, Vythilingam M, Vermetten E, Adil J, Khan S, et al. (2003) Cortisol response to a cognitive stress challenge in posttraumatic stress disorder (PTSD) related to childhood abuse. Psychoneuroendocrinology 28: 733-750.
24. Teicher MH, Andersen SL, Polcari A, Anderson CM, Navalta CP, et al. (2003) The neurobiological consequences of early stress and childhood maltreatment. Neurosci Biobehav Rev 27: 33-44.
25. Kaufman J, Charney D (2001) Effects of early stress on brain structure and function: implications for understanding the relationship between child maltreatment and depression. Dev Psychopathol 13: 451-471.
26. Heim C, Newport DJ, Heit S, Graham YP, Wilcox M, et al. (2000) Pituitary-adrenal and autonomic responses to stress in women after sexual and physical abuse in childhood. JAMA 284: 592-597.
27. Woods SJ, Campbell JC (1993) Posttraumatic stress in battered women: does the diagnosis fit? Issues Ment Health Nurs 14: 173-186.
28. Dobie DJ, Kivlahan DR, Maynard C, Bush KR, Davis TM, et al. (2004) Posttraumatic stress disorder in female veterans: association with self-reported health problems and functional impairment. Arch Intern Med 164: 394-400.
29. Kagan BL, Leskin G, Haas B, Wilkins J, Foy D (1999) Elevated lipid levels in Vietnam veterans with chronic posttraumatic stress disorder. Biol Psychiatry 45: 374-377.
30. Goodwin RD, Davidson JR (2005) Self-reported diabetes and posttraumatic stress disorder among adults in the community. Prev Med 40: 570-574.
31. Buckley TC, Kaloupek DG (2001) A meta-analytic examination of basal cardiovascular activity in posttraumatic stress disorder. Psychosom Med 63: 585-594.
32. von Kanel R, Hepp U, Kraemer B, Traber R, Keel M, et al. (2007) Evidence for low-grade systemic proinflammatory activity in patients with posttraumatic stress disorder. J Psychiatr Res 41: 744-752.
33. McGowan PO (2013) Epigenomic Mechanisms of Early Adversity and HPA Dysfunction: Considerations for PTSD Research. Front Psychiatry 4: 110.
34. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, et al. (2012) Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. International Journal of Epidemiology.
35. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, et al. (2012) Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International Journal of Epidemiology.

Supervisors/Collaborators: Dr. Abigail Fraser, Dr. Laura Howe, Prof. Gene Feder

27 Title: Approaches to data integration and visualisation in multi-omics and epidemiological data

Background:

New technologies mean we are now able to collect data on thousands or millions of different molecular variables in parallel in an individual sample. The application of these in large population samples like the Avon Longitudinal Study of Parents and Childhood (ALSPAC) is creating extremely powerful data resources for understanding the mechanisms of health and disease. In addition, the availability of a wealth of molecular data in public databases enables rich annotation and enhancement of population-based studies. However, data integration and analysis of these high-dimensional datasets presents specific computational challenges, and new methods are required to maximize their potential.

Objectives:

The aim of this bioinformatics project is to develop new approaches to data integration and visualisation in high-dimensional omics datasets.

Plan:

Beginning with methylation, genetic, expression and metabolomic data generated on samples from the ALSPAC cohort, the student will develop methods to integrate these datasets, identify interesting patterns in the data and relate these to data from public databases. The project will initially be focused on one specific mechanism or phenotype (to be agreed with student), but the expectation would be that the methods developed will be widely applicable.

Proposed methods/technologies:

Methods will be developed as part of the project, but amongst many others may involve kernel methods, graph-based data integration, GPGPU approaches, data federation and consolidation.

Supervisors/Collaborators: Dr Tom Gaunt, Dr Colin Campbell, Jose Soeane

28 Title: Patterns of correlation in DNA methylation: inferring functional relationships from temporal and spatial epigenetic changes

Background:

DNA methylation plays an important regulatory role in the expression of genes. Differences in the methylation of a particular part of the genome may occur between different tissues and/or different stages of life and/or different exposure groups. We have measured genome-wide DNA methylation in nearly 1000 children at three time-points, and their mothers at two time-points. We are also generating genome-wide methylation data on a range of tissue samples. We can therefore investigate the networks of genes that correlate in their DNA methylation between different time-points, tissues or exposures to identify potential functional relationships.

Objectives:

The aim of this bioinformatics project is to explore the potential functional relevance of networks of sites in the genome with correlated methylation patterns.

Plan:

The project will initially use correlation analysis of high-dimensional epigenetic data (using high-performance computing). The correlation patterns will then be analysed across a range of different sample groupings to determine common patterns. Data will be linked to public datasources on functionality and epigenomics (eg ENCODE, NIH Roadmap Epigenomics etc), and potentially to relevant phenotypes to determine whether patterns reflect specific functional regulation of genes.

Proposed methods:
Correlation analysis using high-performance computing. Data integration and fusion techniques. High-dimensional and interactive data visualization approaches. There is likely to be a component of methods development and programming in this project.

Supervisors/Collaborators: Dr Tom Gaunt

29 Title: Aggregating methods for analysis of association between methylation and phenotypes or outcomes

Background:

DNA methylation plays an important role in the regulation of genes. We have measured DNA methylation in the peripheral blood of 1000 participants in the Avon Longitudinal Study of Parents and Children (ALSPAC) at three timepoints in childhood. Whilst single-point (i.e single CpG site) association analyses can be informative, they result in a large number of tests and make no use of prior information about the functional importance of the genes in which these sites are based. Aggregating methods which combine information across multiple CpG sites based on genome annotations and pathway information have the potential to reduce the number of tests and increase the power.

Objectives:

The aim of this bioinformatics project is to utilize pathway information to reduce the dimensionality and improve the power of epigenetic association studies.

Plan:

The project will utilize correlation and pathway data to construct defined sets of CpG sites for a specific research question (to be agreed with student). These sets will then be tested for association with the agreed phenotype. A post-hoc analysis will be performed to determine whether a single-point analysis would have been sufficient.

Proposed methods:

Correlation analysis, pathway data extraction and association analyses. Depending on context information from individual CpG sites may be combined before or after association testing.

Supervisors/Collaborators: Dr Tom Gaunt, Dr Colin Campbell, Jose Seoane

30 Title: Linking sleep phenotypes to psychiatric risk genotypes

Background:

This proposal is for a genotype based recall experiment using variation at the genetic locus *ZNF804A* in the ALSPAC cohort and the CRiC sleep labs. *ZNF804a* gene function is unknown, but circumstantial evidence (including in situ mRNA expression in human embryo) indicates roles in neural development. Since genetic variants are largely unrelated to environmental characteristics, the assignment of testing groups by genotype is not subject to the problems of confounding or reverse causality seen in conventional observational studies in patients, in whom antipsychotic medication and profound circadian abnormalities confound sleep parameters.

Testing for a causal relationship between *ZNF804A* variants and detailed sleep neurophysiology in healthy participants has unique potential to shed light on the genetic bases of thalamocortical neural circuit development, function and dysfunction in neuropsychiatry. Our long-term goal is to reveal premorbid biomarkers of disease and to inform novel approaches to improving sleep quality and cognitive function in both healthy and patient populations.

Only the ALSPAC framework provides the unique combination of phenotypic and genotypic data alongside potential for prospective contact with participants required for a recall study. Combining the ALSPAC resources with the polysomnographic facilities in the CRiC Sleep Study Suite consitutes a unique

opportunity to integrate complementary strengths of UoB.

Sleep and rhythmic neuronal activity Rhythmic neuronal activity spanning sub-cortical and cortical regions underpins sleep's functions, with network activity patterns during rapid eye movement (REM) and non-REM (NREM) sleep stages acting in concert to support cognition, particularly memory consolidation[2]. Slow-wave oscillations during NREM sleep act as a clocking signal, coordinating limbic-cortical and thalamo-cortical circuits central to memory processing. Thalamocortical sleep spindles are a second defining feature of NREM sleep; their timing is in turn dependent upon slow-waves, which thereby coordinate spindle activity with hippocampal activity during NREM. These neurophysiological hallmarks are therefore biomarkers of neural circuit function and dysfunction, but genetic determinants of their characteristics in health and disease have not yet been studied.

Sleep and schizophrenia Sleep disturbances are a common complaint in schizophrenia and can cause and/or exacerbate cognitive symptoms[3]. In particular, schizophrenia-associated deficits in attention and memory processing may be attributed to aberrant sleep-related mechanisms. Reductions in the number and power of slow-waves and reductions in sleep spindle density[4] have been correlated with either baseline cognitive deficits or deficits in overnight memory recall in patients[5]. Since current schizophrenia treatments largely fail to treat cognitive symptoms, these sleep abnormalities constitute important targets for novel therapeutic intervention. However, whether they are cause or symptom remains unknown.

ZNF804A and schizophrenia Population variance in risk for schizophrenia is ~80% heritable. Initial GWAS showed evidence for association of common variants at several loci, of which *ZNF804A* was the strongest: a replication study of ~60000 participants by our collaborators in Cardiff demonstrated that the signal SNP rs1344706 had strongest evidence for association[6]. Brain imaging studies show that *ZNF804A* risk variants associate with aberrant hippocampal-cortical connectivity during wakefulness in healthy volunteers[7], but the mechanisms through which *ZNF804A* influences brain function and disease risk are completely unknown.

- **Rationale and objectives**
- Cognitive deficits in schizophrenia respond poorly to current therapies and are likely to be caused or exacerbated by sleep abnormalities.
- Alongside altered sleep architecture, schizophrenia patients present with abnormalities in neural oscillations during sleep, implicating well-defined neural circuits – that also subserve cognition during

wakefulness –in disease aetiology. As an intermediate phenotype, these neurophysiological metrics constitute quantifiable and objective measures relative to, for example, questionnaire-based assessment of sleep quality.

- Variation at the ZNF804A locus is associated with (1) risk of schizophrenia and (2) intermediate neural traits (abnormal fMRI during wakefulness) in healthy volunteers; the mechanisms driving these associations are unknown.
- We propose a genotype based recall experiment using ZNF804a risk allele carriers in the ALSPAC cohort and the CRiC sleep labs. Since genetic variants are largely unrelated to environmental characteristics[1], the assignment of testing groups by genotype is not subject to the problems of confounding or reverse causality seen in conventional observational studies in patients.
- Testing for a causal relationship between ZNF804A variants and detailed sleep neurophysiology has unique potential to shed light on the genetic bases of neural circuit (dys)function in neuropsychiatry, to reveal premorbid biomarkers of disease and to inform novel, sleep-targeting therapeutic approaches.

Methods :

AIMS: To link variation at the *ZN4804a* locus with variation in sleep-dependent neural oscillations quantified using polysomnographic EEG recordings.

HYPOTHESES: Sleep quality has a dramatic impact on quality of life in both the healthy population and in neuropsychiatric patients, many of whom suffer sleep disruption. For example, cognitive deficits in schizophrenia are likely to be caused or exacerbated by sleep abnormalities, which disrupt coordinated activity in hippocampus, thalamus and cortex and thereby impair sleep-dependent memory consolidation. Alongside altered sleep architecture, schizophrenia patients present with abnormalities in neural oscillations during sleep, implicating these well-defined neural circuits – that also subserve cognition during wakefulness – in disease aetiology. As an intermediate phenotype, these neurophysiological metrics constitute quantifiable and objective measures relative to, for example, questionnaire-based assessment of sleep quality.

Variation at the ZNF804A locus is associated with (1) abnormal hippocampal-cortical interactions (inferred from fMRI) during wakefulness in healthy volunteers and (2) increased risk of developing schizophrenia; the mechanisms driving these associations are unknown. The strongest association signal was observed with the SNP rs1344706, located in an intron of ZNF804A that maps to a short region of conserved mammalian sequence on 2q32.1

Based on our work in rodent models (Phillips, Bartsch et al. 2012, Neuron) and a patient cohort (Walmsley et al. 2012 Biological Psychiatry) we hypothesise that carriers of the rs1344706 variant will show (1) reduced densities of thalamocortical sleep spindles during non-REM sleep and (2) attenuated coordinated activity between anterior and posterior cortices during non-REM sleep.

EXPOSURE VARIABLES: ZNF804a SNP rs1344706, homozygous subjects

OUTCOME VARIABLES: EEG-derived measures of sleep neurophysiology including:

Spindle quantification [Spindles are 7-15Hz oscillations in sleep EEG, driven by thalamocortical activity, augmented following learning events and attenuted in schizophrenia. Spindles will be quantified using algorithms written and established by Dr. Ullrich Bartsch in the Jones lab]

Slow-wave activity [SWA is characteristic of non-REM sleep and driven by coordinated transitions of cortical pyramidal neurons between activated 'UP' and quiescent 'DOWN' states. Again, algorithms for SWA analysis are established in the Jones lab]

Anterior-posterior coordination [spindle and SWA predominantly arises in frontal cortices and propagatess posteriorly during non-REM. This propagation will be quantified by assessing the relative timing of antero-posterior EEG activity]

CONFOUNDING: The densities and timing of neural oscillations during non-REM are dependent upon prior waking experience. For example, increased spindle densities occur following novel experiences that recruit memory encoding. Differences in sleep neurophysiology may therefore reflect differences in on-going subject behaviours during wakefulness rather than altered sleep per se. However, published data demonstrate that sleep parameters correlate with behavioural performance and overnight improvement in tasks performed immediately prior to sleep, hence our plan 'normalise' behaviour-dependent effects by training subjects in a finger tapping-based motor sequence task prior to EEG recordings.

Disciplines, research skills and training:

The student will gain detailed working knowledge of neuro-oscillatory function and the phenotyping of neurological activity. This will include working within the sleep laboratories at "CRiC", however there will be opportunity to assess and interrogate working models with Jones and others. There will be a comprehensive introduction and ultimate use of genetic epidemiology approaches containing everything to the handling and undertaking of genomewide association studies to the use of next generation sequence data and the design of new recall by genotype experiments. The candidate will have full access to SSCM short courses and regular contact time with supervisors across the MRC IEU and School of Physiology and Pharmacology.

References

1. Lawlor, D.A., Harbord, R.M., Sterne, J.A., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Statistics in medicine* **27**, 1133-1163 (2008).

2. Phillips, K.G.*, Bartsch, U.*, McCarthy, A.P., Edgar, D.M., Tricklebank, M.D., Wafford, K.A. & Jones, M.W. Decoupling of sleep-dependent cortical and hippocampal interactions in a neurodevelopmental model of schizophrenia. *Neuron* **76**, 526-533 (2012).

3. Wulff, K., Gatti, S., Wettstein, J.G. & Foster, R.G. Sleep and circadian rhythm disruption in psychiatric and neurodegenerative disease. *Nat Rev Neurosci* **11**, 589-599 (2010).

4. Ferrarelli, F., Huber, R., Peterson, M.J., Massimini, M., Murphy, M., Riedner, B.A., Watson, A., Bria, P. & Tononi, G. Reduced sleep spindle activity in schizophrenia patients. *Am J Psychiatry* **164**, 483-492 (2007).

5. Wamsley, E.J., Tucker, M.A., Shinn, A.K., Ono, K.E., McKinley, S.K., Ely, A.V., Goff, D.C., Stickgold, R. & Manoach, D.S. Reduced sleep spindles and spindle coherence in schizophrenia: mechanisms of impaired memory consolidation? *Biol Psychiatry* **71**, 154-161 (2012).

6. Williams, H.J., *et al.* Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Mol Psychiatry* (2011).

7. Esslinger, C., *et al.* Neural mechanisms of a genome-wide supported psychosis variant. *Science* **324**, 605 (2009).

Supervisors/Collaborators: Dr Nic Timpson, Dr Matt Jones

31 Title: Methylation of platelet F2RL3 gene provides a major mechanism for smoking-induced coronary thrombosis and cardiovascular disease risk

Background:

Exposure to tobacco smoke is associated with accelerated atherosclerosis, increased risk of myocardial infarction, stroke, pulmonary artery disease, aortic aneurysm and sudden death[1]. With observational evidence, genomewide association studies concerned principally with the genetics of smoking behaviour have verified a causal role for smoking in cardiovascular risk[2-5]. However, whilst it is known that there is an independent contribution of smoking as a risk factor to cardiovascular disease/ischaemic heart disease (CVD/IHD)[6-15] and it has been possible to characterise the beneficial effects of smoking reduction or cessation[16-21], there has been little progress in identifying the way in which smoking elevates risk.

One of the most compelling possible explanations of the relationship between smoking and CVD risk is the impact of exposure to tobacco smoke on platelet activity and thrombogenesis. Platelet activation is central to thrombosis in myocardial infarction (MI) and ischaemic stroke[22]. Work in the Caerphilly Prospective Study (CaPS) has shown that an index of platelet sensitivity based on whole blood responsiveness to adenosine diphosphate (ADP) induced coagulation and platelet aggregation properties measured in platelet rich plasma are related to a history of MI and ECG diagnosed IHD[23,24]. Elsewhere, spontaneous clotting and platelet counts have shown positive relationships with disease risk (and smoking behaviour)[25,26].

Smoking is detrimental to haemostasis, having effects on the regulation of platelet activity by altering endothelial function and patterns of platelet adhesion and aggregation[27]. There have been observations of acute effects of smoking on platelet aggregation[28], however little detailed examination of the impact of either long-term smoking or cessation on measurements of clotting directly.

In a recent small age-matched study with self-reported smoking data, clear evidence for differential

methylation by smoking status at the Factor 2 receptor-like 3 locus on chromosome 19 was found (cg03636183 ~12% difference in methylation, $p=2.68x10^{-31}$). This association has been independently replicated in over 1,400 individuals from the International COPD Genetics Network[4] and in work undertaken in 185 individuals from the TWINS UK collection[29].

F2RL3 is of relevance for cardiovascular health as it codes for protease-activated receptor 4 (PAR4) which is one of two primary receptors for thrombin expressed on platelets[30]. Recent work supports the postulate that methylation at *F2RL3* is a mediator of the detrimental effects of smoking on coronary heart disease (CHD) and is strongly related to mortality among patients with CHD.

Objectives:

There are three key objectives of the study:

(i) Relationship between methylation status at F2RL3 and cardiovascular intermediates, coagulation, platelet activity and expression - We hypothesise that differential patterns of methylation at the *F2RL3* locus will be associated with patterns of coagulation, platelet specific and cardiovascular risk factors.

(ii) Relationship between methylation status at F2RL3 and cessation/reduction of smoking - We hypothesise that, in part, the beneficial effects of smoking cessation and reduction on vascular outcomes is due to a restoration of normal methylation patterns at *F2RL3*.

(iii) Other effects of smoking on the epigenome - We hypothesise that the impact of exposure to tobacco smoke on methylation will not be restricted to *F2RL3*. In a well powered, association study design based on a dense genomewide methylation array, we will be able to detect further loci which have methylation signatures associated with smoking.

Methods:

(i) Methylation levels at *F2RL3* and an exhaustive cardiovascular diseases risk phenotype profiling. The BWHHS[31] (n=4286) and CaPS[32] (n=2959) have extensive intermediate cardiovascular related health data available from adults aged 47 to 80. DNA samples and smoking status data will be available for 1000 samples from each study for which cardiovascular risk factors, confounding factors, platelet related measurements and induced and spontaneous coagulation challenge test data are available. These data will allow a systematic breakdown of the relationship between smoking, differential methylation at F2RL3 and markers of intermediate cardiovascular risk.

(ii) An examination of the impact of smoking cessation on the patterns of methylation at *F2RL3* in a prospectively collected sub sample of mothers from ALSPAC. Smoking rates and numbers have been assessed from records of smoking exposure taken 11 times during the first 12 years of the study. Samples will be recruited for specific *F2RL3* methylation assessment on the basis of the smoking status of the participant when DNA was extracted. Detailed platelet functional analysis, including assessment of surface receptor levels, receptor signaling, platelet secretion, aggregation and in vitro thrombus formation will be undertaken.

(iii) Other effects of smoking on the epigenome. The inclusion of 300 DNA samples for genomewide methylation analysis (Illumina 450k array) is a crucial part of this investigation as it will allow for exploratory analysis of the impact of smoking on patterns of methylation in the human genome. **Platelet functional analysis will also be performed in further ALSPAC recall studies. Also if appropriate, mouse gene knockout strains will be identified with deletions in genes identified in this part of the study. The Poole laboratory has many years experience analysing platelet function in mouse model systems.**

Disciplines, research skills and training:

The student will gain detailed working knowledge of platelet signalling and functional assays, including a variety of microscopy-based cell biology approaches. This will be provided in the Poole lab, which utilises a wide array of platelet functional analysis approaches.

References:

1. Services USDoHaH. The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. In: U.S. Department of Health and Human Services CfDCaP, National Center for Chronic Disease Pre- vention and Health Promotion, Office on Smoking and Health ed. Atlanta; 2010.
2. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease.[see comment]. *Nature*. 2008;452(7187):638-642.
3. Wang F, Xu C-Q, He Q, et al. Genome-wide association identifies a susceptibility locus for coronary artery disease in the Chinese Han population. *Nature Genetics*. 2011;43(4):345-349.
4. Coronary Artery Disease Genetics C. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nature Genetics*. 2011;43(4):339-344.
5. Schunkert H, Konig IR, Kathiresan S, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics*. 2011;43(4):333-338.
6. Friedman GD, Dales LG, Ury HK. Mortality in middle-aged smokers and nonsmokers. *New England Journal of Medicine*. 1979;300(5):213-217.
7. Shaper AG, Pocock SJ, Walker M, Phillips AN, Whitehead TP, Macfarlane PW. Risk factors for ischaemic heart disease: the prospective phase of the British Regional Heart Study. *Journal of Epidemiology & Community Health*. 1985;39(3):197-209.

8. Criqui MH, Cowan LD, Tyroler HA, et al. Lipoproteins as mediators for the effects of alcohol consumption and cigarette smoking on cardiovascular mortality: results form the Lipid Research Clinics Follow-up Study. *American Journal of Epidemiology*. 1987;126(4):629-637.
9. Ragland DR, Brand RJ. Coronary heart disease mortality in the Western Collaborative Group Study. Follow-up experience of 22 years. *American Journal of Epidemiology*. 1988;127(3):462-475.
10. Shaten BJ, Kuller LH, Neaton JD. Association between baseline risk factors, cigarette smoking, and CHD mortality after 10.5 years. MRFIT Research Group. *Preventive Medicine*. 1991;20(5):655-659.
11. Neaton JD, Wentworth D. Serum cholesterol, blood pressure, cigarette smoking, and death from coronary heart disease. Overall findings and differences by age for 316,099 white men. Multiple Risk Factor Intervention Trial Research Group. *Archives of Internal Medicine*. 1992;152(1):56-64.
12. Freund KM, Belanger AJ, D'Agostino RB, Kannel WB. The health risks of smoking. The Framingham Study: 34 years of follow-up. *Annals of Epidemiology*. 1993;3(4):417-424.
13. Gartside PS, Wang P, Glueck CJ. Prospective assessment of coronary heart disease risk factors: the NHANES I epidemiologic follow-up study (NHEFS) 16-year follow-up. *Journal of the American College of Nutrition*. 1998;17(3):263-269.
14. Wannamethee SG, Shaper AG, Walker M, Ebrahim S. Lifestyle and 15-year survival free of heart attack, stroke, and diabetes in middle-aged British men. *Archives of Internal Medicine*. 1998;158(22):2433-2440.
15. Jacobs DR, Jr., Adachi H, Mulder I, et al. Cigarette smoking and mortality risk: twenty-five-year follow-up of the Seven Countries Study. *Archives of Internal Medicine*. 1999;159(7):733-740.
16. Vlietstra RE, Kronmal RA, Oberman A, Frye RL, Killip T, 3rd. Effect of cigarette smoking on survival of patients with angiographically documented coronary artery disease. Report from the CASS registry. *Jama*. 1986;255(8):1023-1027.
17. Hermanson B, Omenn GS, Kronmal RA, Gersh BJ. Beneficial six-year outcome of smoking cessation in older men and women with coronary artery disease. Results from the CASS registry.[see comment]. *New England Journal of Medicine*. 1988;319(21):1365-1369.
18. Services USDoHaH. The Health Benefits of Smoking Cessation. A Report of the Surgeon General. In: U.S. Department of Health and Human Services PHS, Centers for Disease Control, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health ed; 1990.
19. Lightwood JM, Glantz SA. Short-term economic and health benefits of smoking cessation: myocardial infarction and stroke. *Circulation*. 1997;96(4):1089-1096.
20. Critchley JA, Capewell S. Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: a systematic review. *Jama*. 2003;290(1):86-97.
21. Morita H, Ikeda H, Haramaki N, Eguchi H, Imaizumi T. Only two-week smoking cessation improves platelet aggregability and intraplatelet redox imbalance of long-term smokers. *Journal of the American College of Cardiology*. 2005;45(4):589-594.
22. Meade TW. Haemostatic function, arterial disease and the prevention of arterial thrombosis. *Baillieres Clinical Haematology*. 1994;7(3):733-755.
23. Elwood PC, Beswick AD, Sharp DS, Yarnell JW, Rogers S, Renaud S. Whole blood impedance platelet aggregometry and ischemic heart disease. The Caerphilly Collaborative Heart Disease Study. *Arteriosclerosis*. 1990;10(6):1032-1036.
24. Elwood PC, Renaud S, Sharp DS, Beswick AD, O'Brien JR, Yarnell JW. Ischemic heart disease and platelet aggregation. The Caerphilly Collaborative Heart Disease Study. *Circulation*. 1991;83:38-44.
25. Trip MD, Cats VM, van Capelle FJ, Vreeken J. Platelet hyperreactivity and prognosis in survivors of myocardial infarction. *New England Journal of Medicine*. 1990;322(22):1549-1554.
26. Thaulow E, Erikssen J, Sandvik L, Stormorken H, Cohn PF. Blood platelet count and function are related to total and cardiovascular death in apparently healthy men. *Circulation*. 1991;84(2):613-617.
27. Leone A. Smoking, haemostatic factors, and cardiovascular risk. *Current Pharmaceutical Design*. 2007;13(16):1661-1667.

28. Renaud S, Blache D, Dumont E, Thedenon C, Wissendanger T. Platelet function after cigarette smoking in relation to nicotine and carbon monoxide. *Clin Pharmacol Ther*. 1984;36:389-395.
29. Tsaprouni L, Grundberg E, Kanoni S, et al. DNA methylation patterns are altered in genes associated with smoking status. Epigenomics of Common Diseases. Wellcome Trust Sanger Institute, Cambridge.; 2011.
30. Leger AJ, Covic L, Kuliopulos A. Protease-activated receptors in cardiovascular diseases. *Circulation*. 2006;114(10):1070-1077.
31. Lawlor DA, Bedford C, Taylor M, Ebrahim S. Geographical variation in cardiovascular disease, risk factors, and their control in older women: British Women's Heart and Health Study. *Journal of Epidemiology & Community Health*. 2003;57(2):134-140.
32. Mendall MA, Strachan DP, Butland BK, et al. C-reactive protein: relation to total mortality, cardiovascular mortality and cardiovascular risk factors in men.[see comment]. *European Heart Journal*. 2000;21(19):1584-1590.

Supervisors/Collaborators: Dr Nic Timpson, Prof Alastair Poole

32 Title: Investigating cellular, transcriptomic and physiological consequences of a functional adenylate cyclase 3 variant: reverse translation of epidemiological findings.

Background:

This project involves an innovative collaboration between epidemiologists, cellular biologists and neuroscientists exploring the functional consequences of an adenylate cyclase 3 (ADCY3) coding variant identified in epidemiological studies by MRC CAiTE. Despite strong association of the ADCY3 variant with fat mass in two cohorts of children [1], the cellular mechanisms underpinning the pathophysiological consequences are unclear, preventing assessment of potential therapeutic approaches. A key aspect to this proposal is that it aims to deconstruct existing findings, to understand biological pathways, but also to link these to the formation of complex, clinically relevant, human behaviour.

BMI is a widely acceptable measure of weight for given height. However, it can be complicated by differential performance in certain body types and age groups. We have used genomewide association studies to investigate the contribution of common variants to a measure of weight which precisely accounts for the covariation with height. A missense variant, rs11676272, in *ADCY3* has been found when height is taken into account and this result was replicated in an independent sample.

ADCY3 codes for the membrane-associated enzyme adenyl cyclase 3, which catalyses the synthesis of cAMP from ATP [2]. The ACDY3 protein sequence comprises two clusters of membrane spanning helices, called M_1 and M_2, which interact to bring together a large cytosolic intracellular loop with a region at the C-terminus to form a composite and competent catalytic domain [3]. The serine-to-proline substitution coded by rs11676272 lies within the second transmembrane spanning alpha-helix of the M_1 cluster. In addition to the observed association with reduced expression of *ADCY3*, we suggest that the proline substitution could disrupt the interaction between helix bundle M_1 and M_2 leading to a reduction in adenyl cyclase activity.

In situ hybridization data from the Allen Mouse Brain Atlas [4], which integrates extensive gene expression and neuroanatomical data, show *ADCY3* mRNA expression in the mouse brain within several nuclei of the hypothalamus, including the paraventricular, ventromedial and arcuate nucleus, regions that are involved in central regulation of energy homeostasis. *ADCY3* knockout mice exhibit age-dependent obesity, which was attributed to hyperphagia, low locomotor activity and leptin insensitivity and demonstrated to be most likely due to hypothalamic cAMP reductions [5].

Given the reduction of *ADCY3* expression associated with rs11676272, in addition to a reduction in ADCY3 activity, a concomitant reduction in cAMP and downstream CREB (cAMP response element binding protein) signalling is likely.

Due to the location of the mutation, alteration in ADCY3 activity is strongly predicted, with potential consequences on cAMP levels and thereby CREB-mediated transcriptome regulation. ADCY3's CNS expression pattern further suggests physiological consequences of the variant on central regulation of metabolic balance and potentially circadian mechanisms – an unexplored area in subjects with the variant.

The project will therefore explore cellular/molecular (Tavare), transcriptomic (Balthasar) and physiological consequences (Balthasar – mice / Timpson – human cohorts) of the variant.

Translational focus: Elucidation of the underlying cellular consequences of coding variants identified via GWAS, genome sequencing and Mendelian-randomisation approaches is critical in validating epidemiological observations and increasing confidence in causality, and thus for identifying tractable new targets for forward translation into therapeutic development.

PhD training: The student will receive an innovative multidisciplinary training in epidemiology, cell biology and in vivo metabolic analysis in genetically modified mice and will be well placed to use their skills

to rapidly exploit findings emerging from genome sequencing projects investigating links between genetic variation and disease.

References:

[1] Timpson et al, in preparation

[2]Ludwig MG, Seuwen K. Characterization of the human adenylyl cyclase gene family: cDNA, gene structure, and tissue distribution of the nine isoforms. Journal of receptor and signal transduction research. 2002 Feb-Nov;22(1-4):79-110. PubMed PMID: 12503609. Epub 2002/12/31. eng.

[3] Gu C, Sorkin A, Cooper DM. Persistent interactions between the two transmembrane clusters dictate the targeting and functional assembly of adenylyl cyclase. Current biology : CB. 2001 Feb 6;11(3):185-90. PubMed PMID: 11231154. Epub 2001/03/07. eng.

[4]Allen Institute for Brain Science. Allen Mouse Brain Atlas Seattle (WA)2012 [[29/04/2013]. Available from: http://mouse.brain-map.org/.

[5] Wang Z, Li V, Chan GC, Phan T, Nudelman AS, Xia Z, et al. Adult type 3 adenylyl cyclase-deficient mice are obese. PloS one. 2009;4(9):e6979. PubMed PMID: 19750222. Pubmed Central PMCID: 2735775. Epub 2009/09/15. eng.

[6] RG. Lerner, C. Depatie, GA. Rutter, RA. Screaton, N.Balthasar. Hypothalamic CRTC2 Links Glucose Sensing with cAMP-reponse-element-mediated Gene Transcription. (2009) *EMBO Reports.* Oct;10(10):1175-81.

[7] J. Rossi*, N. Balthasar*, D. Olson, M. Scott, E. Berglund, CE. Lee, MJ. Choi, D. Lauzon, BB. Lowell, JK. Elmquist. (*joint first authors) Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. (2011) *Cell Metabolism*. Feb 2;13(2):195-204.

Extra reading:

Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, et al. (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 316: 889-894.

Smith GD, Lawlor DA, Harbord R, Timpson N, Day I, et al. (2007) Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. PLoS Medicine 4: e352.

Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature Genetics 42: 937-948.

Timpson NJ, Forouhi NG, Brion M-J, Harbord RM, Cook DG, et al. (2010) Genetic variation at the SLC23A1 locus is associated with circulating concentrations of L-ascorbic acid (vitamin C): evidence from 5 independent studies with >15,000 participants. American Journal of Clinical Nutrition 92: 375-382

Seebohm, G., Strutz-Seebohm, N., Ureche, O.N., Henrion, U., Baltaev, R., Mack, A.F., Korniychuk, G., Steinke, K., Tapken, D., Pfeufer, A., Kääb, S., Bucci, C., Attali, B., Merot, J., Tavaré, J.M., Hoppe, U.C., Sanguinetti, M.C. and Lang, F. (2008) Circ Res. 103, 1451-1457. Long QT Syndrome-Associated Mutations in KCNQ1 and KCNE1 Subunits Disrupt Normal Endosomal Recycling of IKs Channels.

Welsh, G.I., Hale, L., Eremina, V., Jeansson, M., Bek, M., Lennon, R., Pons, D., Owen, R., Satchell, S., Miles, M., Caunt, C., McArdle, C., Herzenberg, A., Toth, T., Pavenstädt, H., Tavaré, J.M., Kahn, C.R., Mathieson, P., Quaggin, S.E., Saleem, M., and Coward, R., (2010) Cell Metabolism. 12:329-40. Insulin Sensitivity of the Kidney Podocyte is Critical for Glomerular Function

Vincent, E.E., Elder, D.J.E., Thomas, E.C., Phillips, L., May, M., Morgan, C., Pawade, J., Sohail, M., Hetzel, M.R. & Tavaré, J.M. (2011) Br. J. Cancer. 104:1755-61. Evaluation of Akt phosphorylation on Thr308 and Ser473 as biomarkers of Akt protein kinase activity in human cancer: correlation of the phosphorylation of Thr308, but not Ser473, with Akt activity in vivo

Seebohm, G., Strutz-Seebohm, N., Nicoleta Ureche, O., Preisig-Müller, R., Zuzarte, M., Hill, EV., Bendahhou, S., Fauler, M., Tapken, D., Decher, N., Collins, A., Jurkat-Rott, K., Steinmeyer, K., Lehmann-Horn, F., Daut, J., Tavaré, J.M., Bloch, W., and Lang, F (2011) FASEB J. 26:513-22. Altered stress stimulation of inward rectifier potassium channels in Andersen-Tawil syndrome.

JW. Sohn, LE. Harris, ED. Berglund, T. Liu, L. Vong, BB. Lowell, **N. Balthasar**[+]***,** KW. Williams[+]*, JK. Elmquist* (*joint senior authors, [+]corresponding authors). Melanocortin 4 Receptors Reciprocally Regulate Sympathetic and Parasympathetic -Preganglionic Neurons (2013) *Cell*. Jan 31;152(3):612-9.

HJM. Weir, TK. Murray, PG. Kehoe, S. Love, EM. Verdin, MJ. O'Neill, JD. Lane, **N. Balthasar**. CNS SIRT3 expression is altered by reactive oxygen species and in Alzheimer's disease. (2012) *PLOSOne*;7(11):e48225.

J. Rossi*, **N. Balthasar***, D. Olson, M. Scott, E. Berglund, CE. Lee, MJ. Choi, D. Lauzon, BB. Lowell, JK. Elmquist. (*joint first authors) Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. (2011) *Cell Metabolism*. Feb 2;13(2):195-204.

RG. Lerner, C. Depatie, GA. Rutter, RA. Screaton, **N.Balthasar**. Hypothalamic CRTC2 Links Glucose Sensing with cAMP-reponse-element-mediated Gene Transcription. (2009) *EMBO Reports.* Oct;10(10):1175-81.

Supervisors/Collaborators: Dr Nic Timpson, Prof Jeremy Tavare

33 Title: An epigenome-wide association study of atopic dermatitis

Background:

Atopic dermatitis (AD), or eczema, is a common chronic skin disease with prevalence rates in developed countries of 15-30% in children and 2-10% in adults (1). Although common genetic variants explain some of the risk, other factors are thought to be important. Epigenetic processes are one possible contributory mechanism.

Epigenetic alterations (such as DNA methylation) are an important mechanism in the regulation of gene expression. Epigenome-wide association studies (EWAS), conducted along the same principles as genome-wide association studies, have the potential to identify new genes and biological pathways implicated in disease pathogenesis (2). This project will involve investigating whether methylation differences are associated with the development of AD using data from the Avon Longitudinal Study of Parents & Children (ALSPAC). A comparison of genome-wide DNA methylation in AD cases and controls will be undertaken in children age 7 (post diagnosis) as well as in DNA samples collected from the same children at birth (pre-diagnosis, to help overcome the issues of confounding and reverse causation).

It is widely recognised that epigenetic patterns are tissue specific and signals observed in surrogate tissues (such as peripheral blood, as in the case of ALSPAC) may not highlight the most important differences in methylation associated with the target disease tissue of interest. This project will therefore include analysis of methylation profiles in skin biopsy samples collected from AD cases and controls. EWAS will be undertaken to identify differentially methylated regions of interest associated with AD and results from both peripheral blood and skin derived DNA will be compared.

Objectives:

1. Carry out an epigenome-wide association study for atopic dermatitis, to identify differentially methylated loci.

2. Investigate how methylation of associated sites varies over time.

3. Explore methylation profiles in skin biopsies of AD cases and controls.

4. Validate any differentially methylated sites using pyrosequencing, involving laboratory skills training (if appropriate).

5. Seek replication in an independent sample, informed by the output of the EWAS.

Methods:

ALSPAC is a prospective population-based birth cohort that recruited 14,541 pregnant women resident in Avon, UK with expected deliveries between 1991 and 1992 (http://www.alspac.bris.ac.uk.). As part of the ARIES (Accessible Resource for Integrated Epigenomic Studies) project, 1000 mother-child pairs have genome-wide methylation data (from the HumanMethylation450 (HM450K) BeadChip) at several time points (for the children these are: at birth, age 7 and age 15-17). Analysis could include stratification by sub-types of AD to understand whether regulation of different genes and biological pathways are involved in the evolution of the various forms of AD observed in childhood.

References:

1. Bieber T. NEJM 358:1483-94(2008); 2. Michels K et al. Nature Methods 10:949-955(2013).

Supervisors/Collaborators:

Lavinia Paternoster, Caroline Relton, Sara Brown (collaborator, University of Dundee)

34 Title: Mathematical modeling of the impact of HCV vaccination among people who inject drugs.

Background:

Hepatitis C virus (HCV) is a blood-borne disease that infects 150 million people worldwide and is a leading cause of liver disease. In developed countries, the main mode of transmission is amongst people who inject drugs (PWID) through syringe sharing.

HCV vaccines are currently in development and have shown promise for reducing infectivity or averting chronic infection[1,2]. One major vaccine candidate targeting genotype 1B has been shown to be well tolerated and highly immunogenic to this strain and less so to other genotypes[3]. Further studies are ongoing. Curative treatment for HCV also exists, and although new drugs have high efficacy (>90%) they are expensive and uncertainty exists on whether HCV vaccination will be effective following successful treatment, due to immune exhaustion.

We propose to develop dynamic models of HCV transmission among PWID[4-6] to evaluate the impact of HCV vaccination strategies. Different intervention setting (prison/community) will be considered as will the combined impact with other interventions. The novelty of the work will be in the model incorporating multiple genotype strains and the possible impact of HCV vaccination in two ways- protection against chronic infection and/or reducing infectivity.

Plan:

This project will consist of developing dynamic models of HCV transmission among PWID including vaccination and other interventions. The models will be used to evaluate:

1. The reduction in HCV transmission resulting from different vaccine coverage levels for varying vaccine characteristics (efficacy for different genotypes, duration of protection, and reduction in infectiousness)
2. The required vaccine coverage level needed for elimination in different HCV epidemic settings and for varying vaccine characteristics?
3. How HCV vaccination strategies should be combined with other interventions to increase impact and reduce the vaccine coverage needed for elimination?
4. The utility of different settings for undertaking HCV vaccination strategies amongst PWID, such as prison or addiction centres?
5. The impact of strategies where vaccination is provided post successful treatment? The effect of reduced vaccination efficacy due to immune exhaustion will be considered, as will different delays in vaccination to allow for immune recovery?
6. The implications of a vaccine only protecting for one strain. Is the vaccine's impact reduced due to the unprotected genotypes partially compensating, and how may the genotype distribution change over time?

Methods

This studentship will utilise mathematical modelling to understand HCV transmission, vaccination and prevention interventions. It will include training in how to set up and undertake research projects incorporating dynamic infectious disease transmission modelling. The candidate will gain experience in methods of infectious disease model development, parameterisation and fitting using Bayesian methods, and sensitivity/uncertainty analysis.

References:

1. Beamont, E. and R. P, *Prospects for prophylactic hepatitis C vaccines based on virus-like particles.* Hum Vaccin Immunother, 2013. **13**(9): p. 5.

2. Maher, L., et al., *Candidate hepatitis C vaccine trials and people who inject drugs: Challenges and opportunities.* Vaccine, 2010. **28**(45): p. 7273-7278.

3. Barnes, E., et al., *Novel adenovirus-based vaccines induce broad and sustained T cell responses to HCV in man.* Sci Transl Med, 2012. **4**(115): p. 115ra1.

4. Vickerman, P., et al., *Can needle and syringe programmes and opiate substitution therapy achieve substantial reductions in HCV prevalence? Model projections for different epidemic settings.* Addiction, 2012. **107**: p. 1984-1995.

5. Martin, N., et al., *HCV treatment for prevention among people who inject drugs: modeling treatment scale-up in the age of direct-acting antivirals.* Hepatology, 2013. **58**(5): p. 1598-1609.

6. Martin, N., et al., *Combination interventions to prevent HCV transmission among people who inject drugs: modelling the impact of antiviral treatment, needle and syringe programmes, and opiate substitution therapy.* Clin Infec Dis, 2013. **57**(suppl 2): p. S39-S45.

Supervisors/Collaborators:
Prof Peter Vickerman, Prof Matthew Hickman and Dr Natasha Martin

35 Title: Mathematical modeling of an emerging epidemic: sexual hepatitis C transmission among HIV positive men who have sex with men

Background

Hepatitis C virus (HCV) is a major cause of liver disease worldwide. In developed countries, HCV is predominantly transmitted through injecting drug use[1], and studies have shown that sexual HCV transmission between heterosexual couples is rare.

In recent years there has been a rapid spread of sexually transmitted HCV amongst HIV-positive men who have sex with men (MSM)[2,3], but little amongst HIV-negative MSM[4,5]. It is unknown why this emerging epidemic has occurred and why it is concentrated amongst HIV-infected MSM. This recent epidemic, along with a rise in HIV infections among MSM in the UK has resulted in increasing concern among the public health/medical communities[6].

Despite substantial literature on modeling HIV transmission among MSM[7] and HCV transmission amongst injecting drug users, no one has modeled sexual HCV transmission among MSM, and little is known about the transmission dynamics. This project would aim to develop both deterministic and network models of HIV and HCV transmission among MSM to explore how behavioural characteristics (such as risk heterogeneity, possibly based on HIV status, and assortative sexual mixing, where HIV positive men tend to have riskier sex with other HIV positive men) and biological characteristics (such as increased HCV viral load amongst HIV-HCV coinfected individuals and recent increases in survival of HIV infected men) drive the transmission of each disease and determine its patterns of spread.

The insights gained will increase our understanding of MSM risk behaviour and mixing, which will improve model projections of the coverage requirements for different prevention to reduce the transmission of both infections to low levels. The network models will be used to show how network characteristics determine the speed of the epidemic, and to identify optimal intervention allocation strategies among the network participants. The outputs of this project will shed light on the drivers of the emerging HCV epidemic among MSM, and shape policymaking surrounding intervention strategies.

Objectives

1. Develop novel mathematical models (network and deterministic) of HIV and HCV transmission among MSM and fit to available data
2. Interrogate the drivers of HIV and HCV transmission among MSM, and determine which behavioural or biological characteristics contribute towards the patterns of disease transmission
3. Identify the most effective combination prevention strategies to reduce transmission of both infections to low levels

Methods

This studentship will utilise various mathematical modelling techniques to understand HIV and HCV transmission and prevention interventions. It will include training in dynamic infectious disease transmission modelling including network and deterministic compartmental models. The candidate will also gain experience in methods of infectious disease model parameterisation and fitting using Bayesian methods, and sensitivity/uncertainty analysis.

References:

1. Shepard CW, Finelli L, Alter MJ. Global epidemiology of hepatitis C virus infection. *Lancet Infec Dis* 2005; 5(9): 558-67.

2. Urbanus AT, van de Laar TJ, Stolte IG, et al. Hepatitis C virus infections among HIV-infected men who have sex with men: an expanding epidemic. *AIDS* 2009; 23(12): F1-7.

3. van de Laar T, Pybus O, Bruisten S, et al. Evidence of a large, international network of HCV transmission in HIV-positive men who have sex with men. *Gastroenterology* 2009; 136(5): 1609-17.

4. Price H, Gilson R, Mercey D, et al. Hepatitis C in men who have sex with men in London--a community survey. *HIV Med* 2013; 14(9): 578-80.

5. Yaphe S, Bozinoff N, Kyle R, Shivkumar S, Pai NP, Klein M. Incidence of acute hepatitis C virus infection among men who have sex with men with and without HIV infection: a systematic review. *Sex Transm Infect* 2012; 88(7): 558-64.

6. Kirby T, Thornber-Dunwell M. New HIV diagnoses in London's gay men continue to soar. *The Lancet* 2013; 382(9889): 295.

7. Punyacharoensin N, Edmunds WJ, De Angelis D, White RG. Mathematical models for the study of HIV spread and control amongst men who have sex with men. *Eur J Epidemiol* 2011; 26(9): 695-709.

Supervisors/Collaborators:

Prof Peter Vickerman, Prof Matthew Hickman and Dr Natasha Martin

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

Background:

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. Similarly the effects of low levels of nutrient intake, for example folate and vitamin D, during pregnancy on infant 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. In addition understanding the influence of prenatal risk factors on childhood behaviour and psychiatric disorders requires some knowledge of the biological pathways mediating between these exposures and outcomes. One possibility is that gene expression is altered via modifications to the foetal epigenome, a phenomenon that has become the focus of considerable interest in recent times. Mendelian randomization can be used to interrogate the role of epigenetic processes in causal pathways to psychiatric (and many other) disorders.

Plan:

The plan is to assess associations between genetic variants related to diet among mothers and relate these to offspring development outcomes (including psychiatric and psychological outcomes), and also to determine whether epigenetics is likely to explain any observed associations.

Objectives:

1) Identify genetic polymorphisms, which metabolise nutrients, or affect exposure propensities and which may be related to development in utero.
2) Analyse associations between the above polymorphisms and prenatal development to determine which nutrients are important in a large cohort study.
4) Analyse associations between the above polymorphisms and markers of gene expression
5) Analyse associations between gene expression and outcomes of interest

Supervisors/Collaborators:

Sarah Lewis and Caroline Relton

37 Title: Systematic analysis of the genetic contribution to hemopoietic transplant failure aimed at improving future donor selection.

Background:

Despite major recent advances in tissue typing, blood or marrow stem cell transplantation from sibling or unrelated donors is associated with many potential complications as well as relapse of the disease requiring transplantation. In combination these cause post-transplant mortality of up to 40%. These high failure rates have led scientists to look for genetic factors which impact on transplant outcome apart from the HLA tissue typing genes themselves.

One area of investigation has been "microsatellites", highly variable repeating sequences of DNA found adjacent to genes. Combinations of microsatellites give a unique identity to individuals, as exploited in paternity testing kits. Logically, it seems likely that patients and donors with similar overall patterns of microsatellites are likely to be more closely related ancestrally and hence have more genes of identical sequence. This in turn may minimise immune reactions provoked by donors recognising patient proteins as foreign or vice versa. This led us to develop a scoring system for microsatellite data which showed a statistically significant relationship with death from all causes following transplant. We now need to know whether our score is a genuine measure of similarity across people's genomes or is highlighting particular gene(s) (closely associated with the microsatellites) which influence post-transplant death. We therefore propose to study 30 donor/patient by a technique which analyses the whole genome, comparing those with highest and lowest scores who survive or died post-transplant. This will provide pilot data for a larger study aimed at improved donor selection either by analysing microsatellite patterns or by developing a “gene chip” testing kit.

Microsatellites (Msat) are already of major interest in HSCT as differences between patterns of Msat in patients and donors can be exploited to determine percentage donor cell engraftment at any stage post-transplant ("chimerism analysis"). Dr John Harvey, Principal Clinical Scientist at NHS Blood and Transplant (NHSBT), Bristol has studied these Msats in detail and developed a Weighted Average Allelic Size Difference (WAASD) score to grade donor/recipient Msat mismatch. His composite score takes account of both the numbers of differences between the alleles of 15 separate sets of microsatellites (selected for their high degree of informativity between individuals) and the discrete differences in numbers of repeat units at those microsatellites.

In a retrospective study of 180 sibling and unrelated Bristol donor/recipient pairs this score showed strong correlation with post-transplant mortality, with increasing strength with greater discrepancy. Patient/donor pairs with a WAASD below or above 1.8 have an overall mortality of 27.9% vs 44.4% respectively ($P = 0.045$). If 1.9 is used as the cut-off, mortality rates were 28% vs 56.2% respectively ($P = 0.014$). Both disease relapse and other transplant-related causes of mortality (e.g. viral infection, multi-organ failure) contributed equally.

Despite these probabilistic inferences about global genomic difference, there has been little evidence as to what aspects of genetically determined variation are responsible for variable transplant outcome. Malkki et al [1] studied microsatellites within the HLA complex and found evidence that mismatch between donor and recipient was associated with the incidence of GVHD or mortality in a cohort of 819 HLA-matched HSCT pairs. However, more particular attention has been paid to Msats which flank cytokine genes, especially the interferon-gamma locus on chromosome 12, *IL1RN* and *IL1A*, *TNFd*, *IL10* and *HMOX1*. None has shown a consistent relationship with post-transplant criteria such as the incidence of acute and/or chronic GVHD, TRM or overall survival [2].

The HSCT team at Bristol Royal Hospital for Children (BRHC) and scientists within NHSBT have a long-standing expertise in the use of Msat for chimerism analysis, through collaborations with the Forensic Science Service and EU partners on a Fifth Framework project [3,4]. Together with collaborators at the MRC Integrative Epidemiology Unit (IEU at the School of Social and Community Medicine, University of Bristol), who have considerable experience with the handling and analysis of genomewide data of the type to be collected in this investigation, we have a superb opportunity to extend the on-going analysis of genomic contributions to the outcome of haemopoetic transplants. Furthermore, the findings made by this investigation lead naturally onto a series of larger experiments to investigation the development of a deliverable product for health services.

Within this already recognized group of patients available to us though the HSCT team at BRHC and clinical scientists at NHSBT we wish to further explore the impact of genomic differences on transplant outcome. Our core objective is to understand how the profile of genetic differences between these donor/recipient pairs varies according to the transplant outcome.

Core objectives:

It may be that the nature of this donor/recipient difference is entirely reflected in the probabilistic inference obtained from the use of microsatellites (i.e. as a global level of difference between donor and recipient). In this case, our key aim will be to refine inferences made using Msat data and to provide a platform for single nucleotide polymorphism (SNP) assessment of global differences.

However, it is equally likely that there are specific genomic differences which are common to differential transplant outcome. This data collected in this study will allow us to identify specific differences and their relationship with clinical outcome.

Secondary objectives:

The information obtained in this experiment would act as a pilot to a follow-up investigation which would be engineered around a larger collection through extension into adult transplant outcomes and other collaborating centres via a bid to the NIHR. This would explicitly consider the development of a "designer array" (i.e. a cheap and clinically actionable tool) which will reliably capture important signals informative for the outcome of transplant.

The undertaking of the proposed research will be a joint venture between the two groups represented by the co-PIs of this application (i.e. BRHC and the MRC IEU) as follows:

(a) Selection of 30 donor/recipient pairs based on WAASD score. These 30 pairs will be selected on the following criteria in efforts to maximize potential information (work already completed).

(b) Collection of genomewide SNP based array data on all participants of the 30 pairs (Illumina 2.5-8 array). This is a whole genome SNP array which will collected up to 2.5 million SNPs with maximal coverage of 1000genomes variation (www.genomes.org) down to a minor allele frequency of ~1%. This will be undertaken at the MRC IEU by co-PI, Dr Nicholas Timpson who will also be responsible for cleaning, processing and analysis of genomewide data (work already completed).

(c) Systematic analysis of differences between donors and recipients and the relationship of this status of transplant outcome.

Disciplines and training:

Genomics
Statistical analysis methods
Relevant training as part of the Wellcome Trust Advanced Courses
Short courses on basic and advanced analysis approaches at the School of Social and Community Medicine.

References:

[1] Malkki M, Gooley T, Horowitz M, Petersdorf EW; IHWG HCT Component. MHC class I, II, and III microsatellite marker matching and survival in unrelated donor hematopoietic cell transplantation. Tissue Antigens. 2007 Apr;69 Suppl 1:46-9.

[2] Tiercy JM. Immunogenetics of hematopoietic stem cell transplantation: the contribution of microsatellite polymorphism studies. Int J Immunogenet. 2011 Oct;38(5):365-72.

[3] Hancock JP, Burgess MF, Goulden NJ, Steward CG, Knechtli CJ, Pamphilon DH, Potter MN, Oakhill A. Same-day determination of chimaeric status in the immediate period following allogeneic bone marrow transplantation.Br J Haematol. 1997 Nov;99(2):403-9

[4] Watzinger F, Lion T, Steward CG; Eurochimerism consortium. The RSD code: proposal for a nomenclature of allelic configurations in STR-PCR-based chimerism testing after allogeneic stem cell transplantation. Leukemia. 2006 Aug;20(8):1448-521;83(5):1355-62.

Supervisors/Collaborators
Dr Nic Timpson, Dr Colin Steward, Professor Davey Smith, Dr John Harvey

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

Background:

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

Assessment of scoliosis

Currently, gold standard assessment of scoliosis is achieved through x-ray examination, however this is not ideal for multiple examination and is not always available in large numbers of individuals. Scoliometers can be employed to assess the degree of spinal curvature though a non-invasive, nurse administered, physical examination (the Adams forward bending test). Unfortunately, these measurements are difficult to repeat in a standardised manner and often yield unreliable assessment of scoliosis status across large numbers of participants. Novel automated methods of measuring the Cobb angle for scoliosis diagnosis have been developed in adults that use spinal dual energy X-ray absorptiometry (DXA) taken with the patient lying flat, which has a much reduced radiation exposure compared with spinal radiographs, and correlates well with traditional methods of measuring the Cobb angle ($R2=0.998$). DXA measures also provide an indication of size, site and direction of curve, as well as measures of local bone quality and density without the large radiation exposure of traditional spinal radiograph. In light of this, members of the musculoskeletal centre at Southmead Hospital, Bristol have been developing more accurate measures of spinal deformity and applying these to the Avon Longitudinal Study of Parents and Children (http://www.bristol.ac.uk/alspac/).

Potential causes of scoliosis

Current knowledge about the causes of the initiation or induction of the scoliotic curve is scarce, particularly outside of high income countries. Few epidemiological studies have identified clear predictors of risk of scoliosis, although girls appear to be more affected than boys. One potential

determinant of curve induction that is of great interest at present is a genetic cause, as twin studies indicate that scoliosis may have a genetic component and many potential genetic areas of interest have been identified. However, it is likely, that as for the majority of human disease, scoliosis is due to a complex interplay between environmental and genetic influences. Bone mass, hyperflexibility and neuromotor or growth abnormalities may also contribute towards the development of scoliosis. In addition to improving our knowledge of the epidemiology of scoliosis, the discovery of novel risk factors may lead on to the development of new management techniques. The discovery of accurate prognostic markers may help decisions about timing of surgery or other interventions.

Hyderabad Nutrition Trial

The prevalence and causes of scoliosis can be assessed within an Indian population using the new methods in the assessment of scoliosis. In collaboration with the London School of Hygiene & Tropical Medicine, we at the MRC CAiTE Centre (University of Bristol) are undertaking a large study in India based on participants of a trial of nutritional supplementation called the Hyderabad Nutrition Trial. In this study, over 2000 women

were randomised to receive supplementation during pregnancy and the first five years of their child's life, or to no supplementation. These children were examined at the age of 14-16, and they are now aged 18-21. We are currently undertaking the follow-up of this cohort and all participants are being undergoing whole body DXA scans, as well as being examined for multiple potential correlates (objective and subjective measures of physical activity, work status, quality of life and depression, spirometry, anthropometry, growth, including puberty, bone and muscle parameters, DNA). We anticipate that we will examine 1300 participants from this cohort by the end of 2010, both from supplement and non- supplement arms of this Indian study. The availability of DXA data presents a superb opportunity to assess the presence of scoliosis in the participants and to investigate the influence of environmental and genetic factors on Scoliosis prevalence. It will also be possible to retrace these subjects for further examination.

Objectives:

1. To use a novel method to assess the presence of absence of scoliosis from whole body DXA scans for the participants of the Hyderabad Nutrition Trial.

2. To assess the prevalence of scoliosis within this study.

3. To assess genetic and environmental predictors of scoliosis, and their interaction, in this study.

Plan of investigation:

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

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

Supervisors:

Nicholas Timpson; Emma Clark; Hannah Kuper

39 Title: Systematic screen of *omic* resources for variation with autism symptomatology

Background:

Having friends, laughing about the same jokes and sharing feelings with others are human skills, which most of us take for granted in life. While we all show some variation in how easily we socialise with other people, some individuals affected by Autism Spectrum Disorders (ASDs)[1] are so severely afflicted by behavioural difficulties, that this makes it difficult for them to join their peers in normal social development.

It has been thought that there is a continuum between milder autism-like symptoms in individuals of the general population and the severe expression of the disease in ASD patients. Like ASD, autism-like symptoms are highly heritable[2–4] and highly persistent[5, 6] throughout the course of development. Furthermore, there is no evidence for differences in heritability between the diseased and the non-diseased end of the autistic continuum[7, 8] suggesting that clinical ASD and autism-like symptoms may have a common aetiology.

Little is currently known however about the molecular nature of overarching molecular links across the autistic spectrum, although such knowledge might be crucial for the design and development of new therapeutic drug targets in autism.

Methods:

This PhD at the MRC IEU will involve a systematic screen of variation in *omic* data in relation to the expression of autism-like symptoms during childhood and adolescence investigating children from the Avon Longitudinal Study of Parents and Children (ALSPAC). These symptoms will include for example impairments in social communication and interaction. The search will focus in the first instance on known autism susceptibility loci and might later be extended to the entire *omic* resource. ALSPAC is a large-scale pregnancy cohort with extremely well characterised longitudinal phenotypic data in addition to detailed biological information. Specifically, the work will involve the analysis of (i) whole genome sequence data on 1800 children (ii) genomewide common variant data on ~8000 children and (iii) intermediate phenotypic data from metabolomic, methylomic and transcriptomic data sources on ~ 1000 children. Evidence from multiple *omic* resources will be combined with multivariate statistical analysis techniques and machine learning algorithms. The project will entail furthermore a detailed genetic profiling of the investigated ASD risk loci with respect to autism symptomatology in ALSPAC including a gene-based analysis using SNP-based heritability[9] and the investigation of pairwise genetic correlations between autism-related behaviour and intermediate phenotypes in a search for loci that contribute to multiple traits.

Integrated PhD training plans :

- Local short courses at the School of Social and Community Medicine, University of Bristol (http://www.bristol.ac.uk/social-community-medicine/shortcourse)
- Relevant training as part of the Wellcome Trust advanced training courses (http://www.wellcome.ac.uk/Education-resources/Courses-and-conferences/Advanced-Courses-and-Scientific-Conferences/Advanced-Courses/index.htm)

References:

1. Wing L. The continuum of autistic characteristics. In Schopler E, Mesibov G (eds.). *Diagnosis and assessment in autism* Plenum: New York, 1988, pp 91–110.

2. Hoekstra RA, Bartels M, Verweij CJH, Boomsma DI. Heritability of autistic traits in the general population. *Arch Pediatr Adolesc Med* 2007; **161**: 372–7.
3. Ronald A, Happé F, Bolton P, Butcher LM, Price TS, Wheelwright S, *et al.* Genetic heterogeneity between the three components of the autism spectrum: a twin study. *J Am Acad Child Adolesc Psychiatry* 2006; **45**: 691–9.
4. Skuse D, Mandy W, Scourfield J. Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *Br J Psychiatry* 2005; **187**: 568–72.
5. Constantino JN, Abbacchi AM, Lavesser PD, Reed H, Givens L, Chiang L, *et al.* Developmental Course of Autistic Social Impairment in Males. *Dev Psychopathol* 2009; **21**: 127–38.
6. St Pourcain B, Mandy WP, Heron J, Golding J, Davey Smith G, Skuse DH. Links Between Co-occurring Social-Communication and Hyperactive-Inattentive Trait Trajectories. *J Am Acad Child Adolesc Psychiatry* 2011; **50**: 892–902.e5.
7. Lundström S, Chang Z, Råstam M, Gillberg C, Larsson H, Anckarsäter H, *et al.* Autism spectrum disorders and autistic like traits: similar etiology in the extreme end and the normal variation. *Arch Gen Psychiatry* 2012; **69**: 46–52.
8. Robinson EB, Koenen KC, McCormick MC, Munir K, Hallett V, Happé F, *et al.* Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Arch Gen Psychiatry* 2011; **68**: 1113–21.
9. Speed D, Hemani G, Johnson MR, Balding DJ. Improved Heritability Estimation from Genome-wide SNPs. *The American Journal of Human Genetics* 2012; **91**: 1011–21.

Supervisors:

Beate St Pourcain, George Davey Smith

40 Title: Utilising genome-wide data as an instrument for studying environmental risk factors in complex behaviour

Background:

There is a long-standing debate about the influences of nature versus nurture and their interplay in human behaviour. The phrase was originally coined by Francis Galton contemplating on the impact of heredity (genetic factors) and environment on social advancement. Within the context of behavioural research, the identification of environmental factors (E) plays a key role, as this determines if and how behaviour can be modified. This is important as a modifiable behaviour implicates, for adverse outcomes, also a potentially preventable behaviour. The search for E in complex behaviour may even facilitate the identification of potential risk factors for neuro-developmental disorders. This is because many theories[1] assume that population variation in human behaviour and disease lie at opposite ends of the same continuum, such as for example hypothesised for subtle impairments in social communication and autism spectrum disorder.

Capturing genuine environmental influences however poses a great challenge due to the possibility that observed associations are due to confounding. That is, the association between an environmental risk factor and complex behaviour is due to an unmeasured third factor (confounder), which is correlated with both variables and may for example include lifestyle and socioeconomic influences. For this reason, methods have been developed that utilise genetic information, which is typically free of confounding, to determine the underlying causal relationships between variables (Mendelian Randomisation[2]). Within the era genome-wide data, this principle has now been extended by measuring "environmental influences" through the summative effect of many markers that index biological intermediate phenotypes[3] providing a highly powerful tool for epidemiological investigations.

Methods:

This PhD will focus on the investigation of environmental factors influencing adverse behavioral outcomes in childhood and adolescence including conduct problems, impairments in social communication and deficits in social interaction. It will specifically involve the design of genetic instruments capturing risk factors for these adverse behavioural outcomes such as for example maternal substance use (e.g smoking) during pregnancy[4, 5].

Instruments will be based on known variants and/or polygenic risk scores derived from publicly available genome-wide data sets. Using these instruments in addition to direct measures of exposure, associations with adverse behaviour during development will be investigated within the Avon Longitudinal Study of Children and Parents (ALSPAC). ALSPAC is a large-scale pregnancy cohort where data is now available on extremely well characterised longitudinal, phenotypes, in addition to detailed biological information. Specifically, allele scores will be constructed based on (i) whole genome sequence data on 1800 individuals and (ii) genome-wide common variant array data on ~8000 children and mothers. In addition, novel screening methods[6] will be utilised for the search of Gene x Environment interactions in complex behavior, capturing the modifications of genetic influences through environmental factors and vice versa.

Integrated PhD training plans :

- Local short courses at the School of Social and Community Medicine, University of Bristol (http://www.bristol.ac.uk/social-community-medicine/shortcourse)
- Relevant training as part of the Wellcome Trust advanced training courses (http://www.wellcome.ac.uk/Education-resources/Courses-and-conferences/Advanced-Courses-and-Scientific-Conferences/Advanced-Courses/index.htm)

References:

1. Constantino JN, Todd RD. Intergenerational transmission of subthreshold autistic traits in the general population. *Biol Psychiatry* 2005; **57**: 655–60.
2. Davey-Smith G, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med* 2007; **4**: e352.
3. Evans DM, Brion MJA, Paternoster L, Kemp JP, McMahon G, Munafo M, *et al.* Mining the Human Phenome Using Allelic Scores that Index Biological Intermediates. *PLoS Genet* in press.
4. Hultman CM, Sparén P, Cnattingius S. Perinatal risk factors for infantile autism. *Epidemiology* 2002; **13**: 417–23.
5. St. Pourcain B, Mandy WP, Heron J, Golding J, Davey Smith G, Skuse DH. Links Between Co-occurring Social-Communication and Hyperactive-Inattentive Trait Trajectories. *J Am Acad Child Adolesc Psychiatry* 2011; **50**: 892–902.e5.
6. Paré G, Cook NR, Ridker PM, Chasman DI. On the Use of Variance per Genotype as a Tool to Identify Quantitative Trait Interaction Effects: A Report from the Women's Genome Health Study. *PLoS Genet* 2010; **6**: e1000981.

Supervisors:

Beate St Pourcain, George Davey Smith

41 Title: Robust evaluation of the relationship between non-syndromic forms of cleft, biological intermediates as well as developmental outcomes in later life

Background:

Cleft lip and/or palate (CL/P) is a common congenital disorder, which affects about a thousand children born in the United Kingdom (UK) each year. Approximately 70% of all children with CL/P carry apart from the facial cleft structure no other apparent cognitive or craniofacial structural abnormality and this form of CL/P is commonly termed 'isolated, non-syndromic CL/P' (nsCL/P). Current research suggests that both genetic and environmental factors contribute to the complex aetiology of nsCL/P.

There is considerable interest in the identification of modifiable environmental factors of CL/P as this research potentially provides new avenues for future prevention and intervention efforts. Recent studies support for example the role of zinc deficiency[1], cholesterol deficiency[2] and folate deficiency[3], as potential risk factor for nsCL/P. However, the reliable identification of environmental influences has been proven to be challenging as nsCL/P-related defects arise early during embryological development and the disorder itself has only modest recurrence rates[4].

In addition, nsCL/P has been associated with impairments in growth, speech, hearing, general health and social integration during later life[4]. It is therefore possible that some of the underlying genetic risk factors for nsCL/P are also causally linked with these deficits in later life, especially as subclinical phenotypes have been observed in relatives of patients with CL/P[4]. This includes for example dental anomalies, 3D facial image measurements but also speech and cognitive differences including reading disability and IQ[4,5], and genetic associations have been indeed observed between nsCLP loci and face shape in the general population[6].

This PhD project offers an exciting possibility to unravel robust and potentially causal links between CL/P and biological intermediate phenotypes as well as growth and behavioural outcomes in later life, using modern epidemiological and genetic epidemiological analysis techniques. Specifically, the project will investigate the following research questions:

I) Are genetic instruments for biological CL/P related intermediate phenotypes (e.g. zinc, cholesterol and folate levels) associated within an increased risk for CL/P?
II) Do healthy individuals with a high load of nsCLP risk alleles differ phenotypically compared to the rest of the population?

Methods:

This PhD at the MRC IEU will involve an in depth analysis of currently available cleft resources and a systematic screen of children from the Avon Longitudinal Study of Parents and Children (ALSPAC). The first part of the project will entail the *post hoc* construction of a CL/P case-control sample using cases from publicly available genotyped and phenotyped nsCL/P samples (via dbgap) and control individuals from the WTCCC2 sample, all of European descent. Exploiting currently available genome-wide association study (GWAS) findings, genetic instruments will then be constructed, which explain a large proportion of the phenotypic variation in biological CL/P related intermediate phenotypes (e.g. zinc, cholesterol and folate levels) and are typically free of confounding factors. Finally, these instruments will be used to screen the constructed CL/P case control sample for association with risk for CL/P.

Within the second part of the PhD, polygenic risk scores based on known nsCL/P loci will be constructed within children of the ALSPAC cohort. ALSPAC is a large-scale pregnancy cohort with well characterised longitudinal phenotypic data in addition to detailed biological information. Specifically, the project will investigate whether healthy individuals with high CL/P polygenic risk scores differ from the rest of the population with respect to CL/P-related dimensional subphenotypes (e.g. 3D facial images, tooth eruption as well as cognitive, anthropometric, language-, reading- and hearing-related outcomes). This will allow investigating the relationship between CL/P and developmental outcomes in later life free of confounding influences, which might be associated with the presence and treatment of the facial cleft structure, and may thus uncover potential causal links.

Integrated PhD training plans :

- Local short courses at the School of Social and Community Medicine, University of Bristol (http://www.bristol.ac.uk/social-community-medicine/shortcourse)
- Relevant training as part of the Wellcome Trust advanced training courses (http://www.wellcome.ac.uk/Education-resources/Courses-and-conferences/Advanced-Courses-and-Scientific-Conferences/Advanced-Courses/index.htm)

References:

1. Munger RG, Tamura T, Johnston KE, Feldkamp ML, Pfister R, Carey JC. Plasma zinc concentrations of mothers and the risk of oral clefts in their children in Utah. *Birth Defects Res Part A Clin Mol Teratol* 2009; **85**: 151–5.
2. Porter FD. Cholesterol precursors and facial clefting. *J Clin Invest* 2006; **116**: 2322–5.
3. Wehby G, Murray JC. Folic Acid and Orofacial Clefts: A Review of the Evidence. *Oral Dis* 2010; **16**: 11–9.
4. Dixon MJ, Marazita ML, Beaty TH, Murray JC. Cleft lip and palate: synthesizing genetic and environmental influences. *Nat Rev Genet* 2011; **12**: 167–78.
5. Tereza GPG, Carrara CF de C, Costa B. Tooth abnormalities of number and position in the permanent dentition of patients with complete bilateral cleft lip and palate. *Cleft Palate Craniofac J* 2010; **47**: 247–52.
6. Liu F, van der Lijn F, Schurmann C, Zhu G, Chakravarty MM, Hysi PG, *et al.* A Genome-Wide Association Study Identifies Five Loci Influencing Facial Morphology in Europeans. *PLoS Genet* 2012; **8**: e1002932.

Supervisors:

Beate St Pourcain, George Davey Smith and Jonathan Sandy