ARTICLES
https://doi.org/10.1038/s41588-018-0178-9
nature genetics

Relatedness disequilibrium regression estimates heritability without environmental bias

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

Heritability measures the proportion of trait variation that is due to genetic inheritance. Measurement of heritability is important in the nature-versus-nurture debate. However, existing estimates of heritability may 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 used a sample of 54,888 Icelanders who had 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 may be inflated by environmental effects.

Heritability is a measure of the proportion of trait variation due to genetic inheritance within a population. Estimation of the relative importance of genetic inheritance (nature) versus environment (including nurture) has generated much controversy[1]. Historically, most estimates of heritability for human traits have come from twin studies[2,3]. Some more recent methods estimate heritability by modeling the effects of genome-wide SNPs[4]. We refer to these methods as GREML-SNP methods, which reference inference on genomic relatedness, estimated from SNPs, with restricted maximum likelihood (REML). To decrease the influence of nonadditive genetic effects and environmental effects, samples are pruned so that no pair is related above a low threshold[4].

Instead of modeling the effects of SNPs, heritability can be estimated by examining how phenotypic similarity changes with relatedness. Relatedness is measured as the fraction of the genome that a pair shares in segments inherited from a common ancestor, called identical-by-descent (IBD) segments. (We note that what we call relatedness herein has sometimes been termed 'realized relatedness' to distinguish it from expected relatedness given a pedigree.[5]) 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, thus further implying that IBD-based methods can capture nearly all of the heritability of a trait. In contrast, GREML-SNP methods can capture only the fraction of the heritability explained by genotyped SNPs[4]. 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[4,6].

An IBD-based method, which we call the 'Kinship' method, examines how phenotypic similarity increases with relatedness for all pairs from a population sample[7]. When close relatives have more similar environments than distant relatives, the Kinship method overestimates heritability because it is unable to distinguish between similarity due to genetic effects and environmental effects. To decrease environmental bias, modeling of environmental effects shared between close relatives has been suggested[8,9]. 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 that case, modeling environmental covariance between close relatives alone would not remove environmental bias from the Kinship method. Although an extension to the Kinship method has been developed that models spatially distributed environmental effects[10], 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[5]. 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 resembles 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 resembles whether two independent tosses of a fair coin will both have the same outcome. Therefore, the siblings inherit the same copy of DNA from a parent half 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 fewer[5,11].

Here, we introduce RDR, a novel method for estimating heritability. RDR examines how much more or less related a pair is than would be expected from the relatedness of the parents. We call this deviation relatedness disequilibrium. Relatedness disequilibrium is due to random Mendelian segregations in the parents during meiosis and consequently is independent of almost all environmental effects. Unlike Sib-Regression, RDR can use any pair of individuals, provided that there is genetic information on the parents of the pair.

[1]deCODE genetics/Amgen Inc., Reykjavik, Iceland. [2]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [3]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. [4]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [5]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. *e-mail: alexander.young@bdi.ox.ac.uk; augustine.kong@bdi.ox.ac.uk

By using all pairs from a large sample with both parents genotyped, RDR can provide precise estimates of heritability with negligible bias due to environment. We applied RDR to estimate the heritability of 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)[12–14]. 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[14]. Any allele inherited by the phenotyped individual (proband) would also be present in one of its parents, implying that 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[15–17]. The proband's 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 procedure separates association due to direct genetic effects from association due to environment. Similarly, through use of random segregation, phenotypic variation can be decomposed into variation due to direct genetic effects alone and other components. Assuming that direct genetic effects are additive, and there is no gene-by-environment interaction, the decomposition is as follows (Supplementary Note):

$$\mathrm{var}(Y) = v_g + v_{e\sim g} + c_{g,e} + \mathrm{var}(\epsilon) \quad (1)$$

where v_g is the variance explained by direct genetic effects, and $h^2 = v_g / \mathrm{var}(Y)$ 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 $\mathrm{var}(\epsilon)$ 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 (equation (1)) leads to a decomposition of the covariance matrix of a vector of observations of a phenotype, $\boldsymbol{Y}$. Under certain assumptions (Supplementary Note):

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

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

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

Environmental bias properties of RDR. Through use of random segregation, both RDR and the TDT separate direct genetic effects from environmental effects. The TDT achieves this separation by conditioning on parental genotype, whereas RDR achieves this separation by conditioning on parental relatedness. The expectation of an offspring's genotype given its parents' genotype is one-half 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, $[R_{\mathrm{par}}]_{ij}$, is $[R_{\mathrm{par}}]_{ij}/2$, and any variation around this expectation comes from random segregation (Fig. 1, Supplementary Figure 1 and Supplementary Note). (Of note, this relationship does not hold for pairs in which one is the direct ancestor of the other, such as parent–offspring pairs.)

By fitting R and R_{par} jointly, RDR uses the variation in $[R]_{ij}$ around its expectation, $[R_{\mathrm{par}}]_{ij}/2$, 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 and thus 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 also excludes pairs in which 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 in which one is the direct ancestor of the other can introduce bias because they have an atypical relationship between $[R]_{ij}$ and $[R_{\mathrm{par}}]_{ij}$ (Fig. 1). However, they will comprise only a small fraction of the total pairs in a large population sample, even if multiple generations are genotyped. For our sample, approximately 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, because we were unable to detect bias arising from the inclusion of parent–offspring and grandparent–grandchild pairs (Online Methods and Supplementary Table 1), we did not remove individuals from our sample that also had 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 (i) Sib-Regression, (ii) the Kinship method, and (iii) the Kinship method allowing for an effect of shared family environment, which we denote the 'Kinship F.E.' method. We determined whether pairs shared a family environment according to whether they shared a mother in the deCODE Genealogy Database. The modeling of the environment in the Kinship F.E. model was similar to a recently proposed extension of the Kinship model[8]. We randomly selected 10,000 SNPs to act as causal SNPs in our simulations (Online Methods). The SNPs had a minimum minor allele frequency (MAF) of 0.5% and a median MAF of 22.8%. We simulated traits in a random subsample of 10,000 individuals who had both parents genotyped for all the methods other than Sib-Regression, for which we used all 54,888 individuals who had both parents genotyped.

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

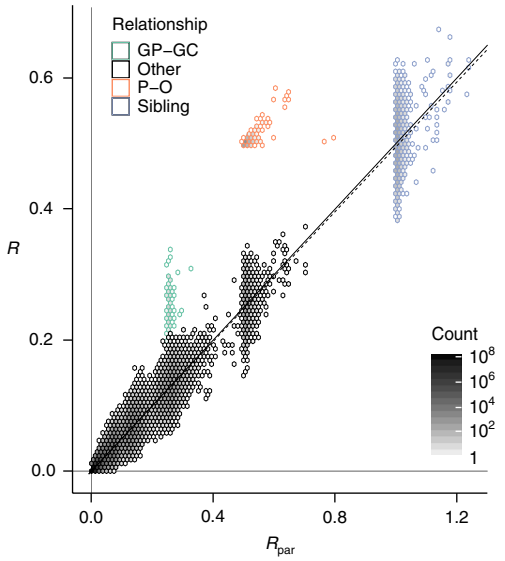

Fig. 1 | Relatedness disequilibrium. For all pairs of individuals *i*, *j* from 20,000 Icelanders with both parents genotyped, the relatedness of *i* and *j*, $[R]_{ij}$, is compared to the relatedness of the parents of *i* and the parents of *j*, $[R_{par}]_{ij}$. The number of pairs in each hexagonal bin is indicated by shading. Relationships determined according to 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 in which 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 -1×10^{-4}, gradient 0.493, and variance explained 84%. The small deviation of the regression line from the theoretical expectation is probably due to some IBD segments that are shared between parents being broken by recombination, thus resulting in a small fraction of segments in the offspring being too small to detect. Relatedness disequilibrium is the variation in $[R]_{ij}$ around $[R_{par}]_{ij}/2$. Relatedness disequilibrium is due to independent, random segregations in the parents, except for pairs in which one is the direct ancestor of the other.

We simulated a trait in which 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 modeling a family environment effect can remove bias from the Kinship method in certain circumstances. However, when indirect genetic effects from relatives are present, modeling the family environment is ineffective at removing bias. As a demonstration, 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, thus yielding 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 $c_{g,e} \approx 28\%$, thereby bringing the total variance explained by parent and offspring genotype to ~78%.

We found that the Kinship method greatly overestimated the heritability of the genetic nurturing trait. Modeling of the family environment only slightly decreased the bias, and, on average, the Kinship F.E. estimates of heritability were more than twice the true value. This result was due to the fact that parental genetic nurturing effects induce correlations between all pairs with nonzero parental relatedness, not just those sharing a family environment, which 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 we adjusted for 20 genetic principal components (regional trait; Table 1). After we adjusted for 100 principal components, the mean Kinship F.E. heritability estimate was 57.6% (s.e. 0.21%), a value still considerably larger than the true heritability of 40%. In contrast, the RDR estimates were approximately unbiased, because relatedness disequilibrium is caused by random segregations and therefore is uncorrelated with regional colocalization.

In some cases, IBD-based methods such as RDR do 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%, and a median MAF of 0.26% (Supplementary Note). RDR captured ~88% of the variance explained by the rare SNPs.

We found that RDR estimates were insensitive to nonadditive 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 nonadditive 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 to which RDR was applied, heritability estimates were obtained from the Kinship and Kinship F.E. methods (Online Methods, Table 2 and Fig. 2). For 11 of the 14 traits, the Kinship F.E. estimate, h^2_{KinFE}, is larger than the RDR estimate, h^2_{RDR} (average $h^2_{\text{KinFE}} - h^2_{\text{RDR}} = 12.1\%$). We found that h^2_{KinFE} was statistically significantly higher than h^2_{RDR} ($P < 0.05$, one-sided Z-test assuming that h^2_{KinFE} and h^2_{RDR} are independent, so *P* values represent an upper bound) for educational attainment ($h^2_{\text{KinFE}} - h^2_{\text{RDR}} = 35.4\%$, $P < 2.2\times10^{-4}$), height ($h^2_{\text{KinFE}} - h^2_{\text{RDR}} = 22.6\%$, $P < 1.3\times10^{-6}$), body mass index (BMI) ($h^2_{\text{KinFE}} - h^2_{\text{RDR}} = 17.8\%$, $P < 4.3\times10^{-3}$), and age at first child in women ($h^2_{\text{KinFE}} - h^2_{\text{RDR}} = 10.9\%$, $P < 0.043$). We found no evidence that differences between h^2_{KinFE} and h^2_{RDR} were driven by atypical properties of the sample with both parents genotyped or by differences in mean trait values among the regions of Iceland (Supplementary Note and Supplementary Table 5).

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

There were 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[19], denoted h^2_{twin}, which were available for 9 of the 14 traits (Online Methods, Table 2, Fig. 2 and Supplementary Table 6). The difference $h^2_{\text{twin}} - h^2_{\text{RDR}}$ was greater than

Table 1 | Comparison of heritability estimates for simulated traits

Trait	RDR		Kinship		Kinship F.E.		Sib-Regression	
	Estimate (%)	s.e. (%)	Estimate (%)	s.e. (%)	Estimate (%)	s.e. (%)	Estimate (%)	s.e. (%)
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

The mean heritability estimates along with their standard errors, expressed as a percentage of the phenotypic variance, from four different methods (RDR, Kinship, Kinship F.E., and Sib-Regression) for different simulated traits. The true (narrow-sense) heritability of each trait was 40%. We simulated 500 replicates of each trait on the basis of Icelandic genetic data from a random subsample of 10,000 individuals with both parents genotyped (Methods), except for Sib-Regression, for which we used all 54,888 individuals. Ten thousand SNPs with a median MAF of 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 shared mothers was added; for the genetic nurturing trait, the genotypes of the parents were also given effects to simulate parental genetic nurturing effects[14]. For the regional trait, the Kinship and Kinship F.E. methods also included adjustment for 20 genetic principal components.

zero and statistically significant ($P<0.05$, one-sided z test) for all nine traits, with an average difference of 33.2%. For Sib-Regression, the average difference $h^2_{\text{twin}}-h^2_{\text{sib}}$ was 26.4%. The finding that both RDR and Sib-Regression estimates were substantially lower than twin-study estimates may 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, generally capture a smaller fraction of the full heritability of a trait than IBD-based methods, such as RDR, thus 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_{par}, and $R_{o,\text{par}}$ are replaced by estimates from a set of SNPs: R^{snp}, $R^{\text{snp}}_{\text{par}}$, and $R^{\text{snp}}_{o,\text{par}}$. The only difference between RDR-SNP and GREML-SNP is that RDR-SNP also fits $R^{\text{snp}}_{\text{par}}$ and $R^{\text{snp}}_{o,\text{par}}$ in addition to R^{snp}.

To compare RDR-SNP to typical GREML-SNP analysis, we simulated traits in a subset of the UK Biobank[20] in which genotype data on both parents were 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, thus leaving 937 individuals (Supplementary Note).

Table 2 | Heritability estimates

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 cholesterol	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	–	

For each trait, the sample size used for the RDR, Kinship F.E., RDR-SNP, and relatedness-thresholded (RELT)-SNP methods is given under *n*, and the sample size for Sib-Regression is given under Sib-pairs. Each heritability estimate (Est.) is expressed as a percentage of the phenotypic variance and is followed by its standard error. RDR, Kinship F.E., RDR-SNP, and RELT-SNP estimates were from the exact same Icelandic samples with both parents genotyped, which 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. To maximize sample size, Sib-Regression estimates are from all genotyped Icelandic sibling pairs available without year-of-birth restrictions. Twin study estimates are from the Swedish Twin Registry[15], apart from for education, which is from a meta-analysis of Scandinavian twin studies[23] (Supplementary Table 6). AFCW, age at first child in women; AFCM, age at first child in men; menarche, age at menarche (years); education, educational attainment (years); HDL, high-density lipoprotein; glucose, fasting glucose; MCH, mean cell hemoglobin; MCHC, mean cell hemoglobin concentration; MCV, mean cell volume.

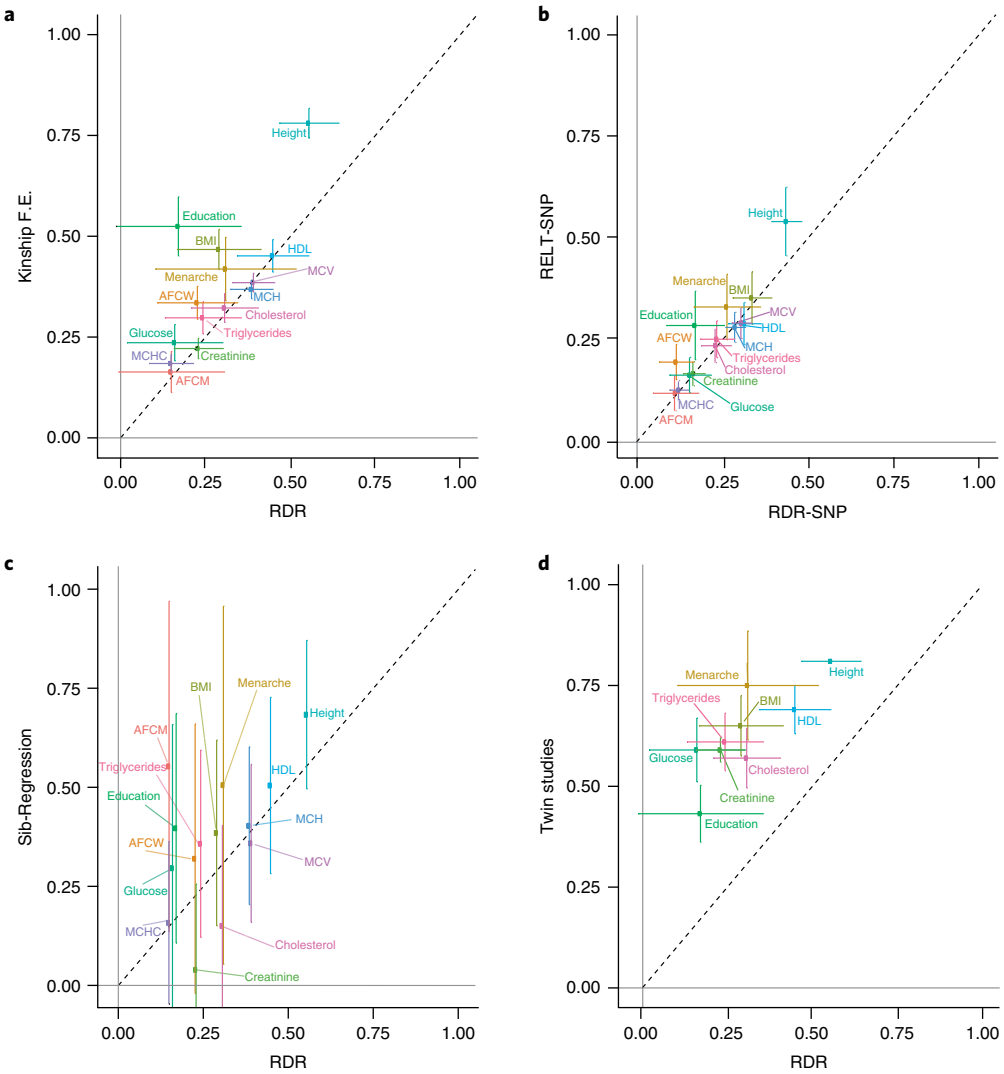

Fig. 2 | Comparison of heritability estimates from different methods. Horizontal intervals show ±1.96 s.e. for the estimates on the *x* axis, and vertical intervals show ±1.96 s.e. for the estimates on the *y* axis. Numerical values are in Table 2. **a**, Comparison of RDR to Kinship F.E. **b**, Comparison of RDR-SNP to RELT-SNP. **c**, Comparison of RDR to Sib-Regression[5] estimates. Intervals for the RDR estimates are not shown to better display Sib-Regression intervals. **d**, Comparison to published twin-study estimates from the Swedish Twin Registry[19], apart from for education, which is from a meta-analysis of Scandinavian twin studies[23] (Supplementary Table 6). 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 hemoglobin; MCHC, mean cell hemoglobin concentration; MCV, mean cell volume.

We randomly sampled 11,771 SNPs to act as causal SNPs, and we calculated R^{snp}, R^{snp}_{par}, and $R^{snp}_{o,par}$ 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 and thus have both direct and parental genetic nurturing effects. Let δ be the direct effect of a SNP, and let η be the parental genetic nurturing effect. The effect of the transmitted allele is therefore $(\delta+\eta)$. GREML-SNP uses only transmitted alleles and consequently is unable to separate the variance from the direct effect alone, which is proportional to δ^2, from the variance explained by the combined direct and parental genetic nurturing effects, which is proportional to $(\delta+\eta)^2$. We performed a theoretical investigation (Supplementary Note) and simulated 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, which is similar to the estimated ratio for educational attainment in Iceland[14]. The direct effects explained 20% of the phenotypic variance, thus implying that the total variance explained by transmitted alleles is $\left(1+\frac{1}{3}\right)^2 \times 20\% \approx 35.56\%$, a value much larger than the heritability of 20%.

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

Evidence of bias in GREML-SNP estimates from analysis of Icelandic data. 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^{snp} only for those pairs whose relatedness is less than the threshold. We call this approach relatedness-thresholded (RELT)-SNP. If the same relatedness threshold is applied, GREML-SNP and RELT-SNP estimates from large samples would be expected to be highly similar under most conditions. By applying RELT-SNP to the simulated traits in the UK Biobank, we showed that RELT-SNP and GREML-SNP gave very similar estimates and exhibited 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 we used the causal variants to calculate R^{snp}, R^{snp}_{par}, and $R^{snp}_{o,par}$, 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.), a value close to the variance explained by combined direct and genetic nurturing effects, ~73.3%. In contrast, the 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 by 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 $h^2_{RELT-SNP}$ was statistically significantly higher than $h^2_{RDR-SNP}$ ($P < 0.05$, one-sided z test under the assumption that $h^2_{RDR-SNP}$ and $h^2_{RELT-SNP}$ are independent, so P values represent an upper bound) for height ($\frac{h^2_{RELT-SNP}}{h^2_{RDR-SNP}} = 1.24$, $P < 0.015$), age at first child in women ($\frac{h^2_{RELT-SNP}}{h^2_{RDR-SNP}} = 1.72$, $P < 7.6 \times 10^{-3}$), and educational attainment (years) ($\frac{h^2_{RELT-SNP}}{h^2_{RDR-SNP}} = 1.69$, $P < 0.027$) (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 showed that removing close relatives does not remove bias due to indirect genetic effects from relatives. Our results suggest that GREML-SNP estimates can 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 of a substantial contribution from indirect genetic effects from parents and siblings[14]. Those results suggest that educational-attainment heritability estimates from GREML-SNP[8,21] and the recently proposed extension to the Kinship model[8] are likely to be upwardly biased. We estimated that GREML-SNP estimates of the heritability of educational attainment may be inflated by a factor of approximately 1.69. This inflation factor is consistent with genetic nurturing effects around 30% the size of direct effects, in agreement with an estimate based on a polygenic score[14] and within-family analyses in other populations[22].

RDR, like other methods using IBD segments[7,8], may underestimate the heritability due to rare variants. However, any underestimation due to rare variants will be less than that for GREML-SNP methods applied to typical genotyping arrays. By using IBD segments, RDR captures substantially more of the variance from rare variants, approximately 88% for variants with MAF between 1% and 0.1%. Consequently, 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 finding that Sib-Regression estimates were close to RDR estimates on average does not suggest substantial underestimation of heritability by RDR as a result of 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 may 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, and this is the main reason that our current study was limited to the Icelandic population. However, our results suggest 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.

URLs. Educational attainment categories, http://uis.unesco.org/en/isced-mappings/.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0178-9.

Received: 2 January 2018; Accepted: 29 June 2018;
Published online: 13 August 2018

References

1. Sesardic, N. *Making Sense of Heritability* (Cambridge Univ. Press, Cambridge, 2005).
2. Polderman, T. J. C. et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
3. Boomsma, D., Busjahn, A. & Peltonen, L. Classical twin studies and beyond. *Nat. Rev. Genet.* **3**, 872–82 (2002).
4. Yang, J., Zeng, J., Goddard, M. E., Wray, N. R. & Visscher, P. M. Concepts, estimation and interpretation of SNP-based heritability. *Nat. Genet.* **49**, 1304–1310 (2017).
5. Visscher, P. M. et al. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet.* **2**, e41 (2006).
6. Speed, D. et al. Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* **49**, 986–992 (2017).
7. Zaitlen, N. et al. Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS Genet.* **9**, e1003520 (2013).
8. Hill, W. D. et al. Genomic analysis of family data reveals additional genetic effects on intelligence and personality. https://doi.org/10.1038/s41380-017-0005-1 (2018).
9. Muñoz, M. et al. Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank. *Nat. Genet.* **48**, 980 (2016).
10. Heckerman, D. et al. Linear mixed model for heritability estimation that explicitly addresses environmental variation. *Proc. Natl. Acad. Sci. USA* **113**, 7377–7382 (2016).
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.* **93**, 865–875 (2013).
12. Baud, A. et al. Genetic variation in the social environment contributes to health and disease. *PLoS Genet.* **13**, e1006498 (2017).

13. Bijma, P. Estimating indirect genetic effects: precision of estimates and optimum designs. *Genetics* **186**, 1013–1028 (2010).
14. Kong, A. et al. The nature of nurture: effects of parental genotypes. *Science* **359**, 424–428 (2018).
15. Spielman, R. S., McGinnis, R. E. & Ewens, W. J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–16 (1993).
16. Ewens, W. J. & Spielman, R. S. The transmission/disequilibrium test: history, subdivision, and admixture. *Am. J. Hum. Genet.* **57**, 455–464 (1995).
17. Thomson, G. Mapping disease genes: family-based association studies. *Am. J. Hum. Genet.* **57**, 487–498 (1995).
18. Carey, G. Sibling imitation and contrast effects. *Behav. Genet.* **16**, 319–341 (1986).
19. Pedersen, N. L., Lichtenstein, P. & Svedberg, P. The Swedish Twin Registry in the Third Millennium. *Twin Res.* **5**, 427–432 (2002).
20. Bycroft, C. et al. Genome-wide genetic data on ~500,000 UK Biobank participants. https://doi.org/10.1101/166298 (2017).
21. Krapohl, E. & Plomin, R. Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol. Psychiatry* **21**, 437–443 (2016).
22. Lee, J. et al. Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. https://doi.org/10.1038/s41588-018-0147-3 (2018).
23. Branigan, A. R., Mccallum, K. J. & Freese, J. Variation in the heritability of educational attainment: an international meta-analysis. *Soc. Forces* **92**, 109–140 (2013).

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.

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 interests

The following authors affiliated with deCODE Genetics are or 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., and A.K.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0178-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.I.Y. or A.K.

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

Methods

Icelandic sample. All participating subjects donating biological samples provided signed informed consent, and the study was approved by the Data Protection Commission of Iceland and the National Bioethics Committee of Iceland. The personal identities associated with the phenotypes and biological samples were encrypted by a third-party system provided by the Icelandic Data Protection Authority.

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

The deCODE Genealogy Database is a comprehensive database including 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 (which has been conducted regularly since the year 1700), church books, and other available information. The database is particularly comprehensive for the past 200 years. The database includes, when known, information on the parents of each individual, sex, year of birth (YOB), and, if applicable, year of death.

We restricted our analyses to genotyped individuals who had both genetic parents genotyped and all four grandparents in the deCODE Genealogy Database. This procedure left 54,888 individuals. A summary of sample restrictions and other information can be found in the Reporting Summary.

The individuals and their parents had all been phased, and segments shared IBD, both within and between individuals, were determined by long-range phasing[25,26]. To decrease bias due to segments being incorrectly called as IBD, we restricted our analyses to segments longer than 5 cM. Of note, sex-chromosomes were not included.

To provide a measure of ascertainment bias, we compared the number of years of education between the individuals with both parents genotyped and the full set of individuals with education data. The mean years of education for individuals with both parents genotyped was 15.07 compared with 13.63 for the whole sample with education data. This was partly because individuals with both parents genotyped were born later than average, and the mean level of education has increased over time. After YOB, YOB^2, and YOB^3 were regressed out, the sample with both parents genotyped still had 0.32 years more education on average, compared with a standard deviation of 3.39 years. Thus, our results were slightly biased toward individuals with higher socioeconomic status, which, for many traits, is expected to increase heritability[27,28].

Trait measurements. To provide a measure of educational attainment, we used information on the number of years of schooling, which was available for 63,508 individuals and originated from questionnaires administered in deCODE's various disease projects and from routine assessments of elderly nursing home residents. Because the data have been gathered over the years for the purpose of descriptive demographics rather than for phenotypic use, the questions were originally not standardized across projects, and many of them have categorical responses. For this study, to make the data as consistent as possible regarding the educational-attainment trait studied in the published meta-analysis[29], we made efforts to map the responses to the questionnaires into the UNESCO ISCED classification (see URLs). In particular, the final quantitative measure used, before sex and YOB adjustments, ranged from a minimum of 10 years to a maximum of 20 years.

Height and 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[30,31]. Approximately 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; and 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 in comparability with other studies that have used one time point only, we took only the first measurement from each individual.

Information on AAFC was extracted from the deCODE Genealogy Database. Age at menarche was determined from the answer to the question 'How old were you when your menstruation started?', as detailed elsewhere[32].

Apart from educational attainment, traits were quantile-normalized within each sex. Educational attainment was not quantile-normalized, because the measurements fall into discrete categories of years of education. The traits were regressed on sex; YOB, YOB^2, and YOB^3; and the interactions of sex with YOB, YOB^2, and YOB^3. The residuals of this regression were then used as the phenotype, $\boldsymbol{Y}$, when fitting the models described below. To ensure that our heritability estimates corresponded to the adult nonelderly population, we further restricted our analysis to individuals born between 1951 and 1995 for height, and between 1951 and 1997 for BMI and the traits measured from blood. (Of note, for Sib-Regression, the YOB restrictions were not applied to maximize the sample size.)

Identification of siblings. For the Sib-Regression estimator, we obtained the relatedness for all pairs of genotyped individuals who shared both parents in the genealogy. To ensure that we used only true full siblings, we clustered the pairs by relatedness into four clusters with k-means clustering: unrelated, half sibling, full sibling, and monozygotic twin. This procedure left 127,264 full-sibling pairs comprising 70,317 unique individuals, whose relatedness distribution had a mean of 0.502 and a standard deviation of 0.0382. To maximize the precision of the Sib-Regression estimator, we did not restrict by YOB or by the number of parents genotyped; thus, the sample used was different from the sample used for the other estimators.

Calculation of IBD relatedness matrices. To calculate R, R_{par}, and $R_{\text{o,par}}$, we used formulae based on the genetic covariance in a population descending from a finite number of ancestors[33] (Supplementary Note):

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

where K_0 is the mean kinship over all pairs in the population, and IBD_{ij}^{kl} is the proportion of the maternally inherited haplotype of i shared IBD with the paternally inherited haplotype of j;

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

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

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

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

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

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

$$R_{\text{par}}^{\text{snp}} = (2l)^{-1}G_{\text{par}}G_{\text{par}}^T; R_{\text{o,par}}^{\text{snp}} = (2l)^{-1}(GG_{\text{par}}^T + G_{\text{par}}G^T)$$

The matrices are calculated in this way to ensure that estimates of v_g, $v_{e\sim g}$, 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 among many of the Illumina genotyping arrays used to genotype the Icelandic sample[24]. We used this set of SNPs 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 a MAF less than 1%, thus leaving 605,966 SNPs. For the simulated traits, we also computed relatedness matrices from only the causal SNPs (Supplementary Note).

Computing RDR estimates. The RDR covariance model is

$$\text{cov}(\boldsymbol{Y}) = v_g R + v_{e\sim g}R_{\text{par}} + c_{g,e}R_{\text{o,par}} + \sigma^2\text{I}$$

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:

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

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, because these pairs violate the relationship between $[R]_{ij}$ and $[R_{\text{par}}]_{ij}$ required for removal of environmental bias from estimation of v_g (Fig. 1 and Supplementary Note). We also investigated fitting the model by unconstrained restricted maximum likelihood in *GCTA*[34], under the assumption that the trait follows a multivariate normal distribution:

$$\boldsymbol{Y} \sim N(\mu, v_g R + v_{e\sim g} R_{\text{par}} + c_{g,e} R_{o,\text{par}} + \sigma^2 \text{I})$$

For the maximum-likelihood method, one can remove only individuals and all the pairs including that individual, not arbitrary pairs. Approximately 30% of the sample with both parents genotyped had an ancestor who also had both parents genotyped. We therefore did not exclude individuals such that no parent–offspring and no grandparent–grandchild pairs would be present, because doing so 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, and there was no consistent advantage in bias evident from fitting the model by least squares after exclusion of 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*:

$$\boldsymbol{Y} \sim N(\mu, v_g R^{\text{snp}} + v_{e\sim g} R^{\text{snp}}_{\text{par}} + c_{g,e} R^{\text{snp}}_{o,\text{par}} + \sigma^2 \text{I})$$

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 $\boldsymbol{Y}$:

$$\boldsymbol{Y} \sim N(X_{\text{kin}} b, v_g R + \sigma^2 \text{I})$$

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

$$\boldsymbol{Y} \sim N(X_{\text{kin}} b, v_g R + v_c C + \sigma^2 \text{I})$$

where $[C]_{ij} = 1$ if i and j shared a mother according to the deCODE Genealogy Database, or $[C]_{ij} = 0$ otherwise. For all of the simulated traits other than the regional trait, X_{kin} was a constant. For the regional trait, it also included the top 20 genetic principal components. In the real-trait analysis, X_{kin} included the top 20 genetic principal components. For both the Kinship and Kinship F.E. methods, we estimated model parameters acccording to unconstrained restricted maximum likelihood in *GCTA*[34].

Computing RELT-SNP estimates. To compute RELT-SNP estimates, we regressed off-diagonal elements of the phenotypic covariance matrix onto elements of R^{snp}, excluding elements of R^{snp} greater than 0.05. Let 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 $\hat{\text{b}}$ be the least-squares estimate of the vector of regression coefficients of the phenotype on X. Then we formed the sample phenotypic covariance matrix as $S = (\boldsymbol{Y} - X\hat{\text{b}})(\boldsymbol{Y} - X\hat{\text{b}})^T$, where $\boldsymbol{Y}$ is the vector of phenotype observations. Estimates of v_g were computed by regressing off-diagonal elements of S, $[S]_{ij}$, on off-diagonal elements of R^{snp}, $[R^{\text{snp}}]_{ij}$, excluding pairs for which $[R^{\text{snp}}]_{ij} > 0.05$.

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[35], which takes into account the dependence between elements of 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 with 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 in which 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 for which the variance explained by the top 20 principal components exceeded 1%: height, age at first child in men and women, and educational attainment. We report 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 the Kinship, GREML-SNP, and RELT-SNP methods. RDR and RDR-SNP, in contrast, do not require such arbitrary decisions, because they separate genetic and environmental effects in a principled way.

Computing Sib-Regression estimates. To obtain Sib-Regression estimates[5], we fit the regression model

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

for all i,j such that i and j are full siblings. We fit the regression model by least squares with custom R code. The estimate of v_g is then minus one-half of the estimated regression coefficient. We compared estimated standard errors with the approximate formula given in the original Sib-Regression paper[5] and estimating standard errors by treating Sib-Regression as a standard univariate linear regression with uncorrelated observations. For the additive simulated trait, both methods 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 with 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 (described above). We filtered the SNPs so that the minimum imputation information was 0.9999, removing approximately half the SNPs. From 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 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 >0.05. We sampled 100 such SNPs from each chromosome, thus resulting in 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 standardized genotypes so that each SNP's genotype vector had a sample mean of zero and a sample variance of one. Let G represent the matrix of standardized genotypes at the 10,000 causal SNPs. We sampled additive effects of SNPs from a normal distribution. Let $\beta \sim N(0, I)$ represent the vector of SNP effects. The additive genetic component, A, was calculated as $A = G\beta$ and then scaled to explain 40% of the phenotypic variance. Details on simulation of environmental components are provided in 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 <5% or >0.5% missing genotypes. This procedure yielded 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 $l = 11,771$. We standardized offspring genotypes so that the genotypes at each SNP had a mean of zero and a variance of 1. Let G be the matrix of standardized offspring genotypes. Here, we describe simulation of the direct, additive genetic component of the traits. Further detail can be found in the Supplementary Note. For each trait, we simulated a normally distributed vector of effects for the l SNPs: $\beta \sim N(0, 0.2l^{-1}I)$. The additive genetic component of the trait, A, was then calculated as $A = G\beta$.

Selection of estimates from twin studies. The Swedish Twin Registry[19] is a large sample of twins from a population with similar cultural and genetic composition to that of Iceland, thus providing the most precise and valid comparison possible based on published data[23,36–40]. For BMI and traits measured from blood, unlike our estimates, the Swedish Twin Registry estimates do not exclude older individuals. This difference is unlikely to account for the higher estimates in the Swedish Twin Registry, because twin correlations and heritability estimates are generally lower in the older population[2].

We took the heritability estimate from the additive-common-environment (ACE) model[2,3] when provided. ACE estimates were not provided for the blood lipid traits, but monozygotic- and dizygotic-twin correlations were provided[39]. We used these values to obtain the moment-based estimate of the heritability under the ACE model with 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, because the ACE estimate was not provided, and neither were the twin correlations, we took the published heritability estimate from the additive-dominance-environment model. The studies and methods used are summarized in Supplementary Table 6. For height, because heritability estimates were provided for males and females separately, 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 difference would be expected to increase our heritability estimates for height and BMI relative to the twin estimates as a result of decreased measurement error. For education, we used a meta-analysis of twin studies in Scandinavian countries, including Sweden, to provide a more precise estimate[23]. We were unable to find published estimates

based on the Swedish Twin Registry for the hemoglobin traits and for age at first child, so we excluded them from the comparison.

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

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/.

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 reasonable request. Applications for access to the UK Biobank data can be made on the UK Biobank website: http://www.ukbiobank.ac.uk/register-apply/.

References

24. Gudbjartsson, D. F. et al. Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
25. Kong, A. et al. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068–75 (2008).
26. Kong, A. et al. Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868–874 (2009).
27. Tuvblad, C., Grann, M. & Lichtenstein, P. Heritability for adolescent antisocial behavior differs with socioeconomic status: gene-environment interaction. *J. Child Psychol. Psychiatry* **47**, 734–743 (2006).
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.* **125**, 392 (1999).
29. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. https://doi.org/10.1038/nature17671 (2016).
30. Gudbjartsson, D. F. et al. Many sequence variants affecting diversity of adult human height. *Nat. Genet.* **40**, 609–15 (2008).
31. Thorleifsson, G. et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).
32. Elks, C. E. et al. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. *Nat. Genet.* **42**, 1077–1085 (2010).
33. Young, A. I. & Durbin, R. Estimation of epistatic variance components and heritability in founder populations and crosses. *Genetics* **198**, 1405–1416 (2014).
34. 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).
35. Tamar, S. Confidence intervals for heritability via Haseman-Elston regression. *Stat. Appl. Genet. Mol. Biol.* **16**, 259 (2017).
36. Carlsson, S., Ahlbom, A., Lichtenstein, P. & Andersson, T. Shared genetic influence of BMI, physical activity and type 2 diabetes: a twin study. *Diabetologia* **56**, 1031–1035 (2013).
37. Silventoinen, K. et al. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Res.* **6**, 399–408 (2003).
38. Baker, J. H., Thornton, L. M., Bulik, C. M., Kendler, K. S. & Lichtenstein, P. Shared genetic effects between age at menarche and disordered eating. *J. Adolesc. Heal.* **51**, 491–496 (2012).
39. Rahman, I. et al. Genetic dominance influences blood biomarker levels in a sample of 12,000 Swedish elderly twins. *Twin Res. Hum. Genet.* **12**, 286–294 (2009).
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.* **4**, e001467 (2015).

nature research

Corresponding author(s): Alexander Young, Augustine Kong

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Not applicable
Data analysis	Custom computer code is available at https://github.com/AlexTISYoung/RDR

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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 the UK Biobank data can be made on the UK Biobank website.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All available Icelandic individuals with genotype data on both parents and all four grandparents in the deCODE Genealogy Database: 54,888.
Data exclusions	We 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.
Replication	Exact replication is not applicable/expected for heritability estimates. However, results consistent with our analysis have been found using within-family analyses in the UK Biobank
Randomization	Not applicable
Blinding	Not applicable

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging