School of Social & Community Medicine

Research student opportunities in Genetics 2014

SUMMARY OF TOPICS

	Supervisor	Title of project
1	Sarah Lewis and Caroline Relton	An investigation of intra-uterine nutrition and prenatal development– applying the principle of Mendelian randomization
2	Santiago Rodriguez & Ian Day	Evaluation of the prevalence and functionality of paucimorphic and private mutations in large epidemiological surveys for cardiovascular risk traits
3	Ian Day & George Davey Smith	Hp genotype as a potential predictor for Hb levels, and investigation of possible correlations with selected phenotypes in mother and child.
4	Santiago Rodriguez & Tom Gaunt	Detection of "non-tagged" CNVs and association analyses with human body weight related traits
5	Ian Day & George Davey Smith	A study of genotype influences on reference ranges for clinical analytes (‘Range Mendelization’)
6	Nic Timpson, Colin Steward, George Davey Smith & John Harvey	Systematic analysis of the genetic contribution to hemopoietic transplant failure aimed at improving future donor selection.
7	Andy Ness, Ian Day	Gene-nutrient interactions in the determination of blood lipid levels and early–stage atherosclerosis in childhood
8	Tom Gaunt and Ian Day	Disentangling instances of causally and pharmacogenetically relevant genomic confounding
9	Santiago Rodriguez and Ian Day	Genetic analyses of individuals expressing extremely low levels of plasma protein biomarkers
10	Beate St Pourcain, George Davey Smith	Systematic screen of *omic* resources for variation with autism symptomatology

11	Nic Timpson, George Davey Smith, Marcus Munafo	The use of genome-wide data for the design and undertaking of 'recall by genotype' experiments
12	Nic Timpson, Oliver Stegle, George Davey Smith	The coordinated analysis of multi-omic data: "Synergomics".
13	Lavinia Paternoster, Caroline Relton, Sara Brown (collaborator, University of Dundee)	An epigenome-wide association study of atopic dermatitis
14	Nic Timpson, Emma Clark & Hannah Kuper	Analysis of environmental and genetic contributions to scoliosis in an Indian population.
15	John Henderson & Jonathan Sterne	Genetic and Environmental precursors of Chronic Obstructive Pulmonary Disease (COPD)
16	Nic Timpson & George Davey Smith	Analysis of common complex traits and genomewide sequence data
17	Kate Tilling, Laura Howe, Abi Fraser, Corrie Macdonald-Wallis, Debbie Lawlor	Diurnal patterns of physical activity, sedentary behaviour, heart rate and blood pressure in adolescents
18	Richard Martin, George Davey-Smith , Caroline Relton	Epigenetic biomarkers of prostate cancer risk
19	Marcus Munafo, Glyn Lewis	Mobile Platform Devices for Measuring Lifestyle-Related Cognitions
20	Caroline Relton, George Davey Smith	DNA Methylation, one carbon metabolism and lung cancer-identifying early biomarkers of disease risk
21	Caroline Relton, Debbie Lawlor, Elliott	Epigenetic variation and type 2 diabetes related traits; investigating causal relationships
22	Caroline Relton, George Davey Smith, Tom Gaunt	Epigenetic variation and childhood body composition; a Mendelian randomization approach to investigating causal relationships

23	John Henderson, Caroline Relton, Kate Tilling	Air pollution exposure in early life and lung function in childhood – a role for epigenetic mechanisms
24	Caroline Relton, George Davey Smith, Debbie Lawlor	Epigenetic variation in children born following assisted reproductive technology
25	George Davey Smith, Laura Howe, Thorkild I Sørensen, Jennifer L Baker	Epigenetic and metabolomic mediators of associations between childhood growth and chronic disease
26	Beate St Pourcain, George Davey Smith	Utilising genome-wide data as an instrument for studying environmental risk factors in complex behaviour
27	Abigail Fraser, Corrie Macdonald-Wallis, Kate Tilling	Maternal haemoglobin in pregnancy: trajectories, risk factors and consequences for offspring. The Avon Longitudinal Study of Parents and Children (ALSPAC)
28	Tom Gaunt, Colin Campbell and Jose Soeane	Approaches to data integration and visualisation in multi-omics and epidemiological data
29	Tom Gaunt	Patterns of correlation in DNA methylation: inferring functional relationships from temporal and spatial epigenetic changes
30	Tom Gaunt, Colin Campbell, Jose Seoane	Aggregating methods for analysis of association between methylation and phenotypes or outcomes
31	Nic Timpson, Matt Jones and Dr Ullrich Bartsch	Linking sleep phenotypes to psychiatric risk genotypes
32	Nic Timpson and Alastair Poole	Methylation of platelet F2RL3 gene provides a major mechanism for smoking-induced coronary thrombosis and cardiovascular disease risk
33	Nic Timpson and Jeremy Tavare	Investigating cellular, transcriptomic and physiological consequences of a functional adenylate cyclase 3 variant: reverse translation of epidemiological findings.
34	John Henderson and Raquel Granell	The Genetics of Asthma Phenotypes

35	Beate St Pourcain, George Davey Smith and Jonathan Sandy	Robust evaluation of the relationship between non-syndromic forms of cleft, biological intermediates as well as developmental outcomes in later life
36	Peter Vickerman, Matthew Hickman and Natasha Martin	Mathematical modeling of the impact of HCV vaccination among people who inject drugs.
37	Peter Vickerman, Matthew Hickman and Natasha Martin	Mathematical modeling of an emerging epidemic: sexual hepatitis C transmission among HIV positive men who have sex with men
38	Luisa Zuccolo, Caroline Relton and Marcus Munafo	Neurodevelopmental consequences of prenatal alcohol exposure: a population based MRI study

1 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:

Sarah Lewis and Caroline Relton

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

Outline of Project:
Background

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

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

Objectives of the PhD

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

Design

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

Techniques and approaches

The scanning of unknown mutations will be performed using a high-throughput mutation

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

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

References

1. Lander,E.S. (1996) The new genomics: global views of biology. Science, 274, 536-539.
2. Pritchard,J.K. (2001) Are rare variants responsible for susceptibility to complex diseases? Am. J Hum Genet, 69, 124-137.
3. Day,I.N.M., Al-Harbi,K., Smith,M., Al Dahmesh,M., Zhang,B., Lotery,A., Cross,N.C.P., Ye,S., Eccles,D., Fox,K., Rodríguez,S. (2004) Paucimorphic alleles versus polymorphic alleles in disease causation: theory, observation and detection. Curr. Genomics, 5, 431-438.
4. Alharbi,K.K., Aldahmesh,M.A., Spanakis,E., Haddad,L., Whittall,R.A., Chen,X.H., Rassoulian,H., Smith,M.J., Sillibourne,J., Ball,N.J., et al. (2005) Mutation scanning by meltMADGE: validations using BRCA1 and LDLR, and demonstration of the potential to identify severe, moderate, silent, rare, and paucimorphic mutations in the general population. Genome Res., 15, 967-977.
5. Alharbi,K.K., Spanakis,E., Tan,K., Smith,M.J., Aldahmesh,M.A., O'Dell,S.D., Sayer,A.A., Lawlor,D.A., Ebrahim,S., Smith,G.D., et al. (2007) Prevalence and functionality of paucimorphic and private MC4R mutations in a large, unselected European British population, scanned by meltMADGE. Hum Mutat., 28, 294-302.
6. Real,J.T., Chaves,F.J., Ejarque,I., Garcia-Garcia,A.B., Valldecabres,C., Ascaso,J.F., Armengod,M.E., Carmena,R. (2003) Influence of LDL receptor gene mutations and the R3500Q mutation of the apoB gene on lipoprotein phenotype of familial hypercholesterolemic patients from a South European population. Eur J Hum Genet, 11, 959-965.
7. Shaw,D.M. (2006) Factor V Leiden: an overview. Clin Lab Sci, 19, 218-221.
8. Lambert,G. (2007) Unravelling the functional significance of PCSK9. Curr Opin Lipidol, 18, 304-309.
9. Gu,D.F., Hinks,L.J., Morton,N.E., Day,I.N. (2000) The use of long PCR to confirm three common alleles at the CYP2A6 locus and the relationship between genotype and smoking habit. Ann. Hum. Genet., 64, 383-390.
10. Huang,S., Cook,D.G., Hinks,L.J., Chen,X.H., Ye,S., Gilg,J.A., Jarvis,M.J., Whincup,P.H., Day,I.N. (2005) CYP2A6, MAOA, DBH, DRD4, and 5HT2A genotypes, smoking behaviour and cotinine levels in 1518 UK adolescents. Pharmacogenet. Genomics, 15, 839-850.

Names of potential supervisors

Dr Santiago Rodriguez
Professor Ian Day

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

Background

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

Objectives

1) To verify that Hp genotype is a robust instrument for predicting Hb levels in mothers and children of the ALSPAC cohort.
2) To test the hypothesis that maternal Hp genotype / variations in maternal

haemoglobin levels during pregnancy are correlated with the following phenotypes:
offspring birth weight (2)
offspring SGA (small for gestational age)(3)
risk of postpartum depression (4)
risk of preterm PROM (premature rupture of membranes)(5)
risk of preterm birth (6)

3) To test the hypothesis that children's Hp genotype / variations in their haemoglobin levels are correlated with differences in IQ and growth.
4) To explore the relevance of Hp genotype to the interpretation of Hb assay in decision cutpoints in clinical situations of anaemia and its management.

Design

1) Tagging SNPs for the duplication, if available, will be selected using HapMap(7). If no suitable tagging SNPs can be found, an in-house liquid-phase copy number assay will be refined and used for genotyping.

2) Hp genotypes of the ALSPAC cohort will be analysed with Hb data to quantify the association with Hb levels.
3) The suitability of Hp genotype as a novel predictor for certain Hb-related phenotypes will be assessed.

Techniques/approaches

Hp genotyping will be outsourced if practicable, or else carried out in our laboratory using

either SNP genotyping or a high-throughput liquid-phase assay developed in-house. Hb levels in mothers and children of the ALSPAC cohort have already been measured. The remainder of the studentship will involve acquiring and processing ALSPAC phenotype data and investigating potential mechanisms for any perceived correlations.

References

1. Langlois,M.R. and Delanghe,J.R. (1996) Biological and clinical significance of haptoglobin polymorphism in humans. Clin Chem, 42, 1589-1600.
2. Steer,P.J. (2000) Maternal hemoglobin concentration and birth weight. American Journal of Clinical Nutrition, 71, 1285S-1287S.
3. Scanlon,K.S., Yip,R., Schieve,L.A. and Cogswell,M.E. (2000) High and low hemoglobin levels during pregnancy: Differential risks for preterm birth and small for gestational age. Obstetrics and Gynecology, 96, 741-748.
4. Corwin,E.J., Murray-Kolb,L.E. and Beard,J.L. (2003) Low hemoglobin level is a risk factor for postpartum depression. Journal of Nutrition, 133, 4139-4142.
5. Ferguson,S.E., Smith,G.N., Salenieks,M.E., Windrim,R. and Walker,M.C. (2002) Preterm premature rupture of membranes: Nutritional and socioeconomic factors. Obstetrics and Gynecology, 100, 1250-1256.
6. Scanlon,K.S., Yip,R., Schieve,L.A. and Cogswell,M.E. (2000) High and low hemoglobin levels during pregnancy: Differential risks for preterm birth and small for gestational age. Obstetrics and Gynecology, 96, 741-748.
7. Gibbs,R.A., Belmont,J.W., Hardenbol,P., Willis,T.D., Yu,F.L., Yang,H.M., Ch'ang,L.Y., Huang,W., Liu,B., Shen,Y. et al. (2003) The International HapMap Project. Nature, 426, 789-796.

Supervisors

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

4 Title: Detection of "non-tagged" CNVs and association analyses with human body weight related traits

Background

Genome-wide association scans (GWAS) have been very successful for the identification of candidate genes for many human complex traits. However, the variance explained by the identified genes does not explain all the heritability observed for these traits. One possible reason to explain the “missing heritability” is that GWAS are based on common genetic variation, in particular on Single Nucleotide Polymorphisms (SNPs). There are other types of genetic variation potentially accounting for the “missing heritability”, including Copy Number Variants (CNVs). There are examples supporting the existence of “non-tagged CNVs”, that is, CNVs that do not show enough pairwise linkage disequilibrium (LD) to be indirectly studied by a highly correlated SNP. It follows that the effects of these CNVs cannot be studied by GWAS and need to be analysed separately. Available Data

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a longitudinal, population-based birth cohort study that recruited 14,541 pregnant women residing in Avon, United Kingdom, and their children (14,062 liveborn children). It is densely phenotyped, including information for anthropometric traits, metabolic traits, cardiovascular risk traits, blood biomarkers, other complex traits and data from questionnaires.

Objectives and Plan

• To identify, from the literature and from databases, CNVs within candidate genes influencing body weight related traits
• To genotype the CNVs in the ALSPAC cohort
• To link the CNVs with SNP genotypes available in ALSPAC
• To characterise pairwise LD between each CNV and neighbouring SNPs
• To perform genetic association analyses in relation to body weight variation Techniques/approaches
• Laboratory analyses to genotype CNVs with the technique 'Amplification Ratio Control System' (ARCS)
• Bioinformatic analyses to characterise LD with the software CubeX and MIDAS
• Statistical analyses to perform standard genetic association analyses"

Supervisors

Dr Santiago Rodriguez and Dr Tom Gaunt

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

Outline of Project:
Background

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

Objectives

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

Design

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

The project is likely focus on a few important clinical analytes and use a combination of approaches drawing from clinical biochemistry (or other laboratory science discipline), from contemporary genome wide association data (the first emerging in 2007) and on a range of informatic, statistical and epidemiological methods also.

References

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

Supervisors

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

6 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

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

Background

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

Objectives and Design

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

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

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

Focussed genotype-selected studies – A group of 400 15-year-old children and their mothers, selected on the basis of genotype, will be invited to a research clinic for collection of fasting blood samples. Rare homozygotes will be oversampled to increase stastistical power. These will be analysed for triglycerides, apoB, total cholesterol, HDL cholesterol, LDL particle size and the LDL subfractions LDLI, LDLII and LDLIII. Measurements of these lipoprotein subfractions can be used to give an indication of postprandial triglyceridaemia and metabolism20. Assessment of recent diet and physical activity will be made, as these have been shown to affect fat tolerance21,22.

Techniques/approaches

Nutritional studies in adolescents (clinical research) – wide variety of aspects
Bioinformatics – mainly genomic and genetic
Statistical and statistical genetic analyses –broad range of methodologies

Supervisors

Andy Ness, Ian Day

8 Title: Disentangling instances of causally and pharmacogenetically relevant genomic confounding

Outline of Project:
Background

The genes for growth hormone (GH1) and angiotensin converting enzyme (ACE) reside within a 450kb region on human chromosome 17. Research by us and others has demonstrated associations between polymorphisms in the ACE (>1000 papers) and GH11 gene region and cardiovascular, metabolic and numerous other risk phenotypes. Since RAS pathway inhibitors are widely used, and recombinant GH is also used, the pathway inferences are potentially of pharmacogenetic significance. We have demonstrated ~20% linkage disequilibrium between the ACE and GH1 genes using our own data2, confirmed with HapMap data (www.hapmap.org)3. Whilst it is apparent that factors in this region influence disease risk, we have demonstrate that pathway misinference may arise due to the high levels of linkage disequilibrium2. The use of HapMap data3 enables the identification of other causally and pharmacogenetically important genomic regions in which LD may cause genomic confounding.

Objectives

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

Design

1. SNPs will be selected using data from the HapMap project3 to select haplotypetagging SNPs.
2. Pharmacogenetically and causally important genomic regions will be identified from the literature and LD across those regions analysed using data from the HapMap project3.
3. Genotyping work will be outsourced using cohort studies based at University of Bristol for which ethical approval is already in place.
4. Haplotype analyses for association with phenotypes will use various programs as described in previous work2,4. Techniques/approaches

The project will be principally bioinformatic and statistical, with the majority of original data being obtained by outsourcing. However, some problem SNPs may require in-house assay development and genotyping, particularly within the GH1 region due to high sequence homology between GH1 and four related genes.

References

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

Supervisors

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

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

Outline of Project:
Background

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

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

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

Objectives of the PhD

1.- To confirm the original PSA measurements in individuals with apparently very low PSA levels, in order to rule out possible measurement or other errors.
2.- To characterise in detail KLK3 for major physical changes (via determination of exonic dose in relation to a reference gene).

3.- To perform sequence level studies of KLK3 in order to look for inactivating mutations responsible for absent or very low expression of KLK3 (e.g. stop codon mutations, deletions, splicing mutations).
4.- To extend this principle to other biomarkers

Techniques and approaches

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

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

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

References

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

8.Rodríguez S, Chen XH, Miller GJ, Day IN (2005) Non-recombining chromosome Y haplogroups and centromeric HindIII RFLP in relation to blood pressure in 2,743 middle-aged Caucasian men from the UK. Hum Genet 116: 311-318.

Supervisors

Dr Santiago Rodriguez (santi.rodriguez@bristol.ac.uk)
Professor Ian Day (ian.day@bristol.ac.uk)

10 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.

11 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. RBG is a study design where the recruitment of a sub-set of participants from an existing study, analysis of their biosamples or collection of new data is undertaken on the basis of measured genotypic variation. Often, the exhaustive collection of phenotypic measurements in large studies is impractical and not the most efficient approach to the allocation of finite resources. Therefore the targeted measurement of dense or precise phenotypic characteristics within selected groups is appealing. This programme will comprehensively assess and test the use of genotypes known to be correlated with features of interest, or which require further examination given unexplained correlations with disease, to define strata or subsamples for intense or directed phenotypic data collection. This will allow examination of detailed phenotypic information in financially and pragmatically feasible sample sizes, with analytical power optimized for measurement depth and precision. The motivation for employing genotypic data in this way is the pursuit of causal relationships based on the Mendelian randomization paradigm. This innovative programme will develop the RBG approach methodologically, test the properties of genotypically assigned groups and apply RBG to a series of worked examples.

There are numerous benefits to the undertaking of RBG experiments which will be the focus of work undertaken 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.

Disciplines and training

Genomics

Causal analysis methods

Statistical methods

Relevant training as part of the programme of Wellcome Trust Advanced Courses

References

Greenland, S. An introduction to instrumental variables for epidemiologists. *International Jounal of Epidemiology* **29**, 722-729 (2000).

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).

Palmer, T.M. *et al.* Using multiple genetic variants as instrumental variables for modifiable risk factors. *Statistical Methods in Medical Research* (2011).

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).

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).

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).

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)

12 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 analyse 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:

To integrate data from multiple omic sources and interrogate their inter-relationship and association with a range of phenotypes or traits.

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:

Wadman, M. Genome miners rush to stake claims. *Nature* **447**, aop (2007).

Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genetics* **44**, 1084-9 (2012).

Wurtz, P. *et al.* Metabolic signatures of insulin resistance in 7,098 young adults. *Diabetes* **61**, 1372-80 (2012).

Tukiainen, T. *et al.* Detailed metabolic and genetic characterization reveals new associations for 30 known lipid loci. *Human Molecular Genetics* **21**, 1444-55 (2012).

Li, J. & Burmeister, M. Genetical genomics: combining genetics with gene expression analysis. *Human Molecular Genetics. 14 Spec No* **2**, R163-9 (2005).

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)

13 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:

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

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

Objectives :

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

Background

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

Assessment of scoliosis

Currently, gold standard assessment of scoliosis is achieved through x-ray examination, however this is not ideal for multiple examination and is not always available in large numbers of individuals. Scoliometers can be employed to assess the degree of spinal curvature 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 ($R^2=0.998$). DXA measures also provide an indication of size, site and direction of curve, as well as measures of local bone quality and density without the large radiation exposure of traditional spinal radiograph. In light of this, members of the musculoskeletal centre at Southmead Hospital, Bristol have been developing more accurate measures of spinal deformity and applying these to the Avon Longitudinal Study of Parents and Children (http://www.bristol.ac.uk/alspac/).

Potential causes of scoliosis

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

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

Hyderabad Nutrition Trial

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

Plan of investigation

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

We anticipate that the student will use data from the Hyderabad Nutrition Trial which will already be available in the 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

15 Title: Genetic and Environmental precursors of Chronic Obstructive Pulmonary Disease (COPD)

Background

Chronic Obstructive Pulmonary Disease (COPD) is a leading cause of morbidity, mortality and disability and a major public health problem worldwide. However, public awareness of COPD is much lower than for other lung diseases and COPD is frequently under-diagnosed, even in high-income countries [1-2]. Due to the slow and progressive nature of the disease a significant and irreversible loss of lung function can develop before symptoms of the disease appear [3-4].

Smoking is the main environmental risk factor that has been associated with COPD, but estimates of the population attributable fraction (PAF) of COPD explained by smoking are generally lower than 80% and the prevalence of COPD in never smokers is between 3-15% [5]. This suggests that hitherto unrecognised genetic and environmental factors may modify individual risk by influencing rates of lung function decline and hence the probability of developing COPD. Apart from smoking, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) identifies as important risk factors for COPD occupational dusts and chemicals, outdoor and indoor air pollution, socioeconomic status, genetic factors, reduced lung growth, airway hyperresponsiveness, and low baseline lung function in early adulthood.

Homozygous inheritance of genes for the deficient state of the protective anti-protease, alpha1 antityipsin is currently the only well-established genetic determinant for severe COPD and rapid lung function decline, especially when combined with smoking [6]. The role of other genetic variants associated with antioxidant defences is less clear but some of these may interact with early exposure to tobacco smoke to influence lung function development in childhood. Bronchial hyperreactivity is another independent determinant for lung function deterioration and progressive airway obstruction in smokers with early COPD [7-8].

All above mentioned factors lead to respiratory dysfunction, which in its turn is an independent prognostic factor for mortality in smokers [9]. Knowledge of early-life determinants of COPD, particularly in relation to external risk-factors may determine the design and development of effective preventive strategies of the disease.

The aim

This study will focus on early-life determinants of low lung function at ages 8 and 15 years, including genetic and other host and behavioural factors. It aims to examine the interrelationships between factors such as genetic background, intrauterine foetal development and growth during early childhood, childhood respiratory disorders, airway hyperresponsiveness and wheezing history, in combination with exposure to tobacco smoke, air pollution and dietary patterns, and their possible role in early-life lung function impairment and development of respiratory symptoms.

We will also address novel hypotheses, including the association between maternal and early childhood diet (including antioxidant intake), distance of residence from main roads as a marker of traffic-related pollution exposure, and interactions between these variables and tobacco smoke exposure and irreversible airways obstruction in adolescence.

Methods

The research will be based in the Avon Longitudinal Study of Parents and Children (ALSPAC), a longitudinal, prospective birth-cohort study recruited in pregnancy that has followed a population of nearly 12,000 children since birth and measured post-bronchodilator pulmonary function at age 15+ years (the main outcome variable for this project) in approximately 5000 of them.

This project will lead to the development of skills in managing, analysing and interpreting lung function data in a large, population-based sample. Data analysis methods will include logistic and linear regression models and the student will also have the chance to develop expertise in the use of longitudinal growth models and in life-course epidemiological methods. There will also be the opportunity to conduct genetic analyses based on the genome-wide data now available for the majority of ALSPAC participants, and emerging data from next-generation sequencing technologies. Key gene-environment interactions postulated in the literature will be investigated in ALSPAC and in collaboration with similar cohorts within Europe.

The project is supervised by a multi-disciplinary team with expertise in respiratory measurement and clinical medicine (Professor Henderson), and epidemiology and biostatistics (Professor Sterne).

Expected results/practical application

The results of the research will be used to guide development of preventive strategies for COPD based on modifiable early-life determinants. The results of the research will provide scientific evidence for the conceptual model for Stepwise Target Group Prevention (STOP) of COPD [10].

References

1. Calverley, P.M., COPD: early detection and intervention. Chest, 2000. 117(5 Suppl 2): p. 365S-71S.
2. Pauwels, R., COPD: the scope of the problem in Europe. Chest, 2000. 117(5 Suppl 2): p. 332S-5S.
3. Fletcher, C. and R. Peto, The natural history of chronic airflow obstruction. Br Med J, 1977. 1(6077): p. 1645-8.
4. Sherman, C.B., et al., Longitudinal lung function decline in subjects with respiratory symptoms. American Review of Respiratory Disease, 1992. 146(4): p. 855-9.
5. Eisner, M.D., et al., An official American Thoracic Society public policy statement: Novel risk factors and the global burden of chronic obstructive pulmonary disease. Am J Respir Crit Care Med, 2010. 182(5): p. 693-718.

6. Corbo, G.M., et al., Passive smoking and lung function in alpha(1)-antitrypsin heterozygote schoolchildren. Thorax, 2003. 58(3): p. 237-241.

7. Jensen, E.J., R. Dahl, and F. Steffensen, Bronchial reactivity to cigarette smoke; relation to lung function, respiratory symptoms, serum-immunoglobulin E and blood eosinophil and leukocyte counts. Respir Med, 2000. 94(2): p. 119-27.
8. Lapperre, T.S., et al., Dissociation of lung function and airway inflammation in chronic obstructive pulmonary disease. Am J Respir Crit Care Med, 2004. 170(5): p. 499-504.
9. Stavem, K., et al., Lung function, smoking and mortality in a 26-year follow-up of healthy middle-aged males. Eur Respir J, 2005. 25(4): p. 618-25.
10. Ghambarian, M.H., et al., COPD: can prevention be improved? Proposal for an integrated intervention strategy. Prev Med, 2004. 39(2): p. 337-43.

Supervisors:

John Henderson & Jonathan Sterne

16 Title: Analysis of common complex traits and genomewide sequence data.

Genetic epidemiology employing genomewide whole sequence data from the Avon Longitudinal Study of Parents and Children (ALSPAC) with diverse phenotypic data collections.

The arrival of genomewide sequence data to the Avon Longitudinal Study of Parents and Children presents an opportunity 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) andfurther 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.

This proposal aims to explore the potential in analyzing the spectrum of genetic variation existing in low read depth genomewide sequence data alongside a rich phenotypic resource, to use this data and imputed data from it within the entire ALSPAC cohort (www.bris.ac.uk/alspac) to assess the impact of intermediate frequency variation on specific phenotypes of interest and to assess the potential for this type of genetic data in the assessment of exomic coding variation.

This work will feed off the growing efforts to analyses like data from 1000 genomes (www.1000genomes.org) and will employ sequence data derived from the 2000 ALSPAC samples part of the UK10K project currently ongoing (www.uk10k.org).

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 or also in the parallel analysis of data from numerous sources including the methylome, the transcriptome the environment.

Objectives:

(i) Characterise low read depth sequence data within the ALSPAC study and use this to refine imputed genotypes to low frequency (~1%) across the entire ALSPAC cohort (including mothers and children)
(ii) Undertake analyses into the relationship between genomic variation within this data set and specific phenotypes within the ALSPAC study
(iii) Assess and or develop methods for the undertaking of genotype/phenotype association studies.

Supervisors

George Davey Smith, Nic Timpson

17 Title: Diurnal patterns of physical activity, sedentary behaviour, heart rate and blood pressure in adolescents

Description: Relatively little is known about the determinants or consequences of different patterns of 24-hour change in blood pressure (BP) in healthy adolescents. There are a number of methodological challenges to simultaneously modeling change data for different measurements, but the potential value of being able to do so is increasingly recognized. The availability of detailed data on 24-hour monitoring of BP and accelerometer assessed activity in healthy 17-18 year olds, together with a wealth of additional data in this birth cohort, provides a unique opportunity to address these gaps.

Below are a number of potential objectives. These are likely to cover more than one distinct PhD. We would anticipate students studying in depth 1-3 of the objectives and indeed adding their own areas of interest in relation to the broad topic of diurnal patterns of physical activity and/or blood pressure in adolescence.

1. To describe diurnal patterns of accelerometer assessed physical activity, sedentary behavior, SBP, DBP and heart rate in healthy 17-18 year olds, measured using 24-hour continuous monitoring.

2. To examine the association of diurnal patterns of accelerometer assessed physical activity and sedentary behaviour with diurnal patterns of SBP, DBP and heart rate.

3. To describe diurnal patterns of central (aortic) BP and compare this with diurnal patterns of peripheral (brachial) BP.

4. To examine whether diurnal variability in BP is associated with other cardiovascular risk factors independently of mean level of BP

5. To examine whether 'dipping' (change in BP between day and night time) is associated with other cardiovascular risk factors independently of mean level of BP

6. To examine the associations of change in weight and height from birth to 17 years and fat and lean mass from 9-17 years with diurnal patterns of SBP, DBP and heart rate at age 17-18

7. To examine the associations of change in physical activity and sedentary behaviour (accelerometer assessed) from age 11-17 with diurnal patterns of SBP, DBP and heart rate at age 17-18

8. To examine the association of hypertensive disorders of pregnancy and change in maternal BP in pregnancy with diurnal patterns of SBP, DBP and heart rate at age 17-18

9. To examine whether genetic variants associated with differences in BP in adults are associated with different diurnal patterns of BP in adolescents

For more information about the actual PhD project (as opposed to adim. Student issues) contact Debbie Lawlor – d.a.lawlor@bristol.ac.uk

Supervisors: There are a number of potential phd projects for this proposal from the very statistical methodology to the much more applied.

Depending on the particular focus of the project supervision would be by 2 or 3 of the following:

Prof Kate Tilling, Dr Laura Howe (University of Bristol), Dr Abi Fraser, Dr Corrie Macdonald-Wallis, Prof Debbie Lawlor

The work will be done in collaboration with colleagues in Imperial (London) and one of either Prof Alun Hughes or Prof Nish Chaturvedi from Imperial will provide additional supervision, in particular on the clinical relevance and meaning of some of the measurements.

18 Title: Epigenetic biomarkers of prostate cancer risk

Outline of project:

Epigenetic changes arise early in the pathogenesis of prostate cancer and provide a very promising avenue for the identification and development of biomarkers which can be applied to early detection, diagnosis, staging, risk stratification and treatment monitoring. Numerous alterations to DNA methylation (the most commonly studied form of epigenetic modification) have been documented in studies of prostate tumour tissue. Recent methodological developments open the possibility of the detection of methylation profiles in low levels of circulating tumour DNA in peripheral blood. Furthermore, the exploration of DNA methylation profiles in peripheral blood DNA as early biomarkers of prostate cancer is recognised as a highly promising area. Much of the work published to date has focused on the analysis of tumour biopsy samples and has been limited to a handful of genes (e.g. *GSTP1*, *APC* and *RASSF1a*). A more comprehensive appraisal of epigenetic variation in a large cohort of prospectively collected biological samples pre-dating diagnosis with detailed information on diagnosis, staging, and treatment would greatly enhance progress in this field and assist in addressing the major problems associated with risk stratification in this form of cancer.

Aims:

The aims of the PhD are to:

1. Undertake a case control study (based on 400 cases and 400 controls) of an epigenome-wide analysis of DNA methylation in prostate cancer to identify epigenetic associations with prostate cancer;
2. Apply novel methodological approaches to assess the causality of observed epigenetic associations with prostate cancer; based on triangulating genetic associations of epigenetic changes with their associations with prostate cancer.
3. Utilise a novel methodology – biomarker demendelization – to purge biomarker measures of the genetic contribution to increase the predictive ability of epigenetic changes.
4. Compare methylation in genomic DNA with benign prostate cancer cores to assess how well the methylation of white blood cell DNA correlates with the benign prostate tissue

Proposed Studies:

Matched case-control study nested within the U.K. population–based Prostate testing for cancer and Treatment (ProtecT) study of PSA detected prostate cancer in men aged 50-69 years (400 cases and 400 controls). Between 2001 and 2009, all (approximately 227,300) men aged 50-69 years in 300 general practices located around nine UK cities (centres) were invited to have a PSA test at a prostate check clinic appointment. Participants with a PSA level between 3.0 and 19.9 ng/mL (approximately 10% of men tested) were invited to attend the centre's urology department for digital rectal examination and 10-core trans-rectal ultrasound-guided biopsy. Men with a PSA level $\geq$ 20 ng/mL were referred as a matter of urgency to a urologist, and were eligible to participate in the treatment trial only if localized cancer was confirmed. A diagnosis of localized prostate cancer was defined as a positive biopsy, clinical stage T1-T2, NX, M0; advanced prostate cancer was defined as positive biopsy, clinical stage T3-T4 or N1 or M1. All men provided written informed consent. Trent Multicentre Research Ethics Committee approved the ProtecT study and allied prostate cancer research under the auspices of ProMPT (Prostate Mechanisms of Prostate cancer and Treatment).

The matched cases and controls will be randomly selected and assessed using a recently released genome-wide DNA methylation array (Illumina HumanMethylation450) which assays 450K CpG

sites in the human genome. The assays will be run in the ALSPAC laboratories in the School of Social and Community Medicine, where an existing project (ARIES, http://www.ariesepigenomics.org.uk/) has already established the hardware and informatics to run this array. Causality will be assessed using novel statistical methods described by co-supervisors, Professors Relton and Davey Smith, in the *Int J Epidemiol 2012;41:161–176*. The PhD students will also utilise a novel methodology – biomarker demendelization – to purge biomarker measures of the genetic contribution to increase predictive ability.

In line with current consensus to modelling power requirements in epigenome wide association studies, we propose to follow the models of Balding suggesting that a sample size of 400 case-control pairs provides good power to detect anticipated effect sizes (OR 1.5) (*Nature Rev Genet*, doi:10.1038/nrg3000).

In a collaboration with Professor David Neal we have the potential to access paired genomic DNA and benign prostate cancer cores to assess how well the methylation of white blood cell DNA correlates with the benign prostate tissue. This will provide important insights into the use of non-invasive blood sampling as a marker of tissue-level methylation. Access to tumour tissue will also be possible to investigate any mechanistic links between perturbed methylation profiles in blood and those observed in prostate tumours.

The proposed project is original, multidisciplinary (oncology, epigenetics, epidemiology) and is founded upon exceptional research resources. The consumables cost to undertake this work (approximately £100K) will be supplemented from other secured sources, providing an

outstanding opportunity for a productive, data rich studentship.

Industrial partnership will undoubtedly enhance the proposed project and will be particularly pertinent to biomarker development, with the potential for the student to enhance the collaboration over and above biomarker identification.

The student will be trained within a vibrant and varied training environment with supervision from academics with strong research profiles in relevant and complementary fields.
Prostate cancer is the most common cancer in men and the second leading cause of cancer death in the UK. Prostate-specific antigen (PSA) screening has dramatically increased the early diagnosis of prostate cancer because it is highly sensitive. However, the specificity of PSA screening is only 25%, resulting in many unnecessary biopsies and, ultimately, substantial overtreatment. A highly specific circulating biomarker (e.g. using blood samples) that complements the traditional PSA test would therefore be an important advance. A specific and non-invasive test would allow patients to avoid the physical pain and discomfort associated with biopsies and avoid the adverse effects and unnecessary costs resulting from overtreatment. As blood sampling is already essential for PSA screening an additional measurement of whole blood biomarkers would not place any extra burden on patients. Whole genome DNA methylation signatures have the potential to elicit a panel of informative biomarkers with predictive utility.

Supervisors:

Richard Martin, Professor of Clinical Epidemiology with expertise in prostate cancer aetiology and screening, and co-I on the ProtecT trial. Supervised 6 PhD students to completion and currently supervising 5 students. Director of Graduate Studies and lead on Wellcome Trust 4 year PhD programme in Lifecourse, Molecular & Genetic Epidemiology (2008-2013, 18 students).

George Davey Smith, Professor of Clinical Epidemiology with expertise in causal inference, Mendelian randomization and epigenetics. Supervised over 20 PhD students to completion and currently supervising 5 students. Director of Wellcome Trust 4-year PhD programme in Lifecourse, Molecular & Genetic Epidemiology.

Caroline Relton, Professor of Epigenetic Epidemiology with expertise both of these disciplines. Supervised 2 PhD students to completion and currently supervising 3 students.

Infrastructure is very well established to support PhD students through doctoral studies. The School of Social & Community Medicine attracts high calibre students and has a very high success rate of timely submission and subsequent employment including many fellowships. We have an internationally renowned Short Course Programme (http://www.bris.ac.uk/social-community-medicine/shortcourse), providing unparalleled access to skills training for all our students.

19 Title: Mobile Platform Devices for Measuring Lifestyle-Related Cognitions

Outline of project:

Impulsive behaviour is a defining feature of various lifestyle behaviours, including tobacco and alcohol use, and overeating / obesity. These behaviours are associated with a cognitive bias or readiness to process behaviour-related stimuli over others, which is triggered by the incentive value associated with such stimuli. Increased levels of impulsivity can provoke relapse to these behaviours and contribute to failed attempts to reduce them. The relationship between cognitions and lifestyle behaviours is therefore well-established, and evidence is also growing that cognitions play a causal role. They may also be amenable to direct modification through cognitive training techniques; for example, it may be possible to "train" attention away from relevant stimuli or improve the inhibition of responses to these stimuli. However, this research is limited by the inability to measure these cognitions dynamically and in naturalistic settings over extended periods of data collection. Most research is conducted in laboratory settings which are highly controlled, typically only assess function at one time point, and do not contain many of the cues to behaviour encountered in everyday life. By taking repeated measurements in natural contexts, we will get an understanding of dynamic changes in cognitive performance, which will help to address the question of whether differences reflect stable trait-like or variable state-dependent variables. Previous studies have shown the value of ecological momentary assessment of this kind, with response inhibition measured in the laboratory having been shown to be related to subsequent smoking behaviour.

Aims:

This studentship will therefore address this gap, by developing mobile platform apps which assess relevant lifestyle-related cognitions (response inhibition, cognitive bias) and then validating these in relevant groups (e.g., cigarette smokers, alcohol consumers, and overweight/obese individuals).

Methods:

The student will work with supervisors in Social and Community Medicine (Lewis) and Experimental Psychology (Munafò), with support from the Department of Computer Science, to develop tasks which capture relevant cognitions and can be delivered via the internet, tablet computers and smartphones. These tasks will then be validated against existing laboratory-based measures in relevant populations (e.g., cigarette smokers). Later stages of the PhD might include piloting the collection of data in a naturalistic, ambulatory setting to investigate the feasibility of this approach.

References:

Luijten M, Littel M, Franken IH. Deficits in inhibitory control in smokers during a Go/NoGo task: an investigation using event-related brain potentials. PLoS One 2011;6(4):e18898.

Batterink L, Yokum S, Stice E. Body mass correlates inversely with inhibitory control in response to food among adolescent girls: an fMRI study. Neuroimage 2010;52(4):1696-703.

Rubio G, Jimenez M, Rodriguez-Jimenez R, Martinez I, Avila C, Ferre F, et al. The role of behavioral impulsivity in the development of alcohol dependence: a 4-year follow-up study. Alcohol Clin Exp Res 2008;32(9):1681-7.

Berkman ET, Falk EB, Lieberman MD. In the trenches of real-world self-control: neural correlates of breaking the link between craving and smoking. Psychol Sci 2011;22(4):498-506.

Penton-Voak IS, Bate H, Lewis G, Munafo MR. Effects of emotion perception training on mood in undergraduate students: randomised controlled trial. Br J Psychiatry 2012;201:71-2.

Supervisors:

Marcus Munafo (Experimental Psychology) and Glyn Lewis (Social and Community Medicine)

20 Title: DNA methylation, one carbon metabolism and lung cancer – identifying early biomarkers of disease risk

Outline of project:

Lung cancer remains one of the most lethal cancers worldwide and is responsible for 1.4 million deaths annually. Little is known about the mechanisms that contribute to lung cancer other than an important role of smoking and some occupational exposures. Prognosis is often poor as diagnosis tends to occur once disease is well established and there is a pressing need to identify early diagnostic biomarkers. We recently demonstrated a pronounced protective effect of elevated vitamin B6, methionine and folate levels (all components of the one carbon metabolism) on lung cancer risk [1]. One carbon metabolism is the major source of methyl donors for the methylation of many biological molecules including DNA. We therefore postulate that deficiencies in one carbon metabolism may be early antecedents of lung cancer and mediate their effect through alteration of DNA methylation patterns and thus gene expression. This project will build on an existing large multi-national collaborative effort to profile one carbon metabolism intermediates in 22 cohorts of lung cancer cases and controls. DNA methylation analysis is being undertaken on a subset of these. This PhD project will complement and extend this initiative through the analysis of data generated using novel causal analysis methods [2]. The project will utilise Mendelian randomization approaches [3] to establish the causal relationship between components of the one carbon metabolism, DNA methylation and lung cancer. This approach will help to decipher whether one carbon metabolites and DNA methylation are causally related to lug cancer and thus potential targets for intervention and prevention or whether they are non-causal biomarkers with potential utility in early diagnosis. The opportunity to spend a period of study in Genetic Epidemiology Unit at the International Agency for Research in Cancer (IARC, Lyon) will be available.

References:

1. Johansson M, Relton C, Ueland PM, et al. Serum B vitamin levels and risk of lung cancer. JAMA. 2010 Jun 16;303(23):2377-85.
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 Feb;41(1):161-76.
3. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003 Feb;32(1):1-22.

Supervisors:

Caroline Relton, George Davey Smith

21: Title: Epigenetic variation and type 2 diabetes related traits; investigating causal relationships

Outline of projects:

Epigenetic variation is believed to play a role in many common complex diseases, including type 2 diabetes (T2D) (1). Epigenetic markings on the genome are involved in the regulation of gene expression and the most widely studied form of modification is DNA methylation. DNA methylation patterns change over time and in response to endogenous and exogenous factors. This means that it is difficult to decipher whether, in a condition such as type 2 diabetes where there is often a protracted pre-clinical phase prior to diagnosis, DNA methylation changes occur as a consequence of the diabetic state or are causally involved in the pathogenesis of the disease. This project will identify associations between DNA methylation and T2D, glucose tolerance and insulin sensitivity. Associated loci will then be interrogated using Mendelian randomization based approached to probe causal relationships.

References:

1. Drong AW, Lindgren CM, McCarthy MI. The Genetic and Epigenetic Basis of Type 2 Diabetes and Obesity. Clin Pharmacol Ther. 2012 Oct 10. doi: 10.1038/clpt.2012.149. [Epub ahead of print]

Supervisors:

Caroline Relton, Debbie Lawlor, Elliott

22 Title: Epigenetic variation and childhood body composition; a Mendelian randomization approach to investigating causal relationships

Outline of project:

The prevention of childhood obesity remains a high research priority. Integral to this is a clearer understanding of the molecular mechanisms involved. Epigenetic marks are established *in utero* and early life and contribute to the regulation of gene expression. Differences in epigenetic patterns in early life have been linked to subsequent differences in body composition in childhood [1,2], however studies to date have been small and inconclusive. This project will explore the potential causal relationship between epigenetic patterns (DNA methylation) and childhood body composition with more rigour. Observational associations between cord blood DNA methylation and indices of body composition (fat and lean mass, height, weight etc) will be investigated. Causal analysis method including the Mendelian randomization approach [3] will be applied to robustly assess whether DNA methylation is a determinant of future childhood body composition or a biomarker with no causal role. This project will use existing data from the Avon Longitudinal Study of Parents and Children and bring together disciplines of molecular epidemiology, bioinformatics, metabolic health, epigenetics and statistics. It will draw on extensive local expertise and mentorship in Mendelian randomization approaches and extend this method to apply it to epigenetic data. The opportunity to incorporate lab work within this project will be provided, although not mandatory.

1. Relton CL, Groom A, St Pourcain B, Sayers AE, Swan DC, Embleton ND, Pearce MS, Ring SM, Northstone K, Tobias JH, Trakalo J, Ness AR, Shaheen SO, Davey Smith G. DNA methylation patterns in cord blood DNA and body size in childhood. PLoS One. 2012;7(3):e31821. Epub 2012 Mar 14.
2. Godfrey KM, Sheppard A, Gluckman PD, Lillycrop KA, Burdge GC, McLean C, Rodford J, Slater-Jefferies JL, Garratt E, Crozier SR, Emerald BS, Gale CR, Inskip HM, Cooper C, Hanson MA. Epigenetic gene promoter methylation at birth is associated with child's later adiposity. Diabetes. 2011 May;60(5):1528-34. Epub 2011 Apr 6.
3. Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol. 2003 Feb;32(1):1

Supervisors:

Caroline Relton, George Davey Smith, Tom Gaunt

23 Title: Air pollution exposure in early life and lung function in childhood – a role for epigenetic mechanisms

Outline of project:

Exposure to particulate air pollution in early life is believed to have adverse effects upon lung development and subsequent lung function, increasing the risk of asthma and other respiratory disorders. One potential molecular mechanism that might mediate the association between air pollution and lung function is epigenetic regulation of gene expression. This project will explore the relationship between DNA methylation patterns, one form of epigenetic marking, and air pollution exposure modelled over the first 7 years of life. DNA methylation variation associated with exposure will then be interrogated with respect to lung function in childhood and adolescence. Initial work will utilise existing data including extensive air pollution measurements, genome-wide DNA methylation and detailed lung function assessment. Training will be provided in longitudinal modelling of exposure (air pollution), outcome (indicators of respiratory function) and intermediate phenotype (DNA methylation) data over time. The opportunity to undertake laboratory based work will be provided through validation and replication of observational associations. Bioinformatic tools will also be used to explore the relationship of differentially methylated regions with other regulatory features of the genome and biological pathways. This project will utilise the world-leading Avon Longitudinal Study of Parents and Children and combine disciplines of molecular epidemiology, respiratory health, epigenetics and statistics.

Supervisors:

John Henderson, Caroline Relton, Kate Tilling

24 Title: Epigenetic variation in children born following assisted reproductive technology

Outline of project:

Epigenetic patterns are integral to normal development and are comprehensively re-programmed post-fertilisation. The profound differences in the immediate post-fertilisation environment of embryos created through assisted reproductive technologies (ART) compared to those conceived naturally raise the possibility that differences in epigenetic patterns may be generated as a consequence. These differences may have long lasting influences on health and development of children. Concern has been raised that ART increases the risk of some imprinting (epigenetic) disorders but little is known about the potentially subtle influences on the epigenome and developmental trajectories. This project will compare epigenetic patterns (genome-wide DNA methylation) in cord blood samples in children born following ART compared to those conceived naturally. ART-associated differences in DNA methylation will be considered as predictors of a range of phenotypic traits in childhood including cognitive and motor function, and body composition. Initial work will utilise existing data with the opportunity to undertake laboratory based work through validation and replication of observational associations. Bioinformatic tools will be used to explore the relationship of differentially methylated regions with other regulatory features of the genome and biological pathways. This project will utilise the world-leading Avon Longitudinal Study of Parents and Children and bring together disciplines of molecular epidemiology, bioinformatics, reproductive health, epigenetics and statistics.

Supervisors:

Caroline Relton, George Davey Smith, Debbie Lawlor

25 Title: Epigenetic and metabolomic mediators of associations between childhood growth and chronic disease

Background

Patterns of growth in childhood are associated with later chronic disease, but the causality and underlying physiological mechanisms of these associations are poorly understood. Higher BMI – measured in childhood or in adulthood – is associated with greater risk of many diseases, including cancers, cardiovascular diseases, and respiratory diseases. The association between height and chronic disease is more complex, with shorter height showing associations with greater risk of cardiovascular and respiratory diseases, but lower risk of some cancers.

This PhD offers the opportunity to explore these associations and the potential causal mechanisms underlying them, using data from two large prospective studies from Denmark and the UK.

Outline of project

The objectives of this PhD are to:

1) investigate the associations between childhood growth and chronic disease events in adulthood, identifying which features of childhood growth trajectories are related to later chronic disease risk
2) assess the physiological (epigenetic, metabolomic, etc) pathways mediating these relationships
3) establish the causality of these pathways using novel statistical methodologies.

Associations between childhood growth and clinical outcomes will be assessed in a large Danish study, with childhood growth (from school records) and clinical outcomes (from linkage to national registers) available for 372,636 children who attended school in Copenhagen, Denmark from 1936 to 2005.[1 2 3] These relationships could also be assessed in other cohorts with data on both child growth and outcomes in adulthood, e.g. the Christ's Hospital study[4 5], and the Barry Caerphilly Growth Study.[6]

The potential physiological (epigenetic, metabolomic, etc) mediators of these associations will be explored in the Avon Longitudinal Study of Parents and Children[7 8], a prospective birth cohort in which almost 14,000 pregnant women were recruited in the early 1990s. Extensive data are available on child growth[9 10], risk factors for later chronic disease[11 12], and physiological factors that may mediate the relationships between child growth and later chronic disease. All of these traits have been measured at multiple time points, allowing for longitudinal analysis.

In order to establish the causality of the role of epigenetic, metabolomic, and other factors in mediating relationships between child growth and chronic disease, novel statistical methodologies will be used to rule out confounding and reverse causality. For example, the dynamic relationships between changes in growth and changes in epigenetic or metabolomic factors will be assessed using multivariate multilevel models[13], and causality of cross-sectional relationships will be assessed using two-step Mendelian Randomisation.[14]

References

1. Baker JL, Sorensen TI. The Copenhagen School Health Records Register. *Scand J Public Health* 2011;**39**:87-90.

2. Baker JL, Sorensen TI. Obesity research based on the Copenhagen School Health Records Register. *Scand J Public Health* 2011;**39**:196-200.

3. Baker JL, Olsen LW, Sorensen TI. Childhood body-mass index and the risk of coronary heart disease in adulthood. *New England Journal of Medicine* 2007;**357**:2329-2337.

4. Cole TJ, Cortina-Borja M, Sandhu J, Kelly FP, Pan H. Nonlinear growth generates age changes in the moments of the frequency distribution: the example of height in puberty. *Biostatistics* 2008;**9**:159-171.

5. Sandhu J, Ben-Shlomo Y, Cole TJ, Holly J, Davey Smith G. The impact of childhood body mass index on timing of puberty, adult stature and obesity: a follow-up study based on adolescent anthropometry recorded at Christ's Hospital (1936-1964). *Int J Obes (Lond)* 2006;**30**:14-22.

6. Ben-Shlomo Y, McCarthy A, Hughes R, Tilling K, Davies D, Davey Smith G. Immediate postnatal growth is associated with blood pressure in young adulthood: the Barry Caerphilly Growth Study. *Hypertension* 2008;**52**:638-644.

7. 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;

8. 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;

9. Howe LD, Tilling K, Galobardes B, Davey Smith G, Ness AR, Lawlor DA. Socioeconomic disparities in trajectories of adiposity across childhood: findings from a prospective birth cohort. *International Journal of Pediatric Obesity* 2010;**6**:e144-e153.

10. Howe LD, Tilling K, Galobardes B, Davey Smith G, Gunnell D, Lawlor DA. Socioeconomic differences in childhood growth trajectories: at what age do height inequalities emerge? *J Epidemiol Community Health* 2012;**66**:143-148.

11. Howe LD, Tilling K, Benfield L, Logue J, Sattar N, Ness AR, 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**:e15186.

12. Lawlor DA, Benfield L, Logue J, Tilling K, Howe LD, Fraser A, et al. Association between general and central adiposity in childhood, and change in these, with cardiovascular risk factors in adolescence: prospective cohort study. *BMJ* 2010;**341**:c6224.

13. 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;**31**:3147-3164.

14. Relton CL, Davey SG. 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.

Supervisors:

George Davey Smith, Laura Howe with additional external supervision from Professor Thorkild I Sørensen and Dr Jennifer L Baker (Institute of Preventive Medicine, Copenhagen, Denmark)

26 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

27 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

1. To describe the average pattern of change in haemoglobin concentrations during pregnancy.
2. To examine associations of potential risk factors with maternal haemoglobin concentrations and/or patterns of changes in haemoglobin across the course of pregnancy.
3. To investigate the associations of maternal haemoglobin concentrations and/or patterns of changes in haemoglobin during pregnancy with the following outcomes in offspring:
    a. patterns of growth in childhood and adolescence
    b. childhood infections
    c. cardiovascular risk factors (e.g. blood pressure, lipids, insulin, glucose) during childhood and adolescence
    d. cognitive and behavioural outcomes in childhood and adolescence
    e. respiratory health in childhood and adolescence
4. To assess whether there is a causal relationship between maternal haemoglobin concentrations in pregnancy and offspring outcomes using a Mendelian randomization approach.
5. 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:

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

28 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:

Tom Gaunt, Colin Campbell, Jose Soeane

29 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:

Tom Gaunt

30 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: Tom Gaunt, Colin Campbell, Jose Seoane

31 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: Nic Timpson, Matt Jones

32 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: Nic Timpson, Alastair Poole

33 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

34 Title: The Genetics of Asthma Phenotypes

Background:

Results from genome-wide association studies (GWAS) have reported a large number of single-nucleotide polymorphisms (SNPs) associated with asthma outcomes[1-4] as well as putative asthma risk factors[5-9]. We hypothesize that allele scores of doctor-diagnosed asthma, lung function, sensitization and FeNO discriminate among different childhood wheezing phenotypes and that Mendelian Randomization analysis can provide robust evidence of causal effects of key exposures (CRP, vitamins A, E, C and D, selenium, omega-3 fatty acids and exposure to smoking) on different asthma phenotypes. We plan to derive allele scores for these asthma outcomes in the context of risk prediction[10] using results from GWAS and to investigate their associations with wheezing and other asthma-related phenotypes in ALSPAC; Furthermore we plan to derive allele scores for the mentioned key exposures and to use them as instruments to estimate their causal effect on an asthma outcome in the context of Mendelian Randomization[11-12]. The potential utility of using genetic risk scores to investigate causal effects in observational studies combined with better segregation of asthma subtypes will enable us to identify causal factors amenable to intervention to prevent or cure asthma.

The student will 1) complete a Masters Epidemiology MSc programme at the London School of Hygiene and Tropical Medicine, a highly-regarded course enabling him to progress rapidly to advanced statistical methods; 2) be co-supervised by mentors from two complementary disciplines; 3) use ALSPAC data readily-available and 4) have free access to a full programme of short courses in epidemiology, genetic and biostatistics.

Hypothesis and Objectives:

Two main hypotheses will be investigated:

- Allele scores of asthma, lung function, FeNO and sensitization discriminate among different wheezing phenotypes
- Mendelian Randomization analysis can provide robust evidence of causal effects of key exposures (CRP, vitamin C and E, vitamin A, selenium, omega-3, vitamin D and ETS) on different asthma phenotypes

The objectives of the project are:

1. to derive allele scores for asthma, lung function, FeNO and sensitization in the context of risk prediction (prediction scores) using results from GWAS excluding ALSPAC data
2. to investigate associations of the prediction scores from objective 1 with wheezing phenotypes and other asthma-related phenotypes in ALSPAC
3. to derive allele scores for CRP, vitamin C and E, vitamin A, selenium, omega-3, vitamin D and ETS
4. to use the allele scores from objective 3 as instruments to estimate their causal effect on an asthma outcome in the context of Mendelian Randomization

References:

1. Moffatt MF, Gut IG, Demenais F, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med 2010;363(13):1211-21.
2. Soler Artigas M, Loth DW, Wain LV, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet. 2011;43(11):1082-90.
3. Bønnelykke K, Matheson MC, Pers TH, et al. Meta-analysis of genome-wide association studies identifies ten loci influencing allergic sensitization. . Nat Genet. 2013;45(8):902-6.

4. Bouzigon E, Nadif R, Thompson E E et al. A Common variant in RAB27A gene is associated with FeNO levels in adults. J Allergy Clin Immunol. In Press.
5. Dehghan A, Dupuis J, Barbalic M, et al. Meta-analysis of genome-wide association studies in >80 000
subjects identifies multiple loci for C-reactive protein levels. Circulation 2011;22(123):7.
6. Nakamura K, Wada K, Sahashi Y, et al. Associations of intake of antioxidant vitamins and fatty acids with asthma in pre-school children. Public Health Nutr. 2012;1:1-6.
7. Major JM, Yu K, Wheeler W, et al. Genome-wide association study identifies common variants associated with circulating vitamin E levels. Hum Mol Genet. 2011;20(19):3876-86.
8. Wang TJ, Zhang F, Richards JB, et al. Common genetic determinants of vitamin D insufficiency: a genomewide association study. Lancet 2010;376(9736):180-8.
9. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. Nat Genet 2010;42(5):436-40.
10. Spycher BD, Silverman M, Brooke AM, et al. Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. European Respiratory Journal 2008;31(5):974-81.
11. Burgess S, Thompson S. Use of allele scores as instrumental variables for Mendelian randomization.
International Journal of Epidemiology 2013;In press.
12. Granell R, Henderson AJ, Evans DM, et al. Causal effects of BMI, fat mass and lean mass on asthma in childhood: Mendelian randomization study. Plos Medicine Under review.

Supervisors/Collaborators: John Henderson and Raquel Granell

35 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

36 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:
Peter Vickerman, Matthew Hickman and Natasha Martin

37 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:
Peter Vickerman, Matthew Hickman and Natasha Martin

38 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.