ABSTRACT BOOKLET

Molecular Epidemiology Group UK

ANNUAL MEETING
Edinburgh 2022

Pathways to improving population health: bringing together industry, academia, patients, and policymakers

#MEGUK2022

Contents

Welcome

Welcome to the 2022 Molecular Epidemiology Group (MEGUK) meeting.

MEGUK was founded in 1996 as a special interest group within the United Kingdom Environmental Mutagen Society. Still going strong today, our aim is to encourage multi-disciplinary links between epidemiologists, molecular biologists, biochemists, geneticists, toxicologists, pathologists, nutritionists, clinician scientists, public health scientists and others studying the role of environmental and genetic factors in the aetiology of chronic disease.

We have a great programme this year including several talks and posters on the use of molecular markers in disease risk prediction and the monitoring of environmental exposures. We hope you enjoy it!

If you are interested in joining our group or want to find out more about us, please have a look at our website: meg-uk.org, follow us on Twitter @megukepi, or join our Linkedin

If you would like to tweet about the meeting, please use our hashtag #MEGUK2022

Our Committee

Hannah Elliott
Chair

Jonine Figueroa
Treasurer

Robert Hillary
Secretary

Robert Young

Sabine Langie

Rachel Lawrence

Natassia Robinson

For more details see meg-uk.org/our-committee

Programme

09.30-10.30	**Registration, poster mounting and viewing**	
10.30-12.15	**Session One**	*chair: Natassia Robinson*
	10.30	**Opening Remarks**
	10.45	**Invited Speaker:** The patient voice in cancer research. Amanda McCann, Chair of the UCD Patient Voice in Cancer Research (PVCR) committee, University College Dublin
	11.15	**Selected abstract:** Using human genetics to evaluate the causal role of circulating inflammatory markers in risk of adult cancer. James Yarmolinsky, MRC Integrative Epidemiology Unit, University of Bristol
	11.30	**Selected abstract:** Measuring DNA damage in circulating blood cells as a potential biomarker for oesophageal adenocarcinoma. Rachel Lawrence Bart's Cancer Institute, Queen Mary University
	11.45	**Selected abstract:** The effect of weight-loss on the colorectal transcriptome and its relation to colorectal cancer risk. Emma Hazelwood, University of Bristol
	12.00	**Selected abstracts:** population cohorts of interest (Generation Scotland; The UK Longitudinal Linkage Collaboration: a trusted research environment for the longitudinal research community; Born in Scotland in the 2020s Pilot Study) Archie Campbell, Robin Flaig, and Rebecca Reynolds, University of Edinburgh
12.30-13.00	**Lunch, John McIntyre Conference Centre Restaurant**	
13.00-13.30	**Poster Viewing**	
13.15-14.30	**Session Two**	*chair: Robert Young*
	13.15	**Invited Speaker:** The BHF Data Science Centre @ Health Data Research UK: how we can help you with population-based research. Cathie Sudlow, Director of the British Heart Foundation Data Science Centre, University of Edinburgh
	14.00	**Selected abstract:** Long-term metabolomic effects of bariatric surgery. Madeleine L. Smith, University of Bristol
	14.15	**Selected abstract:** Augmenting clinical risk prediction of cardiovascular disease through protein and epigenetic biomarkers. Aleksandra D Chybowska, Institute of Genetics and Cancer, University of Edinburgh
14.30-15.00	**Coffee and Poster viewing**	
15.00-15.45	**Session Three**	*chair: Jonine Figueroa*
	15.00	**Invited Speaker:** Our Future Health: a new UK-wide health research programme. Michael Cook, Executive Director of Epidemiology, Our Future Health
	15.30	**Selected abstract:** Phenome-wide association study of genetically determined B vitamins and homocysteine biomarkers with multiple health and disease outcomes: Analysis of the UK Biobank. Lijuan Wang, University of Edinburgh
	15.45	**Selected abstract:** Using PathWAS associated with proteomics to predict pathway function and relationships with complex traits. Sebastian May Wilson, University of Edinburgh
16.00-16.15	**Prizes and Closing Remarks**	

Useful Information

Wifi … Wifi network: VISIT-ED (no password required)

Venue … more information about the John McIntyre Conference Centre be found on their website: https://www.uoecollection.com/ or by asking at reception

General information … Our registration desk is open between 09.30 and 10.30. If you need information or assistance outside of this time, please ask one of our committee.

Invited Speakers

Amanda McCann
The Patient Voice in Cancer Research

Professor Amanda McCann is a Principal Investigator and Senior Conway Fellow in the UCD Conway Institute of Biomolecular and Biomedical Research in University College Dublin, Ireland. Professor McCanns research centres on understanding resistance to chemotherapy in women presenting with triple negative breast cancer. Professor McCann is active in promotion of Patient and Public Involvement (PPI) in research. We welcome Professor McCann to the meeting to discuss her work in this area.

Cathie Sudlow
The BHF Data Science Centre @ Health Data Research UK: how we can help you with population-based research

Professor Sudlow is the Director of the British Heart Foundation Data Science Centre and the Chief Scientist and Deputy Director of Health Data Research UK. Between 2011 and 2019 Professor Sudlow was the Chief Scientist of Biobank UK. Professor Sudlow's research aims to improve cardiovascular health in the population using large-scale population data and she is a proponent of collaborative and open-science projects. In her presentation, Professor Sudlow will describe how the BHF Data Science Centre @ HDR UK can inform population based research initiatives.

Michael Cook
Our Future Health: a new UK-wide health research programme

Dr Cook is the Executive Director of Epidemiology at Our Future Health and was previously a Senior Investigator at the US National Institutes of Health with research interests in chronic disease epidemiology. We welcome Dr Cook to the meeting to present the Our Future Health initiative – the UKs largest health research programme formed via collaboration between public, private and charity sectors.

Abstracts

Notes:

Presenters are shown in **bold**.

Names underlined are eligible for early career researcher prizes which will be awarded in the final session of the meeting.

Platform presentations

Session One

Using human genetics to evaluate the causal role of circulating inflammatory markers in risk of adult cancer

James Yarmolinsky [1,2], Jamie Robinson [1,2], Kostas K. Tsilidis [3,4], Abbas Dehghan [4], Mattias Johansson [5], Daniela Mariosa [5], Marc J. Gunter [6], Lambertus A. Kiemeney [7], George Davey Smith [1,2] , Richard M. Martin [1,2,8]

1MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom
2Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom
3Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece
4Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, St Mary's Campus, London, W2 1PG, United Kingdom
5Genomics Branch, International Agency for Research on Cancer (IARC/WHO), Lyon, France
6Nutrition and Metabolism Branch, International Agency for Research on Cancer (IARC/WHO), Lyon, France
7Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands
8University Hospitals Bristol, NHS Foundation Trust, National Institute for Health Research Bristol Biomedical Research Centre, University of Bristol, Bristol, United Kingdom

Tumour-promoting inflammation is a "hallmark" of cancer and laboratory and epidemiological studies have reported links between various inflammatory markers and cancer risk. The causal nature of these relationships and, thus, the suitability of these markers as intervention targets for cancer prevention is unclear. We meta-analysed 6 genome-wide association studies of circulating inflammatory markers comprising 59,969 participants of European ancestry. We then used combined cis-Mendelian randomization and colocalisation to evaluate the causal role of 75 circulating inflammatory markers in risk of 32 adult cancers in 347,300 cancer cases and up to 1,015,204 controls. Genetic instruments for inflammatory markers were constructed using genome-wide significant (P<5.0x10-8) cis-acting SNPs (i.e. ±250KB from the gene encoding the relevant protein) in weak linkage disequilibrium (LD, r2<0.10). Effect estimates were generated using inverse-variance weighted random-effects models and standard errors were inflated to account for weak LD between variants with reference to the 1000G Phase 3 European panel. We find strong evidence for effects of genetically-proxied circulating adrenomedullin on breast cancer risk (OR:1.19, 95%CI:1.10-1.29, P=2.02x10-5, H4=84.3%), interleukin-23 receptor on pancreatic cancer risk (OR:1.42, 95%CI:1.20-1.69, P=6.72x10-5, H4=73.9%), antithrombin on triple-negative breast cancer risk (OR:3.62, 95%CI:1.70-7.70, P=8.30x10-4, H4=72.7%), and macrophage migration inhibitory factor on bladder cancer risk (OR:1.14, 95%CI:1.05-1.23, P=1.43x10-3, H4=76.1%), among other findings. Our comprehensive analyses represent the largest human genetics evaluation of circulating inflammatory markers in risk of adult cancers to date. Our findings highlight various novel inflammatory markers implicated in cancer development and suggest pharmacological targeting of these markers as a potential strategy for primary cancer prevention.

Measuring DNA damage in circulating blood cells as a potential biomarker for oesophageal adenocarcinoma

Rachel Lawrence [1], Kathryn Munn [2], Rhiannon Wright [2], Hamsa Naser [2], Hasan Haboubi [3], Lisa Williams [4], Gareth Jenkins [2]

1. Bart's Cancer Institute, Queen Mary University, London, UK
2. Swansea University Medical School, Swansea, UK
3. Gastroenterology department, University Hospital Llandough, Cardiff and Vale Health Board, Cardiff, UK
4. Gastroenterology department, Singleton Hospital, Swansea Bay University Health Board, Swansea, UK

Late-stage diagnosis contributes to the poor survival statistics for oesophageal adenocarcinoma (OAC) and better diagnostic tools are urgently required to improve patient outcomes. Our research group has an interest in the development of blood-based biomarkers for early detection. We present data here on the lymphocyte cytokinesis block micronucleus (L-CBMN) assay as a liquid biopsy for OAC.

Blood samples were obtained from 120 participants including healthy volunteers (n=18), patients with gastro-oesophageal reflux disease (GORD) (n=44), Barrett's oesophagus (BO) (n=24) and OAC (N=34). Lymphocytes were isolated and stimulated to divide in culture using phytohemagglutinin. After the addition of cytochalasin B, 1000 binucleate cells were scored in triplicate for the presence of micronuclei (MN).

OAC patients had significantly elevated micronucleus frequencies (MN%) compared to BO patients, GORD patients and healthy volunteers ($p<0.001$) with an average MN% of 1.725% (95% CI 1.0-2.152). The average MN% for healthy volunteers, GORD and BO patients was 0.474% (95% CI 0.267-0.637), 0.85% (95% CI 0.567-1.016) and 0.738% (95% CI 0.533-1.629) respectively. In the non-cancer group (n=86), there was no difference in MN% between males and females ($p=0.138$). There was a positive correlation between MN% and increasing age ($Rho=0.307$, $p=0.005$). Cancer patients had more micronuclei containing whole chromosomes, suggesting an aneugenic mechanism of DNA damage. Furthermore, treatment of TK6 cells with cancer patient plasma showed an increase in MN% compared with plasma from non-cancer controls, suggesting a mechanism of blood-borne genotoxic exposures.

Promising results presented here suggest the L-CBMN assay could be useful as a surrogate marker for OAC diagnosis with MN formation potentially being driven through inflammation. However, further assay development and elucidation of the exact mechanism behind elevated MN% is required to facilitate the acceptance of this biomarker.

The effect of weight-loss on the colorectal transcriptome and its relation to colorectal cancer risk

Emma Hazelwood [1,2], Aayah Nounu [1,2], Vivian Viallon [3], Rebecca J. Beeken [4], Neil Murphy [3], Jane C. Figueiredo [5,6], Graham G. Giles [7,8,9], Mark A. Jenkins [8], Sun-Seog Kweon [10,11], Loic Le Marchand [12], Li Li [13], Rish K. Pai [14], Elizabeth A. Platz [15], Franzel J. B. van Duijnhoven [16], Marc J. Gunter [3*], Emma E. Vincent [1,2,17*] * Marc J. Gunter and Emma E. Vincent contributed equally and are co-senior authors on this work

1. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
2. Bristol Medical School, University of Bristol, Bristol, UK
3. Nutrition and Metabolism Branch, International Agency for Research on Cancer, Lyon, France
4. Leeds Institute of Health Sciences, University of Leeds, Leeds, UK
5. Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA
6. Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, California, USA
7. Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia
8. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
9. Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia
10. Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, Korea
11. Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Hwasun, Korea
12. University of Hawaii Cancer Center, Honolulu, Hawaii, USA
13. Department of Family Medicine, University of Virginia, Charlottesville, Virginia, USA
14. Department of Laboratory Medicine and Pathology, Mayo Clinic Arizona, Scottsdale, Arizona, USA
15. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
16. Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, The Netherlands
17. School of Translational Health Sciences, University of Bristol, Bristol, UK

Colorectal cancer is the second most common cause of cancer-related death. Obesity is a modifiable risk factor for this disease, and weight-loss has been shown to protect from colorectal carcinogenesis. However, the mechanisms underpinning this effect remain largely unknown. Recent research has revealed that weight-loss alters tumourigenic biomarker levels within colorectal tissue, highlighting the importance of tissue-specific effects of weight-loss in relation to cancer risk. However, further research is needed to characterise how adiposity influences colorectal tissue biology, and to investigate how these changes may in turn impact colorectal cancer susceptibility.

In order to identify genes with colorectal expression which is altered by weight-loss, we analysed gene expression microarray data from colorectal biopsies performed in eight individuals before and after a low-energy diet intervention. We then performed Mendelian randomization, a genetic epidemiological approach which uses genetic instruments to estimate causal relationships between traits. This evaluated the potential causal effect of the gene expression changes identified on colorectal cancer risk (GECCO/CCFR/CORECT consortium; 52,775 cases, 45,940 controls). Subsequently, for one gene identified, we conducted lab-based analyses to investigate the mechanism by which changes to expression could cause carcinogenesis.

We identified 288 genes which were differentially expressed following the weight-loss intervention ($P < 9.34 \times 10\text{-}3$, fold change > 1.4). We found evidence for a causal effect of expression of five of these genes on colorectal cancer risk (5% FDR-$P < 0.05$). For one gene, ABHD11, we confirmed that knock-down in two colorectal cancer cell lines resulted in changes to cellular stemness, potentially mediated through α-ketoglutarate dehydrogenase complex activity. These results provide suggestive evidence that weight-loss alters colorectal-specific expression of ABHD11, which may then influence carcinogenesis through the identified mechanism. Our multidisciplinary analysis provides insight into the colorectal-specific effects of weight-loss and highlights ABHD11 as a potential target for colorectal cancer prevention.

Generation Scotland - Linking Health Records for Research

Archie Campbell [1], Robin Flaig [1], Cathie Sudlow [1, 2]

1. University of Edinburgh
2. BHF Data Science Centre, HDR UK, London

Generation Scotland is a family-based genetic epidemiology study of ~24,000 volunteers from ~7000 families recruited across Scotland 2006-2011 with the capacity for follow-up through record linkage and re-contact. Broad consent was obtained for linkage to "medical records" for 98% of the cohort.

Participants completed a demographic, health and lifestyle questionnaire, provided samples, and underwent clinical assessments. The samples, phenotype and genotype data collected form a resource with broad consent for research on the genetics of conditions of public health importance.

This has become a longitudinal dataset by linkage to routine NHS hospital, maternity, laboratory, prescriptions, dentistry, mortality, cancer screening, GP data records, Covid-19 testing and vaccinations. GWAS has been done on quantitative traits and biomarkers, with DNA methylation data and proteomics available for most of the cohort. "CovidLife" surveys collected data on effects of the pandemic.
Researchers can use the linked datasets to find prevalent and incident disease cases, and healthy controls, to test research hypotheses on a stratified population. They can target recruitment of participants to new studies, including recall by genotype, utilising the NHS Scotland CHI Register for current contact details. We have established and validated linkage, overcoming technical and governance issues in the process. Using consented data avoids some limitations of safe havens for analysis. Generation Scotland is a contributor to international consortia, with collaborators from many institutions worldwide, both academic and commercial. New recruits are asked to give consent to linkage to other administrative data, and reuse of samples from routine NHS tests for medical research.
Generation Scotland will extend the linkage process to include scanned images and some administrative data. We are reopening recruitment to double the size of the cohort, including teenagers from age 12+, collecting new data online.

The resources are available to academic and commercial researchers through a managed access process (www.generationscotland.org).

The UK Longitudinal Linkage Collaboration: a trusted research environment for the longitudinal research community

Robin Flaig [1], Jacqueline Oakley [2], Kirsteen Campbell [1], Katharine Evans [2], Stela McLachlan [1], Rich Thomas [2], Emma Turner [2], Andy Boyd [2]

1 UK Longitudinal Linkage Collaboration, Usher Institute, University of Edinburgh
2 UK Longitudinal Linkage Collaboration, Population Health Sciences, University of Bristol

The COVID-19 Longitudinal Health and Wellbeing National Core Study utilises data from Longitudinal Population Studies (LPS) and whole-population NHS databases: yet no resource existed linking multiple LPS to COVID-19 records. A centralised infrastructure was needed to pool LPS data and systematically link participants' routine health, administrative and environmental records. The UK Longitudinal Linkage Collaboration (UK LLC) is a new, unprecedented infrastructure enabling research into the COVID-19 pandemic. The UK LLC integrates data from >20 UK longitudinal studies with systematically linked health, administrative and environmental records to facilitate cross-disciplinaryCOVID-19 research for accredited UK based researchers. Integrated and curated data are made available for pooled analysis within a functionally anonymous Trusted Research Environment (TRE). We commissioned a "Secure eResearch Platform-SeRP" to provide an underlying secure computing system including trusted third-party processing of participant identifiers. We developed a bespoke governance and data curation framework designed collaboratively with LPS data managers and with an access mechanism enabling LPS to retain key decision-making controls. Public/participant involvement guided our approach. Our pipeline/resource/provision of linked data is based on three key innovations: (1) a "Linkage brokerage" model where our trusted third-party processes participant identifiers for many different data owners; (2) a novel longitudinal data pipeline with NHS records which enables linkage and data extraction and update of records over time; (3) a "Delegated & Distributed Access Framework" enabling UKLLC Data Access Panel to consider applications on behalf of data owners (e.g., the NHS) whilst distributing applications to LPS for approval of appropriate data use. UKLLC provides a strategic research-ready platform for longitudinal research: a clear step change from pre-pandemic capability. With sustained investment, and through exploring options to extend linkages and generalise to wider purposes, UKLLC is positioned to inform cross-cutting themes such as understanding health and social inequalities, health-social-environmental interactions, and managing the COVID-19 recovery.

Born in Scotland in the 2020s Pilot Study

Rebecca M Reynolds [1] , on behalf of the Born in Scotland Study Team [1-9]

1. University of Edinburgh, Edinburgh, UK
2. University of Glasgow, Glasgow, UK
3. University of Cambridge, Cambridge, UK
4. Imperial College London, London, UK,
5. University of Bristol, Bristol, UK
6. University of Stirling, Stirling, UK
7. University of Aberdeen, Aberdeen, UK.
8. King's College London, London, UK
9. University of Warwick, Warwick, UK

Pregnancy offers a unique opportunity to positively impact on short- and long-term health outcomes of mother and baby and to influence the next generation's health. With Medical Research Council funding, we are conducting a pilot study in Edinburgh, Glasgow and the Borders to demonstrate the feasibility of establishing a large population study of 100,000 pregnant women and their children, recruited from across the whole of Scotland. "Born in Scotland in the 2020s" will include extensive data linkage from mothers, fathers and children linked to a bioresource. We are using an efficient study design, capitalising on routine data and biological sample collection, avoiding the need for additional participant visits

We are using routine clinical and health service data to record details of:
a) socio-demographic, medical and obstetric history,
b) current pregnancy details including laboratory results, fetal ultrasound scans, pregnancy complications,
c) birth outcomes,
d) healthcare usage,

We are storing and banking biosamples, utilising the Scottish National Health Services Research Scotland Biorepositories, enabling storage of surplus samples which would otherwise be discarded after clinical use.

We are nesting additional studies within the cohort focused on collection of personal data from wearables, pollution exposure, birth samples, DNA analysis and alternative data governance models using a Data Trust.

We plan long-term follow-up through prospective and retrospective record linkage of participants using the Scottish Community Health Index (CHI) unique personal identifier to relevant health and healthcare databases, for example primary and secondary care and educational attainment.

Our multidisciplinary team has expertise in epidemiology, birth cohorts, life-course health, obstetric medicine, neonatology, biobanking, data science, clinical trials, public health and patient & public involvement. The cohort will be a flexible, dynamic and sustainable resource to address fundamental and contemporaneous questions in our understanding of the drivers of long-term maternal and child health, including pregnancy and its complications.

Session Two

Long-term metabolomic effects of bariatric surgery

Madeleine L. Smith [1,2], Lucy J. Goudswaard [1,2], The By-Band-Sleeve Trial Management Group, Laura J. Corbin [1,2], Nicholas J. Timpson [1,2]

1. Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK
2. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

Determining the biological mechanisms that underpin the increased risk of disease associated with variation in body mass index (BMI) and obesity is challenging. These related exposures are associated with a wide range of confounding factors and at the level of the population we have been limited in our ability to alter them. Collecting metabolomics data in studies of weight loss interventions presents an opportunity to describe the biological response to weight loss. Using blood samples collected before and after an intervention, we can assess the impact of weight loss on the metabolome - a read-out of cellular activity encompassing genetic, lifestyle, environmental, microbial and pharmacological influences. Here we use metabolomic data from samples collected before and on average 33 months after bariatric surgery; a subsample of participants from the By-Band-Sleeve trial. Serum samples were analyzed using mass spectrometry (UPLC-MS/MS). We applied a linear mixed model to metabolomic data from 250 paired samples to compare levels of 960 metabolites pre-surgery to post-surgery. 41 metabolites showed evidence of change (Holm-corrected p-value<0.05). Thirty-four metabolites had a higher abundance post-surgery compared to pre-surgery (i.e., they increased post-surgery). The majority (59%) of the altered metabolites belonged to the lipid super-pathway. The strongest association with timepoint was observed for 1-oleoyl-GPC (18:1), a lysophospholipid (beta: 0.28, SE: 0.04, Holm-corrected p-value: 4.16 x 10-10). Xenobiotic metabolites (e.g., drugs) with high missingness (>20%) were transformed to presence/absence data and tested in a generalized linear mixed model. Eight of these 133 xenobiotics were found to be altered post-surgery, including metformin and fluoxetine which were depleted post-surgery, and two tobacco metabolites which were more common in post-surgery samples. These results highlight the utility of metabolomics to measure behavioural and medical effects of surgery as well as cellular effects. Future analyses will include comparisons with metabolomic signatures of non-surgical weight loss interventions.

Augmenting clinical risk prediction of cardiovascular disease through protein and epigenetic biomarkers

Aleksandra D Chybowska [1]

1. Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, EH4 2XU, UK

Cardiovascular disease (CVD) is the leading cause of global mortality. Therefore, it is important to identify CVD risk as early as possible and to translate this knowledge into effective preventative strategies. Here, ASSIGN – a cardiovascular risk calculator recommended for use in Scotland – was examined in tandem with epigenetic and proteomic features in risk prediction models in 12,245 participants in the Generation Scotland cohort. Previously generated DNA methylation-derived epigenetic scores (EpiScores) for 109 protein levels were considered, in addition to both measured levels and an EpiScore for cardiac troponins. There were 42 protein EpiScores that associated with incident CVD (ncases=1,245) independently of ASSIGN (PBonferroni<0.05), over a follow up of up to 15 years of electronic health record linkage. Splitting the cohort into independent training (n=6,643) and testing (n=3,572) subsets composed of unrelated individuals, two composite scores were developed. The CVD ProteinScore (based on the concentration of two cardiac troponins) and CVD EpiScore (based on 50 protein EpiScores) were both associated with CVD risk in Cox models in a test set (Hazard Ratio HR=1.20 and 1.21, P=6.9x10-3 and P=3.2x10-3, respectively). EpiScores for circulating protein levels as well as measured levels of cardiac troponin T and cardiac troponin I have the potential to improve the prediction of CVD and be useful tools for precision medicine.

Session Three

Phenome-wide association study of genetically determined B vitamins and homocysteine biomarkers with multiple health and disease outcomes: Analysis of the UK Biobank

Lijuan Wang [1], Xue Li [1], Azita Montazeri [3], Amanda J. MacFarlane [4], Franco Momoli [3], Susan Duthie [5], Marjanne Senekal [6], Ines Mesa Eguiagaray [2], Ron Munger [7], Derrick Bennett [8], Harry Campbell [2], Michele Rubini [9], Helene McNulty [10], Julian Little [3], Evropi Theodoratou [2,11]

1. School of Public Health and the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
2. Centre for Global Health, Usher Institute, The University of Edinburgh, Edinburgh, UK
3. School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
4. Nutrition Research Division, Health Canada, Ottawa, Ontario, Canada
5. School of Pharmacy and Life Sciences, Robert Gordon University, Aberdeen, UK
6. Department of Human Biology, University of Cape Town, Cape Town, South Africa
7. Department of Nutrition and Food Sciences and the Center for Epidemiologic Studies, Utah State University, Logan, USA
8. Medical Research Council Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK
9. Department of Neuroscience and rehabilitation, University of Ferrara, Ferrara, Italy
10. Nutrition Innovation Centre for Food and Health, Ulster University, Coleraine, Northern Ireland, UK
11. Cancer Research UK Edinburgh Centre, The University of Edinburgh MRC Institute of Genetics and Cancer, Edinburgh, UK

Background: B vitamins are known to be important in many human health outcomes. However, evidence is of uneven quality and volume across outcomes, and there is uncertainty about putative causal relationships.
Objectives: To explore health outcomes in relation to B vitamins and homocysteine based on a large biorepository linking biological samples and electronic medical records.
Methods: First, we performed a phenome-wide association study (PheWAS) to investigate associations of genetically determined plasma concentrations of folate, vitamin B6, vitamin B12 and their metabolite homocysteine with a wide range of disease outcomes in the UK Biobank. Second, we conducted a two-sample Mendelian randomization (MR) analysis based on the inverse variance weighted (IVW) method to triangulate any observed associations. Third, dose-response, mediation and bioinformatics analyses were carried out to examine any linear or non-linear trends and to disentangle the underlying mediating pathophysiological processes for the identified associations.
Results: We identified 33 significant phenotypic associations of B vitamins and homocysteine. Six of them were successfully validated in the two-sample MR analyses, including associations of higher plasma vitamin B6 with lower risk of calculus of kidney (OR=0.64, 95% CI: 0.42-0.97, P=0.033); higher serum vitamin B12 with higher risk of hypertension (OR=1.11, 95% CI: 1.04-1.18, P=0.001), essential hypertension (OR=1.13, 95% CI: 1.06-1.21, P=2.79×10-4), and colorectal cancer (OR=1.13, 95% CI: 1.04-1.67, P=0.024); and higher homocysteine concentration with higher risk of hypercholesterolemia (OR=1.28, 95% CI: 1.04-1.56, P=0.018) and chronic kidney disease (OR=1.32, 95% CI: 1.06-1.63, P=0.012). Significant dose-response relationships were observed for the associations between vitamin B12 and hypertension, and homocysteine and cerebrovascular disease.
Conclusions: This study provides strong evidence for the associations of B vitamins and homocysteine with multiple disease outcomes, including neoplasm, endocrine/metabolic, circulatory and genitourinary disorders.

Using PathWAS associated with proteomics to predict pathway function and relationships with complex traits

Sebastian May-Wilson [1], Linda Repetto [1], Erin Macdonald-Dunlop [2], SCALLOP Consortium, Jim F. Wilson [1], Nicola Pirastu [3]

1. University of Edinburgh, Edinburgh, UK
2. Imperial College London, London, UK
3. Fondazione Human Technopole, Milan, Italy

Association studies are a staple in the analysis of complex traits, used to elucidate phenotype relationships from GWAS loci. These methods however only focus on individual loci or genes in isolation, while it may be more advantageous to instead examine loci in the context of broader biological pathways. Our aim was to create polygenic scores for biological pathway functionality and then use these scores to search for phenotype associations. By combining multiple genes into one network score, we hypothesise that we gain greater insight into the aetiology of complex traits while also aiding in improving power of discovery of novel associations.

Our methodology involves the stratification of transcribed genes by biological pathways, based on curated databases. We use cis-eQTL data from the eQTLgen consortium for the pathway genes in a Mendelian randomisation against specific end-point proteins from the DeCODE proteomics data. In doing so we weight the effect of each gene against a given protein at the end of the pathway, therefore the protein measurement acts as a proxy for the functionality of the pathway. The weights are combined with polygenic scores (PRS) for each gene to create overall PRS for each pathway which were subsequently used in a PheWAS analysis in UK Biobank.

We were able to create over 500 pathway and proteomic end-point combinations and from these identified thousands of significant associations with complex traits. This includes examples such as a pathway for pyruvate metabolism associated with increased risk of heart attacks and the pathway for insulin secretion associated with increased BMI.

Overall, we believe this demonstrates a proof-of-concept for our methodology in finding associations between biological pathways and complex phenotypes and we propose that this methodology may be used to complement GWAS to aid in the discovery of further pathway associations, a vital aspect in precision medicine.

Poster Presentations

The comet assay; a predictive biomarker

Andrew Collins[1]

1. University of Oslo (on behalf of the hCOMET team)

Best known for its widespread use in genotoxicity testing, the comet assay has also been applied since the early days as a tool in human biomonitoring. Modifications of the standard assay have been particularly valuable in this field: the inclusion of a digestion with lesion-specific enzymes has provided information on DNA oxidation damage and antioxidant protection; the use of mini-gels has allowed many more samples to be assayed in one experiment; and the in vitro DNA repair assay has revealed the extent of individual variation in this important biomarker. In the recent European COST Action, hCOMET, we created a database of almost 20,000 individual comet assay datasets and carried out pooled analyses. In several studies, the health of subjects was monitored for years after the DNA damage measurement, and it was possible to look for a correlation between DNA damage and mortality. We have successfully isolated white cells from stored 'buffy coat' samples (commonly stored in biobanks) and so it should be possible to carry out nested case-control studies within epidemiological trials (such as EPIC), to investigate the links between DNA damage (and possibly repair) and risk of cancer and other diseases.

Circulating white blood cell count and colorectal cancer risk: a Mendelian Randomization study

Andrei-Emil Constantinescu [1,2], Caroline J. Bull [1,2], Jeroen R Huyghe [3], Marc J. Gunter [4], Neil Murphy [4], Nicholas J. Timpson [1], Emma E. Vincent [1,2]

1. Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, United Kingdom
2. School of Translational Health Sciences, University of Bristol, Bristol, United Kingdom
3. Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
4. Nutrition and Metabolism Branch, International Agency for Research on Cancer, World Health Organization, Lyon, France

Colorectal cancer (CRC) is one of the most common cancers in the UK and accounts for around 10% of cancer deaths worldwide. Previous observational studies have suggested a role for white blood cell (WBC) subtypes in CRC risk and mortality. Here, we aimed to investigate the effect of circulating WBC count on CRC risk using Mendelian randomization (MR) and observational methods. For the MR, genome-wide association study summary statistics for WBCs were accessed from a comprehensive meta-analysis (N=562,132 Europeans), and for CRC overall and by site (colon, proximal colon, distal colon and rectal) through a large meta-analysis (58,221 cases and 67,694 controls in the Genetics and Epidemiology of Colorectal Cancer Consortium, Colorectal Cancer Transdisciplinary Study, and Colon Cancer Family Registry). For the observational study, individual-level UK Biobank data for 4,043 incident CRC cases and 332,773 controls were available. Univariable (UV) and multivariable (MV) analyses were done to assess the effect of WBCs on CRC risk. The inverse-variance weighted (IVW) UVMR analysis showed evidence of an effect on CRC risk for basophil count (overall CRC - OR: 0.88, CI(95%): 0.78-0.99, P=0.04; observational OR: 1.06, CI(95%): 1.02-1.09, P=4.46e-4) and eosinophil count (overall CRC - OR: 0.93, CI(95%): 0.88-0.98, P=0.01; observational OR: 0.97, CI(95%): 0.94-1.00, P=0.09). Additional sensitivity UVMR analyses (MR-PRESSO, Cochran's Q test, MR-Egger) were done. The IVW MVMR method (adjusting for overall WBC count) provided evidence of a protective effect of eosinophil count (Overall CRC - OR: 0.93, CI(95%): 0.87-0.99, P=0.04; observational OR: 0.96, CI(95%): 0.93-0.99, P=4.81e-3), and a detrimental effect of neutrophil count (Overall CRC - OR: 1.51, CI(95%): 1.18-1.93, P=8.66e-4; observational OR: 1.05, CI(95%): 1.02-1.08, P=2.16e-3) on overall CRC risk, both which were consistent with existing observational evidence. Our study provides evidence that circulating immune cells play a role in CRC aetiology, laying the path for future mechanistic studies.

Effects of dietary Polyunsaturated Fatty Acids on inflammation: a Mendelian randomization study

Daisy C P Crick [1,2], Abigail Fraser [1,2], Hannah Jones [1,2]

1. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.
2. MRC Integrative Epidemiology Unit, University of Bristol. Bristol, UK.

Inflammation can be affected by dietary factors, including polyunsaturated fatty acids (PUFAs). It is believed that omega-3 PUFAs are anti-inflammatory, while omega-6 PUFAs are pro-inflammatory. This has led to diets high in omega-3 (e.g. fish), such as the Mediterranean diet, being recommended. We aimed to investigate the causal effect of total omega-3 (n-3), total omega-6 (n-6), linoleic acid (LA) and docosahexaenoic acid (DHA) on levels of three inflammatory biomarkers: C-reactive protein (CRP), Interleukin-6 (IL-6) and Glycoprotein Acetyls (GlycA).

Causal effects were evaluated using genetic variant-exposure and genetic variant-outcome information from published PUFA and inflammatory biomarker genome-wide association studies and the inverse variance weighted (IVW) two-sample Mendelian randomization (MR) method. Sensitivity analyses included MR-Egger, weighted-median, and weighted-mode. MR-Clust was used to explore whether heterogeneity in individual genetic variant causal effects were due to underlying clusters of similar causal estimates.

We found evidence of positive effects of n-3 (βIVW:0.47; 95% CI:0.31, 0.64), n-6 (βIVW:0.27;0.15, 0.39), DHA (βIVW:0.31; 0.10, 0.52) and LA (βIVW:0.21; 95% CI:0.07,0.35) on GlycA levels. There was evidence of a positive effects of n-6 (βIVW:0.12; 0.03, 0.22) and LA (βIVW:0.12; 0.03, 0.20) on IL-6. However, results were not consistent across sensitivity analyses. We found no effect of DHA or n-3 on IL-6 or of any PUFA on CRP. MR-Clust identified distinct clusters of causal effects within the GlycA-analyses.

Our IVW estimates suggest that both n-3 and n-6 have pro-inflammatory effects; increasing GlycA and IL-6 levels. This is in contrast with evidence suggesting that n-3 are anti-inflammatory. However the lack of consistency across the sensitivity analyses highlights there is still uncertainty about their causal effect and triangulation through the use of addition study types is needed for causal inference. Future work determining the pro-inflammatory effect of dietary PUFAs may inform treatment strategies of chronic inflammatory diseases.

Association between predicted methylation quantitative trait loci scores and colorectal cancer in three case-control studies of colorectal cancer in Scotland

Ines Mesa Eguiagaray [1], Xue Li [2], Andrii Iakovliev [3], Athina Spiliopoulou [3], Malcolm Dunlop [4], Evropi Theodoratou [1,4]

1 Centre for Global Health, Usher Institute, University of Edinburgh, Edinburgh, UK.
2 School of Public Health and the Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China.
3 Centre for Population Health Sciences, Usher Institute, University of Edinburgh, Edinburgh, Scotland.
4 Cancer Research UK Edinburgh Centre, MRC Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK.

Background: To date, most of the genetic variants identified to be associated with CRC risk have not been associated with CRC survival or recurrence. Given that DNA methylation plays an important role in the regulation of gene expression and that it has a genetic component, we hypothesise that investigating methylation loci associated not only with CRC risk but also with survival and recurrence could shed some light into disease aetiology and help identify biomarkers for prevention, therapeutics and prognosis. Methods: Using individual data from three CRC case-control studies in Scotland with 6,821 CRC cases and 14,692 controls we derived methylation quantitative trait loci (mQTLs) based on the largest meta- analysis of DNA methylation from the Genetics of DNA Methylation Consortium (GoDMC). Association analysis between mQTLs and CRC risk, survival (overall and CRC- specific) and recurrence was performed using logistic and Cox regression models and adjusted for age, sex, stage and ten principal components. Date and cause of death were ascertained from the Scottish Cancer Registry and recurrence from the South East Scotland Database. Multiplicity was corrected using Benjamini-Hochberg method for a False Discovery Rate (FDR) of 5%. Results: We identified 19 mQTLs in 10 distinct regions associated with CRC risk. We further investigated the association of those 19 mQTLs and over 2,000 additional mQTLs in previously identified genes associated with CRC risk or survival (from GWAS studies), with survival and recurrence. None of the associations with survival and recurrence were significant after FDR correction. Conclusion: DNA methylation might be one of the mechanisms influencing CRC risk; however, within the limitations of this study (event rate and unmeasured confounders) we found no evidence of its contribution to survival or recurrence. Further research is needed to identify biomarkers for CRC prognosis.

Exploring DNAm proxies of smoking: characteristics and context

Chloë Fabbricatore [1], Ryan Langdon [2]

1. University of St Andrews, St Andrews, UK
2. University of Bristol, Bristol, UK

DNA methylation (DNAm) shows great promise in predicting smoking behaviour. However, most existing studies examine current/never or ever/never smokers and do not compare DNAm to other biomarkers of smoking. Additionally, there has been little appraisal of the characteristics of smoking DNAm proxies capture. Our aims were: 1) appraise discriminative performance of DNAm for different classes of smoking vs cotinine and cadmium 2) investigate which aspects of smoking DNAm is associated with.

Our data comprised 346 pregnant women in ARIES. For current/former, current/non, ever/never and former/never smoking, we compared AUCs (via DeLong Z-test) between cotinine, cadmium and DNAm. DNAm included cg05575921 (AHRR), cg05951221 (ALPPL2) and a weighted score derived by Maas et al, respectively. We performed gene set enrichment analysis (GSEA) of 9 unique genes from the Maas score.

For ever/never and former/never smoking, all DNAm proxies outperformed cotinine ($3.98\times10^{-4}<P<0.02$). For current/former ($P:2.1\times10^{-3}$) and current/non-smoking ($P:8.8\times10^{-3}$), cotinine outperformed cg05951221 (ALPPL2) but not cg05575921 (AHRR) or the Maas score. Interestingly, there was no statistical difference seen between urinary cadmium and DNAm. GSEA returned abnormal B cell count as the top Human Ontology and vitamin D biosynthesis as the top biological process.

Cotinine's short biological half-life (~12hrs) likely causes former to be indistinguishable from never smokers. Contrarywise, DNAm scores contains multiple probes with different half-lives, potentially discriminating former and never smokers better as a result. GSEA findings can largely be substantiated by literature – e.g. smoking-related COPD via B cell impairment and lower vitamin D in smokers vs non-smokers.

Our study compares DNAm, cotinine and cadmium in the same dataset. DNAm performed excellently at discriminating between multiple combinations of smoking classes, statistically outperforming cotinine for ever/never and former/never smoking. DNAm gene regions appear to enrich for smoking-relevant ontologies and biological processes, perhaps highlighting why it explains greater variance than self-report for multiple health outcomes.

Characterising the circulating proteome of adiposity through use of weight loss interventions and Mendelian randomisation

Lucy J. Goudswaard [1], Madeleine L. Smith [1], David A. Hughes [1], Naveed Sattar [2], Roy Taylor [3], Michael E. J. Lean [2], NIHR By-Band-Sleeve Trial Management Group, Jane M. Blazeby [1], Ingeborg Hers [1], Laura J. Corbin [1], Nicholas J. Timpson [1]

1 University of Bristol, Bristol, UK
2 University of Glasgow, Glasgow, UK
3 Newcastle University, Newcastle-upon-Tyne, UK

Adiposity is associated with an increased risk of type 2 diabetes (T2D) and cardiovascular disease. It is likely that a change in circulating proteins plays a role in obesity-related disease risk. This study aimed to triangulate evidence from randomized controlled trials of caloric restriction and bariatric surgery, along with a Mendelian randomisation (MR) study, to characterise the effects of BMI on proteins in circulation. Data were used from the intervention arm of the Diabetes Remission Clinical Trial (DiRECT, N=119). Here, participants with overweight/obesity (baseline mean BMI 35.0 kg/m2, SD 4.6 kg/m2) underwent the CounterWeight Plus programme, consisting of a low energy total diet replacement (TDR). Plasma samples were taken at baseline and after 1-year and SomaLogic was used to quantify plasma proteins. Data were also utilised from a pilot sample release of an ongoing surgically induced weight loss trial (N=118). Serum samples were donated at baseline and 3-years post-surgery; proteins were measured by the Olink Explore 1536 panel. Linear mixed models were used to determine proteins associated with the weight loss interventions. Proteins with consistent evidence across both trials were examined in our previously published MR analysis, which used a UK blood donor cohort, INTERVAL (N=2729), to estimate the causal effect of BMI on circulating plasma proteins. 1496 proteins out of 4601 were altered with the TDR intervention ($p<5.8x10-5$) and 188 proteins out of 1472 were associated with bariatric surgery ($p<6.2x10-5$). 41 unique proteins were altered in the same direction with both interventions, of which 11 also had consistent MR effect estimates. Interventions reduced levels of alcohol dehydrogenase 4, glutathione S-transferase A1 and mannan-binding lectin serine protease 1, with higher BMI being associated with higher levels in an MR framework. Observing consistency across such analyses provides greater confidence that the effects we see reflect genuine physiological responses to differential BMI.

Maternal plant-based diets and neonatal DNA methylation: an epigenome-wide association study

Peiyuan Huang [1], Giulia Mancano [1], Carolina Borges [1], Kate Northstone [1], Gemma Sharp [1,2]

1. University of Bristol, Bristol, UK
2. University of Exeter, Exeter, UK

Objectives: To examine cord blood DNA methylation (DNAm) in relation to maternal adherence to plant-based diets (PBDs) during pregnancy.

Methods: 693 British mother-child pairs with both pregnancy food frequency questionnaire and cord blood DNAm data were included. PBD adherence was assessed using 3 indices: an overall plant-based diet index (PDI), with higher scores representing greater plant food intake; a healthful (hPDI) and unhealthful PDI (uPDI) further distinguished between intakes of healthy and unhealthy plant foods. DNAm was measured by Illumina 450K arrays. Linear regression adjusted for maternal age, education, parity, child sex, cell types, and batch (Model 1). Model 2 further adjusted for maternal lifestyle factors. We excluded polymorphic sites and performed functional analyses and lookups.

Results: At p-value <1×10-5, there were 3 (cg13180232, cg24265806, cg04896381), 1 (cg11760198), and 2 (cg22254580, cg24810917) differentially methylated cytosine-phosphate-guanine (CpG) sites identified in Model 1 in relation to PDI, hPDI, and uPDI, respectively; only cg13180232 passed the threshold for false discovery rate (FDR) correction (FDR-corrected p=0.049). Associations were slightly strengthened in Model 2 for most CpG sites. Neither model identified any overlap between top CpGs for the 3 indices. In EWAS Catalog, adult rheumatoid arthritis and maternal glucose levels have previously been associated with cg04896381 (in TBKBP1) in whole and cord blood, respectively, and clear cell renal carcinoma with cg24810917 (in ODZ4) in adults. cg22254580 (in GOLGA3) was associated with gene expression of ANKLE2, CHFR, ZNF605, and ZNF84 in a publicly available cis-expression quantitative trait methylation database from child blood. Based on our results, we identified 1 differentially methylated region associated with PDI on chromosome 3:128968351-128968543 (in CFAP92, β=-2.55×10-3, SE=4.14×10-4, p=7.09×10-10). No firm evidence of enrichment was found for gene ontology, pathways, or tissue-specific DNaseI hypersensitivity regions.

Conclusions: Adherence to different PBDs during pregnancy was associated with differential cord blood DNAm patterns.

Plasma metabolomic profiles of carotid atherosclerosis in two Scottish populations.

Zhe Huang [1], Lucija Klaric [2], Jim F Wilson [1,2], Jackie F Price [1]

1. Usher Institute, University of Edinburgh, Edinburgh, United Kingdom.
2. MRC Institute of Genetics and Cancer, University of Edinburgh, Western General Hospital, Edinburgh, United Kingdom.

Aims: To describe the association between plasma metabolites and atherosclerosis in carotid arteries in two Scottish populations.

Methods: Nuclear magnetic resonance spectroscopy was used to quantify the metabolomic profile of 1,785 participants from the Orkney Complex Disease Study (ORCADES), with replication in 2,029 participants from VIKING Health Study-Shetland. Carotid atherosclerosis was assessed by carotid intima-media thickness (cIMT) using ultrasonography. Least absolute shrinkage and selection operator (LASSO) was applied to select metabolites associated with cIMT following adjustment for age, sex, smoking, systolic blood pressure, HDL-c, total cholesterol and BMI. Additionally, the association between each individual metabolite and cIMT was assessed using univariate analysis following adjustment for covariates.

Results: Mean (±SD) cIMT in ORCADES and VIKING was 0.57 (±0.14) mm and 0.47 (±0.11) mm respectively. In LASSO analyses, among 223 metabolites, 14 metabolites were associated with cIMT in ORCADES, mainly lipid ratios and amino acids. Only histidine (β = 0.001) was replicated in VIKING (β = 0.005). While glutamine was identified in both ORCADES and VIKING, the direction of association between it and cIMT was opposing (β = 0.0021 vs. -0.0001). Similarly, in univariate analysis, inconsistent associations between glutamine and cIMT were observed in ORCADES [β = 0.0055 (95%CI -0.0002, 0.1125)] and VIKING [β = -0.002 (95%CI -0.006, 0.002)], despite being non-significant in statistics (P>0.05). Histidine showed positive association with cIMT in VIKING in univariate analysis [β = 0.007 (95%CI 0.003, 0.011), P=0.0005], although it became non-significant after Bonferroni correction.

Conclusions: Plasma histidine was associated with cIMT in two Scottish populations. Further investigation is warranted to determine its potential causal or predictive role for carotid atherosclerosis.

Novel pQTL discovered by proteomic genome-wide association of 2000 unstudied proteins

S. Jurgis Kuliesius [1], Lucija Klaric [2], Paul R. H. J. Timmers [1, 2], Pau Navarro [2], James F. Wilson [1, 2]

1. Centre for Global Health Research, Usher Institute, University of Edinburgh
2. MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh

Genome-wide association study (GWAS) power lies in testing hundreds of thousands of single nucleotide polymorphisms (SNPs) across many genomes for associations with a trait. Recent technological developments in high-throughput proteomic assays allow simultaneous quantitative measurement of thousands of proteins in a single sample. Blood plasma proteins effectively provide a glimpse into multiple biological systems. Combining broad-capture proteomics with genomic variation across a population is a valuable resource with implications in elucidating complex traits and disease, drug development or repurposing, and precision medicine. Here, we investigated 6434 plasma proteins using the Somalogic aptamer-based technology in samples from the Viking Health Study - Shetland.

622 significant protein quantitative trait loci (pQTL) were found for 461 proteins in plasma (482 cis ($P < 5e-8$), 140 trans ($P < 6.6e-12$)). Of these, 221 pQTL were for previously unstudied proteins. We leveraged this new resource to perform causal inference using Mendelian randomization against complex traits of biomedical importance.

Dispensing with unnecessary assumptions in population genetics analysis

Olivier Labayle Pabet [1, 2], Kelsey Tetley-Campbell [1, 2], Mark Van Der Laan [4], Chris P. Ponting [1], Sjoerd Viktor Beentjes [1, 3], Ava Khamseh [1, 2]

1. MRC Human Genetics Unit, Institute of Genetics & Cancer, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom.
2. School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, United Kingdom
3. School of Mathematics, University of Edinburgh, Edinburgh EH9 3FD, United Kingdom.
4. Division of Biostatistics, University of California, Berkeley, CA

Parametric assumptions in population genetics analysis – including linearity,
sources of population stratification and the gaussianity and additivity of errors are often made, yet a principled argument for their (approximate) validity is not given. We present a unified statistical workflow, called TarGene, for targeted estimation of effect sizes, as well as two-point and higher-order epistatic interactions of genomic variants on polygenic traits, that dispenses with these unnecessary assumptions. Our approach is founded on Targeted Learning, a framework for estimation that integrates mathematical statistics,
machine learning and causal inference to provide mathematical guarantees and realistic p-values. TarGene defines effect sizes of variants, as well as two-point and higher-order interactions amongst genomic variants on traits in a model-independent manner, thus avoiding all-too-common model-misspecification whilst taking advantage of a library of parametric and state-of-the-art non-parametric algorithms. TarGene data-adaptively incorporates confounders and sources of population stratification, accounts for population dependence structures and controls for multiple hypothesis testing by bounding any desired type I error rate. Extensive simulations demonstrate the necessity of this model-independent approach. We validate the effectiveness of our method by reproducing previously verified effect sizes on UK Biobank data, whilst simultaneously discovering non-linear effect sizes of additional allelic copies on trait or disease. To exemplify this, we demonstrate that for the FTO variant rs1421085 effect size on body mass index (BMI), the addition of one copy of the C allele is associated with 0.77 kg/m2 (95% CI: 0.68 − 0.85) increase, while the addition of the second C copy non-linearly adds 1.31 kg/m2 (95% CI: 1.19 − 1.43) to BMI. TarGene thus extends the reach of current genome-wide association studies by simultaneously (i) allowing for the classification of the types of SNPs and phenotypes for which such non-linearities occur, whilst (ii) data-adaptively incorporating complex non- linear relations between phenotype, genotype, and confounders, as well as (iii) accounting for strong population dependence such as island cohorts. The method provides a platform
for comparative analyses across biobanks, or integration of multiple biobanks and heterogeneous populations to increase power, whilst controlling for population stratification and multiple hypothesis testing.

Application process for UK Longitudinal Linkage Collaboration: a new linked data resource

Stela McLachlan [1], Samantha Berman [2], Katharine Evans [2], Christie Levein [1], Rebecca Whitehorn [1], Jacqueline Oakley [2], Kirsteen Campbell [1], Robin Flaig [1]

1 UK Longitudinal Linkage Collaboration, Usher Institute, University of Edinburgh.
2 UK Longitudinal Linkage Collaboration, Population Health Sciences, University of Bristol.

The UK LLC is an innovative project, developing a new approach to linking 20+ well-established longitudinal studies to routine health and administrative records. Setting up the process included consultation with stakeholders and data community, acknowledging constraints imposed by data controllers and ensuring meaningful public involvement. The initial process had to be flexible because the Trusted Research Environment (TRE) was regularly being updated with new data. This required an adaptive approach where the process was created, adopted, and re-evaluated in real-time. As a result, the UK LLC has a three-stage application and review process, starting with an enquiry through the HDRUK Innovation Gateway. Enquiries are triaged to ensure the applications meet basic requirements (UK-based, COVID-19 related, scientifically sound, and feasible) before the applicants are invited to submit a full application. The full application is then reviewed internally against the Five Safes model, in communication with applicants. Approved applications move on to the third stage which includes a review by Data Access Committees (DACs) for each study whose data is requested, and the UK LLC DAC - comprised of a public review panel advising on lay summaries and public involvement in the proposed project, and, if applicable, an additional linked data review panel on behalf of NHS Digital. Our process is still evolving as new datasets are added and in response to feedback. However, the team has risen to the challenge of balancing this with the needs of 20+ studies, large number of stakeholders, applicants, and participant/public contributors, to create an agile process that can handle complex needs and requirements.

Telomere length in individuals with mood or anxiety disorder

Julian Mutz [1], Cathryn M Lewis [1,2]

1. Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom.
2. Department of Medical and Molecular Genetics, Faculty of Life Sciences & Medicine, King's College London, London, United Kingdom.

Background: Accelerated biological ageing might contribute to the lower life expectancy of individuals with mental disorders. The aim of this study was to examine telomere length, a biological hallmark of ageing, in individuals with and without mental disorders.
Methods: The UK Biobank is a multicentre community-based observational study that recruited >500,000 middle-aged and older adults. Leukocyte telomere length (T/S ratio) was measured using quantitative polymerase chain reaction. Polygenic risk scores (PRS) were calculated for individuals of European ancestry. We estimated group differences in T/S ratio between individuals with anxiety disorder, depression or bipolar disorder and people without mental disorders. We also examined associations with psychotropic medication use, age and PRS for these three disorders.
Results: The analyses included up to 308,725 participants. Individuals with depression had shorter telomeres than people without mental disorders (β = -0.011, 95% CI -0.019 to -0.004, pBonf. = 0.027). Associations between bipolar disorder and telomere length differed by lithium use. There was limited evidence that individuals with anxiety disorder had shorter telomeres. Associations between age and telomere length did not differ between individuals with and without these disorders. PRS for depression, but not anxiety disorder or bipolar disorder, were associated with shorter telomeres (β = -0.006, 95% CI -0.010 to -0.003, pBonf. = 0.001).
Conclusions: Differences in telomere length were observed primarily for individuals with depression or bipolar disorder and in individuals with a higher polygenic risk score for depression. There was no evidence that the association between age and telomere length differed between individuals with and without anxiety disorder, depression or bipolar disorder.

Migration and risk of schizophrenia and bipolar disorder in a Swedish national register study

Natassia Robinson [1], Alexander Ploner [1], Roxana Müller-Eberstein [1], Paul Lichtenstein [1], Kenneth S. Kendler [2] and Sarah E. Bergen [1]

1 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
2 Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia

Background: Many studies have linked migration with increased risk of schizophrenia (SCZ). However, few studies have investigated migration as a risk factor for bipolar disorder (BD), and findings are inconsistent. Demographic characteristics impacting risk have rarely been studied in the context of migration. We investigated the risk of SCZ and BD in migrants and their children compared to those of Swedish ancestry, and whether risk varied by age at migration, region of origin and sex.

Methods: We conducted a population-wide nested case-control study involving 5,220 SCZ cases and 20,107 BD cases diagnosed 1988-2013 in the Swedish National Patient Register. Each case was matched on year of birth and sex to five controls. Conditional logistic regression was used to evaluate the risk of SCZ and BD by migrant status, region of origin, and age at migration, with models also stratified by sex.

Results: First-generation migrants had increased risk of SCZ, and decreased risk of BD. There was a distinct pattern of risk for SCZ by age at migration, which peaked between ages 10-14. In contrast, there was a smooth decline in risk for BD with increasing age at migration. Childhood migrants from all regions had increased risk of SCZ, particularly those from Africa, whilst only child migrants from the Nordic region had significantly higher risk of BD than those of Swedish ancestry. Adult migrants had lower risk of SCZ and BD diagnoses than individuals of Swedish ancestry. Risk was elevated in children of migrants (with risk differing by number of migrant parents) and was generally higher in male migrants (vs. female).

Discussion: Migration significantly impacts risk of SCZ and BD diagnoses, and age at migration, sex, and region of origin affect risk in different ways. Further research is required to determine how migration-related factors influence disease aetiology and the designation of these diagnoses.

Identification of early diagnosis markers of pancreatic ductal adenocarcinoma (PDAC) using publicly available transcriptomic tumour and blood sample data.

Aristeidis Sionakidis [1], Panagiotis-Nikolaos Lalagkas [2], Andigoni Malousi [3], Ioannis S. Vizirianakis [4,5]

1. Institute of Genetics and Cancer, University of Edinburgh, Scotland, UK
2. Department of Biological Sciences, University of Massachusetts, Lowell, Boston, USA
3. Laboratory of Biological Chemistry, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece
4. Laboratory of Pharmacology, School of Pharmacy, Aristotle University of Thessaloniki, Thessaloniki, Greece
5. Department of Life and Health Sciences, University of Nicosia, Nicosia, Cyprus

Introduction: Approximately, 20% of pancreatic ductal adenocarcinoma (PDAC) patients are diagnosed at early stages and less than 7% of patients survive more than 5 years in the UK. We aimed to identify general PDAC disease biomarkers that may improve earlier diagnosis and patient stratification for improved mortality outcomes.
Methods: Publicly available gene expression data from 10 studies with tumour tissue (448 samples) and/or blood samples (128 samples) from PDAC patients prior to treatment were analysed. Validation of markers was performed using Cancer Genome Atlas (TCGA) PDAC expression data. Tissue samples had AJCC (American Joint Committee for Cancer) staging information available. Differential gene expression analysis was carried out to compare tumour and normal samples (stage-specific tissue samples vs. normal tissue samples and PDAC blood samples vs. normal blood samples). Active subnetwork search and miRNA enrichment analysis were used to identify enriched gene networks and miRNA interactions.
Results: We identified 820 consistently deregulated (either up- or down-regulated) genes between tissue samples of all stages and blood samples. These markers were confirmed in TCGA data predicting PDAC outcome (dead/alive status), in the form of custom risk scores. Active subnetwork analysis revealed enriched ribosome, proteasome, adherens junction and cell cycle pathways across all stages and blood samples suggesting biological plausibility. Stage-specific enriched miRNAs were also identified (miR-21, miR-29, miR-124, miR-30, for stages 1-4 respectively).
Discussion: We identified PDAC markers deregulated across all stages and different sample sets. Extensive gene expression deregulation was found in all clinical stages with significant overlap. Additionally, miRNA contribution to PDAC pathology may be important and probably mediated by distinct miRNAs in each stage of PDAC. We therefore present a list of markers and miRNAs that could potentially act as a diagnostic tool for early detection of PDAC onset to be evaluated in other clinical and epidemiologic studies.

Limited effect of Y chromosome variation on coronary artery disease and mortality in UK Biobank

Paul R.H.J. Timmers [1,2], James F. Wilson [1,2]

1. MRC Human Genetics Unit, MRC Institute of Genetics and Cancer, University of Edinburgh, United Kingdom.
2. Centre for Global Health Research, Usher Institute, University of Edinburgh, United Kingdom.

Background: The effect of genetic variation in the male-specific region of the Y chromosome (MSY) on coronary artery disease and cardiovascular risk factors has been disputed. In this study, we systematically assessed the association of MSY genetic variation on these traits using a kin-cohort analysis of family disease history in the largest sample to date.

Methods: We tested 90 MSY haplogroups against coronary artery disease, hypertension, blood pressure, classical lipid levels, and all-cause mortality in up to 152 186 unrelated, genomically British individuals from UK Biobank. Unlike previous studies, we did not adjust for heritable lifestyle factors (to avoid collider bias) and instead adjusted for geographic variables and socioeconomic deprivation, given the link between MSY haplogroups and geography. For family history traits, subject MSY haplogroups were tested against father and mother disease as validation and negative control, respectively.

Results: Our models find little evidence for an effect of any MSY haplogroup on cardiovascular risk in participants. Parental models confirm these findings.

Conclusions: Kin-cohort analysis of the Y chromosome uniquely allows for discoveries in subjects to be validated using family history data. Despite our large sample size, improved models, and parental validation, there is little evidence to suggest cardiovascular risk in UK Biobank is influenced by genetic variation in MSY.

Integrating Mendelian randomization and literature-mined evidence for breast cancer risk factors

Marina Vabistsevits [1,2 *], Tim Robinson [1,2], Ben Elsworth [1,2,3], Yi Liu [1,2], Tom R Gaunt [1,2]

1. Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom
2. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom
3. Our Future Health, Manchester, United Kingdom

Identifying evidence from multiple sources when studying disease risk factors is one of the main challenges in population health research. Biomedical data integration platforms such as EpiGraphDB (epigraphdb.org) can facilitate evidence triangulation from different sources, improving confidence in the causal relationships of interest. In this study, we aimed to integrate Mendelian randomization (MR) and literature-mined evidence from EpiGraphDB to build a comprehensive overview of breast cancer risk factors.

We queried MR-EvE (MR "Everything-vs-Everything") data in EpiGraphDB to generate a list of causal risk factors for breast cancer and extracted literature-mined relationships for these traits to dissect how they may be linked to breast cancer. Integrating these two sources of evidence allowed us to identify potential mediators of the risk factors' effect on breast cancer. We used multivariable MR to separate the direct effects of the traits as a validation step.

We identified 213 lifestyle and molecular traits with evidence of an effect on breast cancer, including traits previously reported in observational and MR studies and novel risk factors. We focused on four risk factor traits as case studies: IGF-1, cardiotrophin-1, childhood obesity, and age at menopause. We reviewed the potential mediators of their effect identified from the MR-EvE data using two-step MR queries and the overlap of literature-mined data between each trait and breast cancer. We determined that the negative effect of cardiotrophin-1 on ER+ breast cancer (OR:0.97[0.95:0.99]) may be mediated via leukaemia inhibitory factor. IGF-1 effect (OR:1.07[1.01:1.13]) may be linked to breast cancer via KIT ligand. None of the detected 38 mediator candidates of childhood obesity protective effect (OR:0.64[0.56:0.73]) explained it.

Our work demonstrates that using MR-EvE to identify disease risk factors is an efficient hypothesis-generating approach. Moreover, we show that integrating MR evidence with literature-mined data may identify causal effect intermediates or uncover mechanisms behind observed effects.

Development and validation of risk prediction models for colorectal cancer in patients with symptoms

Wei Xu[1], Ines Mesa Eguiagaray[1], Theresa Kirkpatrick[1], Jennifer Devlin [23], Stephanie Brogan[4], Patricia Turner[4], Chloe Macdonald[5], Michelle Thornton[6], Xiaomeng Zhang[1], Yazhou He[1], Xue Li[1], Malcolm Dunlop[23], Evropi Theodoratou[13]*

1. Centre for Global Health, Usher Institute, University of Edinburgh, Edinburgh, EH8 9AG, UK.
2. Colon Cancer Genetics Group, Medical Research Council Human Genetics Unit, Medical Research Council Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK.
3. Edinburgh Cancer Research Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK.
4. Clinical Research Team, Oncology Dept, Forth Valley Royal Hospital, Stirling Road, Larbert, FK5 4WR, UK.
5. University Hospital Wishaw & University Hospital Monklands, NHS Lanarkshire, ML6 0JS, UK.
6. Wishaw General Hospital, Wishaw, ML2 0DP, UK.

Background: There is a clear need to develop and validate prediction models for colorectal cancer (CRC) risk in patients with symptoms.

Methods: CRC prediction models were developed with internal validation in Study of Colorectal Cancer in Scotland and Lothian Bowel Symptoms Study [N=1352; Cases: n=818/ Controls: n=534]. Candidate predictors included age, sex, BMI, weighted genetic risk score (wGRS) of 113 SNPs, family history, and symptoms (change of bowel habit, rectal bleeding, weight loss, anaemia, abdominal pain). Models A (baseline model + wGRS) and B (baseline model) were developed based on LASSO regression algorithm to select predictors, whereas Models C (baseline model + wGRS) and D (baseline model) were built using all the variables. Models' prediction performance (calibration, discrimination) were evaluated through Hosmer-Lemeshow (HL) test (calibration curves were plotted) and Harrell's C-statistic. The corrected C-statistic was calculated based on bootstrapping validation (1,000 bootstrap resamples).

Results: Models A and B were constructed using LASSO-selected predictors (age, sex, anaemia, wGRS). Model A [C-statistic=0.718 (corrected: 0.715); HL-P=0.511] had better discrimination and calibration accuracy than Model B [C-statistic=0.707 (corrected: 0.705); HL-P=0.725]. Models C and D were built based on full model approach. Model C [C-statistic=0.743 (corrected: 0.735); HL-P=0.753] demonstrated better discrimination and calibration performance than model D [C-statistic=0.732 (corrected: 0.725); HL-P=0.802].

Models A and C that integrated wGRS in combination with demographic and clinical predictors had better prediction performance, which suggested incremental predictive value had been introduced by the addition of genetic variants. There was no statistical difference in C-statistics of models A and C [P=0.204]. An online CRC risk prediction calculator (A) was built: https://crcpredictionmodel.shinyapps.io/dynnomapp/.

Conclusion: In summary, integration of genetic architecture into CRC classical prediction model could improve prediction performance. The findings merit further investigation through model external validation and model clinical impact.