Supplemental for https://doi.org/10.1123/JAPA.2022-0067

Supplementary Methods S1

Genotyping methods in the Women's Health Initiative (WHI)

Data Sources. Genotyping data were generated from blood specimens collected at WHI enrollment. Genotyping data were aggregated from six WHI GWAS: 1) the Genomics and Randomized Trials Network (GARNET); 2) the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO); 3) the Hip Fracture GWAS (HipFx); 4) Modification of PM-Mediated Arrhythmogenesis in Populations (MOPMAP); 5) the SNP Health Association Resource (SHARe); and 6) the WHI Memory Study (WHIMS). Principal components (PCs) analysis was completed using a common set of 5,665 SNPs to create PCs to adjust for population stratification (Price et al., 2006).

Initial Quality Control. Initial quality control was performed using the GENEVA protocol or similar. Hardy Weinberg Equilibrium (HWE) was assessed using the chi-square test statistic comparing observed and expected genotype distributions. Deviations from HWE suggest that the allele distribution shifted possibly due to decreases in the frequency of deleterious variants due to censoring, among other reasons. HWE was assessed for all three SNPs of the GRS by race/ethnicity in both the WHI and OPACH samples using dosages converted to hard calls (P<10e-4 threshold defined out of HWE) (Shriner, 2011). Additional quality control parameters from the WHI sample are described in the table below.

	HipFx	SHARe	GARNET	WHIMS+	GECCO	MOPMAP
Minimal sample call rate	98%	95%	98%	97%	97%	95%
Minimal SNP call rate	98%	90%	98%	98%	98%	90%
Hardy Weinberg P-value cut-off below which SNPs are excluded	1e-4	1e-6	1e-4	1e-4	1e-4	1e-6
Samples used for	Controls	All	Unrelated	All	Controls	All

Hardy Weinberg calculations	of European-ancestry	samples, separate for Hispanics and African Americans	controls of European-ancestry			
Minimum allele frequency cut-off	1%	1%	None	1%	5%	0.5%

Imputation. All six GWAS were imputed using the 1,000 Genomes Project (1kGP) as a reference panel. Strands of the GWAS data were matched with the 1KGP data by comparing the letters of the alleles (ambiguous A/T or C/G SNPs were excluded) using the 1kGP reference panel (1,092 samples; v2.20101123 for GECCO; v3.20101123 for GARNET, HIPFX, MOPMAP, WHIMS+). The GWAS data were first split into chunks, containing 10,000 SNPs each. Neighboring chunks had 1,000 overlapping SNPs. All chunks were then phased and combined using BEAGLE (Browning & Browning, 2008). An autoclip file was created for minimac to specify the range of the chunks and SNPs to be imputed within the chunk to avoid >1 imputation. All chunks were imputed to the 1kGP imputation panel using minimac. Poorly imputed SNPs (r^2<0.10) were omitted. The present analyses included SNPs with r^2>0.20 imputation quality scores. The SHARe study was independently imputed to 1kGP using MACH (Howie et al., 2012). The X chromosome was not imputed. Genotype data derived from imputation were reported as continuous dosage values between 0 and 2, which accounts for the uncertainty of imputation and represents the expected number of copies of an allele at that SNP, conditional on the directly observed genotypes in the subject and phased haplotype assignments in the 1kGP samples.

Harmonization and duplicates. Data from the six GWAS were harmonized to create a dataset of genetic data from all studies. A panel of 5,665 SNPs was used to check the pairwise

concordance among all samples across all studies. This panel of SNPs was used for principal components (PC) analysis and study samples were projected onto HapMap space to identify ethnicity outliers. The same panel of SNPs were used for evaluating identity-by-descent in PLINK to identify relatedness among samples. Another PC analysis was completed for combined samples after removing of ineligible duplicates in all studies. The resulting PCs were mapped back to samples within each study. As participants for each GWAS were selected independently, duplicates were checked across studies. We removed a small number of samples considered to be duplicates but had a concordance rate <90% and appeared as duplicates but were from unrelated individuals, who were not monozygotic twins. There were 5 pairs of monozygotic twins (further detail outlined in *Relatedness* section).

Genetic ancestry. To identify participants with inconsistencies between genetic ancestry and self-reported ethnicity, PC analysis was performed using a subset of the 5,665 SNPs that were consistent between our samples and the reference panels. Eigenvectors were calculated using Eigenstrat (Price et al., 2006) and 475 publicly available samples from four ancestral populations including the Yourbans from Ibadan, Nigeria (YRI); Utah residents of Northern and Western European ancestry (CEU); the Human Genome Diversity Project (HGDP) East Asian population; and the HGDP Native American populations (Cann et al., 2002; Frazer et al., 2007).

Relatedness. Identity-by-descent analysis was completed using a subset of 5,665 SNPs and the PLINK package to identify parent-offspring pairs, sibling pairs, and first-degree relatives. One relative from each relative pair was retained for this analysis. If there was a group of relatives, one participant out of each group of relatives was randomly selected for inclusion prior to obtaining our analytic dataset. There were 42 parent off-spring pairs, 303 siblings/first degree relatives, and 5 pairs of monozygotic twins. Second and higher degree relatives were not identified.

References for Supplementary Methods S1:

Browning, B. L., & Browning, S. R. (2008). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American Journal of Human Genetics*, *84*(2), 210–223. https://doi.org/10.1016/j.ajhg.2009.01.005

Cann, H. M., de Toma, C., Cazes, L., Legrand, M. F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W. F., Bonne-Tamir, B., Cambon-Thomsen, A., & Chen, Z. (2002). A human genome diversity cell line panel. *Science (New York, NY)*, *296*(5566), 261. https://doi.org/10.1126/science.296.5566.261b

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., … Stewart, J. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, *449*(7164), 851–861. https://doi.org/10.1038/nature06258

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*, *44*(8), 955–959. https://doi.org/10.1038/ng.2354

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909. https://doi.org/10.1038/ng1847

Shriner, D. (2011). Approximate and exact tests of Hardy-Weinberg equilibrium using uncertain genotypes. *Genetic Epidemiology*, *35*(7), 632–637. https://doi.org/10.1002/gepi.20612