nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-019-12283-6 OPEN

Associations of autozygosity with a broad range of human phenotypes

David W Clark et al.#

In many species, the offspring of related parents suffer reduced reproductive success, a phenomenon known as inbreeding depression. In humans, the importance of this effect has remained unclear, partly because reproduction between close relatives is both rare and frequently associated with confounding social factors. Here, using genomic inbreeding coefficients (F_{ROH}) for >1.4 million individuals, we show that F_{ROH} is significantly associated ($p < 0.0005$) with apparently deleterious changes in 32 out of 100 traits analysed. These changes are associated with runs of homozygosity (ROH), but not with common variant homozygosity, suggesting that genetic variants associated with inbreeding depression are predominantly rare. The effect on fertility is striking: F_{ROH} equivalent to the offspring of first cousins is associated with a 55% decrease [95% CI 44–66%] in the odds of having children. Finally, the effects of F_{ROH} are confirmed within full-sibling pairs, where the variation in F_{ROH} is independent of all environmental confounding.

*email: jim.wilson@ed.ac.uk. #A full list of authors and their affiliations appears at the end of the paper.

Given the pervasive impact of purifying selection on all populations, it is expected that genetic variants with large deleterious effects on evolutionary fitness will be both rare and recessive[1]. However, precisely because they are rare, most of these variants have yet to be identified and their recessive impact on the global burden of disease is poorly understood. This is of particular importance for the nearly one billion people living in populations where consanguineous marriages are common[2], and the burden of genetic disease is thought to be disproportionately due to increased homozygosity of rare, recessive variants[3–5]. Although individual recessive variants are difficult to identify, the net directional effect of all recessive variants on phenotypes can be quantified by studying the effect of inbreeding[6], which gives rise to autozygosity (homozygosity due to inheritance of an allele identical-by-descent).

Levels of autozygosity are low in most of the cohorts with genome-wide data[7,8] and consequently very large samples are required to study the phenotypic impact of inbreeding[9]. Here, we meta-analyse results from 119 independent cohorts to quantify the effect of inbreeding on 45 commonly measured complex traits of biomedical or evolutionary importance, and supplement these with analysis of 55 more rarely measured traits included in UK Biobank[10].

Continuous segments of homozygous alleles, or runs of homozygosity (ROH), arise when identical-by-descent haplotypes are inherited down both sides of a family. The fraction of each autosomal genome in ROH > 1.5 Mb (F_{ROH}) correlates well with pedigree-based estimates of inbreeding[11].We estimate F_{ROH} using standard methods and software[6,12] for a total of 1,401,776 individuals in 234 uniform sub-cohorts. The traits measured in each cohort vary according to original study purpose, but together cover a comprehensive range of human phenotypes (Fig. 1, Supplementary Data 7). The five most frequently contributed traits (height, weight, body mass index, systolic and diastolic blood pressure) are measured in >1,000,000 individuals; a further 16 traits are measured >500,000 times.

We find that F_{ROH} is significantly associated with apparently deleterious changes in 32 out of 100 traits analysed. Increased F_{ROH} is associated with reduced reproductive success (decreased number and likelihood of having children, older age at first sex and first birth, decreased number of sexual partners), as well as reduced risk-taking behaviour (alcohol intake, ever-smoked, self-reported risk taking) and increased disease risk (self-reported overall health and risk factors including grip strength and heart rate). We show that the observed effects are predominantly associated with rare (not common) variants and, for a subset of traits, differ between men and women. Finally, we introduce a within-siblings method, which confirms that social confounding of F_{ROH} is modest for most traits. We therefore conclude that inbreeding depression influences a broad range of human phenotypes through the action of rare, recessive variants.

Results

Cohort characteristics. As expected, cohorts with different demographic histories varied widely in mean F_{ROH}. The within-cohort standard deviation of F_{ROH} is strongly correlated with the mean (Pearson's $r = 0.82$; Supplementary Fig. 3), and the most homozygous cohorts provide up to 100 times greater per-sample statistical power than cosmopolitan European-ancestry cohorts (Supplementary Data 5). To categorise cohorts, we plotted mean F_{ROH} against F_{IS} (Fig. 2). F_{IS} measures inbreeding as reflected by non-random mating in the most recent generation, and is calculated as the mean individual departure from Hardy–Weinberg equilibrium (F_{SNP}; see Methods). Cohorts with high rates of consanguinity lie near the $F_{ROH} = F_{IS}$ line, since most excess SNP homozygosity is caused by ROH. In contrast, cohorts with small effective population sizes, such as the Amish and Hutterite isolates of North America, have high average F_{ROH}, often despite avoidance of mating with known relatives, since identical-by-descent haplotypes are carried by many couples, due to a restricted number of possible ancestors.

Traits affected by F_{ROH}. To estimate the effect of inbreeding on each of the 100 phenotypes studied, trait values were regressed on F_{ROH} within each cohort, taking account of covariates including

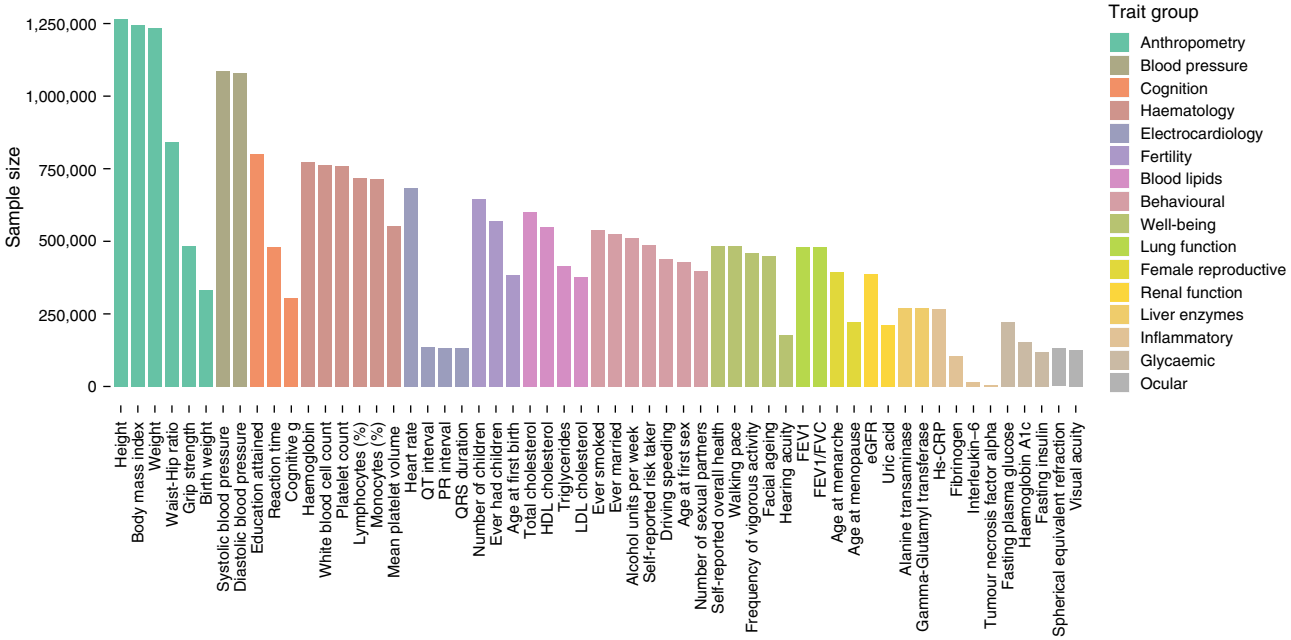


Fig. 1 Census of complex traits. Sample sizes are given for analyses of 57 representative phenotypes, arranged into 16 groups covering major organ systems and disease risk factors. HDL high-density lipoprotein, LDL low-density lipoprotein, hs-CRP high-sensitivity C-reactive protein, TNF-alpha tumour necrosis factor alpha, FEV1 forced expiratory volume in one second, FVC forced vital capacity, eGFR estimated glomerular filtration rate

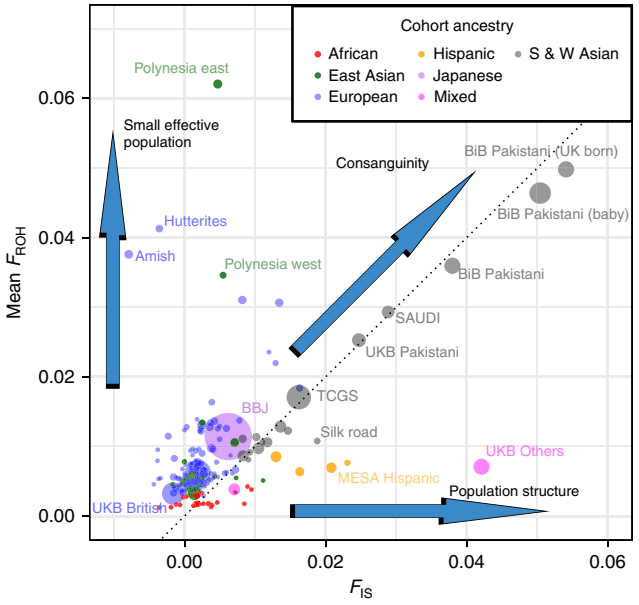


Fig. 2 Mean F_{ROH} and F_{IS} for 234 ROHgen sub-cohorts. Each cohort is represented by a circle whose area is proportional to the approximate statistical power ($N\sigma^2_{F_{ROH}}$) contributed to estimates of $\beta_{F_{ROH}}$. Mean F_{ROH} can be considered as an estimate of total inbreeding relative to an unknown base generation, approximately tens of generations past. F_{IS} measures inbreeding in the current generation, with $F_{IS} = 0$ indicating random mating, $F_{IS} > 0$ indicating consanguinity, and $F_{IS} < 0$ inbreeding avoidance[46]. In cohorts along the *y*-axis, such as the Polynesians and the Anabaptist isolates, autozygosity is primarily caused by small effective population size rather than preferential consanguineous unions. In contrast, in cohorts along the dotted unity line, all excess SNP homozygosity is accounted for by ROH, as expected of consanguinity within a large effective population. A small number of cohorts along the *x*-axis, such as Hispanic and mixed-race groups, show excess SNP homozygosity without elevated mean F_{ROH}, indicating population genetic structuring, caused for instance by admixture and known as the Wahlund effect. A few notable cohorts are labelled. BBJ Biobank Japan, BiB Born in Bradford, UKB UK Biobank, MESA Multiethnic Study of Atherosclerosis, TCGS Tehran Cardiometabolic Genetic Study

age, sex, principal components of ancestry and, in family studies, a genomic relationship matrix (GRM) (Supplementary Data 3). Cross-cohort effect size estimates were then obtained by fixed-effect, inverse variance-weighted meta-analysis of the within-cohort estimates (Supplementary Data 10). Twenty-seven out of 79 quantitative traits and 5 out of 21 binary traits reach experiment-wise significance (0.05/100 or $p < 0.0005$; Fig. 3a, b). Among these are replications of the previously reported effects on reduction in height[13], forced expiratory lung volume in one second, cognition and education attained[6]. We find that the 32 phenotypes affected by inbreeding can be grouped into five broader categories: reproductive success, risky behaviours, cognitive ability, body size, and health.

Despite the greater individual control over reproduction in the modern era, due to contraception and fertility treatments, we find that increased F_{ROH} has significant negative effects on five traits closely related to fertility. For example, an increase of 0.0625 in F_{ROH} (equivalent to the difference between the offspring of first cousins and those of unrelated parents) is associated with having 0.10 fewer children [$\beta_{0.0625} = -0.10 \pm 0.03$ 95% confidence interval (CI), $p = 1.8 \times 10^{-10}$]. This effect is due to increased F_{ROH} being associated with reduced odds of having any children ($OR_{0.0625} = 0.65 \pm 0.04$, $p = 1.7 \times 10^{-32}$) as opposed to fewer children among parents ($\beta_{0.0625} = 0.007 \pm 0.03$, $p = 0.66$). Since autozygosity also decreases the likelihood of having children in the subset of individuals who are, or have been, married, ($OR_{0.0625} = 0.71 \pm 0.09$, $p = 3.8 \times 10^{-8}$) it appears that the cause is a reduced ability or desire to have children, rather than reduced opportunity. Consistent with this interpretation, we observe no significant effect on the likelihood of marriage ($OR_{0.0625} = 0.94 \pm 0.07$, $p = 0.12$) (Fig. 3b). All effect size, odds ratios and 95% CI are stated as the difference between $F_{ROH} = 0$ and $F_{ROH} = 0.0625$.

The effects on fertility may be partly explained by the effect of F_{ROH} on a second group of traits, which capture risky or addictive behaviour. Increased F_{ROH} is associated with later age at first sex ($\beta_{0.0625} = 0.83 \pm 0.19$ years, $p = 5.8 \times 10^{-17}$) and fewer sexual partners ($\beta_{0.0625} = -1.38 \pm 0.38$, $p = 2.0 \times 10^{-12}$) but also reduced alcohol consumption ($\beta_{0.0625} = -0.66 \pm 0.12$ units per week, $p = 1.3 \times 10^{-22}$), decreased likelihood of smoking ($OR_{0.0625} = 0.79 \pm 0.05$, $p = 5.9 \times 10^{-13}$), and a lower probability of being a self-declared risk-taker ($OR_{0.0625} = 0.84 \pm 0.06$, $p = 3.4 \times 10^{-5}$) or exceeding the speed limit on a motorway ($p = 4.0 \times 10^{-8}$). Conservative beliefs are likely to affect these traits, and are known to be confounded with F_{ROH} in some populations[14], however, fitting religious participation as a covariate in UKB reduces, but does not eliminate the reported effects (Supplementary Fig. 10b, Supplementary Data 20). Similarly, fitting educational attainment as an additional covariate reduces 16 of 25 significant effect estimates, but actually increases 9, including age at first sex and number of children (Supplementary Fig. 10a, Supplementary Data 20). This is because reduced educational attainment is associated with earlier age at first sex and increased number of children, which makes it an unlikely confounder for the effects of F_{ROH}, which are in the opposite directions.

A third group of traits relates to cognitive ability. As previously reported, increased autozygosity is associated with decreased general cognitive ability, *g*[6,15] and reduced educational attainment[6]. Here, we also observe an increase in reaction time ($\beta_{0.0625} = 11.6 \pm 3.9$ ms, $p = 6.5 \times 10^{-9}$), a correlate of general cognitive ability (Fig. 3a, Supplementary Data 10).

A fourth group relates to body size. We replicate previously reported decreases in height and forced expiratory volume[6] (Supplementary Data 21) and we find that increased F_{ROH} is correlated with a reduction in weight ($\beta_{0.0625} = 0.86 \pm 0.12$ kg, $p = 3.4 \times 10^{-28}$) and an increase in the waist to hip ratio ($\beta_{0.0625} = 0.004 \pm 0.001$, $p = 1.4 \times 10^{-11}$).

The remaining effects are loosely related to health and frailty; higher F_{ROH} individuals report significantly lower overall health and slower walking pace, have reduced grip strength ($\beta_{0.0625} = -1.24 \pm 0.19$ kg, $p = 6.9 \times 10^{-24}$), accelerated self-reported facial ageing, and poorer eyesight and hearing. Increased F_{ROH} is also associated with faster heart rate ($\beta_{0.0625} = 0.56 \pm 0.24$ bpm, $p = 5.9 \times 10^{-6}$), lower haemoglobin ($\beta_{0.0625} = 0.81 \pm 0.24$ gL^{-1}, $p = 1.6 \times 10^{-11}$), lymphocyte percentage, and total cholesterol ($\beta_{0.0625} = -0.05 \pm 0.015$ mmol L^{-1}, $p = 5.2 \times 10^{-10}$).

Sex-specific effects of F_{ROH}. Intriguingly, for a minority of traits (13/100), the effect of F_{ROH} differs between men and women (Fig. 3c, Supplementary Data 12). For example, men who are the offspring of first cousins have 0.10 mmol L^{-1} [95% CI 0.08–0.12] lower total cholesterol on average, while there is no significant effect in women; LDL shows a similar pattern. More generally, for these traits, the effect in men is often of greater magnitude than the effect in women, perhaps reflecting differing relationships between phenotype and fitness.

Associations most likely caused by rare, recessive variants. The use of ROH to estimate inbreeding coefficients is relatively new in inbreeding research[11,16–19]. Earlier frequency-based estimators

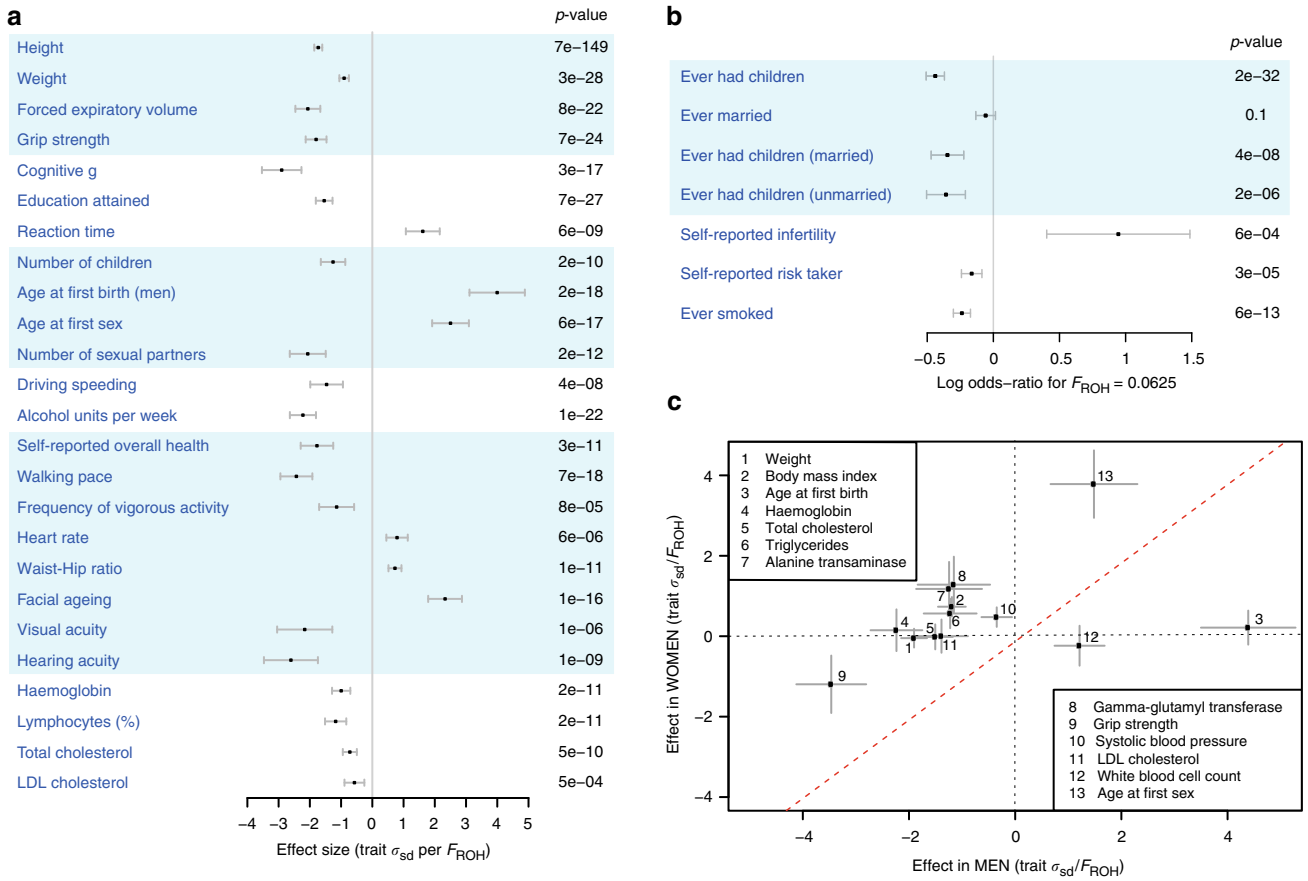


Fig. 3 Scope of inbreeding depression. **a** Effect of F_{ROH} on 25 quantitative traits. To facilitate comparison between traits, effect estimates are presented in units of within-sex standard deviations. Traits shown here reached Bonferroni-corrected significance of $p = 0.0005$ (=0.05/100 traits). Sample sizes, within-sex standard deviations, and effect estimates in measurement units are shown in Supplementary Data 9. FEV1 forced expiratory volume in one second. Traits are grouped by type. **b** Effect of F_{ROH} on eight binary traits with associated p values. Effect estimates are reported as ln(Odds-Ratio) for the offspring of first cousins, for which $E(F_{ROH}) = 0.0625$. Self-declared infertility is shown for information, although this trait does not reach Bonferroni corrected significant ($OR_{0.0625} = 2.6 \pm 1.1$, $p = 0.0006$). Numbers of cases and controls and effect estimates for all binary traits are shown in Supplementary Data 10. **c** Sex-specificity of ROH effects. The effect of F_{ROH} in men versus that in women is shown for 13 traits for which there was evidence of significant differences in the effects between sexes. For 11 of these 13 traits the magnitude of effect is greater in men than in women. Traits such as liver enzymes levels (alanine transaminase, gamma-glutamyl transferase) show sex-specific effects of opposite sign (positive in women, negative in men), which cancel out in the overall analysis. BMI body mass index, LDL low-density lipoprotein. All errors bars represent 95% confidence intervals

such as F_{SNP} and F_{GRM}[20], made use of excess marker homozygosity[21–23] and did not require physical maps. We performed both univariate and multivariate regressions to evaluate the effectiveness of F_{ROH} against these measures. The correlations between them range from 0.13 to 0.99 and are strongest in cohorts with high average inbreeding (Supplementary Data 6, Supplementary Fig. 6). Significantly, univariate regressions of traits on both F_{SNP} and F_{GRM} show attenuated effect estimates relative to F_{ROH} (Supplementary Data 13). This attenuation is greatest in low autozygosity cohorts, suggesting that F_{ROH} is a better estimator of excess homozygosity at the causal loci (Fig. 4c).

To explore this further, we fit bivariate models with F_{ROH} and F_{GRM} as explanatory variables. For all 32 traits that were significant in the univariate analysis, we find that $\widehat{\beta}_{F_{ROH}|F_{GRM}}$ is of greater magnitude than $\widehat{\beta}_{FGRM|FROH}$in the conditional analysis (Fig. 4b, Supplementary Data 22). This suggests that inbreeding depression is predominantly caused by rare, recessive variants made homozygous in ROH, and not by the chance homozygosity of variants in strong LD with common SNPs (Fig. 4d, Supplementary Note 5). We also find that ROH of different lengths have similar effects per unit length (Fig. 4a, Supplementary Fig. 11a), consistent with their having a causal effect on traits and not with confounding by socioeconomic or other factors, as shorter ROH arise from deep in the pedigree are thus less correlated with recent consanguinity.

Quantifying the scope of social confounding. Previous studies have highlighted the potential for F_{ROH} to be confounded by non-genetic factors[6,24]. We therefore estimated the effect of F_{ROH} within various groups, between which confounding might be expected either to differ, or not be present at all.

For example, the effect of F_{ROH} on height is consistent across seven major continental ancestry groups (Supplementary Fig. 1, Supplementary Data 18), despite differing attitudes towards consanguinity, and consequently different burdens and origins of ROH. Similarly, grouping cohorts into consanguineous, more cosmopolitan, admixed and those with homozygosity due to ancient founder effects also shows consistent effects (Supplementary Fig. 2, Supplementary Data 19). Equally, categorising samples into bins of increasing F_{ROH} shows a dose-dependent response of the study traits with increased F_{ROH} (Supplementary Data 17 and

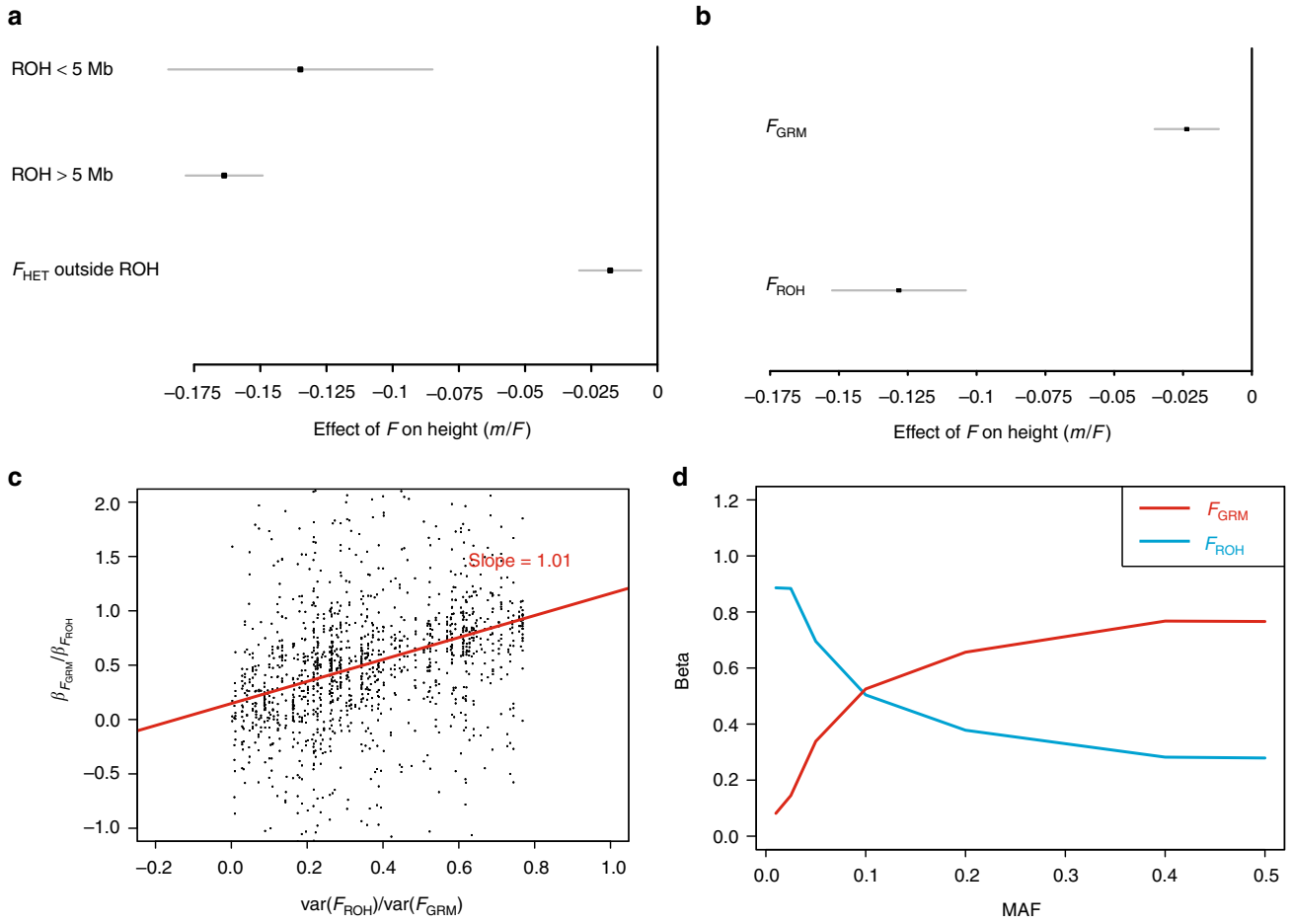


Fig. 4 Inbreeding depression caused by ROH. **a** Effect of different ROH lengths on height, compared with the effect of SNP homozygosity outside of ROH. The effects of shorter (<5 Mb) and longer (>5 Mb) ROH per unit length are similar and strongly negative, whereas the effect of homozygosity outside ROH is much weaker. The pattern is similar for other traits (Supplementary Fig. 11a; Supplementary Data 14). **b** $\boldsymbol{F_{ROH}}$ is more strongly associated than $\boldsymbol{F_{GRM}}$ in a bivariate model of height. Meta-analysed effect estimates, and 95% confidence intervals, are shown for a bivariate model of height (Height $\sim F_{ROH} + F_{GRM}$). The reduction in height is more strongly associated with F_{ROH} than F_{GRM}, as predicted if the causal variants are in weak LD with the common SNPs used to calculate F_{GRM} (Supplementary Note 5). The pattern is similar for other traits (Supplementary Fig. 15a, b; Supplementary Data 22). **c** F_{ROH} is a lower variance estimator of the inbreeding coefficient than F_{GRM}. The ratio of $\beta_{F_{GRM}} : \beta_{F_{ROH}}$ is plotted against $\frac{\mathrm{var}(F_{ROH})}{\mathrm{var}(F_{GRM})}$ for all traits in all cohorts. When the variation of F_{GRM} which is independent of F_{ROH} has no effect on traits, $\hat{\beta}_{F_{GRM}}$ is downwardly biased by a factor of $\frac{\mathrm{var}(F_{ROH})}{\mathrm{var}(F_{GRM})}$ (Supplementary Note 4). A linear maximum likelihood fit, shown in red, has a gradient consistent with unity [1.01; 95% CI 0.84–1.18], as expected when the difference between F_{GRM} and F_{ROH} is not informative about the excess homozygosity at causal variants (Supplementary Note 5). **d** F_{ROH} is a better predictor of rare variant homozygosity than F_{GRM}. The excess homozygosities of SNPs, extracted from UK Biobank imputed genotypes, were calculated at seven discrete minor allele frequencies (F_{MAF}), and regressed on two estimators of inbreeding in a bivariate statistical model (see Supplementary Note 5). The homozygosity of common SNPs is better predicted by F_{GRM}, but rare variant homozygosity is better predicted by F_{ROH}. The results from real data (Fig. 4b, Supplementary Figs 15a, b and Supplementary Data 22) are consistent with those simulated here, if the causal variants are predominantly rare. All errors bars represent 95% confidence intervals

Fig. 5a, b show the response for height and ever having children; Supplementary Figs 9a–f for all significant traits). The proportionality of these effects is consistent with a genetic cause, while it is difficult to envisage a confounder proportionally associated across the *entire* range of observed F_{ROH}. In particular, the highest F_{ROH} group ($F_{ROH} > 0.18$), equivalent to the offspring of first-degree relatives, are found to be, on average, 3.4 [95% CI 2.5–4.3] cm shorter and 3.1 [95% CI 2.5–3.7] times more likely to be childless than an $F_{ROH} = 0$ individual.

Next, we estimated $\beta_{F_{ROH}}$ for 7153 self-declared adopted individuals in UK Biobank, whose genotype is less likely to be confounded by cultural factors associated with the relatedness of their biological parents. For all 26 significant traits measured in this cohort, effect estimates are directionally consistent with the meta-analysis and 3 (height, walking pace and hearing acuity) reach replication significance ($p < 0.004$). In addition, a meta-analysis of the ratio $\widehat{\beta}_{F_{ROH\_ADOPTEE}} : \widehat{\beta}_{F_{ROH}}$ across all traits differs significantly from zero (Fig. 5c; average = 0.78, 95% CI 0.56–1.00, $p = 2 \times 10^{-12}$).

Finally, the effect of F_{ROH} was estimated in up to 118,773 individuals in sibships (full-sibling pairs, trios, etc.: $\widehat{\beta}_{F_{ROH\_wSibs}}$). F_{ROH} differences between siblings are caused entirely by Mendelian segregation, and are thus independent of any reasonable model of confounding. The variation of F_{ROH} among siblings is a small fraction of the population-wide variation[11] (Supplementary Data 5); nevertheless, 23 out of 29 estimates of $\widehat{\beta}_{F_{ROH\_wSibs}}$ are directionally consistent with $\widehat{\beta}_{F_{ROH}}$, and two (self-reported overall health and ever having children) reach replication

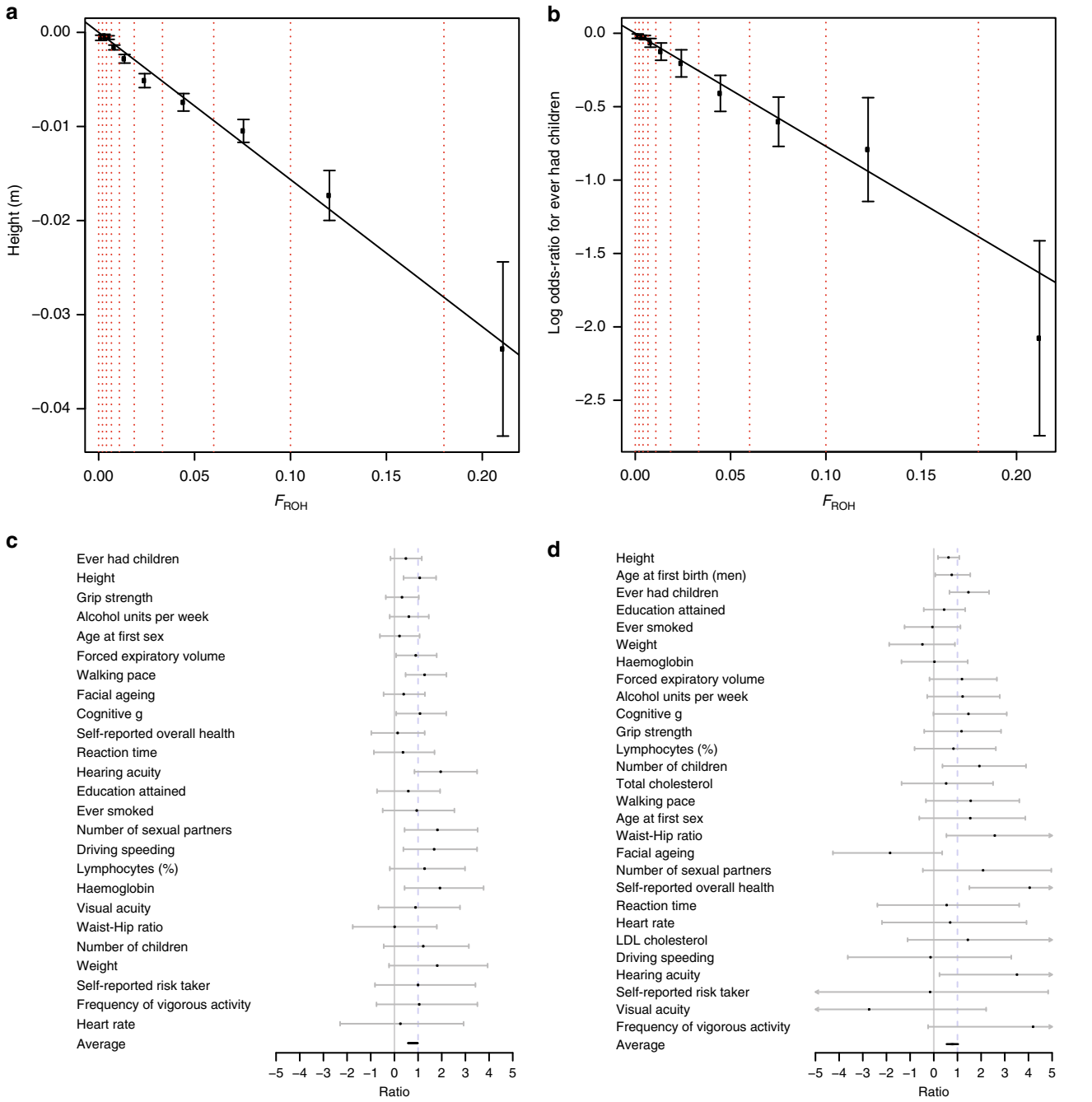


Fig. 5 Evidence ROH effects are un-confounded. **a** Linear decrease in height with increasing F_{ROH}. Average heights (in metres) is plotted in bins of increasing F_{ROH}. The limits of each bin are shown by red dotted lines, and correspond to the offspring of increasing degree unions left-to-right. The overall estimate of $\beta_{F_{ROH}}$ is shown as a solid black line. Subjects with kinship equal to offspring of full-sibling or parent-child unions are significantly shorter than those of avuncular or half-sibling unions who in turn are significantly shorter than those of first-cousin unions. **b** Linear decrease in odds of ever having children with increasing F_{ROH}. Linear model approximations of ln(Odds-Ratio) for ever having children (1 = parous, 0 = childless) are plotted in bins of increasing F_{ROH}. A strong relationship is evident, extending beyond the offspring of first cousins. **c** ROH effects are consistent in adoptees. The ratios of effect estimates, $\beta_{F_{ROH}}$, between adoptees and all individuals are presented by trait. All traits are directionally consistent and overall show a strongly significant difference from zero (average = 0.78, 95% CI 0.56–1.00, $p = 2 \times 10^{-12}$). FEV1 forced expiratory volume in one second. **d** ROH effects are consistent in full siblings. The ratios of effect estimates within full siblings to effects in all individuals ($\beta_{F_{ROH\_wSibs}} : \beta_{F_{ROH}}$) are presented by trait. Twenty-three of 29 estimates are directionally consistent and overall show a significant difference from zero (average = 0.78, 95% CI 0.53–1.04, $p = 7 \times 10^{-10}$). BMI body mass index. All errors bars represent 95% confidence intervals

significance. A meta-analysis of the ratio $\widehat{\beta}_{F_{\mathrm{ROH\_wSibs}}} : \widehat{\beta}_{F_{\mathrm{ROH}}}$ for all traits is significantly greater than zero (Fig. 5d; average = 0.78, 95% CI 0.53–1.04, $p = 7 \times 10^{-10}$), indicating a substantial fraction of these effects is genetic in origin. However, for both adoptees and siblings, the point estimates are less than one, suggesting that non-genetic factors probably contribute a small, but significant, fraction of the observed effects.

Discussion

Our results reveal inbreeding depression to be broad in scope, influencing both complex traits related to evolutionary fitness and others where the pattern of selection is less clear. While studies of couples show optimal fertility for those with distant kinship[25,26], fewer have examined reproductive success as a function of individual inbreeding. Those that did are orders of magnitude smaller in size than the present study, suffer the attendant drawbacks of pedigree analysis, and have found mixed results[27–29]. Our genomic approach also reveals that in addition to socio-demographic factors and individual choice, recessive genetic effects have a significant influence on whether individuals reproduce. The discordant effects on fertility and education demonstrate that this is not just a result of genetic correlations between the two domains[30].

The effects we see on fertility might be partially mediated through a hitherto unknown effect of autozygosity on decreasing the prevalence of risk-taking behaviours. Significant effects of autozygosity are observed for self-reported risk taking, speeding on motorways, alcohol and smoking behaviour, age at first sexual intercourse and number of sexual partners. Independent evidence for a shared genetic architecture between risk-taking and fertility traits comes from analysis of genetic correlations using LD-score regression in UKB (Supplementary Table 1). The core fertility traits, ever had children and number of children, are strongly genetically correlated ($r_G = 0.93$; $p < 10^{-100}$). Genetic correlations with ever-smoking and self-reported risk-taking are lower, but also significant: 0.23–0.27, $p < 10^{-10}$. Age at first sex is strongly genetically correlated both with the fertility traits, ($r_G =$ 0.53–0.57), and number of sexual partners, ever-smoking and risk-taking[30] ($r_G = 0.42$–0.60).

Reproductive traits are understandable targets of natural selection, as might be walking speed, grip strength, overall health, and visual and auditory acuity. While we cannot completely exclude reverse causality, whereby a less risk-taking, more conservative, personality is associated with greater likelihood of consanguineous marriage, we note that the effects are consistent for ROH < 5 Mb, which are less confounded with mate choice, due to their more distant pedigree origins (Supplementary Fig. 11a). This group of traits also shows similar evidence for unconfounded effects in the analysis of adoptees and full siblings (Fig. 5c, d; Supplementary Data 16) and the signals remained after correcting for religious activity or education.

On the other hand, for some traits that we expected to be influenced by ROH, we observed no effect. For example, birth weight is considered a key component of evolutionary fitness in mammals, and is influenced by genomic homozygosity in deer[31]; however, no material effect is apparent here (Supplementary Data 10). Furthermore, in one case, ROH appear to provide a beneficial effect: increasing F_{ROH} significantly decreases total and LDL-cholesterol in men, and may thus be cardio-protective in this regard.

Our multivariate models show that homozygosity at common SNPs outside of ROH has little influence on traits, and that the effect rather comes from ROH over 1.5 Mb in length. This suggests that genetic variants causing inbreeding depression are almost entirely rare, consistent with the dominance hypothesis[1]. The alternative hypothesis of overdominance, whereby positive selection on heterozygotes has brought alleles to intermediate frequencies, would predict that more common homozygous SNPs outside long ROH would also confer an effect. The differential provides evidence in humans that rare recessive mutations underlie the quantitative effects of inbreeding depression.

Previous studies have shown that associations observed between F_{ROH} and traits do not prove a causal relationship[14,24]. Traditional Genome-wide Association Studies (GWAS) can infer causality because, in the absence of population structure, genetic variants (SNPs) are randomly distributed between, and within, different social groups. However, this assumption does not hold in studies of inbreeding depression, where, even within a genetically homogeneous population, social groups may have differing attitudes towards consanguinity, and therefore different average F_{ROH} and, potentially, different average trait values. We therefore present a number of analyses that discount social confounding as a major factor in our results. Firstly, we show that the effects are consistent across diverse populations, including those where ROH burden is driven by founder effects rather than cultural practices regarding marriage. Effects are also consistent across a 20-fold range of F_{ROH}: from low levels, likely unknown to the subject, to extremely high levels only seen in the offspring of first-degree relatives. Secondly, we show that the effects of ROH are consistent in direction and magnitude among adopted individuals, and also for short ROH which are not informative about parental relatedness. Finally, we introduce a within-siblings method, independent of all confounders, that confirms a genetic explanation for most of the observed effects. Variation in F_{ROH} between siblings is caused entirely by random Mendelian segregation; we show that higher F_{ROH} siblings experience poorer overall health and lower reproductive success, as well as other changes consistent with population-wide estimates. Nevertheless, average effect sizes from both adoptees and siblings are 20% smaller than population-wide estimates, confirming the importance of accounting for social confounding in future studies of human inbreeding depression.

Our results reveal five large groups of phenotypes sensitive to inbreeding depression, including some known to be closely linked to evolutionary fitness, but also others where the connection is, with current knowledge, more surprising. The effects are mediated by ROH rather than homozygosity of common SNPs, causally implicating rare recessive variants rather than overdominance as the most important underlying mechanism. Identification of these recessive variants will be challenging, but analysis of regional ROH and in particular using whole-genome sequences in large cohorts with sufficient variance in autozygosity will be the first step. Founder populations or those which prefer consanguineous marriage will provide the most power to understand this fundamental phenomenon.

see Supplementary Data.

Methods

Overview. Our initial aim was to estimate the effect of F_{ROH} on 45 quantitative traits and to assess whether any of these effects differed significantly from zero. Previous work[7,11] has shown that inbreeding coefficients are low in most human populations, and that very large samples are required to reliably estimate the genetic effects of inbreeding[13]. To maximise sample size, a collaborative consortium (ROHgen[6]) was established, and research groups administering cohorts with SNP chip genotyping were invited to participate. To ensure that all participants performed uniform and repeatable analyses, a semi-automated software pipeline was developed and executed locally by each research group. This software pipeline required cohorts to provide only quality-controlled genotypes (in plink binary format) and standardised phenotypes (in plain-text) and used standard software (R, PLINK[12,32], KING[33]) to perform the analyses described below. Results from each cohort were returned to the central ROHgen analysts for meta-analysis.

During the initial meta-analysis, genotypes were released for >500,000 samples from the richly phenotyped UK Biobank (UKB)[10]. It was therefore decided to add a

further 34 quantitative phenotypes and 21 binary traits to the ROHgen analysis. Many of these additional traits were unique to UKB, although 7 were also available in a subset of ROHgen cohorts willing to run additional analyses. In total, the effect of F_{ROH} was tested on 100 traits and therefore experiment-wise significance was defined as 5×10^{-4} (=0.05/100).

Cohort recruitment. In total, 119 independent, genetic epidemiological study cohorts were contributed to ROHgen. Of these, 118 were studies of adults and contributed multiple phenotypes, while 1 was a study of children and contributed only birth weight. To minimise any potential confounding or bias caused by within-study heterogeneity, studies were split into single-ethnicity sub-cohorts wherever applicable. Each sub-cohort was required to use only one genotyping array and be of uniform ancestry and case-status. For example, if a study contained multiple distinct ethnicities, sub-cohorts of each ancestry were created and analysed separately. At minimum, ancestry was defined on a sub-continental scale (i.e. European, African, East Asian, South Asian, West Asian, Japanese, and Hispanic were always analysed separately) but more precise separation was used when deemed necessary, for example, in cohorts with large representation of Ashkenazi Jews. In case-control studies of disease, separate sub-cohorts were created for cases and controls and phenotypes associated with disease status were not analysed in the case cohort: for example, fasting plasma glucose was not analysed in Type 2 diabetes case cohorts. Occasionally, cohorts had been genotyped on different SNP genotyping microarrays and these were also separated into sub-cohorts. There was one exception (deCODE) to the single microarray rule, where the intersection between all arrays used exceeded 150,000 SNPs. In this cohort the genotype data from all arrays was merged since the correspondence between F_{ROH} for the individual arrays and F_{ROH} the intersection dataset was found to be very strong ($\beta_{merged,hap} = 0.98$, $r^2 = 0.98$; $\beta_{merged,omni} = 0.97$, r^2=0.97). Dividing studies using these criteria yielded 234 sub-cohorts. Details of phenotypes contributed by each cohort are available in Supplementary Data 4.

Ethical approval. Data from 119 independent genetic epidemiology studies were included. All subjects gave written informed consent for broad-ranging health and genetic research and all studies were approved by the relevant research ethics committees or boards. PubMed references are given for each study in Supplementary Data 2.

Genotyping. All samples were genotyped on high-density (minimum 250,000 markers), genome-wide SNP microarrays supplied by Illumina or Affymetrix. Genotyping arrays with highly variable genomic coverage (such as Exome chip, Metabochip, or Immunochip) were judged unsuitable for the ROH calling algorithm and were not permitted. Imputed genotypes were also not permitted; only called genotypes in PLINK binary format were accepted. Each study applied their own GWAS quality controls before additional checks were made in the common analysis pipeline: SNPs with >3% missingness or MAF <5% were removed, as were individuals with >3% missing data. Only autosomal genotypes were used for the analyses reported here. Additional, cohort-specific, genotyping information is available in Supplementary Data 2.

Phenotyping. In total, results are reported for 79 quantitative traits and 21 binary traits. These traits were chosen to represent different domains of health and reproductive success, with consideration given to presumed data availability. Many of these traits have been the subject of existing genome-wide association meta-analyses (GWAMA), and phenotype modelling, such as inclusion of relevant covariates, was copied from the relevant consortia (GIANT for anthropometry, EGG for birth weight, ICBP for blood pressures, MAGIC for glycaemic traits, CHARGE-Cognitive, -Inflammation & -Haemostasis working groups for cognitive function, CRP, fibrinogen, CHARGE-CKDgen for eGFR, CHARGE-ReproGen for ages at menarche and menopause, Blood Cell & HaemGen for haematology, GUGC for urate, RRgen, PRIMA, QRS & QT-IGC for electrocardiography, GLGC for classical lipids, CREAM for spherical equivalent refraction, Spirometa for lung function traits, and SSGAC for educational attainment and number of children ever born). Further information about individual phenotype modelling is available in Supplementary Note 1 and Supplementary Data 8.

ROH calling. Runs of homozygosity (ROH) of >1.5 Mb in length were identified using published methods[6,11]. In summary, SNPs with minor allele frequencies below 5% were removed, before continuous ROH SNPs were identified using PLINK with the following parameters: homozyg-window-snp 50; homozyg-snp 50; homozyg-kb 1500; homozyg-gap 1000; homozyg-density 50; homozyg-window-missing 5; homozyg-window-het 1. No linkage disequilibrium pruning was performed. These parameters have been previously shown to call ROH that correspond to autozygous segments in which all SNPs (including those not present on the chip) are homozygous-by-descent, not chance arrangements of independent homozygous SNPs, and inbreeding coefficient estimates calculated by this method (F_{ROH}) correlate well with pedigree-based estimates (F_{PED})[11]. Moreover, they have also been shown to be robust to array choice[6].

Calculating estimators of *F*. For each sample, two estimates of the inbreeding coefficient (F) were calculated, F_{ROH} and F_{SNP}. We also calculated three additional measures of homozygosity: $F_{ROH<5Mb}$, $F_{ROH>5Mb}$ and $F_{SNP\_outsideROH}$.

F_{ROH} is the fraction of each genome in ROH >1.5 Mb. For example, in a sample for which PLINK had identified n ROH of length l_i (in Mb), $i \in \{1..n\}$, then F_{ROH} was then calculated as

$$F_{ROH} = \frac{\sum_{i=1}^{n} l_i}{3Gb}, \quad (1)$$

where $F_{ROH<5Mb}$ and $F_{ROH>5Mb}$ are the genomic fractions in ROH of length >5 Mb, and in ROH of length <5 Mb (but >1.5 Mb), respectively, and the length of the autosomal genome is estimated at 3 gigabases (Gb). It follows from this definition that

$$F_{ROH} = F_{ROH>5Mb} + F_{ROH<5Mb}. \quad (2)$$

Single-point inbreeding coefficients can also be estimated from individual SNP homozygosity without any reference to a genetic map. For comparison with F_{ROH}, a method of moments estimate of inbreeding coefficient was calculated[34], referred to here as F_{SNP}, and implemented in PLINK by the command–het.

$$F_{SNP} = \frac{O(HOM) - E(HOM)}{N - E(HOM)}, \quad (3)$$

where O(HOM) is the observed number of homozygous SNPs, E(HOM) is the expected number of homozygous SNPs, i.e. $\sum_{i=1}^{N}(1 - 2p_iq_i)$, and N is the total number of non-missing genotyped SNPs.

F_{ROH} and F_{SNP} are strongly correlated, especially in cohorts with significant inbreeding, since both are estimates of F. To clarify the conditional effects of F_{ROH} and F_{SNP}, an additional measure of homozygosity,$F_{SNPoutsideROH}$, was calculated to describe the SNP homozygosity observed outside ROH.

$$F_{SNP_{outsideROH}} = \frac{O'(HOM) - E'(HOM)}{N' - E'(HOM)}, \quad (4)$$

where

$$O'(HOM) = O(HOM) - N_{SNP\_ROH}, \quad (5)$$

$$E'(HOM) = \left(\frac{N - N_{ROH}}{N}\right) * E(HOM), \quad (6)$$

$$N' = N - N_{ROH} \quad (7)$$

And N_{SNP_ROH} is the number of homozygous SNPs found in ROH. Note that:

$$F_{SNP\,outsideROH} \approx F_{SNP} - F_{ROH} \quad (8)$$

A further single point estimator of the inbreeding coefficient, described by Yang et al.[20] as $\widehat{F}^{III}$, is implemented in PLINK by the parameter –ibc (Fhat3) and was also calculated for all samples.

$$F_{GRM} = \widehat{F}^{III} = \frac{1}{N}\sum_{i=1}^{N} \frac{\left(x_i^2 - (1 + 2p_i)x_i + 2p_i^2\right)}{2p_i(1 - p_i)}, \quad (9)$$

where N is the number of SNPs, $_{pi}$ is the reference allele frequency of the ith SNP in the sample population and x_i is the number of copies of the reference allele.

Effect size estimates for quantitative traits. In each cohort of n samples, for each of the quantitative traits measured in that cohort, trait values were modelled by

$$y = \beta_{F_{ROH}} * \mathbf{F_{ROH}} + \mathbf{Xb} + \boldsymbol{\varepsilon}, \quad (10)$$

where $\mathbf{y}$ is a vector ($n \times 1$) of measured trait values, $\beta_{F_{ROH}}$ is the unknown scalar effect of F_{ROH} on the trait, $\mathbf{F_{ROH}}$ is a known vector ($n \times 1$) of individual F_{ROH}, $\mathbf{b}$ is a vector ($m \times 1$) of unknown fixed covariate effects (including a mean, μ), $\mathbf{X}$ in a known design matrix ($n \times m$) for the fixed effects, and $\boldsymbol{\varepsilon}$ is an unknown vector ($n \times 1$) of residuals.

The m fixed covariates included in each model were chosen with reference to the leading GWAMA consortium for that trait and are detailed in Supplementary Data 8. For all traits, these covariates included: age (and/or year of birth), sex, and at least the first 10 principal components of the genomic relatedness matrix (GRM). Where necessary, additional adjustments were made for study site, medications, and other relevant covariates (Supplementary Data 3).

For reasons of computational efficiency, it was decided to solve Eq. (10) in two steps. In the first step, the trait ($\mathbf{y}$) was regressed on all fixed covariates to obtain the maximum likelihood solution of the model:

$$\mathbf{y} = \mathbf{Xb} + \boldsymbol{\varepsilon}'. \quad (11)$$

All subsequent analyses were performed using the vector of trait residuals $\boldsymbol{\varepsilon}'$, which may be considered as the trait values corrected for all known covariates.

In cohorts with a high degree of relatedness, mixed-modelling was used to correct for family structure, although, because ROH are not narrow-sense heritable, this was considered less essential than in Genome-Wide Association Studies. Equation (11) becomes

$$\mathbf{y} = \mathbf{Xb} + \mathbf{u} + \boldsymbol{\varepsilon}', \quad (12)$$

where $\mathbf{u}$ is an unknown vector ($n \times 1$) of polygenic effects with multivariate normal distribution of mean 0 and covariance matrix $\sigma_g^2\mathbf{A}$, where $\mathbf{A}$ is the genomic relationship matrix (GRM). In these related cohorts, a GRM was calculated using PLINK v1.9 and Grammar+ residuals of Eq. (12) were estimated using GenABEL[35]. These Grammar+ residuals ($\boldsymbol{\varepsilon}'$) were used in subsequent analyses.

To estimate $\beta_{F_{\mathrm{ROH}}}$ for each trait, trait residuals were regressed on F_{ROH} to obtain the maximum likelihood (ML) solution of the model

$$\boldsymbol{\varepsilon}' = \mu + \beta_{F_{\mathrm{ROH}}} * \mathbf{F_{ROH}} + \boldsymbol{\varepsilon}. \quad (13a)$$

The sex-specific estimates of $\beta_{F_{\mathrm{ROH}}}$ (Supplementary Data 12) were obtained from Eq. (13) applied to the relevant sex.

For all traits, a corresponding estimates of $\beta_{F_{\mathrm{SNP}}}$ and $\beta_{F_{\mathrm{GRM}}}$ were obtained from the models

$$\boldsymbol{\varepsilon}' = \mu + \beta_{F_{\mathrm{SNP}}} * \mathbf{F_{SNP}} + \boldsymbol{\varepsilon}, \quad (13b)$$

$$\boldsymbol{\varepsilon}' = \mu + \beta_{F_{\mathrm{GRM}}} * \mathbf{F_{GRM}} + \boldsymbol{\varepsilon} \quad (14)$$

and the effects of different ROH lengths and of SNP homozygosity (Fig. 4b) were obtained from the model

$$\boldsymbol{\varepsilon}' = \mu + \left(\beta_1 * \mathbf{F_{SNP_{outsideROH}}}\right) + \left(\beta_2 * \mathbf{F_{ROH<5Mb}}\right) + \left(\beta_3 * \mathbf{F_{ROH>5Mb}}\right) + \boldsymbol{\varepsilon} \quad (15)$$

Effect size estimates for binary traits. Binary traits were analysed by two methods. The primary estimates of $\beta_{F_{\mathrm{ROH}}}$ (Fig. 3b and Supplementary Data 10) were obtained from full logistic models:

$$g(E[\mathbf{y}]) = \mathbf{Xb}, \quad (16)$$

where $g()$ is the link function (logit), and where F_{ROH} and all applicable covariates (Supplementary Datas 3, 8) were fitted simultaneously. Mixed modelling for family structure was not attempted in the logistic models since an accepted method was not apparent.

For all subsequent results, $\mathbf{y}$ was scaled by $1/\sigma_y^2$ and analysed by linear models, as for quantitative traits, including mixed-modelling where appropriate for family studies. This method of estimating binary traits with simple linear models gives asymptotically unbiased estimates of $\beta_{F_{\mathrm{ROH}}}$ and $\mathrm{se}(\beta_{F_{\mathrm{ROH}}})$ on the ln(Odds-Ratio) scale[36]. For all significant binary traits, a comparison of $\widehat{\beta}_{F_{\mathrm{ROH}}}$ from the full model with $\widehat{\beta}_{F_{\mathrm{ROH}}}$ from the linear model approximation is presented in Supplementary Fig. 8.

To give $\widehat{\beta}_{F_{\mathrm{ROH}}}$ a more tangible interpretation, effect estimates are frequently quoted in the text as $\beta_{0.0625}$, i.e. the estimated effect in the offspring of first cousins, where 6.25% of the genome is expected to be autozygous.

Religiosity and educational attainment as additional covariates. To assess the importance of potential social confounders, proxy measures of socio-economic status and religiosity were separately included in Eq. (13) as additional covariates. The modified effect estimates ($\widehat{\beta}'_{F_{\mathrm{ROH}}}$) were tested for significance (Supplementary Data 20) and compared to the uncorrected estimates ($\beta_{F_{\mathrm{ROH}}}$) (Supplementary Fig. 10a, b).

Since Educational Attainment (EA) was measured in many cohorts, this was chosen as the most suitable proxy for socio-economic status. However, since F_{ROH} is known to affect EA directly[6] any change in $\beta_{F_{\mathrm{ROH}}}$ when conditioning on EA cannot be assumed to be entirely due to environmental confounding.

The analysis of religiosity was only carried out in UKB, where a rough proxy was available. Although no direct questions about religious beliefs were included, participants were asked about their leisure activities. In response to the question *Which of the following do you attend once a week or more often? (You can select more than one)*, 15.6% of UKB participants selected *Religious Group* from one of the seven options offered. In the models described, religiosity was coded as 1 for those who selected *Religious Group* and 0 for those who did not. Although this is likely to be an imperfect measure of actual religious belief it is currently the best available in a large dataset.

Assortative mating. Humans are known to mate assortatively for a number of traits including height and cognition[37], and so we sought to investigate if this could influence our results, for example, by the trait extremes being more genetically similar and thus the offspring more homozygous. We see no evidence for an effect of assortative mating on autozygosity, however. Firstly, a polygenic risk score for height (see Supplementary Note 1), which explains 18.7% of the phenotypic variance in height, was not associated with F_{ROH} ($p = 0.77$; Supplementary Fig. 5). Secondly, linear relationships between traits and autozygosity extend out to very high F_{ROH} individuals (Supplementary Figs. 9a–f). Samples in the highest F_{ROH} group are offspring of genetically similar parents, very likely first or second degree relatives and, for example, the height of these samples is on average 3.4 cm [95% CI 2.5–4.3] shorter than the population mean. Assortative mating would suggest this height deficit has been inherited from genetically shorter parents, but this would require an implausibly strong relationship between short stature and a propensity to marry a very close relative. Thirdly, the sex-specific effects we observe could only be explained by assortative mating if the additive heritability of these traits also differed by gender.

Average trait values in groups of similar F_{ROH}. In each cohort individuals were allocated to one of ten groups of similar F_{ROH}. The bounds of these groups were the same for all cohorts, specifically {0, 0.002, 0.0041, 0.0067, 0.0108, 0.0186, 0.0333, 0.06, 0.10, 0.18 and 1.0}. Within each group the mean trait residual ($\boldsymbol{\varepsilon}'$) and mean F_{ROH} were calculated, along with their associated standard errors. Within each cohort the expectation of $\boldsymbol{\varepsilon}'$ is zero at the mean F_{ROH}, however as mean F_{ROH} varies between cohorts (Fig. 2, Supplementary Data 5) it was necessary to express $\boldsymbol{\varepsilon}'$ relative to a common F_{ROH} before meta-analysis. Hence, for this analysis only, the trait residuals ($\boldsymbol{\varepsilon}'$) were expressed relative to the $F_{\mathrm{ROH}} = 0$ intercept, i.e. by subtracting μ from Eq. (13).

Effect of F_{ROH} within adoptees. We compared $\beta_{F_{\mathrm{ROH\_ADOPTEE}}}$ to cross-cohort $\beta_{F_{\mathrm{ROH}}}$, not that from UKB alone, as we consider the latter to be a noisy estimate of the former; estimates in UKB are consistent with those from meta-analysis.

Effect of F_{ROH} within full-sibling families. In a subset of cohorts, with substantial numbers of related individuals, further analyses were performed to investigate the effect of F_{ROH} within full-sibling families. In each of these cohorts, all second-degree, or closer, relatives were identified using KING (parameters:–related–degree 2). Full-siblings were then selected as relative pairs with genomic kinship >0.175 and IBS0 >0.001. This definition includes monozygotic twins, who were intentionally considered as part of full-sibling families. Although monozygotic twins are expected to have identical F_{ROH}, they may not have identical trait values, and including additional trait measurements decreases the sampling error of the within-family variance estimate, hence increasing statistical power. Dizygotic twins were also included.

For each individual (j) with identified siblings, the values of F_{ROH} and trait residual ($\boldsymbol{\varepsilon}'$) were calculated relative to their family mean (and called $F_j^{\mathrm{ROH\_wSibs}}$ and $\varepsilon_j^{\mathrm{wSibs}}$, respectively), i.e. for individual j with n full-siblings S_k where $k \in \{1..n\}$

$$F_j^{\mathrm{ROHwSibs}} = F_j^{\mathrm{ROH}} - \frac{1}{(n+1)} \sum_{i \in \{j, S_k\}} F_i^{\mathrm{ROH}}, \quad (17)$$

$$\varepsilon_j^{\mathrm{wSibs}} = \varepsilon_j' - \frac{1}{(n+1)} \sum_{i \in \{j, S_k\}} \varepsilon'_i. \quad (18)$$

The effect of F_{ROH} within-full-siblings ($\beta_{F_{\mathrm{ROH}}\_\mathrm{wSibes}}$) was estimated by linear regression of $\boldsymbol{\varepsilon}^{\mathbf{wSibs}}$ on $\mathbf{F^{ROH\_wSibs}}$.

Importantly, the variation of F_{ROH} within full-siblings is entirely caused by differences in Mendelian segregation, and is therefore completely independent of all possible confounders. Hence, the effect estimates obtained by this method are estimates of the genetic effects of F_{ROH}, unbiased by any possible confounder. Since confounding by social factors is a major concern in this field, methods that can definitively exclude this possibility are of critical importance.

Between-cohort meta-analysis. As is typical in genome-wide association meta-analyses (GWAMA), genetic effects were estimated within single-ethnicity sub-cohorts, and meta-analysis of the within-cohort effect sizes was used to combine results[38]. This established method eliminates any potential confounding caused by between-cohort associations between F_{ROH} and traits.

Each cohort returned estimates and standard errors of: $\beta_{F_{\mathrm{ROH}}}$, $\beta_{F_{\mathrm{SNP}}}$, $\beta_{F_{\mathrm{ROH>Mb}}}$, $\beta_{F_{\mathrm{ROH<Mb}}}$, $\beta_{F\_\mathrm{outsideROH}}$, $\beta_{F_{\mathrm{ROH\_wSibs}}}$, as well as trait means ($\overline{\varepsilon'}$) and standard errors within each of 10 F_{ROH} bins. The between-cohort mean of each of these 16 estimates was then determined by fixed-effect, inverse-variance meta-analysis using the R package metafor[39]. Results shown in Figs. 3–5 are meta-analysed averages of the within-cohort effects.

The meta-analysis was also run for various subsets of cohorts, stratified by ancestry as defined in Supplementary Data 18. Meta-analysis estimates from these groupings are shown in Supplementary Fig. 1.

Median and 95% CI of a ratio. In the analyses of adoptees (Fig. 5c), siblings (Fig. 5d) and potential confounders (Supplementary Figs. 10a, b) we wished to compare the effect estimates ($\beta_{F_{\mathrm{ROH}}}$) from two different methods across a wide range of traits. The units of $\beta_{F_{\mathrm{ROH}}}$ differ by trait so, to allow comparison across all traits, the unitless ratio of effect size estimates was calculated (for example $\beta_{F_{\mathrm{ROH\_wSibs}}} : \beta_{F_{\mathrm{ROH}}}$). Figure 5c, d and Supplementary Figs. 10a, b show the medians and 95% CI of these ratios. These were determined empirically by bootstrap since, although formulae exist for the mean and standard error of a ratio[40], the assumption of normality is violated when $\beta_{F_{\mathrm{ROH}}}/\mathrm{se}(\beta_{F_{\mathrm{ROH}}})$ is not large.

Genetic correlations in UK Biobank. Genetic correlations were calculated using LD-Score Regression[41], implemented in LDSC v1.0.0 (https://github.com/bulik/ldsc). Summary statistics were parsed using default parameters in the LDSC

'munge_sumstats.py' script, extracting only variants present in the HapMap 3 reference panel.

Accuracy of F_{ROH} measures of inbreeding effects. A recent paper suggested that ROH may overestimate inbreeding effects by as much as 162%[42]; however, this could only be the case if F_{ROH} underestimates excess homozygosity at the causal loci by at least 162%. We do not believe this to be the case since the maximum F_{ROH} measured in many cohorts is around 0.25 (the expectation in the offspring off first-degree relatives), and the effect size estimates from these samples are consistent with the overall estimates (Fig. 5c, d and Supplementary Fig. 9a–f). We note that Yengo et al. applied the ROH calling parameters used here to imputed data. These parameters have been validated for called genotype data[6] but not, to our knowledge, for the higher SNP density and error rate of imputed data (see also Supplementary Note 4). The simple method for detecting ROH used here was well suited to our study, since it could be easily implemented on over one million samples, and most of the variation in F_{ROH} is caused by easily-identified long ROH.[43–45]

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The meta-analysed data which support these findings are available as Supplementary Data files. Cohort-level summary statistics underlying all figures and tables are available in a publicly accessible dataset (https://doi.org/10.6084/m9.figshare.9731087). In the majority of cases we do not have consent to share individual-level data, although for UK Biobank this is available on request from https://www.ukbiobank.ac.uk/.

Received: 1 November 2018; Accepted: 30 August 2019;
Published online: 31 October 2019

References

1. Charlesworth, D. & Willis, J. H. The genetics of inbreeding depression. *Nat. Rev. Genet.***10**, 783–796 (2009).
2. Bittles, A. H. & Black, M. L. Consanguinity, human evolution, and complex diseases. *Proc. Natl Acad. Sci.* **107**, 1779–1786 (2010).
3. Nakatsuka, N. et al. The promise of discovering population-specific disease-associated genes in South Asia. *Nat. Genet.***49**, 1403–1407 (2017).
4. Narasimhan, V. M. et al. Health and population effects of rare gene knockouts in adult humans with related parents. *Science* **352**, 474–477 (2016).
5. Saleheen, D. et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature* **544**, 235–239 (2017).
6. Joshi, P. K. et al. Directional dominance on stature and cognition in diverse human populations. *Nature* **523**, 459–462 (2015).
7. Kirin, M. et al. Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **5**, e13996 (2010).
8. Pemberton, T. J. et al. Genomic patterns of homozygosity in worldwide human populations. *Am. J. Hum. Genet.***91**, 275–292 (2012).
9. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.* **19**, 220–234 (2018).
10. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
11. McQuillan, R. et al. Runs of homozygosity in European populations. *Am. J. Hum Genet.* **83**, 359–372 (2008).
12. Purcell, S. & Chang, C. *PLINK 1.9*. www.cog-genomics.org/plink/1.9/.
13. McQuillan, R. et al. Evidence of inbreeding depression on human height. *PLoS Genet.* **8**, e1002655 (2012).
14. Abdellaoui, A. et al. Association between autozygosity and major depression: stratification due to religious assortment. *Behav. Genet.* **43**, 455–467 (2013).
15. Bashi, J. Effects of inbreeding on cognitive performance. *Nature* **266**, 440–442 (1977).
16. Broman, K. W. & Weber, J. L. Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *Am. J. Hum. Genet.* **65**, 1493–1500 (1999).
17. Gibson, J., Morton, N. E. & Collins, A. Extended tracts of homozygosity in outbred human populations. *Hum. Mol.Genet.* **15**, 789–795 (2006).
18. Lencz, T. et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc.Natl. Acad. Sci. USA.* **104**, 19942–19947 (2007).
19. Keller, M. C., Visscher, P. M. & Goddard, M. E. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* **189**, 237–249 (2011).
20. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
21. Coulson, T., Albon, S., Jon, S. & Pemberton, J. Microsatellite loci reveal sex-dependent responses to inbreeding and outbreeding in Red Deer calves. *Evolution* **53**, 1951–1960 (1999).
22. Slate, J. & Pemberton, J. M. Comparing molecular measures for detecting inbreeding depression. *J.Evolutionary Biology* **15**, 20–31 (2002).
23. Campbell, H. et al. Effects of genome-wide heterozygosity on a range of biomedically relevant human quantitative traits. *Hum. Mol. Genet.* **16**, 233–241 (2007).
24. Abdellaoui, A. et al. Educational attainment influences levels of homozygosity through migration and assortative mating. *PLoS ONE* **10**, e0118935 (2015).
25. Helgason, A., Pálsson, S., Guobjartsson, D. F., Kristjánsson, P. & Stefánsson, K. An association between the kinship and fertility of human couples. *Science* **319**, 813–816 (2008).
26. Bittles, A. H., Grant, J. C., Sullivan, S. G. & Hussain, R. Does inbreeding lead to decreased human fertility? *Ann. Hum. Biol.***29**, 111–130 (2002).
27. Alvarez, G., Ceballos, F. C. & Berra, T. M. Darwin was right: inbreeding depression on male fertility in the Darwin family. *Biol. J.Linn.Soc.* **114**, 474–483 (2015).
28. Ober, C., Hyslop, T. & Hauck, W. W. Inbreeding effects on fertility in humans: evidence for reproductive compensation. *Am. J. Hum. Genet.* **64**, 225–231 (1999).
29. Robert, A., Toupance, B., Tremblay, M. & Heyer, E. Impact of inbreeding on fertility in a pre-industrial population. *Eur. J. Hum. Genet.* **17**, 673–681 (2009).
30. Day, F. R. et al. Physical and neuro-behavioural determinants of reproductive onset and success Europe PMC Funders Group. *Nat. Genet.* **48**, 617–623 (2016).
31. Huisman, J., Kruuk, L. E. B., Ellis, P. A., Clutton-Brock, T. & Pemberton, J. M. Inbreeding depression across the lifespan in a wild mammal population. *Proc. Natl. Acad. Sci.* **113**, 3585–3590 (2016).
32. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
33. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
34. Wright, S. Genetics of populations. *Encyclopedia Britannica* Vol. 10, 111-A-D-112 (1948).
35. Belonogova, N. M., Svishcheva, G. R., van Duijn, C. M., Aulchenko, Y. S. & Axenovich, T. I. Region-based association analysis of human quantitative traits in related individuals. *PLoS ONE* **8**, e65395 (2013).
36. Cook, J. P., Mahajan, A. & Morris, A. P. Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur. J. Hum. Genet.* **25**, 240–245 (2017).
37. Buss, D. *The Evolution of Desire: Strategies of Human Mating*. Vol. 15 (Basic Books, 2003).
38. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
39. Viechtbauer, W. Conducting meta-analyses in R with the metafor Package. *J. Stat. Softw.* **36**, 1–48 (2010).
40. Fieller, E. C. Some problems in interval estimation. *J. R. Stat. Soc. Ser. B* **16**, 175–185 (1954).
41. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
42. Yengo, L. et al. Detection and quantification of inbreeding depression for complex traits from SNP data. *Proc. Natl. Acad. Sci.* **114**, 8602–8607 (2017).
43. Kardos, M., Nietlisbach, P. & Hedrick, P. W. How should we compare different genomic estimates of the strength of inbreeding depression? *Proc. Natl. Acad. Sci. USA* **115**, E2492–E2493 (2018).
44. Gazal, S. et al. Inbreeding coefficient estimation with dense SNP data: comparison of strategies and application to HapMap III. *Hum. Hered.* **77**, 49–62 (2014).
45. Johnson, E. C., Evans, L. M. & Keller, M. C. Relationships between estimated autozygosity and complex traits in the UK Biobank. *PLoS Genet.* **14**, e1007556 (2018).
46. Keller, L. F. & Waller, D. M. Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17**, 230–241 (2002).

Acknowledgements

This paper is the work of the ROHgen consortium. We thank the Sigma T2D Consortium, whose members are detailed in Supplementary Note 3. We thank the UK Biobank Resource, approved under application 19655; we acknowledge funding from the UK Medical Research Council Human Genetics Unit and MRC Doctoral Training Programme in Precision Medicine. We also thank Neil Robertson, Wellcome Trust Centre for Human Genetics, Oxford, for use of his author details management software, Authorial. Finally, we thank all the participants, researchers and funders of ROHgen cohorts. Cohort-specific acknowledgements are in Supplementary Data 2; personal acknowledgements and disclosures are in Supplementary Note 2. We thank Rachel Edwards for administrative assistance.

Author Contributions

Directed consortium: T.E., P.K.J. and J.F.W. Central analyst: D.W.C., Y.O., N. Pirastu and P.K.J. Cohort PI: C.A.A.-S., M.L.A., T.A., F.A., E.B., G.J.d.B., E.P.B., S. Chanock, Y.-D.I.C., Z.C., R.M.v.D., M.S.D., N.D., P.E., B.I.F., C.A. Haiman, D.A.v.H., I.H., M.A.I., P.J., T. Kessler, K.-T.K., D.P.V.d.K., W.-P.K., J. Kuusisto, C.L., D.A.L., L.L., T.A.M., Y. Murakami, K.K.O., L.O., O.P., N. Poulter, P.P.P., L.Q.-M., K.R., D.C.R., S.S.R., P.M.R., L.J.S., P.J.S., W.H.-H.S., A. Stanton, J.M.S., L. Straker, T.T.-L., J.H.V., V.V., Y.X.W., N.J.W., C.S. Y., J.-M.Y., F.W.A., S.J.L.B., D.M.B., D.A.B., L.H.v.d.B., S.I.B., D.B., C. Bouchard, M.J.C., J.C.C., G.R.C., C.-Y.C., M. Ciullo, M. Cornelis, D. Cusi, G.D.-S., I.D., C.M.v.D., D.E., J. Erdmann, J.G.E., E.E., M.K.E., B.F., M.F., A.F., Y.F., P. Gasparini, C. Gieger, C. Gonzalez, S.F.A.G., L.R.G., L.G., V.G., U.G., A. Hamsten, P.v.d.H., C.-K.H., H. Hochner, S.C.H., V.W.V.J., Å.J., J.B.J., J.W.J., J.J., J. Kaprio, S.L.K., F.K., M. Kumari, M. Laakso, M. Laudes, W.L., N.G.M., W. März, G. Matullo, M.I.M., T.R.M., A. Metspalu, B.F.M., K.L. M., G.W.M., D.M.-K., P.B.M., K.E.N., C.O., A.J.O., C. Palmer, G.G.P., E.P., C.E.P., L.P., M. Pirastu, D.J.P., D.P., B.M.P., C.R., J.I.R., I.R., D.K.S., R.S., H. Schunkert, A.R.S., N. Small, E.-S.T., N.J.T., D.T., T.T., P.V., D.R.W., T.-Y.W., J.W., A.B.Z., M. Perola, P.K.M., A.G.U., J.S.K., D.I.C., R.J.F.L., N.F., C.H., J.R.B.P., T.E., K. Stefansson, M. Kubo, J.F.W. Cohort analyst: D.W.C., Y.O., K.H.S.M., D.M., I.G., H.M., K.L., J.H.Z., P.D., R.R., C. Schurmann, X.G., F. Giulianini, W. Zhang, C.M.-G., R.K., Y.B., T.M.B., C. Baumbach, G.B., M.J.B., M. Brumat, J.-F.C., D.L.C., D.A.E., C.F., H.G., M. Germain, S.D.G., H.G.d.H., S.E.H., E.H., A.H.-C., C.I., I.E.J., Y.J., T. Kacprowski, T. Karlsson, M.E.K., S.A.L., R.L.-G., A. Mahajan, W. Meng, M.E.M., P.J.v.d.M., M.Munz, T.N., T.P., G. Prasad, R.B.P., T.D.S.P., F. Rizzi, E.S., B.R.S., D.S., L. Skotte, A.V.S., A.v.d.S., C.N.S., R.J.S., S.M.T., S.T., C.T., N.V., C.V., L.W., H.R.W., R.E.W., L.R.Y., J. Yao, N.A.Y., W. Zhao, A.A.A., S.A., M. Akiyama, M. Alver, G. Chen, M. Cocca, M.P.C., G. Cugliari, F.R.D., G.E.D., G.G., A.G., M. Gögele, M. Graff, E.G.-H., A. Halevy, D.A.v.H., J.H., Y.K., M. Kanai, N.D.K., M. Loh, S.L., Y.L., J'a. L., N.M., X.W.M., M. Mezzavilla, A. Moore, H.M.-M., M.A.N., C.A.R., A.R., D.R., M.S.-L., W.R.S., B.S., J.v.S., S.S., S.R.S., T.T.S., A. Tillander, E.V., L. Zeng, N.A., L. Benjamin, L.F.B., J.P.B., J.A.B., S. Carmi, G.R.C., M. Cornelis, D. Cusi, R.D., D.E., J.D.F., M.F., S.F., P. Goyette, S.C.H., Å.J., S.W.v.d.L., J. Lahti, R.A.L., S.E.M., K.E.N., J.R.O'c., E.P., L.P., J.D. R., F. Rivadeneira, C.R., R.A.S., X.S., J.A.S., C.A.W., J. Yang, L.Y., D.I.C., N.F., R.G.W., J.R. B.P., T.E., A. Helgason. Provided data: Y.O., C. Schurmann, W. Zhang, C.M.-G., Y.B., A.M.D., K.R.v.E., C.F., H.G., M.E.K., K. Matsuda, R.B.P., F. Rizzi, E.S., M.C.S., A.V.S., S.M. T., N.V., A.A.A., C.A.A.-S., M.L.A., M.A.A., A.R.B., E.B., J.B.B., G.J.d.B., E.P.B., L. Broer, H.C., S. Chanock, M.-L.C., G. Chen, Y.-D.I.C., Y.-F.C., J.C., M.S.D., K.D., M.D., A.P.D., N.D., S.S.E., J. Elliott, R.E., J.F.F., K.F., B.I.F., M. Gögele, M.O.G., S.G., D.F.G., K.G., B.G., Y.G., S.P.H., C.A. Haiman, T.B.H., M. Hedayati, M. Hirata, I.H., C.A. Hsiung, Y.-J.H., M.A.I., A.J., P.J., Y.K., C.C.K., W.-P.K., I.K., B.K.K., J. Kuusisto, L.J.L., D.A.L., I.-T.L., W.-J. L., M.M.L., J. Liu, S.J.L., R.M., A.W.M., P.M., G. Másson, C.M., T. Meitinger, L.M., I.Y.M., Y. Momozawa, T.A.M., A.C.M., T. Muka, A.D.M., R.d.M., J.C.M., M.A.N., M.N., M.J.N., I.M.N., L.O., S.P., G. Pálsson, J.S.P., C. Pattaro, A.P., O.P., N. Poulter, L.Q.-M., K.R., S.R., D.C.R., W.v.R., F.J.A.v.R., C. Sabanayagam, C.F.S., V.S., K. Sandow, B.S.-K., P.J.S., W.H.-H.S., Y. Shi, S.R.S., J.K.S., J.R.S., B.H.S., A. Stanton, L. Stefansdottir, L. Straker, P.S., G.S., M.A.S., A.M.T., K.D.T., N.T., Y.-C.T., G.T., U.T., R.P.T., T.T.-L., I.T., S.V., J.H.V., V.V., U.V., E.V., S.M.W., M.W., G.S.W., S.W., C.S.Y., J.-M.Y., L. Zhang, J.Z., S.J.L.B., D.M.B., S.I.B., A.C., M.J.C., D. Cesarini, J.C.C., G.R.C., M. Cornelis, D. Cusi, G.D.-S., I.D., R.D., D.E., J. Erdmann, J.G.E., E.E., M.K.E., B.F., M.F., Y.F., P. Gasparini, C. Gonzalez, S.F.A.G., L.G., V.G., C.-K.H., A.A.H., H. Hochner, H. Huikuri, S.C.H., V.W.V.J., P.L.D.J., M.J., J.W. J., J. Kaprio, S.L.K., M. Laakso, S.W.v.d.L., J.Lahti, W.L., N.G.M., G. Matullo, B.F.M., K.L. M., G.W.M., P.B.M., D.R.N., A.J.O., W.P., C. Palmer, C.E.P., L.P., P.A.P., T.J.P., D.J.P., D.P., B.M.P., J.D.R., F. Rivadeneira, C.R., J.I.R., I.R., D.K.S., N. Sattar, H. Schunkert, A. Teumer, N.J.T., T.T., D.R.W., J.B.W., C.W., J.W., A.B.Z., M. Perola, P.K.M., A.G.U., J.S.K., D.I.C., L.F., C.S.H., C.H., R.G.W., T.E. Contributed to manuscript: D.W.C., C. Schurmann, S.E.H., S.M.T., E.P.B., J.C., A.P.D., X.G., S.P.H., D.A.v.H., P.J., W.-P.K., D.A.L., S.J.L., A.P., J.M.S., A.M.T., J.-M.Y., J.Z., I.D., R.D., Y.F., J.W.J., C.R., J.I.R., M.Perola, R.J.F.L., P.K.J., J.F.W. Wrote manuscript: D.W.C. and J.F.W.

Competing interests

M.L.A. is an employee of Genentech, a member of The Roche Group. D.A.L. has received support from several national and international government and charity funders, as well as Roche Diagnostics and Medtronic for work unrelated to this publication. M.I.M.: The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. He has served on advisory panels for Pfizer, NovoNordisk, Zoe Global; has received honoraria from Merck, Pfizer, Novo-Nordisk and Eli Lilly; has stock options in Zoe Global; has received research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, Novo-Nordisk, Pfizer, Roche, Sanofi Aventis, Servier & Takeda. As of June 2019, M.Mc.C. is an employee of Genentech, and holds stock in Roche. T. Muka is now working as medical specialist at Novo Nordisk. O.P. is owner of Gen-info Ltd. Gen-info Ltd provided support in the form of salaries and financial gains for author O.P., but did not have any additional role in selection of the journal or preparation of this manuscript. N. Poulter received financial support from several pharmaceutical companies which manufacture either blood pressure lowering or lipid lowering agents, or both, and consultancy fees. V.S. has participated in a congress trip sponsored By Novo Nordisk. P.J.S. has received research awards from Pfizer Inc. M.J.C. is Chief Scientist for Genomics England, a UK government company. B.M.P. serves on the DSMB of a clinical trial funded by Zoll LifeCor and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. A.R.S. is an employee of Regeneron Pharmaceutical Inc. The remaining authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41467-019-12283-6.

Correspondence and requests for materials should be addressed to J.F.W.

Peer review information *Nature Communications* thanks Jun Li, Matthew Robinson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

David W Clark[1], Yukinori Okada[2,3,4], Kristjan H S Moore[5], Dan Mason[6], Nicola Pirastu[1], Ilaria Gandin[7,8], Hannele Mattsson[9,10], Catriona L K Barnes[1], Kuang Lin[11], Jing Hua Zhao[12,13], Patrick Deelen[14], Rebecca Rohde[15], Claudia Schurmann[16], Xiuqing Guo[17], Franco Giulianini[18], Weihua Zhang[19,20], Carolina Medina-Gomez[21,22,23], Robert Karlsson[24], Yanchun Bao[25], Traci M Bartz[26], Clemens Baumbach[27], Ginevra Biino[28], Matthew J Bixley[29], Marco Brumat[8], Jin-Fang Chai[30], Tanguy Corre[31,32,33], Diana L Cousminer[34,35], Annelot M Dekker[36], David A Eccles[37,38], Kristel R van Eijk[36], Christian Fuchsberger[39], He Gao[19,40], Marine Germain[41,42], Scott D Gordon[43], Hugoline G de Haan[44], Sarah E Harris[45,46], Edith Hofer[47,48], Alicia Huerta-Chagoya[49], Catherine Igartua[50], Iris E Jansen[51,52], Yucheng Jia[17,327], Tim Kacprowski[53,54], Torgny Karlsson[55], Marcus E Kleber[56], Shengchao Alfred Li[57], Ruifang Li-Gao[44], Anubha Mahajan[58], Koichi Matsuda[59], Karina Meidtner[60,61], Weihua Meng[62],

May E Montasser[63,64], Peter J van der Most[65], Matthias Munz[66,67,68,69], Teresa Nutile[70], Teemu Palviainen[71], Gauri Prasad[72], Rashmi B Prasad[73], Tallapragada Divya Sri Priyanka[74], Federica Rizzi[75,76], Erika Salvi[76,77], Bishwa R Sapkota[78], Daniel Shriner[79], Line Skotte[80], Melissa C Smart[25], Albert Vernon Smith[81,82], Ashley van der Spek[22], Cassandra N Spracklen[83], Rona J Strawbridge[84,85], Salman M Tajuddin[86], Stella Trompet[87,88], Constance Turman[89,90], Niek Verweij[91], Clara Viberti[92], Lihua Wang[93], Helen R Warren[94,95], Robyn E Wootton[96,97], Lisa R Yanek[98], Jie Yao[17], Noha A Yousri[99,100], Wei Zhao[101], Adebowale A Adeyemo[79], Saima Afaq[19], Carlos Alberto Aguilar-Salinas[102,103], Masato Akiyama[3,104], Matthew L Albert[105,106,107,108], Matthew A Allison[109], Maris Alver[110], Tin Aung[111,112,113], Fereidoun Azizi[114], Amy R Bentley[79], Heiner Boeing[115], Eric Boerwinkle[116], Judith B Borja[117], Gert J de Borst[118], Erwin P Bottinger[16,119], Linda Broer[21], Harry Campbell[1], Stephen Chanock[120], Miao-Li Chee[111], Guanjie Chen[79], Yii-Der I Chen[17], Zhengming Chen[11], Yen-Feng Chiu[121], Massimiliano Cocca[122], Francis S Collins[123], Maria Pina Concas[122], Janie Corley[45,124], Giovanni Cugliari[92], Rob M van Dam[30,125,126], Anna Damulina[47], Maryam S Daneshpour[127], Felix R Day[12], Graciela E Delgado[56], Klodian Dhana[22,126,128], Alexander S F Doney[129], Marcus Dörr[130,131], Ayo P Doumatey[79], Nduna Dzimiri[132], S Sunna Ebenesersdóttir[5,133], Joshua Elliott[19], Paul Elliott[19,40,134,135,136], Ralf Ewert[130], Janine F Felix[22,23,137], Krista Fischer[110], Barry I Freedman[138], Giorgia Girotto[8,139], Anuj Goel[58,140], Martin Gögele[39], Mark O Goodarzi[141], Mariaelisa Graff[15], Einat Granot-Hershkovitz[142], Francine Grodstein[89], Simonetta Guarrera[92], Daniel F Gudbjartsson[5,143], Kamran Guity[127], Bjarni Gunnarsson[5], Yu Guo[144], Saskia P Hagenaars[45,124,145], Christopher A Haiman[146], Avner Halevy[142], Tamara B Harris[86], Mehdi Hedayati[127], David A van Heel[147], Makoto Hirata[148], Imo Höfer[149], Chao Agnes Hsiung[121], Jinyan Huang[150], Yi-Jen Hung[151,152], M Arfan Ikram[22], Anuradha Jagadeesan[5,133], Pekka Jousilahti[153], Yoichiro Kamatani[3,154], Masahiro Kanai[2,3,155], Nicola D Kerrison[12], Thorsten Kessler[156], Kay-Tee Khaw[157], Chiea Chuen Khor[111,158], Dominique P V de Kleijn[118], Woon-Puay Koh[30,159], Ivana Kolcic[160], Peter Kraft[126], Bernhard K Krämer[56], Zoltán Kutalik[32,33], Johanna Kuusisto[161,162], Claudia Langenberg[12], Lenore J Launer[86], Deborah A Lawlor[96,163,164], I-Te Lee[165,166,167], Wen-Jane Lee[168], Markus M Lerch[169], Liming Li[170], Jianjun Liu[125,158], Marie Loh[19,171,172], Stephanie J London[173], Stephanie Loomis[174], Yingchang Lu[16], Jian'an Luan[12], Reedik Mägi[110], Ani W Manichaikul[175], Paolo Manunta[176], Gísli Másson[5], Nana Matoba[3], Xue W Mei[11], Christa Meisinger[177], Thomas Meitinger[178,179,180], Massimo Mezzavilla[139], Lili Milani[181], Iona Y Millwood[11], Yukihide Momozawa[182], Amy Moore[120], Pierre-Emmanuel Morange[183,184], Hortensia Moreno-Macías[185], Trevor A Mori[186], Alanna C Morrison[187], Taulant Muka[22,188], Yoshinori Murakami[189], Alison D Murray[190], Renée de Mutsert[44], Josyf C Mychaleckyj[175], Mike A Nalls[191,192], Matthias Nauck[131,193], Matt J Neville[194,195], Ilja M Nolte[65], Ken K Ong[12,196], Lorena Orozco[197], Sandosh Padmanabhan[198], Gunnar Pálsson[5], James S Pankow[199], Cristian Pattaro[39], Alison Pattie[124], Ozren Polasek[160,200], Neil Poulter[201,202], Peter P Pramstaller[39], Lluis Quintana-Murci[203,204,205], Katri Räikkönen[206], Sarju Ralhan[207], Dabeeru C Rao[208], Wouter van Rheenen[36], Stephen S Rich[175], Paul M Ridker[18,209], Cornelius A Rietveld[210,211], Antonietta Robino[122], Frank J A van Rooij[22], Daniela Ruggiero[70,212], Yasaman Saba[213], Charumathi Sabanayagam[111,112,113], Maria Sabater-Lleal[85,214], Cinzia Felicita Sala[215], Veikko Salomaa[216], Kevin Sandow[17], Helena Schmidt[213], Laura J Scott[217], William R Scott[19], Bahareh Sedaghati-Khayat[127], Bengt Sennblad[85,218], Jessica van Setten[219], Peter J Sever[201], Wayne H-H Sheu[152,165,220,221], Yuan Shi[111], Smeeta Shrestha[74,222], Sharvari Rahul Shukla[223,224], Jon K Sigurdsson[5], Timo Tonis Sikka[110], Jai Rup Singh[225], Blair H Smith[226], Alena Stančáková[161],

Alice Stanton[227], John M Starr[45,228,327], Lilja Stefansdottir[5], Leon Straker[229], Patrick Sulem[5], Gardar Sveinbjornsson[5], Morris A Swertz[14], Adele M Taylor[124], Kent D Taylor[17], Natalie Terzikhan[22,230], Yih-Chung Tham[111,112], Gudmar Thorleifsson[5], Unnur Thorsteinsdottir[5,82], Annika Tillander[24], Russell P Tracy[231], Teresa Tusié-Luna[232,233], Ioanna Tzoulaki[19,40,234], Simona Vaccargiu[235], Jagadish Vangipurapu[161], Jan H Veldink[36], Veronique Vitart[236], Uwe Völker[53,131], Eero Vuoksimaa[237], Salma M Wakil[132], Melanie Waldenberger[27], Gurpreet S Wander[238], Ya Xing Wang[239], Nicholas J Wareham[12], Sarah Wild[240], Chittaranjan S Yajnik[241], Jian-Min Yuan[242], Lingyao Zeng[156], Liang Zhang[111], Jie Zhou[79], Najaf Amin[22], Folkert W Asselbergs[243,244,245,246], Stephan J L Bakker[247], Diane M Becker[98], Benjamin Lehne[19], David A Bennett[248,249], Leonard H van den Berg[36], Sonja I Berndt[120], Dwaipayan Bharadwaj[250], Lawrence F Bielak[101], Murielle Bochud[32], Mike Boehnke[217], Claude Bouchard[251], Jonathan P Bradfield[252,253], Jennifer A Brody[254], Archie Campbell[46], Shai Carmi[142], Mark J Caulfield[94,95], David Cesarini[255,256], John C Chambers[19,20,40,257,258], Giriraj Ratan Chandak[74], Ching-Yu Cheng[111,112,113], Marina Ciullo[70,212], Marilyn Cornelis[259], Daniele Cusi[76,260,261], George Davey Smith[96,164], Ian J Deary[45,124], Rajkumar Dorajoo[158], Cornelia M van Duijn[11,22], David Ellinghaus[262], Jeanette Erdmann[66], Johan G Eriksson[263,264,265,266,267], Evangelos Evangelou[19,234], Michele K Evans[86], Jessica D Faul[268], Bjarke Feenstra[80], Mary Feitosa[93], Sylvain Foisy[269], Andre Franke[262], Yechiel Friedlander[142], Paolo Gasparini[8,139], Christian Gieger[27,61], Clicerio Gonzalez[270], Philippe Goyette[269], Struan F A Grant[35,252,271], Lyn R Griffiths[37], Leif Groop[71,73], Vilmundur Gudnason[81,82], Ulf Gyllensten[55], Hakon Hakonarson[252,271], Anders Hamsten[272], Pim van der Harst[91], Chew-Kiat Heng[273,274], Andrew A Hicks[39], Hagit Hochner[142], Heikki Huikuri[275], Steven C Hunt[99,276], Vincent W V Jaddoe[22,23,137], Philip L De Jager[277,278], Magnus Johannesson[279], Åsa Johansson[55], Jost B Jonas[239,280], J Wouter Jukema[87], Juhani Junttila[275], Jaakko Kaprio[71,281], Sharon L.R. Kardia[101], Fredrik Karpe[194,282], Meena Kumari[25], Markku Laakso[161,162], Sander W van der Laan[283], Jari Lahti[206,284], Matthias Laudes[285], Rodney A Lea[37], Wolfgang Lieb[286], Thomas Lumley[287], Nicholas G Martin[43], Winfried März[56,288,289], Giuseppe Matullo[92], Mark I McCarthy[58,194,282], Sarah E Medland[43], Tony R Merriman[29], Andres Metspalu[110], Brian F Meyer[290], Karen L Mohlke[83], Grant W Montgomery[43,291], Dennis Mook-Kanamori[44,292], Patricia B Munroe[94,95], Kari E North[15], Dale R Nyholt[43,293], Jeffery R O'connell[63,64], Carole Ober[50], Albertine J Oldehinkel[294], Walter Palmas[295], Colin Palmer[296], Gerard G Pasterkamp[149], Etienne Patin[203,204,205], Craig E Pennell[297,298], Louis Perusse[299,300], Patricia A Peyser[101], Mario Pirastu[301], Tinca J.C. Polderman[51], David J Porteous[45,46], Danielle Posthuma[51,302], Bruce M Psaty[303,304], John D Rioux[269,305], Fernando Rivadeneira[21,22,23], Charles Rotimi[79], Jerome I Rotter[17], Igor Rudan[1], Hester M Den Ruijter[306], Dharambir K Sanghera[78,307], Naveed Sattar[198], Reinhold Schmidt[47], Matthias B Schulze[60,61,308], Heribert Schunkert[156,309], Robert A Scott[12], Alan R Shuldiner[63,64,310], Xueling Sim[30], Neil Small[311], Jennifer A Smith[101,268], Nona Sotoodehnia[312], E-Shyong Tai[30,125,313], Alexander Teumer[131,314], Nicholas J Timpson[96,315,316], Daniela Toniolo[215], David-Alexandre Tregouet[41], Tiinamaija Tuomi[10,317,318,319], Peter Vollenweider[320], Carol A Wang[297,298], David R Weir[268], John B Whitfield[43], Cisca Wijmenga[14], Tien-Yin Wong[111,112], John Wright[6], Jingyun Yang[248,249], Lei Yu[248,249], Babette S Zemel[271,321], Alan B Zonderman[86], Markus Perola[322], Patrik K.E. Magnusson[24], André G Uitterlinden[21,22,23], Jaspal S Kooner[20,40,258,323], Daniel I Chasman[18,209], Ruth J.F. Loos[16,324], Nora Franceschini[15], Lude Franke[14], Chris S Haley[236,325], Caroline Hayward[236], Robin G Walters[11],

John R.B. Perry[12], Tõnu Esko[110,326], Agnar Helgason[5,133], Kari Stefansson[5,82], Peter K Joshi[1], Michiaki Kubo[182] & James F Wilson[1,236*]

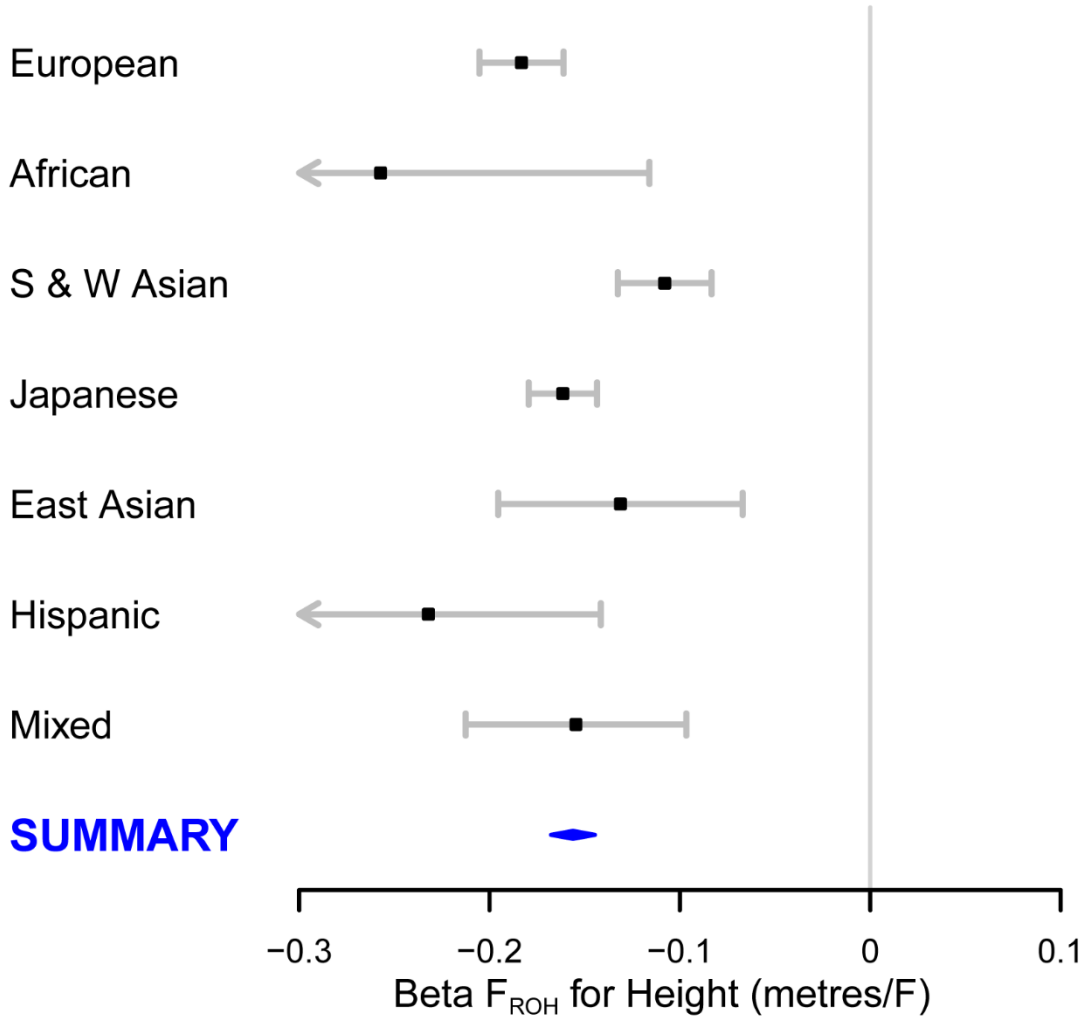
[1]Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh EH8 9AG, Scotland. [2]Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Osaka 565-0871, Japan. [3]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan. [4]Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Osaka 565-0871, Japan. [5]deCODE genetics/Amgen Inc., Reykjavik 101, Iceland. [6]Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Trust, Bradford BD96RJ, UK. [7]Research Unit, Area Science Park, Trieste 34149, Italy. [8]Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy. [9]Unit of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland. [10]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [11]Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK. [12]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge CB2 0QQ, UK. [13]Cardiovascular Epidemiology Unit, Department of Public health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. [14]Department of Genetics, University Medical Centre Groningen, University of Groningen, Groningen, the Netherlands, Groningen, Groningen 9700 RB, The Netherlands. [15]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27514, USA. [16]The Charles Bronfman Institute for Personalized Medicine, Ichan School of Medicine at Mount Sinai, New York, NY 10029, USA. [17]Division of Genomic Outcomes, Department of Pediatrics, The Institute for Translational Genomics and Population Sciences, LABioMed at Harbor-UCLA Medical Center, Torrance, California 90502, USA. [18]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02215, USA. [19]Department of Epidemiology and Biostatistics, Imperial College London, London W2 1PG, UK. [20]Department of Cardiology, Ealing Hospital, Middlesex, Middlesex UB1 3HW, UK. [21]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam 3015 CN, Netherlands. [22]Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3015 CN, Netherlands. [23]The Generation R Study Group, Erasmus University Medical Center, Rotterdam 3015 CN, The Netherlands. [24]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm 17177, Sweden. [25]Institute for Social and Economic Research, University of Essex, Colchester CO4 3SQ, UK. [26]Cardiovascular Health Research Unit, Departments of Biostatistics and Medicine, University of Washington, Seattle, WA 98101, USA. [27]Research Unit of Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg 85764, Germany. [28]Institute of Molecular Genetics, National Research Council of Italy, Pavia 27100, Italy. [29]Department of Biochemistry, University of Otago, Dunedin 9054, New Zealand. [30]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore 117549, Singapore. [31]Department of Computational Biology, University of Lausanne, Lausanne 1011, Switzerland. [32]Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland. [33]Swiss Institute of Bioinformatics, Lausanne 1015, Switzerland. [34]Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA. [35]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [36]Department of Neurology, Brain Centre Rudolf Magnus, University Medical Centre Utrecht, Utrecht University, Utrecht 3584 CX, The Netherlands. [37]Genomics Research Centre, School of Biomedical Sciences, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland 4059, Australia. [38]Malaghan Institute of Medical Research, Wellington 6242, New Zealand. [39]Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Bolzano 39100, Italy. [40]MRC-PHE Centre for Environment and Health, Imperial College London, London W2 1PG, UK. [41]INSERM UMR_S 1166, Sorbonne Universités, Paris 75013, France. [42]ICAN Institute for Cardiometabolism and Nutrition, Paris 75013, France. [43]QIMR Berghofer Institute of Medical Research, Brisbane, Australia. [44]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands. [45]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh EH8 9JZ, UK. [46]Centre for Genomic & Experimental Medicine, Institute of Genetics & Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK. [47]Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz 8036, Austria. [48]Institute of Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz 8036, Austria. [49]CONACyT, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico 03940, México. [50]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. [51]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam 1081 HV, The Netherlands. [52]Alzheimer Center Department of Neurology, VU University Medical Center, Amsterdam Neuroscience, Amsterdam 1081HV, The Netherlands. [53]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald 17475, Germany. [54]Chair of Experimental Bioinformatics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising-Weihenstephan 85354, Germany. [55]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, 75108 Uppsala, Sweden. [56]Vth Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, Heidelberg University, Mannheim 68167, Germany. [57]Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick National Lab for Cancer Research, Frederick, MD, USA. [58]Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [59]Department of Computational Biology and Medical Sciences, Graduate school of Frontier Sciences, The University of Tokyo, Tokyo 108-8639, Japan. [60]Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany. [61]German Center for Diabetes Research (DZD), München-Neuherberg, Germany. [62]Medical Research Institute, Ninewells Hospital and School of Medicine, University of Dundee, Dundee, UK. [63]Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland, School of Medicine, Baltimore, MD 21201, USA. [64]Program for Personalized and Genomic Medicine, Department of Medicine, University of Maryland, School of Medicine, Baltimore, MD 21201, USA. [65]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, The Netherlands. [66]Institute for Cardiogenetics, University of Lübeck, Lübeck 23562, Germany. [67]DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, Lübeck 23562, Germany. [68]University Heart Center Luebeck, Lübeck 23562, Germany. [69]Charité – University Medicine Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute for Dental and Craniofacial Sciences, Department of Periodontology and Synoptic Dentistry, Berlin, Germany. [70]Institute of Genetics and Biophysics A. Buzzati-Traverso - CNR, Naples 80131, Italy. [71]Finnish Institute for Molecular Medicine, University of Helsinki, Helsinki, Finland. [72]Genomics and Molecular Medicine Unit, CSIR-Institute of Genomics and Integrative Biology, New Delhi 110020, India. [73]Department of Clinical Sciences, Diabetes and Endocrinology, Lund University Diabetes Centre, Lund University, Skåne University Hospital, Malmö 20502, Sweden. [74]Genomic Research on Complex diseases (GRC Group), CSIR-Centre for Cellular and Molecular Biology, Hyderabad, Telangana 500007, India. [75]ePhood Scientific Unit, ePhood SRL, Bresso (Milano) 20091, Italy. [76]Department of Health Sciences, University of Milano, Milano 20139, Italy. [77]Neuroalgology Unit, IRCCS Foundation Carlo Besta Neurological Institute, Milano 20133, Italy. [78]Department of Pediatrics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA. [79]Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892-5635, USA. [80]Department of Epidemiology Research, Statens Serum Institut, Copenhagen DK-2300, Denmark. [81]Icelandic Heart Association, Kopavogur 201, Iceland. [82]Faculty of Medicine, School of Health Sciences, University of Iceland,

Reykjavik 101, Iceland. [83]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA. [84]Institute of Health and Wellbeing, University of Glasgow, Glasgow G12 8RZ, UK. [85]Cardiovascular Medicine Unit, Department of Medicine Solna, Centre for Molecular Medicine, Karolinska Institutet, Stockholm 171 76, Sweden. [86]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore City, Maryland 21224, USA. [87]Department of Cardiology, Leiden University Medical Center, Leiden 2300 RC, the Netherlands. [88]Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Leiden 2300RC, the Netherlands. [89]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [90]Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [91]University of Groningen, University Medical Center Groningen, Department of Cardiology, Ther Netherlands, Groningen 9713 GZ, the Netherlands. [92]Italian Institute for Genomic Medicine (IIGM) and Dept. Medical Sciences, University of Turin, Italy, Turin 10126, Italy. [93]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, Saint Louis, MO 63110-1093, USA. [94]NIHR Barts Cardiovascular Biomedical Research Centre, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. [95]Department of Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK. [96]MRC Integrative Epidemiology Unit at the University of Bristol, Bristol BS8 2BN, UK. [97]School of Psychological Science, University of Bristol, Bristol BS8 1TU, UK. [98]Department of Medicine, GeneSTAR Research Program, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA. [99]Department of Genetic Medicine, Weill Cornell Medicine Qatar, Doha, Qatar. [100]Computer and Systems Engineering, Alexandria University, Alexandria, Egypt. [101]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI 48109, USA. [102]Departamento de Endocrinología y Metabolismo, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico 14080, México. [103]Unidad de Investigacion de Enfermades Metabolicas, Tecnologico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Monterrey, N.L. 64710, México. [104]Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Fukuoka 812-8582, Japan. [105]Immunobiology of Dendritic Cells, Institut Pasteur, Paris 75015, France. [106]Inserm U1223, Paris 75015, France. [107]Centre for Translational Research, Institut Pasteur, Paris 75015, France. [108]Department of Cancer Immunology, Genentech Inc, San Francisco, California 94080, USA. [109]Division of Preventive Medicine, Department of Family Medicine and Public Health, UC San Diego School of Medicine, La Jolla, California 92093, USA. [110]Estonian Genome Center, University of Tartu, University of Tartu, Tartu 51010, Estonia. [111]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore 169856, Singapore. [112]Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, Singapore 169857, Singapore. [113]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore 119228 SG, Singapore. [114]Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran 19839-63113, Iran. [115]Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany. [116]Health Science Center at Houston, UTHealth School of Public Health, University of Texas, Houston, TX 77030, USA. [117]USC-Office of Population Studies Foundation, Inc., Department of Nutrition and Dietetics, Talamban, University of San Carlos, Cebu City 6000 Cebu, Philippines. [118]Department of Vascular Surgery, Division of Surgical Specialties, University Medical Center Utrecht, University of Utrecht, Utrecht, Utrecht 3584 CX, Netherlands. [119]Digital Health Center, Hasso Plattner Institute, Universität Potsdam, Potsdam 14482, Germany. [120]Division of Cancer Epidemiology & Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA. [121]Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan, Taiwan. [122]Institute for Maternal and Child Health - IRCCS Burlo Garofolo, Trieste 34137, Italy. [123]National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [124]Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK. [125]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228 SG, Singapore. [126]Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA. [127]Cellular and Molecular Endocrine Research Center, Research Institute for Endocrine Sciences, Shahid Beheshti University of Medical Sciences, Tehran 19839-63113, Iran. [128]Department of Internal Medicine, Rush University Medical Center, Chicago, Illinois, USA. [129]MEMO Research, Molecular and Clinical Medicine, University of Dundee, Dundee DD19SY, UK. [130]Department of Internal Medicine B, University Medicine Greifswald, Greifswald 17475, Germany. [131]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, Greifswald 17475, Germany. [132]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh, KSA 12713, Saudi Arabia. [133]Department of Anthropology, University of Iceland, Reykjavik 101, Iceland. [134]National Institute for Health Research Imperial Biomedical Research Centre, Imperial College Healthcare NHS Trust and Imperial College London, London, UK. [135]UK Dementia Research Institute (UK DRI) at Imperial College London, London, UK. [136]Health Data Research UK - London, London, England. [137]Department of Pediatrics, Erasmus University Medical Center, Rotterdam 3015CN, The Netherlands. [138]Section on Nephrology, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC 27101, US. [139]Medical Genetics, Institute for Maternal and Child Health - IRCCS Burlo Garofolo, Trieste, Italy. [140]Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, UK. [141]Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, California 90048, USA. [142]Braun School of Public Health, Hebrew University-Hadassah Medical Center, Jerusalem, Israel. [143]School of Engineering and Natural Sciences, University of Iceland, Reykjavik 101, Iceland. [144]Chinese Academy of Medical Sciences, Beijing 100730, China. [145]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London SE5 8AF, UK. [146]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90089, USA. [147]Blizard Institute, Queen Mary University of London, London E1 2AT, UK. [148]Laboratory of Genome Technology, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. [149]Laboratory of Clinical Chemistry and Hematology, Division Laboratories and Pharmacy, University Medical Center Utrecht, University of Utrecht, Utrecht, Utrecht 3584 CX, Netherlands. [150]Shanghai Institute of Hematology, State Key Laboratory Of Medical Genomics, Rui-jin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China 200025, China. [151]Division of Endocrine and Metabolism, Tri-Service General Hospital Songshan branch, Taipei, Taiwan, Taiwan. [152]School of Medicine, National Defense Medical Center, Taipei, Taiwan, Taiwan. [153]Unit of Public Health Promotion, National Institute for Health and Welfare, Helsinki, Finland. [154]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan. [155]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA. [156]Deutsches Herzzentrum München, Klinik für Herz- und Kreislauferkrankungen, Technische Universität München, Munich 80636, Germany. [157]Department of Public Health and Primary Care, University of Cambridge, Cambridge CB2 0SR, UK. [158]Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore 138672, Singapore. [159]Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore 169857. [160]Centre for Global Health, Faculty of Medicine, University of Split, Split, Croatia. [161]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland. [162]Kuopio University Hospital, Kuopio, Finland. [163]Bristol NIHR Biomedical Research Centre, Bristol BS8 2BN, UK. [164]Population Health Science, Bristol Medical School, Bristol BS8 2BY, UK. [165]Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan, Taiwan. [166]School of Medicine, National Yang-Ming University, Taipei, Taiwan, Taipei 112, Taiwan. [167]School of Medicine, Chung Shan Medical University, Taichung, Taiwan, Taichung City 402, Taiwan. [168]Department of Medical Research, Taichung Veterans General Hospital, Taichung, Taiwan, Taiwan. [169]Department of Internal Medicine A, University Medicine Greifswald, Greifswald 17475, Germany. [170]Department of Epidemiology and Biostatistics, Peking University

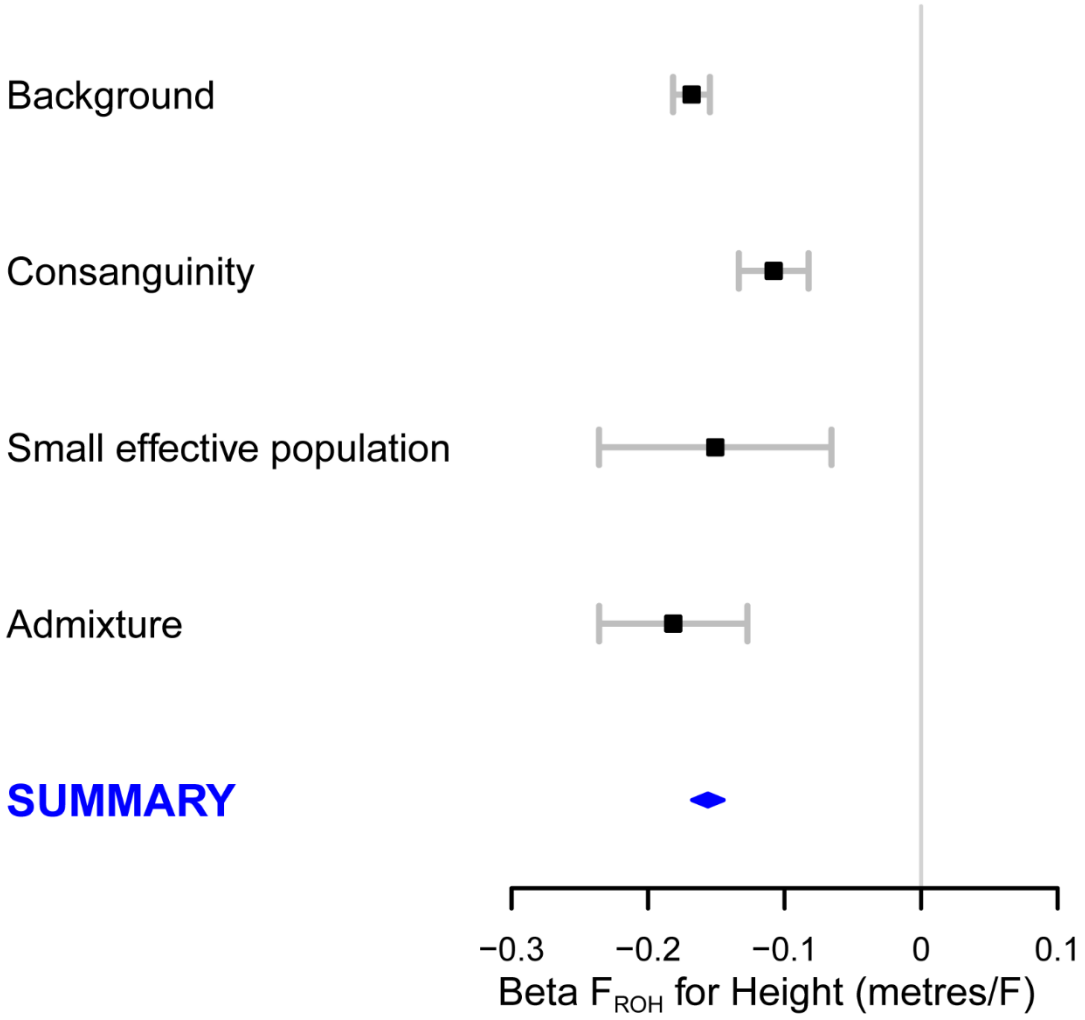
Health Science Centre, Peking University, Beijing 100191, China. [171]Translational Laboratory in Genetic Medicine, Agency for Science, Technology and Research, Singapore (A*STAR), Singapore 138648, Singapore. [172]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore 117596, Singapore. [173]National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, Durham, NC 27709, USA. [174]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA. [175]Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA 22908, USA. [176]Genomics of Renal Diseases and Hypertension Unit, IRCCS San Raffaele Scientific Institute, Università Vita Salute San Raffaele, Milano 20132, Italy. [177]Helmholtz Zentrum München, Independent Research Group Clinical Epidemiology, Neuherberg 85764, Germany. [178]Institute of Human Genetics, Helmholtz Zentrum Muenchen, Neuherberg 85764, Germany. [179]Institute of Human Genetics, Technical University of Munich, Munich 81675, Germany. [180]DZHK (German Center for Cardiovascular Research), partner site Munich Heart Alliance, Munich 80802, Germany. [181]Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu 51010, Estonia. [182]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa 230-0045, Japan. [183]Laboratory of Haematology, La Timone Hospital, Marseille, France. [184]INSERM UMR_S 1263, Center for CardioVascular and Nutrition research (C2VN), Aix-Marseille University, Marseille, France. [185]Departamento de Economía, Universidad Autónoma Metropolitana, Mexico 09340, México. [186]Medical School, The University of Western Australia, Perth, Western Australia/Australia 6009, Australia. [187]The University of Texas Health Science Center at Houston, School of Public Health, Department of Epidemiology, Human Genetics and Environmental Sciences, Houston, Texas 77030, USA. [188]Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland. [189]Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan. [190]The Institute of Medical Sciences, Aberdeen Biomedical Imaging Centre, University of Aberdeen, Aberdeen AB25 2ZD, UK. [191]Laboratory of Neurogenetics, Bethesda, MD 20892, USA. [192]Data Tecnica International LLC, Glen Echo, MD 20812, USA. [193]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald 17475, Germany. [194]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Headington, Oxford OX3 7LJ, UK. [195]Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK. [196]Department of Paediatrics, University of Cambridge School of Clinical Medicine, Cambridge CB2 0QQ, UK. [197]Instituto Nacional de Medicina Genómica, Mexico 14610, México. [198]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow G12 8TA, UK. [199]Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, MN 55454, USA. [200]Gen-info Ltd, Zagreb, Croatia, Zagreb, Select a Province 10000, Croatia. [201]International Centre for Circulatory Health, Imperial College London, London W2 1PG, UK. [202]Imperial Clinical Trials Unit, Imperial College London, London, London W12 7TA, UK. [203]Human Evolutionary Genetics Unit, Institut Pasteur, Paris 75015, France. [204]Centre National de la Recherche Scientifique (CNRS) UMR2000, Paris 75015, France. [205]Center of Bioinformatics, Biostatistics and Integrative Biology, Institut Pasteur, Paris 75015, France. [206]Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, University of Helsinki, Helsinki 00014, Finland. [207]Hero Heart Institute and Dyanand Medical College and Hospital, Ludhiana, Punjab, India. [208]Division of Biostatistics, Washington University School of Medicine, St. Louis, Missouri, USA. [209]Harvard Medical School, Boston, MA 02115, USA. [210]Department of Applied Economics, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam 3062 PA, The Netherlands. [211]Erasmus University Rotterdam Institute for Behavior and Biology, Erasmus University Rotterdam, Rotterdam 3062 PA, The Netherlands. [212]IRCCS Neuromed, Pozzilli (IS) 86077, Italy. [213]Gottfried Schatz Research Center (for Cell Signaling, Metabolism and Aging), Division of Molecular Biology and Biochemistry, Medical University of Graz, 8010 Graz, Austria. [214]Unit of Genomics of Complex Diseases, Institut de Recerca Hospital de la Santa Creu i Sant Pau, IIB-Sant Pau, Barcelona, Spain. [215]San Raffaele Research Institute, Milano, Italy. [216]Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki FI-00271, Finland. [217]Department of Biostatistics, and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. [218]Dept of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala, SE-752 37 Uppsala, Sweden. [219]Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, University of Utrecht, Utrecht, Utrecht 3485 CX, Netherlands. [220]School of Medicine, National Yang-Ming University, Taipei, Taiwan, Taiwan. [221]Institute of Medical Technology, National Chung-Hsing University, Taichung, Taiwan, Taiwan. [222]School of Basic and Applied Sciences, Dayananda Sagar University, Bangalore, Karnataka 560078, India. [223]Diabetes Unit, KEM Hospital and Research Centre, Pune, Maharashtra 411101, India. [224]Symbiosis Statistical Institute, Symbiosis International University, Pune, Maharashtra 411007, India. [225]Panjab University, Chandigarh, India. [226]Division of Population Health Sciences, Ninewells Hospital and Medical School, University of Dundee, Dundee DD1 9SY, UK. [227]RCSI Molecular & Cellular Therapeutics (MCT), Royal College of Surgeons in Ireland, RCSI Education & Research Centre, Beaumont Hospital, Dublin 9, Ireland. [228]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh EH8 9JZ, Scotland. [229]School of Physiotherapy and Exercise Science, Faculty of Health Sciences, Curtin University, Perth, Western Australia/Australia 6102, Australia. [230]Department of Respiratory Medicine, Ghent University Hospital, Ghent 9000, Belgium. [231]Department of Pathology, University of Vermont, Colchester, VT 05446, USA. [232]Departamento de Medicina Genómica y Toxicología Ambiental, Instituto de Investigaciones Biomédicas, UNAM, Mexico 04510, México. [233]Unidad De Biología Molecular y Medicina Genómica, Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico 14080, México. [234]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina 45110, Greece. [235]Institute of Genetic and Biomedical Research - Support Unity, National Research Council of Italy, Rome, Italy. [236]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, Scotland. [237]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki FI-00014, Finland. [238]Department of Cardiology, Hero DMC Heart Institute, Dayanand Medical College & Hospital, Ludhiana, Punjab 141001, India. [239]Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing Tongren Hospital, Capital Medical University, Beijing Ophthalmology and Visual Science Key Lab, Beijing, China 100005, China. [240]Centre for Population Health Sciences, Usher Institute of Population Health and Informatics, University of Edinburgh, Edinburgh EH8 9AG, Scotland. [241]Diabetes Unit, K.E.M. Hospital Research Centre, Pune, MAH 411011, India. [242]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [243]Department of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, University of Utrecht, Utrecht, Utrecht 3584 CX, Netherlands. [244]Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London WC1E 6DD, UK. [245]Durrer Center for Cardiovascular Research, Netherlands Heart Institute, Utrecht, Netherlands. [246]Farr Institute of Health Informatics Research and Institute of Health Informatics, University College London, London, UK. [247]Department of Internal Medicine, University Medical Center Groningen, University of Groningen, Groningen 9713GZ, The Netherlands. [248]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL 60612, USA. [249]Department of Neurological Sciences, Rush University Medical Center, Chicago, IL 60612, USA. [250]Systems Genomics Laboratory, School of Biotechnology, Jawaharlal Nehru University, New Delhi 110067, India. [251]Pennington Biomedical Research Center, Baton Rouge, Louisiane 70808, USA. [252]Center for Applied Genomics, Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA. [253]Quantinuum Research LLC, San Diego, CA 92101, USA. [254]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA. [255]Center for Experimental Social Science, Department of Economics, New York University, New York, New York 10012, USA. [256]Research Institute for Industrial Economics (IFN), Stockholm 102 15, Sweden. [257]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore. [258]Imperial College Healthcare NHS Trust, London, London W12 0HS, UK. [259]Department of Preventive Medicine, Northwestern University

Feinberg School of Medicine, Chicago, IL 60611, USA. [260]Institute of Biomedical Technologies Milano, National Research Council of Italy (CNR), Segrate (Milano) 20090, Italy. [261]Bio4Dreams Scientific Unit, Bio4Dreams SRL, Bio4Dreams - business nursery for life sciences, Milano 20121, Italy. [262]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, 24105 Kiel, Germany. [263]Department of General Practice and Primary health Care, University of Helsinki, Tukholmankatu 8 B, Helsinki 00014, Finland. [264]National Institute for Health and Welfare, Helsinki, Finland. [265]Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland. [266]Folkhälsan Research Centre, Helsinki, Finland. [267]Vasa Central Hospital, Vaasa, Finland. [268]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI 48014, USA. [269]Montreal Heart Institute, Montreal, QC, Canada. [270]Centro de Estudios en Diabetes, Unidad de Investigacion en Diabetes y Riesgo Cardiovascular, Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Publica, Cuernavaca 01120, México. [271]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [272]Cardiovascular Medicine Unit, Department of Medicine Solna, Centre for Molecular Medicine, Stockholm 171 76, Sweden. [273]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [274]Khoo Teck Puat - National University Children's Medical Institute, National University Health System, Singapore, Singapore. [275]Research Unit of Internal Medicine, Medical Research Center Oulu, University of Oulu and Oulu University Hospital, Oulu 90014, Finland. [276]Division of Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah 84108, USA. [277]Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, 650 West 168th street, PH19-311, Newyork, NY 10032, USA. [278]Cell Circuits Program, Broad Institute, Cambridge, MA 02142, USA. [279]Department of Economics, Stockholm School of Economics, Stockholm SE-113 83, Sweden. [280]Department of Ophthalmology, Medical Faculty Mannheim of the Ruprecht-Karls-University of Heidelberg, Mannheim 698167, Germany. [281]Department of Public Health, University of Helsinki, Helsinki FI-00014, Finland. [282]Oxford NIHR Biomedical Research Centre, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford OX3 9DU, UK. [283]Laboratory of Clinical Chemistry and Hematology, Division Laboratories and Pharmacy, University Medical Center Utrecht, University of Utrecht, Utrecht 3584 CX, Netherlands. [284]Helsinki Collegium for Advanced Studies, University of Helsinki, University of Helsinki, Helsinki 00014, Finland. [285]University Hospital Schleswig-Holstein (UKSH), Campus Kiel, Kiel 24105, Germany. [286]Institute of Epidemiology and PopGen Biobank, University of Kiel, Kiel, Schleswig Holstein 24105, Germany. [287]Department of Statistics, University of Auckland, Auckland, New Zealand. [288]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz, Austria. [289]Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany. [290]Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh 11211, Saudi Arabia. [291]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia. [292]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden 2333 ZA, The Netherlands. [293]School of Biomedical Sciences, Institute of Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, QLD 4059, Australia. [294]Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, The Netherlands. [295]Department of Medicine, Columbia University Medical Center, New York, New York, USA. [296]Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, The School of Medicine, University of Dundee, Dundee DD1 9SY, UK. [297]School of Medicine and Public Health, Faculty of Medicine and Health, The University of Newcastle, Newcastle, New South Wales, Australia. [298]Division of Obstetrics and Gynaecology, The University of Western Australia, Perth, Western Australia/Australia 6009, Australia. [299]Department of kinesiology, Laval University, Quebec, QC G1V 0A6, Canada. [300]Institute of Nutrition and Functional Foods, Laval University, Quebec, QC G1V 0A6, Canada. [301]Institute of Genetic and Biomedical Research - Support Unity, National Research Council of Italy, Sassari 07100, Italy. [302]Department of Clinical Genetics, Amsterdam Neuroscience, VU Medical Centre, Amsterdam 1081HV, The Netherlands. [303]Cardiovascular Health Research Unit, Departments of Epidemiology, Medicine and Health Services, University of Washington, Seattle, WA 98101, USA. [304]Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101, USA. [305]Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec H3T 1J4, Canada. [306]Laboratory of Experimental Cardiology, Division Heart & Lungs, University Medical Center Utrecht, University of Utrecht, Utrecht, Utrecht 3584 CX, Netherlands. [307]Oklahoma Center for Neuroscience, Oklahoma City, OK 73104, USA. [308]Institute of Nutritional Sciences, University of Potsdam, Nuthetal, Germany. [309]Deutsches Zentrum für Herz- und Kreislauferkrankungen (DZHK), Munich Heart Alliance, Munich 80636, Germany. [310]Regeneron Genetics Center, Regeneron Pharmaceuticals, Inc, Tarrytown, NY 10591-6607, USA. [311]Faculty of Health Studies, University of Bradford, Bradford, West Yorkshire BD7 1DP, UK. [312]Cardiovascular Health Research Unit, Division of Cardiology, University of Washington, Seattle, WA 98101, USA. [313]Duke-NUS Medical School, National University of Singapore, Singapore, Singapore 169857 SG, Singapore. [314]Institute for Community Medicine, University Medicine Greifswald, Greifswald 17475, Germany. [315]Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2PR, UK. [316]Avon Longitudinal Study of Parents and Children (ALSPAC), University of Bristol, Bristol BS8 2PR, UK. [317]Endocrinology, Abdominal Centre, University of Helsinki, Helsinki University Hospital, Helsinki, Finland. [318]Folkhalsan Research Center, Helsinki, Finland. [319]Research Program of Diabetes and Endocrinology, University of Helsinki, Helsinki, Finland. [320]Department of Medicine, Internal Medicine, Lausanne University Hospital, Lausanne 1011, Switzerland. [321]Division of Gastroenterology, Hepatology and Nutrition, Children's Hospital of Philadelphia, Philadelphia, PA 19146, USA. [322]Unit of Genomics and Biomarkers, National Institute for Health and Welfare, Helsinki 00271, Finland. [323]National Heart and Lung Institute, Imperial College London, London W12 0NN, UK. [324]The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [325]Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, Scotland. [326]Program in Medical and Population Genetics, Broad Institute, Broad Institute, Cambridge, MASSACHUSETTS 02142, USA. [327]Deceased: Yucheng Jia, John M. Starr.

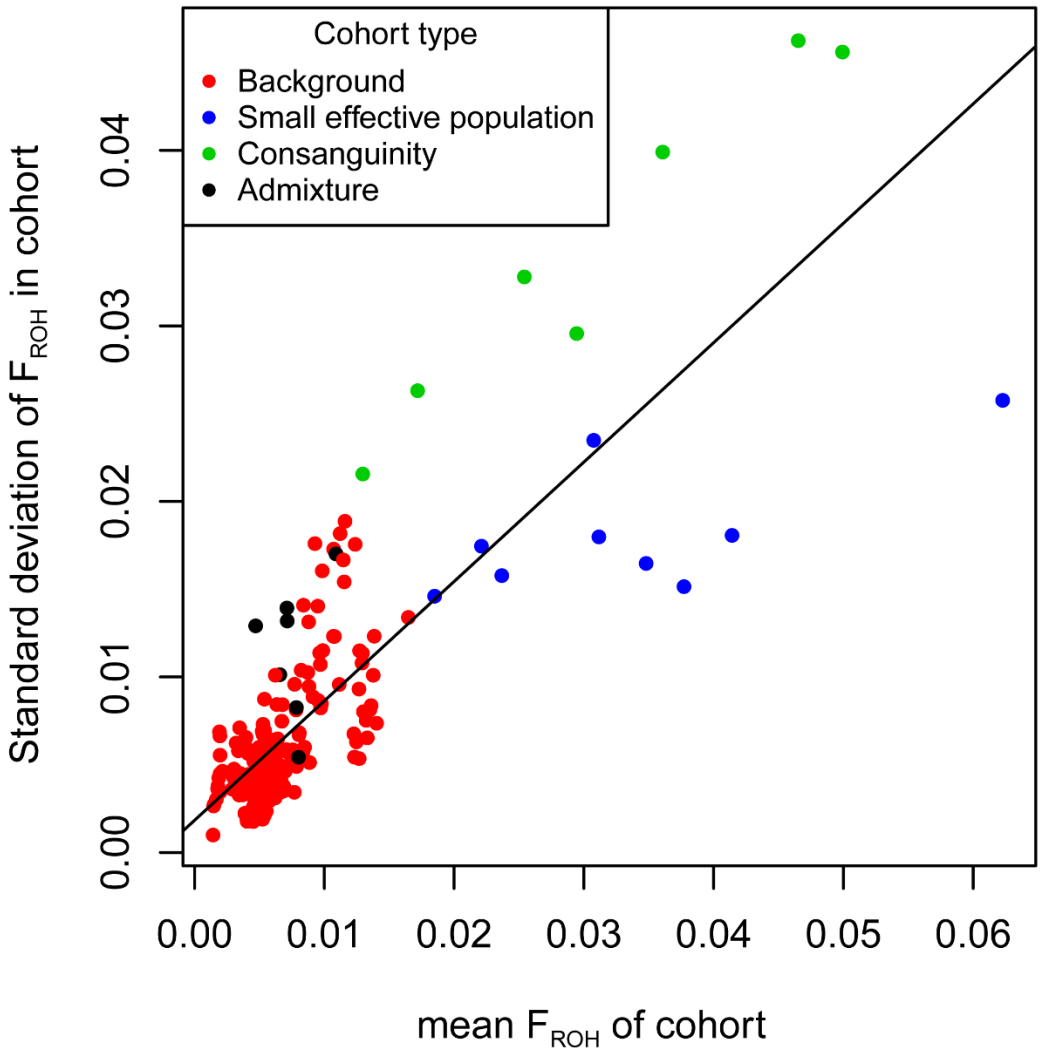
SUPPLEMENTARY FIGURES



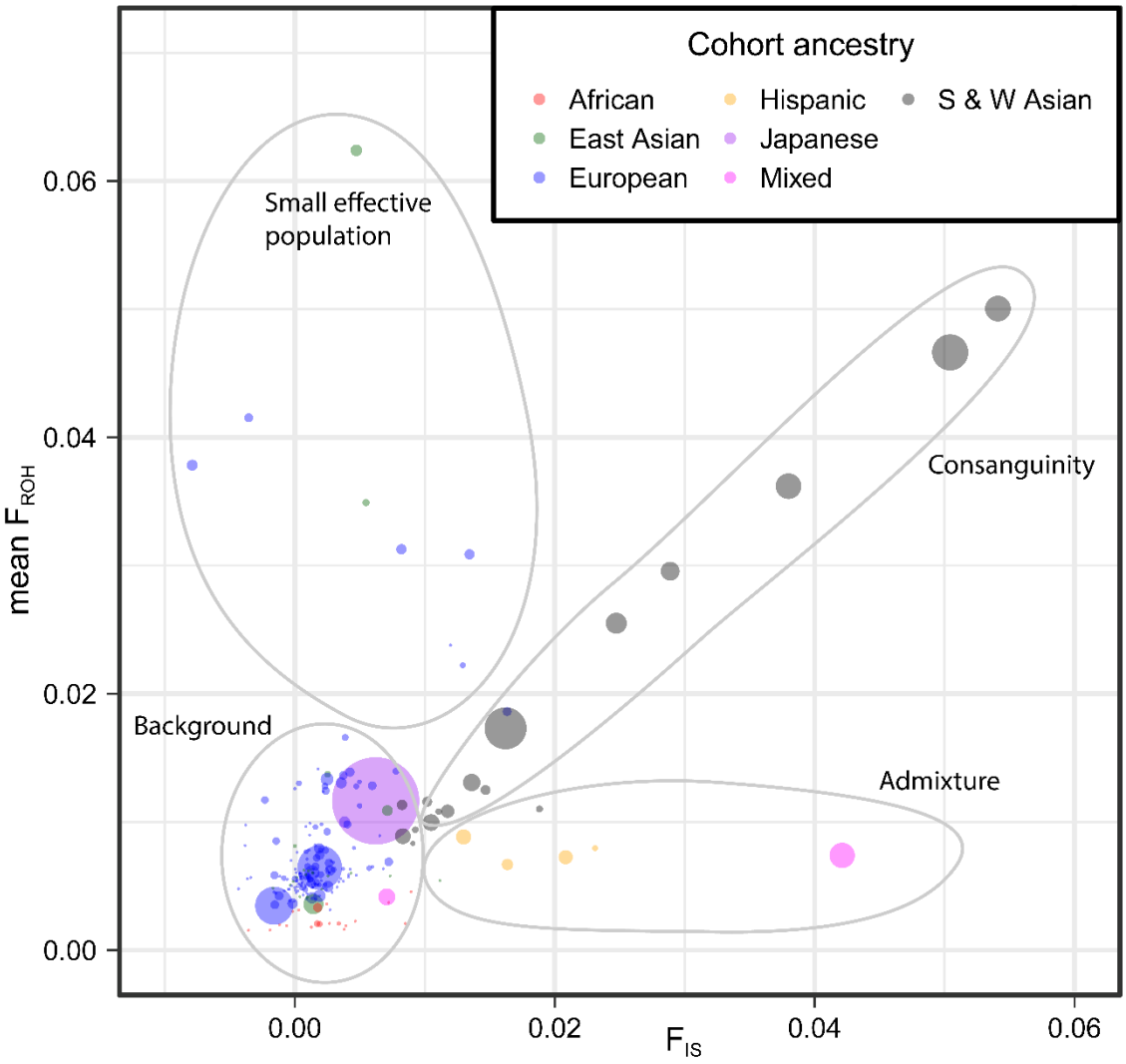
Supplementary Figure 1: Effect of F_{ROH} on height is robust to stratification by ancestral group. Cohorts were divided into eight broad ancestral groups (Supplementary Data Table 1) and meta-analysed separately. Although some heterogeneity is observed (heterogeneity *p*-value = 3 x 10^{-4}), $\beta_{F_{\mathrm{ROH}}}$ is directionally consistent and differs significantly from 0 in all ancestral groups. All errors bars represent 95% confidence intervals.



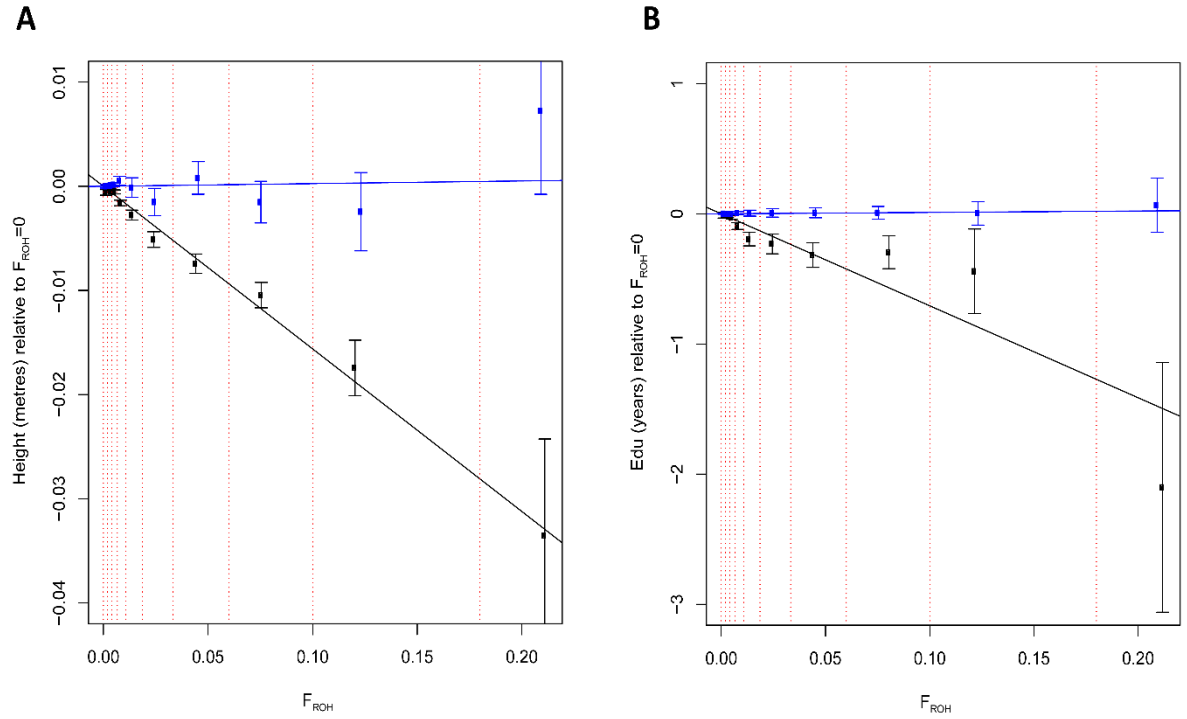
Supplementary Figure 2: Effect of F_{ROH} on height is robust to stratification by inferred demographic history. Cohorts were divided by inferred demographic history (Supplementary Fig. 4, Supplementary Data Table 1) and meta-analysed separately. A small amount of heterogeneity is observed (heterogeneity *p*-value = 0.008), but $\beta_{F_{\mathrm{ROH}}}$ is directionally consistent and differs significantly from 0 in all groups. In particular, in the small effective population size cohorts, where the variation of F_{ROH} is believed to be caused variations in cryptic relatedness between parents, $\beta_{F_{\mathrm{ROH}}}$ [-0.15, 95% CI -0.07 -0.23, *p*-value 3 x 10^{-4}] is consistent with the global estimate. All errors bars represent 95% confidence intervals.



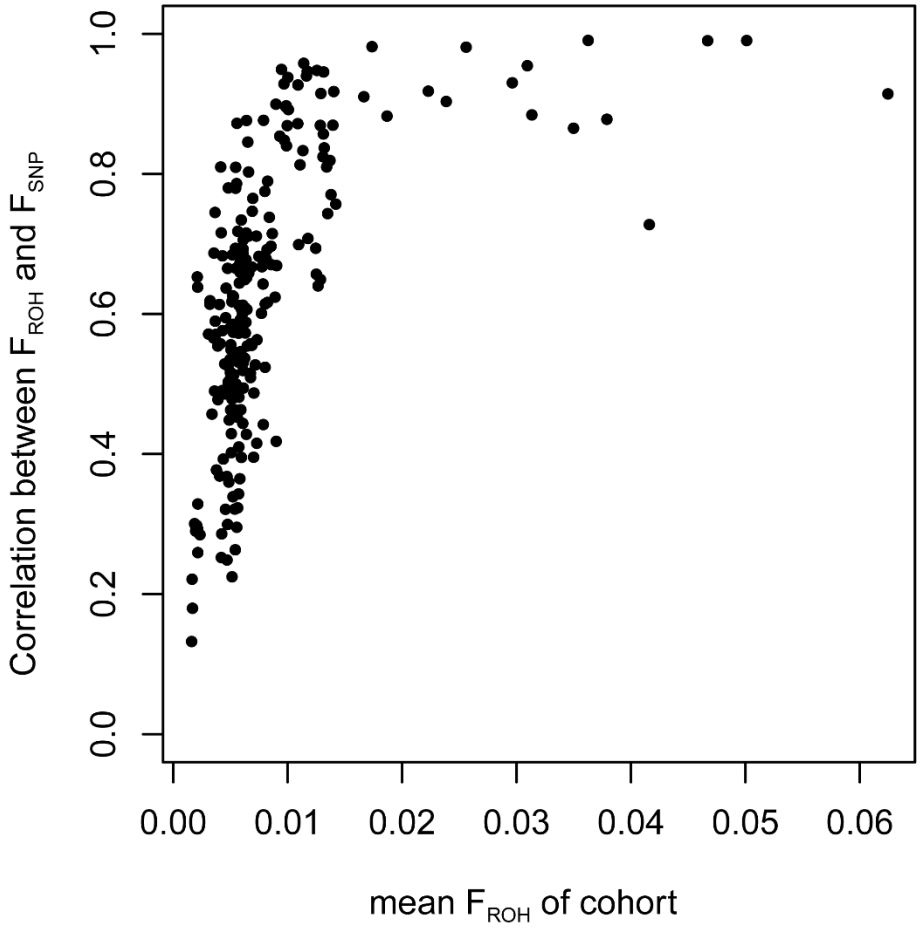
Supplementary Figure 3: A strong correlation (r=0.82, p-value = 9 x 10^{-103}) is observed between $\sigma_{F_{\mathrm{ROH}}}$ and mean F_{ROH}. The standard deviation of F_{ROH} ($\sigma_{F_{\mathrm{ROH}}}$) is plotted against mean F_{ROH} for all cohorts. In regressions on F_{ROH} the statistical power is approximately proportional to $\sigma^2_{F_{\mathrm{ROH}}}$ and cohorts with high mean F_{ROH} generally provide greater per-sample statistical power. Also, for a given mean F_{ROH}, cohorts where ROH are primarily attributable to consanguinity rather than small effective population size provide greater statistical power.



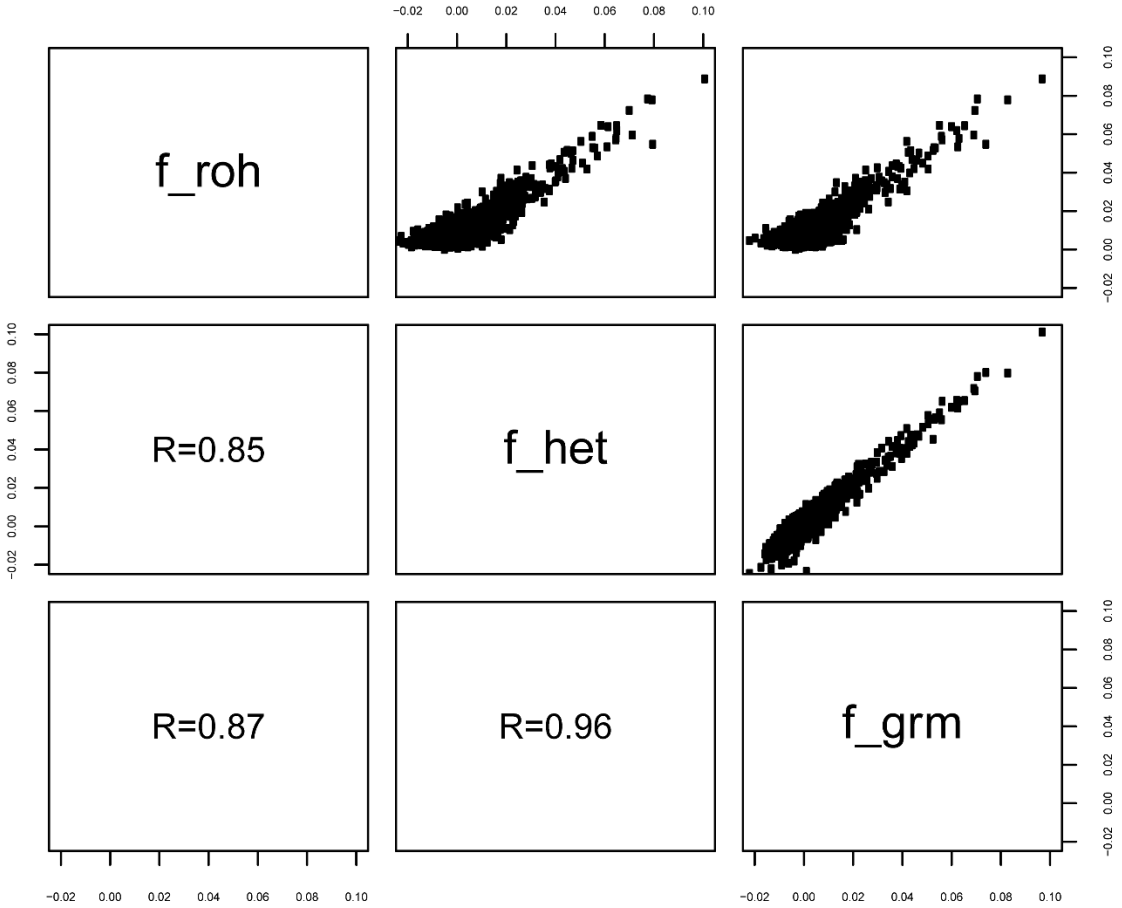
Supplementary Figure 4: Assignment of cohorts to one of four inferred demographic histories. Fig. 2 is replicated (see also Fig. 2 legend) and used to empirically assign cohorts to one of four inferred demographic histories. In cohorts where $F_{IS} > 0.1$, but the Cartesian distance to the 1:1 line was < 0.005, consanguinity was inferred to be the main origin of ROH. Cohorts which had not been defined as consanguineous but had mean $F_{ROH} > 0.02$ were consider to have a small effective population. Cohorts with $F_{IS} > 0.1$, but not consanguineous nor small effective population, were defined as admixed and the remaining cohorts were described as *background*.



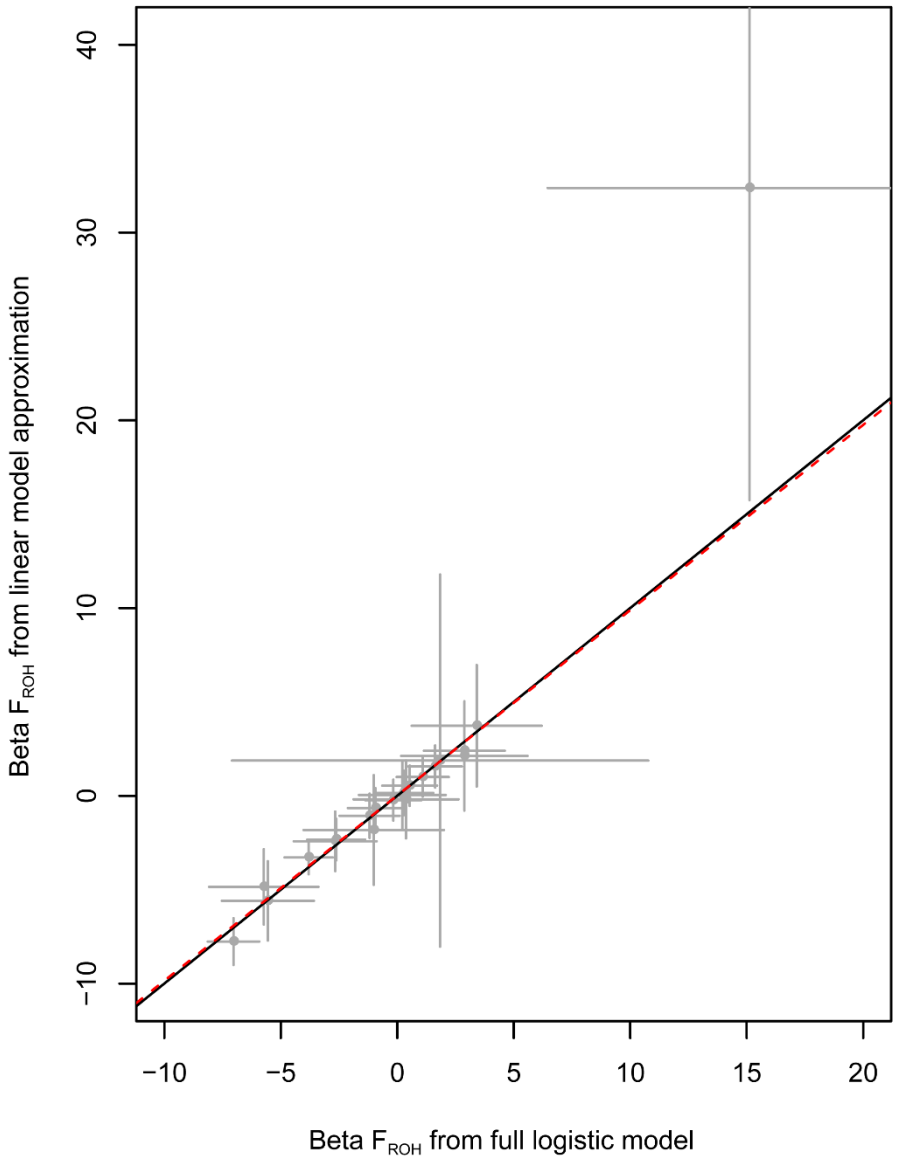
Supplementary Figure 5: Effect of assortative mating on height and educational attainment (a): Linear decrease in height with increasing F_{ROH} but no decrease in a polygenic score for height. In black, average height (in metres) is plotted in bins of increasing F_{ROH}. In blue, averages of a polygenic risk score for height are plotted in the same bins. Increased F_{ROH} is not associated with decreased polygenic score for height, providing evidence against a hypothesis of assortative mating generating the relationship with height. **(a): Decrease in education attained (EA) with increasing F_{ROH} but no decrease in a polygenic score for EA.** In black, average EA (in years) is plotted in bins of increasing F_{ROH}. In blue, average polygenic risk score for EA are plotted in the same bins. Increased F_{ROH} is not associated with decreased polygenic score for EA, providing evidence against a hypothesis of assortative mating generating the relationship with EA. All errors bars represent 95% confidence intervals.



Supplementary Figure 6: Strong correlations between F_{ROH} and F_{SNP} are observed in cohorts with high average F_{ROH}. The correlation between F_{ROH} and F_{SNP} is plotted against mean F_{ROH} for all cohorts. In low autozygosity cohorts the correlation between F_{ROH} and F_{SNP} is weak to moderate, as only a small fraction of homozygous SNPs is found in ROH. In contrast, in higher autozygosity cohorts, ROH represent a larger fraction of homozygous SNPs and the correlation between F_{ROH} and F_{SNP} is stronger.

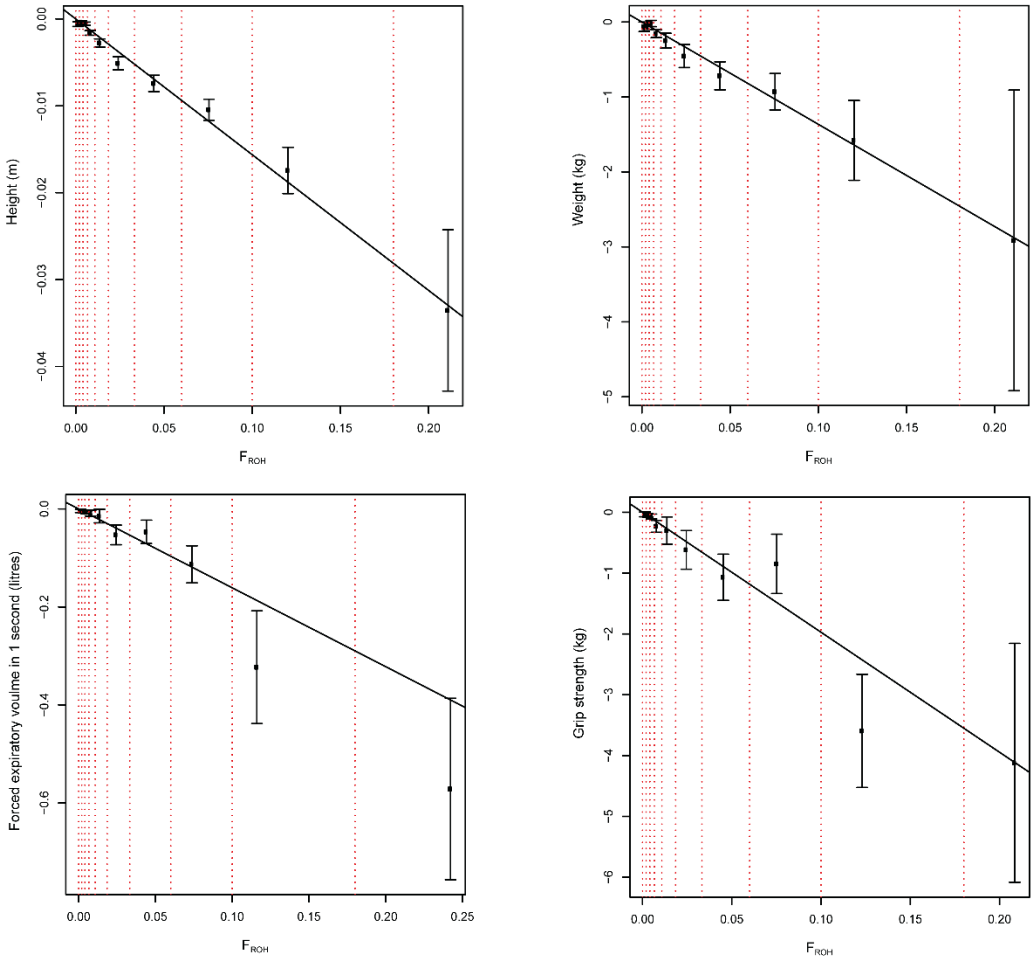


Supplementary Figure 7: Scatter plots of F_{ROH} plotted against F_{SNP} and F_{GRM} in a single cohort (VIKING). Scatter plots of three estimates of inbreeding coefficient (F_{ROH}, F_{SNP} and F_{GRM}) are shown in the upper right panels. The correlation between these estimates is shown in the lower left panels.

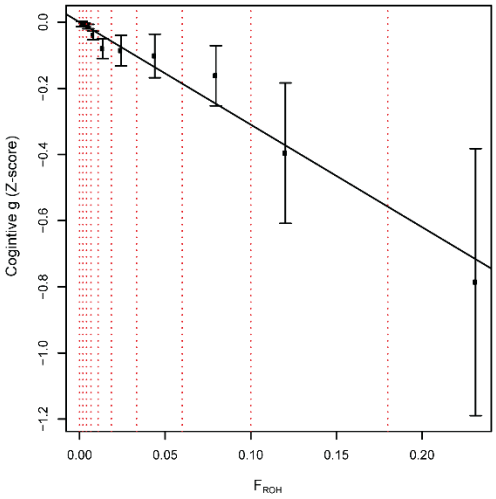


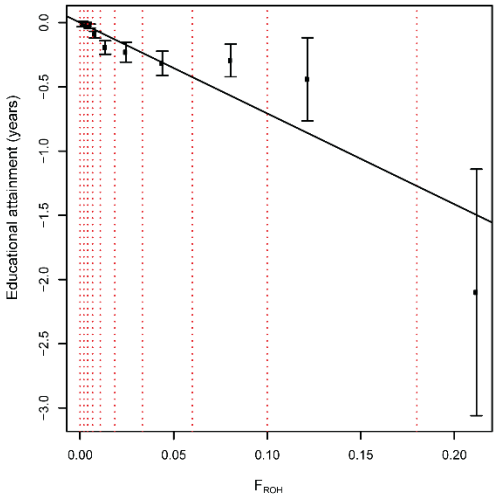
Supplementary Figure 8: A linear model approximation of the full logistic model gives relatively unbiased estimates of $\boldsymbol{\beta}_{\boldsymbol{F}_{\mathbf{ROH}}}$. For all 22 binary traits analysed, estimates obtained from a two-step linear model approximation are plotted against estimates obtained from the full logistic model (See Methods). Estimates of $\boldsymbol{\beta}_{\boldsymbol{F}_{\mathbf{ROH}}}$ are shown in grey. The 1:1 unity line is shown in red, and a linear least-squares fit is shown in black. The gradient of the best fit line (1.02 95% CI 0.99-1.02) does not differ significantly from the unbiased expectation of 1 (*p*-value 0.87). For all but one trait, the linear model approximation is consistent with the full logistic model estimate of $\boldsymbol{\beta}_{\boldsymbol{F}_{\mathbf{ROH}}}$. Self-declared infertility has the most extreme case:control ratio (632:472544) of any of the binary traits analysed and for this trait only the linear model significantly overestimates $\boldsymbol{\beta}_{\boldsymbol{F}_{\mathbf{ROH}}}$. The linear model estimates are therefore marked with an asterisk where they appear in Supplementary Data Tables 12-14, 16-21. All errors bars represent 95% confidence intervals.

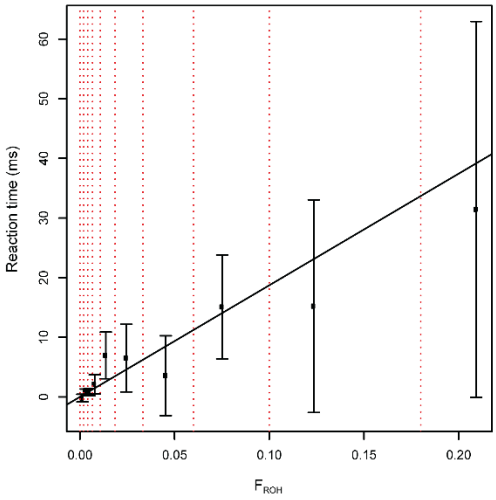
9a


Height (m)
0.00
−0.01
−0.02
−0.03
−0.04
0.00
0.05
0.10
0.15
0.20
F_ROH
Weight (kg)
0
−1
−2
−3
−4
−5
0.00
0.05
0.10
0.15
0.20
F_ROH
Forced expiratory voulme in 1 second (litres)
0.0
−0.2
−0.4
−0.6
0.00
0.05
0.10
0.15
0.20
0.25
F_ROH
Grip strength (kg)
0
−1
−2
−3
−4
−5
−6
0.00
0.05
0.10
0.15
0.20
F_ROH

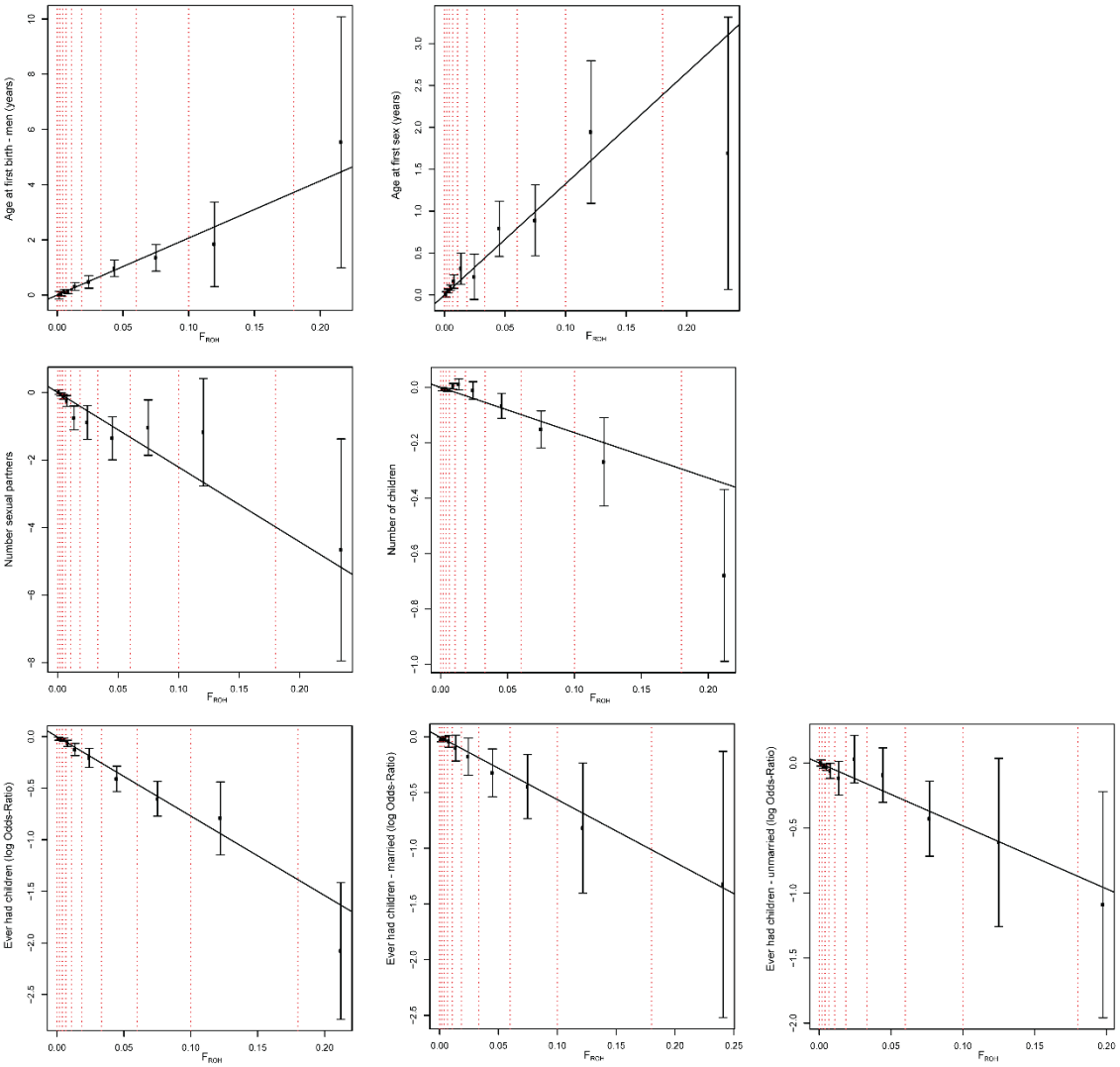
9b


Cogintive g (Z-score)
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
−1.2
0.00
0.05
0.10
0.15
0.20
F_ROH

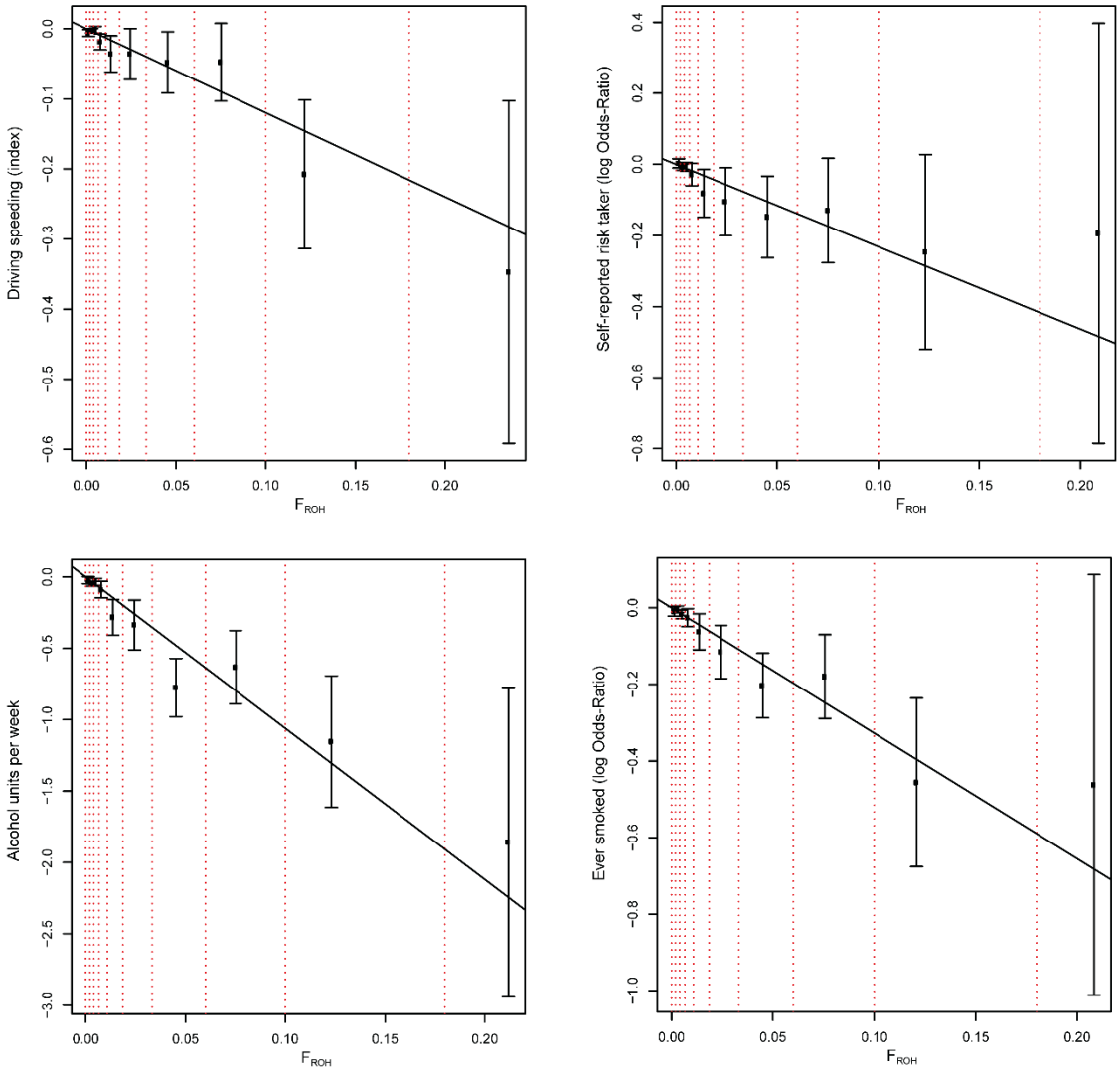

Educational attainment (years)
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
0.00
0.05
0.10
0.15
0.20
F_ROH


Reaction time (ms)
60
50
40
30
20
10
0
0.00
0.05
0.10
0.15
0.20
F_ROH

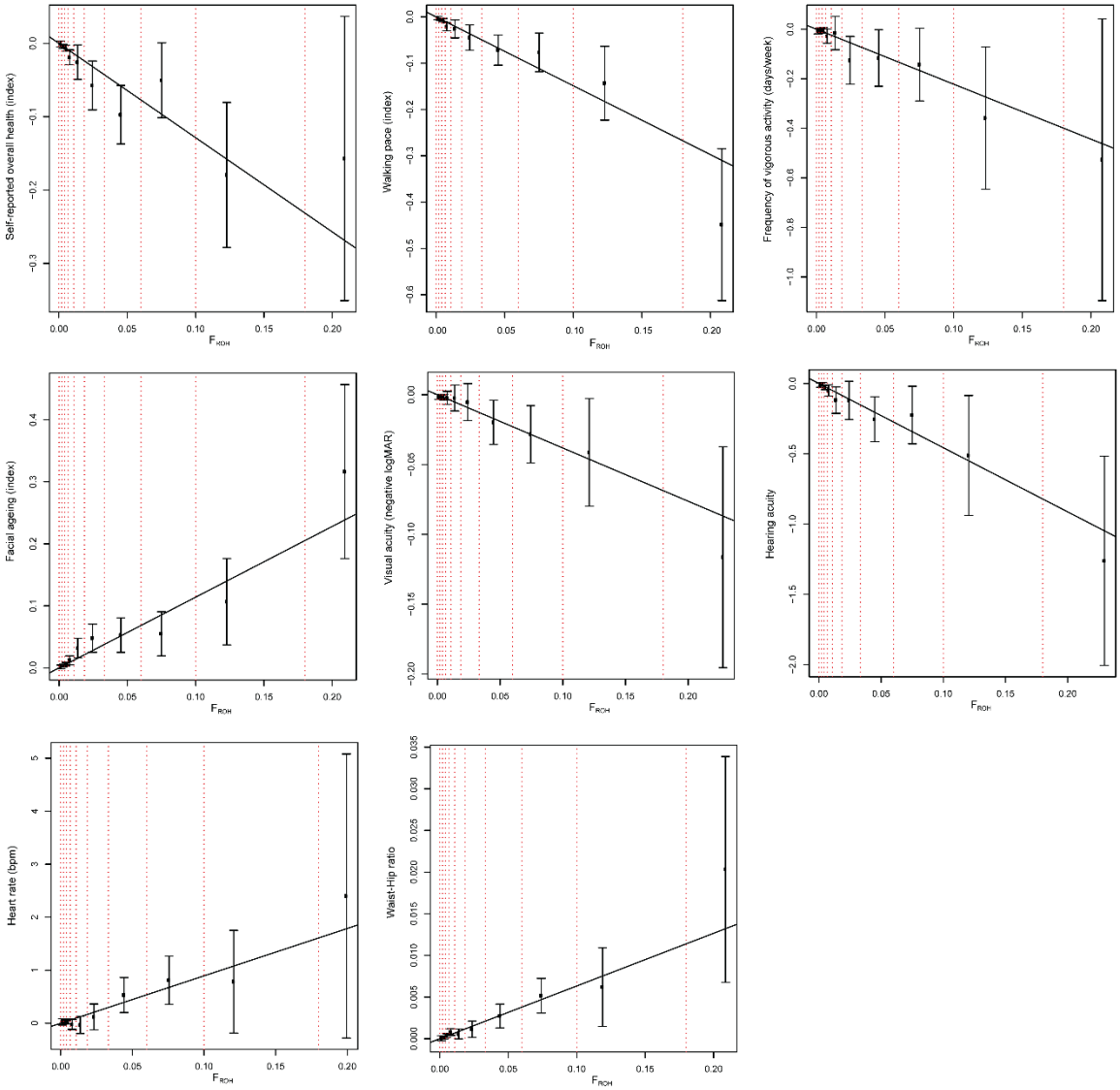
9c


Age at first birth - men (years)
Age at first sex (years)
Number sexual partners
Number of children
Ever had children (log Odds-Ratio)
Ever had children - married (log Odds-Ratio)
Ever had children - unmarried (log Odds-Ratio)
F_ROH

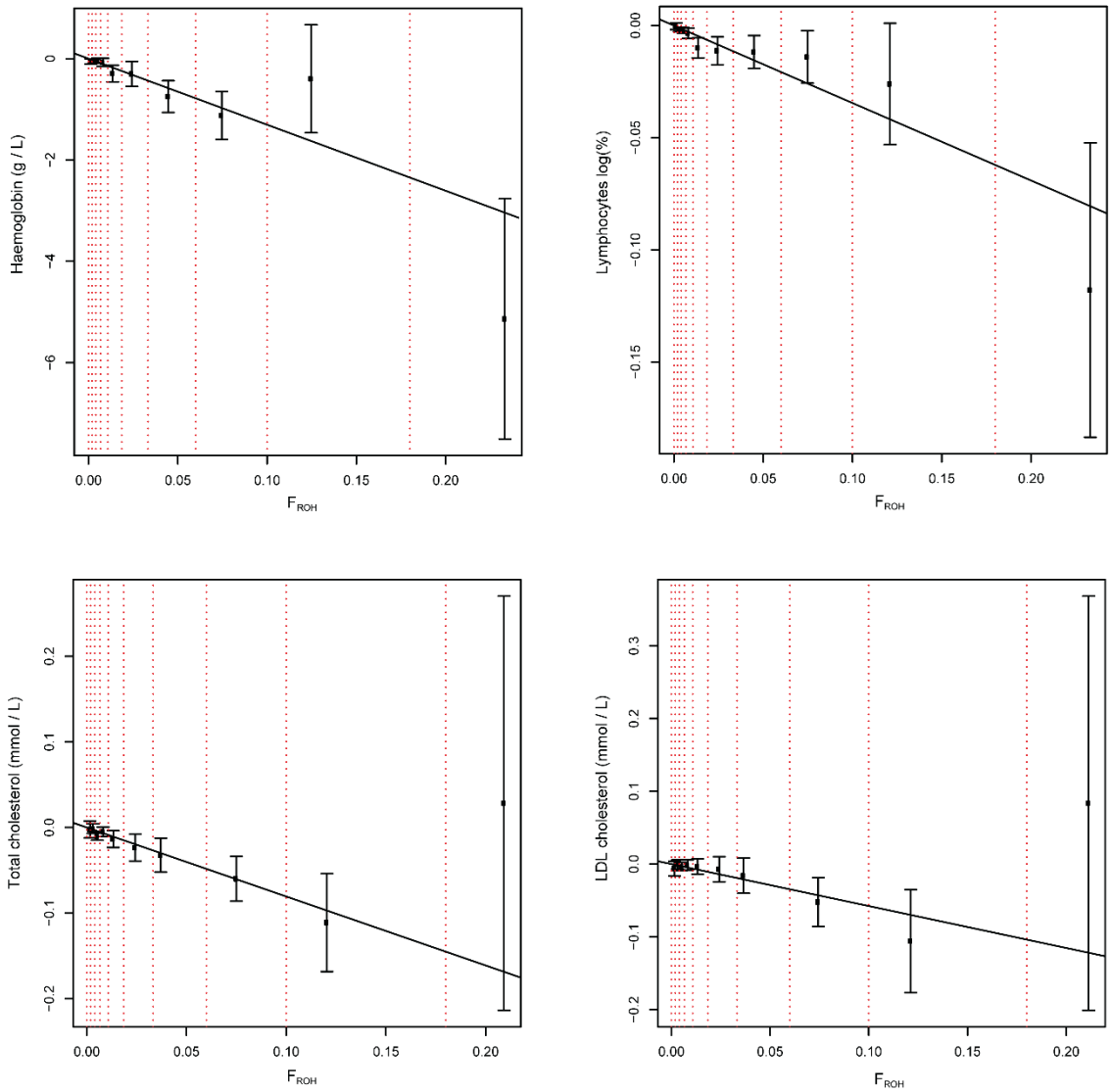
9d


Driving speeding (index)
Self-reported risk taker (log Odds-Ratio)
Alcohol units per week
Ever smoked (log Odds-Ratio)
F_ROH
0.00
0.05
0.10
0.15
0.20

9e

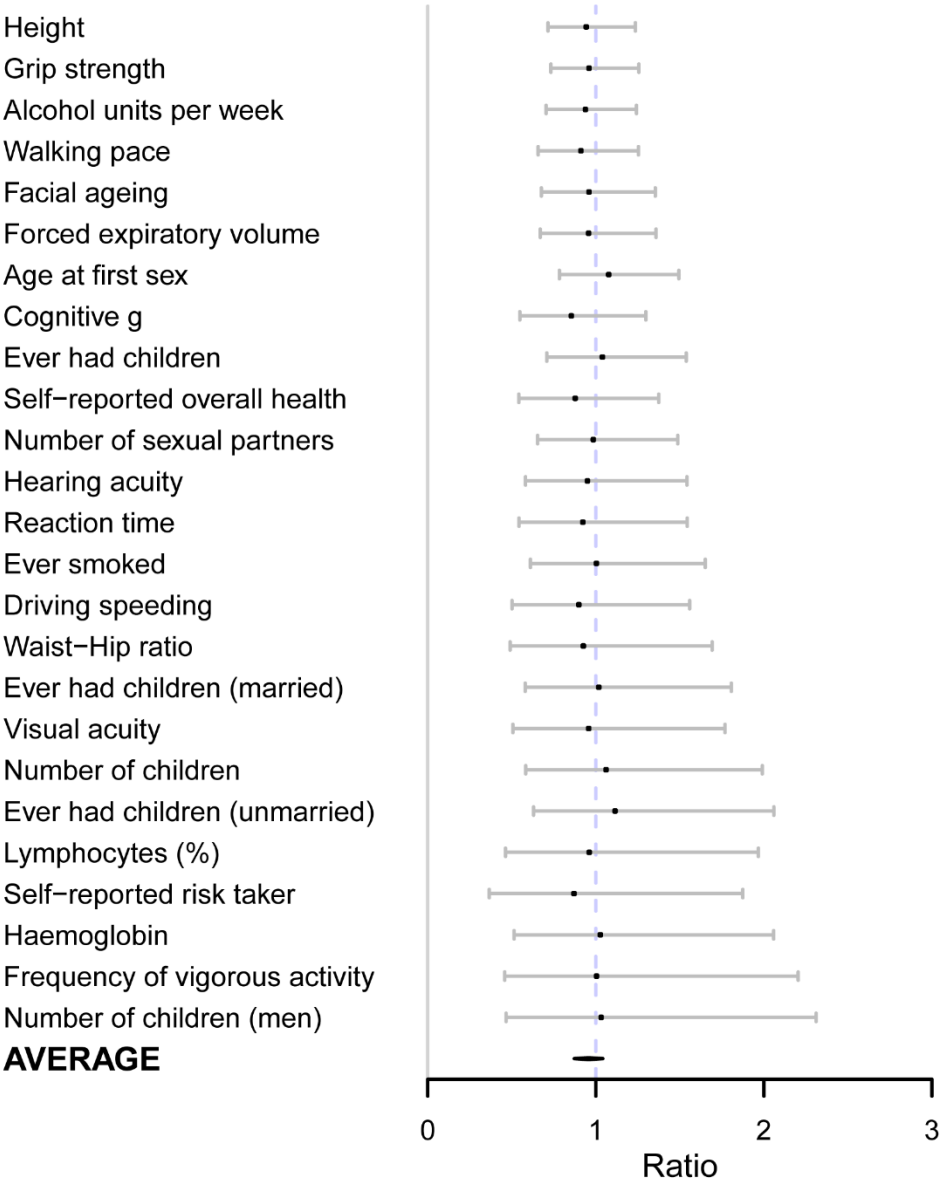

Self-reported overall health (index)
Walking pace (index)
Frequency of vigorous activity (days/week)
Facial ageing (index)
Visual acuity (negative logMAR)
Hearing acuity
Heart rate (bpm)
Waist-Hip ratio
F_ROH

9f

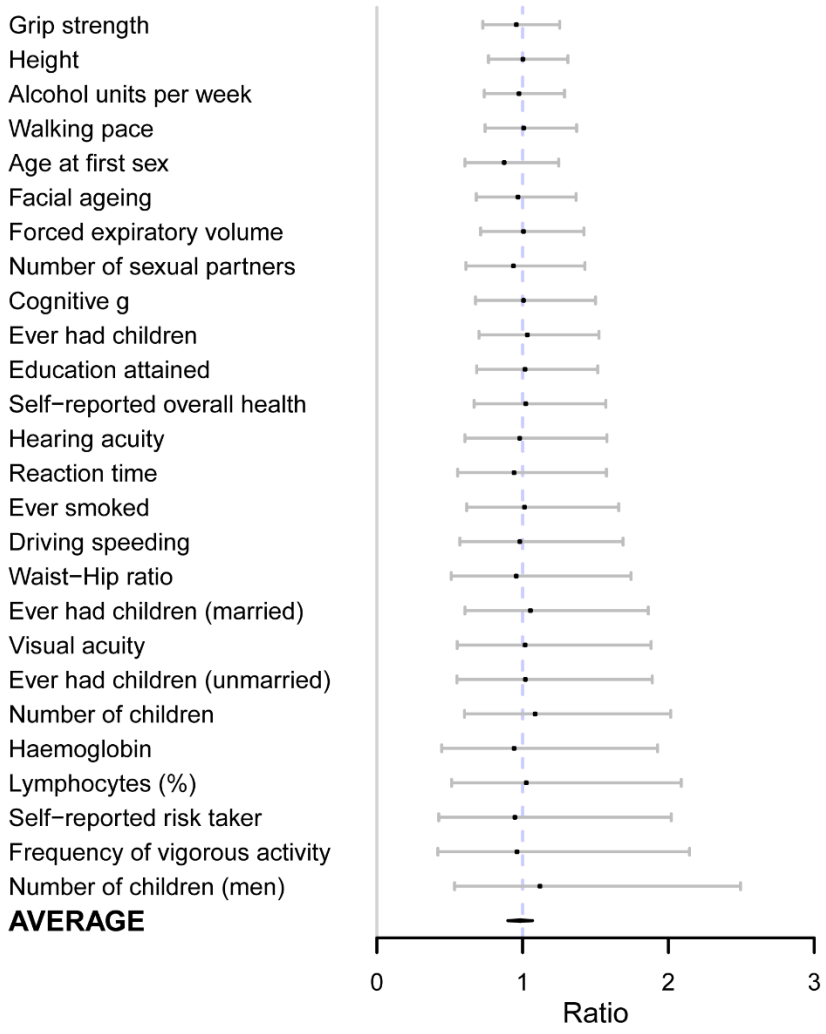


Supplementary Figure 9: Significant traits show a dosed response to increasing F_{ROH}. For all traits that reach experiment-wise significance in the meta-analysis, mean trait residuals are plotted in bins of increasing F_{ROH} (Methods) as also shown for height and Ever had children in Fig. 5a, b. Traits have been grouped into six categories (**a) Anthropometry, (b) Cognition, (c) Reproduction, (d) Risk-taking behaviour, (e) Well-being/Frailty, (f) Blood.** Although significant heterogeneity is observed for three traits (Height heterogeneity *p*-value = 7 x 10^{-8}, Educational Attainment heterogeneity *p*-value = 2 x 10^{-8}, Number ever born heterogeneity *p*-value = 7 x 10^{-5}) there is otherwise a dosed repose to increasing F_{ROH} for all traits. A dosed response across a wide range of F_{ROH} would be expected of a causal genetic effect, but not necessarily of environmental confounding. Although the effect of a confounder on a trait may be proportional, there is no a priori reason to expect a linear association between any confounder and F_{ROH}, especially extending to the large effects seen in very high F_{ROH} samples (F_{ROH}>0.18). All errors bars represent 95% confidence intervals.

10a

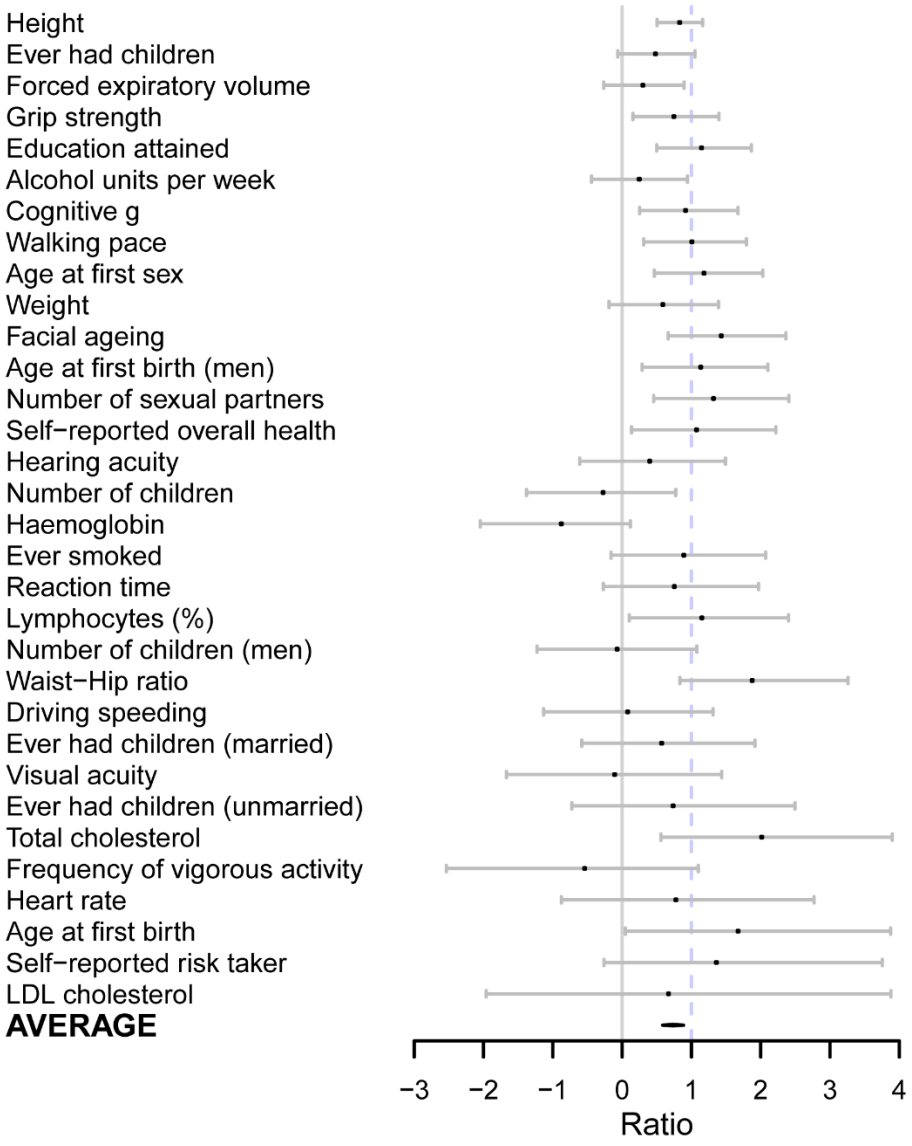

Height
Grip strength
Alcohol units per week
Walking pace
Facial ageing
Forced expiratory volume
Age at first sex
Cognitive g
Ever had children
Self−reported overall health
Number of sexual partners
Hearing acuity
Reaction time
Ever smoked
Driving speeding
Waist−Hip ratio
Ever had children (married)
Visual acuity
Number of children
Ever had children (unmarried)
Lymphocytes (%)
Self−reported risk taker
Haemoglobin
Frequency of vigorous activity
Number of children (men)
AVERAGE
0
1
2
3
Ratio

10b

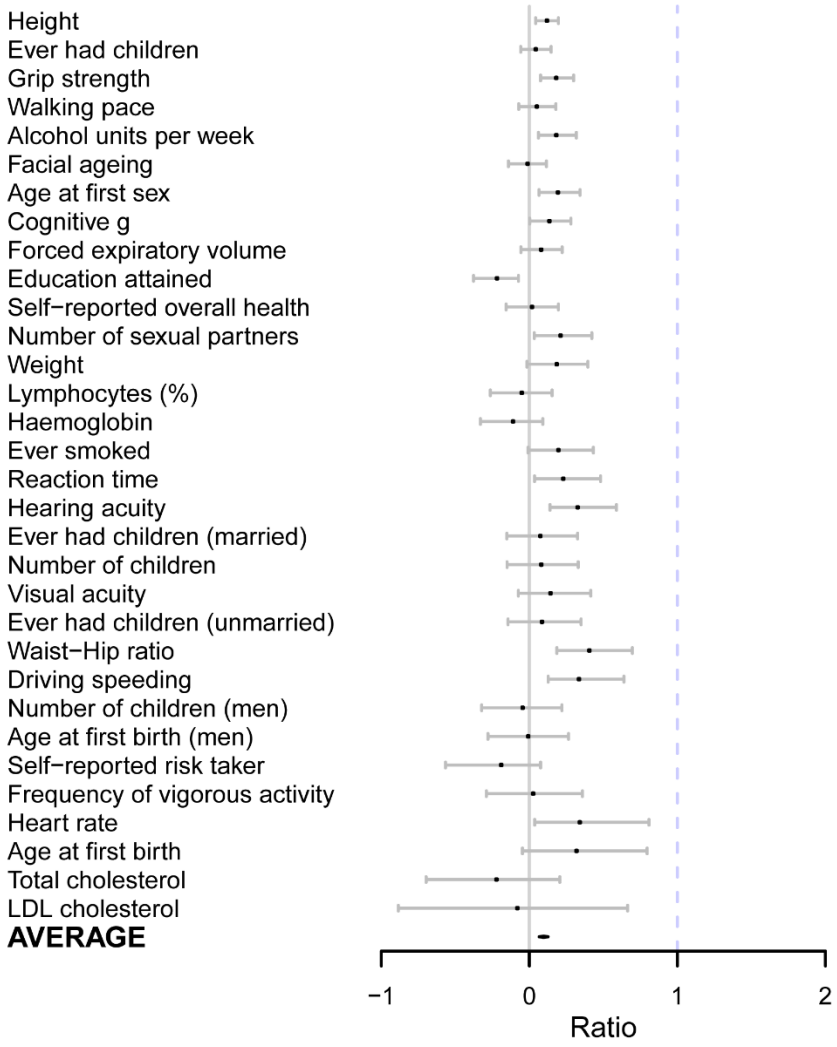


Supplementary Figure 10: Effect of fitting potential confounders as covariates. (a) Educational Attainment. For all traits that reach significance in UK Biobank, the ratio of effect size estimates with Educational Attainment (EA) fitted as an additional covariate ($\beta_{F_{\mathrm{ROH}}}^{+\mathrm{EA}}$) to the corresponding effect size estimates without EA ($\beta_{F_{\mathrm{ROH}}}$) are shown. The largest change is seen in Cognition (g) where fitting EA reduces $\beta_{F_{\mathrm{ROH}}}$ by 14.6%. However, since F_{ROH} is known to directly influence both g and EA, this change is not necessarily evidence of non-genetic effects. Overall fitting EA reduces the magnitude of $\beta_{F_{\mathrm{ROH}}}$ for 16 traits, but increases it for 9 traits, including number and likelihood of having children. **(b) Religious participation.** For the same traits (plus EA), the ratio of effect size estimates with a measure of religious participation (see Methods) fitted as an additional covariate ($\beta_{F_{\mathrm{ROH}}}^{+\mathrm{R}}$) to the corresponding effect size estimates without religious participation ($\beta_{F_{\mathrm{ROH}}}$) are shown. The largest reductions in $\beta_{F_{\mathrm{ROH}}}$ are seen for age at first sex (-12.7%) and number of sexual partners (-6.2%), suggesting that these traits may be partially confounded by social associations between F_{ROH} and religious beliefs. However, overall, fitting religious participation as a covariate increases $\beta_{F_{\mathrm{ROH}}}$ for 14 of 26 traits, again including number and likelihood of having children. All errors bars represent 95% confidence intervals.

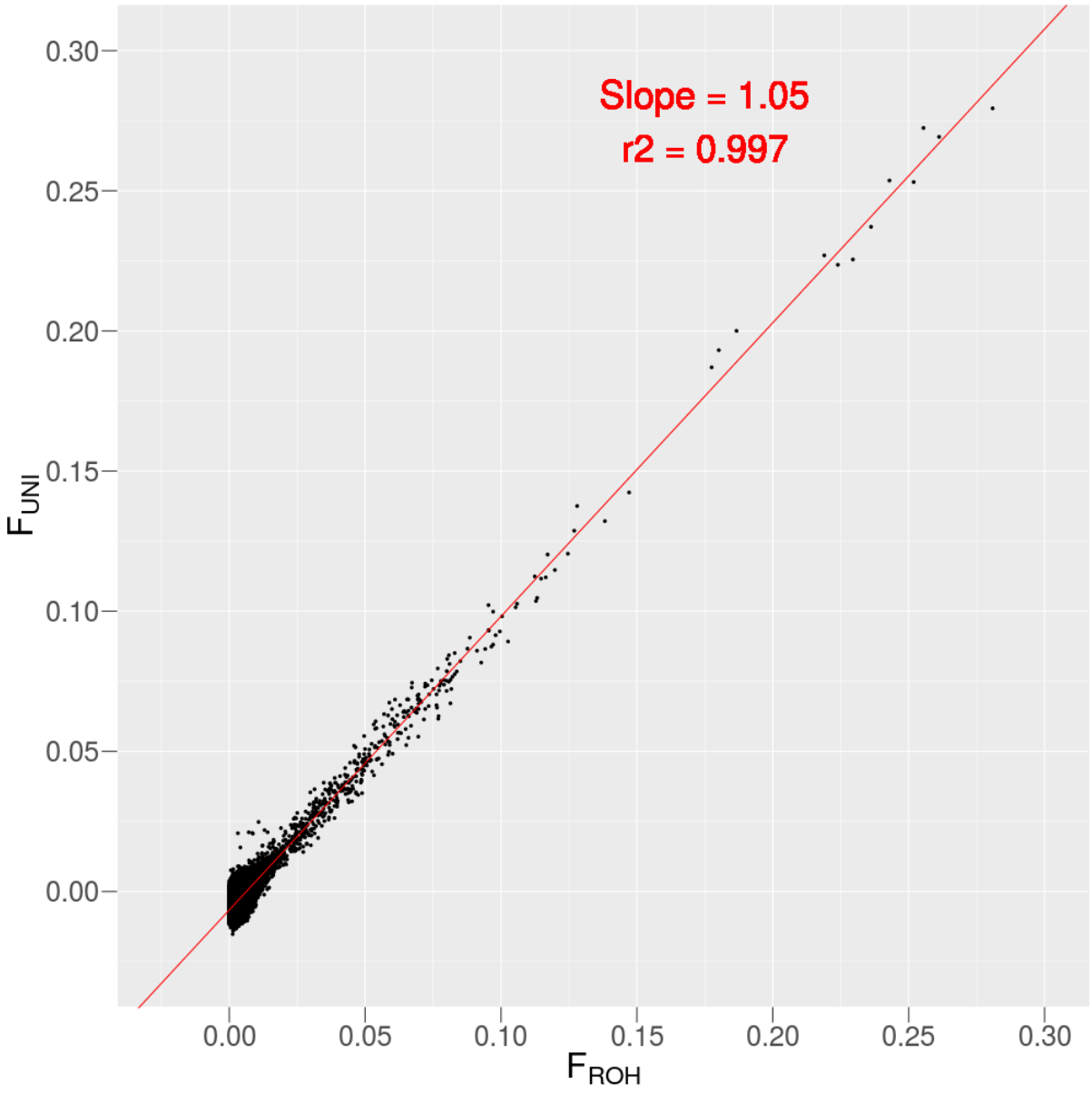
11a


Height
Ever had children
Forced expiratory volume
Grip strength
Education attained
Alcohol units per week
Cognitive g
Walking pace
Age at first sex
Weight
Facial ageing
Age at first birth (men)
Number of sexual partners
Self−reported overall health
Hearing acuity
Number of children
Haemoglobin
Ever smoked
Reaction time
Lymphocytes (%)
Number of children (men)
Waist−Hip ratio
Driving speeding
Ever had children (married)
Visual acuity
Ever had children (unmarried)
Total cholesterol
Frequency of vigorous activity
Heart rate
Age at first birth
Self−reported risk taker
LDL cholesterol
AVERAGE
−3
−2
−1
0
1
2
3
4
Ratio

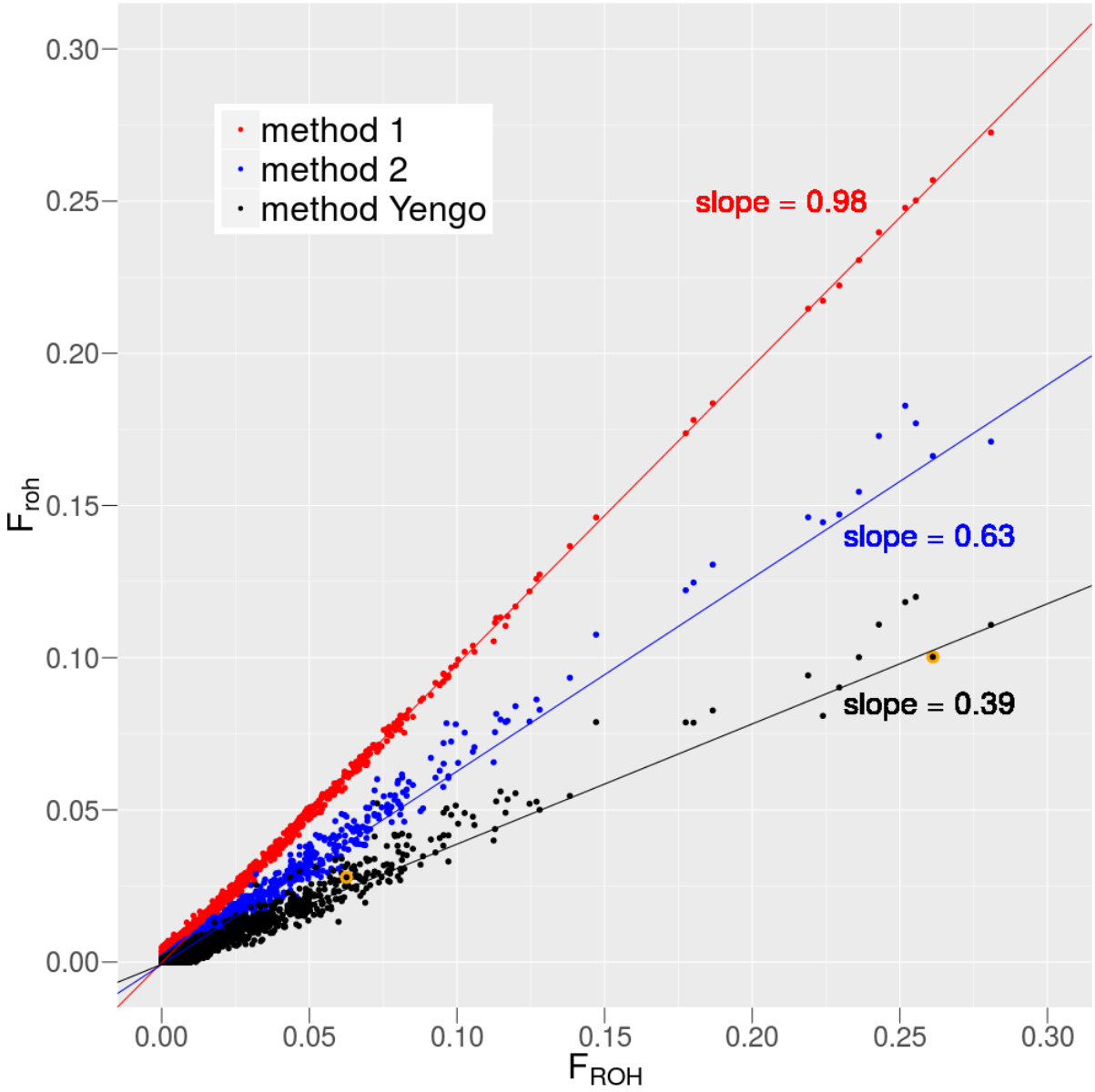
11b



Supplementary Figure 11: Conditional effects of ROH < 5Mb and SNP homozygosity outside ROH. Multivariate models were run for all traits including 3 different measures for homozygosity: $F_{\text{SNP\_OusideROH}}$, $F_{\text{ROH<5Mb}}$ and $F_{\text{ROH>5Mb}}$ (See Methods). **(a) Relative effect of ROH < 5Mb**. The conditional effect of $F_{\text{ROH<5Mb}}$ divided by $\beta_{F_{\text{ROH}}}$ is shown for all significant traits. Across all traits, the meta-analysed average of this ratio is 0.74 [95% CI 0.59-0.89, *p*-value 5 x 10^{-22}, heterogeneity *p*-value 0.132]. ROH of length less than 5 Mb are believed to be largely unconfounded by recent consanguinity, supporting the hypothesis that environmental confounding is responsible for only a small fraction (approximately 25%) of the reported effects. **(b) Relative effect of SNP homozygosity outside ROH.** The conditional effect of $F_{\text{SNP\_OusideROH}}$ divided by $\beta_{F_{\text{ROH}}}$ is shown for all significant traits. Although there is some heterogeneity, as might be expected from different trait architectures, for all traits the effect of $F_{\text{SNP\_OusideROH}}$ is significantly less than the effect of F_{ROH}. Averaging across all traits, the meta-analysed average of $\beta_{F_{\text{SNP\_OutsideROH}}}$: $\beta_{F_{\text{ROH}}}$ is 0.12 [95% CI 0.04-0.20, *p*-value 2 x 10^{-10}, heterogeneity *p*-value 0.001], showing that ROH have a larger effect on inbreeding depression on complex traits than does common SNP homozygosity outside ROH. All errors bars represent 95% confidence intervals.

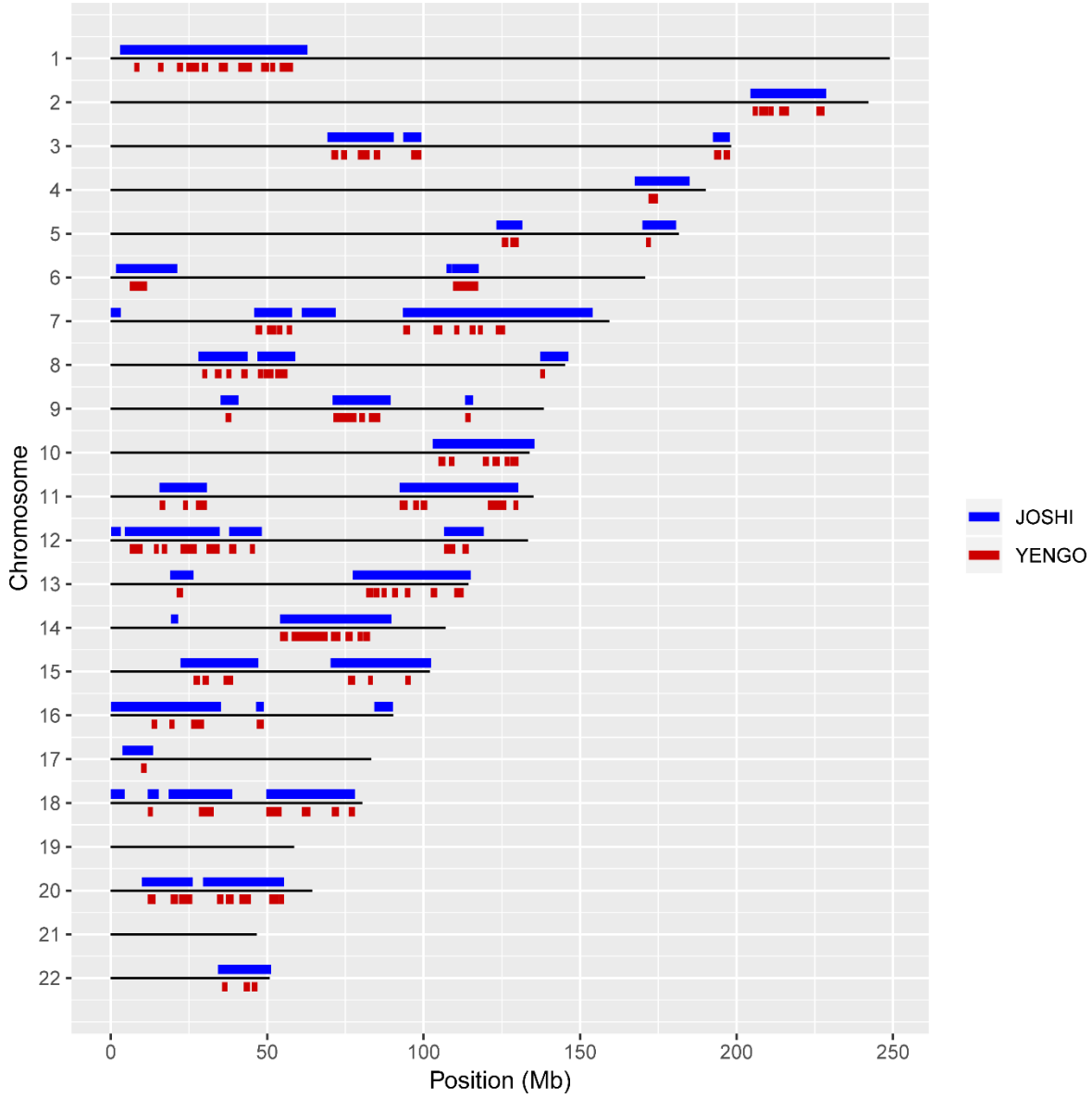


Supplementary Figure 12: Good correspondence between F_{UNI} and F_{ROH}. For 141,774 British samples in UK Biobank F_{UNI}, calculated from excess homozygosity of imputed genotypes, is plotted against F_{ROH}, calculated from SNP array genotypes. A weighted linear regression line is shown in red. Because average inbreeding coefficients are low ($\bar{F}_{\mathrm{ROH}} = 0.003$ in this population), it is high F individuals who contribute most of the statistical power to estimates of β_F. Weighting the regression by an estimate of power contribution $(F_i^{\mathrm{ROH}} - \bar{F}_{\mathrm{ROH}})^2$, we estimate $\beta_{F_{\mathrm{UNI}},F_{\mathrm{ROH}}} = 1.05$ and $r^2 = 0.997$. The good correspondence between F_{UNI} and F_{ROH} suggests both have minimal bias in estimating F.

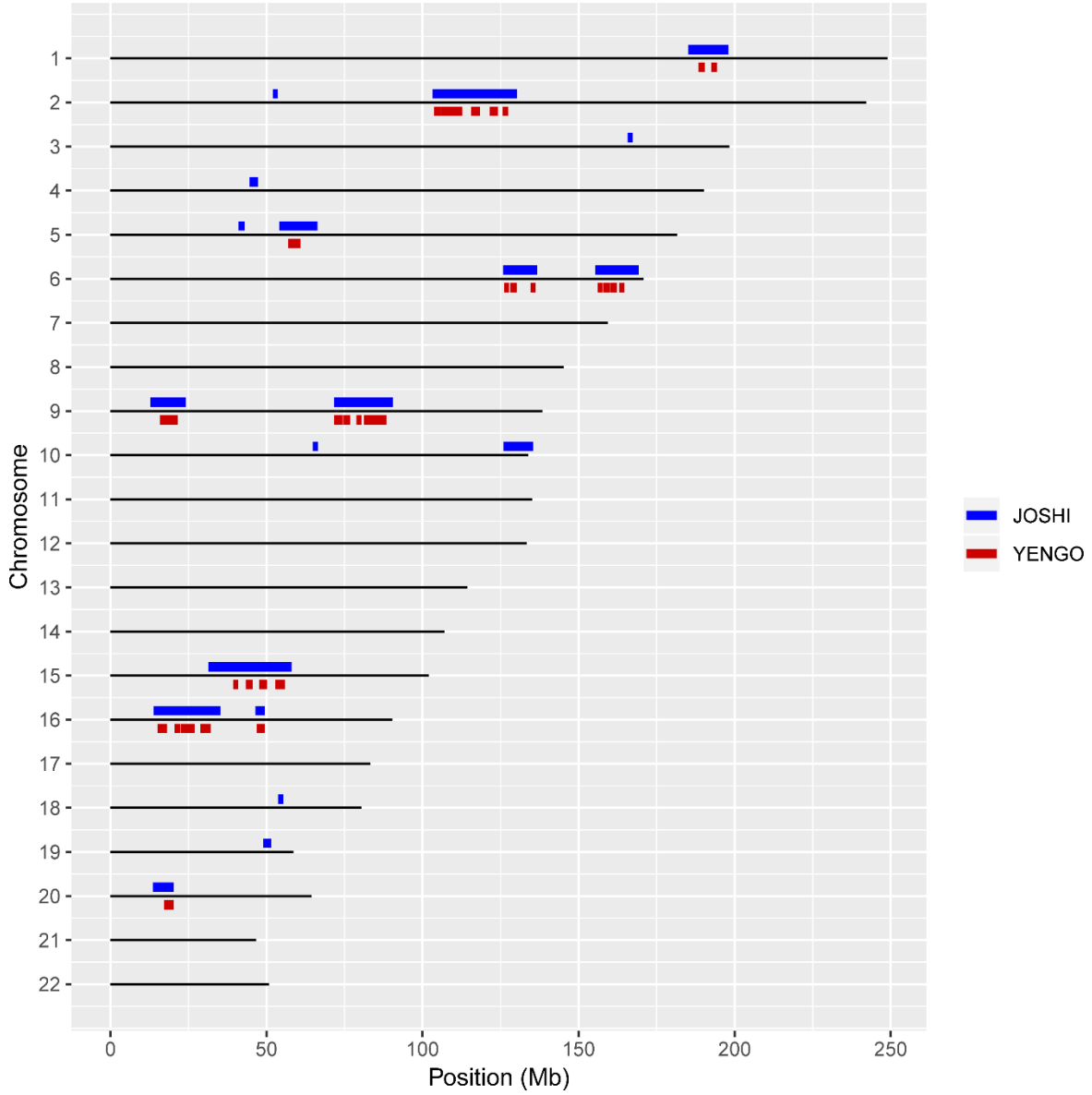


Supplementary Figure 13: Comparison of F_{roh} from imputed data and F_{ROH} from SNP array genotypes. For 141,774 British samples in UK Biobank, F_{roh} calculated from imputed genotypes is plotted against F_{ROH} calculated from SNP array genotypes for three methods of imputed genotype preparation. In method 1, in red, uncertain genotypes are removed. In method 2, in blue, uncertain genotypes are set to missing and in method Yengo, in black, no genotype filtering is performed. Increasingly permissive treatments of uncertain genotypes introduce increasing downward bias in F_{roh}. Two high F_{ROH} are highlighted in orange and further explored in Supplementary Figures 14a,b.

14a

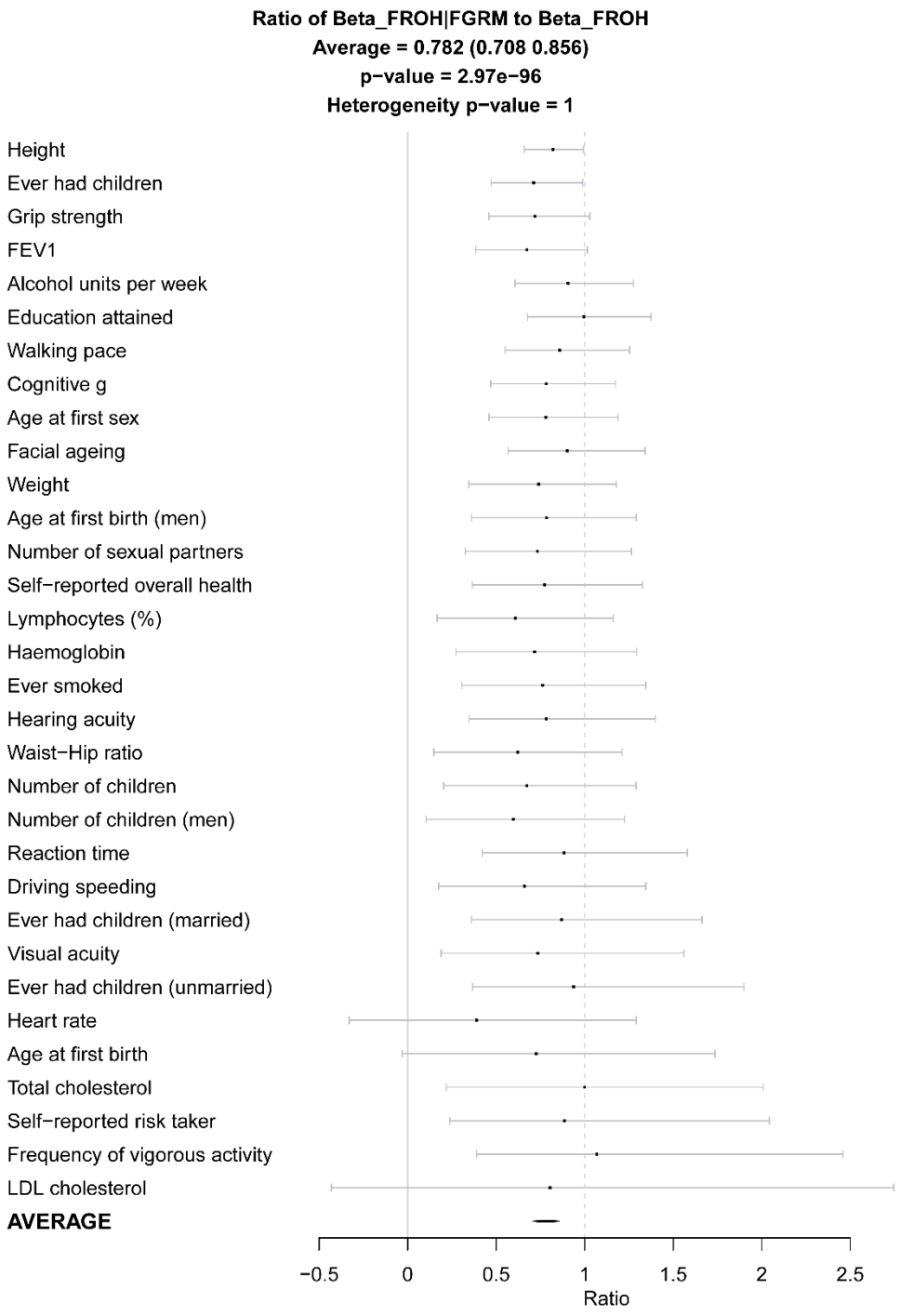

Chromosome
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
0
50
100
150
200
250
Position (Mb)
JOSHI
YENGO

14b

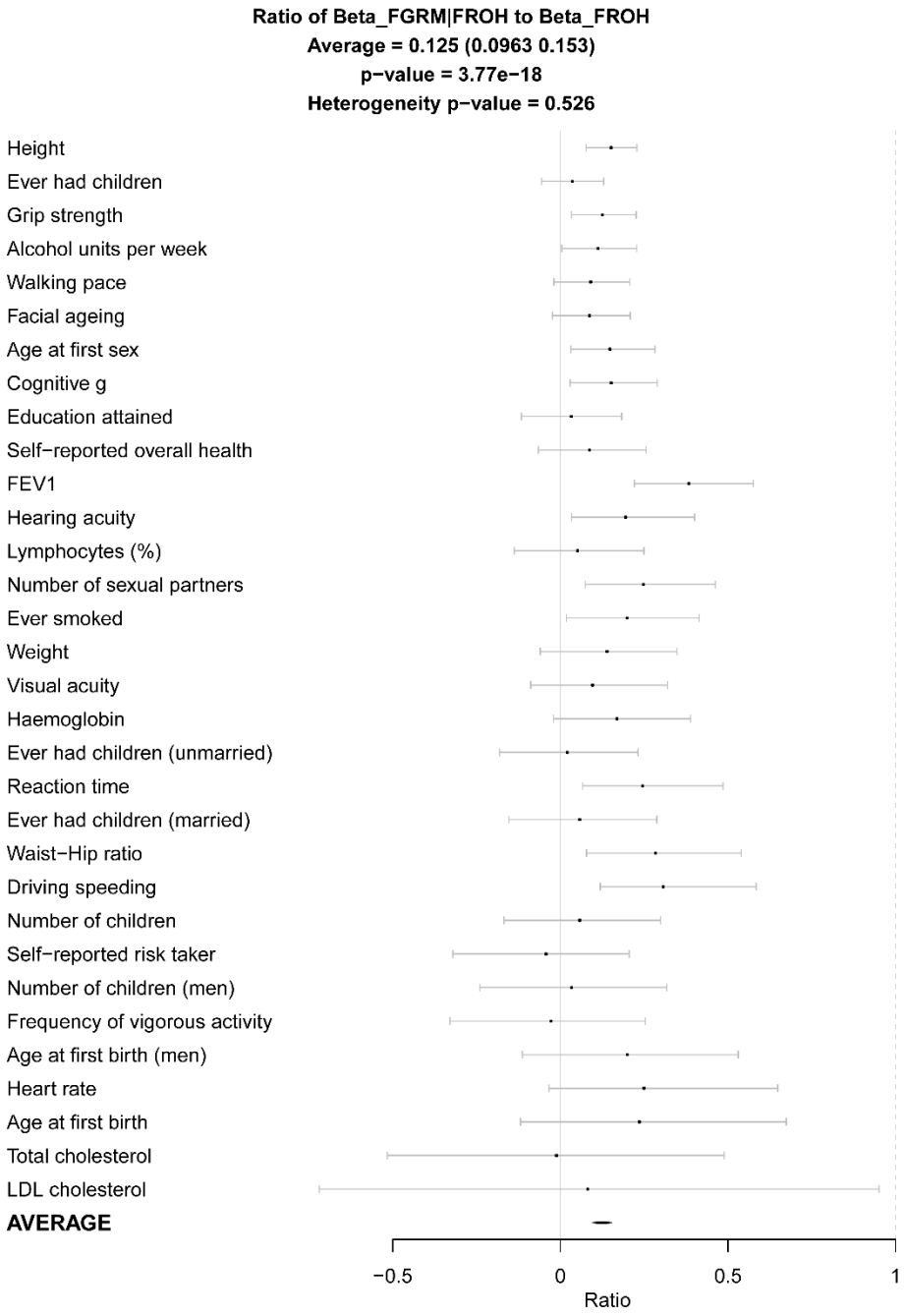


Supplementary Figure 14: Comparing ROH calling from SNP array genotypes and imputed dosages. For two high F_{ROH} individuals, the locations of called ROH are compared for two methods. The method shown in blue calls ROH from SNP array genotypes using the parameters used in Joshi et al. (2015). The method shown in red calls ROH from hard called imputed dosages following the method described in Yengo et al (2017). In both individuals the long ROH detected in SNP array genotypes, and which are thought to be autozygous segments, are broken up in the imputed data method by the presence of miscalled heterozygotes. **(a) Individual with $F_{\mathrm{ROH}} = 0.261$ thought to be offspring of first-degree relatives. (b) Individual with $F_{\mathrm{ROH}} = 0.0626$ thought to be offspring of third-degree relatives.**

15a

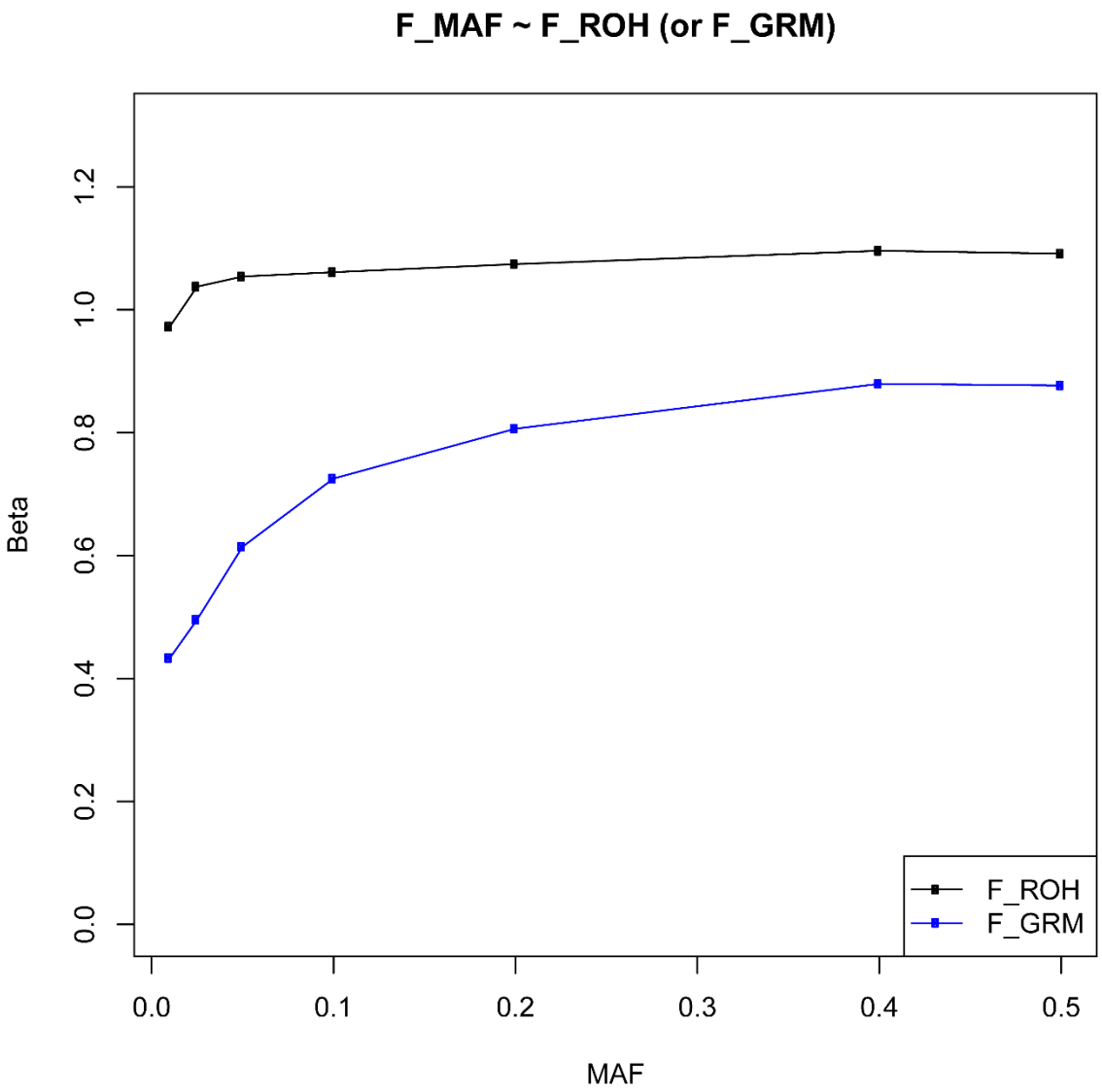

Ratio of Beta_FROH|FGRM to Beta_FROH
Average = 0.782 (0.708 0.856)
p−value = 2.97e−96
Heterogeneity p−value = 1
Height
Ever had children
Grip strength
FEV1
Alcohol units per week
Education attained
Walking pace
Cognitive g
Age at first sex
Facial ageing
Weight
Age at first birth (men)
Number of sexual partners
Self−reported overall health
Lymphocytes (%)
Haemoglobin
Ever smoked
Hearing acuity
Waist−Hip ratio
Number of children
Number of children (men)
Reaction time
Driving speeding
Ever had children (married)
Visual acuity
Ever had children (unmarried)
Heart rate
Age at first birth
Total cholesterol
Self−reported risk taker
Frequency of vigorous activity
LDL cholesterol
AVERAGE
−0.5
0
0.5
1
1.5
2
2.5
Ratio

15b

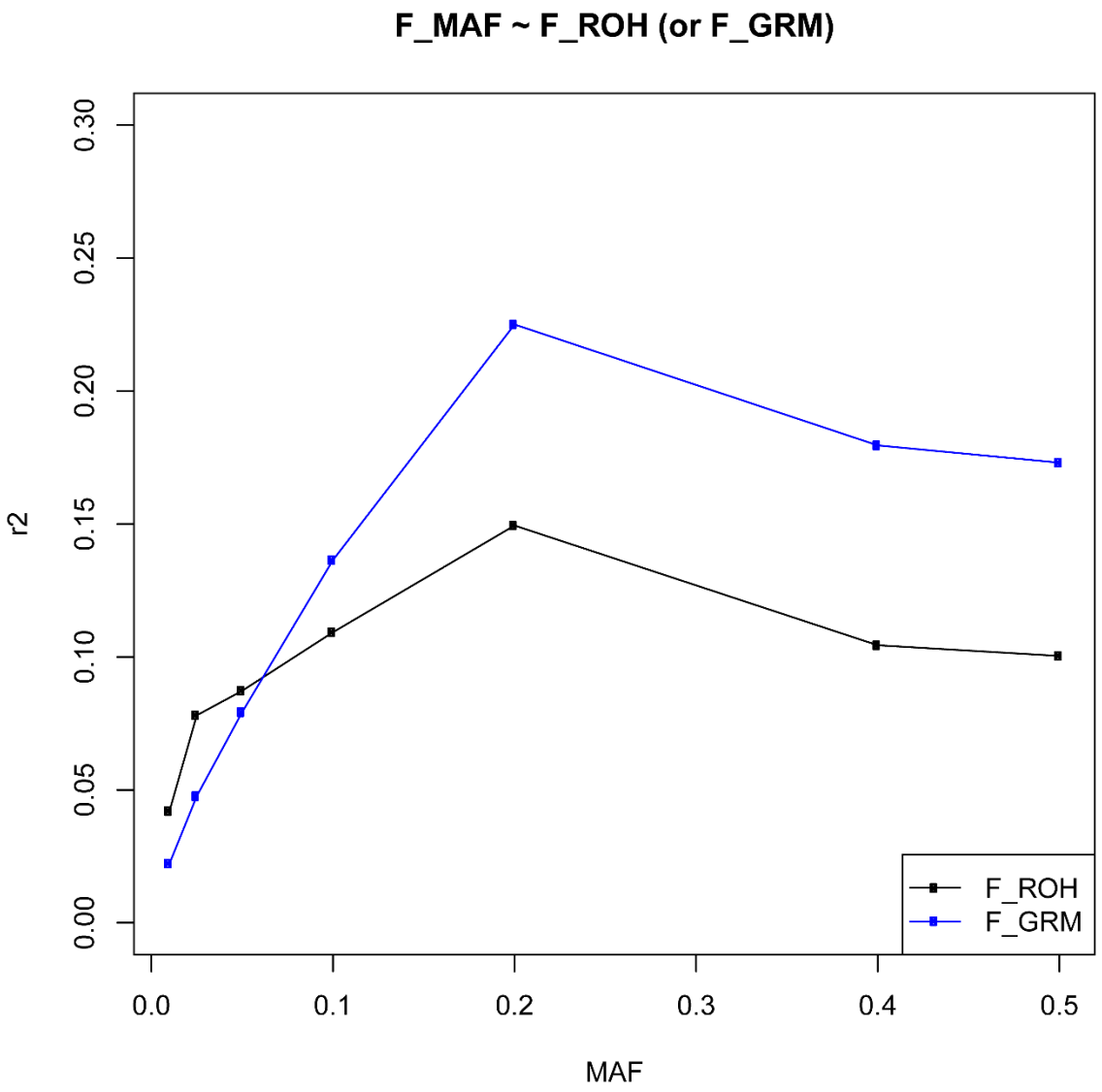


Supplementary Figure 15: Comparing effect estimates from bivariate models to the equivalent univariate estimates. For all significant traits, effect estimates were obtained from bivariate models of $Trait \sim F_{\mathrm{ROH}} + F_{\mathrm{GRM}}$ and compared to univariate estimates from the model $Trait \sim F_{\mathrm{ROH}}$. **(a) Ratio of $\boldsymbol{\beta}_{F_{\mathrm{ROH}}|F_{\mathrm{GRM}}}$ to $\boldsymbol{\beta}_{F_{\mathrm{ROH}}}$.** For all significant traits the ratio $\frac{\boldsymbol{\beta}_{F_{\mathrm{ROH}}|F_{\mathrm{GRM}}}}{\boldsymbol{\beta}_{F_{\mathrm{ROH}}}}$ is plotted. A meta-analysis across all traits gives an average ratio of 0.78 [95% CI 0.71-0.86]. **(b) Ratio of $\boldsymbol{\beta}_{F_{\mathrm{GRM}}|F_{\mathrm{ROH}}}$ to $\boldsymbol{\beta}_{F_{\mathrm{ROH}}}$.** For all significant traits the ratio $\frac{\boldsymbol{\beta}_{F_{\mathrm{GRM}}|F_{\mathrm{ROH}}}}{\boldsymbol{\beta}_{F_{\mathrm{ROH}}}}$ is plotted. A meta-analysis across all traits gives an average ratio of 0.12 [95% CI 0.10-0.15]. All errors bars represent 95% confidence intervals.

16a


F_MAF ~ F_ROH (or F_GRM)
Beta
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.0
0.1
0.2
0.3
0.4
0.5
MAF
F_ROH
F_GRM

16b



Supplementary Figure 16: Univariate relationships between estimates of inbreeding coefficient ($\boldsymbol{F_{ROH}}$, $\boldsymbol{F_{GRM}}$) and the excess homozygosity at specific allele frequencies. The excess homozygosity of SNPs at seven allele frequencies (F_{MAF}) was calculated for 402,559 genetically British samples in the phase 2 UKB imputation. **(a) Effect estimates of $\boldsymbol{F_{ROH}}$ and $\boldsymbol{F_{GRM}}$ on $\boldsymbol{F_{MAF}}$.** Univariate models of F_{MAF}~F_{ROH} and F_{MAF}~F_{GRM} were fitted at each allele frequency. A one unit increases in F_{ROH} is associated with a one unit increase in F_{MAF} across all allele frequencies. In contrast, the slope of $\beta_{F_{MAF},F_{GRM}}$ is downwardly biased at all allele frequencies. **(b) Correlations between of $\boldsymbol{F_{ROH}}$, $\boldsymbol{F_{GRM}}$ and $\boldsymbol{F_{MAF}}$.** Univariate models of F_{MAF}~F_{ROH} and F_{MAF}~F_{GRM} were fitted at each allele frequency. Despite the downward bias of its effect estimate, F_{GRM} is more strongly correlated with F_{MAF} at most allele frequencies.

SUPPLEMENTARY TABLES

Supplementary Table 1: Genetic correlations between risk and reproductive traits. Genetic correlations estimated in UKB by LD score regression and their corresponding *p*-values. 5 reproductive traits and 4 risk traits are shown. The sign of age at first sex has been reversed so that larger trait values are associated with higher reproductive output. Positive correlations are shown in blue, with darker shades signifying Bonferroni-corrected significance. Unsurprisingly, positive genetic correlations are found within the groups of risk and reproductive success but, perhaps more unexpectedly, the genetic correlations between risk and reproductive traits are also most often positive. This is particularly true for smoking and self-declared risk taking.

		Ever had children	Number of children	***Earlier*** age at first sex	Number of sexual partners	Alcohol units	Ever smoked	Driving speed
Number of children	Rg	0.93						
	P	0						
Earlier age at first sex	Rg	0.57	0.53					
	P	3E-175	2E-200					
Number of sexual partners	Rg	0.08	0.10	0.52				
	P	1E-02	1E-03	2E-168				
Alcohol units	Rg	0.15	0.11	0.14	0.37			
	P	5E-02	1E-01	1E-02	7E-07			
Ever smoked	Rg	0.27	0.28	0.60	0.49	0.37		
	P	4E-26	2E-30	0	1E-119	3E-08		
Driving speed	Rg	0.00	0.03	0.01	0.29	0.2192	0.14	
	P	1	0.40	0.80	8E-24	2E-03	3E-07	
Self-declared risk taking	Rg	0.23	0.27	0.42	0.57	0.28	0.33	0.39
	P	5E-11	2E-16	9E-77	2E-137	3E-04	5E-46	3E-41

Supplementary Table 2: Number of SNPs extracted from UKB imputation by allele frequency. The excess homozygosity of SNPs at seven allele frequencies (F_{MAF}) was calculated for 402,559 genetically British samples in the phase 2 UKB imputation. The number of SNPs used at each frequency is shown.

MAF	Number of SNPs used in calculation of F_MAF
0.01	84835
0.025	122498
0.05	198310
0.1	261504
0.2	369777
0.4	253826
0.5	301191

SUPPLEMENTARY NOTE 1: Trait descriptions

45 quantitative traits were initially chosen for analysis in a potentially wide range of cohorts from the ROHgen consortium. During the initial meta-analysis of these traits, the full release of >500,000 samples from UK Biobank (UKB) became available and, it was decided to include a further 55 less-commonly measured traits available in UKB. Of these new traits, 21 were binary, requiring an extension to the existing analysis plan. 7 of the UKB traits were also measured in some ROHgen cohorts and were thus analysed in a subset of ROHgen cohorts willing to rerun the new analysis plan. In summary, a total of 100 complex traits were analysed; 45 in a potentially wide range of ROHgen cohorts, 7 in a subset of ROHgen cohorts and 48 in UKB only. All are defined below, under headings in the format **short name – full name – units**.

afb - Age at first birth – years. Age of subject (either male or female) when their first child was born. Nulliparous samples and reported ages less than 12 or greater than 80 were excluded.

afb_men - Age at first birth (men) – years. Men only. Unlike most other traits, age at first birth was treated as a separate trait for men and women, and the full set of analyses was therefore performed on both sexes separately. Age at first birth (men) is the age of a male subject when their first child was born. Nulliparous samples and reported ages less than 12 or greater than 80 were excluded.

afb_women - Age at first birth (women) – years. Women only. Unlike most other traits, age at first birth was treated as a separate trait for men and women, and the full set of analyses was therefore performed on both sexes separately. Age at first birth (women) is the age of a female subject when their first child was born. Nulliparous samples and reported ages less than 12 or greater than 60 were excluded.

age_menarche - Age at menarche – years. Women only. Reported age at menarche. Women with age at menarche less than 5 or greater than 25 were excluded.

age_menopause - age at menopause (years). Women only. Age at natural menopause. Women whose menopause was due to surgical operations (hysterectomy/ovariectomy), cancer treatment

(radiation, chemotherapy) or on HRT before menopause were excluded. Responses below 35 or greater than 70 were also excluded.

birth_weight – Birth weight – kg. Individual's own weight at birth. Participants who were known to be part of a multiple birth (twins, triplets, etc.) were set to NA. Values less than 0.5 kg or greater than 7 kg were excluded.

bmi – Body mass index – kgm^{-2}. Weight in kilograms divided by height in metres squared. Values less than 10 or greater than 150 were excluded.

dp_dia – Diastolic blood pressure – mmHg. Averaged readings taken during a single session. Guidance was to take the unweighted mean of second and third readings although cohorts were given discretion to use best judgement where appropriate. Participants known to be on hypertension medication had 10mmHg added to their readings. Values less than 20 or greater than 200 were excluded.

bp_sys – Systolic blood pressure – mmHg. Averaged readings taken during a single session. Guidance was to take the unweighted mean of second and third readings although cohorts were given discretion to use best judgement where appropriate. Participants known to be on hypertension medication had 15mmHg added to their readings. Values less than 50 or greater than 300 were excluded.

edu – Education Attained – years. Based on SSGAC, Education Attained was defined in accordance with the ISCED 1997 classification(UNESCO), relating to seven categories of educational attainment that are internationally comparable. Subjects age <30 were excluded as were values $\notin$ {1,7,10,13,15,19,22}.

Definition	US years of schooling
Pre-primary education	1
Primary education or first stage of basic education	7
Lower secondary or second stage of basic education	10
(Upper) secondary education	13
Post-secondary non-tertiary education	15
First stage of tertiary education (not leading directly to an advanced research qualification)	19
Second stage of tertiary education (leading directly to an advanced research qualification e.g. PhD)	22

fev1 – Forced expiratory volume in 1 second – Litres. Where multiple blows were available the maximum valid reading was used. Values less than 0 or greater than 10 were excluded.

fev1perfvc – Forced expiratory volume in 1 second / forced vital capacity – no units. Values less than 0 or greater than 15 were excluded.

fpg – Fasting plasma glucose – $mmolL^{-1}$. Known diabetic subjects were excluded, as were subjects with fpg > 7 or hba1c > 6.5. Measurements made in whole blood (not plasma) were multiplied by 1.13 to estimate fpg.

g – Cognitive g – z-score. The first unrotated principal component of three or more tests of different domains of cognition. Care was taken to ensure this was in the direction of larger values being associated with greater cognition. Specifically, the sign of the correlation between Cognitive g and

Education attained was ensured to be positive in all cohorts. This trait was rank-normalised and values less than -10 or greater than 8 were excluded.

hb – Haemoglobin – gL^{-1}. Concentration of haemoglobin. Values less than 0 or greater than 500 were set to NA.

hba1c – Glycosylated haemoglobin – % of hb (DCCT)[2]. Set to NA for known diabetics and all subjects for whom HbA1c > 6.5 or fpg > 7. Also, set to NA for subjects with known major blood abnormalities (thalassaemia, sickle cell anaemia, etc.), subjects who have had a blood transfusion in the previous 3 months.

hdl – High-density lipoprotein cholesterol – $mmolL^{-1}$. Taken only from fasted or semi-fasted subjects. If semi-fasted a covariate specifying fasting time was required. Values greater than 5.17 were set to NA.

height – Height – meters. Standing height in meters. Values less than 1.2 or greater than 2.5 were set to NA.

hr – Heart rate – beats per minute. Participants on cardiac medications (Beta blockers, antiarrhythmics) were exclude as were those with previous myocardial infarction or heart failure. Values less than 20 or greater than 150 were set to NA.

ldl – Low-density lipoprotein cholesterol – $mmolL^{-1}$. Taken only from fasted or semi-fasted subjects. If semi-fasted a covariate specifying fasting time was required. If HDL cholesterol, total cholesterol and Triglycerides were all provided, LDL cholesterol was calculated using Friedewald's equation. Alternatively, LDL cholesterol could be supplied if directly measured. Samples known to be on lipid lowering medication were adjusted by dividing by a factor of 0.7. Values less than 0 or greater than 10.34 were set to NA.

log.egfr – Estimated glomerular filtration rate – $mLmin^{-1}1.73m^{-2}$. Glomerular filtration rate was estimated from measured creatinine (in $mgdL^{-1}$) using the formula $186*creatinine^{-1.154}*age^{-0.203}$. In cohorts with African or African-American ancestry these values were multiplied by a correction factor of 1.21. Values of creatinine greater than 20 or eGFR greater than 200 were set to NA.

log.fast_ins – Fasting insulin – $pmolL^{-1}$. Known diabetic samples, as well as samples with fpg > 7 or HbA1c < 6.5 were excluded. Values of fasting insulin greater than 1000 were set to NA.

log.fibrinogen – Fibrinogen – gL^{-1}. Plasma fibrinogen levels. Values greater than 20 were set to NA.

log.hscrp – high sensitivity C-reactive protein – $nmolL^{-1}$. Serum levels of C-reactive protein (CRP) detected with high sensitivity systems (lower detection limit around 1 $nmolL^{-1}$). Samples known to be on anti-inflammatory drugs (ATC codes L01, L03, L04, L02A, L02B) were set to NA, as were values greater than 952 $nmolL^{-1}$.

log.il6 – Interleukin-6 – $pgmL^{-1}$. Serum levels of Interleukin-6. Samples known to be on anti-inflammatory drugs (ATC codes L01, L03, L04, L02A, L02B) were set to NA, as were values greater than 100 $pgmL^{-1}$.

log.lymphoc – Lymphocytes – %. Percentage of lymphocytes per white blood cell count. Values greater than 100 were set to NA.

log.mpv – Mean platelet volume – fL. Mean platelet volume in femtolitres. Values greater than 30 were set to NA.

log.tnfa – Tumour necrosis factor alpha – pgmL^{-1}. Samples known to be on anti-inflammatory drugs (ATC codes L01, L03, L04, L02A, L02B) were set to NA, as were values greater than 100 pgmL^{-1}.

log.triglyc – Triglycerides – mmolL^{-1}. Taken only from fasted or semi-fasted subjects. If semi-fasted a covariate specifying fasting time was required. Values of Triglycerides greater than 33.9 mmolL^{-1} were set to NA.

log.wbc – White blood cell count – 10^9 per Litre. Values greater than 30 x 10^9 per Litre were excluded.

log10.alt – Alanine transaminase – IUper Litre. Plasma concentrations of Alanine transaminase (also called Glutamic-pyruvate transaminase). Values greater than 500 IU per Litre were set to NA.

log10.ggt – Gamma-Glutamyl Transferase – IU per Litre. Plasma concentrations of Gamma-Glutamyl Transferase. Values greater than 1000 IU per Litre were set to NA.

monoc – Monocytes - %. Percentage of monocytes in white blood cell count. Values greater than 40 were set to NA.

neb – Number ever born – count. Number of children the subject has brought into being. Subjects aged less than 45 were excluded from this analysis.

neb_men – Number ever born (men) – count. Men only. Unlike most other traits, number ever born was treated as a separate trait for men and women, and the full set of analyses was therefore performed on both sexes separately. Number ever born (men) is the number of children fathered by a male subject. Subjects aged less than 45 were excluded from this analysis.

neb_women – Number ever born (women) – count. Women only. Unlike most other traits, number ever born was treated as a separate trait for men and women, and the full set of analyses was therefore performed on both sexes separately. Number ever born (women) is the number of children given birth to by a female subject. Subjects aged less than 45 were excluded from this analysis.

plt – Platelet count - 10^9 per Litre. Platelet count in whole blood. Values less than 20 or greater than 1000 were set to NA.

pr – PR interval – ms. Electrocardiographic PR interval. Participants on cardiac medications (Beta blockers, antiarrhythmics) were exclude as were those with previous myocardial infarction or heart failure. Values less than 80 or greater than 320 were set to NA.

qrs – QRS duration – ms. Electrocardiographic QRS duration. Participants on cardiac medications (Beta blockers, antiarrhythmics) were exclude as were those with previous myocardial infarction or heart failure. Values less than 30 or greater than 120 were set to NA.

qt – QT interval – ms. Electrocardiographic QT interval. Participants on cardiac medications (Beta blockers, antiarrhythmics) were exclude as were those with previous myocardial infarction or heart failure. Values less than 200 or greater than 700 were set to NA.

ser – Spherical equivalent refraction – no units. The mean of left and right eyes calculated from spherical and cylindrical power of each eye by the standard formula ser = sphere + 0.5*cylinder. Samples known to have had eye surgery were set to NA, as were values less than -15 or greater than +15.

tot_chol – Total cholesterol – mmolL^{-1}. Taken only from fasted or semi-fasted subjects. If semi-fasted a covariate specifying fasting time was required. Samples known to be on lipid lowering medication were adjusted by dividing by a factor of 0.8. Values greater than 16.8 molL^{-1} were set to NA.

uric – Uric acid – umolL^{-1}. Serum urate concentration. Values greater than 1190 umolL^{-1} were set to NA.

weight – Weight – kg. Weight in kilograms. Values less than 20 kg or greater than 250 kg were set to NA.

whr – Waist : Hip ratio – no units. Calculated when both waist and hip circumference were available in centimetres. Values of waist or hip circumference less than 20 or greater than 300 were set to NA, as were values of waist : hip ratio less than 0.3 or greater than 2.

alcohol_units - Alcohol units per week – UK units per week. Self-declared alcohol consumption in UK units (10 ml of ethanol) per week. Where necessary this was estimated from alcohol intake frequency and average drink sizes. Values were capped to 100 units per week.

ever_married_glm – Ever married – TRUE/FALSE. Subjects who were known to be (or have been) married or in a long-term cohabiting relationship were encoded as 1 while all others were encoded as 0.

ever_smoked_glm – Ever smoked – TRUE/FALSE. Subjects who reported either being current smokers or having previously smoked on all or most days were encoded as 1, while those who had never or only occasionally smoked were encoded as 0.

neb_parous – Number ever born (parous) – count. In samples with one or more children, number ever born (parous) is equal to number of children ever born (neb). All other samples are set to NA.

parous_glm – Ever had children – TRUE/FALSE. For all samples with a non-missing value of number ever born (neb) a trait was defined with value 1 for samples with neb>0 and value 0 for samples with neb=0.

parous_married_glm – Ever had children (married) – TRUE/FALSE. Ever had children defined only for samples for whom ever married = 1. Set to NA for all samples where ever married is 0 or NA.

parous_unmarried_glm – Ever had children (unmarried) – TRUE/FALSE. Ever had children defined only for samples for whom ever married = 0. Set to NA for all samples where ever married is 1 or NA.

age_at_first_sex – Age at first sex – years. Response to the question *What was your age when you first had sexual intercourse?* Participants who declined to answer, or who gave an answer less than 3 were excluded. Participants who answered *Never had sex* were set to their current age. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2139.

age_facial_hair – Relative age of facial hair – index. Men only. Response to the question *When did you start to grow facial hair?*. Participants were given five options: *Younger than average*, *About average age*, *Older than average*, *Do not know* and *Prefer not to answer* which were encoded -1, 0, 1, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2375.

age_voice_broke – Relative age voice broke – index. Men only. Response to the question *When did your voice break?*. Participants were given five options: *Younger than average*, *About average age*,

Older than average, *Do not know* and *Prefer not to answer* which were encoded -1, 0, 1, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2385.

alb – Age at last birth – years. Women only. Response to the question *How old were you when you had your LAST child?*. Responses less than 8 or greater than 65 were excluded. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2764.

ankle_width – Mean ankle width – mm. Average of left and right ankle width as measured by the spacing between measurement transducer pads on each heel. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4100.

any_pain_glm – Any reported pain – TRUE/FALSE. Participants were asked the question *In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer)* and given ten options: *Headache*, *Facial pain*, *Neck or shoulder pain*, *Back pain*, *Stomach or abdominal pain*, *Hip pain*, *Knee pain*, *Pain all over the body*, *None of the above* and *Prefer not to answer*. Participants who selected any of the first eight options were coded as 1, those who selected only *None of the above* were coded as 0 and the remainder were treated as NA. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6159.

any_same_sex_glm – Any same-sex partners – TRUE/FALSE. Participants were asked the question *Have you ever had sexual intercourse with someone of the same sex?* and given the options *Yes*, *No* and *Prefer not to answer* which were encoded 1, 0 and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/coding.cgi?id=100352.

back_pain_glm – Backpain – TRUE/FALSE. Participants were asked the question *In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer)* and given ten options: *Headache*, *Facial pain*, *Neck or shoulder pain*, *Back pain*, *Stomach or abdominal pain*, *Hip pain*, *Knee pain*, *Pain all over the body*, *None of the above* and *Prefer not to answer*. Participants who selected *Back pain* were coded as 1, those who selected only *Prefer not to answer* were coded as NA and the remainder set to 0. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6159.

baldness – Baldness pattern – index. Men only. Male participants were asked the question *Which of the following best describes your hair/balding pattern?* and shown four images of increasing hair loss (patterns 1 to 4). Responses were coded 1 to 4, where 4 represents most hair loss. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2395.

body_pain_glm – Whole body pain – TRUE/FALSE. Participants were asked the question *In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer)* and given ten options: *Headache*, *Facial pain*, *Neck or shoulder pain*, *Back pain*, *Stomach or abdominal pain*, *Hip pain*, *Knee pain*, *Pain all over the body*, *None of the above* and *Prefer not to answer*. Participants who selected *Pain all over the body* were coded as 1, those who selected only *Prefer not to answer* were coded as NA and the remainder set to 0. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6159.

broken_bones_glm – Broken bones – TRUE/FALSE. Participants were asked the question *Have you fractured/broken any bones in the last 5 years?* and given the options *Yes*, *No*, *Do not know* and *Prefer not to answer* which were encoded 1, 0, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2463.

cancer_diagnosis_glm – Cancer diagnosis – TRUE/FALSE. Participants were asked the question *Has a doctor ever told you that you have had cancer?* and given the options *Yes*, *No*, *Do not know* and

Prefer not to answer which were encoded 1, 0, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2453.

dead_glm – Dead – TRUE/FALSE. Death records in UKB are periodically updated by linkage to national death registries. At data download on 13/12/2017, 13739 participants had record dates of death and were thus encoded at 1. Those without a death register entry were encoded as 0. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=40000.

depression_glm – Self-reported mood disorder – TRUE/FALSE. Participants were asked the question *Have you ever seen a general practitioner (GP) for nerves, anxiety, tension or depression?* and given the options *Yes*, *No*, *Do not know* and *Prefer not to answer* which were encoded 1, 0, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2090.

diabetes_diagnosis_glm – Diabetes diagnosis – TRUE/FALSE. Participants were asked the question *Has a doctor ever told you that you have diabetes?* and given the options *Yes*, *No*, *Do not know* and *Prefer not to answer* which were encoded 1, 0, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2443.

facial_ageing – Facial ageing – index. Participants were asked the question *Do people say that you look:* and given the options *Younger than you are*, *Older than you are*, *About your age*, *Do not know* and *Prefer not to answer* which were encoded -1, 1, 0, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=1757.

family_satisfaction – Family satisfaction – index. Participants were asked the question *In general how satisfied are you with your family relationships?* and given the options *Extremely happy*, *Very happy*, *Moderately happy*, *Moderately unhappy*, *Very unhappy*, *Extremely unhappy*, *Do not know* and P*refer not to answer* which were encoded 6, 5, 4, 3, 2, 1, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4559.

fat_pc – Body fat percentage – %. Body composition estimated by impedance measurement. Values less than 1% or greater than 75% were excluded. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=23099.

financial_satisfaction – Financial satisfaction – index. Participants were asked the question *In general how satisfied are you with your financial situation?* and given the options *Extremely happy*, *Very happy*, *Moderately happy*, *Moderately unhappy*, *Very unhappy*, *Extremely unhappy*, *Do not know* and *Prefer not to answer* which were encoded 6, 5, 4, 3, 2, 1, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4581.

grip_strength – Grip strength – kg. Average of left and right hand grip strength as measured by a hydraulic hand dynamometer. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=46.

handedness – Left-handed – index. Participants were asked the question *Are you right or left handed?* and given the options *Right-handed*, *Left-handed*, *Use both right and left hands equally* and *Prefer not to answer* which were encoded -1, 1, 0 and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=1707.

happiness – Self-reported happiness – index. Participants were asked the question *In general how happy are you?* and given the options *Extremely happy*, *Very happy*, *Moderately happy*, *Moderately unhappy*, *Very unhappy*, *Extremely unhappy*, *Do not know* and *Prefer not to answer* which were

encoded 6, 5, 4, 3, 2, 1, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20458.

headache – Headaches – TRUE/FALSE. Participants were asked the question *In the last month have you experienced any of the following that interfered with your usual activities? (You can select more than one answer)* and given ten options: *Headache*, *Facial pain*, *Neck or shoulder pain*, *Back pain*, Stom*ach or abdominal pain*, *Hip pain*, *Knee pain*, *Pain all over the body*, *None of the above* and *Prefer not to answer*. Participants who selected *Headache* were coded as 1, those who selected only *Prefer not to answer* were coded as NA and the remainder set to 0. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=6159.

health_satisfaction – Health satisfaction – index. Participants were asked the question *In general how satisfied are you with your health?* and given the options *Extremely happy*, *Very happy*, *Moderately happy*, *Moderately unhappy*, *Very unhappy*, *Extremely unhappy*, *Do not know* and *Prefer not to answer* which were encoded 6, 5, 4, 3, 2, 1, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4548.

hearing_acuity – Hearing acuity – no units. Mean of left and right ear Speech Reception Threshold (SRT), defined here as the signal-to-noise ratio at which half of the presented speech can be understood correctly. This value was multiplied by -1 so that larger values correspond to better hearing. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20019.

heelbone_density – Heelbone density – Z-score. Mean of left and right heelbone density T-score calculated from an ultrasound heel Bone Mineral Density measurement and normalised within each sex. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4106.

infertility_self_declared_glm – Self-reported infertility – TRUE/FALSE. UKB participants were asked in a verbal interview with a trained nurse to describe any serious illness or disabilities. Responses were classified in a tree-structured list used by clinic nurses to code non-cancer illnesses. The values 1403 and 1404 correspond to female and male infertility respectively and participants with these either of these responses were encoded 1. All other participants who completed the verbal interview were encoded 0. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20002.

irritability_glm – Self-reported irritability – TRUE/FALSE. Participants were asked the question *Are you an irritable person?* and given the options *Yes*, *No*, *Do not know* and *Prefer not to answer* which were encoded 1, 0, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=1940.

job_satisfaction – Job satisfaction – index. Participants were asked the question *In general how satisfied are you with the work that you do?* and given the options *Extremely happy*, *Very happy*, *Moderately happy*, Mod*erately unhappy*, *Very unhappy*, *Extremely unhappy*, *Do not know* and *Prefer not to answer* which were encoded 6, 5, 4, 3, 2, 1, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4537.

match_time – Reaction time – ms. Participants were shown two cards at a time on a touchscreen and instructed to press a button as quickly as possible when the symbols on the cards match. This field is the mean duration to first press of snap-button summed over rounds in which both cards matched. It gives a crude measure of the raw processing + reaction speed of a participant. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20023.

memory – Memory – count. The participant was shown a 2-digit number to remember. The number then disappeared and after a short while they were asked to enter the number onto the screen. The number became one digit longer each time they remembered correctly (up to a maximum of 12 digits). This trait is the longest number correctly recalled during the numeric memory test. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=4282.

miscarriage – Miscarriage – TRUE/FALSE. Women only. Female participants were asked the question *Have you ever had any stillbirths, spontaneous miscarriages or terminations?* and given the options *Yes*, *No*, *Do not know* and *Prefer not to answer* which were encoded 1, 0, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2774.

moderate_activity – Frequency of moderate activity – count. Participants were asked the question *In a typical week, on how many days did you do 10 minutes or more of moderate physical activities like carrying light loads, cycling at normal pace? (Do not include walking)*. Values less than 0 or greater than 7 were rejected. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=884.

moody_glm – Moody – TRUE/FALSE. Participants were asked the question *Does your mood often go up and down?* and given the options *Yes*, *No*, *Do not know* and *Prefer not to answer* which were encoded 1, 0, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=1920.

motorway_speeding – Driving speed – index. Participants were asked the question *How often do you drive faster than the speed limit on the motorway?* and given the options *Never/rarely*, *Sometimes*, *Often*, *Most of the time* ,*Do not drive on the motorway*, *Do not know* and *Prefer not to answer* which were encoded 0, 1, 2, 3, NA, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=1100.

neuroticism – Neuroticism – index. An externally derived summary score of neuroticism, based on 12 neurotic behaviour domains reported in UKB. Values range from 0 to 12 with higher scores corresponding to increased neurotic behaviour. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20127.

number_sexual_partners – Number sexual partners – count. Participants were asked the question *About how many sexual partners have you had in your lifetime*?. Subjects who answered *Do not know* or *Prefer not to answer* were set to NA, otherwise values were capped at 100. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2149.

overall_health – Self-reported overall health – index. Participants were asked the question *In general how would you rate your overall health?* and given the options *Excellent*, *Good*, *Fair*, *Poor*, *Do not know* and *Prefer not to answer* which were encoded 3, 2, 1, 0, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2178.

pacemaker_glm – Pacemaker – TRUE/FALSE. Participants were asked by an interviewer if they have a pace-maker before the body impedance measures. Those that answered *Yes* were encode 1, otherwise 0. http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=3079.

pgrs_edu – Polygenic score for Education Attained – years. A polygenic score for Education Attained (EA) calculated from 159 genome-wide significant SNPs reported in a GWAS of Education Attained [Okbay et al. 2016] and imputed in UKB using the UK10K + 1000 Genomes panel. This polygenic score explains 0.9% of the residual variance of EA in the UKB British cohort after conditioning on sex and age.

pgrs_height – Polygenic score for Height – metres. A polygenic score for height calculated from 697 genome-wide significant SNPs reported in a GWAS of height [Wood et al. 2014] and imputed in UKB using the UK10K + 1000 Genomes panel. This polygenic score explains 18.7% of the residual variance of height in the UKB British cohort after conditioning on sex and age.

potassium – Urinary Potassium – mML^{-1}. Potassium in urine measured by ISE (ion selective electrode) analysis on a Beckman Coulter AU5400. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=30520.

risk_glm – Self-reported risk taker – TRUE/FALSE. Participants were asked the question *Would you describe yourself as someone who takes risks?* and given the options *Yes*, *No*, *Do not know* and *Prefer not to answer* which were encoded 1, 0, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2040.

sleep_duration – Sleep duration – hours. Participants were asked the question *About how many hours sleep do you get in every 24 hours? (please include naps)*. Subjects who answered *Do not know* or *Prefer not to answer* were set to NA, as were values less than 1 or greater than 23. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=1160.

sodium – Urinary Sodium – mML^{-1}. Sodium in urine measured by ISE (ion selective electrode) analysis on a Beckman Coulter AU5400. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=30530.

vigorous_activity – Frequency of vigorous activity – count. Participants were asked the question *In a typical week, how many days did you do 10 minutes or more of vigorous physical activity? (These are activities that make you sweat or breathe hard such as fast cycling, aerobics, heavy lifting)*. Values less than 0 or greater than 7 were rejected. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=904.

visual_acuity – Visual acuity – negative log(MAR). Mean of left and right visual acuity as defined by the smallest size letters that can be reliably identified at a 4 metres. The UK Biobank system is based on a traditional LogMar chart. This log(MAR) value was multiplied by -1 so that larger values correspond to better vision. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=5201.

walking_pace – Walking pace – index. Participants were asked the question *How would you describe your usual walking pace?* and given the options *Slow pace*, *Steady average pace*, *Brisk pace*, *None of the above* and *Prefer not to answer* which were encoded 0, 1, 2, NA and NA respectively. For more details see http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=924.

SUPPLEMENTARY NOTE 2: Personal acknowledgments

Yukinori Okada Y.O. BioBank Japan was supported by the Tailor-Made Medical Treatment program (the BioBank Japan Project) of the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), the Japan Agency for Medical Research and Development (AMED).
Xiuqing Guo X.G. MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was

provided by NHLBI Contract N02-HL-64278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Carolina Medina-Gomez C.M.-G. Supported by the Netherlands Organization for Health Research and Development (ZonMw VIDI 016.136.367).
Annelot M Dekker A.M.D. ALS Foundation the Netherlands
He Gao H.G. The Airwave Health Monitoring Study is funded by the Home Office (grant number 780-TETRA) with additional support from the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre (BRC). The study uses the computing resources of the UK MEDical BIOinformatics partnership - aggregation, integration, visualization and analysis of large, complex data (UK MED-BIO), which is supported by the Medical Research Council (MR/L01632X/1).
May E Montasser M.E.M. The Amish studies are supported by grants and contracts from the NIH, including R01 AG18728, R01 HL088119, U01 GM074518, U01 HL072515-06, U01 HL84756, R01 DK54261, U01 HL137181, , the University of Maryland General Clinical Research Center, grant M01 RR 16500, the Mid-Atlantic Nutrition Obesity Research Center grant P30 DK72488, and by the Baltimore Diabetes Research and Training Center grant P60DK79637, the American Heart Association: 12SDG9280031 and 17GRNT33661168. In addition, this project was supported by National Research Initiative Competitive Grant no. 2007-35205-17883 from the USDA National Institute of Food and Agriculture. We gratefully thank our Amish community and research volunteers for their long-standing partnership in research, and acknowledge the dedication of our Amish liaisons, field workers and the Amish Research Clinic staff, without which these studies would not have been possible.
Daniel Shriner D.S. The authors acknowledge with thanks the participants in the AADM project, their families and their physicians. The study was supported in part by the Intramural Research Program of the National Institutes of Health in the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute (NHGRI), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Center for Information Technology, and the Office of the Director at the National Institutes of Health (1ZIAHG200362). Support for participant recruitment and initial genetic studies of the AADM study was provided by NIH grant No. 3T37TW00041-03S2 from the Office of Research on Minority Health.
Cassandra N Spracklen C.N.S. American Heart Association 17POST3650016
Rona J Strawbridge R.J.S. RJS is supported by a UKRI Innovation- HDR-UK Fellowship (MR/S003061/1)
Niek Verweij N.V. NWO-VENI (016.186.125)
Helen R Warren H.R.W. This work was funded by the Medical Research Council of Great Britain (grant number: G9521010D). The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. We wish to acknowledge the support of the NIHR Cardiovascular Biomedical Research Unit at Barts and Queen Mary University of London, UK
Robyn E Wootton R.E.W. Funded by the MRC [MC_UU_12013/6] and the NIHR Biomedical Research Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol.
Lisa R Yanek L.R.Y. GeneSTAR was supported by grants from the National Institutes of Health/National Heart, Lung, and Blood Institute (U01 HL72518,HL087698, HL49762, HL59684, HL58625, HL071025, HL092165, HL099747, K23HL105897),
Lisa R Yanek L.R.Y. National Institutes of Health/National Institute of Nursing Research (NR0224103, NR008153),
Lisa R Yanek L.R.Y. National Institutes of Health/National Institute of Neurological Disorders and Stroke (NS062059)
Lisa R Yanek L.R.Y. National Institutes of Health/National Center for Research Resources (M01-RR000052) to the Johns Hopkins General Clinical Research Center,

Lisa R Yanek L.R.Y. and National Institutes of Health/National Center for Research Resources/ National Center for Advancing Translational Sciences (UL1 RR 025005) to the Johns Hopkins Institute for Clinical & Translational Research
Jie Yao J.Yao MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Adebowale A Adeyemo A.A.A. The authors acknowledge with thanks the participants in the AADM project, their families and their physicians. The study was supported in part by the Intramural Research Program of the National Institutes of Health in the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute (NHGRI), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Center for Information Technology, and the Office of the Director at the National Institutes of Health (1ZIAHG200362). Support for participant recruitment and initial genetic studies of the AADM study was provided by NIH grant No. 3T37TW00041-03S2 from the Office of Research on Minority Health.
Matthew L Albert M.L.A. The Milieu Intérieur cohort was supported by the French government's program Investissement d'Avenir, managed by the Agence Nationale de la Recherche (reference 10-LABX-69-01).
Matthew A Allison M.A.A. MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Maris Alver M.AlverEGCUT was supported by Estonian Research Council [IUT20-60, IUT24-6, PUT1660 to T.E and PUT1665 to K.F.; European Union Horizon 2020 [692145]; European Union through the European Regional Development Fund [2014-2020.4.01.15-0012 GENTRANSMED];
Amy R Bentley A.R.B. The AADM and HUFS studies were supported in part by the Intramural Research Program of the National Institutes of Health in the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute (NHGRI), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Center for Information Technology, and the Office of the Director at the National Institutes of Health (1ZIAHG200362).
Guanjie Chen G.Chen The AADM and HUFS studies were supported in part by the Intramural Research Program of the National Institutes of Health in the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute (NHGRI), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Center for Information Technology, and the Office of the Director at the National Institutes of Health (1ZIAHG200362).

Yii-Der I Chen Y.-D.I.C.TUDR (Taiwan-US Diabetic Retinopathy): This study was supported by the National Eye Institute of the National Institutes of Health (EY014684 to J.I.R. and Y.-D.I.C.) and ARRA Supplement (EY014684-03S1, -04S1), the Eye Birth Defects Foundation Inc., the National Science Council, Taiwan (NSC 98-2314-B-075A-002-MY3 to W.H.S.) and the Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1003001C to W.H.S .). DNA handling and genotyping were supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001880 and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Yen-Feng Chiu Y.-F.C. The TAICHI study was supported by grants from the National Health Research Institutes, Taiwan (PH-099-PP-03, PH-100-PP-03, and PH-101-PP-03); the National Science Council, Taiwan (NSC 101-2314-B-075A-006-MY3, MOST 104-2314-B-075A-006-MY3, MOST 104-2314-B-075A-007, and MOST 105-2314-B-075A-003); and the Taichung Veterans General Hospital, Taiwan (TCVGH-1020101C, TCVGH-020102D, TCVGH-1023102B, TCVGH-1023107D, TCVGH-1030101C, TCVGH-1030105D, TCVGH-1033503C, TCVGH-1033102B, TCVGH-1033108D, TCVGH-1040101C, TCVGH-1040102D, TCVGH-1043504C, and TCVGH-1043104B); it was also supported in part by the National Center for Advancing Translational Sciences (CTSI grant UL1TR001881).
Ayo P Doumatey A.P.D. The authors acknowledge with thanks the participants in the AADM project, their families and their physicians. The study was supported in part by the Intramural Research Program of the National Institutes of Health in the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute (NHGRI), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Center for Information Technology, and the Office of the Director at the National Institutes of Health (1ZIAHG200362). Support for participant recruitment and initial genetic studies of the AADM study was provided by NIH grant No. 3T37TW00041-03S2 from the Office of Research on Minority Health.
Nduna Dzimiri N.D. The authors would like to thank Editha Andres, Nejat Mazher and Dr. Maie Alshaid for their assistance in patient sample and clinical data collection.
Paul Elliott P.E. P.E. acknowledges support from the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre, the NIHR Health Protection Research Unit in Health Impact of Environmental Hazards (HPRU-2012-10141), and the Medical Research Council (MRC) and Public Health England (PHE) Centre for Environment and Health (MR/L01341X/1). This work used the computing resources of the UK MEDical BIOinformatics partnership UK MED-BIO)supported by the Medical Research Council (MR/L01632X/1). P.E. is a UK Dementia Research Institute (DRI) professor, UK DRI at Imperial College London, funded by the MRC, Alzheimer's Society and Alzheimer's Research UK; and associate director, Health Data Research UK - London funded by a consortium led by the MRC.
Janine F Felix J.F.F. Received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 633595 (DynaHEALTH) and No 733206 (LifeCycle).
Krista Fischer K.F. EGCUT was supported by Estonian Research Council [IUT20-60, IUT24-6, PUT1660 to T.E and PUT1665 to K.F.; European Union Horizon 2020 [692145]; European Union through the European Regional Development Fund [2014-2020.4.01.15-0012 GENTRANSMED];
Anuj Goel A.G. Anuj Goel acknowledges support from BHF, Wellcome Trust, European Community Sixth Framework Program (LSHM-CT- 2007-037273), European Union Seventh Framework Programme FP7/2007-2013 under grant agreement no. HEALTH-F2-2013-601456 (CVGenes@Target), TriPartite Immunometabolism Consortium [TrIC]- Novo Nordisk Foundation's Grant number NNF15CC0018486
Mark O Goodarzi M.O.G. MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-

95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Chao Agnes Hsiung C.A.Hsiung The THRV study is supported by National Heart, Lung, and Blood Institute grant R01HL111249.
Jinyan Huang J.H. the National Natural Science Foundation of China (Grants 81570122, 81770205);
Yi-Jen Hung Y.-J.H. The TAICHI study was supported by grants from the National Health Research Institutes, Taiwan (PH-099-PP-03, PH-100-PP-03, and PH-101-PP-03); the National Science Council, Taiwan (NSC 101-2314-B-075A-006-MY3, MOST 104-2314-B-075A-006-MY3, MOST 104-2314-B-075A-007, and MOST 105-2314-B-075A-003); and the Taichung Veterans General Hospital, Taiwan (TCVGH-1020101C, TCVGH-020102D, TCVGH-1023102B, TCVGH-1023107D, TCVGH-1030101C, TCVGH-1030105D, TCVGH-1033503C, TCVGH-1033102B, TCVGH-1033108D, TCVGH-1040101C, TCVGH-1040102D, TCVGH-1043504C, and TCVGH-1043104B); it was also supported in part by the National Center for Advancing Translational Sciences (CTSI grant UL1TR001881).
Zoltán Kutalik Z.K. Swiss National Science Foundation (31003A_169929)
Zoltán Kutalik Z.K. SystemsX.ch (51RTP0_151019)
Deborah A Lawlor D.A.L. Born in Bradford (BiB) receives core infrastructure funding from the Wellcome Trust (WT101597MA) a joint grant from the UK Medical Research Council (MRC) and UK Economic and Social Science Research Council (ESRC) (MR/N024397/1) and the National Institute for Health Research (NIHR) under its Collaboration for Applied Health Research and Care (CLAHRC) for Yorkshire and Humber. The research presented in this paper is supported by the British Heart Foundation (CS/16/4/32482), US National Institute of Health (R01 DK10324), the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 669545), and the NIHR Biomedical Centre at the University Hospitals Bristol NHS Foundation Trust and the University of Bristol. D.A.L. works in a unit that receives UK MRC funding (MC_UU_12013/5) and D.A.L. is an NIHR senior investigator (NF-SI-0611-10196). The funders had no role in the design of the study, the collection, analysis, or interpretation of the data; the writing of the manuscript, or the decision to submit the manuscript for publication
Wen-Jane Lee W.-J.L. The THRV study is supported by National Heart, Lung, and Blood Institute grant R01HL111249.
Stephanie J London S.J.L. Supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences, NIH ZO1 ES043012
Reedik Mägi R.M. EGCUT was supported by Estonian Research Council [IUT20-60, IUT24-6, PUT1660 to T.E and PUT1665 to K.F.; European Union Horizon 2020 [692145]; European Union through the European Regional Development Fund [2014-2020.4.01.15-0012 GENTRANSMED];
Ani W Manichaikul A.W.M. MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

Lili Milani L.M. EGCUT was supported by Estonian Research Council [IUT20-60, IUT24-6, PUT1660 to T.E and PUT1665 to K.F.; European Union Horizon 2020 [692145]; European Union through the European Regional Development Fund [2014-2020.4.01.15-0012 GENTRANSMED];
Alison D Murray A.D.M. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award STratifying Resilience and Depression Longitudinally (STRADL) Reference 104036/Z/14/Z).
Josyf C Mychaleckyj J.C.M. MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Sandosh Padmanabhan S.P. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award STratifying Resilience and Depression Longitudinally (STRADL) Reference 104036/Z/14/Z).
Ozren Polasek O.P. CROATIA cohorts were funded by the Medical Research Council UK, The Croatian Ministry of Science, Education and Sports (grant 216-1080315-0302), the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947), the Croatian Science Foundation (grant 8875) and the Centre of Excellence in Personalized Health Care.
Neil Poulter N.Poulter This work was funded by the National Institutes for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Barts Biomedical Research Centre and the NIHR Biomedical Research Centre at Imperial College, the International Centre for Circulatory Health Charity and the Medical Research Council through G952010. We thank all ASCOT trial participants, physicians, nurses, and practices in the participating countries for their important contribution to the study.
Lluis Quintana-Murci L.Q.-M. The Milieu Intérieur cohort was supported by the French government's program Investissement d'Avenir, managed by the Agence Nationale de la Recherche (reference 10-LABX-69-01).
Dabeeru C Rao D.C.R. The THRV study is supported by National Heart, Lung, and Blood Institute grant R01HL111249.
Stephen S Rich S.S.R. MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

Cornelius A Rietveld C.A.R. Netherlands Organisation for Scientific Research Veni grant 016.165.004
Maria Sabater-Lleal M.S.-L. MS-L is supported by Miguel Servet contract (ISCIII CP17/00142) from the Spanish Ministry of Health, and acknowledges funding from the Swedish Heart and Lung Foundation (#20160290).
Veikko Salomaa V.S. VS was supported By the Finnish Foundation for Cardiovascular Research.
Kevin Sandow K.Sandow The TAICHI study was supported by grants from the National Health Research Institutes, Taiwan (PH-099-PP-03, PH-100-PP-03, and PH-101-PP-03); the National Science Council, Taiwan (NSC 101-2314-B-075A-006-MY3, MOST 104-2314-B-075A-006-MY3, MOST 104-2314-B-075A-007, and MOST 105-2314-B-075A-003); and the Taichung Veterans General Hospital, Taiwan (TCVGH-1020101C, TCVGH-020102D, TCVGH-1023102B, TCVGH-1023107D, TCVGH-1030101C, TCVGH-1030105D, TCVGH-1033503C, TCVGH-1033102B, TCVGH-1033108D, TCVGH-1040101C, TCVGH-1040102D, TCVGH-1043504C, and TCVGH-1043104B); it was also supported in part by the National Center for Advancing Translational Sciences (CTSI grant UL1TR001881).
Bengt Sennblad B.S. Bengt Sennblad is financially supported by the Knut and Alice Wallenberg Foundation as part of the National Bioinformatics Infrastructure Sweden at SciLifeLab.
Peter J Sever P.J.S. This work was funded by the National Institutes for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Barts Biomedical Research Centre and the NIHR Biomedical Research Centre at Imperial College, the International Centre for Circulatory Health Charity and the Medical Research Council through G952010. We thank all ASCOT trial participants, physicians, nurses, and practices in the participating countries for their important contribution to the study.
Wayne H-H Sheu W.H.-H.S. TUDR (Taiwan-US Diabetic Retinopathy): This study was supported by the National Eye Institute of the National Institutes of Health (EY014684 to J.I.R. and Y.-D.I.C.) and ARRA Supplement (EY014684-03S1, -04S1), the Eye Birth Defects Foundation Inc., the National Science Council, Taiwan (NSC 98-2314-B-075A-002-MY3 to W.H.S.) and the Taichung Veterans General Hospital, Taichung, Taiwan (TCVGH-1003001C to W.H.S .). DNA handling and genotyping were supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001880 and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Timo Tonis Sikka T.T.S. EGCUT was supported by Estonian Research Council [IUT20-60, IUT24-6, PUT1660 to T.E and PUT1665 to K.F.; European Union Horizon 2020 [692145]; European Union through the European Regional Development Fund [2014-2020.4.01.15-0012 GENTRANSMED];
Blair H Smith B.H.S. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award STratifying Resilience and Depression Longitudinally (STRADL) Reference 104036/Z/14/Z).
Alice Stanton A.Stanton This work was funded by the National Institutes for Health Research (NIHR) as part of the portfolio of translational research of the NIHR Barts Biomedical Research Unit and the NIHR Biomedical Research Centre at Imperial College, the International Centre for Circulatory Health Charity and the Medical Research Council through G952010. We thank all ASCOT trial participants, physicians, nurses, and practices in the participating countries for their important contribution to the study.
Kent D Taylor K.D.T. MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-

95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Russell P Tracy R.P.T. MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Jan H Veldink J.H.V. ALS foundation the Netherlands
Ya Xing Wang Y.X.W. no
Nicholas J Wareham N.J.W. MRC Epidemiology Unit, Fenland study, EPIC-Norfolk case-cohort study funding: this study was funded by the United Kingdom's Medical Research Council through grants MC_UU_12015/1, MC_PC_13046, MC_PC_13048 and MR/L00002/1
Chittaranjan S Yajnik C.S.Y. The Wellcome Trust, London, UK; MRC, UK, Department of Biotechnology, Government of India, New Delhi, India
Jie Zhou J.Z. he AADM and HUFS studies were supported in part by the Intramural Research Program of the National Institutes of Health in the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute (NHGRI), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Center for Information Technology, and the Office of the Director at the National Institutes of Health (1ZIAHG200362).
Folkert W Asselbergs F.W.A. Folkert W. Asselbergs is supported by UCL Hospitals NIHR Biomedical Research Centre.
Stephan J L Bakker S.J.L.B. The establishment of the PREVEND study was primarily funded by the Dutch Kidney Foundation.
Leonard H van den Berg L.H.v.d.B. Supported by The Netherlands ALS Foundation
Sonja I Berndt S.I.B. Pegasus was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH.
Dwaipayan Bharadwaj D.B. Funding source:Council of Scientific and Industrial Research [CSIR], Government of India through Centre for Cardiovascular and Metabolic Disease Research [CARDIOMED] project [Grant No: BSC0122).
Archie Campbell A.C. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award STratifying Resilience and Depression Longitudinally (STRADL) Reference 104036/Z/14/Z).
Mark J Caulfield M.J.C. This work was funded by the Medical Research Council of Great Britain (grant number: G9521010D). The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. This work forms part of the research themes contributing to the translational research portfolio for the NIHR Barts Cardiovascular Biomedical Research Centre.

Giriraj Ratan Chandak G.R.C. This research was supported by Council of Scientific and Industrial Research (CSIR), Ministry of Science and Technology, Government of India, India, under their XII-Five Year Plan titled CARDIOMED. The cohorts were supported by funds from Department of Biotechnology, Ministry of Science and Technology, Government of India, India and Wellcome Trust, London, UK
Marina Ciullo M.Ciullo This work was supported by the Italian Ministry of Universities and CNR (Interomics Flagship Project, PON03PE_00060_7), the Assessorato Ricerca Regione Campania, the Fondazione con il SUD (2011-PDR-13), and the Istituto Banco di Napoli - Fondazione to MC.
Daniele Cusi D.Cusi HYPERGENES grant from EU: FP7-HEALTH-F4-2007-201550
Daniele Cusi D.Cusi InterOmics flagship grant from CNR (Italian National Institute of Research): PB05_SP3
Daniele Cusi D.Cusi ATHENA grant from EU: FP7, EU Grant Agreement 245121
George Davey-Smith G.D.-S. Funded by the MRC: http://www.mrc.ac.uk [MC_UU_12013/1]
Bjarke Feenstra B.F. Oak Foundation fellowship and Novo Nordisk Foundation grant (12955)
Mary Feitosa M.F. The FamHS was supported by RO1HL117078 from the National Heart, Lung, and Blood Institute, and R01DK08925607 from the National Institute of Diabetes and Digestive and Kidney Diseases.
Andre Franke A.F. FoCus received infrastructure support from the DFG Cluster of Excellence Inflammation at Interfaces.
Paolo Gasparini P.Gasparini Acknowledgements: All the populations recruited that took part in the study. Funds: RC-5x1000 Ministry of Health (to PG).
Christian GiegerC.Gieger The KORA study was initiated and financed by the Helmholtz Zentrum München – German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC-Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. This study was supported by the German Center for Diabetes Research (DZD e.V.). For this publication, biosamples from the KORA Biobank as part of the Joint Biobank Munich have been used.
Struan F A Grant S.F.A.G. CHOP: We thank the network of primary care clinicians, their patients and families for their contribution to this project and clinical research facilitated through the Pediatric Research Consortium (PeRC) at The Children's Hospital of Philadelphia. Rosetta Chiavacci, Elvira Dabaghyan, Hope Thomas, Kisha Harden, Andrew Hill, Kenya Fain, Crystal Johnson-Honesty, Cynthia Drummond, Shanell Harrison and Sarah Wildrick, Cecilia Kim, Edward Frackelton, George Otieno, Kelly Thomas, Cuiping Hou, Kelly Thomas and Maria L. Garris provided expert assistance with genotyping or data collection and management. We would also like to thank Smari Kristinsson, Larus Arni Hermannsson and Asbjörn Krisbjörnssonof Raförninn ehf for their extensive software design and contribution. This research was financially supported by the Daniel B. Burke Endowed Chair for Diabetes Research, an Institute Development Award from the Children's Hospital of Philadelphia, a Research Development Award from the Cotswold Foundation and NIH grant R01-HD056465.
Struan F A Grant S.F.A.G. BMDCS: The study was funded by R01 HD58886; the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) contracts (N01-HD-1-3228, -3329, -3330, -3331,-3332, -3333); and the CTSA program Grant 8 UL1 TR000077. The funders had no role in the design and conduct of the study; collection, management,analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. We appreciate the dedication of the study participants and their families, and the support of Dr Karen Winer, Scientific Director of the Bone Mineral Density in Childhood Study.
Lyn R Griffiths L.R.G. The Norfolk Island Health Study was supported by funding from National Health and Medical Research Council of Australia (NHMRC) Project Grants. Our appreciation goes to the Norfolk Islanders who volunteered for this study.

Leif Groop L.G. The Botnia studies (L.G) have been financially supported by grants from the Sigrid Juselius Foundation, The Academy of Finland (grants no. 263401, 267882, 312063 to LG), Nordic Center of Excellence in Disease Genetics. The research leading to these results has also received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 269045.
Leif Groop L.G. The study has also been supported by the Ministry of Education in Finland, Municipal Heath Care Center and Hospital in Jakobstad and Health Care Centers in Vasa, Närpes and Korsholm. The skillful assistance of the Botnia Study Group is gratefully acknowledged.
Vilmundur Gudnason V.G. The AGES Reykjavik study has been funded by NIH contracts N01-AG-1-2100 and 271201200022C, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association), and the Althingi (the Icelandic Parliament)
Ulf Gyllensten U.G. Swedish Research Council, European Union Framework Grant,
Anders Hamsten A.Hamsten PROCARDIS was funded by the European Commission Framework 6 (FP6) program (LSHM-CT-2007-037273), the British Heart Foundation, AstaZeneca, the Swedish Research Council (8691), the Knut and Alice Wallenberg Foundation, the Swedish Heart-Lung Foundation and the Torsten and Ragnar Söderberg Foundation.
Andrew A Hicks A.A.H. In South Tyrol, the study was supported by the Ministry of Health and Department of Educational Assistance, University and Research of the Autonomous Province of Bolzano.
Hagit Hochner H.Hochner This work was supported by the National Institutes of Health [grants R01HL088884, K01HL103174, 2T32HD052462-06, T32HL007902]; the Israeli Science Foundation [grant numbers 1252/07, 552/12]; the Israel Ministry of Health [IMH49267]; and the NUS-HUJ CREATE Programme of the National Research Foundation, Singapore [Project number 370062002].
Steven C Hunt S.C.H. The hypertension genetic epidemiology network was funded by cooperative agreements (U10) with NHLBI: HL54471, HL54472, HL54473, HL54495, HL54496, HL54497, HL54509, HL54515, and 2 R01 HL55673-12. Funding also came from the Biomedical Research Program, Qatar Foundation, Qatar.
Vincent W V Jaddoe V.W.V.J. Supported by the Netherlands Organization for Health Research and Development (ZonMw VIDI 016.136.361), a Consolidator Grant from the European Research Council (ERC-2014-CoG-648916) and the European Union's Horizon 2020 research and innovation programme under grant agreements No 633595 (DynaHEALTH) and No 733206 (LifeCycle).
Magnus Johannesson M.J. The Jan Wallander and Tom Hedelius Foundation (P2015-0001:1), The Swedish Research Council (421-2013-1061)
Jost B Jonas J.B.J. Supported by National Natural Science Foundation of China (grant # 81570835)
Meena Kumari M.Kumari The UK Household Longitudinal Study is led by the Institute for Social and Economic Research at the University of Essex. The UK Household Longitudinal Study is funded by the Economic and Social Research Council (ES/H029745/1). The survey was conducted by NatCen and the genome-wide scan data were analysed and deposited by the Wellcome Trust Sanger Institute (WT098051). Information on how to access the data can be found on the Understanding Society website https://www.understandingsociety.ac.uk/.
Markku Laakso M.Laakso The study was supported by grants from the Academy of Finland and the Sigrid Juselius Foundation.
Jari Lahti J.Lahti Academy of Finland and University of Helsinki
Wolfgang Lieb W.L. The popgen 2.0 network has been financed by the German Ministry for Education and Research; 01EY1103
Giuseppe Matullo G.Matullo The work was supported by the Compagnia di San Paolo for the EPIC-Italy and EPICOR projects, the Italian Institute for Genomic Medicine (IIGM, formerly Human Genetics Foundation-Torino, HuGeF, Turin, Italy) and the MIUR ex60% grant. EPIC-Italy is further supported by a grant from the Associazione Italiana per la Ricerca sul Cancro (AIRC, Milan).

Mark Ian McCarthy M.I.M. MMcC is a Wellcome Senior Investigator and an NIHR Senior Investigator.
Sarah E Medland S.E.M. SEM was funded by an NHMRC Senior Reseach Fellowship (APP1103623)
Andres Metspalu A.Metspalu EGCUT was supported by Estonian Research Council [IUT20-60, IUT24-6, PUT1660 to T.E and PUT1665 to K.F.; European Union Horizon 2020 [692145]; European Union through the European Regional Development Fund [2014-2020.4.01.15-0012 GENTRANSMED];
Grant W Montgomery G.W.M. G.W.M. was supported by the NHMRC Fellowships Scheme (339446, 619667)
Dennis Mook-Kanamori D.M.-K. The authors of the NEO study thank all individuals who participated in the Netherlands Epidemiology in Obesity study, all participating general practitioners for inviting eligible participants and all research nurses for collection of the data. We thank the NEO study group, Pat van Beelen, Petra Noordijk and Ingeborg de Jonge for the coordination, lab and data management of the NEO study. The genotyping in the NEO study was supported by the Centre National de Génotypage (Paris, France), headed by Jean-Francois Deleuze. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Center, and by the Leiden University, Research Profile Area Vascular and Regenerative Medicine. Dennis Mook-Kanamori is supported by Dutch Science Organization (ZonMW-VENI Grant 916.14.023).
Patricia B Munroe P.B.M. This work was funded by the Medical Research Council of Great Britain (grant number: G9521010D). The BRIGHT study is extremely grateful to all the patients who participated in the study and the BRIGHT nursing team. We wish to acknowledge the support of the NIHR Cardiovascular Biomedical Research Centre at Barts and Queen Mary University of London, UK
Kari E North K.E.N. see ARIC disclosure
Jeffery R O'connell J.R.O'c. The Amish studies are supported by grants and contracts from the NIH, including R01 AG18728, R01 HL088119, U01 GM074518, U01 HL072515-06, U01 HL84756, R01 DK54261, U01 HL137181, , the University of Maryland General Clinical Research Center, grant M01 RR 16500, the Mid-Atlantic Nutrition Obesity Research Center grant P30 DK72488, and by the Baltimore Diabetes Research and Training Center grant P60DK79637, the American Heart Association: 12SDG9280031 and 17GRNT33661168. In addition, this project was supported by National Research Initiative Competitive Grant no. 2007-35205-17883 from the USDA National Institute of Food and Agriculture. We gratefully thank our Amish community and research volunteers for their long-standing partnership in research, and acknowledge the dedication of our Amish liaisons, field workers and the Amish Research Clinic staff, without which these studies would not have been possible.
Carole Ober C.O. NIH Grants R01 HL085197, T32 GM007197, and F31 HL123289.
Walter Palmas W.P. MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Colin Palmer C.Palmer We are grateful to all the participants in this study, the general practitioners, the Scottish School of Primary Care for their help in recruiting the participants, and to the whole team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The study complies with the

Declaration of Helsinki. We acknowledge the support of the Health Informatics Centre, University of Dundee for managing and supplying the anonymised data and NHS Tayside, the original data owner. The Wellcome Trust United Kingdom Type 2 Diabetes Case Control Collection (GoDARTS) was funded by The Wellcome Trust (072960/Z/03/Z, 084726/Z/08/Z, 084727/Z/08/Z, 085475/Z/08/Z, 085475/B/08/Z) and as part of the EU IMI-SUMMIT program.
Etienne Patin E.P. The Milieu Intérieur cohort was supported by the French government's program Investissement d'Avenir, managed by the Agence Nationale de la Recherche (reference 10-LABX-69-01).
Louis Perusse L.P. The Quebec Family Study was funded by multiple grants from the Medical Research Council of Canada and the Canadian Institutes for Health Research. This work was supported by a team grant from the Canadian Institutes for Health Research (FRN-CCT-83028).
Mario Pirastu M.Pirastu We thank the Ogliastra population and all the individuals who participated in this study. We are very grateful to the municipal administrators for their collaboration to the project and for economic and logistic support. This research was supported by grants from the Italian Ministry of Education, University and Research (MIUR) no.5571/DSPAR/2002 and (FIRB) D. M. no. 718/Ric/2005.
David J Porteous D.J.P. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award STratifying Resilience and Depression Longitudinally (STRADL) Reference 104036/Z/14/Z).
Danielle Posthuma D.P. The Netherlands Organization for Scientific Research (NWO VICI 435-14-005. Analyses were carried out on the Genetic Cluster Computer, which is financed by the Netherlands Scientific Organization (NWO: 480-05-003 awarded to DP), by the VU University, Amsterdam, the Netherlands, and by the Dutch Brain Foundation, and is hosted by the Dutch National Computing and Networking Services SurfSARA.
Bruce M Psaty B.M.P. Cardiovascular Health Study: This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268201800001C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, R01HL120393, and R01HL130114 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on Aging (NIA). A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The provision of genotyping data was supported in part by UL1TR000124, and DK063491.
Bruce M Psaty B.M.P. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
John D Rioux J.D.R. JDR acknowledges financial support from the Montreal Heart Institute Foundation
Fernando Rivadeneira F.Rivadeneira The Netherlands Organization for Health Research and Development supported FR (ZonMw VIDI 016.136.367)
Charles Rotimi C.R. The AADM and HUFS studies were supported in part by the Intramural Research Program of the National Institutes of Health in the Center for Research on Genomics and Global Health (CRGGH). The CRGGH is supported by the National Human Genome Research Institute (NHGRI), the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Center for Information Technology, and the Office of the Director at the National Institutes of Health (1ZIAHG200362).
Jerome I Rotter J.I.R. MESA (Multi-Ethnic Study of Atherosclerosis): MESA and the MESA SHARe project are supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-

95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1-TR-001881, and DK063491. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-64278. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
Igor Rudan I.R. None
Dharambir K Sanghera D.K.S. This work was supported by NIH grants -R01DK082766 funded by the National Institute of Health (NIDDK) and NOT-HG-11-009 funded by NHGRI, VPR Bridge Grant and Enrichment Grant from Harold Hamm Diabetes Center from University of Oklahoma Health Sciences Center. Authors thank all the participants of AIDHS/SDS and are grateful for their contribution in this study.
Matthias B Schulze M.B.S. The study was supported in part by a grant from the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD e.V.) and the State of Brandenburg. The recruitment phase of the EPIC-Potsdam study was supported by the Federal Ministry of Science, Germany (01 EA 9401) and the European Union (SOC 95201408 05 F02). The follow-up of the EPIC-Potsdam study was supported by German Cancer Aid (70-2488-Ha I) and the European Community (SOC 98200769 05 F02).
Matthias B Schulze M.B.S. Exome chip genotyping of EPIC-Potsdam samples was carried out under supervision of Per Hoffmann and Stefan Herms at Life & Brain GmbH, Bonn. We are grateful to the Human Study Centre (HSC) of the German Institute of Human Nutrition Potsdam-Rehbrücke, namely the trustee and the data hub for the processing, and the participants for the provision of the data, the biobank for the processing of the biological samples and the head of the HSC, Manuela Bergmann, for the contribution to the study design and leading the underlying processes of data generation.
Alan R Shuldiner A.R.S. The Amish studies are supported by grants and contracts from the NIH, including R01 AG18728, R01 HL088119, U01 GM074518, U01 HL072515-06, U01 HL84756, R01 DK54261, U01 HL137181, , the University of Maryland General Clinical Research Center, grant M01 RR 16500, the Mid-Atlantic Nutrition Obesity Research Center grant P30 DK72488, and by the Baltimore Diabetes Research and Training Center grant P60DK79637, the American Heart Association: 12SDG9280031 and 17GRNT33661168. In addition, this project was supported by National Research Initiative Competitive Grant no. 2007-35205-17883 from the USDA National Institute of Food and Agriculture. We gratefully thank our Amish community and research volunteers for their long-standing partnership in research, and acknowledge the dedication of our Amish liaisons, field workers and the Amish Research Clinic staff, without which these studies would not have been possible.
Nicholas J Timpson N.J.T. NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 102215/2/13/2), is supported by the University of Bristol NIHR Biomedical Research Centre (BRC) and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A19169).
Daniela Toniolo D.T. We thank all the participants to the Val Borbera project, prof. Clara Camaschella who coordinated the data collection, Corrado Masciullo and Massimiliano Cocca for the database informatics. The research was supported by funds from Compagnia di San Paolo, Torino, Italy; Fondazione Cariplo, Italy; Ministry of Health, Ricerca Finalizzata 2011-2012 and Public Health Genomics Project 2010.
David-Alexandre Tregouet D.-A.T. The MARTHA genetics project is supported by the GENMED Laboratory of Excellence on Medical Genomics (ANR-10-LABX-0013), the French Clinical Research Infrastructure Network on Venous Thrombo-Embolism (F-CRIN INNOVTE) and the ICAN Institute for Cardiometabolism and Nutrition (ANR-10-IAHU-05), three research programs managed by the National Research Agency (ANR) as part of the French Investment for the Future initiative.

Tiinamaija Tuomi T.T. The Botnia studies (T.T) Folkhälsan Research Foundation, The Academy of Finland (grant no. 312072), Ollqvist Foundation, Swedish Cultural Foundation in Finland, Finnish Diabetes Research Foundation, Foundation for Life and Health in Finland, Signe and Ane Gyllenberg Foundation, Finnish Medical Society, Paavo Nurmi Foundation, Helsinki University Central Hospital Research Foundation, Perklén Foundation, Närpes Health Care Foundation and Ahokas Foundation. The study has also been supported by the Ministry of Education in Finland, Municipal Heath Care Center and Hospital in Jakobstad and Health Care Centers in Vasa, Närpes and Korsholm. The skillful assistance of the Botnia Study Group is gratefully acknowledged.
Peter Vollenweider P.V. CoLaus (Cohorte Lausannoise): The CoLaus study was and is supported by research grants from GlaxoSmithKline, the Faculty of Biology and Medicine of Lausanne, and the Swiss National Science Foundation (grants 33CSCO-122661, 33CS30-139468 and 33CS30-148401).
Carol A Wang C.A.W. This study was supported by the National Health and Medical Research Council of Australia [grant numbers 572613, 403981 and 003209] and the Canadian Institutes of Health Research [grant number MOP-82893]. The authors are grateful to the Raine Study participants and their families, and to the Raine Study research staff for cohort coordination and data collection. The authors gratefully acknowledge the NH&MRC for their long term contribution to funding the study over the last 29 years and also the following Institutions for providing funding for Core Management of the Raine Study: The University of Western Australia (UWA), Curtin University, Raine Medical Research Foundation, UWA Faculty of Medicine, Dentistry and Health Sciences, The Telethon Institute for Child Health ResearchKids Institute, and Women and Infants Research Foundation (King Edward Memorial Hospital), Murdoch University, The University of Notre Dame (Australia), and Edith Cowan University.. The authors gratefully acknowledge the assistance of the Western Australian DNA Bank (National Health and Medical Research Council of Australia National Enabling Facility). This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia
David R Weir D.R.W. NIA U01AG009740
John Wright J.W. BiB receives core infrastructure funding from Wellcome (WT101597MA), a joint grant from the Medical Research Council (MR/N024397/1)and the National Institute for Health Research Collaboration for Applied Health Research and Care for Yorkshire and Humber.
Markus Perola M.Perola The EU FP7 under grant agreement HZ2020 633589 (Ageing with Elegans)
Markus Perola M.Perola The Finnish Academy grant no. 269517
Markus Perola M.Perola The Yrjö Jahnsson Foundation
Markus Perola M.Perola The Finnish Foundation for Cardiovascular Research
Patrik KE Magnusson P.K.M. The Swedish Twin Registry is managed by Karolinska Institutet and receives funding through the Swedish Research Council under the grant no 2017-00641.
Daniel I Chasman D.I.C. The WGHS is supported by the National Heart, Lung, and Blood Institute (HL043851 and HL080467) and the National Cancer Institute (CA047988 and UM1CA182913) with funding for genotyping provided by Amgen.
Ruth J F Loos R.J.F.L. Ruth Loos is supported by the National Institutes of Health [R01 DK107786, R01 DK110113, U01HG007417].
Ruth J F Loos R.J.F.L. The Mount Sinai BioMe Biobank is supported by The Andrea and Charles Bronfman Philanthropies.
Nora Franceschini N.F. The ARIC study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number

UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. NF is supported by the National Institute of Health R01-MD012765,R21-HL140385,R21-HL123677 and R56-DK104806.
Caroline Hayward C.H. We would like to acknowledge the contributions of Professor Pavao Rudan and staff of the Institute for Anthropological Research in Zagreb, the Croatian Centre for Global Health, University of Split, the recruitment teams in Vis and Korcula, the administrative teams in Croatia and Edinburgh and the people of Vis and Korcula. The SNP genotyping for the CROATIA_Korcula cohort was performed either by Helmholtz Zentrum München, Neuherberg, Germany or the Clinical Research Facility, University of Edinburgh. The SNP genotyping for the CROATIA_Vis cohort was performed by the Clinical Research Facility, University of Edinburgh. The work was funded by the Medical Research Council (UK) core funding to the QTL in Health and Disease programme, the European Union framework program 6 EUROSPAN project (contract no. LSHG-CT-2006-018947) and the Ministry of Science, Education and Sport in the Republic of Croatia (number 108-1080315-0302).
Tõnu Esko T.E. EGCUT was supported by Estonian Research Council [IUT20-60, IUT24-6, PUT1660 to T.E and PUT1665 to K.F.; European Union Horizon 2020 [692145]; European Union through the European Regional Development Fund [2014-2020.4.01.15-0012 GENTRANSMED];

SUPPLEMENTARY NOTE 3: The SIGMA Type 2 Diabetes Genetics Consortium members

Genetic analyses: Josep M. Mercader[1, 2, 3], Alicia Huerta-Chagoya[4], Humberto García-Ortiz[5], Hortensia Moreno-Macías[4, 6], Alisa Manning[3, 7, 8], Lizz Caulkins[3], Noël P. Burtt[3], Jason Flannick[3, 9], Nick Patterson[10], Carlos A. Aguilar-Salinas[4], Teresa Tusié-Luna[4, 11], David Altshuler[3, 9, 12], Jose C. Florez[1, 3, 8, 13]

Study cohorts: *Diabetes in Mexico Study:* Humberto García-Ortiz[5], Angélica Martínez-Hernández[5], Federico Centeno-Cruz[5], Francisco Martin Barajas-Olmos[5], Carlos Zerrweck[14], Cecilia Contreras-Cubas[5], Elvia Mendoza-Caamal[5], Cristina Revilla-Monsalve[15], Sergio Islas-Andrade[15], Emilio Córdova[5], Xavier Soberón[5], Lorena Orozco[5]. *Mexico City Diabetes Study:* Clicerio González-Villalpando[16], María Elena González-Villalpando[16]. *Multiethnic Cohort Study:* Christopher A. Haiman[17], Lynne Wilkens[18], Loic Le Marchand[18], Kristine Monroe[17], Laurence Kolonel[18]. *UNAM/INCMNSZ Diabetes Study:* Olimpia Arellano-Campos[4], Alicia Huerta-Chagoya[4], Maria L. Ordóñez-Sánchez[4], Maribel Rodríguez-Torres[4], Yayoi Segura-Kato[4], Rosario Rodríguez-Guillén[4], Ivette Cruz-Bautista[4], Linda Liliana Muñoz-Hernandez[4], Tamara Sáenz[4], Donají Gómez[4], Ulices Alvirde[4], Paloma Almeda-Valdés[4], Hortensia Moreno-Macías[4, 6], Teresa Tusié-Luna[4, 11], Carlos A. Aguilar-Salinas[4]

Scientific and project management: Noël P. Burtt[3], Lizz Caulkins[3], Maria L. Cortes[10]

Steering committee: David Altshuler[3, 9, 12], Jose C. Florez[1, 3, 8, 13], Christopher A. Haiman[17], Carlos A. Aguilar-Salinas[4], Clicerio González-Villalpando[16], Lorena Orozco[5], Teresa Tusié-Luna[4, 11]

[1] Diabetes Unit and Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA.

[2] Joint BSC-CRG-IRB Research Program in Computational Biology. Barcelona Supercomputing Center, 08034 Barcelona.

[3] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, 02142, USA.

[4] Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Sección XVI, Tlalpan, 14000 Mexico City, Mexico.

[5] Instituto Nacional de Medicina Genómica, Tlalpan, 14610, Mexico City, Mexico.

[6] Universidad Autónoma Metropolitana, Tlalpan 14387, Mexico City, Mexico.

[7] Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA.

[8] Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA.

[9] Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA.

[10] Broad Institute of Harvard and MIT, Cambridge, Massachusetts, 02142, USA.

[11] Instituto de Investigaciones Biomédicas, UNAM Unidad de Biología Molecular y Medicina Genómica, UNAM/INCMNSZ, Coyoacán, 04510 Mexico City, Mexico.

[12] Department of Genetics, Harvard Medical School, Boston, Massachusetts, 02115, USA.

[13] Metabolism Program, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, 02142, USA.

[14] Clínica de Integral de Cirugía para la Obesidad y Enfermedades Metabólicas, Hospital General Tláhuac, Secretaría de Salud del GDF. México City.

[15] Instituto Mexicano del Seguro Social SXXI, Mexico City, Mexico.

[16] Centro de Estudios en Diabetes, Unidad de Investigacion en Diabetes y Riesgo Cardiovascular, Centro de Investigacion en Salud Poblacional, Instituto Nacional de Salud Publica, Mexico City, Mexico.

[17] Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, 90033, USA.

[18] Epidemiology Program, University of Hawaii Cancer Center, Honolulu, Hawaii, USA.

SUPPLEMENTARY NOTE 4: Comparison of inbreeding coefficient estimates

Introduction. Accurate estimation of the Inbreeding Depression (β_F) requires precise and unbiased estimates of individual inbreeding coefficients (F). Numerous methods have been proposed for estimating inbreeding coefficients from dense SNP marker data. Broadly, these can be grouped into two categories: methods that consider the excess homozygosity at a large number of, preferably independent, markers (e.g. F_{HOM}, F_{UNI}) and methods that identify autozygous genomic segments from unbroken tracts of homozygous genotypes, which are unlikely to occur by chance (e.g. F_{ROH}).

To maximize statistical power and minimize bias in $\hat{\beta}_F$, an estimator of F should be both unbiased, ($E[\hat{F}] = F$) and precise, $MSE(\hat{F} - F) \ll var(F)$. Extensive comparisons of different estimators have been made by previous studies. Many of these conclude that F_{ROH} calculated from appropriately parameterized ROH calling gives estimates of F with minimal bias and lower variance than independent SNP methods[1–4]. In contrast, Yengo et al. (PNAS 2017) claim that $\hat{\beta}_{F_{\mathrm{ROH}}}$ may be upwardly biased by as much as 162%, and provide apparent evidence of this in both simulated and real data[5]. Furthermore, Yengo et al. show, in both theory and simulation, that two independent SNP methods (F_{UNI} and F_{HOM}) give unbiased estimates of β_F when causal SNPs are a random subset of all genotyped SNPs. This would cast doubt on the validity of using F_{ROH} to estimate β_F, and suggests F_{UNI} might be a more appropriate measure. To understand the apparent contradictions between different studies, we have repeated and extended the investigations described in Yengo et al.

Yengo et al. base their study on genotype data derived from the first phase of the UK Biobank (UKB) imputation. 9,493,148 SNPs with a minor allele frequency (MAF) of >1%, INFO score of > 0.3 and

HWE p-value > 1e-06 were selected from the imputation of 140,720 British individuals. SNP dosages were rounded to the nearest whole genotype and LD pruning r^2> 0.9 was performed to reduce the number of SNPs to 3,857,369. F_{UNI} was calculated by the formula presented by Yang et al. (2011)[6] as $\hat{F}^{III}$ (implemented in PLINK[7] by parameters --ibc Fhat3) and ROH were called by two PLINK parameterizations, including the values proposed in Joshi et al. (2015)[8]. Joshi et al. validated the PLINK parameters for moderately dense SNP chips, but not for the different characteristics (SNP density, error rate) of genotypes called from imputed dosages. In particular, the PLINK method allows only one heterozygote or 5 missing genotypes within each ROH.

F_{ROH} calculated from SNP chip genotypes agrees well with F_{UNI} indicating minimal bias. We followed the method described in Yengo et al. to calculate both F_{UNI} and F_{ROH} from the UKB imputation, but initially compared Yengo's F_{UNI} to our F_{ROH} calculated from SNP chip data available for the same phase one individuals. From now on, we use lowercase roh to refer to measures based on imputed data and uppercase ROH to refer to those from SNP-chip genotypes. We find good correspondence between F_{UNI} and F_{ROH} (Supplementary Fig. 12). Since F_{UNI} is believed to be an unbiased estimator of F, the correspondence between F_{UNI} and F_{ROH} places limits on the possible bias of F_{ROH}. Specifically, if we model F_{ROH} as $F_i^{\mathrm{ROH}} = \Theta F_i + \epsilon_i$, then

$$\frac{r^2}{\beta_{F_{\mathrm{UNI}},F_{\mathrm{ROH}}}} \leq \Theta \leq \frac{1}{\beta_{F_{\mathrm{UNI}},F_{\mathrm{ROH}}}} \qquad (19)$$

(see Supplementary note 6). Substituting the regression values from Supplementary Fig. 12 gives $0.94 \leq \Theta \leq 0.95$. I.e. the good correspondence between F_{UNI} and F_{ROH} limits the potential error in F_{ROH} to a small downward bias. This very small underestimate of F may be caused by autozygosity not captured in F_{ROH}, namely ROH of less than 1.5 Mb length or ROH in sparsely genotyped regions.

Calculating F_{roh} from imputed genotypes can introduce large downward bias. To call ROH from the UKB imputation we followed the method described by Yengo et al., but identified only 9,067,605 SNPs (out of a total of 72,355,667) matching their inclusion criteria (MAF > 1%, INFO > 0.3, HWE exact p-value > 1e-6). We then used PLINK 2.0 to convert the imputed genotype probabilities to the hard-called genotypes required for LD pruning and ROH calling. PLINK first converts genotype probabilities to an estimated dosage which is then rounded to the nearest genotype ϵ [0 1 2]. By default, a genotype is recorded if the estimated dosage is within 0.15 of any of [0 1 2], otherwise the genotype is recorded as missing. Using these default parameters introduces 3.9% missingness into our dataset. We nevertheless proceeded to LD-prune this dataset, both with these missing data (method 2), and having removed all SNPs with a missing fraction > 0.03 (method 1). However, Yengo et al. state that *Imputed SNPs were called to the genotypes having the largest posterior probability* which is achieved by changing the PLINK hardcall parameter from 0.15 to 0.499999 (method Yengo). After LD pruning 3,061,484 SNPs remain for 141,774 British individuals.

For all three methods of data preparation, we called ROH and calculated F_{roh} using the Joshi et al. PLINK parameters also used by Yengo et al. We find that calling ROH from hard-called imputed dosages (method Yengo) gives F_{roh} with an expected value of just 0.39 of the F_{ROH} obtained from SNP chip genotypes (Supplementary Figure 13). More stringent treatments of uncertain genotype probabilities (methods 2 and 1) give progressively less downwards bias. To understand the cause of this downward bias in F_{roh} we plotted the genome wide distribution of ROH for two high F individuals, highlighted in orange in Supplementary Figure 13.

For these two individuals all ROH called from SNP chip genotypes (in blue) and the Yengo et al. imputed data method (in red) are shown in Supplementary Figs 14a,b. Individual 1 has an $F_{\mathrm{ROH}} =$

0.261, amongst the highest observed in UKB, and is most likely the progeny of 1st degree relatives where $E[F] = 0.25$. Individual 2 has $F_{\mathrm{ROH}} = 0.0626$ and is most likely the offspring of first cousins (3nd degree relatives) where $E[F] = 0.0625$. Calling ROH from hard-called imputed dosages fragments, and consequently fails to identify, many of the ROH found in SNP chip data. The resultant downward bias in F_{roh} is sufficient to explain an upward bias of up to 156% (1/0.39) in $\hat{\beta}_{F_{\mathrm{roh}}}$.

In summary, calculating F_{ROH} from dense SNP chip genotypes, with the parameters used in this study, gives valid estimates of inbreeding coefficients. In contrast, calculating F_{roh} from unfiltered imputed genotypes, as done in Yengo et al., introduces a large bias which appears to be responsible for the poor performance of F_{roh} in that study.

$\hat{\boldsymbol{\beta}}_{F_{\mathbf{GRM}}}$ is downwardly biased in real data. Yengo et al. also show, in both theory and simulation, that $\hat{\beta}_{F_{\mathrm{UNI}}}$ is an unbiased estimate of β_F in certain conditions, for example, when causal SNPs are a random subset of all observed SNPs. In real UKB data they find that $\hat{\beta}_{F_{\mathrm{roh}}}$is systematically of greater magnitude than $\hat{\beta}_{F_{\mathrm{UNI}}}$, which they therefore interpret as empirical evidence that $\hat{\beta}_{F_{\mathrm{ROH}}}$is upwardly biased. We have already shown that calling ROH from imputed data may cause an upward bias of $\hat{\beta}_{F_{\mathrm{roh}}}$, however, interestingly, we also observe that estimates obtained from unbiased SNP-chip genotypes ($\hat{\beta}_{F_{\mathrm{ROH}}}$) are systematically larger than estimates obtained from frequency-based measures ($\hat{\beta}_{F_{\mathrm{SNP}}}$ and $\hat{\beta}_{F_{\mathrm{GRM}}}$) (Supplementary Data Table 13). Note, we use the nomenclature F_{GRM} to refer the $\hat{F}^{III}$ calculation used in the ROHgen consortium. Although F_{GRM} and F_{UNI} are identical calculations (PLINK –ibc Fhat3), F_{GRM} is calculated from SNP-chip genotypes with a minimum MAF of 5%, while Yengo et al. calculated F_{UNI} from hard called imputed dosages with a minimum MAF of 1%. We explain the differences between $\hat{\beta}_{F_{\mathrm{ROH}}}$ and $\hat{\beta}_{F_{\mathrm{GRM}}}$ below.

Causal variants for Inbreeding Depression are not in strong LD with common SNPs. For all traits, we fit bivariate models with F_{ROH} and F_{GRM} as explanatory variables. For all 32 traits that were significant in the univariate analysis, we find that $\hat{\beta}_{F_{\mathrm{ROH}}|F_{\mathrm{GRM}}}$ is of greater magnitude than $\hat{\beta}_{F_{\mathrm{GRM}}|F_{\mathrm{ROH}}}$in the conditional analysis (Supplementary Data Table 22). Furthermore, for 30 of these traits $\hat{\beta}_{F_{\mathrm{GRM}}|F_{\mathrm{ROH}}}$ does not differ significantly from zero. I.e., for many traits, the variation of F_{GRM} which is independent of F_{ROH} is not associated with any change in trait values. In Supplementary Note 5 we show that these results are consistent with inbreeding depression caused by rare, but not common, variants. Furthermore, we observe that the downward bias of $\hat{\beta}_{F_{\mathrm{GRM}}}$ is proportional to the ratio $\frac{var(F_{\mathrm{ROH}})}{var(F_{\mathrm{GRM}})}$ (Fig 4c), as expected when the difference between F_{GRM} and F_{ROH} can be considered as estimation error (See Supplementary Note 7).

In summary, Yengo et al showed that $\hat{\beta}_{F_{\mathrm{UNI}}}$ is unbiased when causal variants are a random subset of the observed SNPs. Although we agree with this statement, we find the evidence does not support the assumption of a random sample, but reveals the importance of rare variants, whose excess homozygosity is well predicted by F_{ROH} (Supplementary Fig 16a).

Comparison of genomic measures of inbreeding with genealogy. As a further assessment of the relative abilities of F_{ROH}, F_{SNP} and F_{GRM} to capture inbreeding, we analysed Pearson's product-moment correlations between the genomic inbreeding measures and pedigree inbreeding (F_{PED}) for 47,927 Icelanders with mostly-complete (info score > 0.6)[9] 10 generation pedigrees. To decrease the confounding effects of pedigree mis-specification, a small number of individuals (n=20) with extreme discrepancies between genetics and genealogy ($F_{\mathrm{ROH}} > 0.05$ & $F_{\mathrm{SNP}} < 0.001$) were removed. The

correlation was highest for F_{ROH} (r = 0.779), lowest for F_{SNP} (0.632) and intermediate for F_{GRM} (0.682), further validating the utility of F_{ROH} as the most accurate genomic measure of inbreeding.

SUPPLEMENTARY NOTE 5: Interpretation of $\boldsymbol{Trait \sim F_{ROH} + F_{GRM}}$ models.

Are inbreeding effects caused by rare or common variants? F_{ROH} is an estimate of autozygosity, which increases the homozygosity of all variants, both common and rare. In contrast, F_{GRM} is calculated from common SNPs (>5% MAF) and correlates well with the homozygosity of common SNPs, but less well with rare SNPs which may be in weak Linkage Disequilibrium (LD). We therefore performed bivariate models of all traits in real data ($Trait \sim F_{\mathrm{ROH}} + F_{\mathrm{GRM}}$) to establish whether the observed inbreeding effects associate more strongly with F_{ROH} or F_{GRM}. For all significant traits, we find the observed associations more attributable to F_{ROH} (Supplementary Data Table 22; Supplementary Figs 15a,b) suggesting inbreeding effects are caused by rare genetic variants. A recent study[10] found evidence for a similar conclusion, but to further support this interpretation we investigate below how both F_{ROH} and F_{GRM} predict the excess homozygosity of SNPs at a range of allele frequencies.

Relationships between $\boldsymbol{F_{ROH}}$, $\boldsymbol{F_{GRM}}$ and excess homozygosity at different allele frequencies. For any trait exhibiting inbreeding depression, the degree of depression will be related to the excess homozygosity (above Hardy-Weinberg expectation) of the causal variants. In Supplementary note 8, we show that inbreeding depression, which is equal to the sum of the dominance deviations at the causal loci, is proportional to the inbreeding coefficient (F_{QTL}) defined in equation (39) below.

$$ID_i = \sum_{i=1}^{m} \delta_i = \beta_F * F_{\mathrm{QTL}} \tag{38}$$

Where

$$F_{\mathrm{QTL}} = \frac{1}{m}\sum_{i=1}^{m} \frac{w_i\left(x_i^2 - (1+2p_i)x_i + 2p_i^2\right)}{2p_iq_i} \tag{39}$$

and

$$w_i = \frac{2p_iq_id_i}{\frac{1}{m}\sum_{i=1}^{m} 2p_iq_id_i} \tag{40}$$

We note that the unweighted form of equation (39) is identical to $\hat{F}^{III}$ introduced by Yang et al (2011)[6], and implemented in PLINK by the parameters –ibc Fhat3. This is the same formula used to calculate F_{GRM} and F_{UNI} from different sets of marker SNPs. In summary, if the causal loci and effect sizes are known, a weighted calculation of $\hat{F}^{III}$ at the causal loci is directly proportional to the degree of inbreeding depression. We have used this, below, to simplify the simulation of inbreeding depression caused by variants at specific allele frequencies.

If we imagine inbreeding depression caused exclusively by variants at one allele frequency then, in the absence of strong selection or assortative mating on the causal loci in the current generation, the expectation of F_{QTL} will be equal to $\hat{F}^{III}$ calculated at marker variants of the same allele frequency (henceforth called F_{MAF}).

To calculate F_{MAF} across a range of allele frequencies we extracted SNPs at seven frequencies (MAF=0.01, 0.025, 0.05, 0.1, 0.2, 0.4 & 0.5) from 402,559 genetically British samples in the phase 2 UKB imputation. Selected SNPs were required to have a minor allele frequency (AF) within 10% of the specified MAF (0.9*MAF < AF < 1.1*MAF) and HWE p-value > 1e-6. The numbers of SNPs retained at each MAF are reported in Supplementary Table 2. F_{ROH} and F_{GRM} had previously been calculated, from SNP-chip genotypes, as part of the ROHgen meta-analysis.

To investigate the relationships between F_{ROH}, F_{GRM} and F_{MAF} we fit univariate ($F_{\mathrm{MAF}} \sim F_{\mathrm{ROH}}$ and $F_{\mathrm{MAF}} \sim F_{\mathrm{GRM}}$) and bivariate models ($F_{\mathrm{MAF}} \sim F_{\mathrm{ROH}} + F_{\mathrm{GRM}}$) at each allele frequency. In the univariate models we find F_{ROH} to be an unbiased predictor of F_{MAF} across the entire frequency spectrum, while F_{GRM} is downwardly biased, particularly at low MAF (Supplementary Fig. 16a). Despite this downward bias, F_{GRM} is more strongly correlated than F_{ROH} at all MAF > 5% (Supplementary Figure 16b). In the bivariate model F_{GRM} is a stronger predictor of the homozygosity of common SNPs (>10%), but F_{ROH} is a stronger predictor for rare SNPs (Fig. 4d).

Observed associations consistent with the homozygosity of rare, not common, SNPs. In real data models of $Trait \sim F_{\mathrm{ROH}} + F_{\mathrm{GRM}}$, we consistently find the observed associations are preferentially attributed to F_{ROH} rather than F_{GRM} (Supplementary Data Table 22, Figure 4c, Supplementary Figs 15a,b). In light of Figure 4d, these results are compatible with the action of rare, not common, causal variants.

SUPPLEMENTARY NOTE 6: Limits of bias in F_{ROH}

If F_{UNI} is an unbiased estimate of F then it can be expressed as

$$F_{\mathrm{UNI}} = F + \varepsilon \tag{20}$$

If F_{ROH} is a potentially biased estimate of F then it can be expressed as

$$F_{\mathrm{ROH}} = \Theta F + \Theta\varepsilon' \tag{21}$$

The regression slope (β) of F_{UNI} on F_{ROH} is known, and

$$\beta = \frac{cov(F_{UNI}, F_{ROH})}{var(F_{ROH})}$$

Substituting (20) and (21) and assuming independent errors gives

$$\beta = \frac{\Theta var(F)}{\Theta^2 var(F) + \Theta^2 var(\varepsilon')} \tag{22}$$

Rearranging (22) gives

$$\Theta = \left(\frac{1}{\beta}\right)\left(1 + \frac{var(\varepsilon')}{var(F)}\right)^{-1} \tag{23}$$

The range of $\frac{var(\varepsilon')}{var(F)}$ is limited by the correlation between F_{UNI} and F_{ROH}, and we can put $\left(1 + \frac{var(\varepsilon')}{var(F)}\right)^{-1}$ in terms $\frac{var(\varepsilon)}{var(\varepsilon')}$ of by considering that

$$\frac{var(F_{\mathrm{ROH}})}{var(F_{\mathrm{GRM}})} = \frac{r^2}{\beta^2} \tag{24}$$

Again substituting (20) and (21) in equation (24) gives

$$\frac{\Theta^2 var(F) + \Theta^2 var(\varepsilon')}{var(F) + var(\varepsilon)} = \frac{r^2}{\beta^2} \tag{25}$$

Rearranging (25) gives

$$\frac{var(\varepsilon')}{var(F)} = \frac{\Theta^2\beta^2 - r^2}{r^2\frac{var(\varepsilon)}{var(\varepsilon')} - \Theta^2\beta^2} \tag{26}$$

Substituting equation (26) into equation (23)

$$\Theta^2 = \left(\frac{1}{\beta}\right)\left(\frac{r^2\frac{var(\varepsilon)}{var(\varepsilon')} - \Theta^2\beta^2}{r^2\left(\frac{var(\varepsilon)}{var(\varepsilon')} - 1\right)}\right) \tag{27}$$

As $\frac{var(\varepsilon)}{var(\varepsilon')} \to 0$, i.e. if F_{UNI} is precise and estimation errors entirely on F_{ROH} then equation (27) →

$$\Theta = \left(\frac{1}{\beta}\right)\left(\frac{\Theta^2\beta^2}{r^2}\right) \tag{28}$$

$$\Theta = \frac{r^2}{\beta} \tag{29}$$

As $\frac{var(\varepsilon)}{var(\varepsilon')} \to \infty$, i.e. if F_{ROH} is precise and estimation errors entirely on F_{UNI} then equation (27) →

$$\Theta = \frac{1}{\beta} \tag{30}$$

Therefore, from the bounds of $\frac{var(\varepsilon)}{var(\varepsilon')}$ and equations (29) and (30)

$$\frac{r^2}{\beta_{F_{\mathrm{UNI}},F_{\mathrm{ROH}}}} \leq \Theta \leq \frac{1}{\beta_{F_{\mathrm{UNI}},F_{\mathrm{ROH}}}} \tag{31}$$

SUPPLEMENTARY NOTE 7: Expected attenuation bias in $\widehat{\boldsymbol{\beta}}_{\boldsymbol{F_{\mathrm{GRM}}}}$

If F_{GRM} varies around F_{ROH} and the difference (ε) has no effect on the trait (y) then

$$F_{\mathrm{GRM}} = F_{\mathrm{ROH}} + \varepsilon \tag{32}$$

And

$$\beta_{F_{\mathrm{GRM}}} = \frac{cov(F_{\mathrm{GRM}}, y)}{var(F_{\mathrm{GRM}})} \tag{33}$$

$$\beta_{F_{\mathrm{GRM}}} = \frac{cov(F_{\mathrm{ROH}} + \varepsilon, y)}{var(F_{\mathrm{GRM}})} \tag{34}$$

Because ε has no effect on the trait (y)

$$\beta_{F_{\mathrm{GRM}}} = \frac{cov(F_{\mathrm{ROH}}, y)}{var(F_{\mathrm{GRM}})} \tag{35}$$

$$\beta_{F_{\mathrm{GRM}}} = \beta_{F_{\mathrm{ROH}}} * \frac{var(F_{\mathrm{ROH}})}{var(F_{\mathrm{GRM}})} \tag{36}$$

$$\frac{\beta_{F_{\mathrm{GRM}}}}{\beta_{F_{\mathrm{ROH}}}} = \frac{var(F_{\mathrm{ROH}})}{var(F_{\mathrm{GRM}})} \tag{37}$$

SUPPLEMENTARY NOTE 8: Calculation of F_{QTL} at known causal loci.

If $x_i \epsilon [0,1,2]$ is the number of copies of the reference allele at locus i of m causal loci, then the number of reference homozygotes at the locus is $\frac{x_i(x_i-1)}{2}$, the number of heterozygotes is $-x_i(x_i-2)$, and the number of alternate homozygotes is $\frac{(x_i-2)(x_i-1)}{2}$.

If p_i is the frequency of the reference allele, and q_i is the frequency of the alternate allele, then the inbreeding depression with complete inbreeding (β) is

$$\beta = -\sum_{i=1}^{m} 2p_i q_i d_i \tag{41}$$

Where d_i is the difference between the heterozygote and mean homozygote value. We wish to define an inbreeding coefficient (F_{QTL}) which is directly proportional to realised inbreeding depression (the sum of the dominance deviations). I.e.

$$\beta F_{\mathrm{QTL}} = \sum_{i=1}^{m} \delta_i \tag{42}$$

The dominance deviations (δ_i) for the three genotypes at a locus can be written in terms of d_i: $\delta_i \epsilon \left[-2q_i^2 d_i, 2p_i q_i d_i, -2p_i^2 d_i\right]$. Substituting these dominance deviations and the genotype counts into equation (42) gives

$$\beta F_{\mathrm{QTL}} = \sum_{i=1}^{m} -\frac{x_i(x_i-1)}{2} 2q_i^2 d_i - x_i(x_i-2)2p_i q_i d_i - \frac{(x_i-2)(x_i-1)}{2} 2p_i^2 d_i \tag{43}$$

Rearranging equation (43) gives

$$\beta F_{\mathrm{QTL}} = \sum_{i=1}^{m} -d_i\left(x_i^2 - (1+2)p_i x_i + 2p_i^2\right) \tag{44}$$

Substituting for β from equation (41) gives

$$F_{\mathrm{QTL}} = \frac{1}{m}\sum_{i=1}^{m} w_i \frac{x_i^2 - (1+2p_i)x_i + 2p_i^2}{2p_i q_i} \tag{45}$$

Where

$$w_i = \frac{2p_i q_i d_i}{\frac{1}{m}\sum_{i=1}^{m} 2p_i q_i d_i} \quad (46)$$

SUPPLEMENTARY REFERENCES

1. McQuillan, R. *et al.* Runs of Homozygosity in European Populations. *American Journal of Human Genetics* **83**, 359–372 (2008).

2. Keller, M. C., Visscher, P. M. & Goddard, M. E. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* **189**, 237–249 (2011).

3. Kardos, M., Nietlisbach, P. & Hedrick, P. W. How should we compare different genomic estimates of the strength of inbreeding depression? *Proceedings of the National Academy of Sciences* **115**, E2492–E2493 (2018).

4. Gazal, S. *et al.* Inbreeding coefficient estimation with dense SNP data: Comparison of strategies and application to HapMap III. *Human Heredity* **77**, 49–62 (2014).

5. Yengo, L. *et al.* Detection and quantification of inbreeding depression for complex traits from SNP data. *Proceedings of the National Academy of Sciences* **114**, 8602–8607 (2017).

6. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics* **88**, 76–82 (2011).

7. Purcell, S. & Chang, C. PLINK 1.9. Available at: www.cog-genomics.org/plink/1.9/.

8. Joshi, P. K. *et al.* Directional dominance on stature and cognition in diverse human populations. *Nature* **523**, 459–462 (2015).

9. Helgason, A., Pálsson, S., Guobjartsson, D. F., Kristjánsson, P. & Stefánsson, K. An association between the kinship and fertility of human couples. *Science* **319**, 813–816 (2008).

10. Johnson, E. C., Evans, L. M. & Keller, M. C. Relationships between estimated autozygosity and complex traits in the UK Biobank. *PLoS Genetics* **14**, (2018).