Study population

The Candidate Gene Association Resource (CARe) established by the National Heart, Lung, and Blood Institute (NHLBI) is composed of > 40,000 individuals representing 4 ethnic groups in 9 community-based cohorts: the Atherosclerosis Risk in Communities (ARIC) study, the Coronary Artery Risk Development in Young Adults study (CARDIA), the Cleveland Family Study (CFS), the Cardiovascular Health Study (CHS), the Cooperative Study of Sickle Cell Disease (CSSCD), the Framingham Heart Study (FHS), the Jackson Heart Study (JHS), the Multi-Ethnic Study of Atherosclerosis (MESA), and the Sleep Heart Health Study (SHHS) (Lettre et al. 2011; Musunuru et al. 2010). The six non-patient based cohorts with self-identified European American participants (ARIC, CARDIA, CFS, CHS, FHS, MESA) were used in the present study. Individuals were genotyped with the IBC chip and processed through quality control measures by CARe, as described elsewhere (Lo et al. 2011). In general, duplicate individuals, population outliers as determined by principal component analysis, cryptically related individuals in unrelated panels, and poorly genotyped individuals were removed from further analysis.

Phenotype definition

The reproductive traits examined here include age at natural menopause, number of children, age at first child, and age at last child. Descriptive statistics for each of the reproductive traits can be found in Table S2.

Age at natural menopause was treated as a continuous variable, defined as the age at the last menstrual period, after at least 12 consecutive months of amenorrhea. For the purpose of this study, we only included women with age at natural menopause between 40 and 60 years of age and excluded individuals whose menopause was induced by irradiation, hysterectomy and/or bilateral ovariectomy, or who used hormone replacement therapy before menopause. Residuals for each study were created separately by regressing age at natural menopause by cohort or study center where applicable.

For number of children, female participants who had not yet reached natural menopause or who had hysterectomy or oophorectomy before age 45 were removed. For the remaining women, the number of children was counted for each woman based on the self-reported number of live births. Number of children was analyzed as categorical variable using Poisson regression. Covariates included in the model were age, marital status, birth control use, education level, family income, study site, and the first 10 principal components. As the association statistics appeared to be unstable for low minor allele frequency (MAF) SNPs, we further restricted our analysis to SNPs with MAF > 0.02.
For age at last child, female participants who had not yet reached natural menopause or who had hysterectomy or oophorectomy before age 45 were also removed. If age at last term pregnancy was not reported, the age at last child was calculated based on the difference of the mother’s age and the age of the youngest child.

For age at first child, no exclusion based on menopausal status or surgery was applied. If age at first term pregnancy was not reported, the age at first child was calculated based on the difference of the mother’s age and the age of the oldest child.

Note that as it is difficult to differentiate biological children from step or adopted children in self-reported surveys, no such distinctions were stringently made but information on biological children were always used if available (FHS and MESA for age at first child; FHS for age at last child).

Age at first and last child were analyzed as continuous variables. Age at first child was log-transformed. Stratified by cohort, multivariate linear regression models were constructed for both phenotypes controlling for birth control pill use, education status, marital status, study site, income status, and age, where applicable. The residuals were normalized to a standard normal distribution. The associations of hcSNPs to these two traits were tested using linear regression, with the top 10 principal components as additional covariates.

The overall fitness traits examined here include longevity, BMI and height.

For longevity, cases (long-lived individuals) were defined as individuals surviving to \( \geq 85 \) years of age at time of death or last contact, and controls (short-lived individuals) were defined as individuals who died \( \leq 75 \) years of age. Longevity was analyzed as a dichotomous trait using logistic regression model, which included gender, study site, and the top 10 principal components as covariates.

For BMI, individuals \(< 20 \) years of age were excluded. Stratified by cohort and gender, raw BMI was regressed on age, age\(^2\), and study site within cohort, if applicable. The residuals were then fit to a standard normal distribution before combining the gender-specific residuals within cohort. BMI was analyzed as continuous trait using linear regression, with the top 10 principal components included as covariates.

For height, we excluded men \(< 23 \) years of age and women \(< 21 \) years of age, as well as individuals \( > 85 \) years of age. Stratified by cohort and gender, we regressed height on age and study site, when available. Residuals were normalized to a standard normal distribution. Outliers \((> 4 \) SD or \(< -4 \) SD\) were excluded from the analyses. Height was analyzed as continuous trait using linear regression, with the top 10 principal components included as covariates.

The number of individuals analyzed for each phenotype can be found in Table 1, organized by each of the CARe cohorts.
REFERENCES

