DNA methylation meta-analysis reveals cellular alterations in psychosis and markers of treatment-resistant schizophrenia

Dr. Eilis Hannon PhD[1], Dr. Emma L Dempster PhD[1], Miss. Georgina Mansell BSc[1], Dr. Joe Burrage PhD[1], Dr. Nick Bass PhD[2], Dr. Marc M Bohlken PhD[3], Prof. Aiden Corvin PhD[4], Charles J Curtis MSc[5,6], David Dempster BSc[5], Dr. Marta Di Forti PhD[5,7,8], Prof. Timothy G Dinan MD,PhD[9], Prof. Gary Donohoe PhD[10], Dr. Fiona Gaughran MB,MD[11,12], Prof. Michael Gill MD[13], Dr. Amy Gillespie PhD[14,11], Dr. Cerisse Gunasinghe PhD[5], Prof. Hilleke E Hulshoff PhD[15], Prof. Christina M Hultman PhD[16], Dr. Viktoria Johansson MD, PhD[17,18], Prof. Rene S Kahn MD,PhD[19,20], Prof. Jaakko Kaprio MD, PhD[21,22], Dr. Gunter Kenis PhD[23], Dr. Kaarina Kowalec PhD[16,24], Prof. James MacCabe MD,PhD[5], Prof. Colm McDonald PhD[25], Prof. Andew McQuillin PhD[26], Dr. Derek W Morris PhD[10], Prof. Kieran C Murphy PhD[27], Dr. Collette Mustard PhD[28], Prof. Igor Nenadic[29,30], Prof. Michael C O'Donovan PhD, FRCPsych[31], Dr. Diego Quattrone MD[5,7], Dr. Alexander L Richards PhD[31], Dr. Bart PF Rutten MD,PhD[32], Prof. David St Clair PhD[33], Dr. Sebastian Therman PhD[34], Prof. Timothea Toulopoulou PhD[35], Prof. Jim Van Os PhD[19], Prof. John L Waddington PhD[36], Wellcome Trust Case Control Consortium 2[23], CRESTAR consortium[23], Prof. Patrick Sullivan MD,PhD[16,37], Dr. Evangelos Vassos[5], Prof. Gerome Breen PhD[5,6], Prof. David Andrew Collier PhD[38], Prof. Robin Murray MD,PhD[39,7], Prof. Leonard S Schalkwyk PhD[40], Prof. Jonathan Mill PhD[1]

[1]University of Exeter Medical School, University of Exeter Medical School, University of Exeter, Barrack Road, Exeter, Devon, EX2 5DW, UK, [2]Division of Psychiatry, University College London, Gower Street, London, London, WC1E 6BT, [3]Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Heidelberglaan, Utrecht, Utrecht, The Netherlands, [4]Department of Psychiatry and Neuropsychiatric Genetics Research Group, Trinity Translational Medicine Institute, Trinity College Dublin, St. James Hospital, James's Street, Dublin, Dublin, Dublin 8, Ireland, [5]Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience (IoPPN), King's College London, De Crespigny Park, London, London, SE5 8AF,

UK, [6]NIHR BioResource Centre Maudsley, South London and Maudsley NHS Foundation Trust (SLaM), King's College London, De Crespigny Park, London, London, SE5 8AF, UK, [7]South London and Maudsley NHS Mental Health Foundation Trust, London, UK, [8]National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre, South London and Maudsley NHS Foundation Trust and King's College London, London, UK, [9]APC Microbiome Ireland, University College Cork, Cork, Cork, T12 YN60, Ireland, [10]Centre for Neuroimaging and Cognitive Genomics (NICOG), School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Galway, Galway, H91 CF50, Ireland, [11]Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience (IoPPN), King's College London, De Crespigny Park, London, London, SE5 8AF, UK, [12]National Psychosis Service, South London and Maudsley NHS Foundation Trust, Edward Street, London, London, SE8 5HA, UK, [13]Department of Psychiatry and Neuropsychiatric Genetics Research Group, Trinity Translational Medicine Institute, Trinity College Dublin, College Green, Dublin, Dublin, Ireland, [14]Department of Psychiatry, Medical Sciences Division, University of Oxford, Warneford Lane, Oxford, Oxfordshire, OX3 0JS, UK, [15]Department of Psychiatry, University Medical Center Utrecht, Heidelberglaan, Utrecht, Utrecht, The Netherlands, [16]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Stockholm, 171 77, Sweden, [17]Department of Medical Epidemiology and Biostatistics Sweden., Karolinska Institutet, Stockholm, Stockholm, 171 77, Sweden, [18]Department of Clinical Neuroscience, Center for psychiatry research, Karolinska Institutet & Stockholm Health Care Services, Stockholm, Stockholm, 171 77, Sweden, [19]Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Universiteitsweg, Utrecht, Utrecht, The Netherlands, [20]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, Gustave L. Levy Place, New York, NY, USA, [21]Institute for Molecular Medicine FIMM, University of Helsinki, Tukholmankatu 8, Helsinki, Helsinki, FI-00014, Finland, [22]Department of Public Health, University of Helsinki, Tukholmankatu 8B, Helsinki, Helsinki, Finland, [23](No affiliation data provided), [24]College of Pharmacy, University of Manitoba, Winnipeg, Manitoba, Canada, [25]Centre for Neuroimaging and Cognitive Genomics (NICOG), School of Medicine, National University of Ireland Galway, Galway, Galway, H91 CF50, Ireland, [26]Division of Psychiatry, University College London, Gower Street, London, London, WC1E 6BT, UK,

[27]Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin, Ireland, [28]Division of Biomedical Sciences, Institute of Health Research and Innovation, University of the Highlands and Islands, Old Perth Road, Inverness, Inverness, IV2 3JH, UK, [29]Department of Psychiatry and Psychotherapy, Jena University Hospital, Philosophenweg, Jena, Germany, [30]Department of Psychiatry and Psychotherapy, Philipps University Marburg/ Marburg University Hospital UKGM, Rudolf-Bultmann-Str, Marburg, Germany, [31]MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Maindy Road, Cardiff, Cardiff, CF24 4HQ, UK, [32]Department of Psychiatry and Neuropsychology, Faculty of Health, Medicine and Life Sciences, Maastricht University, Minderbroedersberg, Maastricht, Maastricht, The Netherlands, [33]The Institute of Medical Sciences, Univeristy of Aberdeen, Aberdeen, Aberdeenshire, UK, [34]Department of Public Health Solutions, Mental Health Unit, National Institute for Health and Welfare, P.O. Box 30, Helsinki, FI-00271, Finland, [35]Department of Psychology & National Magnetic Resonance Research Center (UMRAM), Aysel Sabuncu Brain Research Centre (ASBAM), Bilkent University, Ankara, Turkey, [36]Molecular and Cellular Therapeutics, Royal College of Surgeons in Ireland, Dublin, Dublin, Ireland, [37]Departments of Genetics and Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA, [38]Neuroscience Genetics, Eli Lilly and Company, Sunninghill Rd, Windlesham, Surrey, GU20 6PH, UK, [39]Department of Psychosis Studies, Institute of Psychiatry, King's College London, De Crespigny Park, London, London, SE5 8AF, UK, [40]School of Life Sciences, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK

Corresponding author: Jonathan Mill, University of Exeter Medical School, RILD building, Royal Devon & Exeter Hospital, Barrack Road, Exeter. EX2 5DW. UK. E-mail: j.mill@exeter.ac.uk

Abstract

We performed a systematic analysis of blood DNA methylation profiles from 4,483 participants from seven independent cohorts identifying differentially methylated positions (DMPs) associated with psychosis, schizophrenia and treatment-resistant schizophrenia. Psychosis cases were characterized by significant differences in measures of blood cell proportions and elevated smoking exposure derived from the DNA methylation data, with the largest differences seen in treatment-resistant schizophrenia patients. We implemented a stringent pipeline to meta-analyze epigenome-wide association study (EWAS) results across datasets, identifying 95 DMPs associated with psychosis and 1,048 DMPs associated with schizophrenia, with evidence of colocalization to regions nominated by genetic association studies of disease. Many schizophrenia-associated DNA methylation differences were only present in patients with treatment-resistant schizophrenia, potentially reflecting exposure to the atypical antipsychotic clozapine. Our results highlight how DNA methylation data can be leveraged to identify physiological (e.g., differential cell counts) and environmental (e.g., smoking) factors associated with psychosis and molecular biomarkers of treatment-resistant schizophrenia.

Introduction

Psychosis is a complex and heterogeneous neuropsychiatric condition characterized by a loss of contact with reality, whose symptoms can include delusions and hallucinations. Episodic psychosis and altered cognitive function are major features of schizophrenia, a severe neurodevelopmental disorder that contributes significantly to the global burden of disease (Whiteford et al., 2013). Schizophrenia is highly heritable (Hilker et al., 2018; Sullivan, Kendler, & Neale, 2003) and recent genetic studies have indicated a complex polygenic architecture involving hundreds of genetic variants that individually confer a minimal increase on the overall risk of developing the disorder(Purcell et al., 2009). Large-scale genome-wide association studies (GWAS) have identified approximately 160 regions of the genome harboring common variants robustly associated with the diagnosis of schizophrenia, with evidence for a substantial polygenic component in signals that individually fall below genome-wide levels of significance (Pardiñas et al., 2018; Schizophrenia Working Group of the PGC et al., 2014). As the majority of schizophrenia-associated variants do not directly index coding changes affecting protein structure, there remains uncertainty about the causal genes involved in disease pathogenesis, and how their function is dysregulated (Maurano et al., 2012).

A major hypothesis is that GWAS variants predominantly act to influence the regulation of gene expression. This hypothesis is supported by an enrichment of schizophrenia associated variants in core regulatory domains (e.g. active promotors and enhancers)(Hannon, Marzi, Schalkwyk, & Mill, 2019). As a consequence, there has been growing interest in the role of epigenetic variation in the molecular etiology of schizophrenia. DNA methylation is the best-characterized epigenetic modification, acting to influence gene expression via disruption of transcription factor binding and recruitment of methyl-binding proteins that initiate chromatin compaction and gene silencing. Despite being traditionally regarded as a mechanism of transcriptional repression, DNA methylation is actually associated with both increased and decreased gene expression(Wagner et al., 2014), and other genomic functions including alternative splicing and promoter usage (Maunakea et al., 2010). We previously demonstrated how DNA methylation is under local genetic control(Hannon, Gorrie-Stone, et al., 2018; Hannon, Spiers, et al., 2015), identifying an enrichment of DNA methylation quantitative

trait loci (mQTL) among genomic regions associated with schizophrenia(Hannon, Spiers, et al., 2015). Furthermore, we have used mQTL associations to identify discrete sites of regulatory variation associated with schizophrenia risk variants implicating specific genes within these regions (Hannon et al., 2016; Hannon, Gorrie-Stone, et al., 2018; Hannon, Spiers, et al., 2015; Hannon, Weedon, Bray, O'Donovan, & Mill, 2017). Of note, epigenetic variation induced by environmental exposures has been hypothesized as another mechanism by which non-genetic factors can affect risk for neuropsychiatric disorders including schizophrenia(E. Dempster, Viana, Pidsley, & Mill, 2013).

The development of standardized assays for quantifying DNA methylation at specific sites across the genome has enabled the systematic analysis of associations between methylomic variation and environmental exposures or disease(Murphy & Mill, 2014). Because DNA methylation is a dynamic process, these epigenome-wide association studies (EWAS) are more complex to design and interpret than GWAS(Mill & Heijmans, 2013; Rakyan, Down, Balding, & Beck, 2011; Relton & Davey Smith, 2010). As for observational epidemiological studies of exposures and outcomes, a number of potentially important confounding factors (e.g. tissue- or cell-type, age, sex, lifestyle exposures, medication, and disorder-associated exposures) that can directly influence DNA methylation need to be considered along with the possibility of reverse causation. Despite these difficulties, recent studies have identified schizophrenia-associated DNA methylation differences in analyses of post-mortem brain tissue(Jaffe et al., 2015; Pidsley et al., 2014; Viana et al., 2016; Wockner et al., 2014), and also detected disease-associated variation in peripheral blood samples from both schizophrenia-discordant monozygotic twin pairs (E. L. Dempster et al., 2011) and clinically-ascertained case-control cohorts (Aberg et al., 2014; Hannon et al., 2016; Kinoshita et al., 2014). We previously reported an EWAS of variable DNA methylation associated with schizophrenia in >1,700 individuals, meta-analyzing data from three independent cohorts and identifying methylomic biomarkers of disease(Hannon et al., 2016).Together these data support a role for differential DNA methylation in the molecular etiology of schizophrenia, although it is not clear whether disease-associated methylation differences are themselves secondary to the disorder itself, or a result of other schizophrenia-associated factors.

In this study we extend our previous analysis, quantifying DNA methylation across the genome in a total of 4,483 participants from seven independent case-control cohorts including patients with schizophrenia or first-episode psychosis (FEP) (**Figure 1**). This represents the largest EWAS of schizophrenia and psychosis, and one of the largest case-control studies of DNA methylation for any disease phenotype. In each cohort, genomic DNA was isolated from whole blood and DNA methylation was quantified across the genome using either the Illumina Infinium HumanMethylation450 microarray ("450K array") or the HumanMethylationEPIC microarray ("EPIC array") (see **Methods**). We implemented a stringent pipeline to meta-analyze EWAS results across datasets to identify associations between psychosis cases and variation in DNA methylation. We show how DNA methylation data can be leveraged to identify biological (e.g. differential cell counts) and environmental (e.g. smoking) factors associated with psychosis, and present evidence for molecular variation associated with clozapine exposure in patients with treatment-resistant schizophrenia.

Results

Study overview and cohort characteristics

We quantified DNA methylation in samples derived from peripheral venous whole blood in seven independent psychosis case-control cohorts (total n = 4,483; 2,379 cases and 2,104 controls). These cohorts represent a range of study designs and recruitment strategies and were initially designed to explore different clinical and etiological aspects of schizophrenia (see **Methods** and **Table 1**); they include studies of first episode psychosis (EU-GEI and IoPPN), established schizophrenia and/or clozapine usage (UCL, Aberdeen, Dublin, IoPPN), mortality in schizophrenia (Sweden), and a study of twins from monozygotic pairs discordant for schizophrenia (Twins). All cohorts were characterised by a higher proportion of male participants (range = 52.1–71.1% male, pooled mean = 62.6% male, **Table 1**) than females. Although there was an overall significantly higher proportion of males amongst cases compared to controls ($\chi^2 = 37.5$, $P = 9.35 \times 10^{-10}$), consistent with reported incidence rates (Aleman, Kahn, & Selten, 2003; van der Werf et al., 2014), there was significant heterogeneity in the sex by diagnosis proportions across different cohorts ($\chi^2 = 348$, $P = 4.86 \times 10^{-63}$) with the overall excess of male patients driven by two cohorts (UCL ($\chi^2 = 52.7$, $P = 3.81 \times 10^{-13}$) and EU-GEI ($\chi^2 =$

25.9, P = 3.68×10^{-7})). Most cohorts were enriched for young and middle-aged adults although there was considerable heterogeneity across the studies reflecting the differing sampling strategies (**Table 1**). For example, the IoPPN cohort has the lowest average age, reflecting the inclusion of a large number of first episode psychosis (FEP) patients (mean = 34.9 years; SD = 12.42 years)(Di Forti et al., 2009). In contrast, individuals in the Sweden cohort were older (mean = 60.0 years; SD = 8.9 years)(Kowalec et al., 2019). There was no overall difference in mean age between cases and controls (mean difference = 0.076 years, P = 0.975) (**Figure 1 – supplement 1**), although differences were apparent in individual cohorts; in UCL (mean difference = 6.8 years; P = 6.55×10^{-9}) and IoPPN (mean difference = 6.2 years; P = 1.46×10^{-11}) patients were significantly older than controls, while in the EU-GEI (mean difference = -7.9 years; P = 1.24×10^{-22}) and the Sweden cohort (mean difference = -7.3 years; P = 1.05×10^{-16}) the cases were significantly younger. With the exception of individuals in the IoPPN and EU-GEI cohorts, which are more ethnically diverse, individuals included in this study were predominantly Caucasian. SNP array data from each donor was merged with HapMap Phase 3 data, and genetic principal components (PCs) were calculated with GCTA (Yang, Lee, Goddard, & Visscher, 2011) to further confirm the ethnicity of each sample (**Figure 1 – supplement 2**).

Psychosis patients are characterized by differential blood cell proportions and smoking levels using measures derived from DNA methylation data

A number of robust statistical classifiers have been developed to derive estimates of both biological phenotypes (e.g. age (Hannum et al., 2013; Horvath, 2013; Zhang et al., 2019) and the proportion of different blood cell types in a whole blood sample (Houseman et al., 2012; Koestler et al., 2013)) and environmental exposures (e.g. tobacco smoking (Elliott et al., 2014; Sugden et al., 2019)) from DNA methylation data. These estimates can be used to identify differences between groups and are often included as covariates in EWAS analyses where empirically-measured data is not available. For each individual included in this study we calculated two measures of “epigenetic age” from the DNA methylation data; DNAmAge using the Horvath multi-tissue clock, which was developed to predict chronological age (Horvath, 2013), and PhenoAge, which was developed as biomarker of advanced biological aging (Levine et al., 2018). We found a strong correlation between reported age and both

derived age estimates across the cohorts (Pearson correlation coefficient range 0.821-0.928 for DNAmAge and 0.795-0.910 for PhenoAge) and no evidence for age acceleration - i.e. the difference between epigenetic age and chronological age - between patients with psychosis and controls (Kowalec et al., 2019) (**Figure 1 - supplement 3 and 4**).

Because of the importance of considering variation in the composition of the constituent cell types in analyses of complex cellular mixtures (Mill & Heijmans, 2013; Relton & Davey Smith, 2010), we used established methods to estimate the proportion (Houseman et al., 2012; Koestler et al., 2013) and abundance (Horvath, 2013) of specific cell-types in whole blood. Using a random effects meta-analysis to combine the results across the seven cohorts, which were adjusted for age, sex and DNAm smoking score, we found that psychosis cases had elevated estimated proportions of granulocytes (mean difference = 0.0431; $P = 5.09x10^{-4}$) and monocytes (mean difference = 0.00320; $P = 1.15x10^{-4}$), and significantly lower proportions of $CD4^+$ T-cells (mean difference = -0.0177; P = 0.00144), $CD8^+$ T-cells (mean difference = -0.0144; P = 0.00159) and natural killer cells (mean difference = -0.0113; P = 0.00322) (**Table 2** and **Figure 2**). Interestingly, the differences in granulocytes, natural killer cells, $CD4^+$ T-cells and $CD8^+$ T-cells were most apparent in cohorts comprising patients with a diagnosis of schizophrenia (**Figure 2**), with cohorts including FEP patients characterized by weaker or null effects. Limiting the analysis of derived blood cell estimates to a comparison of schizophrenia cases and controls did not perceivably change the estimated differences of our random effects model but did reduce the magnitude of heterogeneity compared to including the FEP cases (**Supplementary Table 1**). This indicates that changes in blood cell proportions may reflect a consequence of diagnosis, reflecting the fact that people with schizophrenia are likely to have been exposed to a variety of medications, social adversities and somatic ill-health - and for longer periods - than FEP patients. Finally, we used an established algorithm to derive a quantitative DNA methylation "smoking score" for each individual (Elliott et al., 2014), building on our previous work demonstrating the utility of this variable for characterizing differences in smoking exposure between schizophrenia patients and controls, and using it as a covariate in an EWAS (Hannon et al., 2016). We observed a significantly increased DNA methylation smoking score (**Figure 3**) in psychosis patients

compared to controls across all cohorts (mean difference = 3.89; $P = 2.88x10^{-11}$). Although of smaller effect, this difference was also present when comparing FEP and controls in the EU-GEI cohort (mean difference = 2.38; $P = 2.68x10^{-8}$). As expected, for individuals where self-reported smoking data was available, the DNA methylation smoking score was significantly elevated in current and former smokers compared to never smokers (**Figure 3 – supplement 1**).

An epigenome-wide association study meta-analysis identifies DNA methylation differences associated with psychosis

To identify differentially methylated positions (DMPs) in blood associated with psychosis, we performed an association analysis within each of the seven schizophrenia and FEP cohorts controlling for age, sex, derived cellular composition variables (from DNA methylation data), derived smoking score (from DNA methylation data), and experimental batch (see **Methods**). We used a Bayesian method to control P-value inflation using the R package *bacon* (van Iterson, van Zwet, Heijmans, & Consortium, 2017) before combining the estimated effect sizes and standard errors across cohorts with a random effects meta-analysis, including all autosomal and X-chromosome DNA methylation sites analyzed in at least two cohorts (n = 839,131 DNA methylation sites) (see **Methods**). Using an experiment-wide significance threshold derived for the Illumina EPIC array (Mansell et al., 2019) ($P < 9x10^{-8}$), we identified 95 psychosis-associated DMPs mapping to 93 independent loci and annotated to 68 genes (**Figure 4A** and **Supplementary Table 2)**. Across these DMPs, the mean difference in DNA methylation between cases and controls was relatively small (0.789%, SD = 0.226%) and there was a striking enrichment of hypermethylated DMPs in psychosis cases (n = 91 DMPs (95.8%) hypermethylated, $P = 1.68x10^{-22}$). A number of the top-ranked DMPs are annotated to genes that have direct relevance to the etiology of psychosis including the GABA transporter *SLC6A12*(Park et al., 2011) (cg00517261, mean difference = 0.663%, $P = 1.53x10^{-8}$), the GABA receptor *GABBR1*(Le-Niculescu et al., 2007) (cg00667298, mean difference = 0.619%, $P = 5.07x10^{-9}$), and the calcium voltage-gated channel subunit gene *CACNA1C* (cg01833890, mean difference = 0.458%, $P = 8.42x10^{-9}$) that is strongly associated with schizophrenia and bipolar disorder (Consortium, 2013; Psychiatric

GWAS Consortium Bipolar Disorder Working Group, 2011; Schizophrenia Working Group of the PGC et al., 2011) (**Figure 5**).

A specific focus on clinically-diagnosed schizophrenia cases identifies more widespread DNA methylation differences

We next repeated the EWAS focussing specifically on the subset of psychosis cases with diagnosed schizophrenia (schizophrenia cases = 1,681, controls = 1,583). Compared to our EWAS of psychosis we identified more widespread differences in DNA methylation (**Figure 4B**), with 1,048 schizophrenia associated DMPs ($P < 9x10^{-8}$) representing 1,013 loci and annotated to 692 genes (**Supplementary Table 3**). Although the list of schizophrenia-associated DMPs included 61 (64.21%) of the psychosis associated DMPs, the total number of significant differences was much larger, potentially reflecting the less heterogeneous clinical characteristics of the cases. Schizophrenia-associated DMPs had a mean difference of 0.789% (SD = 0.204%), and like the psychosis-associated differences, were significantly enriched for sites that were hypermethylated in cases compared to controls (n = 897 (87.4%), $P = 1.27x10^{-129}$)). A number of the top-ranked DMPs are annotated to genes that have direct relevance to the etiology of schizophrenia and gene ontology (GO) analysis highlighted multiple pathways previously implicated in schizophrenia including several related to the extracellular matrix(Berretta, 2012) and cell-cell adhesion(O'Dushlaine et al., 2011) (**Supplementary Table 4**). Given the large range of ages across the samples included in this study, we tested whether there was evidence for a relationship between age and differential DNA methylation at the 1,048 schizophrenia DMPs by refitting our analysis model using an additional interaction term between age and schizophrenia status individually for each cohort prior to the interaction effects being meta-analysed (see **Methods**). Overall, we found limited evidence for a relationship between age and DNA methylation at schizophrenia-associated DMPs; controlling for multiple testing ($P < 0.00004771$), only two (0.002%) DMPs were identified as showing a significant interaction with age (**Supplementary Table 5**). We used the same approach to explore for an interaction between sex and DNA methylation, finding no evidence for sex differences at these sites or evidence for a significant interaction between sex and DNA methylation ($P < 0.00004771$) (**Supplementary Table 6**). Finally,

288 although most of the cohorts included in this study were predominantly Caucasian, there was some
289 ethnic heterogeneity in the IoPPN and EU-GEI cohorts. To explore the extent to which this diversity
290 might be influencing our results we merged SNP array data from each donor with HapMap Phase 3
291 data and calculated genetic PCs using GCTA (Yang et al., 2011) (**Figure 1 – supplement 2**). We
292 reanalyzed data from individual cohorts including increasing numbers of genetic PCs to the model,
293 finding that even in the most ethnically diverse cohort (IoPPN) the inclusion of up to five genetic PCs
294 had negligible effects, with a very strong correlation in test statistics between models (**Figure 4 –**
295 **supplement 1**).
296
297
298 *Schizophrenia-associated DNA methylation differences show overlap with previous analyses of*
299 *schizophrenia and other traits*
300 Two of our experiment-wide significant SZ-associated DMPs (cg00390724 and cg09868768)
301 overlapped with those reported in a previous smaller whole blood schizophrenia EWAS performed by
302 Montano and colleagues (Montano et al., 2016) with the same direction of effect; of note, 119
303 (71.3%) of the 167 replicated DMPs reported by this study were characterized by a consistent
304 direction of effect in our meta-analysis, representing a significantly higher rate than expected by
305 chance ($P = 3.83x10^{-8}$). Unfortunately, we could not check the extent to which our schizophrenia-
306 associated DMPs were replicated in the Montano et al dataset because the full results from their
307 analysis are not publicly available. We next compared our results with those from a prefrontal cortex
308 (PFC) EWAS meta-analysis of schizophrenia also performed by our group (Viana et al., 2017),
309 finding that 627 (60.2%) of the 1,042 DMPs tested in both analyses had the same direction of effect, a
310 significantly higher rate than expected by chance ($P = 5.43X10^{-11}$). Finally, we also explored the
311 extent to which DMPs associated with schizophrenia overlapped with other traits using the database
312 of results in the online EWAS catalog (http://ewascatalog.org/); across EWAS studies undertaken
313 using blood DNA (isolated from whole blood or cord blood) this resource includes 101,091
314 significant DMPs (at $P < 1X10^{-7}$) associated with 87 traits. Of the 1,048 schizophrenia-associated
315 DMPs identified in our meta-analysis, 219 (20.9%) were present in the database and significantly

associated with 18 different traits (**Supplementary Table 7**). Where effect sizes for individual DMPs were available in the EWAS catalog, we tested for an enrichment of consistent (or discordant) associations to those identified with schizophrenia. Schizophrenia DMPs also associated with C-reactive protein (CRP) and gestational age, for example, were significantly enriched for a consistent direction of effect (CRP: 10 overlapping DMPs, 10 consistent direction of effect, P = 0.001953; gestational age: 105 overlapping DMPs, 72 consistent direction of effect, P = 0.000178). In contrast, schizophrenia DMPs also associated with age and high-density lipoprotein (HDL) cholesterol were enriched for discordant effect directions (age: 30 overlapping DMPs, 28 same direction of effect, P = $8.68X10^{-7}$; HDL: 12 overlapping DMPs, 12 same direction of effect, P = 0.00049) (**Figure 6**).

Schizophrenia-associated DMPs colocalize to regions nominated by genetic association studies

As the etiology of schizophrenia has a large genetic component, we next sought to explore the extent to which DNA methylation at schizophrenia-associated DMPs is influenced by genetic variation. Using results from a quantitative genetic analysis of DNA methylation in monozygotic and dizygotic twins (Hannon, Knox, et al., 2018), we found that DNA methylation at schizophrenia-associated DMPs is more strongly influenced by additive genetic factors compared to non-associated sites matched for comparable means and standard deviations (**Figure 7**) (mean additive genetic component across DMPs = 23.0%; SD = 16.8%; P = $1.61x10^{-87}$). Using a database of blood DNA methylation quantitative trait loci (mQTL) previously generated by our group (Hannon, Gorrie-Stone, et al., 2018) we identified common genetic variants associated with 256 (24.4%) of the schizophrenia-associated DMPs. Across these 256 schizophrenia-associated DMPs there were a total of 455 independent genetic associations with 448 genetic variants, indicating that some of these DMPs are under polygenic control with multiple genetic variants associated. Of note, 31 of these genetic variants are located within 12 schizophrenia-associated GWAS regions (**Supplementary Table 8**) with 19 genetic variants associated with schizophrenia DMPs located in the MHC region on chromosome 6. To further support an overlap between GWAS and EWAS signals for schizophrenia, we compared the list of genes identified in this study with those from the largest GWAS meta-analysis of schizophrenia (Pardiñas et al., 2018) identifying 21 schizophrenia-associated DMPs located in 11 different GWAS

regions. To more formally test for an enrichment of differential DNA methylation across schizophrenia-associated GWAS regions, we calculated a combined EWAS P-value for each of the GWAS associated regions using all DNA methylation sites within each region identifying 21 significant regions ($P < 3.16x10^{-4}$, corrected for testing 158 regions; **Supplementary Table 9**). Three of these regions also contained a significant schizophrenia-associated DMP and a genetic variant associated with that schizophrenia-associated DMP. These include a region located within the MHC, another located on chromosome 17 containing *DLG2*, *TOM1L2* and overlapping the Smith-Magenis syndrome deletion, and another on chromosome 16 containing *CENPT*, and *PRMT7*.

Schizophrenia-associated patterns of DNA methylation are observed in individuals with first-episode psychosis

To explore whether schizophrenia-associated differences in DNA methylation are present before a formal diagnosis of schizophrenia we next performed an EWAS of FEP in the IoPPN and EUGEI cohorts (total n = 698 FEP cases and 724 controls), meta-analysing the results across 384,217 common DNAm sites. Although we identified no significant DMPs at our stringent experiment-wide significance threshold, this is not surprising given the greatly attenuated sample size and the high phenotypic heterogeneity amongst individuals with FEP compared to diagnosed schizophrenia; both factors negatively influence power to detect effects. We next repeated our EWAS of diagnosed schizophrenia, excluding the IoPPN cohort to ensure that there were no overlapping samples between the schizophrenia vs control analysis and the FEP vs control analysis, identifying 125 significant DMPs of which 101 were also tested in the FEP EWAS. To see if there was any evidence for differential DNAm at these sites prior to a diagnosis of schizophrenia, we compared the estimated differences between schizophrenia cases and controls and FEP cases and controls (**Supplementary Table 10**). Strikingly, 96 (95.0%) of the tested DMPs had a consistent direction of effect in the FEP EWAS, a significantly higher rate than expected by chance (P = 6.58 x10-23). While this result is consistent with schizophrenia-associated differences being present prior to diagnosis, it is not sufficient to state that they are causal; they may still reflect some underlying environmental risk factor or be a consequence of FEP (e.g. medication exposure).

372

373 *Treatment-resistant schizophrenia cases differ from treatment-responsive schizophrenia patients for*
374 *blood cell proportion estimates and smoking score derived from DNA methylation data*

375 Up to 25% of schizophrenia patients are resistant to the most commonly prescribed antipsychotic
376 medications, and clozapine is a second-generation antipsychotic often prescribed to patients with such
377 treatment-resistant schizophrenia (TRS) who may represent a more severe subgroup (Ajnakina et al.,
378 2018). Using data from four cohorts for which medication records were available (UCL, Aberdeen,
379 IoPPN, and Sweden), we performed a within-schizophrenia analysis comparing schizophrenia patients
380 prescribed clozapine (described as TRS cases) and those prescribed standard antipsychotic
381 medications (total n = 399 TRS and 636 non-TRS). Across each of the four cohorts the proportion of
382 males prescribed clozapine was slightly higher than the proportion of males on other medications (χ^2
383 = 7.04, P = 7.96×10^{-3}; **Supplementary Table 11**) consistent with findings from epidemiological
384 studies that report increased rates of clozapine prescription in males(Bachmann et al., 2017), although
385 there was statistically significant heterogeneity in the sex distribution between groups across cohorts
386 (χ^2 = 20.5, P =0.0150). TRS cases were significantly younger than non-TRS cases (mean difference =
387 -5.48 years, P = 0.00533), although there was significant heterogeneity between the cohorts (I^2 = 89%;
388 P = 7.40×10^{-32}). There was no evidence of accelerated epigenetic aging between TRS and non-TRS
389 patients (**Figure 1 – supplement 5** and **Figure 1 – supplement 6**). Interestingly, cellular composition
390 variables derived from the DNA methylation data suggests that TRS cases are characterized by a
391 significantly higher proportion of granulocytes (meta-analysis mean difference = 0.00283; P =
392 8.10×10^{-6}) and lower proportions of $CD8^+$ T-cells (mean difference = -0.0115; P = 4.37×10^{-5}
393 (**Supplementary Table 12** and **Figure 2 – supplement 1**) compared to non-TRS cases. Given the
394 finding of higher derived granulocyte and lower $CD8^+$ T-cell levels in the combined psychosis patient
395 group compared to controls described above, a finding driven primarily by patients with
396 schizophrenia, we performed a multiple regression analysis of granulocyte proportion to partition the
397 effects associated with schizophrenia status from effects associated with TRS status. After including a
398 covariate for TRS, schizophrenia status was not significantly associated with granulocyte proportion
399 using a random effects model (P = 0.210) but there was significant heterogeneity of effects across the

four cohorts (I^2 = 91%, P = 4.93x10^{-7}). Within the group of patients with schizophrenia, however, there were notable differences between TRS and non-TRS groups (mean difference = 0.0275; P = 3.22x10^{-6}; **Figure 2 – supplement 2)**. In contrast a multiple regression analysis found that both schizophrenia status (mean difference = -0.0113; P = 0.00818) and TRS status (mean difference = -0.0116; P = 2.82x10^{-5}) had independent additive effects on $CD8^+$ T-cell proportion (**Figure 2 – supplement 3**). Finally, TRS was also associated with significantly higher DNA methylation-derived smoking scores than non-TRS in all four cohorts (mean difference = 2.16; P = 7.79x10^{-5}; **Figure 3 – supplement 2**). Testing both schizophrenia diagnosis status and TRS status simultaneously, we found that both remained significant; schizophrenia diagnosis was associated with a significant increase in smoking score (mean difference = 3.98, P = 2.19x10^{-8}) with TRS status associated with an additional increase within cases (mean difference = 2.15, P = 2.22x10^{-7}) (**Figure 3 – supplement 3**).

There are widespread DMPs between treatment-resistant schizophrenia patients and treatment-responsive patients

We next performed an EWAS within schizophrenia patients comparing TRS cases to non-TRS cases, including each autosomal and X-chromosome DNA methylation site analyzed in at least two cohorts (n = 431,659 DNA methylation sites). We identified seven DMPs associated with clozapine exposure (P < 9x10^{-8}; **Supplementary Table 13**) with a mean difference of 1.47% (SD = 0.242%) and all sites being characterized by elevated DNA methylation in TRS cases (P = 0.0156). We were interested in whether the DNA methylation differences associated with TRS overlapped with those identified between all schizophrenia cases and non-psychiatric controls. Although there was no direct overlap between the clozapine associated DMPs and the schizophrenia associated DMPs identified for each analysis, the direction of effects across the 1,048 schizophrenia-associated DMPs were enriched for consistent effects (n = 738 (70.4%) DMPs with consistent direction; P = 7.57x10^{-41}). Given these observations, we formally tested whether the schizophrenia-associated differences are driven by the subset of TRS cases on clozapine by fitting a model that simultaneously estimates the effect of schizophrenia status and TRS status across all 1,048 sites (**Supplementary Table 14**). While the vast majority of schizophrenia associated DMPs remained at least nominally significant (n = 1,003 95.7%,

$P < 0.05$) between schizophrenia patients and controls, amongst those that didn't 25 (2.39%) had a significant effect associated with TRS status. For example, differential DNA methylation at the schizophrenia-associated DMP cg16322565, located in the *NR1L2* gene on chromosome 3 (schizophrenia EWAS meta-analysis: mean DNA methylation difference = 0.907%, $P = 3.52x10^{-9}$), is driven primarily by cases with TRS (**Figure 8**; multiple regression analysis mean DNA methylation difference between schizophrenia cases and controls = 0.323%, P = 0.123, mean DNA methylation difference between TRS cases and non-TRS controls = 1.01%, $P = 8.71x10^{-5}$). 152 (14.5%) of the schizophrenia associated DMPs were associated with a significant effect between schizophrenia cases and controls and a significant affect within schizophrenia patients between TRS and non-TRS patients, with the majority (128 (84.2%)) characterized by the same direction of effect in both groups and indicative of an additive effect of both schizophrenia diagnosis and TRS status (e.g. **Figure 8 – supplement 1**). Of particular interest are 24 DMPs which are significantly associated with both schizophrenia and TRS but with an opposite direction of effect, highlighting how that at some DNA methylation sites, TRS counteracts changes induced by schizophrenia (e.g. **Figure 8 – supplement 2)**. Taken together, 177 (16.9%) of the schizophrenia-associated DMPs identified in our EWAS meta-analysis are influenced by TRS and reflect either differences induced by exposure to a specific antipsychotic therapy or other differences (e.g. treatment resistance) in individuals who are prescribed clozapine.

448 **Discussion**

449 We report the most comprehensive study of methylomic variation associated with psychosis and
450 schizophrenia, profiling DNA methylation across the genome in peripheral blood samples from 2,379
451 cases and 2,104 controls. We show how DNA methylation data can be leveraged to derive measures
452 of blood cell counts and smoking that are associated with psychosis. Using a stringent pipeline to
453 meta-analyze EWAS results across datasets, we identify novel DMPs associated with both psychosis
454 and a more refined diagnosis of schizophrenia. Of note, we show evidence for the co-localization of
455 genetic associations for schizophrenia and differential DNA methylation. Finally, we present evidence
456 for differential methylation associated with treatment-resistant schizophrenia, potentially reflecting
457 differences in DNA methylation associated with exposure to the atypical antipsychotic drug
458 clozapine.
459
460 We identify robust psychosis-associated differences in cellular composition estimates derived from
461 DNA methylation data, with cases having increased proportions of monocytes and granulocytes and
462 decreased proportions of natural killer cells, $CD4^+$ T-cells and $CD8^+$ T-cells compared to non-
463 psychiatric controls. This analysis extends previous work based on a subset of these data, which
464 reported a decrease in the proportion of natural killer cells and increase in the proportion of
465 granulocytes in schizophrenia patients, with the large number of samples enabling us to identify
466 additional associations with other cell types. We also confirm findings from an independent study of
467 schizophrenia which reported significantly increased proportions of granulocytes and monocytes, and
468 decreased proportions of $CD8^+$ T-cells using estimates derived from DNA methylation data (Montano
469 et al., 2016). Of note, because we can only derive proportion of cell types from whole blood DNA
470 methylation data, and not actual counts, an increase in one or more cell types must be balanced by a
471 decrease in one or more other cell types and an apparent change in the proportion of one specific cell
472 type does not mean that the actual abundance of that cell type is altered. Despite this, the results from
473 DNA methylation-derived cell proportions are consistent with previous studies based on empirical
474 cell abundance measures which have reported increased monocyte counts(Beumer et al., 2012;
475 Moody & Miller, 2018), increased neutrophil counts(Garcia-Rizo et al., 2019; Núñez et al., 2019),

increased monocyte to lymphocyte ratio(Mazza, Lucchi, Rossetti, & Clerici, 2019; Steiner et al., 2019) and increased neutrophil to lymphocyte ratio (Karageorgiou, Milas, & Michopoulos, 2019; Mazza et al., 2019) in both schizophrenia and FEP patients compared to controls. Previous studies have also shown that higher neutrophil counts in schizophrenia patients correlate with a greater burden of positive symptoms(Núñez et al., 2019) suggesting that variations in the number of neutrophils is a potential marker of disease severity(Steiner et al., 2019). Our sub-analysis of treatment-resistant schizophrenia, which is associated with a higher number of positive symptoms (Bachmann et al., 2017), found that the increase in granulocytes was primarily driven by those with the more severe phenotype, supporting this hypothesis. Importantly, the differences we observe may actually reflect the effects of various antipsychotic medications that have been previously shown to influence cell proportions in blood(Steiner et al., 2019) or a recruitment bias whereby patients with low levels of granulocytes are not prescribed clozapine given the risk of agranulocytosis.

We also identified a highly-significant increase in a DNA methylation-derived smoking score in patients with schizophrenia, replicating our previous finding (Hannon et al., 2016). The smoking score captures multiple aspects of tobacco smoking behaviour including both current smoking status and the quantity of cigarettes smoked; our results therefore reflect existing epidemiological evidence demonstrating that schizophrenia patients not only smoke more, but also smoke more heavily (de Leon, Becoña, Gurpegui, Gonzalez-Pinto, & Diaz, 2002; de Leon & Diaz, 2005; McClave, McKnight-Eily, Davis, & Dube, 2010). We also report an increased smoking score in patients with FEP, although not to the same extent as seen in schizophrenia, consistent with a meta-analysis reporting high levels of smoking in FEP (Myles et al., 2012). In the subset of treatment-resistant patients, we found that there was an additional increase in smoking score relative to schizophrenia cases prescribed alternative medications, supporting evidence for higher rates of smoking in TRS groups relative to treatment-responsive schizophrenia patients(Kennedy, Altar, Taylor, Degtiar, & Hornberger, 2014). These results not only highlight physiological (i.e. cell proportions) and environmental (i.e. smoking) differences associated with psychosis and schizophrenia and the utility of DNA methylation data for deriving these variables in epidemiological studies, but also highlight

the importance of controlling for these differences as potential confounders in analyses of disease-associated DNA methylation differences.

Our epigenome-wide association study, building on our previous analysis on a subset of the sample cohorts profiled here (Hannon et al., 2016), identified 95 DMPs associated with psychosis that are robust to differences in measured smoking exposure and heterogeneity in blood cellular composition derived from DNA methylation data. Of note, we identified a dramatic increase in sites characterized by an increase in DNA methylation in patients. A key strength of our study is the inclusion of the full spectrum of schizophrenia diagnoses, from FEP through to treatment-resistant cases prescribed clozapine. While this may introduce heterogeneity into our primary analyses, we used a random effects meta-analysis to identify consistent effects across all cohorts and diagnostic subtypes. We also performed an additional analysis focused specifically on cases with a more refined diagnosis of schizophrenia excluding those with FEP, which identified over 1,000 DMPs. A number of the top-ranked DMPs are annotated to genes that have direct relevance to the etiology of schizophrenia and gene ontology (GO) analysis highlighted multiple pathways previously implicated in schizophrenia including several related to the extracellular matrix (Berretta, 2012) and cell-cell adhesion (O'Dushlaine et al., 2011). Given the known genetic component to the etiology of schizophrenia, it is interesting that schizophrenia-associated DMPs were found to colocalize to several regions nominated by genetic association studies. Our results suggest that this analysis of a more specific phenotype in a smaller number of samples is potentially more powerful and that schizophrenia cases have a more discrete molecular phenotype that might reflect both etiological factors but also factors associated with a diagnosis of schizophrenia (e.g. medications, stress, etc). The mean difference in DNA methylation between cases and controls for both psychosis and schizophrenia was small, consistent with other blood-based EWAS of schizophrenia (Montano et al., 2016) and complex traits (Hannon, Schendel, et al., 2018; Hannon, Schendel, et al., 2019; Marioni et al., 2018) in general. While individually they may be too small to have a strong predictive power as a biomarker, together they may have utility as a molecular classifier (Chen et al., 2020).

To explore whether schizophrenia-associated differences in DNA methylation are present before a formal diagnosis of schizophrenia we also performed an EWAS of individuals with first-episode psychosis. Strikingly, the majority of our schizophrenia-associated DMPs were found to have a consistent direction of effect in the EWAS of individuals with FEP. While this result is consistent with schizophrenia-associated differences being present prior to a formal diagnosis of schizophrenia, it is not sufficient to state that they are causal; they may still reflect some underlying environmental risk factors or be a consequence of having FEP (e.g. medication exposure or other psychiatric condition). Further work is needed to explore the extent to which the DMPs associated with psychosis and schizophrenia in this meta-analysis might have a causal role in disease.

Finally, we also report the first systematic analysis of individuals with TRS, identifying seven DMPs at which differential DNA methylation was significantly different in the subset of schizophrenia cases prescribed clozapine. These data are informative for the interpretation of our schizophrenia-associated differences, because a number of these DMPs are driven by the subset of patients on clozapine. Furthermore, a number of sites show opposite effects in our analyses of TRS vs our analysis of schizophrenia, suggesting they might represent important differences between diagnostic groups. Because the prescription of clozapine is generally only undertaken in patients with treatment-resistant schizophrenia, we are unable to separate the effects of clozapine exposure from differences associated with a more severe sub-type of schizophrenia such as the influence of polypharmaceutical treatment.

Our results should be considered in light of a number of important limitations. First, our analyses were constrained by the technical limitations of the Illumina 450K and EPIC arrays which only assays ~ 3% of CpG sites in the genome. Second, this is a cross-sectional study and was not possible to distinguish cause from effect. It is possible, and indeed likely, for example, that the differences associated with both schizophrenia and TRS reflect the effects of medication exposure or other consequences of having schizophrenia, e.g. living more stressful lives, poorer diet and health. The importance of such confounding variables is demonstrated by our findings of differential smoking score and blood cell proportions derived directly from the DNA methylation data, although these

examples also highlight the potential utility of such effects for molecular epidemiology. Third, although our aim was not to make inferences about mechanistic changes in the brain associated with psychosis, it is important to note that our study analyzed DNA methylation profiled in peripheral blood and therefore can provide only limited information about variation in the primary tissue associated with disease(Hannon, Lunnon, Schalkwyk, & Mill, 2015). Although this limits mechanistic conclusions about the role of DNA methylation in schizophrenia, biomarkers, by definition, need to be measured in an easily accessible tissue and don't need to reflect the underlying pathogenic process. Furthermore, because most classifiers used to quantify variables such as smoking exposure and age have been trained in blood, this represents the optimal tissue in which to derive these measures. Of course, blood may also be an appropriate choice for investigating medication effects, particularly given the known effects on white blood cell counts associated with taking clozapine(Alvir, Lieberman, Safferman, Schwimmer, & Schaaf, 1993). Fourth, while we have explored the potential effects of clozapine on DNA methylation by assessing a sub-group of individuals with TRS, this is just one of a range of antipsychotics schizophrenia and psychosis patients are prescribed. The fact that the TRS group show more extreme differences for many of the schizophrenia-associated DMPs suggests that the polypharmaceutical treatment regimens often prescribed to schizophrenia patients may produce specific DNA methylation signatures in patients, akin to the effect seen for smoking. Fifth, although we found no evidence for a significant interaction between sex and DNA methylation at DMPs associated with schizophrenia, it is possible that there are other DNA methylation differences associated with disease only in males or females. Finally, although we found some evidence that schizophrenia-associated DMPs colocalize to regions nominated by GWAS, the integration of our DNA methylation data with genetic data was beyond the scope of this analysis. Of note, we have previously used mQTL associations to identify discrete sites of regulatory variation associated with schizophrenia risk variants to prioritize specific genes within broad GWAS regions (Hannon et al., 2016; Hannon, Gorrie-Stone, et al., 2018; Hannon, Spiers, et al., 2015; Hannon et al., 2017) and future work will aim to further explore explore interactions between genetic and epigenetic risk factors.

In conclusion, our analysis of 4,483 participants represents the largest study of blood-based DNA-methylation in schizophrenia and psychosis yet performed, and one of the largest EWAS studies for *any* disease phenotype. Our study also includes the first within-case analysis of treatment-resistant schizophrenia yet performed, providing important molecular insights into genomic differences associated with poor outcome to standard therapeutic approaches. Our results highlight differences in measures of blood cellular composition and smoking behaviour derived from methylomic dats between not just cases and controls, but also between treatment-resistant schizophrenia patients prescribed clozapine and those prescribed alternative medications. We report widespread differences in DNA methylation in psychosis and schizophrenia, a subset of which are driven by the more severe treatment-resistant subset of patients. On a practical level, our study demonstrates the utility of DNA methylation data for deriving measures of specific physiological phenotypes (e.g. blood cell-type proportions) and environmental exposures (e.g. exposure to tobacco smoke) that can be used to identify epidemiological associations with health and disease, but also highlights the importance of properly controlling for these potential confounders in EWAS analyses. Our results are important because they suggest there are also clear molecular signatures of schizophrenia and psychosis that can be identified in whole blood DNA. Although it is unlikely these differences are mechanistically related to neuropathological changes in the brain, they may have utility as diagnostic and prognostic biomarkers in individuals with FEP and may potentially be used to differentiate individuals with TRS at an early stage of disease. Future work should aim to prospectively profile DNA methylation in individuals at risk for FEP and schizophrenia to explore how methylomic variation at baseline can predict outcome and the extent to which longitudinal changes at psychosis-associated DMPs map on to clinical trajectories.

Materials and Methods:

Cohort descriptions

University College London (UCL) samples

447 schizophrenia cases and 456 controls from the University College London schizophrenia sample cohort were selected for DNA methylation profiling. A full description of this cohort can be found elsewhere(Datta et al., 2010) but briefly comprises of unrelated ancestrally matched cases and controls from the United Kingdom. Case participants were recruited from UK NHS mental health services with a clinical ICD-10 diagnosis of schizophrenia. All case participants were interviewed with the Schedule for Affective Disorders and Schizophrenia-Lifetime Version (SADS-L)(Spitzer & Endicott, 1977) to confirm Research Diagnostic Criteria (RDC) diagnosis. A control sample screened for an absence of mental health problems was recruited. Each control subject was interviewed to confirm that they did not have a personal history of an RDC defined mental disorder or a family history of schizophrenia, bipolar disorder, or alcohol dependence. UK National Health Service multicentre and local research ethics approval was obtained and all subjects signed an approved consent form after reading an information sheet.

Aberdeen samples

482 schizophrenia cases and 468 controls from the Aberdeen schizophrenia sample were selected for DNA methylation profiling. The Aberdeen case-control sample has been fully described elsewhere (International Schizophrenia Consortium, 2008) but briefly contains schizophrenia cases and controls who have self-identified as born in the British Isles (95% in Scotland). All cases met the Diagnostic and Statistical Manual for Mental Disorders-IV edition (DSM-IV) and International Classification of Diseases 10th edition (ICD-10) criteria for schizophrenia. Diagnosis was made by Operational Criteria Checklist (OPCRIT). Controls were volunteers recruited through general practices in Scotland. Practice lists were screened for potentially suitable volunteers by age and sex and by exclusion of subjects with major mental illness or use of neuroleptic medication. Volunteers who replied to a written invitation were interviewed using a short questionnaire to exclude major mental

illness in individual themselves and first-degree relatives. All cases and controls gave informed consent. The study was approved by both local and multiregional academic ethical committees.

Monozygotic twins discordant for schizophrenia

The monozygotic twin cohort is a multi-centre collaborative project aimed at identifying DNA methylation differences in monozygotic-twin pairs discordant for a diagnosis of schizophrenia. 96 informative twin-pairs (n = 192 individuals) were identified from European twin studies based in Utrecht (The Netherlands), Helsinki (Finland), London (United Kingdom), Stockholm (Sweden), and Jena (Germany). Of the monozygotic twin pairs utilized in the analysis, 75 were discordant for diagnosed schizophrenia, 6 were concordant for schizophrenia and 15 twin pairs were free of any psychiatric disease. Each twin study has been approved; ethical permission was given by the relevant local ethics committee and the participating twins have provided written informed consent.

Dublin samples

361 schizophrenia cases and 346 controls were selected from the Irish Schizophrenia Genomics consortium, a detailed description of this cohort can be found in the Morris et al manuscript (Morris et al., 2014). Briefly, participants, from the Republic of Ireland or Northern Ireland, were interviewed using a structured clinical interview and diagnosis of schizophrenia or a related disorder [schizoaffective disorder; schizophreniform disorder] was made by the consensus lifetime best estimate method using DSM-IV criteria. Control subjects were ascertained with written informed consent from the Irish GeneBank and represented blood donors from the Irish Blood Transfusion Service. Ethics Committee approval for the study was obtained from all participating hospitals and centres.

IoPPN samples

The IoPPN cohort comprises of 290 schizophrenia cases, 308 first episode psychosis (FEP) patients and 203 non-psychiatric controls recruited from the same geographical area into three studies via the South London & Maudsley Mental Health National Health Service (NHS) Foundation Trust.

Established schizophrenia cases were recruited to the Improving Physical Health and Reducing Substance Use in Severe Mental Illness (IMPACT) study from three English mental health NHS services (Gaughran et al., 2019). First episode psychosis patients were recruited to the GAP study(Di Forti et al., 2015) via in-patient and early intervention in psychosis community mental health teams. All patients aged 18–65 years who presented with a first episode of psychosis to the Lambeth, Southwark and Croydon adult in-patient units of the South London & Maudsley Mental Health NHS Foundation Trust between May 1, 2005, and May 31, 2011 who met ICD–10 criteria for a diagnosis of psychosis (codes F20–F29 and F30–F33). Clinical diagnosis was validated by administering the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). Cases with a diagnosis of organic psychosis were excluded. Healthy controls were recruited into the GAP study from the local population living in the area served by the South London & Maudsley Mental Health NHS Foundation Trust, by means of internet and newspaper advertisements, and distribution of leaflets at train stations, shops and job centres. Those who agreed to participate were administered the Psychosis Screening Questionnaire(Bebbington & Nayani, 1995) and excluded if they met criteria for a psychotic disorder or reported to have received a previous diagnosis of psychotic illness. All participants were included in the study only after giving written, confirmed consent. The study protocol and ethical permission was granted by the Joint South London and Maudsley and the Institute of Psychiatry NHS Research Ethics Committee (17/NI/0011).

Sweden

190 schizophrenia cases and 190 controls from the Sweden Schizophrenia Study (S3) [31] were selected for DNA methylation profiling details of which have been described previously [2]. Briefly, S3 is a population-based cohort of individuals born in Sweden including 4,936 SCZ cases and 6,321 healthy controls recruited between 2004 and 2010. SCZ cases were identified from the Sweden Hospital Discharge Register [32, 33] with ≥2 hospitalizations with an ICD discharge diagnosis of SCZ or schizoaffective disorder (SAD) [34]. This operational definition of SCZ was validated in clinical, epidemiological, genetic epidemiological, and genetic studies [31]. More generally, the Hospital Discharge Register has high agreement with medical [32, 33] and psychiatric diagnoses [35]. Controls

were also selected through Swedish Registers and were group-matched by age, sex and county of residence and had no lifetime diagnoses of SCZ, SAD, or bipolar disorder or antipsychotic prescriptions. Blood samples were drawn at enrolment. All subjects were 18 years of age or older and provided written informed consent. Ethical permission was obtained from the Karolinska Institutet Ethical Review Committee in Stockholm, Sweden.

The European Network of National Schizophrenia Networks Studying Gene-Environment Interactions (EU-GEI) cohort

458 first-episode psychosis (FEP) cases and 558 controls from the incidence and case-control work package (WP2) of the European Network of National Schizophrenia Networks Studying Gene-Environment Interactions (EU-GEI) cohort were selected for DNA methylation profiling (Jongsma et al., 2018). Patients presenting with FEP were identified, between 1/5/2010 and 1/4/2015, by trained researchers who carried out regular checks across the 17 catchment area Mental Health Services across 6 European countries. FEP were included if a) age 18-64 years and b) resident within the study catchment areas at the time of their first presentation, and with a diagnosis of psychosis (ICD-10 F20-33). Using the Operational Criteria Checklist algorithm (McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin, Farmer, & Harvey, 1991; Quattrone et al., 2018)(McGuffin et al., 1991, Quattrone et al., 2018) all cases interviewed received a research-based diagnosis. FEPs were excluded if a) previously treated for psychosis, b) they met criteria for organic psychosis (ICD-10: F09), or for a diagnosis of transient psychotic symptoms resulting from acute intoxication (ICD-10: F1X.5). FEP

were approached via their clinical team and invited to participate in the assessment. Random and Quota sampling strategies were adopted to guide the recruitment of controls from each of the sites. The most accurate local demographic data available were used to set quotas for controls to ensure the samples' representativeness of each catchment area's population at risk. Controls were excluded if they had received a diagnosis of and/or treatment for, a psychotic disorder. All participants provided informed, written consent. Ethical approval was provided by relevant research ethics committees in each of the study sites.

Genome-wide quantification of DNA methylation

Approximately 500ng of blood-derived DNA from each sample was treated with sodium bisulfite in duplicate, using the EZ-96 DNA methylation kit (Zymo Research, CA, USA). DNA methylation was quantified using either the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc, CA, USA) or Illumina Infinium HumanMethylationEPIC BeadChip (Illumina Inc, CA, USA) run on an Illumina iScan System (Illumina, CA, USA) using the manufacturers' standard protocol. Samples were batched by cohort and randomly assigned to chips and plates to ensure equal distribution of cases and controls across arrays and minimize batch effects. For the monozygotic Twin cohort, both members of the same twin pair were run on the same chip. A fully methylated control sample (CpG Methylated HeLa Genomic DNA; New England BioLabs, MA, USA) was included in a random position on each plate to facilitate plate tracking. Signal intensities were imported in R programming environment using the *methylumIDAT* function in the *methylumi* package (Davis, Du, Bilke, Triche, & Bootwalla, 2015). Our stringent quality control pipeline included the following steps: 1) checking methylated and unmethylated signal intensities, excluding samples where this was < 2500; 2) using the control probes to ensure the sodium bisulfite conversion was successful, excluding any samples with median < 90; 3) identifying the fully methylated control sample was in the correct location; 4) all tissues predicted as of blood origin using the tissue prediction from the Epigenetic Clock software (https://DNAmAge.genetics.ucla.edu/) (Horvath, 2013); 5) multidimensional scaling of sites on X and Y chromosomes separately to confirm reported gender; 6) comparison with genotype data across SNP probes; 7) *pfilter* function from wateRmelon package (Pidsley et al., 2013) to exclude samples with >

1% of probes with detection *P*-value > 0.05 and probes with > 1% of samples with detection *P*-value > 0.05. PCs were used (calculated across all probes) to identify outliers, samples > 2 standard deviations from the mean for both PC1 and PC2 were removed. An additional QC step was performed in the Twins cohort using the 65 SNP probes to confirm that twins were genetically identical. Normalization of the DNA methylation data was performed used the *dasen* function in the *wateRmelon* package(Pidsley et al., 2013). As cell count data were not available for these DNA samples these were estimated from the 450K DNA methylation data using both the Epigenetic Clock software (Horvath, 2013) and Houseman algorithm (Houseman et al., 2012; Koestler et al., 2013), including the seven variables recommended in the documentation for the Epigenetic Clock in the regression analysis. For cohorts with the EPIC array DNA methylation data, we were only able to generate the six cellular composition variables using the Houseman algorithm(Houseman et al., 2012; Koestler et al., 2013), which were included as covariates. Similarly as smoking data was incomplete for the majority of cohorts, we calculated a smoking score from the data using the method described by Elliot et al(Elliott et al., 2014) and successfully used in our previous (Phase 1) analyses(Hannon et al., 2016). Raw and processed data for the UCL, Aberdeen, Dublin, IoPPN and EU-GEI cohorts are available through GEO accession numbers GSE84727, GSE80417, GSE147221, GSE152027 and GSE152026 respectively.

Data analysis

All analyses were performed with the statistical language R unless otherwise stated. Custom code for all steps of the analysis are available on GitHub: (https://github.com/ejh243/SCZEWAS/tree/master/Phase2).

Comparison of estimates of cellular composition and tobacco smoking derived from DNA methylation data

A linear regression model was used to test for differences in ten cellular composition variables estimated from the DNA methylation data, reflecting either proportion or abundance of blood cell types. These estimated cellular composition variables were regressed against case/control status with

779 covariates for age, sex and smoking. Estimated effects and standard errors were combined across the
780 cohorts using a random effect meta-analysis implemented with the meta package(Schwarzer, 2007).
781 The same methodology was used to test for differences in the smoking score derived from DNA
782 methylation data between cases and controls including covariates for age and sex. P values are from
783 two-sided tests.
784
785 *Within-cohort EWAS analysis*
786 A linear regression model was used to test for differentially methylated sites associated with
787 schizophrenia or first episode psychosis. DNA methylation values for each probe were regressed
788 against case/control status with covariates for age, sex, derived cellular composition scores (from the
789 DNA methylation data), derived smoking score (from the DNA methylation data) and experimental
790 batch. For the EU-GEI cohort there was an additional covariate for contributing study. For the Twins
791 cohort, a linear model was used to generate regression coefficients, but *P*-values were calculated with
792 clustered standards errors using the *plm* package (Croissant & Millo, 2008), recognising individuals
793 from the same twin pair.
794
795 *Within-patient EWAS of clozapine prescription*
796 Four individual cohorts (UCL, Aberdeen, IoPPN and Sweden) had information on medication and/or
797 clozapine exposure and were included in the treatment-resistant schizophrenia (TRS) EWAS. TRS
798 patients were defined as any case that had ever been prescribed clozapine, and non-TRS patients were
799 defined as schizophrenia cases that had no record of being prescribed clozapine. Within each cohort
800 DNA methylation values for each probe were regressed against TRS status with covariates for age,
801 sex, cell composition, smoking status, and batch as described for the case control EWAS.
802
803 *Multiple regression analysis of schizophrenia and clozapine prescription*
804 Using the four cohorts that were included in the TRS EWAS (UCL, Aberdeen, IoPPN and Sweden),
805 we fitted a multiple regression model with two binary indicator variables: one that identified the
806 schizophrenia patients and a second that identified the TRS schizophrenia patients. Within each

cohort DNA methylation values for each probe were regressed against these two binary variables, with covariates for age, sex, derived cellular composition scores (from the DNA methylation data), derived smoking score (from the DNA methylation data) and experimental batch as described above for the other EWAS analyses.

Meta-analysis

The EWAS results from each cohort were processed using the *bacon* R package(van Iterson et al., 2017), which uses a Bayesian method to adjust for inflation in EWAS P-values. All probes analysed in at least two studies were taken forward for meta-analysis. This was performed using the *metagen* function in the R package meta(Schwarzer, 2007), using the effect sizes and standard errors adjusted for inflation from each individual cohort to calculate weighted pooled estimates and test for significance. P-values are from two-sided tests and significant DMPs were identified from a random effects model at a significance threshold of $9x10^{-8}$, which controls for the number of independent tests performed when analysis data generated with the EPIC array(Mansell et al., 2019). DNA methylation sites were annotated with location information for genome build hg19 using the Illumina manifest files (CHR and MAPINFO).

Overlap with schizophrenia GWAS loci

The GWAS regions were taken from the largest published schizophrenia GWAS to date by Pãrdinas and colleagues (Pardiñas et al., 2018) made available through the Psychiatric Genomics Consortium (PGC) website (https://www.med.unc.edu/pgc/results-and-downloads). Briefly, regions were defined by performing a "clumping" procedure on the GWAS *P*-values to collapse multiple correlated signals (due to linkage disequilibrium) surrounding the index SNP (i.e. with the smallest P-value) into a single associated region. To define physically distinct loci, those within 250kb of each other were subsequently merged to obtain the final set of GWAS regions. The outermost SNPs of each associated region defined the start and stop parameters of the region. Using the set of 158 schizophrenia-associated genomic loci we used Brown's method (Brown, 1975) to calculate a combined P-value across all probes located within each region (based on hg19) using the probe-level P-values and

correlation coefficients between all pairs of probes calculated from the DNA methylation values. Briefly, correlation statistics were calculated and (along with the P values) were inputted into Brown's formula. As correlations between probes could only be calculated using probes profiled on the same array, this analysis was limited to probes included on the EPIC array. Correlations between probes were calculated within the EU-GEI cohort as this had the largest number of samples.

Enrichment analyses

Enrichment of the heritability statistics of DMPs was performed against a background set of probes selected to match the distribution of the test set for both mean and standard deviation. This was achieved by splitting all probes into 10 equally sized bins based on their mean methylation level and ten equally sized bins based on their standard deviation, to create a matrix of 100 bins. After counting the number of DMPs within each bin, we selected the same number of probes from each bin for the background comparison set. This was repeated multiple times, without replacement, until all the probes from at least one bin were selected giving the maximum possible number of background probes (n = 42,968) such that they matched the characteristics of the test set of DMPs.

Gene ontology (GO) analysis

Illumina UCSC gene annotation, which is derived from the genomic overlap of probes with RefSeq genes or up to 1500bp of the transcription start site of a gene, was used to create a test gene list from the DMPs for pathway analysis. Where probes were not annotated to any gene (i.e. in the case of intergenic locations) they were omitted from this analysis, and where probes were annotated to multiple genes, all were included. A logistic regression approach was used to test if genes in this list predicted pathway membership, while controlling for the number of probes that passed quality control (i.e. were tested) annotated to each gene. Pathways were downloaded from the GO website (http://geneontology.org/) and mapped to genes including all parent ontology terms. All genes with at least one 450K probe annotated and mapped to at least one GO pathway were considered. Pathways were filtered to those containing between 10 and 2000 genes. After applying this method to all pathways, the list of significant pathways ($P < 0.05$) was refined by grouping to control for the effect

of overlapping genes. This was achieved by taking the most significant pathway, and retesting all remaining significant pathways while controlling additionally for the best term. If the test genes no longer predicted the pathway, the term was said to be explained by the more significant pathway, and hence these pathways were grouped together. This algorithm was repeated, taking the next most significant term, until all pathways were considered as the most significant or found to be explained by a more significant term.

869 **Figure Legends**

870

871 **Figure 1: Overview of the sample cohorts and analytical approaches used in this study of altered**
872 **DNA methylation in psychosis and schizophrenia.**

873

874 **Figure 1 – supplement 1: Forest plot showing the difference in mean age between psychosis**
875 **cases and controls across each cohort.** TE – treatment effect (i.e. the mean difference between cases
876 and controls), seTE – standard error of the treatment effect.

877

878 **Figure 1 – supplement 2**: **Scatterplot of the relationship between the first two genetic principal**
879 **components merged with HapMap Phase 3 data for individual cohorts.** With the exception of the
880 IoPPN and EUGEI cohorts, there is little ethnic heterogeneity in each of the cohorts with samples
881 being predominantly of Caucasian origin.

882

883 **Figure 1 – supplement 3**: **Scatterplots of DNAmAge derived from the DNA methylation data**
884 **against actual chronological age for each of the cohorts.** DNAmAge was calculated using the
885 algorithm described by Horvath (Horvath, 2013). Each point represents an individual and is coloured
886 by psychosis status (blue = psychosis, red = control). The solid diagonal line depicts x=y, i.e. where
887 the estimated and actual values are the same. The dashed diagonal line depicts the line of best fit.
888 Presented at the top of the graph is the Pearson's correlation coefficient (r) between the estimated and
889 actual age across all samples in that cohort. Also shown in the bottom right hand corner of each panel
890 is an interaction P value from a test for different correlations between DNAmAge and actual age
891 between psychosis cases and controls.

892

893 **Figure 1 – supplement 4: Scatterplots of PhenoAge derived from DNA methylation data against**
894 **actual chronological age for each of the cohorts.** PhenoAge was calculated using the algorithm
895 described by Levine et al. (Levine et al., 2018). Each point represents an individual and is coloured by
896 psychosis status (blue = psychosis, red = control). The solid diagonal line depicts x=y, i.e. where the

estimated and actual values are the same. The dashed diagonal line depicts the line of best fit. Presented at the top of the graph is the Pearson's correlation coefficient (r) between the estimated and actual age across all samples in that cohort. Also shown in the bottom right hand corner of each panel is an interaction P value from a test for different correlations between PhenoAge and actual age between psychosis cases and controls.

Figure 1 – supplement 5: Scatterplots of DNAmAge derived from the DNA methylation data against actual chronological age for each of the cohorts. DNAmAge was calculated using the algorithm described by Horvath (Horvath, 2013). Each point represents an individual and is coloured by medication status (yellow = schizophrenia cases not prescribed clozapine, green = treatment-resistant schizophrenia cases prescribed clozapine). The solid diagonal line depicts x=y, i.e. where the estimated and actual values are the same. The dashed diagonal line depicts the line of best fit. Presented at the top of the graph is the Pearson's correlation coefficient (r) between the estimated and actual age across all samples in that cohort. Also shown in the bottom right hand corner of each panel is an interaction P value from a test for different correlations between DNAmAge and actual age for schizophrenia patients prescribed clozapine and schizophrenia patients prescribed alternative medications.

Figure 1 – supplement 6: Scatterplots of PhenoAge derived from the DNA methylation data against actual chronological age for each of the cohorts. PhenoAge was calculated using the algorithm described by (Levine et al., 2018). Each point represents an individual and is coloured by schizophrenia status (yellow = schizophrenia cases not prescribed clozapine, green = treatment-resistant schizophrenia cases prescribed clozapine). The solid diagonal line depicts x=y, i.e. where the estimated and actual values are the same. The dashed diagonal line depicts the line of best fit. Presented at the top of the graph is the Pearson's correlation coefficient (r) between the estimated and actual age across all samples in that cohort. Also shown in the bottom right hand corner of each panel is an interaction P value from a test for different correlations between PhenoAge and actual age for

schizophrenia patients prescribed clozapine and schizophrenia patients prescribed alternative medications.

Figure 2 Blood cell-type proportions derived from DNA methylation data are altered in psychosis. Shown are forest plots from meta-analyses of differences in blood cell proportions derived from DNA methylation data between psychosis patients and controls for **A**) monocytes **B)** granulocytes **C)** natural killer cells **D**) CD4+ T-cells and **E**) CD8+ T-cells. TE – treatment effect (i.e. the mean difference between cases and controls), seTE – standard error of the treatment effect.

Figure 2 – supplement 1: Treatment-resistant schizophrenia patients prescribed clozapine are characterized by altered blood cell proportions. Shown are forest plots from meta-analyses of differences in estimated blood cell proportions derived from DNA methylation data between treatment-resistant schizophrenia patients prescribed clozapine and schizophrenia patients prescribed other medications for granulocytes, CD8+ T-cells. TE – treatment effect (i.e. the mean difference between cases and controls), seTE – standard error of the treatment effect.

Figure 2 – supplement 2: Additive effect of schizophrenia and treatment-resistance on granulocyte proportions. Shown are forest plots from meta-analyses of differences in estimated granulocyte proportions derived from DNA methylation data between **A)** schizophrenia patients and controls and **B)** treatment-resistant schizophrenia patients prescribed clozapine and schizophrenia patients prescribed other medications. TE – treatment effect (i.e. the mean difference between cases and controls), seTE – standard error of the treatment effect.

Figure 2 – supplement 3: Additive effect of schizophrenia and treatment-resistance on CD8+ T-cell proportions. Shown are forest plots from meta-analyses of differences in estimated granulocyte proportions derived from DNA methylation data between **A)** schizophrenia patients and controls and **B)** treatment-resistant schizophrenia patients prescribed clozapine and schizophrenia patients

prescribed other medications. TE – treatment effect (i.e. the mean difference between cases and controls), seTE – standard error of the treatment effect.

Figure 3: Smoking scores derived from DNA methylation data highlight that psychosis patients are characterized by an elevated exposure to tobacco smoking. Forest plot from a meta-analysis of differences in smoking score derived from DNA methylation data between psychosis patients and controls. The smoking score was calculated from DNA methylation data using the method described by Elliott and colleagues (Elliott et al., 2014). TE – treatment effect (i.e. the mean difference between cases and controls), seTE – standard error of the treatment effect.

Figure 3 – supplement 1: Current and former smokers are characterized by a significantly higher smoking score derived from DNA methylation data than non-smokers. Shown is the DNA methylation smoking score (y-axis) from individuals in the IoPPN cohort for whom self-reported smoking data was available regarding current (left panel) and former (right panel) smoking behavior. 0 = no, 1 = yes.

Figure 3 – supplement 2: Treatment resistant schizophrenia is associated with significantly higher DNA methylation-derived smoking scores. Forest plot from meta-analyses of differences in smoking derived from DNA methylation data between treatment-resistant schizophrenia patients prescribed clozapine and schizophrenia patients prescribed other medications. TE – treatment effect (i.e. the mean difference between cases and controls), seTE – standard error of the treatment effect.

Figure 3 – supplement 3: Treatment-resistant schizophrenia patients show an elevated exposure to tobacco smoking relative to non-treatment-resistant schizophrenia and controls in a model testing both schizophrenia diagnosis status and TRS status simultaneously. A) schizophrenia diagnosis was associated with a significant increase in smoking score (mean difference = 3.98, P = 2.19x10-8) with **B)** TRS status associated with an additional increase within cases (mean difference =

978 2.15, P = 2.22x10-7). TE – treatment effect (i.e. the mean difference between cases and controls),
979 seTE – standard error of the treatment effect.
980
981 **Figure 4: Differential DNA methylation at multiple loci across the genome is associated with**
982 **psychosis and schizophrenia.** Manhattan plots depicting the –log10 P value from the EWAS meta-
983 analysis (y-axis) against genomic location (x-axis). Panel **A)** presents results from the analysis
984 comparing psychosis patients and controls, and panel **B)** presents results from the analysis comparing
985 diagnosed schizophrenia cases and controls.
986
987 **Figure 4 – supplement 1: Including genetic principal components PCs into DNA methylation**
988 **analysis models has little effect on the results in ethnically heterogeneous cohorts.** Shown is a
989 scatterplot of statistics (–log10(P-value)) from an EWAS of psychosis in the IoPPN cohort without
990 the inclusion of any genetic principal components in the analysis model (x-axis) compared to an
991 EWAS of psychosis including five genetic principal components in the analysis model (y-axis).
992
993 **Figure 5: Psychosis-associated differential DNA methylation at sites annotated to genes**
994 **previously implicated in disease etiology.** Shown are forest plots for DMPs annotated to the GABA
995 transporter *SLC6A12* (cg00517261, P = $1.53x10^{-8}$), the GABA receptor *GABBR1* (cg00667298, P =
996 $5.07x10^{-9}$), and the calcium voltage-gated channel subunit gene *CACNA1C* (cg01833890, P =
997 $8.42x10^{-9}$). TE – treatment effect (i.e. the mean difference between cases and controls), seTE –
998 standard error of the treatment effect.
999
1000 **Figure 6: Comparison of effect sizes for schizophrenia-associated DMPs overlapping with**
1001 **EWAS results for other traits.** Shown for each overlapping DMP is the association effect size for
1002 the other trait (x-axis) taken from the online EWAS catalog (http://ewascatalog.org/) compared to the
1003 effect size identified in our meta-analysis of schizophrenia (y-axis).
1004

Figure 7: DNA methylation at sites associated with schizophrenia is more strongly influenced by genetic factors and common environmental influences than equivalent matched sites across the genome. A series of density plots for estimates of additive genetic effects (A, left), common environmental effects (C, middle), and non-shared environmental effects (E, right) derived using data from a dataset generated by Hannon and colleagues (Hannon, Knox, et al., 2018) schizophrenia DMPs (red) and matched background sites (green).

Figure 8: Differences in DNA methylation between schizophrenia cases and controls are partially influenced by a subset of cases with treatment resistant schizophrenia. Forest plots from a meta-analysis of differences in DNA methylation at cg16322565 located in the *NR1L2* gene on chromosome 3 between **A**) schizophrenia patients and controls and **B**) TRS patients prescribed clozapine and non-TRS prescribed other medications. TE – treatment effect (i.e. the mean difference between cases and controls), seTE – standard error of the treatment effect.

Figure 8 – supplement 1: Forest plot of a site where DNA methylation is significantly associated with schizophrenia and within cases, with treatment-resistant schizophrenia. TE – treatment effect (i.e. the mean difference between cases and controls), seTE – standard error of the treatment effect.

Figure 8 – supplement 2: Forest plot of a site where DNA methylation is significantly associated with schizophrenia and within cases, with treatment-resistant schizophrenia but with an opposite directions of effect. TE – treatment effect (i.e. the mean difference between cases and controls), seTE – standard error of the treatment effect.

1028 **Tables**

Cohort		UCL	Aberdeen	Twins	IoPPN	Dublin	EU-GEI	Sweden	Total
Total sample		675	847	192	800	679	912	378	4483
% Cases		52.3	48.9	45.3	74.6	51.3	42.9	50.0	53.1
% Schizophrenia		52.3	48.9	45.3	36.3	51.3	0.0	50.0	37.5
% First episode psychosis		0.0	0.0	0.0	38.4	0.0	42.9	0.0	15.6
% Males	All	58.7	71.1	52.1	63.0	71.0	54.4	59.5	62.6
	Cases	72.0	68.4	54.0	65.3	71.6	64.2	60.3	66.8
	Controls	44.1	73.7	50.5	56.2	70.4	47.0	58.7	57.8
	Chi-square test P value	3.81E-13	0.103	0.730	0.024	0.804	3.68E-07	0.834	9.35E-10
Age (years)	Mean	40.4	44.6	35.3	28.8	41.7	35.3	60.0	40.5
	SD	15.0	12.9	10.8	9.46	12.0	12.8	8.86	14.7
	Mean in controls	43.7	44.2	37.9	27.8	41.4	30.7	56.3	41.6
	Mean in cases	36.8	44.9	33.3	30.3	42.0	38.7	63.7	39.4
	T-test P value	6.55E-09	0.529	0.033	0.007	0.505	1.24E-22	1.05E-16	

1029

1030 **Table 1. Summary of cohort demographics included in the psychosis EWAS meta-analysis.**

Cell type	Measure type	Number of cohorts	Random effects model			Fixed effects model			Heterogeneity P value
			Mean difference	SE	P value	Mean difference	SE	P value	
Monocytes	Proportion	7	0.00320	0.00083	0.000115	0.00320	0.00083	0.000115	0.6490
Granulocytes	Proportion	7	0.04312	0.01241	0.000509	0.03930	0.00315	1.21E-35	2.22E-16
Natural Killer cells	Proportion	7	-0.01135	0.00385	0.003221	-0.00827	0.00133	4.48E-10	2.43E-08
CD4+ T-cells	Proportion	7	-0.01767	0.00555	0.00144	-0.01569	0.00196	1.15E-15	1.23E-07
CD8+ T-cells	Proportion	7	-0.01444	0.00457	0.001586	-0.01443	0.00148	1.31E-22	8.13E-10
B-cells	Proportion	7	-0.00495	0.00280	0.077103	-0.00477	0.00102	2.75E-06	2.25E-07
PlasmaBlast	Abundance	5	0.05626	0.02987	0.059671	0.05332	0.00722	1.55E-13	8.45E-13
CD8pCD28nCD45RAn	Abundance	5	0.06280	0.22674	0.781792	0.10797	0.14981	0.4711	0.0826
CD8.naive T-cells	Abundance	5	7.21687	3.12594	0.02096	8.03957	1.89169	2.14E-05	0.0443
CD4.naive T-cells	Abundance	5	11.77240	4.72532	0.012726	11.77240	4.72532	0.0127	0.824

Table 2. Results of a meta-analysis of differences in blood cell compositionestimates derived from DNA methylation data between schizophrenia cases and controls.

Acknowledgments

This work was primarily supported by grants from the UK Medical Research Council (MRC; MR/K013807/1 and MR/R005176/1) to J.M. High-performance computing was supported by MRC Clinical Research Infrastructure Funding (MR/M008924/1). The Finnish Twin study was supported by the Academy of Finland Centre of Excellence in Complex Disease Genetics (grant numbers: 213506, 129680), and J.K. by the Academy of Finland grants 265240, 263278 and 312073. Financial support for the Sweden twin study was provided by the Karolinska Institutet (ALF 20090183 and ALF 20100305 to Hultman) and NIH (R01 MH52857). Collection of the Sweden case control samples was supported by the Sweden Research Council (Vetenskapsrådet, award D0886501 to PFS) and the NIMH (R01MH077139). Collection of the Irish case control samples was funded by the Wellcome Trust Case Control Consortium 2 project (085475/B/08/Z and 085475/Z/08/Z), the Wellcome Trust (072894/Z/03/Z, 090532/Z/09/Z and 075491/Z/04/B), and Science Foundation Ireland (08/IN.1/B1916). The European Network of National Schizophrenia Networks Studying Gene-Environment Interactions (EU-GEI) Project is funded by grant agreement HEALTH-F2-2010-241909 (Project EU-GEI) from the European Community's Seventh Framework Programme. The IMPaCT programme at King's College London and the South London and Maudsley NHS Foundation Trust is funded by the National Institute for Health Research (RP-PG-0606-1049). The CRESTAR project received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement 279227 (CRET AR Consortium). EH, ED, LS and JM were supported by MRC grant K013807 to JM. Cardiff University researchers were supported by Medical Research Council (MRC) Centre (G0800509) and Programme Grant (G0801418). Bart PF Rutten is supported by a VIDI grant (number 91718336) from the Netherlands Organisation for Scientific Research. FG is in part supported by the National Institute for Health Research's (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, the Stanley Medical Research Institute, the Maudsley Charity and the National Institute for Health Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. MDF and DQ are funded by an MRC fellowship to MDF (MR/M008436/1). We gratefully acknowledge capital

equipment funding from the Maudsley Charity (Grant Ref. 980) and Guy's and St Thomas's Charity (Grant Ref. R130505). This study presents independent research supported by the National Institute for Health Research NIHR BioResource Centre Maudsley at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, NIHR, Department of Health and Social Care or King's College London.

Disclosures. DC is a full-time employee and stockholder of Eli Lilly and Company. FG has received honoraria from Lundbeck, Otsuka, and Sunovion, and has a family member with professional links to Lilly and GSK, including shares. KK has consulted with Emerald Lake Safety Ltd. (2017-2018) and has received speaker honoraria from Biogen/Fraser Health Multiple Sclerosis Clinic (2018). MDF has received personal fees from Janssen. MOD is supported by a collaborative research grant from Takeda Pharmaceuticals. PS has received research funding from Lundbeck and has served or is currently serving on the scientific advisory board of Pfizer and Lundbeck. RM reports personal fees from Janssen, Lundbeck, Sunovion, Recordati and Otsuka. JM has received research funding from Eli Lilly and Company. JMac has received research funding from Lundbeck. All of these relationships are outside the remit of the submitted work.EH, GM, MB, TD, GD, VJ, JK, CM, AM, DM, IN, DQ, TT, JV, JW, LS, ED, JB, NB, AC, CC, DD, MG, AG, CG, HH, CH, RK, GK, KM, CM, AR, BR, DS, GB, and JM report no financial relationships with commercial interests.

Supplementary Files

Supplementary File 1 – Supplementary Tables 1-14

References

Aberg, K. A., McClay, J. L., Nerella, S., Clark, S., Kumar, G., Chen, W., . . . van den Oord, E. J. (2014). Methylome-wide association study of schizophrenia: identifying blood biomarker signatures of environmental insults. *JAMA Psychiatry, 71*(3), 255-264. doi:10.1001/jamapsychiatry.2013.3730

Ajnakina, O., Horsdal, H. T., Lally, J., MacCabe, J. H., Murray, R. M., Gasse, C., & Wimberley, T. (2018). Validation of an algorithm-based definition of treatment resistance in patients with schizophrenia. *Schizophr Res, 197*, 294-297. doi:10.1016/j.schres.2018.02.017

Aleman, A., Kahn, R. S., & Selten, J. P. (2003). Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Arch Gen Psychiatry, 60*(6), 565-571. doi:10.1001/archpsyc.60.6.565

Alvir, J. M., Lieberman, J. A., Safferman, A. Z., Schwimmer, J. L., & Schaaf, J. A. (1993). Clozapine-induced agranulocytosis. Incidence and risk factors in the United States. *N Engl J Med, 329*(3), 162-167. doi:10.1056/NEJM199307153290303

Bachmann, C. J., Aagaard, L., Bernardo, M., Brandt, L., Cartabia, M., Clavenna, A., . . . Taylor, D. (2017). International trends in clozapine use: a study in 17 countries. *Acta Psychiatr Scand, 136*(1), 37-51. doi:10.1111/acps.12742

Bebbington, P., & Nayani, T. (1995). The Psychosis Screening Questionnaire. *Int J Methods Psychiatr Res*(5), 11–19.

Berretta, S. (2012). Extracellular matrix abnormalities in schizophrenia. *Neuropharmacology, 62*(3), 1584-1597. doi:10.1016/j.neuropharm.2011.08.010

Beumer, W., Gibney, S. M., Drexhage, R. C., Pont-Lezica, L., Doorduin, J., Klein, H. C., . . . Drexhage, H. A. (2012). The immune theory of psychiatric diseases: a key role for activated microglia and circulating monocytes. *J Leukoc Biol, 92*(5), 959-975. doi:10.1189/jlb.0212100

Brown, M. B. (1975). A Method for Combining Non-Independent, One-Sided Tests of Significance. *Biometrics, 31*(4), 987-992.

Chen, J., Zang, Z., Braun, U., Schwarz, K., Harneit, A., Kremer, T., . . . Schwarz, E. (2020). Association of a Reproducible Epigenetic Risk Profile for Schizophrenia With Brain Methylation and Function. *JAMA Psychiatry*. doi:10.1001/jamapsychiatry.2019.4792

1115 Consortium, C.-D. G. o. t. P. G. (2013). Identification of risk loci with shared effects on five major
1116 psychiatric disorders: a genome-wide analysis. *Lancet, 381*(9875), 1371-1379.
1117 doi:10.1016/S0140-6736(12)62129-1
1118 Croissant, Y., & Millo, G. (2008). Panel Data Econometrics in R: The plm Package. *Journal of*
1119 *Statistical Software, 27*(2). Retrieved from http://www.jstatsoft.org/v27/i02/
1120 Datta, S. R., McQuillin, A., Rizig, M., Blaveri, E., Thirumalai, S., Kalsi, G., . . . Gurling, H. M.
1121 (2010). A threonine to isoleucine missense mutation in the pericentriolar material 1 gene is
1122 strongly associated with schizophrenia. *Mol Psychiatry, 15*(6), 615-628.
1123 doi:10.1038/mp.2008.128
1124 Davis, S., Du, P., Bilke, S., Triche, J., & Bootwalla, M. (2015). *methylumi: Handle Illumina*
1125 *methylation data. R package version 2.14.0.*
1126 de Leon, J., Becoña, E., Gurpegui, M., Gonzalez-Pinto, A., & Diaz, F. J. (2002). The association
1127 between high nicotine dependence and severe mental illness may be consistent across
1128 countries. *J Clin Psychiatry, 63*(9), 812-816. Retrieved from
1129 http://www.ncbi.nlm.nih.gov/pubmed/12363123
1130 de Leon, J., & Diaz, F. J. (2005). A meta-analysis of worldwide studies demonstrates an association
1131 between schizophrenia and tobacco smoking behaviors. *Schizophr Res, 76*(2-3), 135-157.
1132 doi:10.1016/j.schres.2005.02.010
1133 Dempster, E., Viana, J., Pidsley, R., & Mill, J. (2013). Epigenetic studies of schizophrenia: progress,
1134 predicaments, and promises for the future. *Schizophr Bull, 39*(1), 11-16.
1135 doi:10.1093/schbul/sbs139
1136 Dempster, E. L., Pidsley, R., Schalkwyk, L. C., Owens, S., Georgiades, A., Kane, F., . . . Mill, J.
1137 (2011). Disease-associated epigenetic changes in monozygotic twins discordant for
1138 schizophrenia and bipolar disorder. *Hum Mol Genet, 20*(24), 4786-4796.
1139 doi:10.1093/hmg/ddr416
1140 Di Forti, M., Marconi, A., Carra, E., Fraietta, S., Trotta, A., Bonomo, M., . . . Murray, R. M. (2015).
1141 Proportion of patients in south London with first-episode psychosis attributable to use of high

potency cannabis: a case-control study. *Lancet Psychiatry, 2*(3), 233-238. doi:10.1016/S2215-0366(14)00117-5

Di Forti, M., Morgan, C., Dazzan, P., Pariante, C., Mondelli, V., Marques, T. R., . . . Murray, R. M. (2009). High-potency cannabis and the risk of psychosis. *Br J Psychiatry, 195*(6), 488-491. doi:10.1192/bjp.bp.109.064220

Elliott, H. R., Tillin, T., McArdle, W. L., Ho, K., Duggirala, A., Frayling, T. M., . . . Relton, C. L. (2014). Differences in smoking associated DNA methylation patterns in South Asians and Europeans. *Clin Epigenetics, 6*(1), 4. doi:10.1186/1868-7083-6-4

Garcia-Rizo, C., Casanovas, M., Fernandez-Egea, E., Oliveira, C., Meseguer, A., Cabrera, B., . . . Bernardo, M. (2019). Blood cell count in antipsychotic-naive patients with non-affective psychosis. *Early Interv Psychiatry, 13*(1), 95-100. doi:10.1111/eip.12456

Gaughran, F., Stahl, D., Stringer, D., Hopkins, D., Atakan, Z., Greenwood, K., . . . team, I. (2019). Effect of lifestyle, medication and ethnicity on cardiometabolic risk in the year following the first episode of psychosis: prospective cohort study. *Br J Psychiatry*, 1-8. doi:10.1192/bjp.2019.159

Hannon, E., Dempster, E., Viana, J., Burrage, J., Smith, A. R., Macdonald, R., . . . Mill, J. (2016). An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biology, 17*(1), 176. doi:10.1186/s13059-016-1041-x

Hannon, E., Gorrie-Stone, T. J., Smart, M. C., Burrage, J., Hughes, A., Bao, Y., . . . Mill, J. (2018). Leveraging DNA-Methylation Quantitative-Trait Loci to Characterize the Relationship between Methylomic Variation, Gene Expression, and Complex Traits. *Am J Hum Genet, 103*(5), 654-665. doi:10.1016/j.ajhg.2018.09.007

Hannon, E., Knox, O., Sugden, K., Burrage, J., Wong, C. C. Y., Belsky, D. W., . . . Mill, J. (2018). Characterizing genetic and environmental influences on variable DNA methylation using monozygotic and dizygotic twins. *PLoS Genet, 14*(8), e1007544. doi:10.1371/journal.pgen.1007544

1169 Hannon, E., Lunnon, K., Schalkwyk, L., & Mill, J. (2015). Interindividual methylomic variation
1170 across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and
1171 neuropsychiatric phenotypes. *Epigenetics, 10*(11), 1024-1032.
1172 doi:10.1080/15592294.2015.1100786
1173 Hannon, E., Marzi, S. J., Schalkwyk, L. S., & Mill, J. (2019). Genetic risk variants for brain disorders
1174 are enriched in cortical H3K27ac domains. *Mol Brain, 12*(1), 7. doi:10.1186/s13041-019-
1175 0429-4
1176 Hannon, E., Schendel, D., Ladd-Acosta, C., Grove, J., Hansen, C. S., Andrews, S. V., . . . Group, i.-B.
1177 A. (2018). Elevated polygenic burden for autism is associated with differential DNA
1178 methylation at birth. *Genome Med, 10*(1), 19. doi:10.1186/s13073-018-0527-4
1179 Hannon, E., Schendel, D., Ladd-Acosta, C., Grove, J., Hansen, C. S., Hougaard, D. M., . . . Group, i.-
1180 B. A. (2019). Variable DNA methylation in neonates mediates the association between
1181 prenatal smoking and birth weight. *Philos Trans R Soc Lond B Biol Sci, 374*(1770),
1182 20180120. doi:10.1098/rstb.2018.0120
1183 Hannon, E., Spiers, H., Viana, J., Pidsley, R., Burrage, J., Murphy, T. M., . . . Mill, J. (2015).
1184 Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci.
1185 *Nat Neurosci*. doi:10.1038/nn.4182
1186 Hannon, E., Weedon, M., Bray, N., O'Donovan, M., & Mill, J. (2017). Pleiotropic Effects of Trait-
1187 Associated Genetic Variation on DNA Methylation: Utility for Refining GWAS Loci. *Am J*
1188 *Hum Genet*. doi:10.1016/j.ajhg.2017.04.013
1189 Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., . . . Zhang, K. (2013). Genome-
1190 wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell, 49*(2),
1191 359-367. doi:10.1016/j.molcel.2012.10.016
1192 Hilker, R., Helenius, D., Fagerlund, B., Skytthe, A., Christensen, K., Werge, T. M., . . . Glenthøj, B.
1193 (2018). Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide
1194 Danish Twin Register. *Biol Psychiatry, 83*(6), 492-498. doi:10.1016/j.biopsych.2017.08.017
1195 Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biol, 14*(10),
1196 R115. doi:10.1186/gb-2013-14-10-r115

1197 Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H.,
1198 . . . Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture
1199 distribution. *BMC Bioinformatics, 13*, 86. doi:10.1186/1471-2105-13-86
1200 International Schizophrenia Consortium. (2008). Rare chromosomal deletions and duplications
1201 increase risk of schizophrenia. *Nature, 455*(7210), 237-241. doi:10.1038/nature07239
1202 Jaffe, A. E., Gao, Y., Deep-Soboslay, A., Tao, R., Hyde, T. M., Weinberger, D. R., & Kleinman, J. E.
1203 (2015). Mapping DNA methylation across development, genotype and schizophrenia in the
1204 human frontal cortex. *Nat Neurosci*. doi:10.1038/nn.4181
1205 Jongsma, H. E., Gayer-Anderson, C., Lasalvia, A., Quattrone, D., Mulè, A., Szöke, A., . . . Group, E.
1206 N. o. N. S. N. S. G.-E. I. W. P. E.-G. W. (2018). Treated Incidence of Psychotic Disorders in
1207 the Multinational EU-GEI Study. *JAMA Psychiatry, 75*(1), 36-46.
1208 doi:10.1001/jamapsychiatry.2017.3554
1209 Karageorgiou, V., Milas, G. P., & Michopoulos, I. (2019). Neutrophil-to-lymphocyte ratio in
1210 schizophrenia: A systematic review and meta-analysis. *Schizophr Res, 206*, 4-12.
1211 doi:10.1016/j.schres.2018.12.017
1212 Kennedy, J. L., Altar, C. A., Taylor, D. L., Degtiar, I., & Hornberger, J. C. (2014). The social and
1213 economic burden of treatment-resistant schizophrenia: a systematic literature review. *Int Clin*
1214 *Psychopharmacol, 29*(2), 63-76. doi:10.1097/YIC.0b013e32836508e6
1215 Kinoshita, M., Numata, S., Tajima, A., Ohi, K., Hashimoto, R., Shimodera, S., . . . Ohmori, T. (2014).
1216 Aberrant DNA methylation of blood in schizophrenia by adjusting for estimated cellular
1217 proportions. *Neuromolecular Med, 16*(4), 697-703. doi:10.1007/s12017-014-8319-5
1218 Koestler, D. C., Christensen, B., Karagas, M. R., Marsit, C. J., Langevin, S. M., Kelsey, K. T., . . .
1219 Houseman, E. A. (2013). Blood-based profiles of DNA methylation predict the underlying
1220 distribution of cell types: a validation analysis. *Epigenetics, 8*(8), 816-826.
1221 doi:10.4161/epi.25430
1222 Kowalec, K., Hannon, E., Mansell, G., Burrage, J., Ori, A. P. S., Ophoff, R. A., . . . Sullivan, P. F.
1223 (2019). Methylation age acceleration does not predict mortality in schizophrenia. *Transl*
1224 *Psychiatry, 9*(1), 157. doi:10.1038/s41398-019-0489-3

Le-Niculescu, H., Balaraman, Y., Patel, S., Tan, J., Sidhu, K., Jerome, R. E., . . . Niculescu, A. B. (2007). Towards understanding the schizophrenia code: an expanded convergent functional genomics approach. *Am J Med Genet B Neuropsychiatr Genet, 144B*(2), 129-158. doi:10.1002/ajmg.b.30481

Levine, M. E., Lu, A. T., Quach, A., Chen, B. H., Assimes, T. L., Bandinelli, S., . . . Horvath, S. (2018). An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY), 10*(4), 573-591. doi:10.18632/aging.101414

Mansell, G., Gorrie-Stone, T. J., Bao, Y., Kumari, M., Schalkwyk, L. S., Mill, J., & Hannon, E. (2019). Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics, 20*(1), 366. doi:10.1186/s12864-019-5761-7

Marioni, R. E., McRae, A. F., Bressler, J., Colicino, E., Hannon, E., Li, S., . . . Deary, I. J. (2018). Meta-analysis of epigenome-wide association studies of cognitive abilities. *Mol Psychiatry*. doi:10.1038/s41380-017-0008-y

Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., . . . Costello, J. F. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature, 466*(7303), 253-257. doi:10.1038/nature09165

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., . . . Stamatoyannopoulos, J. A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science, 337*(6099), 1190-1195. doi:10.1126/science.1222794

Mazza, M. G., Lucchi, S., Rossetti, A., & Clerici, M. (2019). Neutrophil-lymphocyte ratio, monocyte-lymphocyte ratio and platelet-lymphocyte ratio in non-affective psychosis: A meta-analysis and systematic review. *World J Biol Psychiatry*, 1-13. doi:10.1080/15622975.2019.1583371

McClave, A. K., McKnight-Eily, L. R., Davis, S. P., & Dube, S. R. (2010). Smoking characteristics of adults with selected lifetime mental illnesses: results from the 2007 National Health Interview Survey. *Am J Public Health, 100*(12), 2464-2472. doi:10.2105/AJPH.2009.188136

McGuffin, P., Farmer, A., & Harvey, I. (1991). A polydiagnostic application of operational criteria in studies of psychotic illness: Development and reliability of the opcrit system. *Archives of General Psychiatry, 48*(8), 764-770. doi:10.1001/archpsyc.1991.01810320088015

1253 Mill, J., & Heijmans, B. T. (2013). From promises to practical strategies in epigenetic epidemiology.
1254 *Nat Rev Genet, 14*(8), 585-594. doi:10.1038/nrg3405
1255 Montano, C., Taub, M. A., Jaffe, A., Briem, E., Feinberg, J. I., Trygvadottir, R., . . . Feinberg, A. P.
1256 (2016). Association of DNA Methylation Differences With Schizophrenia in an Epigenome-
1257 Wide Association Study. *JAMA Psychiatry, 73*(5), 506-514.
1258 doi:10.1001/jamapsychiatry.2016.0144
1259 Moody, G., & Miller, B. J. (2018). Total and differential white blood cell counts and hemodynamic
1260 parameters in first-episode psychosis. *Psychiatry Res, 260*, 307-312.
1261 doi:10.1016/j.psychres.2017.11.086
1262 Morris, D. W., Pearson, R. D., Cormican, P., Kenny, E. M., O'Dushlaine, C. T., Perreault, L. P., . . . 2,
1263 W. T. C. C. C. (2014). An inherited duplication at the gene p21 Protein-Activated Kinase 7
1264 (PAK7) is a risk factor for psychosis. *Hum Mol Genet, 23*(12), 3316-3326.
1265 doi:10.1093/hmg/ddu025
1266 Murphy, T. M., & Mill, J. (2014). Epigenetics in health and disease: heralding the EWAS era. *Lancet,*
1267 *383*(9933), 1952-1954. doi:10.1016/S0140-6736(14)60269-5
1268 Myles, N., Newall, H. D., Curtis, J., Nielssen, O., Shiers, D., & Large, M. (2012). Tobacco use before,
1269 at, and after first-episode psychosis: a systematic meta-analysis. *J Clin Psychiatry, 73*(4),
1270 468-475. doi:10.4088/JCP.11r07222
1271 Núñez, C., Stephan-Otto, C., Usall, J., Bioque, M., Lobo, A., González-Pinto, A., . . . group, P.
1272 (2019). Neutrophil Count Is Associated With Reduced Gray Matter and Enlarged Ventricles
1273 in First-Episode Psychosis. *Schizophr Bull, 45*(4), 846-858. doi:10.1093/schbul/sby113
1274 O'Dushlaine, C., Kenny, E., Heron, E., Donohoe, G., Gill, M., Morris, D., . . . Consortium, I. S.
1275 (2011). Molecular pathways involved in neuronal cell adhesion and membrane scaffolding
1276 contribute to schizophrenia and bipolar disorder susceptibility. *Mol Psychiatry, 16*(3), 286-
1277 292. doi:10.1038/mp.2010.7
1278 Pardiñas, A. F., Holmans, P., Pocklington, A. J., Escott-Price, V., Ripke, S., Carrera, N., . . .
1279 Consortium, C. (2018). Common schizophrenia alleles are enriched in mutation-intolerant

genes and in regions under strong background selection. *Nat Genet, 50*(3), 381-389. doi:10.1038/s41588-018-0059-2

Park, H. J., Kim, J. W., Lee, S. K., Kim, S. K., Park, J. K., Cho, A. R., . . . Song, J. Y. (2011). Association between the SLC6A12 gene and negative symptoms of schizophrenia in a Korean population. *Psychiatry Res, 189*(3), 478-479. doi:10.1016/j.psychres.2011.01.023

Pidsley, R., Viana, J., Hannon, E., Spiers, H. H., Troakes, C., Al-Saraj, S., . . . Mill, J. (2014). Methylomic profiling of human brain tissue supports a neurodevelopmental origin for schizophrenia. *Genome Biol, 15*(10), 483. doi:10.1186/PREACCEPT-1621721621132088

Pidsley, R., Y Wong, C. C., Volta, M., Lunnon, K., Mill, J., & Schalkwyk, L. C. (2013). A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics, 14*, 293. doi:10.1186/1471-2164-14-293

Psychiatric GWAS Consortium Bipolar Disorder Working Group. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet, 43*(10), 977-983. doi:10.1038/ng.943

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., . . . Consortium, I. S. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460*(7256), 748-752. doi:10.1038/nature08185

Quattrone, D., Di Forti, M., Gayer-Anderson, C., Ferraro, L., Jongsma, H. E., Tripoli, G., . . . Reininghaus, U. (2018). Transdiagnostic dimensions of psychopathology at first episode psychosis: findings from the multinational EU-GEI study. *Psychological Medicine*, 1-14. doi:10.1017/S0033291718002131

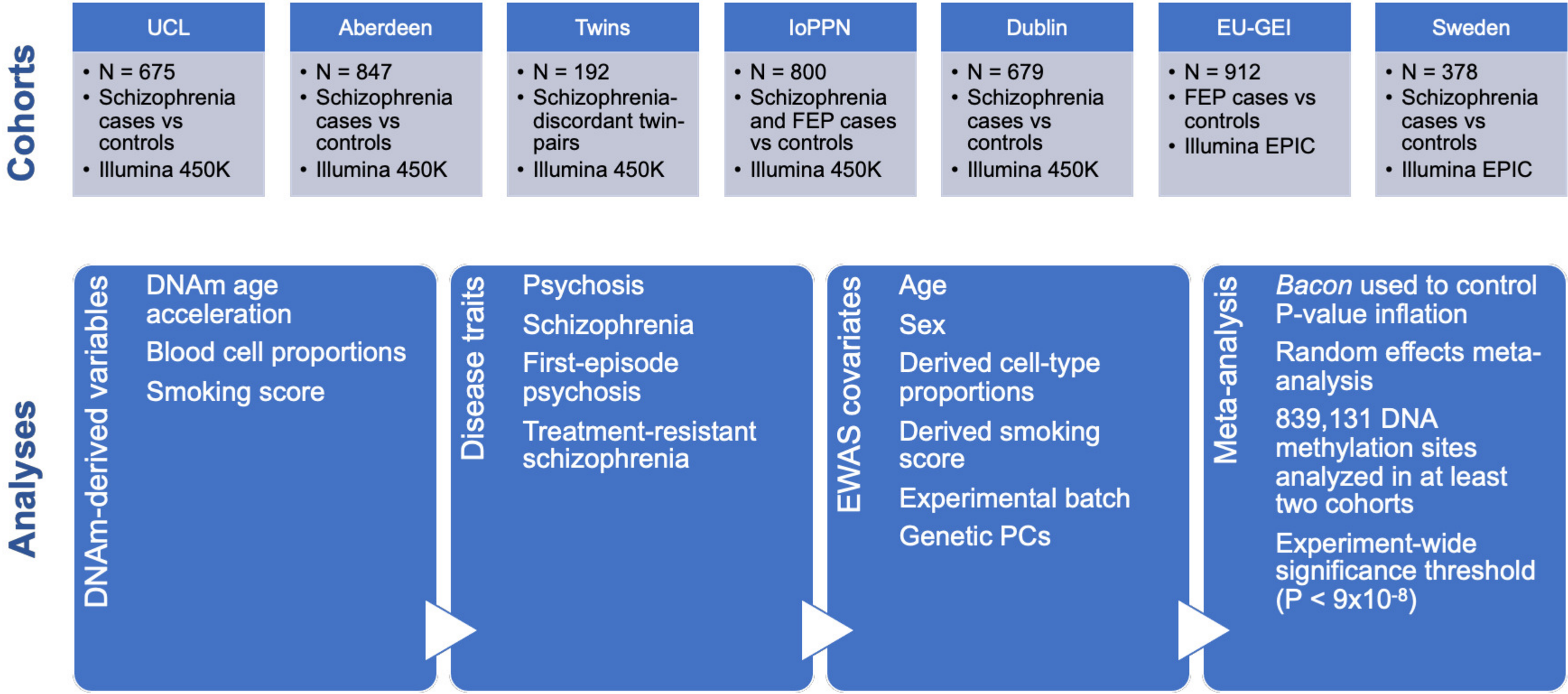
Rakyan, V. K., Down, T. A., Balding, D. J., & Beck, S. (2011). Epigenome-wide association studies for common human diseases. *Nat Rev Genet, 12*(8), 529-541. doi:10.1038/nrg3000

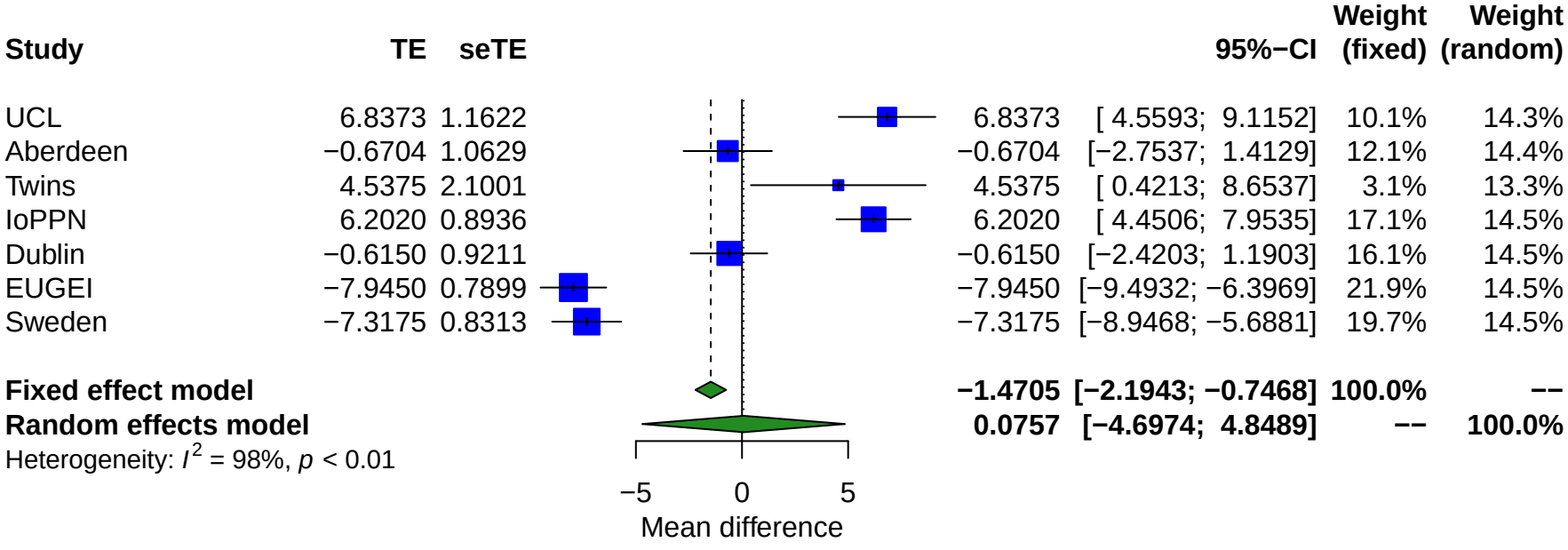
Relton, C. L., & Davey Smith, G. (2010). Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. *PLoS Med, 7*(10), e1000356. doi:10.1371/journal.pmed.1000356

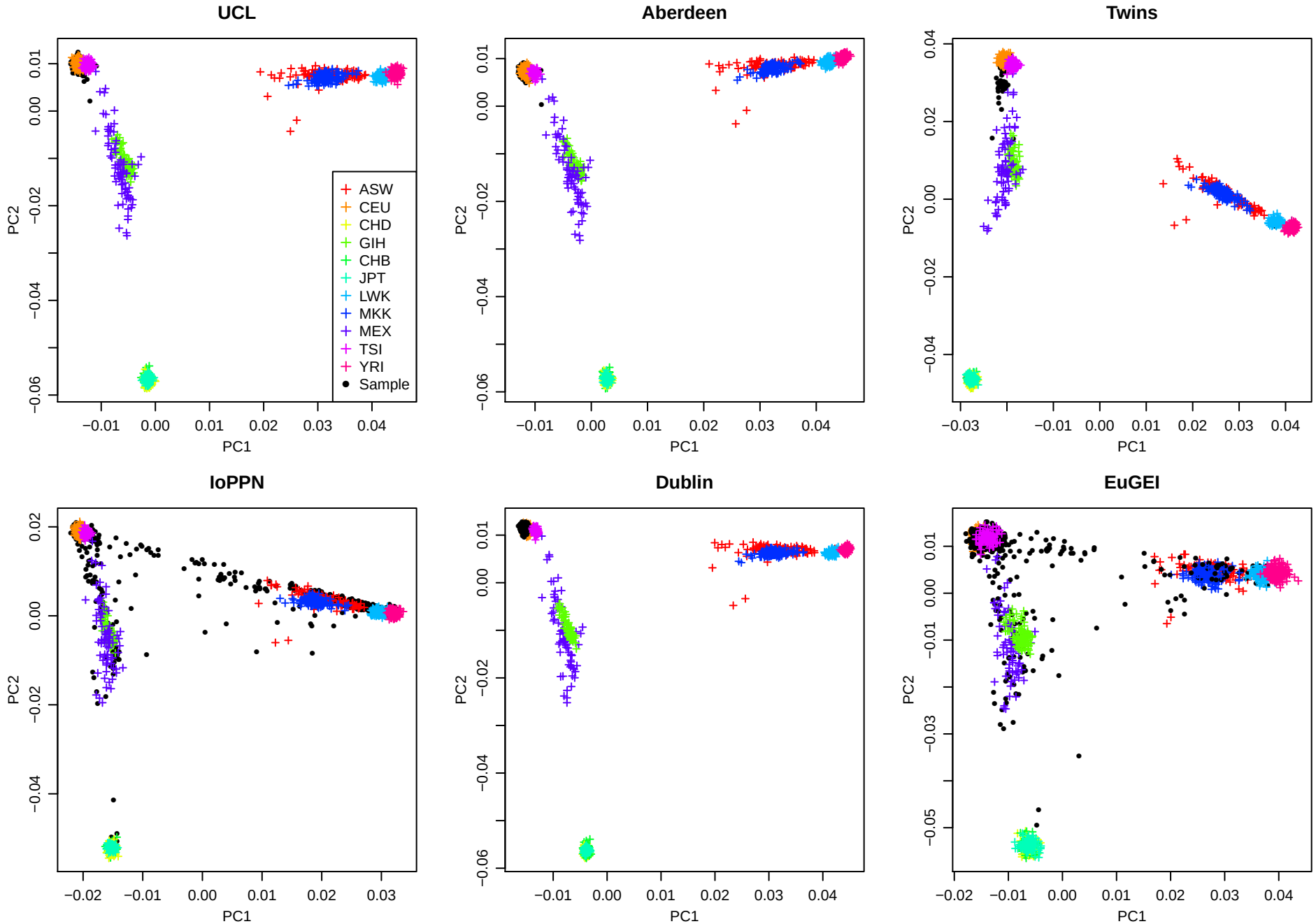
1306 Schizophrenia Working Group of the PGC, Ripke, S., Neale, B., Corvin, A., Walters, J., Farh, K., . . .
1307 Consor, W. T. C.-C. (2014). Biological insights from 108 schizophrenia-associated genetic
1308 loci. *Nature, 511*(7510), 421-+. doi:10.1038/nature13595
1309 Schizophrenia Working Group of the PGC, Ripke, S., Sanders, A., Kendler, K., Levinson, D., Sklar,
1310 P., . . . Genome-Wide, S. P. (2011). Genome-wide association study identifies five new
1311 schizophrenia loci. *Nature Genetics, 43*(10), 969-U977. doi:10.1038/ng.940
1312 Schwarzer, G. (2007). meta: An R Package for meta-analysis. *R News, 7*, 40-45.
1313 Spitzer, R., & Endicott, J. (1977). *The Schedule for Affective Disorders and Schizophrenia, Lifetime*
1314 *Version* (3 ed.). New York State Psychiatric Institute, New York.
1315 Steiner, J., Frodl, T., Schiltz, K., Dobrowolny, H., Jacobs, R., Fernandes, B. S., . . . Bernstein, H. G.
1316 (2019). Innate Immune Cells and C-Reactive Protein in Acute First-Episode Psychosis and
1317 Schizophrenia: Relationship to Psychopathology and Treatment. *Schizophr Bull*.
1318 doi:10.1093/schbul/sbz068
1319 Sugden, K., Hannon, E. J., Arseneault, L., Belsky, D. W., Broadbent, J. M., Corcoran, D. L., . . .
1320 Caspi, A. (2019). Establishing a generalized polyepigenetic biomarker for tobacco smoking.
1321 *Transl Psychiatry, 9*(1), 92. doi:10.1038/s41398-019-0430-9
1322 Sullivan, P. F., Kendler, K. S., & Neale, M. C. (2003). Schizophrenia as a complex trait: evidence
1323 from a meta-analysis of twin studies. *Arch Gen Psychiatry, 60*(12), 1187-1192.
1324 doi:10.1001/archpsyc.60.12.1187
1325 van der Werf, M., Hanssen, M., Köhler, S., Verkaaik, M., Verhey, F. R., van Winkel, R., . . .
1326 Investigators, R. (2014). Systematic review and collaborative recalculation of 133,693
1327 incident cases of schizophrenia. *Psychol Med, 44*(1), 9-16. doi:10.1017/S0033291712002796
1328 van Iterson, M., van Zwet, E. W., Heijmans, B. T., & Consortium, B. (2017). Controlling bias and
1329 inflation in epigenome- and transcriptome-wide association studies using the empirical null
1330 distribution. *Genome Biol, 18*(1), 19. doi:10.1186/s13059-016-1131-9
1331 Viana, J., Hannon, E., Dempster, E., Pidsley, R., Macdonald, R., Knox, O., . . . Mill, J. (2016).
1332 Schizophrenia-associated methylomic variation: molecular signatures of disease and

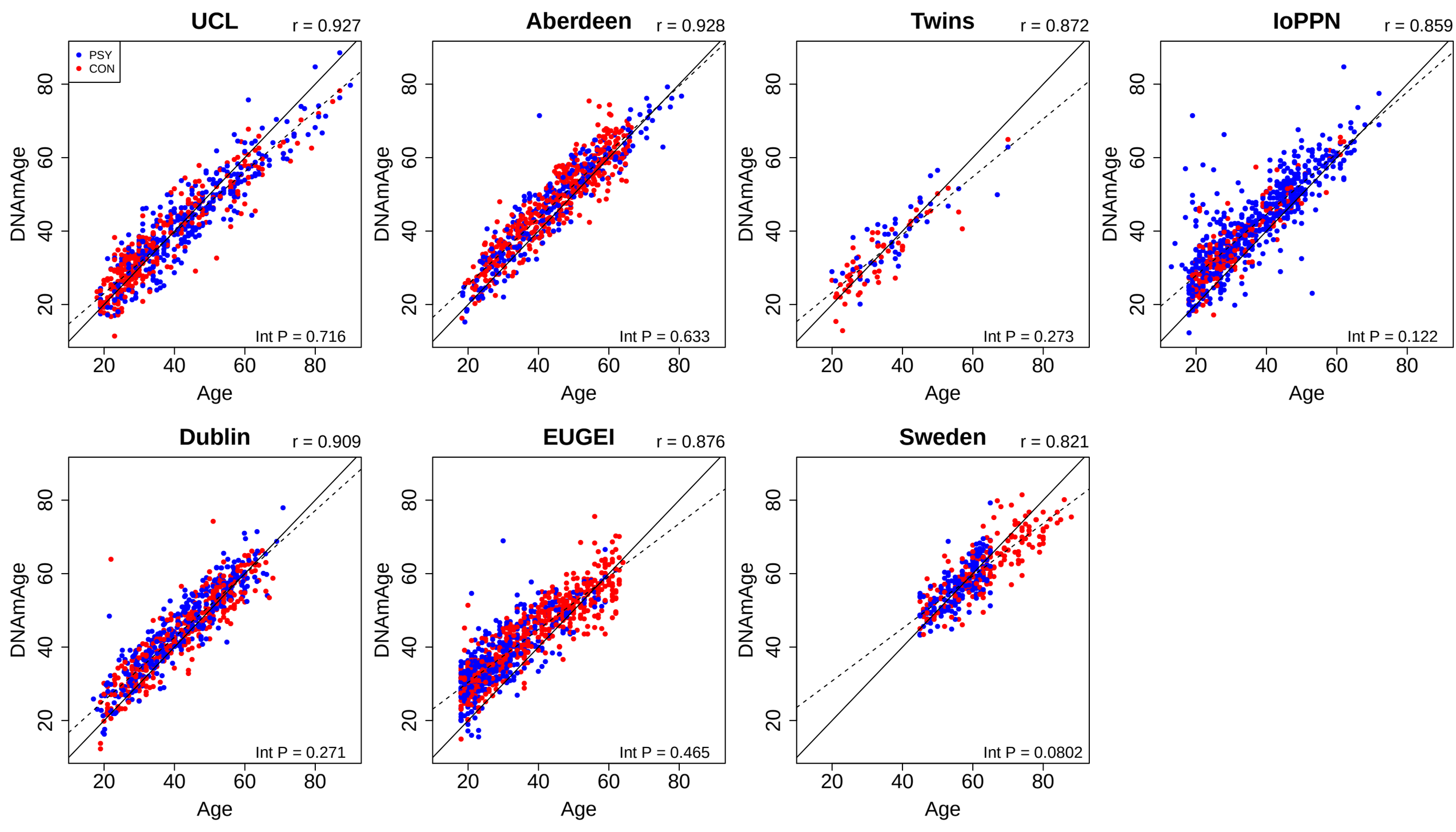
1333 polygenic risk burden across multiple brain regions. *Hum Mol Genet*.
1334 doi:10.1093/hmg/ddw373
1335 Viana, J., Hannon, E., Dempster, E., Pidsley, R., Macdonald, R., Knox, O., . . . Mill, J. (2017).
1336 Schizophrenia-associated methylomic variation: molecular signatures of disease and
1337 polygenic risk burden across multiple brain regions. *Hum Mol Genet, 26*(1), 210-225.
1338 doi:10.1093/hmg/ddw373
1339 Wagner, J. R., Busche, S., Ge, B., Kwan, T., Pastinen, T., & Blanchette, M. (2014). The relationship
1340 between DNA methylation, genetic and expression inter-individual variation in
1341 untransformed human fibroblasts. *Genome Biol, 15*(2), R37. doi:10.1186/gb-2014-15-2-r37
1342 Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., . . . Vos, T.
1343 (2013). Global burden of disease attributable to mental and substance use disorders: findings
1344 from the Global Burden of Disease Study 2010. *Lancet, 382*(9904), 1575-1586.
1345 doi:10.1016/S0140-6736(13)61611-6
1346 Wockner, L. F., Noble, E. P., Lawford, B. R., Young, R. M., Morris, C. P., Whitehall, V. L., &
1347 Voisey, J. (2014). Genome-wide DNA methylation analysis of human brain tissue from
1348 schizophrenia patients. *Transl Psychiatry, 4*, e339. doi:10.1038/tp.2013.111
1349 Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide
1350 complex trait analysis. *Am J Hum Genet, 88*(1), 76-82. doi:10.1016/j.ajhg.2010.11.011
1351 Zhang, Q., Vallerga, C. L., Walker, R. M., Lin, T., Henders, A. K., Montgomery, G. W., . . . Visscher,
1352 P. M. (2019). Improved precision of epigenetic clock estimates across tissues and its
1353 implication for biological ageing. *Genome Med, 11*(1), 54. doi:10.1186/s13073-019-0667-1

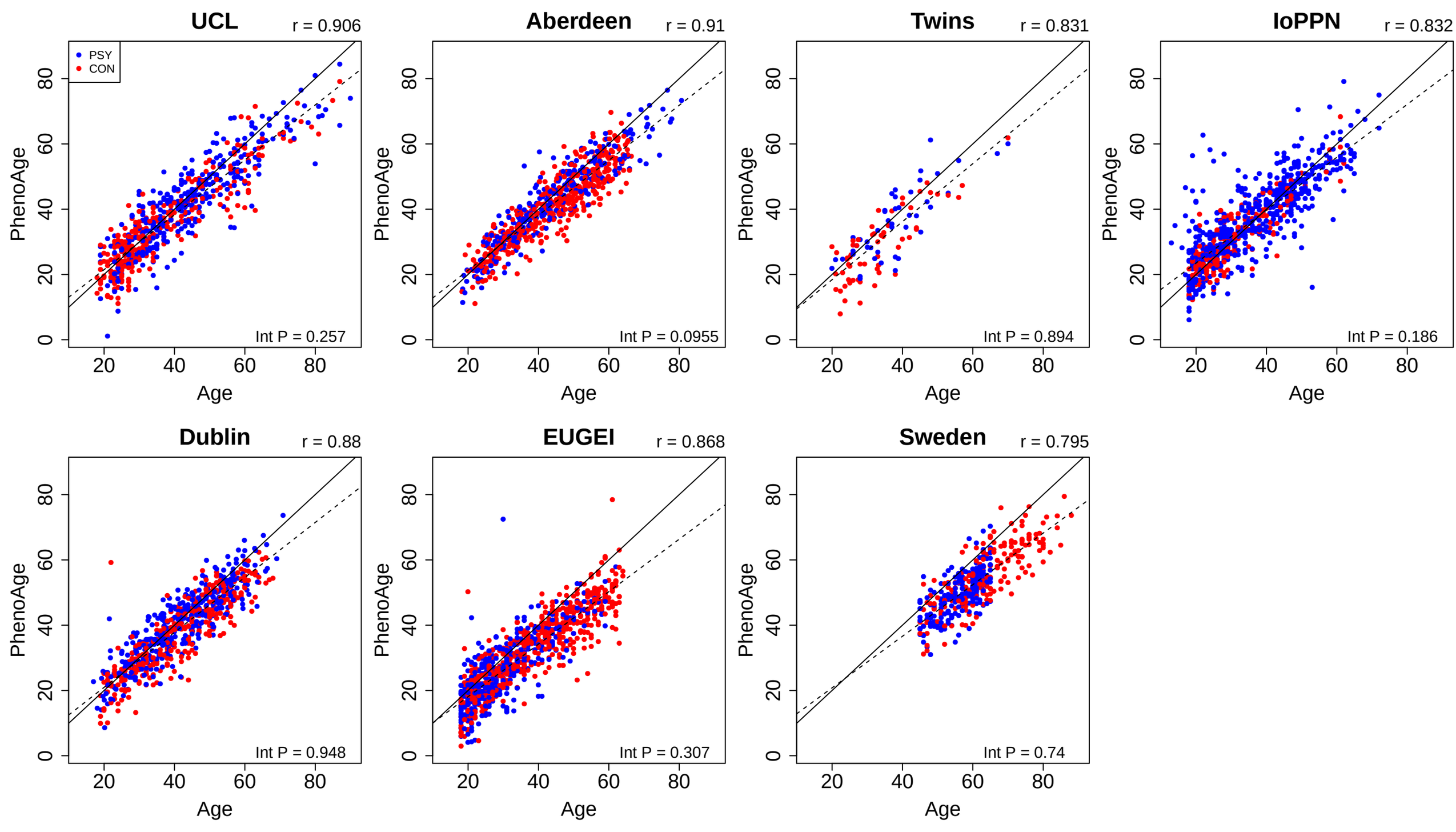
1354

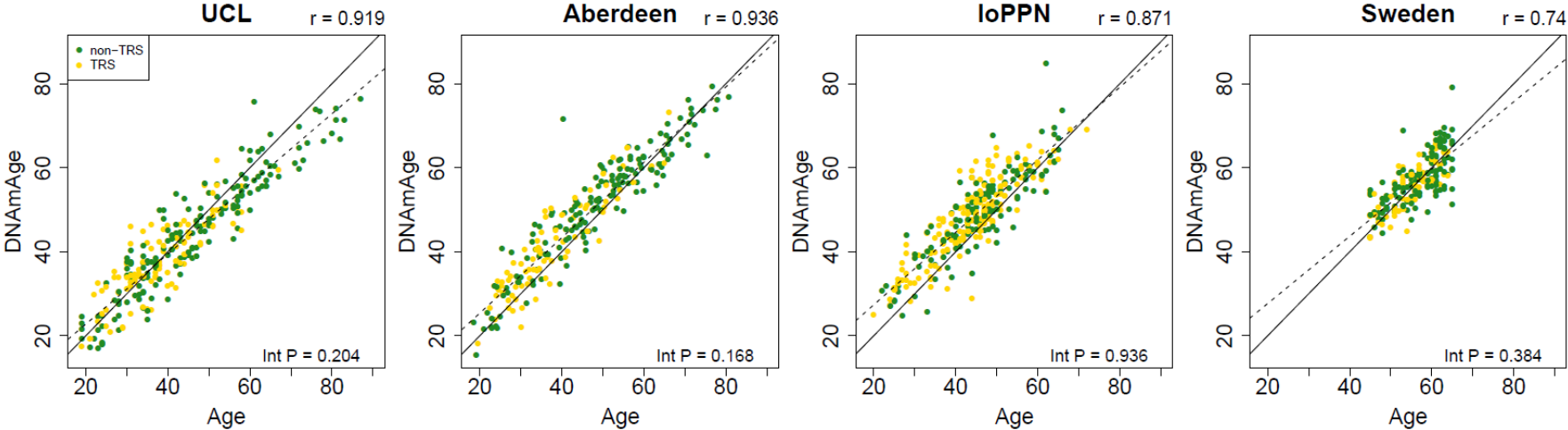

Cohorts
UCL
• N = 675
• Schizophrenia cases vs controls
• Illumina 450K
Aberdeen
• N = 847
• Schizophrenia cases vs controls
• Illumina 450K
Twins
• N = 192
• Schizophrenia-discordant twin-pairs
• Illumina 450K
IoPPN
• N = 800
• Schizophrenia and FEP cases vs controls
• Illumina 450K
Dublin
• N = 679
• Schizophrenia cases vs controls
• Illumina 450K
EU-GEI
• N = 912
• FEP cases vs controls
• Illumina EPIC
Sweden
• N = 378
• Schizophrenia cases vs controls
• Illumina EPIC
Analyses
DNAm-derived variables
DNAm age acceleration
Blood cell proportions
Smoking score
Disease traits
Psychosis
Schizophrenia
First-episode psychosis
Treatment-resistant schizophrenia
EWAS covariates
Age
Sex
Derived cell-type proportions
Derived smoking score
Experimental batch
Genetic PCs
Meta-analysis
Bacon used to control P-value inflation
Random effects meta-analysis
839,131 DNA methylation sites analyzed in at least two cohorts
Experiment-wide significance threshold ($P < 9x10^{-8}$)

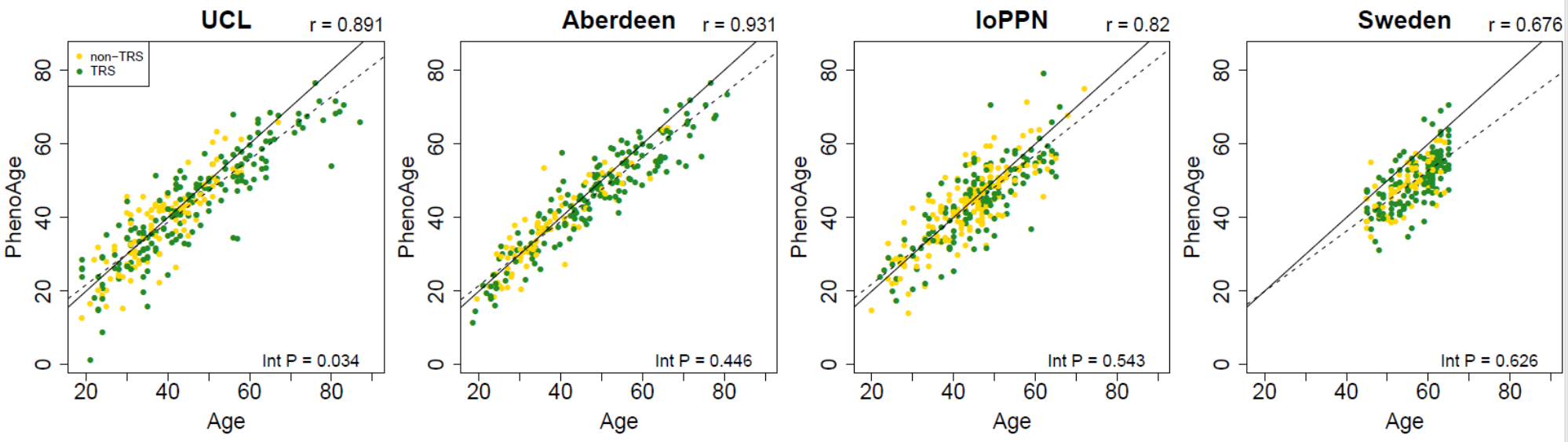

Study
TE
seTE
95%−CI
Weight (fixed)
Weight (random)
UCL 6.8373 1.1622 6.8373 [4.5593; 9.1152] 10.1% 14.3%
Aberdeen −0.6704 1.0629 −0.6704 [−2.7537; 1.4129] 12.1% 14.4%
Twins 4.5375 2.1001 4.5375 [0.4213; 8.6537] 3.1% 13.3%
IoPPN 6.2020 0.8936 6.2020 [4.4506; 7.9535] 17.1% 14.5%
Dublin −0.6150 0.9211 −0.6150 [−2.4203; 1.1903] 16.1% 14.5%
EUGEI −7.9450 0.7899 −7.9450 [−9.4932; −6.3969] 21.9% 14.5%
Sweden −7.3175 0.8313 −7.3175 [−8.9468; −5.6881] 19.7% 14.5%
Fixed effect model −1.4705 [−2.1943; −0.7468] 100.0% ––
Random effects model 0.0757 [−4.6974; 4.8489] –– 100.0%
Heterogeneity: I^2 = 98%, p < 0.01
−5
0
5
Mean difference


UCL
Aberdeen
Twins
IoPPN
Dublin
EuGEI
PC1
PC2
ASW
CEU
CHD
GIH
CHB
JPT
LWK
MKK
MEX
TSI
YRI
Sample


UCL
r = 0.927
PSY
CON
Int P = 0.716
Aberdeen
r = 0.928
Int P = 0.633
Twins
r = 0.872
Int P = 0.273
IoPPN
r = 0.859
Int P = 0.122
Dublin
r = 0.909
Int P = 0.271
EUGEI
r = 0.876
Int P = 0.465
Sweden
r = 0.821
Int P = 0.0802
DNAmAge
Age
20
40
60
80


UCL
r = 0.906
PSY
CON
Int P = 0.257
Aberdeen
r = 0.91
Int P = 0.0955
Twins
r = 0.831
Int P = 0.894
IoPPN
r = 0.832
Int P = 0.186
Dublin
r = 0.88
Int P = 0.948
EUGEI
r = 0.868
Int P = 0.307
Sweden
r = 0.795
Int P = 0.74
PhenoAge
Age
0
20
40
60
80


UCL
r = 0.919
non−TRS
TRS
DNAmAge
Age
Int P = 0.204
Aberdeen
r = 0.936
Int P = 0.168
IoPPN
r = 0.871
Int P = 0.936
Sweden
r = 0.74
Int P = 0.384

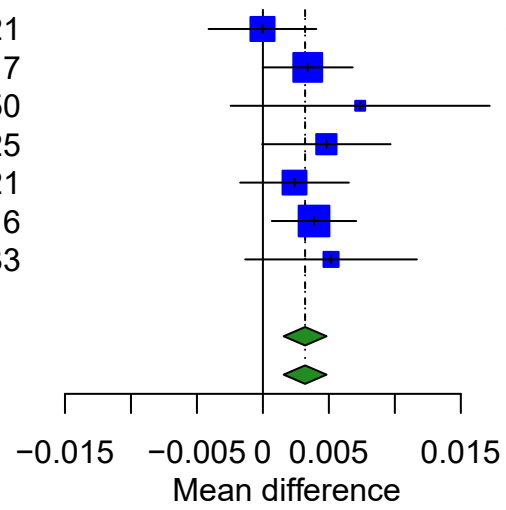

UCL
r = 0.891
non−TRS
TRS
Int P = 0.034
Aberdeen
r = 0.931
Int P = 0.446
IoPPN
r = 0.82
Int P = 0.543
Sweden
r = 0.676
Int P = 0.626
PhenoAge
Age
0
20
40
60
80
20
40
60
80

A

Monocytes

Study	TE	seTE	95%−CI	Weight (fixed)	Weight (random)
UCL	−0.0000	0.0021	−0.0000 [−0.0041; 0.0040]	15.8%	15.8%
Aberdeen	0.0034	0.0017	0.0034 [0.0000; 0.0068]	22.7%	22.7%
Twins	0.0074	0.0050	0.0074 [−0.0024; 0.0172]	2.7%	2.7%
IoPPN	0.0048	0.0025	0.0048 [−0.0001; 0.0097]	11.2%	11.2%
Dublin	0.0024	0.0021	0.0024 [−0.0017; 0.0065]	15.6%	15.6%
EUGEI	0.0039	0.0016	0.0039 [0.0007; 0.0071]	25.8%	25.8%
Sweden	0.0052	0.0033	0.0052 [−0.0013; 0.0117]	6.2%	6.2%
Fixed effect model			**0.0032 [0.0016; 0.0048]**	**100.0%**	**--**
Random effects model			**0.0032 [0.0016; 0.0048]**	**--**	**100.0%**

Heterogeneity: I^2 = 0%, p = 0.65

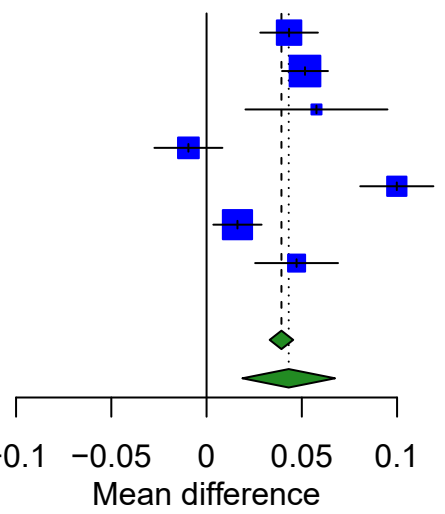


B

Granulocytes

Study	TE	seTE	95%−CI	Weight (fixed)	Weight (random)
UCL	0.0433	0.0077	0.0433 [0.0282; 0.0584]	16.7%	14.9%
Aberdeen	0.0517	0.0061	0.0517 [0.0396; 0.0637]	26.3%	15.3%
Twins	0.0577	0.0190	0.0577 [0.0204; 0.0950]	2.8%	11.5%
IoPPN	−0.0095	0.0091	−0.0095 [−0.0274; 0.0083]	12.0%	14.6%
Dublin	0.0999	0.0098	0.0999 [0.0807; 0.1191]	10.4%	14.4%
EUGEI	0.0162	0.0065	0.0162 [0.0035; 0.0289]	23.8%	15.2%
Sweden	0.0472	0.0111	0.0472 [0.0255; 0.0690]	8.1%	14.1%
Fixed effect model			**0.0393 [0.0331; 0.0455]**	**100.0%**	**--**
Random effects model			**0.0431 [0.0188; 0.0674]**	**--**	**100.0%**

Heterogeneity: I^2 = 93%, p < 0.01

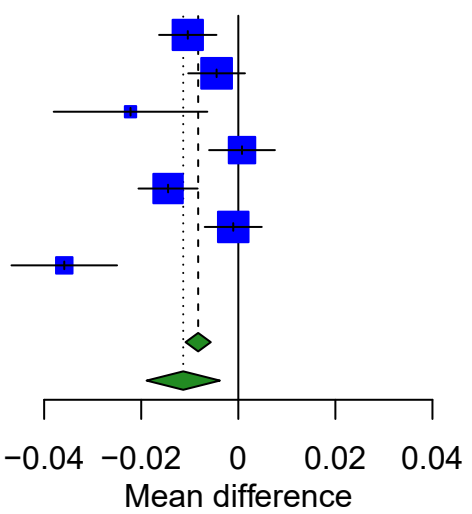


C

Natural Killer

Study	TE	seTE	95%−CI	Weight (fixed)	Weight (random)
UCL	−0.0104	0.0030	−0.0104 [−0.0163; −0.0045]	19.3%	15.6%
Aberdeen	−0.0045	0.0030	−0.0045 [−0.0103; 0.0014]	19.8%	15.6%
Twins	−0.0222	0.0081	−0.0222 [−0.0380; −0.0064]	2.7%	9.8%
IoPPN	0.0008	0.0035	0.0008 [−0.0060; 0.0075]	14.7%	15.1%
Dublin	−0.0145	0.0031	−0.0145 [−0.0206; −0.0084]	18.2%	15.5%
EUGEI	−0.0010	0.0030	−0.0010 [−0.0069; 0.0048]	19.6%	15.6%
Sweden	−0.0359	0.0055	−0.0359 [−0.0467; −0.0250]	5.7%	12.7%
Fixed effect model			**−0.0083 [−0.0109; −0.0057]**	**100.0%**	**--**
Random effects model			**−0.0113 [−0.0189; −0.0038]**	**--**	**100.0%**

Heterogeneity: I^2 = 87%, p < 0.01

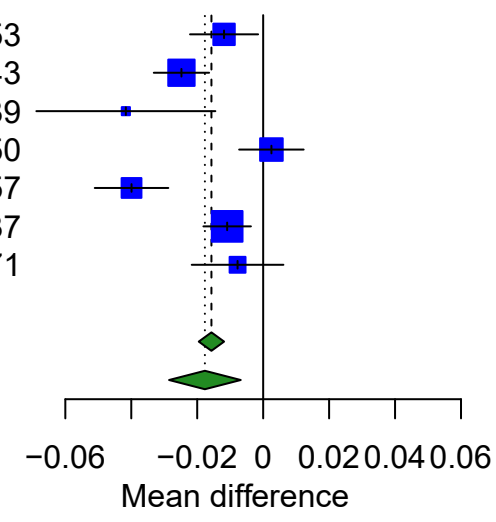


D

$CD4^+$ T-cells

Study	TE	seTE	95%−CI	Weight (fixed)	Weight (random)
UCL	−0.0119	0.0053	−0.0119 [−0.0222; −0.0015]	13.8%	15.2%
Aberdeen	−0.0248	0.0043	−0.0248 [−0.0332; −0.0164]	20.9%	16.0%
Twins	−0.0416	0.0139	−0.0416 [−0.0688; −0.0145]	2.0%	8.4%
IoPPN	0.0025	0.0050	0.0025 [−0.0073; 0.0123]	15.4%	15.4%
Dublin	−0.0399	0.0057	−0.0399 [−0.0511; −0.0288]	11.8%	14.9%
EUGEI	−0.0109	0.0037	−0.0109 [−0.0181; −0.0037]	28.4%	16.4%
Sweden	−0.0078	0.0071	−0.0078 [−0.0217; 0.0061]	7.6%	13.7%
Fixed effect model			**−0.0157 [−0.0195; −0.0119]**	**100.0%**	**--**
Random effects model			**−0.0177 [−0.0285; −0.0068]**	**--**	**100.0%**

Heterogeneity: I^2 = 86%, p < 0.01

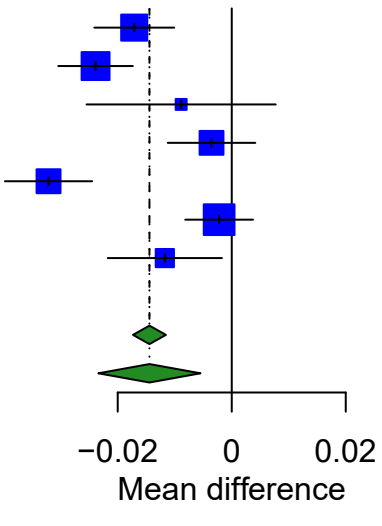


E

$CD8^+$ T-cells

Study	TE	seTE	95%−CI	Weight (fixed)	Weight (random)
UCL	−0.0171	0.0036	−0.0171 [−0.0241; −0.0101]	17.0%	15.1%
Aberdeen	−0.0239	0.0033	−0.0239 [−0.0304; −0.0173]	19.5%	15.3%
Twins	−0.0089	0.0085	−0.0089 [−0.0255; 0.0077]	3.0%	10.6%
IoPPN	−0.0035	0.0039	−0.0035 [−0.0112; 0.0041]	14.2%	14.8%
Dublin	−0.0321	0.0039	−0.0321 [−0.0398; −0.0245]	14.3%	14.9%
EUGEI	−0.0022	0.0030	−0.0022 [−0.0082; 0.0037]	23.6%	15.5%
Sweden	−0.0117	0.0051	−0.0117 [−0.0217; −0.0017]	8.3%	13.8%
Fixed effect model			**−0.0144 [−0.0173; −0.0115]**	**100.0%**	**--**
Random effects model			**−0.0144 [−0.0234; −0.0055]**	**--**	**100.0%**

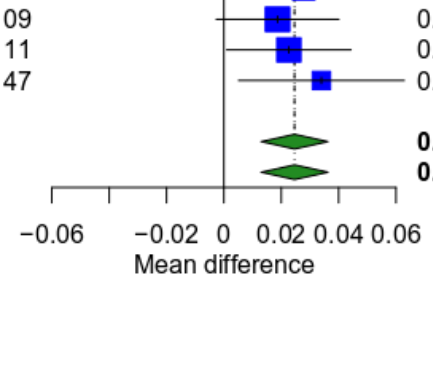
Heterogeneity: I^2 = 89%, p < 0.01



Granulocytes

Study	TE	seTE	95%−CI	Weight (fixed)	Weight (random)
UCL	0.0278	0.0119	0.0278 [0.0045; 0.0512]	24.9%	24.9%
Aberdeen	0.0187	0.0109	0.0187 [−0.0027; 0.0401]	29.8%	29.8%
IoPPN	0.0226	0.0111	0.0226 [0.0010; 0.0443]	29.0%	29.0%
Sweden	0.0340	0.0147	0.0340 [0.0051; 0.0629]	16.3%	16.3%
Fixed effect model			**0.0246 [0.0129; 0.0363]**	**100.0%**	**--**
Random effects model			**0.0246 [0.0129; 0.0363]**	**--**	**100.0%**

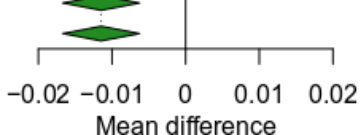
Heterogeneity: I^2 = 0%, p = 0.85



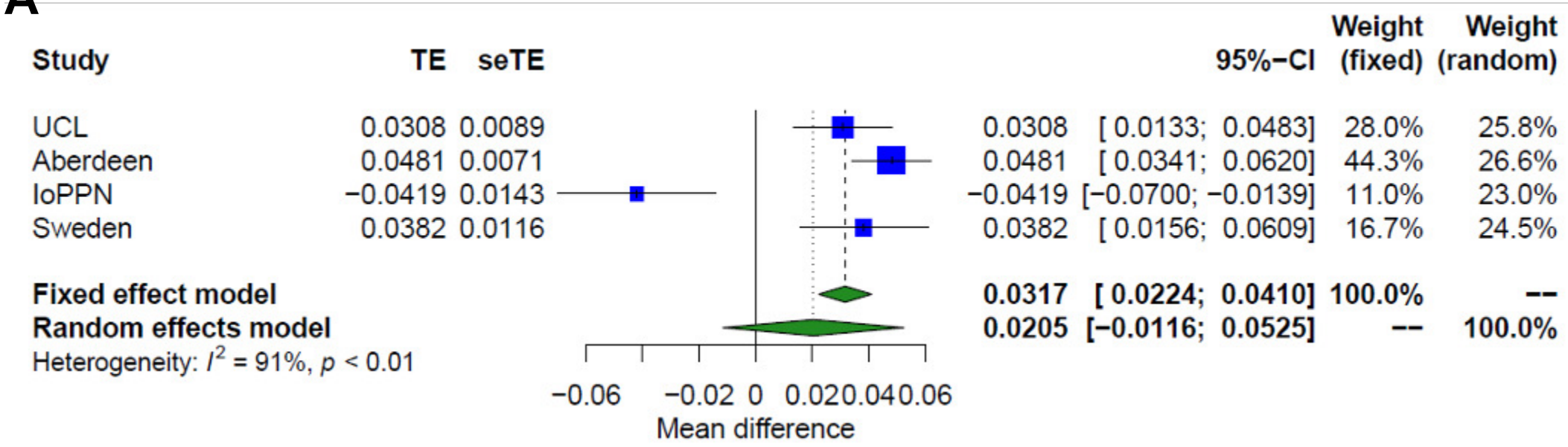
CD8$^+$T-cells

Study	TE	seTE	95%−CI	Weight (fixed)	Weight (random)
UCL	−0.0148	0.0049	−0.0148 [−0.0245; −0.0051]	29.1%	29.1%
Aberdeen	−0.0103	0.0056	−0.0103 [−0.0212; 0.0007]	22.8%	22.8%
IoPPN	−0.0100	0.0047	−0.0100 [−0.0191; −0.0008]	32.5%	32.5%
Sweden	−0.0102	0.0067	−0.0102 [−0.0234; 0.0030]	15.6%	15.6%
Fixed effect model			**−0.0115 [−0.0167; −0.0063]**	**100.0%**	**--**
Random effects model			**−0.0115 [−0.0167; −0.0063]**	**--**	**100.0%**

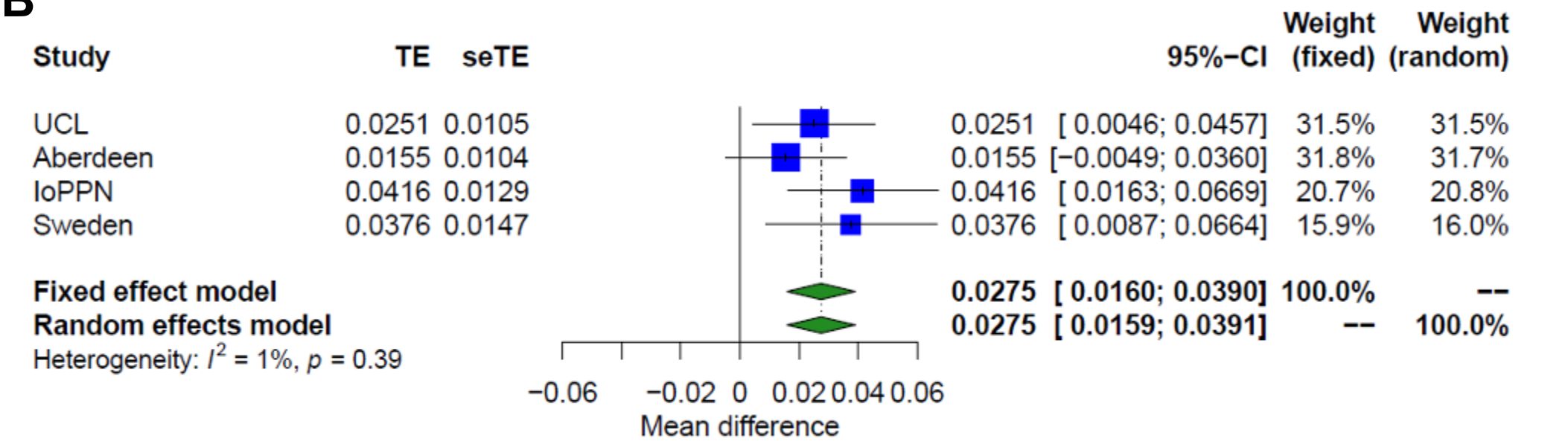
Heterogeneity: I^2 = 0%, p = 0.89



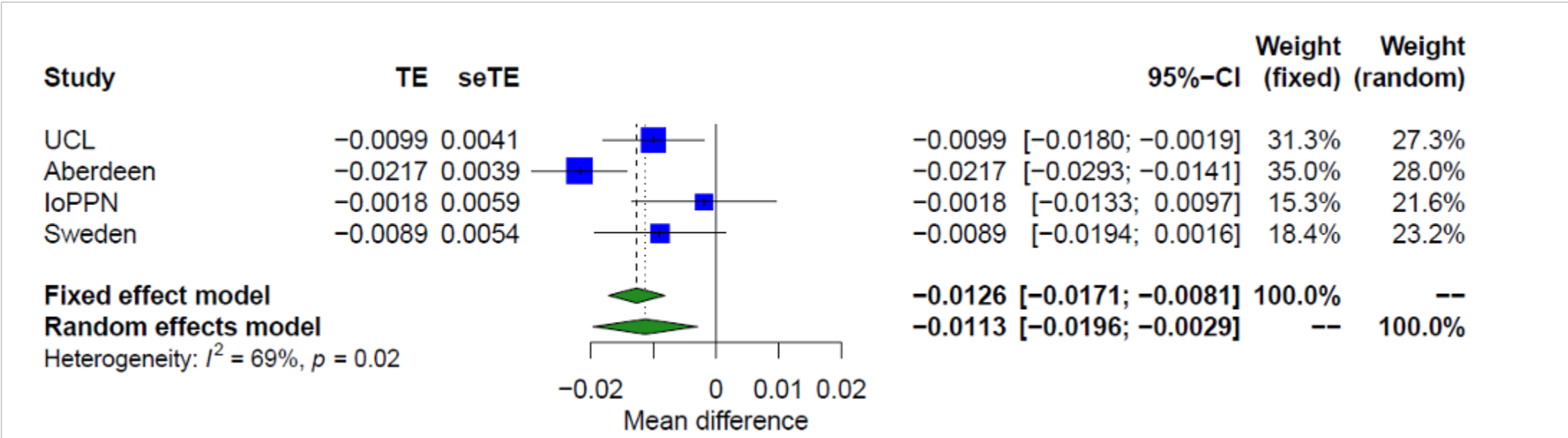
A


Study
TE
seTE
95%−CI
Weight (fixed)
Weight (random)
UCL 0.0308 0.0089 0.0308 [0.0133; 0.0483] 28.0% 25.8%
Aberdeen 0.0481 0.0071 0.0481 [0.0341; 0.0620] 44.3% 26.6%
IoPPN −0.0419 0.0143 −0.0419 [−0.0700; −0.0139] 11.0% 23.0%
Sweden 0.0382 0.0116 0.0382 [0.0156; 0.0609] 16.7% 24.5%
Fixed effect model 0.0317 [0.0224; 0.0410] 100.0% --
Random effects model 0.0205 [−0.0116; 0.0525] -- 100.0%
Heterogeneity: I^2 = 91%, p < 0.01
−0.06 −0.02 0 0.02 0.04 0.06
Mean difference

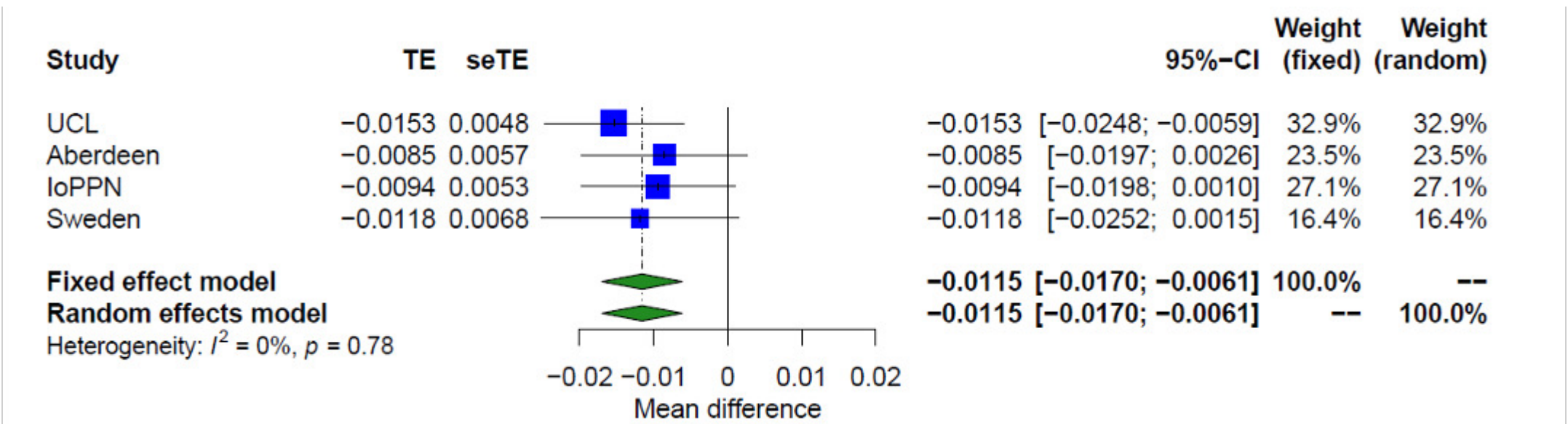
B

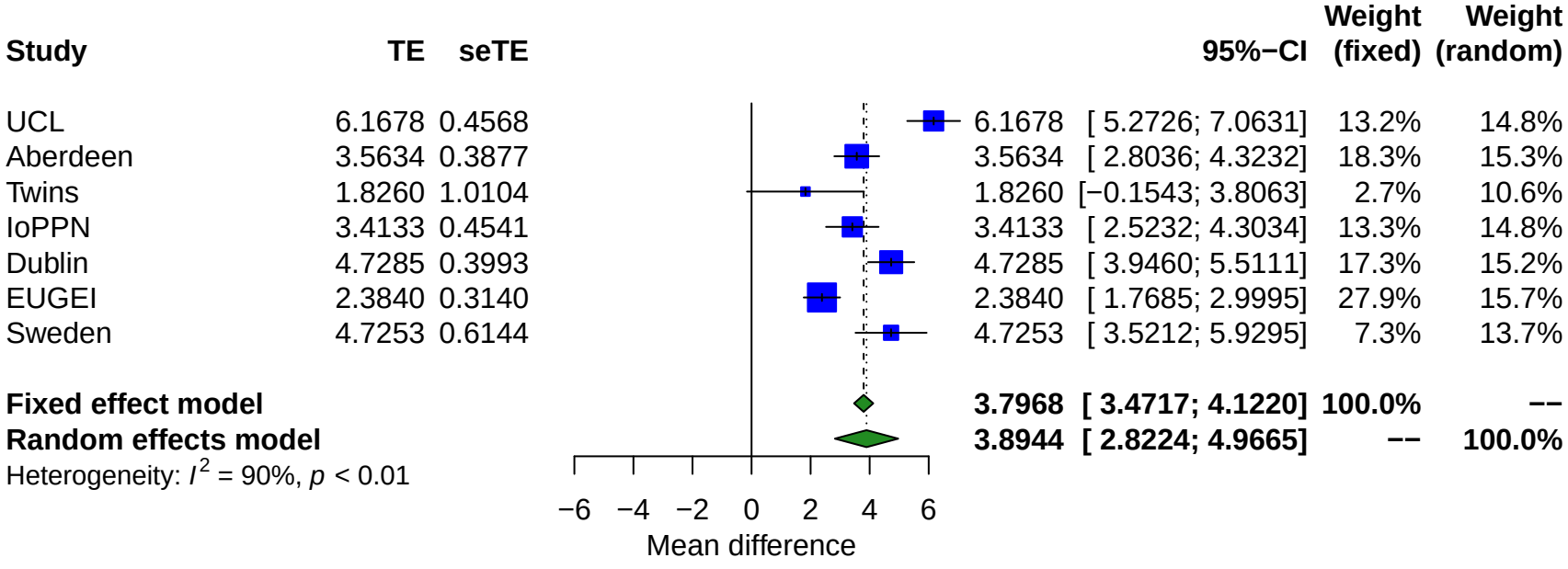

Study
TE
seTE
95%−CI
Weight (fixed)
Weight (random)
UCL 0.0251 0.0105 0.0251 [0.0046; 0.0457] 31.5% 31.5%
Aberdeen 0.0155 0.0104 0.0155 [−0.0049; 0.0360] 31.8% 31.7%
IoPPN 0.0416 0.0129 0.0416 [0.0163; 0.0669] 20.7% 20.8%
Sweden 0.0376 0.0147 0.0376 [0.0087; 0.0664] 15.9% 16.0%
Fixed effect model 0.0275 [0.0160; 0.0390] 100.0% --
Random effects model 0.0275 [0.0159; 0.0391] -- 100.0%
Heterogeneity: I^2 = 1%, p = 0.39
−0.06 −0.02 0 0.02 0.04 0.06
Mean difference

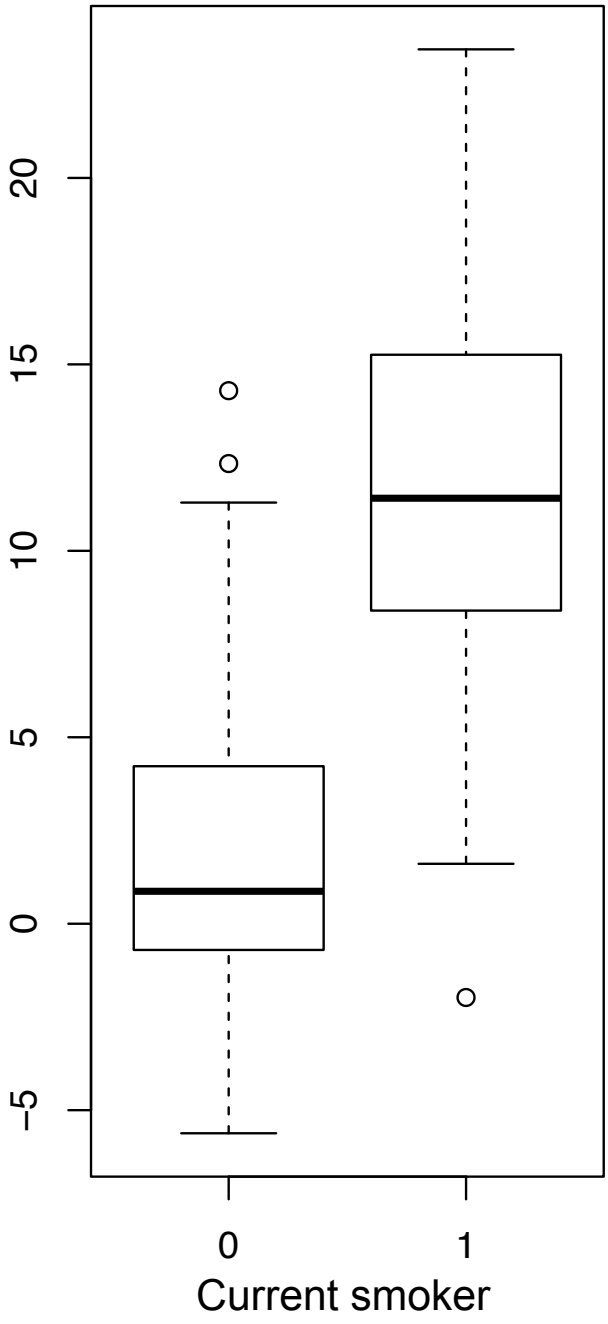
A

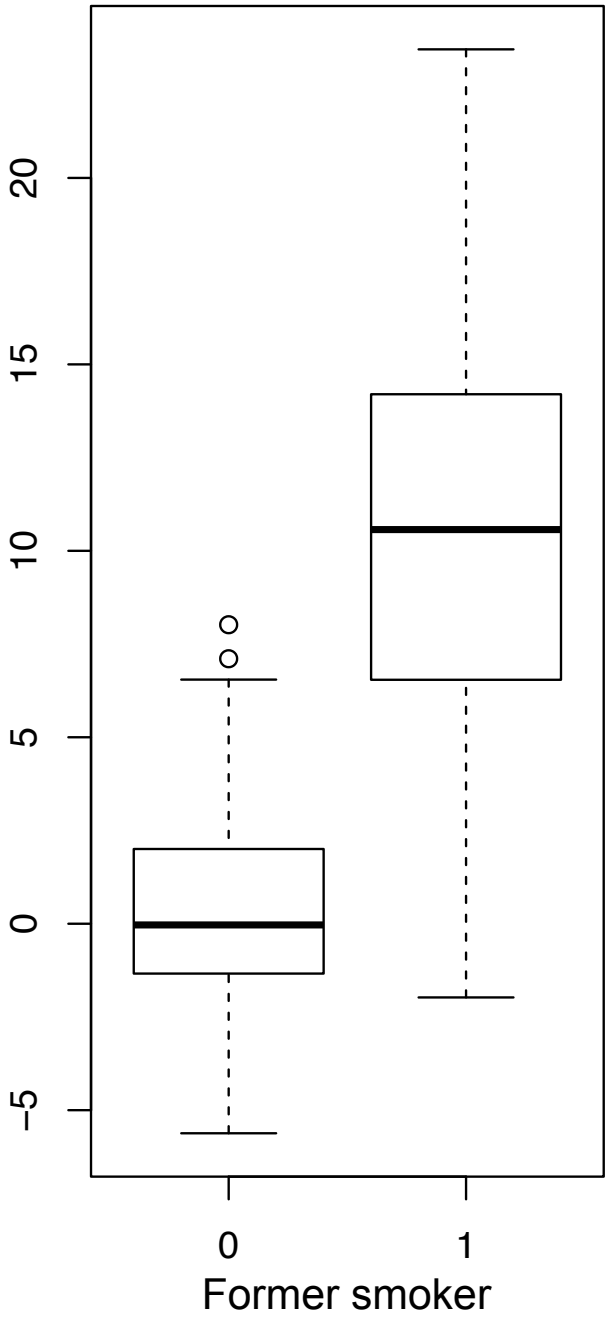

Study
TE
seTE
95%−CI
Weight (fixed)
Weight (random)
UCL −0.0099 0.0041 −0.0099 [−0.0180; −0.0019] 31.3% 27.3%
Aberdeen −0.0217 0.0039 −0.0217 [−0.0293; −0.0141] 35.0% 28.0%
IoPPN −0.0018 0.0059 −0.0018 [−0.0133; 0.0097] 15.3% 21.6%
Sweden −0.0089 0.0054 −0.0089 [−0.0194; 0.0016] 18.4% 23.2%
Fixed effect model −0.0126 [−0.0171; −0.0081] 100.0% --
Random effects model −0.0113 [−0.0196; −0.0029] -- 100.0%
Heterogeneity: I^2 = 69%, p = 0.02
−0.02
0
0.01
0.02
Mean difference

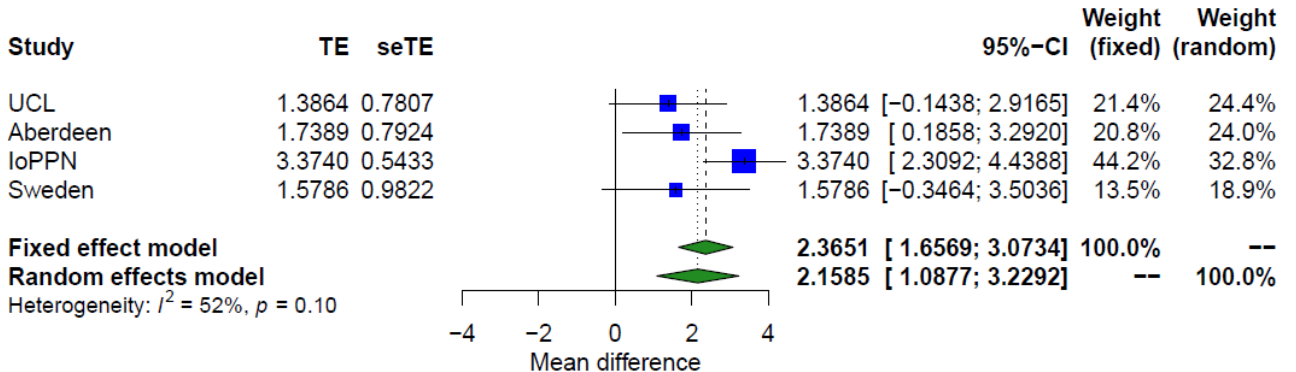
B

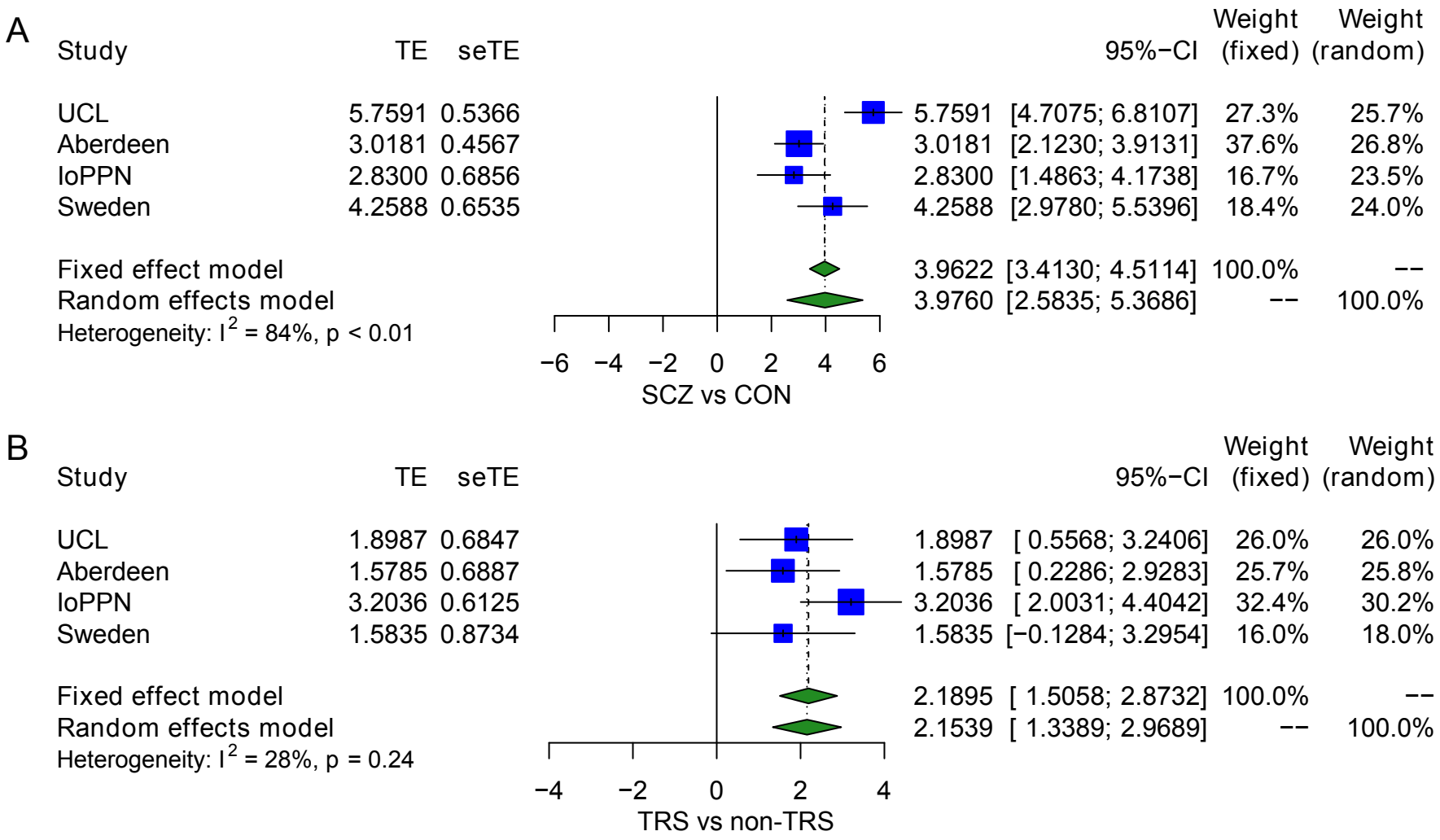

Study
TE
seTE
95%−CI
Weight (fixed)
Weight (random)
UCL −0.0153 0.0048 −0.0153 [−0.0248; −0.0059] 32.9% 32.9%
Aberdeen −0.0085 0.0057 −0.0085 [−0.0197; 0.0026] 23.5% 23.5%
IoPPN −0.0094 0.0053 −0.0094 [−0.0198; 0.0010] 27.1% 27.1%
Sweden −0.0118 0.0068 −0.0118 [−0.0252; 0.0015] 16.4% 16.4%
Fixed effect model −0.0115 [−0.0170; −0.0061] 100.0% --
Random effects model −0.0115 [−0.0170; −0.0061] -- 100.0%
Heterogeneity: I^2 = 0%, p = 0.78
−0.02
−0.01
0
0.01
0.02
Mean difference

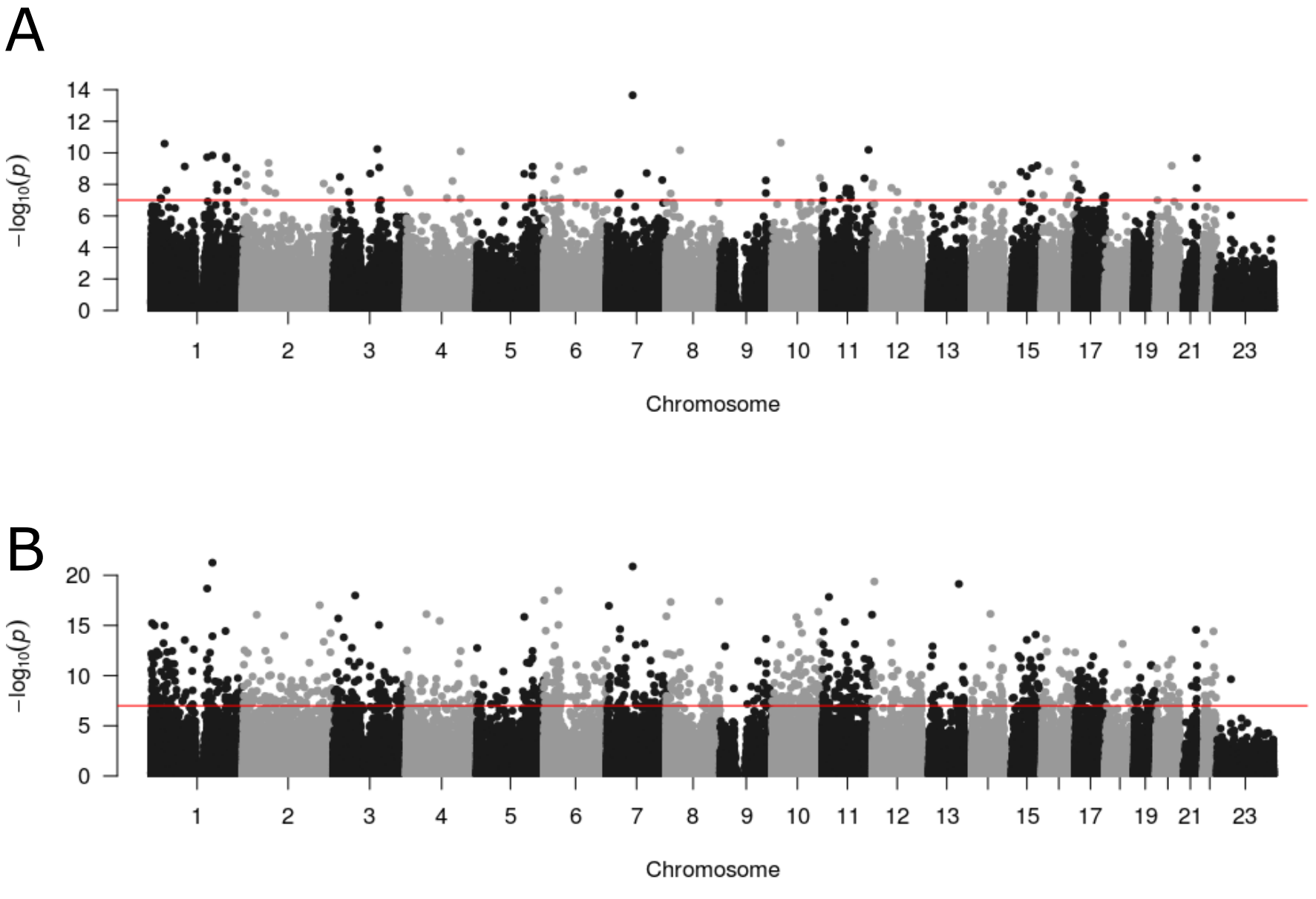

Study
TE
seTE
95%−CI
Weight (fixed)
Weight (random)
UCL 6.1678 0.4568 6.1678 [5.2726; 7.0631] 13.2% 14.8%
Aberdeen 3.5634 0.3877 3.5634 [2.8036; 4.3232] 18.3% 15.3%
Twins 1.8260 1.0104 1.8260 [−0.1543; 3.8063] 2.7% 10.6%
IoPPN 3.4133 0.4541 3.4133 [2.5232; 4.3034] 13.3% 14.8%
Dublin 4.7285 0.3993 4.7285 [3.9460; 5.5111] 17.3% 15.2%
EUGEI 2.3840 0.3140 2.3840 [1.7685; 2.9995] 27.9% 15.7%
Sweden 4.7253 0.6144 4.7253 [3.5212; 5.9295] 7.3% 13.7%
Fixed effect model 3.7968 [3.4717; 4.1220] 100.0% ––
Random effects model 3.8944 [2.8224; 4.9665] –– 100.0%
Heterogeneity: I^2 = 90%, p < 0.01
−6 −4 −2 0 2 4 6
Mean difference

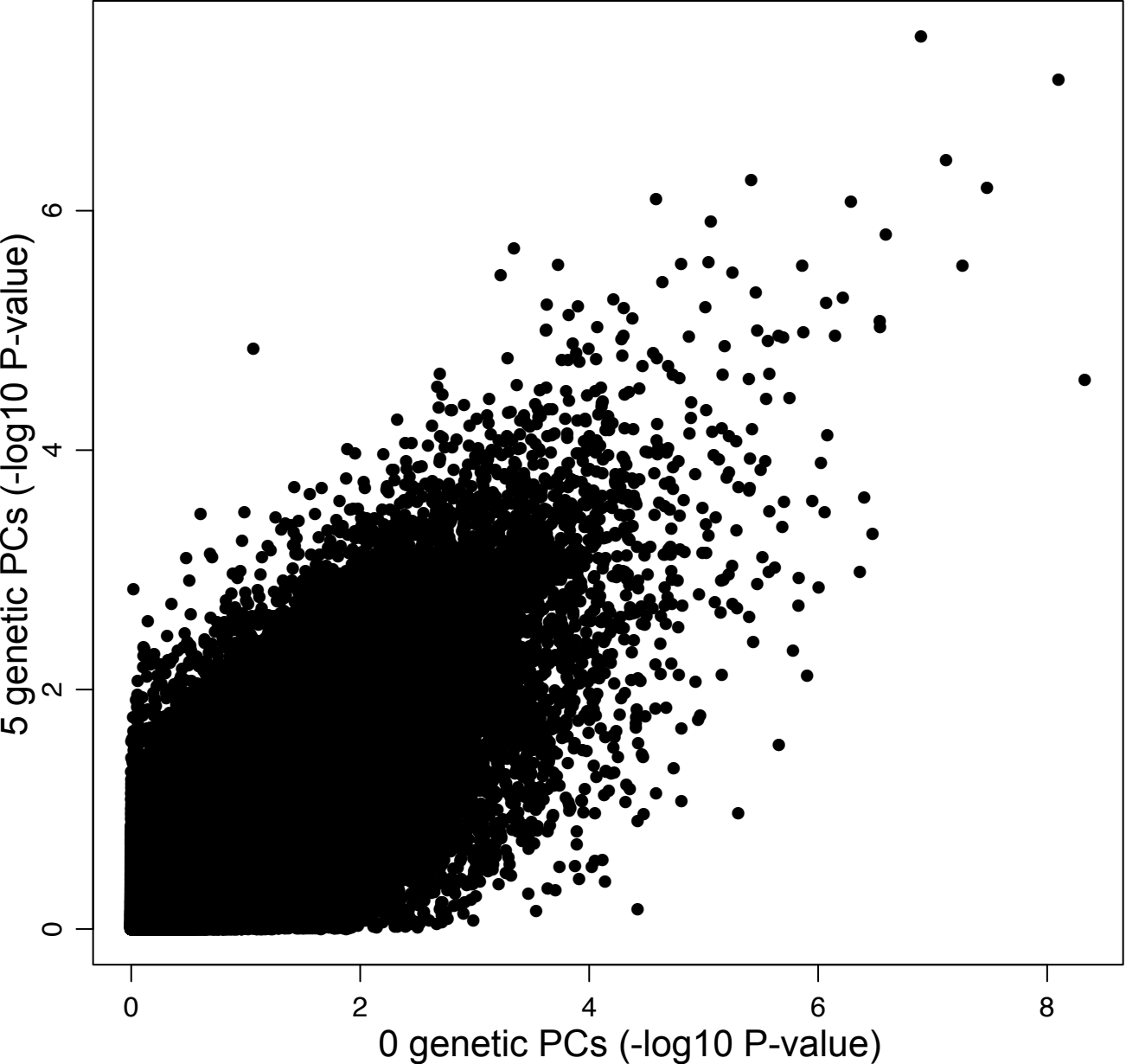

20
15
10
5
0
−5
0
1
Current smoker

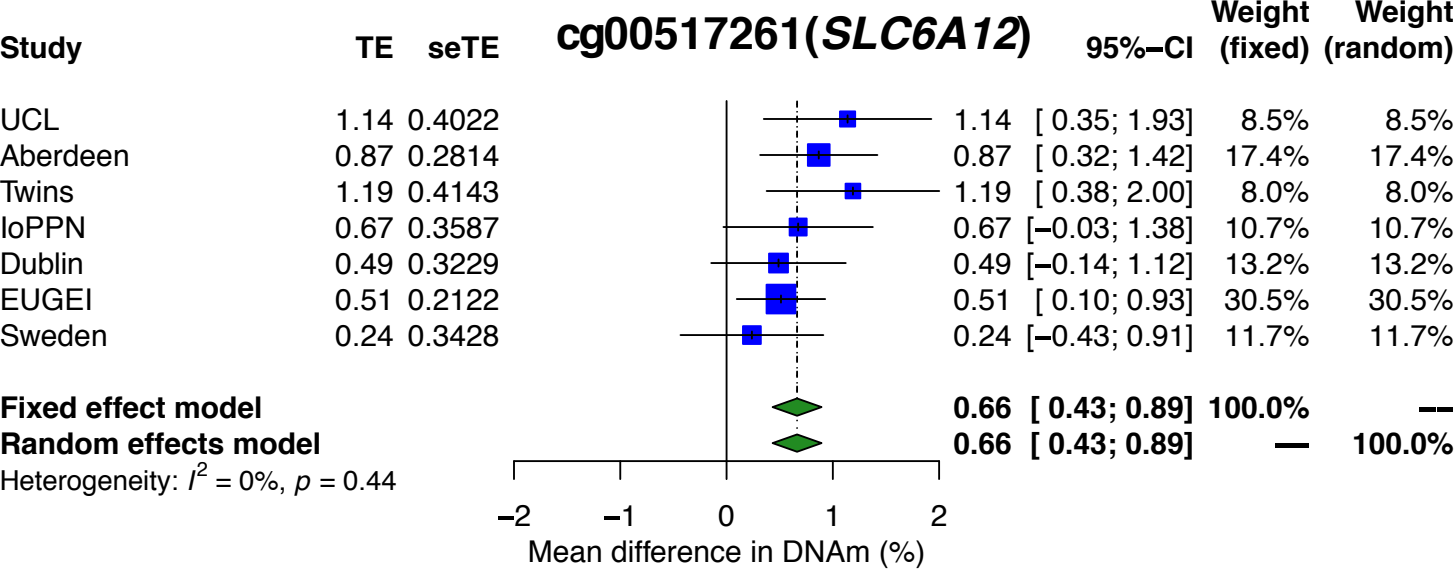

20
15
10
5
0
−5
0
1
Former smoker

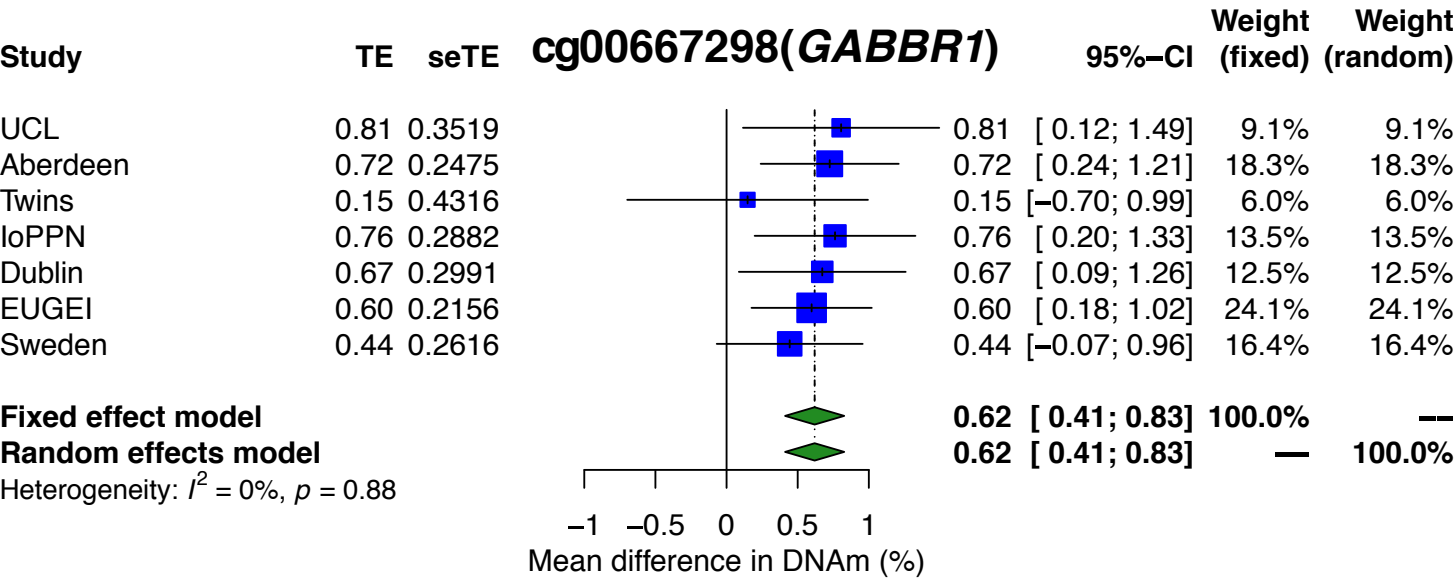

Study
TE
seTE
95%−CI
Weight (fixed)
Weight (random)
UCL 1.3864 0.7807 1.3864 [−0.1438; 2.9165] 21.4% 24.4%
Aberdeen 1.7389 0.7924 1.7389 [0.1858; 3.2920] 20.8% 24.0%
IoPPN 3.3740 0.5433 3.3740 [2.3092; 4.4388] 44.2% 32.8%
Sweden 1.5786 0.9822 1.5786 [−0.3464; 3.5036] 13.5% 18.9%
Fixed effect model 2.3651 [1.6569; 3.0734] 100.0% --
Random effects model 2.1585 [1.0877; 3.2292] -- 100.0%
Heterogeneity: I^2 = 52%, p = 0.10
−4
−2
0
2
4
Mean difference

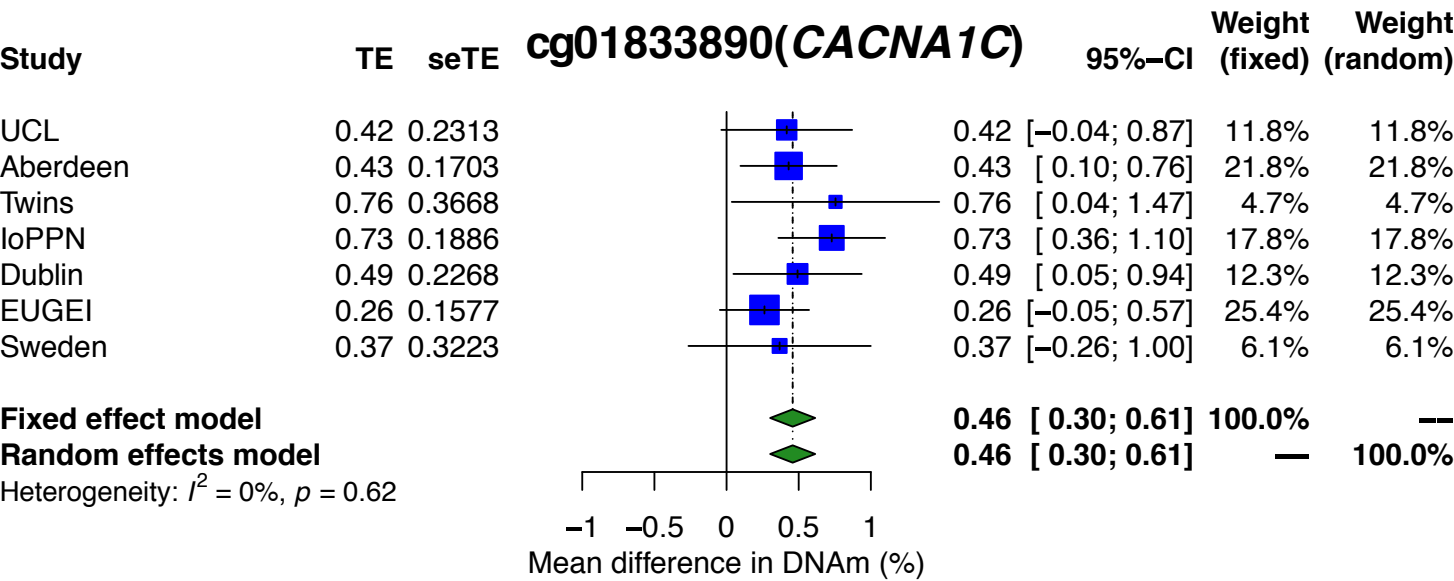

A
Study
TE
seTE
95%−CI
Weight (fixed)
Weight (random)
UCL 5.7591 0.5366 5.7591 [4.7075; 6.8107] 27.3% 25.7%
Aberdeen 3.0181 0.4567 3.0181 [2.1230; 3.9131] 37.6% 26.8%
IoPPN 2.8300 0.6856 2.8300 [1.4863; 4.1738] 16.7% 23.5%
Sweden 4.2588 0.6535 4.2588 [2.9780; 5.5396] 18.4% 24.0%
Fixed effect model 3.9622 [3.4130; 4.5114] 100.0% --
Random effects model 3.9760 [2.5835; 5.3686] -- 100.0%
Heterogeneity: I^2 = 84%, p < 0.01
−6 −4 −2 0 2 4 6
SCZ vs CON
B
Study
TE
seTE
95%−CI
Weight (fixed)
Weight (random)
UCL 1.8987 0.6847 1.8987 [0.5568; 3.2406] 26.0% 26.0%
Aberdeen 1.5785 0.6887 1.5785 [0.2286; 2.9283] 25.7% 25.8%
IoPPN 3.2036 0.6125 3.2036 [2.0031; 4.4042] 32.4% 30.2%
Sweden 1.5835 0.8734 1.5835 [−0.1284; 3.2954] 16.0% 18.0%
Fixed effect model 2.1895 [1.5058; 2.8732] 100.0% --
Random effects model 2.1539 [1.3389; 2.9689] -- 100.0%
Heterogeneity: I^2 = 28%, p = 0.24
−4 −2 0 2 4
TRS vs non-TRS

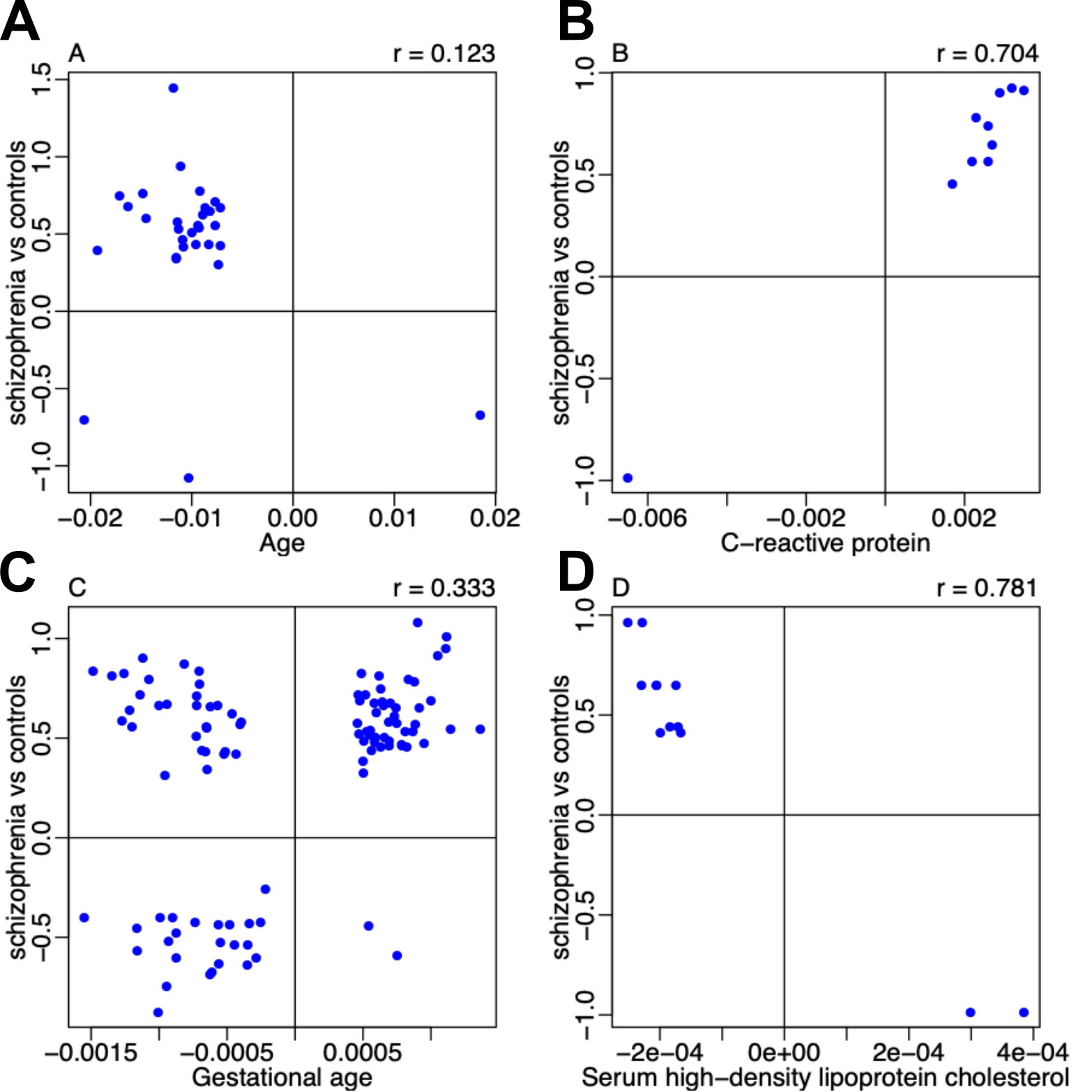

A
14
12
10
8
6
4
2
0
−log10(p)
1
2
3
4
5
6
7
8
9
10
11
12
13
15
17
19
21
23
Chromosome
B
20
15
10
5
0
−log10(p)
1
2
3
4
5
6
7
8
9
10
11
12
13
15
17
19
21
23
Chromosome

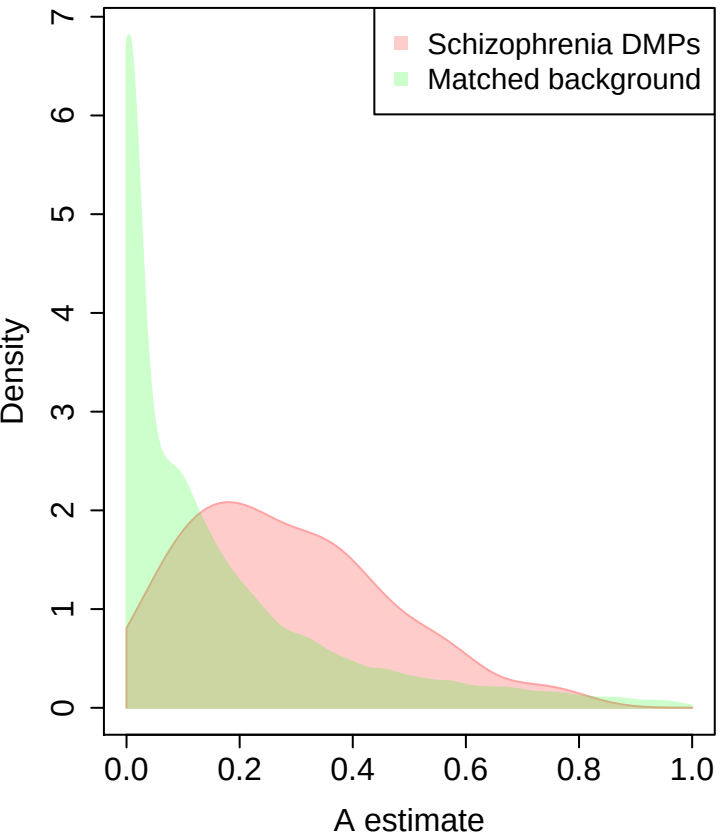

5 genetic PCs (-log10 P-value)
0 genetic PCs (-log10 P-value)
0
2
4
6
8

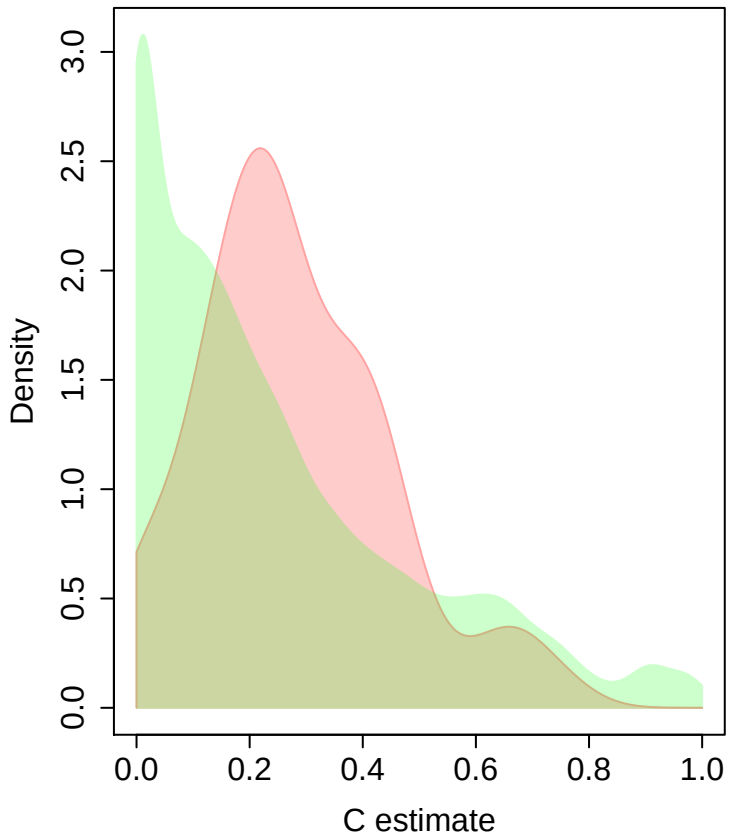

Study
TE
seTE
cg00517261(SLC6A12)
95%–CI
Weight (fixed)
Weight (random)
UCL 1.14 0.4022 1.14 [0.35; 1.93] 8.5% 8.5%
Aberdeen 0.87 0.2814 0.87 [0.32; 1.42] 17.4% 17.4%
Twins 1.19 0.4143 1.19 [0.38; 2.00] 8.0% 8.0%
IoPPN 0.67 0.3587 0.67 [–0.03; 1.38] 10.7% 10.7%
Dublin 0.49 0.3229 0.49 [–0.14; 1.12] 13.2% 13.2%
EUGEI 0.51 0.2122 0.51 [0.10; 0.93] 30.5% 30.5%
Sweden 0.24 0.3428 0.24 [–0.43; 0.91] 11.7% 11.7%
Fixed effect model 0.66 [0.43; 0.89] 100.0% —
Random effects model 0.66 [0.43; 0.89] — 100.0%
Heterogeneity: I^2 = 0%, p = 0.44
–2 –1 0 1 2
Mean difference in DNAm (%)

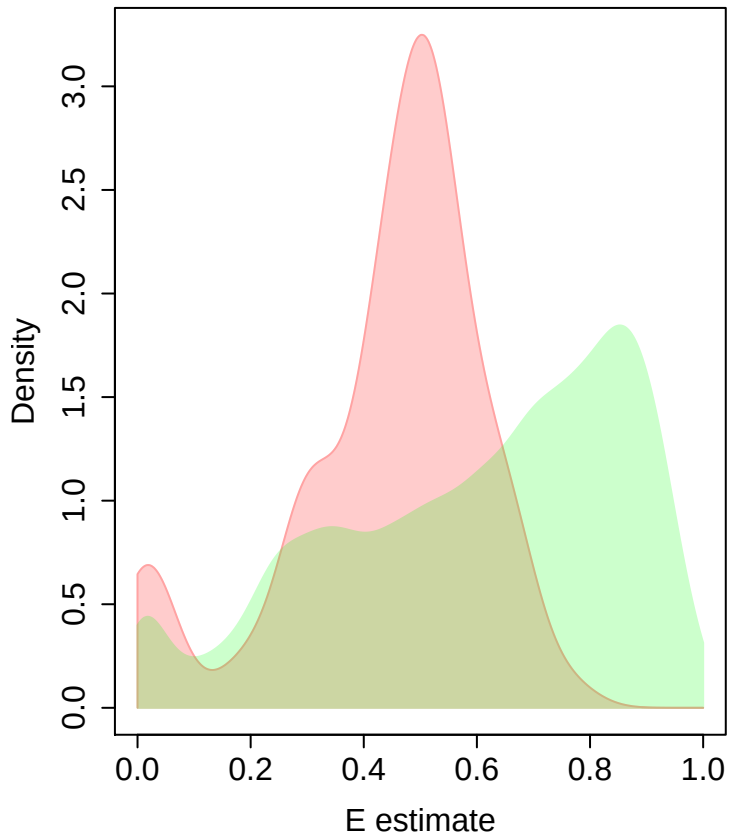

Study
TE
seTE
cg00667298(GABBR1)
95%–CI
Weight (fixed)
Weight (random)
UCL 0.81 0.3519 0.81 [0.12; 1.49] 9.1% 9.1%
Aberdeen 0.72 0.2475 0.72 [0.24; 1.21] 18.3% 18.3%
Twins 0.15 0.4316 0.15 [–0.70; 0.99] 6.0% 6.0%
IoPPN 0.76 0.2882 0.76 [0.20; 1.33] 13.5% 13.5%
Dublin 0.67 0.2991 0.67 [0.09; 1.26] 12.5% 12.5%
EUGEI 0.60 0.2156 0.60 [0.18; 1.02] 24.1% 24.1%
Sweden 0.44 0.2616 0.44 [–0.07; 0.96] 16.4% 16.4%
Fixed effect model 0.62 [0.41; 0.83] 100.0% —
Random effects model 0.62 [0.41; 0.83] — 100.0%
Heterogeneity: I^2 = 0%, p = 0.88
–1 –0.5 0 0.5 1
Mean difference in DNAm (%)


Study
TE
seTE
cg01833890(CACNA1C)
95%–CI
Weight (fixed)
Weight (random)
UCL 0.42 0.2313 0.42 [–0.04; 0.87] 11.8% 11.8%
Aberdeen 0.43 0.1703 0.43 [0.10; 0.76] 21.8% 21.8%
Twins 0.76 0.3668 0.76 [0.04; 1.47] 4.7% 4.7%
IoPPN 0.73 0.1886 0.73 [0.36; 1.10] 17.8% 17.8%
Dublin 0.49 0.2268 0.49 [0.05; 0.94] 12.3% 12.3%
EUGEI 0.26 0.1577 0.26 [–0.05; 0.57] 25.4% 25.4%
Sweden 0.37 0.3223 0.37 [–0.26; 1.00] 6.1% 6.1%
Fixed effect model 0.46 [0.30; 0.61] 100.0% —
Random effects model 0.46 [0.30; 0.61] — 100.0%
Heterogeneity: I^2 = 0%, p = 0.62
–1 –0.5 0 0.5 1
Mean difference in DNAm (%)


A
A
r = 0.123
schizophrenia vs controls
1.5
1.0
0.5
0.0
−0.5
−1.0
−0.02
−0.01
0.00
0.01
0.02
Age
B
B
r = 0.704
schizophrenia vs controls
1.0
0.5
0.0
−0.5
−1.0
−0.006
−0.002
0.002
C−reactive protein
C
C
r = 0.333
schizophrenia vs controls
1.0
0.5
0.0
−0.5
−0.0015
−0.0005
0.0005
Gestational age
D
D
r = 0.781
schizophrenia vs controls
1.0
0.5
0.0
−0.5
−1.0
−2e−04
0e+00
2e−04
4e−04
Serum high−density lipoprotein cholesterol


Schizophrenia DMPs
Matched background
Density
A estimate
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
7

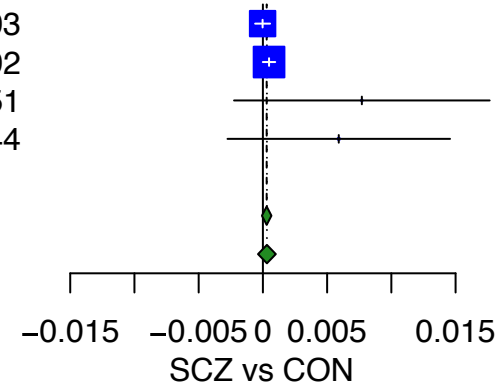

Density
C estimate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Density
E estimate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0

cg16322565(*NR1L2*)

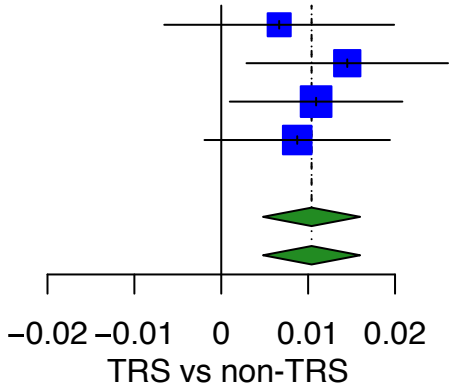
Study	TE	seTE	95%–CI	Weight (fixed)	Weight (random)
UCL	−0.0000	0.0003	−0.0000 [−0.0006; 0.0006]	40.6%	46.7%
Aberdeen	0.0005	0.0002	0.0005 [0.0000; 0.0010]	59.1%	52.1%
IoPPN	0.0077	0.0051	0.0077 [−0.0022; 0.0177]	0.1%	0.5%
Sweden	0.0059	0.0044	0.0059 [−0.0028; 0.0146]	0.2%	0.6%
Fixed effect model			**0.0003 [−0.0001; 0.0007]**	**100.0%**	**––**
Random effects model			**0.0003 [−0.0004; 0.0010]**	**––**	**100.0%**

Heterogeneity: I^2 = 45%, p = 0.14



Study	TE	seTE	95%–CI	Weight (fixed)	Weight (random)
UCL	0.0067	0.0067	0.0067 [−0.0066; 0.0199]	17.8%	17.8%
Aberdeen	0.0145	0.0059	0.0145 [0.0029; 0.0261]	23.2%	23.2%
IoPPN	0.0109	0.0051	0.0109 [0.0010; 0.0208]	31.6%	31.6%
Sweden	0.0087	0.0054	0.0087 [−0.0019; 0.0194]	27.4%	27.4%
Fixed effect model			**0.0104 [0.0048; 0.0160]**	**100.0%**	**––**
Random effects model			**0.0104 [0.0048; 0.0160]**	**––**	**100.0%**

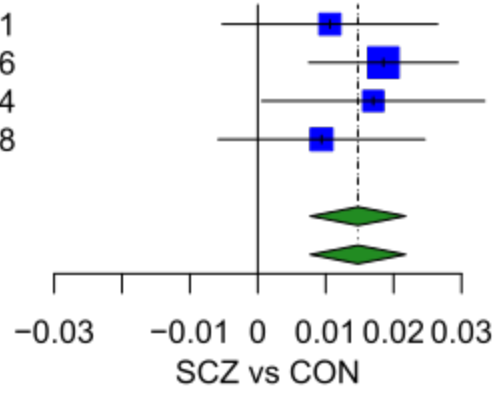
Heterogeneity: I^2 = 0%, p = 0.83



cg04173586

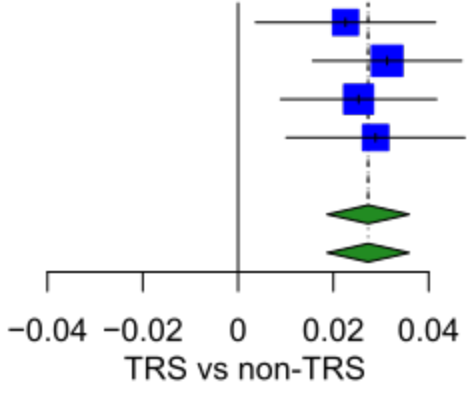
Study	TE	seTE	95%−CI	Weight (fixed)	Weight (random)
UCL	0.0106	0.0081	0.0106 [−0.0052; 0.0265]	19.6%	19.6%
Aberdeen	0.0185	0.0056	0.0185 [0.0075; 0.0295]	40.9%	40.9%
IoPPN	0.0170	0.0084	0.0170 [0.0006; 0.0334]	18.3%	18.3%
Sweden	0.0094	0.0078	0.0094 [−0.0058; 0.0246]	21.3%	21.3%
Fixed effect model			**0.0147 [0.0077; 0.0218]**	**100.0%**	**--**
Random effects model			**0.0147 [0.0077; 0.0218]**	**--**	**100.0%**

Heterogeneity: I^2 = 0%, p = 0.74

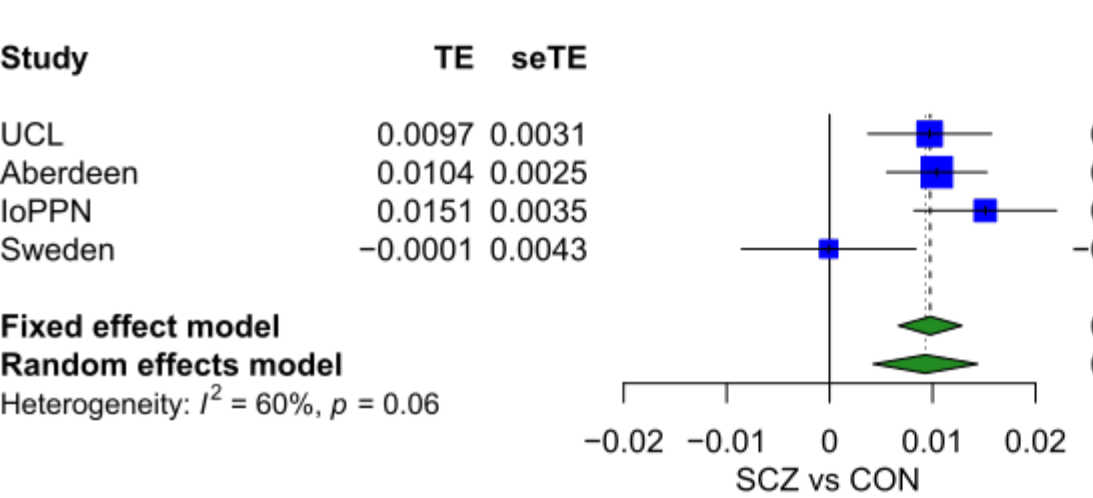


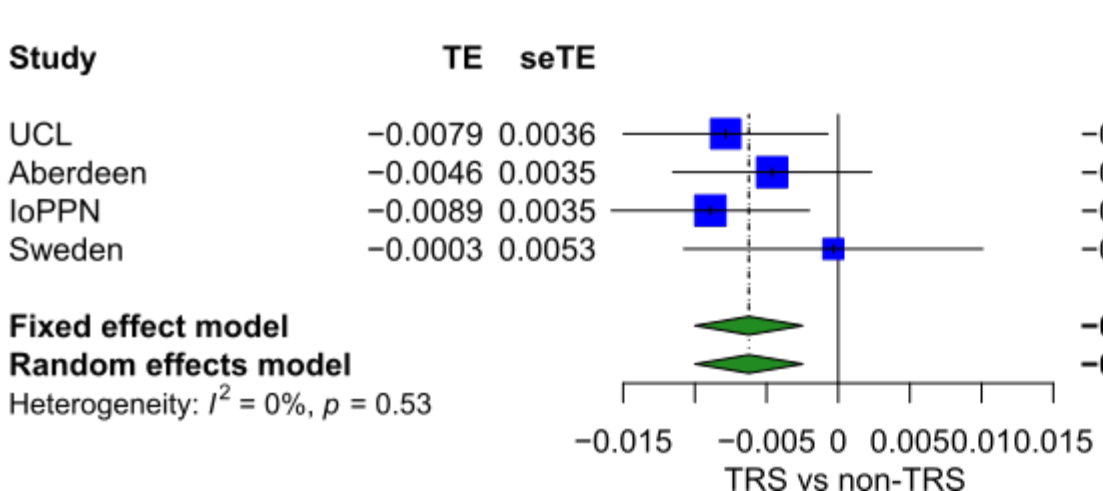
Study	TE	seTE	95%−CI	Weight (fixed)	Weight (random)
UCL	0.0226	0.0096	0.0226 [0.0037; 0.0414]	20.8%	20.8%
Aberdeen	0.0313	0.0080	0.0313 [0.0157; 0.0469]	30.4%	30.4%
IoPPN	0.0253	0.0084	0.0253 [0.0089; 0.0417]	27.7%	27.7%
Sweden	0.0289	0.0096	0.0289 [0.0101; 0.0476]	21.1%	21.1%
Fixed effect model			**0.0273 [0.0187; 0.0359]**	**100.0%**	**--**
Random effects model			**0.0273 [0.0187; 0.0359]**	**--**	**100.0%**

Heterogeneity: I^2 = 0%, p = 0.90



cg26263239


Study
TE
seTE
95%−CI
Weight (fixed)
Weight (random)
UCL 0.0097 0.0031 0.0097 [0.0037; 0.0157] 26.5% 26.5%
Aberdeen 0.0104 0.0025 0.0104 [0.0056; 0.0153] 40.3% 30.4%
IoPPN 0.0151 0.0035 0.0151 [0.0082; 0.0220] 19.9% 23.7%
Sweden −0.0001 0.0043 −0.0001 [−0.0085; 0.0084] 13.3% 19.4%
Fixed effect model 0.0098 [0.0067; 0.0129] 100.0% --
Random effects model 0.0093 [0.0042; 0.0144] -- 100.0%
Heterogeneity: I^2 = 60%, p = 0.06
−0.02 −0.01 0 0.01 0.02
SCZ vs CON


Study
TE
seTE
95%−CI
Weight (fixed)
Weight (random)
UCL −0.0079 0.0036 −0.0079 [−0.0150; −0.0007] 27.8% 27.8%
Aberdeen −0.0046 0.0035 −0.0046 [−0.0115; 0.0023] 29.5% 29.5%
IoPPN −0.0089 0.0035 −0.0089 [−0.0158; −0.0020] 29.7% 29.7%
Sweden −0.0003 0.0053 −0.0003 [−0.0108; 0.0101] 13.0% 13.0%
Fixed effect model −0.0062 [−0.0100; −0.0025] 100.0% --
Random effects model −0.0062 [−0.0100; −0.0025] -- 100.0%
Heterogeneity: I^2 = 0%, p = 0.53
−0.015 −0.005 0 0.005 0.010 0.015
TRS vs non-TRS