Figure S1. Participant Flow Diagram

Visited the data center between 2006-2010 (baseline) and provide full informed consent to participate in the cohort (N=502,616)

Excluded for not providing data on recent caffeine drinking, habitual coffee and tea intake and data for at least one biomarker at baseline (N=54,822)

Participants in the main analysis (N=447,794)

Excluded sample outliers based on heterozygosity and missingness, participants with sex discrepancies between the self-reported and X-chromosome heterozygosity, those potentially related to other participants based on estimated kinship coefficients for all pairs of samples, those not self-reporting “British”, or those without similar (British) ancestral backgrounds based on results of principal component analysis (N=140,098)

Participants in the genetic analysis (N=307,696)