

Published in final edited form as:

*Nat Genet.* 2018 September ; 50(9): 1304–1310. doi:10.1038/s41588-018-0178-9.

## Relatedness disequilibrium regression estimates heritability without environmental bias

Alexander I. Young<sup>1,2,3,\*</sup>, Michael L. Frigge<sup>1</sup>, Daniel F. Gudbjartsson<sup>1,4</sup>, Gudmar Thorleifsson<sup>1</sup>, Gyda Bjornsdottir<sup>1</sup>, Patrick Sulem<sup>1</sup>, Gisli Masson<sup>1</sup>, Unnur Thorsteinsdottir<sup>1,5</sup>, Kari Stefansson<sup>1,5</sup>, and Augustine Kong<sup>1,3,4,\*</sup>

<sup>1</sup>deCODE genetics/Amgen Inc., Reykjavik 101, Iceland <sup>2</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, U.K <sup>3</sup>Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK <sup>4</sup>School of Engineering and Natural Sciences, University of Iceland, Reykjavik 101, Iceland <sup>5</sup>Faculty of Medicine, University of Iceland, Reykjavik 101, Iceland

### Abstract

Heritability measures the proportion of trait variation that is due to genetic inheritance. Measurement of heritability is of importance to the nature-versus-nurture debate. However, existing estimates of heritability could be biased by environmental effects. Here we introduce relatedness disequilibrium regression (RDR), a novel method for estimating heritability. RDR avoids most sources of environmental bias by exploiting variation in relatedness due to random Mendelian segregation. We use a sample of 54,888 Icelanders with both parents genotyped to estimate the heritability of 14 traits, including height (55.4%, S.E. 4.4%) and educational attainment (17.0%, S.E. 9.4%). Our results suggest that some other estimates of heritability could be inflated by environmental effects.

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Correspondence to: alexander.young@bdi.ox.ac.uk or kong@decode.is.

#### Data Availability

The authors declare that the Icelandic data supporting the findings of this study are available within the article, its supplementary information files and upon request. Applications for access to the UK Biobank data can be made on the UK Biobank website: <http://www.ukbiobank.ac.uk/register-apply/>.

#### Code Availability

Code used for estimating heritability by RELT-SNP and Sib-Regression is freely available under an MIT license at <https://github.com/AlexTISYoung/RDR>.

**Author Contributions** A.I.Y. conceived and designed the study, performed statistical analyses, contributed analysis tools, developed theoretical results, and wrote the paper. M.L.F. performed statistical analyses and contributed analysis tools. D.F.G. contributed analysis tools, processed raw genotype/sequencing data, and collected and processed phenotype data. G.T. contributed analysis tools, and collected and processed phenotype data. G.B. collected and processed phenotype data. P.S. collected and processed phenotype data. G.M. processed raw genotype/sequence data. U.T. supervised generation of genotype/sequence data and phenotype data. K.S. jointly supervised research and wrote the paper. A.K. conceived and designed the study, jointly supervised research, and wrote the paper.

**Competing Financial Interests** The authors affiliated with deCODE Genetics are/were employed by the company, which is owned by Amgen, Inc: A.I.Y., M.L.F., D.F.G., G.T., G.B., P.S., U.T., K.S., A.K.

## Introduction

Heritability measures the proportion of trait variation in a population that is due to genetic inheritance. Estimation of the relative importance of genetic inheritance (nature) versus environment (including nurture) has generated much controversy<sup>1</sup>. Historically, most estimates of heritability for human traits have come from twin studies<sup>2,3</sup>. Some more recent methods propose to estimate heritability by modelling the effects of genome-wide single nucleotide polymorphisms (SNPs)<sup>4</sup>. We refer to these methods as GREML-SNP methods, referencing inference on genomic relatedness (GR), estimated from SNPs, using restricted maximum likelihood (REML). In order to reduce the influence of non-additive genetic effects and environmental effects, samples are pruned so that no pair is related above some low threshold<sup>4</sup>, typically 0.025 or 0.05.

Instead of modelling the effects of SNPs, heritability can be estimated by examining how phenotypic similarity changes with relatedness. Relatedness is measured by the fraction of the genome a pair shares in segments inherited from a common ancestor, called IBD (identical-by-descent) segments. (We note that what we call ‘relatedness’ here has sometimes been termed ‘realised relatedness’ to distinguish it from expected relatedness given a pedigree<sup>5</sup>). Sharing of an IBD segment implies sharing of all genetic variants in that segment, except for mutations that occurred since the last common ancestor of the segment. This implies that IBD based methods can capture nearly all of the heritability of a trait. In contrast, GREML-SNP methods can only capture the fraction of the heritability explained by genotyped SNPs<sup>4</sup>. Another advantage of IBD based methods over GREML-SNP methods is that they do not make assumptions about the distribution of SNP effect sizes. Violation of these assumptions has been shown to introduce bias to GREML-SNP estimates of heritability<sup>4,6</sup>.

An IBD based method, which we call the ‘Kinship’ method, examines how phenotypic similarity increases with relatedness for all pairs from a population sample<sup>7</sup>. When close relatives have more similar environments than distant relatives, the Kinship method will overestimate heritability, as it is unable to distinguish between similarity due to genetic effects and environmental effects. To reduce environmental bias, modelling of environmental effects shared between close relatives has been suggested<sup>8,9</sup>. However, environmental similarity may increase with relatedness across much of the relatedness spectrum: siblings may have more similar environments than cousins, and so on, down to distant relatives. In this case, modelling environmental covariance between close relatives alone will not remove environmental bias from the Kinship method. While an extension to the Kinship method has been developed that models spatially distributed environmental effects<sup>10</sup>, most environmental effects do not follow a simple, spatial distribution.

A different IBD based method, which we call ‘Sib-Regression’, restricts the analysis to sibling pairs<sup>5</sup>. There are two copies of each piece of DNA in each parent. Whether a sibling inherits one or the other copy of a piece of DNA from a parent is like the outcome of a fair coin toss. The coin toss represents the outcome of random Mendelian segregation of DNA in the parent during meiosis. Whether both siblings inherit the same copy of a piece of DNA is like whether two independent tosses of a fair coin both come out the same. Therefore, the

siblings inherit the same copy of DNA from a parent half of the time on average. Most of the variation around the average relatedness is due to random segregations in the parents of the siblings. The random segregations are independent of almost all environmental effects. Sib-Regression therefore avoids most sources of environmental bias. However, Sib-Regression requires hundreds of thousands of genotyped sibling pairs to obtain precise heritability estimates, whereas existing applications have used ~20,000 sibling pairs or less<sup>5,11</sup>.

Here we introduce a novel method for estimating heritability, relatedness disequilibrium regression (RDR). RDR looks at how much more or less related a pair is than would be expected from the relatedness of the parents. This deviation we call ‘relatedness disequilibrium’. Relatedness disequilibrium is due to random Mendelian segregations in the parents during meiosis, so is independent of almost all environmental effects. Unlike Sib-Regression, RDR can use any pair of individuals, provided there is genetic information on the parents of the pair. By using all pairs from a large sample with both parents genotyped, RDR can obtain precise estimates of heritability with negligible bias due to environment. We apply RDR to estimate heritability for 14 quantitative traits in Iceland.

## Results

### Defining heritability through random segregation

We first distinguish direct genetic effects and indirect genetic effects: a direct genetic effect is the effect of genetic material in a body on that body, whereas an indirect genetic effect is the effect on another body (Supplementary Note)<sup>12–14</sup>. For example, if parenting affects the educational attainment of offspring, then there could be indirect genetic effects from parent to offspring, which we term ‘parental genetic nurturing effects’<sup>14</sup>. Any allele inherited by the phenotyped individual (proband) was also present in one of its parents, implying the allele can have both direct and parental genetic nurturing effects on the proband. However, parental genetic nurturing effects, and other indirect genetic effects, are environmental effects from the perspective of the individual whose trait is affected. The heritability of the trait is thus defined as the fraction of trait variation in the population that is explained by direct genetic effects alone.

To separate variation due to direct genetic effects (heritability) from variation explained by the environment, we use random segregation during meiosis. This approach is analogous to the transmission disequilibrium test (TDT) for a direct genetic effect of an allele on a phenotype<sup>15–17</sup>. Proband genotype is determined by the genotypes of the proband’s parents and random segregations. The TDT looks for an association between the phenotype and the variation in proband genotype caused by random segregations in the parents. This separates association due to direct genetic effects from association due to environment. Similarly, by using random segregation, phenotypic variation can be decomposed into variation due to direct genetic effects alone and other components. Assuming direct genetic effects are additive and there is no gene-by-environment interaction, the decomposition is (Supplementary Note):

$$\text{Var}(Y) = v_g + v_{e \sim g} + c_{g,e} + \text{Var}(\epsilon); \quad (1)$$

where  $v_g$  is the variance explained by direct genetic effects, and  $h^2 = v_g/\text{Var}(Y)$  is the heritability;  $v_{e\sim g}$  is the variance of the part of the environmental component of the phenotype that is correlated with parental genotype, which includes the variance explained by (additive) parental genetic nurturing effects;  $c_{g,e}$  is the covariance between direct genetic effects and environmental effects; and  $\text{Var}(\epsilon)$  is the variance of the component of the phenotype that is uncorrelated with both proband genotype and parental genotype.

### RDR covariance model

The variance decomposition (1) leads to a decomposition of the covariance matrix of a vector of observations of a phenotype,  $Y$ . Under certain assumptions (Supplementary Note):

$$\text{Cov}(Y) = v_g R + v_{e\sim g} R_{\text{par}} + c_{g,e} R_{\text{o,par}} + \text{Cov}(\epsilon), \quad (2)$$

where  $[R]_{ij}$  is the relatedness of individual  $i$  and individual  $j$ ,  $[R_{\text{par}}]_{ij}$  is the relatedness of the parents of  $i$  and the parents of  $j$ ;  $[R_{\text{o,par}}]_{ij}$  is the relatedness of  $i$  and the parents of  $j$  and  $j$  and the parents of  $i$  (Online Methods). In general,  $\text{Cov}(\epsilon)$  is unknown and can be similar to  $R$ . For example, family environment effects that are independent of genetics will cause closely related pairs to be more similar than distantly related pairs. Furthermore, pairs that are more related than the average are more likely to be from the same region and thereby have more similar environments<sup>10</sup>.

To fit the RDR covariance model, we make the simplifying assumption that  $\text{Cov}(\epsilon) = \sigma^2 I$ . Importantly, violation of the assumption that  $\text{Cov}(\epsilon) = \sigma^2 I$  does not introduce bias to RDR estimates of heritability, as we outline below.

### Environmental bias properties of RDR

By using random segregation, both RDR and the TDT separate direct genetic effects from environmental effects. The TDT achieves this by conditioning on parental genotype, whereas RDR achieves this by conditioning on parental relatedness. The expectation of offspring genotype given its parents' genotypes is one half of the sum of the parents' genotypes, and any variation around this expectation comes from random segregation. Similarly, the expectation of offspring relatedness,  $[R]_{ij}$ , given parental relatedness,  $[R_{\text{par}}]_{ij}$ , is  $[R_{\text{par}}]_{ij}/2$ , and any variation around this expectation comes from random segregation (Figure 1, Supplementary Figure 1, and Supplementary Note). (Note that this relationship does not hold for pairs where one is the direct ancestor of the other, such as parent-offspring pairs.)

By fitting  $R$  and  $R_{\text{par}}$  jointly, RDR uses the variation in  $[R]_{ij}$  around its expectation,  $[R_{\text{par}}]_{ij}/2$ , to estimate heritability. We call this variation relatedness disequilibrium. For a pair, relatedness disequilibrium is caused by random segregations in the parents of the pair, so is independent of sharing of all environmental effects apart from indirect genetic effects between the pair. This insight forms the basis of a mathematical proof that heritability estimates from RDR converge to the true heritability, when the sample excludes pairs that have indirect genetic effects on each other and excludes pairs where one is the direct ancestor of the other (Supplementary Note). If indirect genetic effects are restricted to close

relatives, the bias is likely to be small for RDR because close relatives comprise only a small fraction of the pairs in a large population sample. The bias due to indirect genetic effects could be much larger for methods that rely on close relatives, such as Sib-Regression and twin studies.

Pairs where one is the direct ancestor of the other can introduce bias because they have an atypical relationship between  $[R]_{ij}$  and  $[R_{\text{par}}]_{ij}$  (Figure 1). However, they will comprise a small fraction of the total pairs in a large population sample, even if multiple generations are genotyped. For our sample, around 30% also have a parent or grandparent in our sample, but parent-offspring and grandparent-grandchild pairs comprise only 0.0014% of all pairs. In simulations, we could not detect bias due to inclusion of parent-offspring and grandparent-grandchild pairs (Online Methods and Supplementary Table 1), so we did not remove individuals from our sample that also have a parent or grandparent in our sample.

### Simulation of RDR heritability estimation

We tested RDR for simulated traits in our sample and compared RDR to Sib-Regression, the Kinship method, and the Kinship method allowing for an effect of shared family environment, which we call the ‘Kinship F.E.’ method. We determined whether pairs shared a family environment by whether they shared a mother according to the deCODE Genealogy Database. The modelling of the environment in the Kinship F.E. model is similar to a recently proposed extension of the Kinship model<sup>8</sup>. We randomly selected 10,000 SNPs to act as causal SNPs for our simulations (Online Methods). The SNPs had a minimum minor allele frequency (MAF) of 0.5% and median MAF of 22.8%. We simulated traits in a random subsample of 10,000 individuals with both parents genotyped for all the methods other than Sib-Regression, where we used all 54,888 individuals with both parents genotyped.

We first confirmed that heritability estimates for all the methods were approximately unbiased for traits determined by additive, direct genetic effects and random noise (‘additive’ trait, Table 1, Supplementary Tables 2 and 3).

We simulated a trait where individuals who shared a mother shared a random environmental effect. We found that the Kinship method greatly overestimated the heritability of this trait (‘maternal environment’ trait, Table 1). However, the Kinship F.E. estimates of heritability were approximately unbiased. Both Sib-Regression and RDR estimates were approximately unbiased.

The results for the ‘maternal environment’ trait show that modelling a family environment effect can remove bias from the Kinship method in certain circumstances. However, when indirect genetic effects from relatives are present, modelling the family environment is ineffective at removing bias. To show this, we simulated a trait determined by direct genetic effects, parental genetic nurturing effects, and random noise (‘genetic nurturing’ trait, Table 1). For the simulated trait, the genetic nurturing effect of each SNP was a fixed fraction of its direct effect, generating a substantial covariance term,  $c_{g,e}$ . The variance components as a percentage of the phenotypic variance were:  $v_g = 40\%$ ,  $v_{e\sim g} = 10\%$ , and  $c_{g,e} \approx 28\%$ , bringing the total variance explained by parent and offspring genotype to  $\sim 78\%$ .

We found that the Kinship method greatly overestimated the heritability of the ‘genetic nurturing’ trait. Modelling of the family environment only slightly reduced the bias, with the Kinship F.E. estimates of heritability over twice the true value. The reason for this is that parental genetic nurturing effects induce correlations between all pairs with non-zero parental relatedness, not just those that share a family environment. This leads to an increase in environmental similarity with relatedness across the relatedness spectrum.

We simulated a trait affected by population stratification. For this trait, each region of Iceland had a different mean trait value (Supplementary Note). We found that the Kinship and Kinship F.E. estimates of heritability were upwardly biased when adjusting for 20 genetic principal components (‘regional’ trait, Table 1). Adjusting for 100 principal components, the mean Kinship F.E. heritability estimate was 57.6% (S.E. 0.21%), still considerably larger than the true heritability, 40%. In contrast, RDR estimates were approximately unbiased. This is because relatedness disequilibrium is caused by random segregations so is uncorrelated with regional co-localization.

In some cases, IBD based methods such as RDR will not capture the phenotypic variance explained by recent mutations, which are rare in the population. To measure how well RDR captures variance from rare variants, we simulated a trait determined by additive, direct effects of SNPs with MAFs between 1% and 0.1%, with median MAF 0.26% (Supplementary Note). RDR captured ~88% of the variance explained by the rare SNPs.

We found that RDR estimates were insensitive to non-additive genetic effects. The mean RDR estimates were close to the true narrow-sense heritability (40%) for traits influenced by both pairwise interactions between SNPs and dominance effects (Table 1). In contrast, the mean Sib-Regression estimates were close to the sum of the variance explained by additive and non-additive genetic effects (Table 1).

### RDR estimates of heritability for 14 human traits

We estimated the variance components of the RDR covariance model for 14 quantitative traits (Online Methods, Table 2, Supplementary Table 4, and Supplementary Figure 2). For the exact same probands that RDR was applied to, heritability estimates were obtained from the Kinship and Kinship F.E. methods (Online Methods, Table 2, Figure 2). For 11 of the 14 traits, the Kinship F.E. estimate,  $h_{\text{kinFE}}^2$ , is bigger than the RDR estimate,  $h_{\text{RDR}}^2$  (average  $h_{\text{kinFE}}^2 - h_{\text{RDR}}^2 = 12.1\%$ ). We found that  $h_{\text{kinFE}}^2$  was statistically significantly higher than  $h_{\text{RDR}}^2$  ( $p < 0.05$ , one sided test assuming  $h_{\text{kinFE}}^2$  and  $h_{\text{RDR}}^2$  are independent, so p-values represent an upper bound) for educational attainment ( $h_{\text{kinFE}}^2 - h_{\text{RDR}}^2 = 35.4\%$ ,  $p < 2.2 \times 10^{-4}$ ), height ( $h_{\text{kinFE}}^2 - h_{\text{RDR}}^2 = 22.6\%$ ,  $p < 1.3 \times 10^{-6}$ ), body mass index (BMI) ( $h_{\text{kinFE}}^2 - h_{\text{RDR}}^2 = 17.8\%$ ,  $p < 4.3 \times 10^{-3}$ ), and age at first child in women ( $h_{\text{kinFE}}^2 - h_{\text{RDR}}^2 = 10.9\%$ ,  $p < 0.043$ ). We found no evidence that differences between  $h_{\text{kinFE}}^2$  and  $h_{\text{RDR}}^2$  were driven by atypical properties of the sample with both parents genotyped or

by differences in mean trait values between the regions of Iceland (Supplementary Note and Supplementary Table 5).

Using Icelandic data, but without limiting to probands with parents genotyped, Sib-Regression estimates of heritability, denoted by  $h_{\text{sib}}^2$ , were computed (Online Methods, Table 2 and Figure 2). RDR estimates were more precise than Sib-Regression estimates for every trait, and, on average, the estimated standard errors for  $h_{\text{sib}}^2$  were 2.5 times larger than those for  $h_{\text{RDR}}^2$ , implying the effective sample size for RDR is around 6.25 times higher than for Sib-Regression. If a difference between RDR and Sib-Regression exists, it could be a consequence of indirect genetic effects between siblings<sup>18</sup>, epistasis, dominance, and/or rare variants. However, the lack of precision in Sib-Regression estimates implies that the power to detect differences is low, and we did not find any statistically significant differences.

There are not enough monozygotic twins in the Icelandic data to obtain precise twin estimates of heritability. To compare RDR results with twin studies from a similar population, we took estimates from the Swedish Twin Registry<sup>19</sup> denoted by  $h_{\text{twin}}^2$ , which were available for nine of the fourteen traits (Online Methods, Table 2, Figure 2, and Supplementary Table 6). The difference  $h_{\text{twin}}^2 - h_{\text{RDR}}^2$  was above zero and statistically significant ( $p < 0.05$ ) for all nine traits, with an average difference of 33.2%. For Sib-Regression, the average difference  $h_{\text{twin}}^2 - h_{\text{sib}}^2$  was 26.4%. The fact that both RDR and Sib-Regression estimates are substantially lower than twin studies estimates could be due to differences in heritability between our sample and the samples of twins and/or overestimation of heritability by twin studies.

### GREML-SNP estimates are biased by genetic nurturing

SNP based methods, such as GREML-SNP, will generally capture a smaller fraction of the full heritability of a trait than IBD based methods, such as RDR, making direct comparison of environmental bias difficult. We therefore introduce RDR-SNP, which uses SNPs to estimate the three relatedness matrices of the RDR covariance model (Online Methods). In other words,  $R$ ,  $R_{\text{par}}$ , and  $R_{\text{o,par}}$  are replaced by estimates from a set of SNPs:  $R^{\text{SNP}}$ ,  $R_{\text{par}}^{\text{SNP}}$ , and  $R_{\text{o,par}}^{\text{SNP}}$ . The only difference between RDR-SNP and GREML-SNP is that RDR-SNP also fits  $R_{\text{par}}^{\text{SNP}}$  and  $R_{\text{o,par}}^{\text{SNP}}$  in addition to  $R^{\text{SNP}}$ .

In order to compare RDR-SNP to typical GREML-SNP analysis, we simulated traits in a subset of the UK Biobank<sup>20</sup> where genotype data on both parents was available ( $n=973$ ). As in typical GREML-SNP analysis, we pruned the sample so that no pair of individuals had relatedness greater than 0.05, leaving 937 individuals (Supplementary Note).

We randomly sampled 11,771 SNPs to act as causal SNPs, and we calculated  $R^{\text{SNP}}$ ,  $R_{\text{par}}^{\text{SNP}}$ , and  $R_{\text{o,par}}^{\text{SNP}}$  from this set (Supplementary Note). We simulated a trait determined only by additive, direct effects of SNPs and random noise. Both GREML-SNP and RDR-SNP estimated the

true heritability, 20%, without detectable bias: mean estimate 19.76% (0.15% S.E.) for GREML-SNP and 19.70% (0.30% S.E.) for RDR-SNP.

Alleles transmitted to offspring are also present in the parents, so have both direct and parental genetic nurturing effects. Let  $\delta$  be the direct effect of a SNP, and let  $\eta$  be the parental genetic nurturing effect. The effect of the transmitted allele is therefore  $(\delta + \eta)$ . GREML-SNP uses only transmitted alleles, so is unable to separate the variance from the direct effect alone, proportional to  $\delta^2$ , from the variance explained by the combined direct and parental genetic nurturing effects, proportional to  $(\delta + \eta)^2$ . We investigated this theoretically (Supplementary Note) and by simulating a trait with both direct and genetic nurturing effects. We set the genetic nurturing effect of each variant to be one third of its direct effect, similar to the estimated ratio for educational attainment in Iceland<sup>14</sup>. The direct effects explained 20% of the phenotypic variance, implying that the total variance explained by transmitted alleles is  $(1 + \frac{1}{3})^2 \times 20\% \approx 35.56\%$ , much larger than the heritability, 20%.

The mean GREML-SNP heritability estimate was 35.15% (0.16% S.E.), very close to the total variance explained by combined direct and indirect effects of transmitted alleles, 35.56%, and in close agreement with our theoretical analysis (Supplementary Note). In contrast, RDR-SNP estimated heritability without detectable bias: mean estimate 19.70% (S.E. 0.30%).

### Evidence from Icelandic data for bias in GREML-SNP estimates

For typical GREML-SNP analysis, the sample is pruned so that no pair has relatedness above a low threshold, usually 0.025 or 0.05. When a large fraction of the sample is related to another person in the sample above threshold levels, such as in our Icelandic sample, this approach entails a large loss of sample size. A similar approach that avoids a large loss in sample size is to regress elements of the sample phenotypic covariance matrix onto  $R^{\text{SNP}}$  only for those pairs whose relatedness is less than the threshold. We call this approach RELT-SNP, with RELT standing for ‘relatedness thresholded’. If the same relatedness threshold is applied, GREML-SNP and RELT-SNP estimates from large samples would be expected to be very similar under most conditions. By applying RELT-SNP to the simulated traits in the UK Biobank, we showed RELT-SNP and GREML-SNP give very similar estimates and exhibit the same bias due to parental genetic nurturing effects (Supplementary Note).

In the Icelandic sample, we compared RDR-SNP to RELT-SNP with a relatedness threshold of 0.05 (Online Methods). We first computed RDR-SNP and RELT-SNP estimates for the simulated traits, whose true heritability was 40% (Supplementary Table 7). When using the causal variants to calculate  $R^{\text{SNP}}$ ,  $R_{\text{par}}^{\text{SNP}}$ , and  $R_{\text{o,par}}^{\text{SNP}}$ , both RDR-SNP and RELT-SNP gave approximately unbiased estimates of heritability for the ‘additive’, ‘maternal’, ‘epistatic’ and ‘dominance’ traits. For the ‘genetic nurturing’ trait, the average RELT-SNP estimate was 74.1% (0.14% S.E.), close to the variance explained by combined direct and genetic nurturing effects, ~73.3%. In contrast, RDR-SNP estimates were approximately unbiased ( $h^2 = 40.1\%$ , 0.07% S.E.). When the causal variants differed from the variants used to calculate



the relatedness matrices, a bias was introduced to RDR-SNP, RELT-SNP, and GREML-SNP estimates (Supplementary Table 7).

For the real traits, we estimated heritability using relatedness matrices calculated from 605,966 genome-wide SNPs typically found on Illumina genotyping arrays (Online Methods, Table 2 and Supplementary Table 8). We found that  $h_{\text{RELT-SNP}}^2$  was statistically significantly higher than  $h_{\text{RDR-SNP}}^2$  ( $p < 0.05$ , one sided test assuming  $h_{\text{RDR-SNP}}^2$  and  $h_{\text{RELT-SNP}}^2$  are independent, so p-values represent an upper bound) for height

$$\left(\frac{h_{\text{RELT-SNP}}^2}{h_{\text{RDR-SNP}}^2} = 1.24, p < 0.015\right), \text{ age at first child in women } \left(\frac{h_{\text{RELT-SNP}}^2}{h_{\text{RDR-SNP}}^2} = 1.72, p < 7.6 \times 10^{-3}\right),$$

$$\text{and educational attainment (years)} \left(\frac{h_{\text{RELT-SNP}}^2}{h_{\text{RDR-SNP}}^2} = 1.69, p < 0.027\right) \text{ (Methods).}$$

## Discussion

We introduced RDR, a novel heritability estimation method, and used it to estimate heritability for 14 quantitative traits in Iceland. Through mathematical investigations and simulations, we demonstrated that RDR estimates of heritability have negligible bias due to environment. In contrast, GREML-SNP, the Kinship method, and the Kinship F.E. method showed substantial bias due to indirect genetic effects from relatives. The GREML-SNP simulations show that removing close relatives does not remove bias due to indirect genetic effects from relatives. Our results suggest that GREML-SNP estimates could be interpreted as estimates of the variance explained by the combined direct and indirect effects of transmitted alleles, rather than the heritability.

For educational attainment, there is evidence for a substantial contribution from indirect genetic effects from parents and siblings<sup>14</sup>. This implies that educational attainment heritability estimates from GREML-SNP<sup>8,21</sup>, and the recently proposed extension to the Kinship model<sup>8</sup>, are likely to be upwardly biased. We estimated that GREML-SNP estimates of the heritability of educational attainment could be inflated by a factor of around 1.69. This inflation factor is consistent with genetic nurturing effects around 30% the size of direct effects, which is consistent an estimate using a polygenic score<sup>14</sup> and within-family analyses in other populations<sup>22</sup>.

RDR, like other methods using IBD segments<sup>7,8</sup>, may underestimate the heritability due to rare variants. However, any underestimation due to rare variants will be less than for GREML-SNP methods applied to typical genotyping arrays. By using IBD segments, RDR captures substantially more of the variance from rare variants, around 88% for variants with MAF between 1% and 0.1%. This implies that the underestimation of heritability by RDR will be small unless very rare variants, especially de-novo mutations, explain a large fraction of the phenotypic variance. Furthermore, Sib-Regression captures variance from all variants other than de-novo mutations not shared by siblings. The fact that Sib-Regression estimates are close to RDR estimates on average argues against substantial underestimation of heritability by RDR due to very rare variants.

Heritability estimates from Swedish twin studies were substantially higher than RDR and Sib-Regression estimates for almost all traits. Some of the difference between could be due to differences in heritability between our Icelandic sample and the Swedish twin samples. Other possible explanations are: overestimation of heritability by twin studies and/or very rare variants, especially de-novo mutations, explaining a substantial fraction of the phenotypic variance.

The RDR method requires parents of probands to be genotyped. Large datasets with this property are currently rare, which is the main reason our current study is limited to the Icelandic population. However, our results argue that large, genotyped samples including close relatives are essential for disentangling nature and nurture. As large population samples become more common, large amounts of family data will inevitably be collected. We therefore expect RDR and related methods to become more widely applied.

## Online Methods

### Icelandic Sample

All participating subjects donating biological samples signed informed consents and the study was approved by the Data Protection Commission of Iceland (DPC) and the National Bioethics Committee of Iceland. Personal identities of the phenotypes and biological samples were encrypted by a third party system provided by the Icelandic Data Protection Authority.

The Icelandic samples were genotyped using Illumina microarrays as previously described<sup>24</sup>. The whole genomes of 2,636 Icelanders were sequenced using Illumina technology to a mean depth of at least 10X (median 20X)<sup>24</sup>. A total of 35.5 million autosomal SNPs and indels were identified using the Genome Analysis Toolkit version 2.3.9.

The deCODE genealogy database is a comprehensive database that includes information on more than 800,000 Icelandic individuals, deceased and living, dating back to the settlement of Iceland 1,200 years ago. The database is constructed from a nationwide census, conducted regularly from the year 1700, church books and other available information, and is particularly complete for the last 200 years. The database includes, when known, information on parents of each individual, gender, year of birth and, if applicable, year of death.

We restricted our analyses to genotyped individuals with both genetic parents genotyped and all four grandparents in the deCODE genealogy database. This left 54,888 individuals. See the Life Sciences Reporting Summary for a summary of sample restrictions and other information.

The individuals and their parents had all been phased and segments shared identical-by-descent (IBD), both within and between individuals, determined by long-range phasing<sup>25,26</sup>. To reduce bias due to segments incorrectly called as identical-by-descent, we

restricted our analyses to segments of length greater than 5 centi-Morgans. Note that sex-chromosomes were not included.

As a measure of ascertainment bias, we compared years of education between the individuals with both parents genotyped and the full set of individuals with education data. Mean years of education for those with both parents genotyped was 15.07 compared to 13.63 for the whole sample with education data. Part of this is due to the fact that those with both parents genotyped were born later than average, and mean levels of education have increased over time. After regressing out year-of-birth (YOB), YOB<sup>2</sup>, YOB<sup>3</sup>, the sample with both parents genotyped still had 0.32 years more education on average, compared to a standard deviation of 3.39 years. This shows that our results are slightly biased towards those with higher socio-economic status, which, for many traits, is expected to increase heritability<sup>27,28</sup>.

### Trait measurements

As a measure of educational attainment, we used information on years of schooling, available for 63,508 individuals, that originated from questionnaires administered in deCODE's various disease projects and from routine assessments of elderly nursing home residents. As the data have been gathered over the years for the purpose of descriptive demographics rather than for phenotype use, the questions were originally not standardized across projects and many of them have categorical responses. For this study, to make it as consistent as is possible when it comes to the educational attainment trait studied in the published meta-analysis<sup>29</sup>, efforts were put into mapping the responses to the questionnaires into the UNESCO ISCED classification (see [URLs](#)). In particular, the final quantitative measure used, before sex and year-of-birth adjustments, ranges from a minimum of 10 years to a maximum of 20 years.

Height and body mass index (BMI) information, collected primarily through deCODE's genetic studies on cardiovascular disease, obesity and cancer, were available for 89,615 and 77,285 adult individuals, respectively<sup>30,31</sup>. About 20% of the information was self-reported.

Blood measurements were collected from three of the largest laboratories in Iceland: Landspítali - The National University Hospital of Iceland, Reykjavík; The Laboratory in Mjódd, Reykjavík; Akureyri Hospital, The Regional Hospital in North Iceland, Akureyri; in addition to the Icelandic Heart Association. For many individuals, multiple blood samples had been taken at different time points. To aid comparability with other studies that have used one time-point only, we took only the first measurement of each individual.

Information on 'age at first child' (AAFC) was extracted from the deCODE genealogy database. Age at menarche was determined by the answer to the question 'How old were you when your menstruation started?' as detailed elsewhere<sup>32</sup>.

---

#### URLs

Educational attainment categories: <http://uis.unesco.org/en/isced-mappings>.

Apart from educational attainment, traits were quantile-normalised within each sex. Educational attainment was not quantile-normalised as the measurements fall into discrete categories of years of education. The traits were regressed on year-of-birth (YOB), sex, YOB, YOB<sup>2</sup>, YOB<sup>3</sup>, and the interactions of sex with YOB<sup>2</sup>, YOB<sup>3</sup>. The residuals of this regression were then used as the phenotype,  $Y$ , when fitting the models described below. To ensure our heritability estimates correspond to the adult but not elderly population, we further restricted our analysis to those born between 1951 and 1995 for height, and between 1951 and 1997 for BMI and the traits measured from blood. (Note that for Sib-Regression, the year of birth restrictions were not applied to maximise the sample size.)

### Identification of siblings

For the Sib-Regression estimator, we obtained the relatedness for all pairs of genotyped individuals who share both parents in the genealogy. To ensure we only used true full-siblings, we clustered the pairs by relatedness into four clusters using k-means clustering: unrelated, half-sibling, full-sibling, and monozygotic twins. This left 127,264 full-sibling pairs, comprised of 70,317 unique individuals, whose relatedness distribution had a mean of 0.502 and a standard deviation of 0.0382. To maximise the precision of the Sib-Regression estimator, we did not restrict by year-of-birth or by number of parents genotyped, so the sample used is different to the sample used for the other estimators.

### Calculation of IBD relatedness matrices

To calculate  $\mathbf{R}$ ,  $\mathbf{R}_{\text{par}}$ , and  $\mathbf{R}_{\text{o,par}}$ , we used formulae based on the genetic covariance in a population descending from a finite number of ancestors<sup>33</sup> (Supplementary Note):

$$[\mathbf{R}]_{ij} = \frac{1}{2} \sum_{k,l=m,p} (\text{IBD}_{ij}^{kl} - K_0) / (1 - K_0)$$

where  $K_0$  is the mean kinship over all pairs in the population, and  $\text{IBD}_{ij}^{kl}$  is the proportion of the maternally inherited haplotype of  $i$  shared identical-by-descent (IBD) with the paternally inherited haplotype of  $j$ ;

$$[\mathbf{R}_{\text{par}}]_{ij} = \frac{K_{p(i)p(j)} + K_{p(i)m(j)} + K_{m(i)p(j)} + K_{m(i)m(j)} - 4K_0}{(1 - K_0)}$$

where  $K_{p(i)m(j)}$  is the kinship between the father of  $i$  and the mother of  $j$ ;

$$[\mathbf{R}_{\text{o,par}}]_{ij} = \frac{K_{ip(j)} + K_{im(j)} + K_{m(i)j} + K_{p(i)j} - 4K_0}{(1 - K_0)}$$

where  $K_{im(j)}$  is the kinship between  $i$  and the mother of  $j$ , etc.

## Calculation of SNP relatedness matrices

To perform RDR-SNP analysis, we calculated relatedness matrices from SNPs ( $\mathbf{R}^{\text{SNP}}, \mathbf{R}_{\text{par}}^{\text{SNP}}, \mathbf{R}_{\text{o,par}}^{\text{SNP}}$ ) that are analogous to the IBD relatedness matrices used in RDR ( $\mathbf{R}, \mathbf{R}_{\text{par}}$ , and  $\mathbf{R}_{\text{o,par}}$ ). Consider a sample of  $n$  individuals genotyped at  $l$  bi-allelic SNPs, where the genotype is expressed as the copy number (0,1, or 2) of one of the two alleles. Let  $\mathbf{G}$  be the  $[n \times l]$  matrix of genotypes standardised to have mean zero and variance 1. The matrix  $\mathbf{R}^{\text{SNP}}$  is equivalent to that used in standard GREML-SNP analysis, and is calculated as  $\mathbf{R}^{\text{SNP}} = l^{-1}\mathbf{G}\mathbf{G}^T$ .

To calculate  $\mathbf{R}_{\text{par}}^{\text{SNP}}$  and  $\mathbf{R}_{\text{o,par}}^{\text{SNP}}$ , we first have to form parental genotypes. Let  $\mathbf{G}_{\text{m}}$  be the  $[n \times l]$  matrix of genotypes of the mothers of the  $n$  individuals in the sample, and let  $\mathbf{G}_{\text{p}}$  be the  $[n \times l]$  matrix of the genotypes of the fathers. Then  $\mathbf{G}_{\text{par}} = \mathbf{G}_{\text{m}} + \mathbf{G}_{\text{p}}$  is the parental genotype matrix, with entries from  $\{0,1,2,3,4\}$ . We normalised the columns of  $\mathbf{G}_{\text{par}}$  to have mean zero and variance two. The variance is naturally twice that of the offspring genotype in an outbred population as each entry is the sum of maternal and paternal genotypes. Then

$$\mathbf{R}_{\text{par}}^{\text{SNP}} = (2l)^{-1}\mathbf{G}_{\text{par}}\mathbf{G}_{\text{par}}^T; \mathbf{R}_{\text{o,par}}^{\text{SNP}} = (2l)^{-1}(\mathbf{G}\mathbf{G}_{\text{par}}^T + \mathbf{G}_{\text{par}}\mathbf{G}^T).$$

The matrices are calculated in this way to ensure estimates of  $v_g$ ,  $v_{e-g}$ , and  $c_{g,e}$  are properly calibrated. These equations can be derived from a random effects model (Supplementary Note).

For the analysis of the real traits, we computed relatedness matrices from SNPs from the Illumina Framework SNP set. The Illumina Framework SNP set is a set of 611,173 SNPs shared between many of the Illumina genotyping arrays used to genotype the Icelandic sample<sup>24</sup>. We used this set of SNPs in order to make our analysis comparable to applications of GREML-SNP to data from typical genotyping arrays. Before computing relatedness matrices, we removed SNPs with imputation information below 0.9999 and MAF less than 1%, leaving 605,966 SNPs. For the simulated traits, we also computed relatedness matrices from only the causal SNPs (Supplementary Note).

## Computing Relatedness Disequilibrium Regression (RDR) estimates

The relatedness disequilibrium regression (RDR) covariance model is

$$\text{Cov}(Y) = v_g \mathbf{R} + v_{e-g} \mathbf{R}_{\text{par}} + c_{g,e} \mathbf{R}_{\text{o,par}} + \sigma^2 \mathbf{I}.$$

We investigated fitting this model by least squares regression of the off-diagonal elements of the sample phenotypic covariance matrix on the off-diagonal elements of the relatedness matrices:

$$(y_i - \bar{y})(y_j - \bar{y}) \sim [\mathbf{R}]_{ij} + [\mathbf{R}_{\text{par}}]_{ij} + [\mathbf{R}_{\text{o,par}}]_{ij}.$$

where  $y_i$  is the phenotype observation for individual  $i$ , and  $\bar{y}$  is the sample phenotype mean. We excluded both parent-offspring and grandparent-grandchild pairs from the regression, as these pairs violate the relationship between  $[R]_{ij}$  and  $[R_{\text{par}}]_{ij}$  required for removal of environmental bias from estimation of  $v_g$  (Figure 1 and Supplementary Note). We also investigated fitting the model by unconstrained restricted maximum likelihood in *GCTA34*, under the assumption the trait follows a multivariate normal distribution:

$$Y \sim N\left(\mu, v_g R + v_e \sim_g R_{\text{par}} + c_{g,e} R_{\text{o,par}} + \sigma^2 \mathbf{1}\right).$$

For the maximum likelihood method, one can only remove individuals, and all the pairs including that individual, not arbitrary pairs. Around 30% of the sample with both parents genotyped have an ancestor who also has both parents genotyped. We therefore did not exclude individuals so that no parent-offspring and no grandparent-grandchild pairs were present, as this would have resulted in a large loss of sample size.

In our simulations, we found that RDR estimates from maximum likelihood and RDR estimates from least-squares were both approximately unbiased, with no consistent advantage in bias evident from fitting the model by least-squares after excluding parent-offspring and grandparent-grandchild pairs (Supplementary Table 1). However, least-squares estimates were considerably less precise than those from maximum likelihood. We therefore used maximum likelihood without exclusion of parent-offspring and grandparent-grandchild pairs for all analyses in the main text. For the real traits, the results from least-squares were consistent with the results from maximum likelihood, but the least-squares estimates were considerably less precise (Supplementary Table 5).

To obtain RDR-SNP estimates, we fitted the following model by restricted maximum likelihood in *GCTA*:

$$Y \sim N\left(\mu, v_g R^{\text{SNP}} + v_e \sim_g R_{\text{par}}^{\text{SNP}} + c_{g,e} R_{\text{o,par}}^{\text{SNP}} + \sigma^2 \mathbf{1}\right).$$

### Computing Kinship and Kinship F.E. estimates

To obtain heritability estimates from the Kinship method, we fitted the following model for a vector of phenotype observations  $\mathbf{Y}$ :

$$Y \sim N\left(X_{\text{kin}} b, v_g R + \sigma^2 \mathbf{1}\right).$$

For the Kinship F.E. model, we added a variance component that modelled shared family environment:

$$Y \sim N\left(X_{\text{kin}} b, v_g R + v_c C + \sigma^2 \mathbf{1}\right),$$

where  $[C]_{ij} = \mathbf{1}$  if  $i$  and  $j$  shared a mother according to the deCODE genealogy database, otherwise  $[C]_{ij} = \mathbf{0}$ . For all of the simulated traits other than the ‘regional’ trait,  $X_{\text{kin}}$  was a

constant. For the ‘regional’ trait, it also included the top 20 genetic principal components. In the real trait analysis,  $\mathbf{X}_{\text{kin}}$  included the top 20 genetic principal components. For both the Kinship and Kinship F.E. methods, we estimated model parameters by unconstrained restricted maximum likelihood in *GCTA34*.

### Computing RELT-SNP estimates

To compute RELT-SNP estimates, we regressed off-diagonal elements of the phenotypic covariance matrix onto elements of  $\mathbf{R}^{\text{SNP}}$ , excluding elements of  $\mathbf{R}^{\text{SNP}}$  greater than 0.05. Let  $\mathbf{X}$  be a matrix whose first column has every entry equal to one, and whose other columns are covariates and/or positions on genetic principal components. Let  $\hat{\mathbf{b}}$  be the least-squares estimate of vector of regression coefficients of the phenotype on  $\mathbf{X}$ . Then we formed the sample phenotypic covariance matrix as  $\mathbf{S} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{b}})(\mathbf{Y} - \mathbf{X}\hat{\mathbf{b}})^{\text{T}}$ , where  $\mathbf{Y}$  is the vector of phenotype observations. Estimates of  $v_g$  were computed by regressing off-diagonal elements of  $\mathbf{S}$ ,  $[\mathbf{S}]_{ij}$ , on off-diagonal elements of  $\mathbf{R}^{\text{SNP}}$ ,  $[\mathbf{R}^{\text{SNP}}]_{ij}$ , excluding pairs where  $[\mathbf{R}^{\text{SNP}}]_{ij} > 0.05$ .

We describe a computational procedure for computing RELT-SNP estimates and their standard errors in the Supplementary Note. This procedure builds on previous work expressing the Haseman-Elston regression as a quadratic form<sup>35</sup>, which takes into account the dependence between elements of  $\mathbf{S}$ . We found our standard error estimates to be accurate in simulations, with a mean error of 4.3% across the simulated traits (Supplementary Table 9). The RELT-SNP estimates and standard errors were computed using custom Python code.

For the ‘regional’ trait, RELT-SNP was upwardly biased ( $h^2 = 45.7\%$ , 0.23% S.E.) but became approximately unbiased ( $h^2 = 39.3\%$ , 0.10% S.E.) when the trait was adjusted for 20 genetic principal components. However, we found that adjustment for 20 genetic principal components resulted in a downward bias for the ‘additive’ trait ( $h^2 = 38.7\%$ , 0.09% S.E.). We therefore decided to take an approach where we adjusted for principal components only for those traits that exhibited substantial stratification. For the results in Table 2, we adjusted for 20 principal components only for the traits where the variance explained by the top 20 principal components exceeded 1%: height, age at first child in men and women, and educational attainment. We give results with and without control for principal components for all traits in Supplementary Table 8. The choice of 1% was somewhat arbitrary. Arbitrary decisions about how many principal components to control for are a disadvantage of Kinship, GREML-SNP, and RELT-SNP methods. RDR and RDR-SNP, in contrast, do not require such arbitrary decisions, as they separate genetic and environmental effects in a principled way.

### Computing Sib-Regression estimates

To obtain Sib-Regression estimates<sup>5</sup>, we fit the regression model

$$(y_i - y_j)^2 \sim [\mathbf{R}]_{ij}$$

for all  $i, j$  such that  $i$  and  $j$  are full-siblings. We fit the regression model by least-squares using custom *R* code. The estimate of  $v_g$  is then minus one half of the estimated regression

coefficient. We compared estimating standard errors by the approximate formula given in the original Sib-Regression paper<sup>5</sup> (equation 17) and estimating standard errors by treating Sib-Regression as a standard univariate linear regression with uncorrelated observations. For the ‘additive’ simulated trait, both gave almost exactly the same estimated standard error, which underestimated the standard error by approximately 9%. We used standard errors estimated from treating Sib-Regression as a standard univariate linear regression with uncorrelated observations for all other results.

### Simulations using deCODE data

For all traits other than the ‘rare SNPs’ trait, we used imputed genotypes at **611, 173** SNPs from the Illumina Framework SNP set (see above). We filtered the SNPs so that the minimum imputation information was 0.9999, removing around half of the SNPs. Out of the remaining SNPs passing the filter, we randomly sampled 10,000 SNPs to use as the causal SNPs in our simulations. In the 10,000 selected SNPs, the median imputation information was 1.0000, the minimum minor allele frequency (MAF) was 0.52%, and the median MAF was 22.8%. For the ‘rare SNPs’ trait, we randomly sampled SNPs from all imputed SNPs with MAF between 1% and 0.1% and with imputation information at least 0.9999 and p-value for Hardy-Weinberg deviation greater than 0.05. We sampled 100 such SNPs from each chromosome, giving 2,200 SNPs in total.

For each type of trait, we simulated 500 independent replicates. We briefly describe the simulation of the direct, additive genetic component of each trait, which explained 40% of the phenotypic variance. Apart from for the ‘rare SNPs’ trait, we standardised genotypes so that each SNP’s genotype vector had sample mean zero and sample variance one. Let  $\mathbf{G}$  represent the matrix of standardised genotypes at the 10,000 causal SNPs. We sampled additive effects of SNPs from a normal distribution. Let  $\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{I})$  represent the vector of SNP effects. The additive genetic component,  $\mathbf{A}$ , was calculated as  $\mathbf{A} = \mathbf{G}\boldsymbol{\beta}$ , and then scaled to explain 40% of the phenotypic variance. For details on simulation of environmental components, see the Supplementary Note.

### Simulations in the UK Biobank

To select causal SNPs for phenotype simulation, for each chromosome we randomly sampled 1,500 SNPs then removed those with MAF less than 5% or more than 0.5% missing genotypes. This gave a set of 11,771 SNPs. We mean imputed missing genotypes for both parents and offspring. We simulated 10,000 independent replications of each trait. Let  $\mathbf{l} = \mathbf{11,771}$ . We standardised offspring genotypes so that the genotypes at each SNP had mean zero and variance 1. Let  $\mathbf{G}$  be the matrix of standardised offspring genotypes. Here, we describe simulation of the direct, additive genetic component of the traits – for further details, see the Supplementary Note. For each trait, we simulated a normally distributed vector of effects for the  $\mathbf{l}$  SNPs:  $\boldsymbol{\beta} \sim N(\mathbf{0}, 0.2\mathbf{l}^{-1}\mathbf{I})$ . The additive genetic component of the trait,  $\mathbf{A}$ , was then calculated as  $\mathbf{A} = \mathbf{G}\boldsymbol{\beta}$ .

### Selection of estimates from twin studies

The Swedish Twin Registry<sup>19</sup> is a large sample of twins from a population of similar cultural and genetic composition to Iceland, giving the most precise and valid comparison



possible based on published data<sup>36–40</sup>. The exception is for education, where we used a meta-analysis of Scandinavian twin studies for increased precision<sup>23</sup>. For BMI and traits measured from blood, unlike our estimates, the Swedish Twin Registry estimates did not exclude elderly individuals. This is unlikely to account for the higher estimates in the Swedish Twin Registry, as twin correlations and heritability estimates are generally lower in the elderly population<sup>2</sup>.

We took the heritability estimate from the additive-common-environment (ACE) model<sup>2,3</sup> when provided. ACE estimates were not provided for the blood lipid traits, but monozygotic and dizygotic twin correlations were<sup>39</sup>. We used these to obtain the moment based estimate of the heritability under the ACE model by the formula:  $2(r_{MZ} - r_{DZ})$ , where  $r_{MZ}$  is the phenotypic correlation for monozygotic twins, and  $r_{DZ}$  is the phenotypic correlation for dizygotic twins. We took the weighted average of the same-sex and opposite-sex dizygotic twin correlations to estimate  $r_{DZ}$ . For creatinine, the ACE estimate was not provided, and neither were the twin correlations, so we took the published heritability estimate from the ADE model (additive-dominance-environment). The studies used and methods used are summarised in Supplementary Table 6. For height, heritability estimates were only provided for males and females separately, so we took the average estimate. The standard error was not provided. Height and weight estimates were based on self-reported data, whereas our estimates were based on approximately 80% measured and 20% self-reported data. This would be expected to increase our heritability estimates for height and BMI relative to the twin estimates due to a reduction in measurement error. For education, we used a meta-analysis of twin studies in Scandinavian countries, including Sweden, to give a more precise estimate<sup>23</sup>. We could not find published estimates based on the Swedish Twin Registry for the haemoglobin traits and for age at first child, so we excluded them from the comparison.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

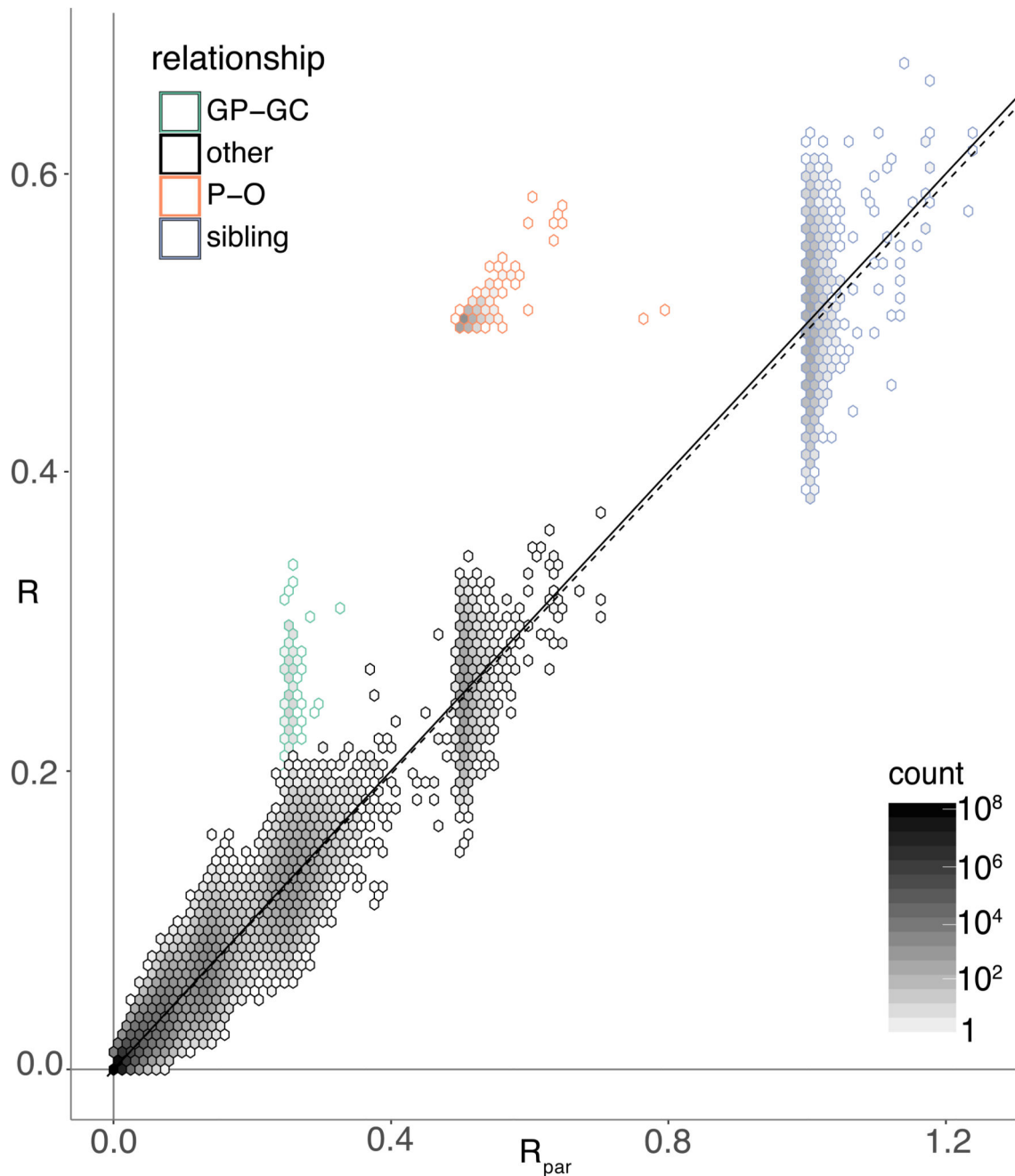
A.I.Y. was supported by a Wellcome Trust Doctoral Studentship (099670/Z/12/Z) for part of this project. A.I.Y. and A.K. were supported by the Li Ka Shing Foundation for part of this project.

## References

1. Sesardic N. Making Sense of Heritability. Cambridge University Press; 2005.
2. Polderman TJC, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet.* 2015; 47:702–709. [PubMed: 25985137]
3. Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nat Rev Genet.* 2002; 3:872–82. [PubMed: 12415317]
4. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. *Nat Genet.* 2017; 49
5. Visscher PM, et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* 2006; 2:e41. [PubMed: 16565746]
6. Speed D, et al. Reevaluation of SNP heritability in complex human traits. *Nat Genet.* 2017; 49:986–992. [PubMed: 28530675]

7. Zaitlen N, et al. Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. *PLoS Genet.* 2013; 9:e1003520. [PubMed: 23737753]
8. Hill WD, et al. Genomic analysis of family data reveals additional genetic effects on intelligence and personality. *Mol Psychiatry.* 2018; 44
9. Muñoz M, et al. Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat Genet.* 2016; 48:980. [PubMed: 27428752]
10. Heckerman D, et al. Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc Natl Acad Sci.* 2016; 113:7377–7382. [PubMed: 27382152]
11. Hemani G, et al. Inference of the genetic architecture underlying bmi and height with the use of 20,240 sibling pairs. *Am J Hum Genet.* 2013; 93:865–875. [PubMed: 24183453]
12. Baud A, et al. Genetic Variation in the Social Environment Contributes to Health and Disease. *PLOS Genet.* 2017; 13:e1006498. [PubMed: 28121987]
13. Bijma P. Estimating indirect genetic effects: Precision of estimates and optimum designs. *Genetics.* 2010; 186:1013–1028. [PubMed: 20713688]
14. Kong A, et al. The nature of nurture: Effects of parental genotypes. *Science* (80). 2018; 359:424–428.
15. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet.* 1993; 52:506–16. [PubMed: 8447318]
16. Ewens WJ, Spielman RS. The transmission/disequilibrium test: history, subdivision, and admixture. *Am J Hum Genet.* 1995; 57:455–464. [PubMed: 7668272]
17. Thomson G. Mapping disease genes: family-based association studies. *Am J Hum Genet.* 1995; 57:487–498. [PubMed: 7668275]
18. Carey G. Sibling imitation and contrast effects. *Behav Genet.* 1986; 16:319–341. [PubMed: 3753368]
19. Pedersen NL, Lichtenstein P, Svedberg P. The Swedish Twin Registry in the Third Millennium. *Twin Res.* 2002; 5:427–432. [PubMed: 12537870]
20. Bycroft C, et al. Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv.* 2017; doi: 10.1101/166298
21. Krapohl E, Plomin R. Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol Psychiatry.* 2016; 21:437–443. [PubMed: 25754083]
22. Lee J, et al. Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nat Genet.* (in Press).
23. Branigan AR, Mccallum KJ, Freese J. Variation in the heritability of educational attainment: An international meta-analysis. *Soc Forces.* 2013; 92:109–140.
24. Gudbjartsson DF, et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet.* 2015; 47:435–444. [PubMed: 25807286]
25. Kong A, et al. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet.* 2008; 40:1068–75. [PubMed: 19165921]
26. Kong A, et al. Parental origin of sequence variants associated with complex diseases. *Nature.* 2009; 462:868–874. [PubMed: 20016592]
27. Tuvblad C, Grann M, Lichtenstein P. Heritability for adolescent antisocial behavior differs with socioeconomic status: Gene-environment interaction. *J Child Psychol Psychiatry Allied Discip.* 2006; 47:734–743.
28. Stoolmiller M. Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior--genetic adoption studies. *Psychol Bull.* 1999; 125:392. [PubMed: 10414224]
29. Okbay A, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature.* 2016; doi: 10.1038/nature17671
30. Gudbjartsson DF, et al. Many sequence variants affecting diversity of adult human height. *Nat Genet.* 2008; 40:609–15. [PubMed: 18391951]

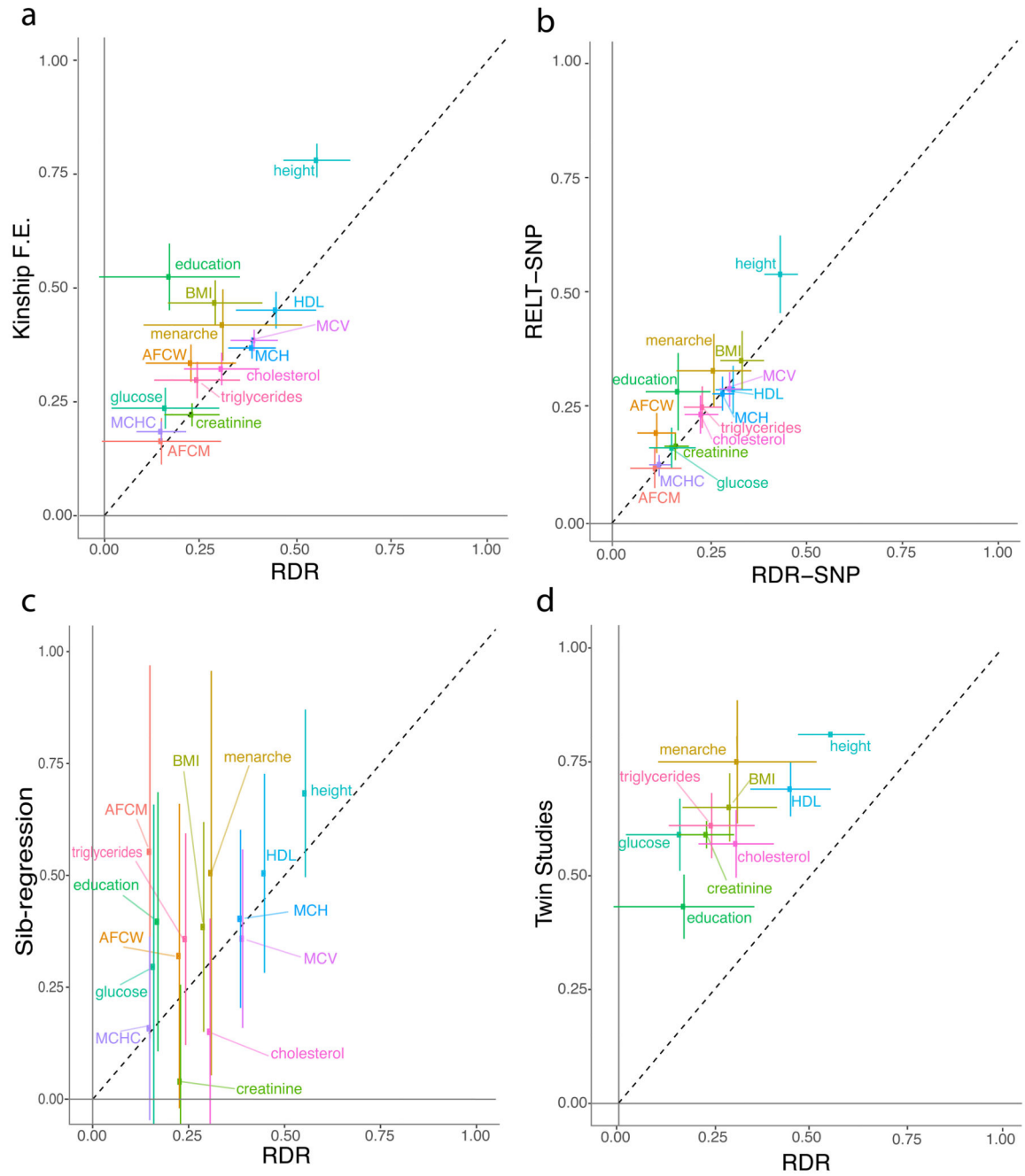
31. Thorleifsson G, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet.* 2009; 41:18–24. [PubMed: 19079260]
32. Elks CE, et al. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat Genet.* 2010; 42:1077–1085. [PubMed: 21102462]
33. Young AI, Durbin R. Estimation of Epistatic Variance Components and Heritability in Founder Populations and Crosses. *Genetics.* 2014; 198:1405–1416. [PubMed: 25326236]
34. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011; 88:76–82. [PubMed: 21167468]
35. Tamar S. Confidence intervals for heritability via Haseman-Elston regression. *Statistical Applications in Genetics and Molecular Biology.* 2017; 16:259. [PubMed: 28862991]
36. Carlsson S, Ahlbom A, Lichtenstein P, Andersson T. Shared genetic influence of BMI, physical activity and type 2 diabetes: A twin study. *Diabetologia.* 2013; 56:1031–1035. [PubMed: 23404445]
37. Silventoinen K, et al. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res.* 2003; 6:399–408. [PubMed: 14624724]
38. Baker JH, Thornton LM, Bulik CM, Kendler KS, Lichtenstein P. Shared genetic effects between age at menarche and disordered eating. *J Adolesc Heal.* 2012; 51:491–496.
39. Rahman I, et al. Genetic dominance influences blood biomarker levels in a sample of 12,000 Swedish elderly twins. *Twin Res Hum Genet.* 2009; 12:286–294. [PubMed: 19456221]
40. Arpegård J, et al. Comparison of heritability of Cystatin C- and creatinine-based estimates of kidney function and their relation to heritability of cardiovascular disease. *J Am Heart Assoc.* 2015; 4:e001467. [PubMed: 25589536]



**Figure 1. Relatedness disequilibrium**

For all pairs of individuals  $i, j$  from 20,000 Icelanders with both parents genotyped, the relatedness of  $i$  and  $j$ ,  $[R]_{ij}$ , is compared to the relatedness of the parents of  $i$  and the parents of  $j$ ,  $[R_{par}]_{ij}$ . The number of pairs in each hexagonal bin is indicated by shading. Relationships determined by the deCODE Genealogy database are indicated: GP-GC, grandparent-grandchild; P-O, parent-offspring; and sibling. The solid diagonal line indicates the expectation of  $[R]_{ij}$ , which is  $[R_{par}]_{ij}/2$ , except for pairs where one is a direct ancestor of the other (Supplementary Note). The dashed diagonal line indicates the regression line

(excluding parent-offspring and grandparent-grandchild pairs), with intercept  $-1 \times 10^{-4}$ , gradient 0.493, and variance explained 84%. The small deviation of the regression line from the theoretical expectation is likely due to some IBD segments shared between parents being broken up by recombination, resulting in a small fraction of segments in the offspring being too small to detect. Relatedness disequilibrium is the variation in  $[R]_{ij}$  around  $[\mathcal{R}_{\text{par}}]_{ij}/2$ . Relatedness disequilibrium is due to independent, random segregations in the parents, except for pairs where one is the direct ancestor of the other.



**Fig. 2. Comparison of heritability estimates from different methods.**

Horizontal intervals show +/-1.96 standard errors for the estimates on the x-axis, and vertical intervals show +/-1.96 standard errors for the estimates on the y-axis. See Table 2 for numerical values. A) Comparison of RDR to ‘Kinship F.E.’. B) Comparison of RDR-SNP to RELT-SNP. C) Comparison of RDR to Sib-Regression5 estimates. Intervals for the RDR estimates are not shown to better display Sib-Regression intervals. D) Comparison to published twin studies estimates from the Swedish Twin Registry<sup>19</sup>, apart from for education, which is from a meta-analysis of Scandinavian twin studies<sup>23</sup> (Supplementary

Table 6). Trait abbreviations: BMI, body mass index; AFCW, age at first child in women; AFCM, age at first child in men; education, educational attainment (years); cholesterol, total cholesterol; HDL, high density lipoprotein; glucose, fasting glucose; MCH, mean cell haemoglobin; MCHC, mean cell haemoglobin concentration; MCV, mean cell volume.

**Table 1**  
**Comparison of heritability estimates for simulated traits.**

The mean heritability estimates, expressed as a % of the phenotypic variance, from four different methods (RDR, Kinship, Kinship F.E., Sib-Regression) for different simulated traits. The true (narrow-sense) heritability of each trait was 40%. We simulated 500 replicates of each trait based on Icelandic genetic data from a random subsample of 10,000 individuals with both parents genotyped (Methods) --- apart from for Sib-Regression, where we used all 54,888 individuals. Ten thousand SNPs with median minor allele frequency (MAF) 22.8% were given additive effects for all the traits other than the ‘rare SNPs’ trait, for which 2,200 SNPs with MAF between 0.1% and 1% (median 0.26%) were used. To the additive genetic component, only noise was added for the ‘additive’ trait and the ‘rare SNPs’ trait. For the ‘epistatic’ trait, 10% of the phenotypic variance was due to pairwise interactions between SNPs. For the ‘dominance’ trait, 10% of the phenotypic variance was due to dominance effects. For the other traits, effects representing different sources of environmental confounding were added in addition to noise and the additive genetic component. For the ‘regional’ trait, each region of Iceland (sysla) was given an effect; for the ‘maternal environment’ trait, an environmental effect shared between those who share mothers was added; for the ‘genetic nurturing trait’, the genotypes of the parents were also given effects to simulate ‘parental genetic nurturing’ effects<sup>14</sup>. For the ‘regional’ trait, the Kinship and Kinship F.E. methods also included adjustment for 20 genetic principal components.

Trait	RDR		Kinship		Kinship F.E.		Sib-Regression	
	Estimate (%)	Standard Error (%)	Estimate (%)	Standard Error (%)	Estimate (%)	Standard Error (%)	Estimate (%)	Standard Error (%)
additive	39.3	0.62	40.4	0.15	40.5	0.18	41.2	0.69
genetic nurturing	39.4	0.49	92.7	0.09	82.8	0.14	40.4	0.37
maternal environment	38.9	0.58	76.3	0.17	39.9	0.18	41.1	0.37
regional	38.3	0.60	59.0	0.17	58.3	0.20	32.1	0.63
rare SNPs	35.0	0.64	39.5	0.15	39.4	0.19	39.7	0.67
epistatic	41.3	0.60	44.2	0.16	43.3	0.19	50.1	0.63
dominance	40.5	0.63	42.7	0.15	41.1	0.19	50.5	0.71



**Table 2****Heritability estimates.**

For each trait, the sample size used for the RDR, Kinship F.E., RDR-SNP, and RELT-SNP methods is given under 'n', and the sample size for Sib-Regression ('Sib-Reg.') given under 'sib-pairs'. Each heritability estimate is expressed as a percentage of the phenotypic variance and is followed by its standard error in brackets. RDR, Kinship F.E., RDR-SNP, and RELT-SNP estimates are from the exact same Icelandic samples with both parents genotyped, and samples were restricted to those born between 1951 and 1997 for BMI and traits measured from blood, and samples were restricted to those born between 1951 and 1995 for height. In order to maximise sample size, Sib-Regression estimates are from all genotyped Icelandic sibling pairs available without year-of-birth restrictions. Twin studies estimates are from the Swedish Twin Registry<sup>19</sup>, apart from for education, which is from a meta-analysis of Scandinavian twin studies<sup>23</sup> (Supplementary Table 6). Trait abbreviations: BMI, body mass index; AFCW, age at first child in women; AFCM, age at first child in men; menarche, age at menarche (years); education, educational attainment (years); total chol., total cholesterol; HDL, high density lipoprotein; glucose, fasting glucose; MCH, mean cell haemoglobin; MCHC, mean cell haemoglobin concentration; MCV, mean cell volume.

Trait	n	RDR		Kinship F.E.		RDR-SNP		RELT-SNP		Sib-Regression			Twin	
		Est. (%)	S.E. (%)	Est. (%)	S.E. (%)	Est. (%)	S.E. (%)	Est. (%)	S.E. (%)	sib-pairs	Est. (%)	S.E. (%)	Est. (%)	S.E. (%)
BMI	19,589	28.9	6.3	46.7	2.5	34.2	2.9	36.1	3.4	56,461	38.5	12.0	65	3.8
height	21,802	55.4	4.4	78.0	1.9	44.5	2.3	55.2	4.4	64,847	68.4	9.6	81	-
AFCW	22,367	22.6	6.0	33.5	2.1	11.7	2.6	20.1	2.3	30,582	32.0	17.4	-	-
AFCM	17,117	14.9	7.9	16.3	2.6	11.5	3.4	12.3	2.2	21,729	55.3	21.3	-	-
menarche	11,242	30.9	10.5	41.9	4.0	26.8	5.0	33.9	4.2	16,621	50.6	23.1	75	6.9
education	12,035	17.0	9.4	52.4	3.7	17.3	4.4	29.2	4.4	32,542	39.7	14.8	43	3.6
total chol.	27,320	30.6	5.0	32.2	1.8	23.5	2.3	24.2	2.2	74,271	15.1	12.9	57	3.8
HDL	24,570	44.8	5.3	45.1	2.1	32.0	2.5	29.7	2.7	67,894	50.5	11.4	69	3.1
triglycerides	24,099	24.2	5.7	29.8	2.0	23.8	2.6	25.8	2.4	62,746	35.8	12.1	61	3.7
glucose	19,500	15.9	7.2	23.6	2.3	15.8	3.1	16.8	2.3	36,469	29.6	18.5	59	4.0
creatinine	38,929	22.9	3.7	22.2	1.3	16.9	1.6	17.2	1.6	98,385	4.0	11.1	59	1.5
MCH	43,917	38.5	3.2	36.8	1.2	29.3	1.5	28.7	1.9	107,711	40.3	10.2	-	-
MCHC	43,963	14.9	3.3	18.4	1.1	12.5	1.5	13.0	1.2	107,833	15.8	10.5	-	-
MCV	43,919	39.1	3.1	38.5	1.2	31.1	1.5	29.8	2.0	107,702	35.9	10.2	-	-