Supplemental Digital Content 1: text appendix – Detailed description of methods used for DNA extraction, DNA methylation assessment and data processing

Genomic DNA was extracted from peripheral blood mononuclear cells and buffy coat samples using the QIAamp 96 DNA blood kit (QIAGEN, Germany), and from dried blood spots on Guthrie cards as described previously.[1] Bisulphite conversion was performed using the EZ DNA Methylation-Gold kit (Zymo research, USA), according to the manufacturer's instructions. DNA methylation was measured using the Illumina Infinium HumanMethylation450K BeadChip (HM450K) array (Illumina, Inc., USA). Samples from each study (i.e. seven case-control studies and longitudinal re-analysis at follow-up in a subset of baseline control participants; Figure 1) were randomly assigned to chips and assayed according to the manufacturer's protocol in non-overlapping time periods. Samples were processed in batches of 96 samples per plate, and 12 samples per chip. All laboratory work was done at the Genetic Epidemiology Laboratory, The University of Melbourne.

Processing and normalization of DNA methylation data was conducted using the R package *minfi*.[2] The manufacturer's background correction based on internal control probes was applied. The subset-quantile within-array normalisation (SWAN) method was also applied to correct for technical discrepancies between the type I and type II probes of the HM450K array.[3] The sex of each sample was predicted using the 'getSex' function of the *minfi* package, and samples with discrepant predicted sex compared to recorded sex were excluded. For each sample, CpG sites with a detection *P*-value>0.01 were set to missing. Samples were excluded if more than 5% of probes had missing values, and, subsequently, CpG sites with more than 20% of samples missing were excluded. The exclusion was performed separately for baseline and follow-up data, and 484,830 and 484,826 CpG sites were included in the baseline and follow-up analysis, respectively. Duplicated samples were excluded on the basis of number of missing values (i.e. sample with least missing values was included). For each

CpG site, β-values were calculated, ranging from zero to one and corresponding to the proportion of methylation. The β-values were transformed into M-values using the following formula:

$$M = \log_2 \frac{\beta}{1 - \beta}$$

M-values are statistically more valid for differential methylation analysis.[4]

References

1. Joo JE, Wong EM, Baglietto L, et al. The use of DNA from archival dried blood spots with the Infinium HumanMethylation450 array. *BMC Biotechnol*. 2013;13:23.
2. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-9.
3. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol*. 2012;13(6):R44.
4. Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11:587.