CAMBRIDGE
UNIVERSITY PRESS

Article

TwinsUK: The UK Adult Twin Registry Update

Serena Verdi[1], Golboo Abbasian[1], Ruth C. E. Bowyer[1], Genevieve Lachance[1], Darioush Yarand[1], Paraskevi Christofidou[1], Massimo Mangino[1,2], Cristina Menni[1], Jordana T. Bell[1], Mario Falchi[1], Kerrin S. Small[1], Frances M. K. Williams[1], Christopher J. Hammond[1], Deborah J. Hart[1], Timothy D. Spector[1] and Claire J. Steves[1,3]

[1]The Department of Twin Research, King's College London, St Thomas' Hospital, London, UK, [2]NIHR Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London, UK and [3]Clinical Age Research Unit, Department of Clinical Gerontology, King's College Hospital, NHS Foundation Trust, London, UK

Abstract

TwinsUK is the largest cohort of community-dwelling adult twins in the UK. The registry comprises over 14,000 volunteer twins (14,838 including mixed, single and triplets); it is predominantly female (82%) and middle-aged (mean age 59). In addition, over 1800 parents and siblings of twins are registered volunteers. During the last 27 years, TwinsUK has collected numerous questionnaire responses, physical/cognitive measures and biological measures on over 8500 subjects. Data were collected alongside four comprehensive phenotyping clinical visits to the Department of Twin Research and Genetic Epidemiology, King's College London. Such collection methods have resulted in very detailed longitudinal clinical, biochemical, behavioral, dietary and socioeconomic cohort characterization; it provides a multidisciplinary platform for the study of complex disease during the adult life course, including the process of healthy aging. The major strength of TwinsUK is the availability of several 'omic' technologies for a range of sample types from participants, which includes genomewide scans of single-nucleotide variants, next-generation sequencing, metabolomic profiles, microbiomics, exome sequencing, epigenetic markers, gene expression arrays, RNA sequencing and telomere length measures. TwinsUK facilitates and actively encourages sharing the 'TwinsUK' resource with the scientific community — interested researchers may request data via the TwinsUK website (http://twinsuk.ac.uk/resources-for-researchers/access-our-data/) for their own use or future collaboration with the study team. In addition, further cohort data collection is planned via the Wellcome Open Research gateway (https://wellcomeopenresearch.org/gateways). The current article presents an up-to-date report on the application of technological advances, new study procedures in the cohort and future direction of TwinsUK.

Keywords: TwinsUK; aging; omics; microbiome; pain; ophthalmology; epidemiology

(Received 31 May 2019; accepted 24 June 2019; First Published online 17 September 2019)

TwinsUK is the largest adult twin registry in the UK and is one of the most deeply phenotyped and genotyped datasets in the world. It provides a multidisciplinary platform to research both health- and social-related questions, with the overarching aim of understanding the etiology of complex disease and the aging process. The registry was started in 1992, with the initial intention to investigate osteoporosis and osteoarthritis. Such conditions are highly prevalent in women, and consequently, several hundred middle-aged women were recruited and formed the core of the initial register. Success from these early studies led to a rapid expansion of TwinsUK, and to date the cohort consists of 14,000 community-dwelling twins, male and female, aged over 18 years. Current research areas of interest include the genetics of metabolic syndrome, cardiovascular disease, the musculoskeletal system, sensory impairment and aging, as well as how the microbiome affects human health. Details of the registry's progression have been described previously (Moayyeri et al., 2013; Spector & Williams, 2006). To date, the TwinsUK

registry has contributed to over 850 publications and 800 international collaborations. More detailed description of research outputs may be accessed through the study website: http://www.twinsuk.ac.uk.

The Collection

Over the last 27 years, the TwinsUK registry has been enhanced through over 80 studies, some of which have been repeated over time. This has resulted in clinically rich, longitudinal phenotype information (Table 1), which may be categorized into four distinct time points (Verdi et al., 2019). Recruitment strategies have predominantly involved media campaigns. These have offered opportunities for adult twin pairs to join the registry and participate in unspecific research investigating various common diseases, without selecting for particular diseases or traits. At baseline (1992–2004), over 7000 twins responded to annual questionnaires, and approximately 5500 twins attended a full comprehensive clinical visit, which included several project-led studies. Age-matched characteristics of these volunteer twins were found not to differ from a singleton population-based cohort of British women (Chingford study; Andrew et al., 2001), apart from a lifelong lower weight in monozygotic (MZ) twins of approximately 1 kg. The follow-up visit occurred between April 2004 and May 2007, in which

3725 twins in the registry attended a full day clinical visit, and an additional 1299 twins posted blood taken via their GPs for DNA sampling. Participants ranged aged between 18 and 82 years (mean 52.5 ± 13 years) and the majority were female (89%). Protocols for the baseline and initial follow-up visit have been described previously (Spector & Williams, 2006).

The second wave of follow-up visits (August 2007–April 2012) aimed to investigate the aging process; Healthy Ageing Twin Study (HATS). Inclusion criteria were women aged ≥40 years with at least one previous clinical visit ($n = 4610$). In total, 3125 women (mean age 59.6 ± 9 years) attended the clinical visit. Follow-up time between first and last visits ranged between 6.1 and 17.4 years, with over 600 of the participants having 4 or more previous clinical visits. HATS outcomes have previously been described (Moayyeri et al., 2012), including details of data collection (Moayyeri et al., 2013).

The third wave of follow-up visits (May 2012 – May 2018) was performed to understand the interactions in disease processes between genes and the environment, as part of the Biomedical Research Centre (BRC) study. All participants of the TwinsUK registry were invited to attend a comprehensive clinical visit, which included collection of bone density/whole-body scan, cognitive and lung function, hearing and eye tests, fitness assessment (gait speed, chair stands and grip strength) and collection of blood, urine, stool and salivary samples. In total, 6686 clinical visits were made, with 3620 volunteers attending at least once and 1531 volunteers attended the clinic on 2 occasions with an average of 4 years between visits. In addition to the clinical visit, 6300 questionnaires were returned, complementing clinical data collected during the visit. Since February 2019, a further wave of follow-up visits has commenced that aims to continue the longitudinal data collection and adds further dynamic phenotyping and blood measurement over a 6-h visit incorporating standardized meals.

Longitudinal Data

Detailed clinical and biochemical phenotypes have been collected using harmonized protocols at each visit stage. A summary of a selection of clinical phenotypes is outlined in Table 2. In addition, questionnaire data have been collected on an annual basis and during visits, some which measure incident clinical endpoints such as cardiovascular accident, type 2 diabetes, chronic obstructive pulmonary disease, which have previously been described (Verdi et al., 2019). Three main comprehensive questionnaires ('TwinsUK Baseline Health', 'Baseline Core' and 'Longitudinal Core') were collected between 2004 and 2018 (detailed in Table 3). These were in paper format, completed at respondents' addresses and returned to the research facility. Over 2500 participants completed all 3 main questionnaires and 2300 completed either 2 of the main questionnaires. Furthermore, the demographic of the cohort provides an excellent resource to study aging where longitudinal changes are important to consider. Table 4 provides summaries of the key cognitive and frailty phenotypes we have acquired to explore questions in this area.

Alongside regular visits and questionnaires, TwinsUK has data linkage to official cancer and mortality data for retrospective analysis and future follow-up. Additional links to national health, education and environmental records to our own database are being established at present.

Novel Molecular and Genetic Phenotypes

In addition to epidemiological and clinical phenotypes collected from clinical visits, numerous biological samples, including body fluids (blood, urine, saliva, stool and sebum) and tissue (hair follicle, colonic mucosa, fat and skin biopsy), have been generously donated. Details of the samples collected are summarized in Table 1 and in Verdi et al. (2019). Collection methods have been described in their respective research publications, which can be found on the TwinsUK website (http://www.twinsuk.ac.uk). Here we describe the omic techniques (genomewide association studies, epigenomics, next-generation sequencing, metagenomics, metabolomics and microbiomics) that have been employed on biological samples and phenotypes from TwinsUK data. Details of some phenotypes collected prior to 2012 (e.g., telomere length) have previously been described (Moayyeri et al., 2013).

Genomewide Association Studies

TwinsUK has contributed to many international consortia for genomewide association analysis of various phenotypes (Mills & Rahal, 2018). Genomewide scan data using 2 chips (Illumina HumanHap300 BeadChip and Illumina HumanHap610 QuadChip) are available for 5654 (both MZ and dizygotic [DZ]) twins. The data have been fully imputed using '1000 Genomes' and 'Haplotype Reference Consortium – (HRC)' reference panels. TwinsUK is a member of many ongoing international consortia for meta-analysis of various traits such as height, BMI, lipids, obesity, blood pressure and back pain phenotypes. Some of the main publications from these collaborations can be found in the TwinsUK website. Our genomewide data are also being used to compile polygenic risk scores to isolate loci for various traits (Mills et al., 2018).

Epigenetic Markers

The first large-scale genomewide epigenetic assessment in TwinsUK was performed on DNA methylation patterns profiled on the Illumina HumanMethylation27 BeadChip in a whole blood sample of 172 female twins. This array examines 27,578 promoter CpG sites that map uniquely across the genome and some of these sites were found to be associated with age and age-related phenotypes (Bell et al., 2012). Subsequently, the Illumina Infinium HumanMethylation450 BeadChip was additionally applied up to 1000 blood samples to generate higher resolution genomewide DNA methylation profiles (Kurushima et al., 2019; Zhang et al., 2015), as well as in 322 skin (Roos et al., 2017) and 648 adipose (Grundberg et al., 2013) tissue biopsy samples from twins. More recently, the Illumina Infinium MethylationEPIC array is being profiled in additional blood samples from over 400 twins. Further epigenetic datasets in TwinsUK cohort have also been generated as part of the EpiTwin study (http://www.epitwin.eu), which in collaboration with the Beijing Genomics Institute, assayed epigenomic sequencing profiles in up 5000 samples from twins aged 16–85 years. The results include methylated DNA immunoprecipitation sequencing profiles in whole blood samples from twins discordant and concordant for a wide variety of diseases and environmental exposures (Bell et al., 2014, 2016; Davies et al., 2014; Yuan et al., 2014).

Gene Expression Markers

During the HATS visit, 856 twins with detailed clinical profiles underwent biopsies of multiple tissues as part of the Multiple Tissue Human Expression Resource project. This was a Wellcome Trust-funded study designed to investigate gene

Table 1. TwinsUK registry update

Register name	TwinsUK
Country	UK
Ascertainment type	Unselected volunteers
Opposite sex twins	Yes
Number of pairs separated by year of birth and sex	1900–1909: 2 FF
	1910–1919: 14 FF
	1920–1929: 304 FF; 48 MM; 4 FM
	1930–1939: 1418 FF; 196 MM; 36 FM
	1940–1949: 2352 FF; 312 MM; 64 FM
	1950–1959: 2380 FF; 486 MM; 98 FM
	1960–1969: 2104 FF; 502 MM; 70 FM
	1970–1979: 1702 FF; 506 MM; 56 FM
	1980–1989: 950 FF; 186 MM; 58 FM
	1990–1999: 554 FF; 162 MM; 56 FM
	2000–2001: 50 FF; 12 MM; 4 FM
Total	11,830 FF; 2410 MM, 446FM
	14,686 (7343 pairs)
Zygosity classification	51% MZ; 49% DZ. Assessed by the 'peas in a pod' questionnaire during registration (Jarrar et al., 2018) and confirmed via genotyping / genomewide association studies
Major interests	Common complex diseases, aging traits, nutrition and microbiome
Traits measured	Questionnaires and clinical examinations for an extensive range of traits, within the categories: twinship and family, demographics, behavior: dietary, sleep, physical activity, smoking, alcohol use; health care. Health: allergy and respiratory conditions; rheumatology; cardiology; operations, gastroenterology; oncology; urology and nephrology; neurology, psychiatry and mental health; dermatology; ophthalmology; oral health, hearing; endocrinology; obstetrics, gynecology and fertility, wellbeing; medication histories
Blood samples	8332 DNA; 8726 serum; 8614 plasma; 5685 peripheral blood lymphocytes
Other samples	8807 urine; 3206 saliva; 4576 stool; 1095 fat, muscle and skin biopsies; 5204 hair; 2711 peripheral blood mononuclear cells; 282 TRUculture (whole blood immune cell culture system)
Comments on omics	5654 genomewide association data
	2000 (UK10K), 2377 (HLI) next-gen sequenced data
	5000 (Epitwin) DNA methylation data
	Data available for transcriptome across 800 individuals in 3 tissues
Main sources of funding	The Wellcome Trust, the Medical Research Council and the Chronic Disease Research Foundation currently support TwinsUK. Since 2012, the GSTT/KCL NIHR Biomedical Research Centre has also provided support for twin visits. Several smaller charitable bodies and commercial companies have also contributed smaller grants
Contact	Prof. Timothy D. Spector
Address	The Department of Twin Research and Epidemiology, St. Thomas Hospital, London, SE1 7EH
Email	tim.spector@kcl.ac.uk or Victoria.vazquez@kcl.ac.uk
Website	http://twinsuk.ac.uk/

FF, female–female; MM, male–male; FM, female–male.

expression across multiple tissues simultaneously with the aim of examining mechanisms involved in common trait susceptibility. Gene expression in 3 tissues and derived cells, fat, skin and lymphoblastoid cell lines (LCL) was determined using Illumina whole genome expression array (HumanHT-12 version 3) comprising 48,803 probes in 3 technical replicates (Grundberg et al., 2012). The same skin, fat and LCL RNA samples plus an additional 400 whole blood samples were RNA sequenced as part of the EuroBATS project (Biomarkers of Ageing using whole Transcriptome Sequencing) a European (EU-FP7) study (Buil et al., 2015).

Whole Genome Sequencing

Whole genome sequencing (WGS) of 2000 healthy, deeply phenotyped twins formed part of the UK10K project, which used state-of-the-art next-generation sequencing methods to uncover rare genetic variants associated with health and disease. The data have been used extensively to describe population structure and functional annotation of rare and low-frequency variants (UK10K Consortium et al., 2015); further details can be accessed at: www.uk10k.org. In addition, approximately 1000 exome sequences at 30–60 × depth have been ascertained as part of

Table 2. Longitudinal data available in the TwinsUK registry participants

Phenotype	No. of measurements	No. of participants	No. of participants with ≥2 measurements	No. of participants with ≥4 measurements	Maximum number of visits	Maximum duration of follow-up (year)
Insulin	19,048	8298	4829	1767	12	24.2
Glucose	23,079	9227	5444	2421	13	24.7
Blood pressure	22,252	9422	5405	2299	8	25
Lipid profile	18,933	8496	4754	1703	9	24.7
Respiratory function	18,665	8880	4790	1494	7	24.2
Hip BMD	18,671	8713	4798	1563	7	21.3
Spine BMD	18,263	8497	4665	1553	7	19.5
Whole-body DXA	17,690	8148	4566	1529	7	21.5
Grip strength	13,551	6780	4008	577	4	13.4
BMI	20,881	8076	5288	2328	8	24.4

DXA, dual-energy X-ray absorptiometry; BMD, bone mineral density; BMI, body mass index.

Table 3. Longitudinal measures of main TwinsUK questionnaires

Questionnaire	Year	No. of responses	No. of participants	Maximum duration of follow-up (year)
TwinsUK Baseline Health Questionnaire	2004–2014	7000	5400	10
Baseline Core Questionnaire	2012–2018	6300	6300	NA
Longitudinal Core Questionnaire	2017–2018	3300	3300	NA

NA, not applicable.
Note: The Baseline Core and Longitudinal Core were collected 4 years apart.

Table 4. Longitudinal cognitive and frailty measures available in the TwinsUK registry participants

Phenotype	No. of measurements	No. of participants	No. of participants with two measurements	Mean follow-up (year)	Maximum duration of follow-up (year)
Mini-Mental State Examination	5426	4951	475	3.9	4.4
Standing balance	3941	3938	NA	NA	2.3
Gait speed	5387	4918	469	3.9	4.4
Chair stand	3928	3927	NA	NA	2.3
Verbal fluency	5413	4941	472	3.9	4.4
Deary-Liewald Reaction Time Test	3227	3226	NA	NA	NA
CANTAB Paired Associates Learning	2091	2091	NA	NA	NA

CANTAB, Cambridge Neuropsychological Test Automated Battery; NA, not applicable.

projects with GoT2D consortium and Pfizer Inc. More recently, WGS of >30× coverage was carried out through collaboration with Human Longevity, Inc. (HLI) for 2377 individuals from the TwinsUK cohort. DNA samples were sequenced on an Illumina HiSeqX sequencer using a 150-base paired-end single-index read format. The data have been used to disentangle to contribution of rare variants to the blood metabolome (Long et al., 2017), and are now under investigation to identify rare variants associated with complex diseases and traits, and for the inference of structural variants.

Metabolomics Profile

Fasting circulating metabolites levels (serum/plasma) have been assessed on TwinsUK study participants using different platforms:

(1) Biocrates AbsoluteIDQ (163 metabolites for 1052 twins) (Menni, Zhai et al., 2013); (2) Metabolon Inc. (Research Triangle Park, NC, USA; 280 known metabolites and 175 unknown metabolites for 6055 twins, 591 known and 165 unknown for 2069 twins at 3 time points); (3) Nightingale Health Ltd. (27 metabolic traits, 143 metabolite concentrations, 80 lipid ratios, 3 lipoprotein particle sizes and a semiquantitative measure of albumin for 2000 twins at 3 time points). Genomewide association studies have identified several genes involved in metabolic individuality (Long et al., 2017; Shin et al., 2014), as well as many health traits (see TwinsUK publications list: http://twinsuk.ac.uk/our-research/publications/). Nightingale Health Ltd. (Helsinki, Finland; previously known as Brainshake Ltd.) is a targeted Nuclear magnetic resonance spectroscopy platform that has been extensively applied by us and others for biomarker profiling in epidemiological studies (Barrios et al., 2018).

More recently, metabolomics profiling (Metabolon, Inc.) has been conducted on fecal ($n = 1016$; Zierer et al., 2018) and salivary samples (Nag et al., 2019).

Glycans

Glycosylation is the most common form of posttranscriptional protein modification and it is a putative mechanism in the modulation of the inflammatory response. The technology to assess glycosylation has recently become high throughput, and glycosylation of immunoglobulin G has been measured on 4900 twins while N-glycans in human serum glycoproteins have been measured in 1800 twins. Using this, we have found that glycans are highly heritable (Menni, Keser et al., 2013) and we have been the first to observe a number of associations between glycans and important age-related traits (Barrios et al., 2015; Menni, Gudelj et al., 2018).

Microbiome

Alongside the BRC study (third follow-up), over 5000 fecal samples have been collected for microbiome analysis. Twin volunteers provided stool samples, stored on site (St Thomas' Hospital, London) at −80°C. DNA extraction and 16S rRNA sequencing using the V4 variable region of nearly 3000 samples have been completed in collaboration with Cornell University using a multiplexed approach on the Illumina MiSeq platform. Smaller subsets of twins have also been sequenced with complementary methods by the BRC Genomics Facility at King's College London. In addition, plain saliva (700) and midstream urine (1600) specimens have undergone similar 16S amplicon sequencing using the same primers in collaboration with University of California San Diego and Stanford University.

Diversity metrics, taxonomic levels from genus through to phylum and relative abundances of operational taxonomic units (OTUs) have been used to assess microbiota associations within the TwinsUK data. Associations have been observed with a number of health deficits and medication usage (Jackson, Goodrich et al., 2016, Jackson et al., 2018; Le Roy et al., 2017), and age-related traits, including frailty and cognition (Jackson, Jeffrey et al., 2016; Verdi et al., 2018), among others (Menni, Lin et al., 2018). In addition, microbiota associations with diet (Menni, Jackson et al., 2017; Menni, Zierer et al., 2017; Ni Lochlainn et al., 2018) and socioeconomic status (Bowyer et al., 2019) have been found. More recently, amplicon sequence variants, also known as exact sequence, have been generated from 3345 stool samples. This approach offers a higher resolution than the OTU, allowing for greater sensitivity and specificity in identifying the taxonomic associations with traits (Wells et al., 2019).

Metagenomics

Whole metagenomic shotgun sequencing (WMGS) has been performed on fecal samples in 2 batches comprising 250 and 1004 volunteers from the TwinsUK registry. This larger dataset, including 161 MZ twin pairs, 201 DZ twin pairs and 280 singletons generated an average of 39M high-quality microbial reads per sample. Taxonomic and functional information have been inferred from the WMGS data. These results are being studied to determine the influence of the microbiome on the fecal and host metabolome, and to identify bacterial species and function mediating microbiome-associated increased risk for common disease.

Table 5. Food Frequency Questionnaire collection

Date of collection	No. of individuals	Longitudinal data comments
1993–2001	4470	Two or three time points in 890 individuals (Teucher et al., 2007)
2007	3370	NA (one time point only)
2014–2018	5440	Two time points (4 years apart) in 127 individuals

NA, not applicable.

Dietary Phenotypes

TwinsUK has detailed datasets on dietary habits, which have been collected since the inception of the registry. Data vary and include dietary indices on >5000 participants (e.g., Mediterranean Diet Score, Healthy Eating Index — 2010 and the Healthy Food Diversity Index; Bowyer et al., 2018). Dietary patterns, which are measured by category of foodstuff, have also been assessed through a Food Frequency Questionnaire previously used in the EPIC Study (Bingham et al., 2008). For details of collection, see Table 5.

Socioeconomic Data

The historical research focus of TwinsUK has shaped the main demographic of the twin cohort having middle socioeconomic status and education typical of a volunteer group (Moayyeri et al., 2013; Steves et al., 2013). Socioeconomic status of the twin volunteers has been collected since the registry's inception through self-reported questions (e.g., highest educational qualification status). More recently, the Index of Multiple Deprivation (IMD) has been compiled for all volunteers having UK postal codes, and data are to be linked to national databases for retrospective and future collection.

Index of Multiple Deprivation

Datasets from online government data repositories were combined, representing four of the UK's administrative countries: England (IMD version 2015), Scotland (IMD version 2016), Wales (IMD version 2014) and Northern Ireland (IMD version 2017). The IMD is a composite measure of area-level deprivation and considers the following domains: income, employment, education, skills and training, health deprivation, crime, barriers to housing and services, and the living environment. As methods may vary between the countries, and ranks are inappropriate (given the differing numbers of administrative districts in each country), the decile score was combined as a relative measure of deprivation. Datasets were matched to postcodes or Lower Layer Super Output Area (LSOA) codes at 17,498 time points for 12,041 individuals. Mean IMD decile score (considering all time points) was 6.49.

Future Directions and Collaborations

Longitudinal and detailed clinical, biochemical, behavioral, socioeconomic and deep omics (including multitissue characterization) of participants for nearly 30 years has provided a unique resource to study complex diseases and domains of healthy aging in the TwinsUK population. These, in conjunction with novel dynamic testing at study visits and lifestyle intervention studies, offer a unique opportunity to explore personalized medicine. High-quality data collection, database management, biological sample storage and statistical quality control enhance the resource. In addition, a

key strength of the resource lies in the highly engaged and loyal population; this is evident from the high retention levels of participation across studies. Blood, urine, DNA and multiple tissue samples are available for future measurements. Online questionnaires and active engagement with our twin participants using text messages, emails and social networking enable responsive and agile data collection. Our Volunteer Advisory Panel is key to developing new strategies and governance of participants, informing on decisions about the ethics, practicalities and appropriateness of potential studies.

The TwinsUK registry has a history of numerous successful scientific collaborations, and we remain committed to providing the scientific community with access to the phenotype data from the TwinsUK Resource. TwinsUK has an exemplary record for data sharing with over 800 data access requests, 150,000 samples shared to over 100 collaborators and over 600 publications in the past 6 years. Detailed descriptions for researchers of data and samples are on the data access pages of the website (http://www.twinsuk. ac.uk/data-access/cohortdata-description/); here, over 10,000 phenotypes can be searched. Longitudinal population studies funding from the Wellcome Trust continues to fund the core functions of TwinsUK and opens up the resource to successful cross-cohort collaborations. Over the next 5 years, TwinsUK will integrate electronic health records into an enhanced deep tissue omics resource and continue dynamic phenotypic testing into clinical visits. In addition, we will extend the age range of the registry to include volunteer twins from birth to adulthood, thus opening up the resource to study unique twin gene–environment interactions across the life course. New efficient broad consent will ensure that the communication with participating twins is ethical and proportionate. New annual sociological questionnaires will harmonize with English Longitudinal Study of Ageing and other LPS (1946/1958). We will also standardize mental health phenotypes between the complementary Twins Early Development Study (TEDS) such that, together, TwinsUK and TEDS cohorts will be an unparalleled twin resource across the life course. These developments will ensure TwinsUK will be a unique global resource of longitudinal omics and twin research across the life course, with immense potential for future scientific exploitation.

References

Andrew, T., Hart, D. J., Snieder, H., de Lange, M., Spector, T. D., & MacGregor, A. J. (2001). Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women. *Twin Research*, 4, 464–477.

Barrios, C., Zierer, J., Gudelj, I., Štambuk, J., Ugrina, I., Rodríguez, E., ... Menni, C. (2015). Glycosylation profile of IgG in moderate kidney dysfunction. *Journal of the American Society of Nephrology*, 27, 933–941.

Barrios, C., Zierer, J., Würtz, P., Haller, T., Metspalu, A., Gieger, C., ... Menni, C. (2018). Circulating metabolic biomarkers of renal function in diabetic and non-diabetic populations. *Scientific Reports*, 8, 15249. doi: 10.1038/s41598-018-33507-7

Bell, C. G., Xia, Y., Yuan, W., Gao, F., Ward, K., Roos, L., ... Spector, T. D. (2016). Novel regional age-associated DNA methylation changes within human common disease-associated loci. *Genome Biology*, 17, 193.

Bell, J. T., Loomis, A. K., Butcher, L. M., Gao, F., Zhang, B., Hyde, C. L., ... Spector, T. D. (2014). Differential methylation of the TRPA1 promoter in pain sensitivity. *Nature Communications*, 5, 2978.

Bell, J. T., Tsai, P. C., Yang, T. P., Pidsley, R., Nisbet, J., Glass, D., ... Deloukas, P. (2012). Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genetics*, 8, e1002629.

Bingham, S. A., Welch, A. A., McTaggart, A., Mulligan, A. A., Runswick, S. A., Luben, R., ... Day, N. E. (2008). Nutritional methods in the European Prospective Investigation of Cancer in Norfolk. *Public Health Nutrition*, 4, 847–858.

Bowyer, R. C., Jackson, M. A., Le Roy, C. I., Ni Lochlainn, M., Spector, T. D., Dowd, J. B., & Steves, C. J. (2019). Socioeconomic status and the Gut Microbiome: a TwinsUK cohort study. *Microorganisms*, 7, 17.

Bowyer, R. C. E., Jackson, M. A., Pallister, T., Skinner, J., Spector, T. D., Welch, A. A., & Steves, C. J. (2018). Use of dietary indices to control for diet in human gut microbiota studies. *Microbiome*, 6, 77.

Buil, A., Brown, A. A., Lappalainen, T., Vinuela, A., Davies, M. N., Zheng, H.-F., ... Dermitzakis, E. T. (2015). Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. *Nature Genetics*, 47, 88–91.

Davies, M. N., Krause, L., Bell, J. T., Gao, F., Ward, K. J., Wu, H., ... Wang, J. (2014). Hypermethylation in the ZBTB20 gene is associated with major depressive disorder. *Genome Biology*, 15, R56.

Grundberg, E., Meduri, E., Sandling, J. K., Hedman, Å. K., Keildson, S., Buil, A., ... Deloukas, P. (2013). Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *American Journal of Human Genetics*, 93, 876–890.

Grundberg, E., Small, K. S., Hedman, Å. K., Nica, A. C., Buil, A., Keildson, S., ... Spector, T. D. (2012). Mapping cis-and trans-regulatory effects across multiple tissues in twins. *Nature Genetics*, 44, 1084–1089.

Jackson, M. A., Goodrich, J. K., Maxan, M. E., Freedberg, D. E., Abrams, J. A., Poole, A. C., ... Steves, C. J. (2016). Proton pump inhibitors alter the composition of the gut microbiota. *Gut*, 65, 749–756.

Jackson, M. A., Jeffery, I. B., Beaumont, M., Bell, J. T., Clark, A. G., ... Steves, C. J. (2016). Signatures of early frailty in the gut microbiota. *Genome Medicine*, 8, 8.

Jackson, M. A., Verdi, S., Maxan, M. E., Shin, C. M., Zierer, J., Bowyer, R. C. E., ... Steves, C. J. (2018). Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nature Communications*, 9, 2655.

Jarrar, Z. A., Ward, K. J., Mangino, M., Cherkas, L. F., Gill, R., Gillham-Nasenya, I., ... Spector, T. D. (2018). Definitive zygosity scores in the Peas in the Pod Questionnaire is a sensitive and accurate assessment of the zygosity of adult twins. *Twin Research and Human Genetics*, 21, 146–154.

Kurushima, Y., Tsai, P. C., Castillo-Fernandez, J., Couto Alves, A., El-Sayed Moustafa, J. S., Le Roy, C., ... Bell, J. T. (2019). Epigenetic findings in periodontitis in UK twins: A cross-sectional study. *Clinical Epigenetics*, 11, 27.

Le Roy, C. I., Beaumont, M., Jackson, M. A., Steves, C. J., Spector, T. D., & Bell, J. T. (2017). Heritable components of the human fecal microbiome are associated with visceral fat. *Gut Microbes*, 9, 61–67.

Long, T., Hicks, M., Yu, H. C., Biggs, W. H., Kirkness, E. F., Menni, C., ... Telenti, A. (2017). Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nature Genetics*, 49, 568–578.

Menni, C., Gudelj, I., MacDonald-Dunlop, E., Mangino, M., Zierer, J., Bešić, E., ... Valdes, A. M. (2018). Glycosylation profile of immunoglobulin G is cross-sectionally associated with cardiovascular disease risk score and subclinical atherosclerosis in two independent cohorts. *Circulation Research*, 122, 1555–1564.

Menni, C., Jackson, M. A., Pallister, T., Steves, C. J., Spector, T. D., & Valdes, A. M. (2017). Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. *International Journal of Obesity*, 41, 1099–1105.

Menni, C., Keser, T., Mangino, M., Bell, J. T., Erte, I., Akmačić, I., ... Valdes, A. M. (2013). Glycosylation of immunoglobulin G: Role of genetic and epigenetic influences. *PLoS One*, 8, e82558.

Menni, C., Lin, C., Cecelja, M., Mangino, M., Matey-Hernandez, M. L., Keehn, L., ... Valdes, A. M. (2018). Gut microbial diversity is associated with lower arterial stiffness in women. *European Heart Journal*, 39, 2390–2397.

Menni, C., Zhai, G., MacGregor, A., Prehn, C., Römisch-Margl, W., Suhre, K., ... **Valdes, A. M.** (2013). Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics*, *9*, 506–514.

Menni, C., Zierer, J., Pallister, T., Jackson, M. A., Long, T., Mohney, R. P., ... **Valdes, A. M.** (2017). Omega-3 fatty acids correlate with gut microbiome diversity and production of N-carbamylglutamate in middle aged and elderly women. *Scientific Reports*, *7*, 11079.

Mills, M. C., Barban, N., & Tropf, F. C. (2018). The sociogenomics of polygenic scores of reproductive behavior and their relationship to other fertility traits. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *4*, 122–136.

Mills, M. C., & Rahal, C. (2018). A scientometric review of genome-wide association studies. *Communications Biology*, *2*, Article no. 9.

Moayyeri, A., Hammond, C. J., Hart, D. J., & Spector, T. D. (2013). The UK Adult Twin Registry (TwinsUK resource). *Twin Research and Human Genetics*, *16*, 144–149.

Moayyeri, A., Hammond, C. J., Valdes, A. M., & Spector, T. D. (2012). Cohort profile: TwinsUK and healthy ageing twin study. *International Journal of Epidemiology*, *42*, 76–85.

Nag, A., Kurushima, Y., Bowyer, R. C. E., Wells, P., Weiss, S., Pietzner, M., Steves, C. J. (2019). Genome-wide scan identifies novel genetic loci regulating salivary metabolite levels. *bioRxiv*. doi: 10.1101/687350

Ni Lochlainn, M., Bowyer, R. C. E., & Steves, C. J. (2018). Dietary protein and muscle in ageing people: The potential role of the gut microbiome and its metabolome. *European Geriatric Medicine*, *10*, E929.

Roos, L., Sandling, J. K., Bell, C. G., Glass, D., Mangino, M., Spector, T. D., ... Bell, J. T. (2017). Higher nevus count exhibits a distinct DNA methylation signature in healthy human skin: Implications for melanoma. *Journal of Investigative Dermatology*, *137*, 910–920.

Shin, S.-Y., Fauman, E. B., Petersen, A.-K., Krumsiek, J., Santos, R., Huang, J., ... Soranzo, N. (2014). An atlas of genetic influences on human blood metabolites. *Nature Genetics*, *46*, 543–550.

Spector, T. D., & Williams, F. M. (2006). The UK adult twin registry (TwinsUK). *Twin Research and Human Genetics*, *9*, 899–906.

Steves, C. J., Jackson, S. H. D., & Spector, T. D. (2013). Cognitive change in older women using a computerised battery: A longitudinal quantitative genetic twin study. *Behavior Genetics*, *43*, 468–479.

Teucher, B., Skinner, J., Skidmore, P. M. L., Cassidy, A., Fairweather-Tait, S. J., Hooper, L., ... MacGregor, A. J. (2007). Dietary patterns and heritability of food choice in a UK female twin cohort. *Twin Research and Human Genetics*, *10*, 734–748.

UK10K Consortium, Walter, K., Min, J. L., Huang, J., Crooks, L., Memari, Y., ... Zhang, W. (2015). The UK10K project identifies rare variants in health and disease. *Nature*, *526*, 82–90.

Verdi, S., Lachance, G., Bowyer, R. C., Abbasian, G., Yarand, D., Paraskevi, C., ... Steves, C. J. (2019). Cohort profile: TwinsUK. Manuscript in preparation.

Verdi, S., Jackson, M. A., Beaumont, M., Bowyer, R. C. E., Bell, J. T., Spector, T. D., & Steves, C. J. (2018). An investigation into physical frailty as a link between the gut microbiome and cognitive health. *Frontiers in Aging Neuroscience*, *10*, 398.

Wells, P. M., Williams, F. M., Matey-Hernandez, M. L., Menni, C., & Steves, C. J. (2019). RA and the microbiome: Do host genetic factors provide the link? *Journal of Autoimmunity*, *99*, 104–115.

Yuan, W., Xia, Y., Bell, C. G., Yet, I., Ferreira, T., Ward, K. J., ... Spector, T. D. (2014). An integrated epigenomic analysis for type 2 diabetes susceptibility loci in monozygotic twins. *Nature Communications*, *5*, 5719.

Zhang, W., Spector, T. D., Deloukas, P., Bell, J. T., & Engelhardt, B. E. (2015). Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements. *Genome Biology*, *16*, 14.

Zierer, J., Jackson, M. A., Kastenmüller, G., Mangino, M., Long, T., Telenti, A., ... Menni, C. (2018). The fecal metabolome as a functional readout of the gut microbiome. *Nature Genetics*, *50*, 790–795.