A quantitative genetic model for indirect genetic effects and genomic imprinting under random and assortative mating

Ilse Krätschmer[1*], Matthew R. Robinson[1*]

[1]Institute of Science and Technology Austria, Klosterneuburg, Austria

[*]To whom correspondence should be addressed:
ilse.kraetschmer@ist.ac.at, matthew.robinson@ist.ac.at

An individual's phenotype reflects a complex interplay of the direct effects of their DNA, epigenetic modifications of their DNA induced by their parents, and indirect effects of their parents' DNA. Here, we derive how the genetic variance within a population is changed under the influence of indirect maternal, paternal and parent-of-origin effects under random mating. We also consider indirect effects of a sibling, in particular how the genetic variance is altered when looking at the phenotypic difference between two siblings. The calculations are then extended to include assortative mating (AM), which alters the variance by inducing increased homozygosity and correlations within and across loci. AM likely leads to covariance of parental genetic effects, a measure of the similarity of parents in the indirect effects they have on their children. We propose that this assortment for parental characteristics, where biological parents create similar environments for their children, can create shared parental effects across traits and the appearance of cross-trait AM. Interestingly, the genetic variances is increased under AM for the child-mother-father design, while it is decreased for the sibling difference. Our results demonstrate that it is not possible to get unbiased estimates of direct genetic effects without controlling for parental and parent-of-origin effects.

Keywords: quantitative genetics model, epigenetic, parent-of-origin, indirect effects

Introduction

In humans, parental influence persists long beyond childhood, shaping adult health and disease[1]. A basic principle of genetics is that observable characteristics result from two sources: the expression of an individual's DNA and the environment that they experience[2]. Parental characteristics shape a child's environment, giving rise to indirect genetic effects, where the genotypes of both the mother and father influence the traits of their children, beyond simply the alleles that are inherited[2], as displayed in Figure 1. Many studies have now shown that failing to separate the effects of an individual's DNA from the indirect effects of their parental genotypes results in confounding within genetic association studies[3,4,5,6]. Moreover, it has been shown that maternal effects cause the appearance of parent-of-origin effects that can mimic genomic imprinting, and vice versa[7].

Previous quantitative genetics models derived the genetic population variance under the influence of maternal effects[8,6], or imprinting[9] or both[10,7,11] under random mating. *Wolf et.al.* demonstrated how indirect maternal effects contribute to the variance under inbreeding[8]. Other studies[12,13,14] provided a general framework to show the influence of confounding in family-based genome-wide association studies (GWAS), and how polygenic risk scores (PGS) are biased under assortative mating in the presence of indirect parental effects.

But so far, none of the models considered direct, indirect maternal and paternal, and parent-of-origin genetic effects jointly. Here, we first derive the genetic variance for a single locus under random mating assuming that all four effects (direct, indirect maternal and paternal, and parent-of-origin) influence an individual's phenotype. We then extend the model to include also the indirect effects of a sibling and consider another common family-based design in human studies: the phenotypic difference between two siblings. Next, we show how the genetic variance changes in the presence of assortative mating (or inbreeding). Lastly, we demonstrate how to include multiple loci in the model. We then also verify our theoretical model with forward-in-time simulations, based on genotypes from the 1000 Genomes project.

Though we primarily consider humans and use the corresponding terms, we want to

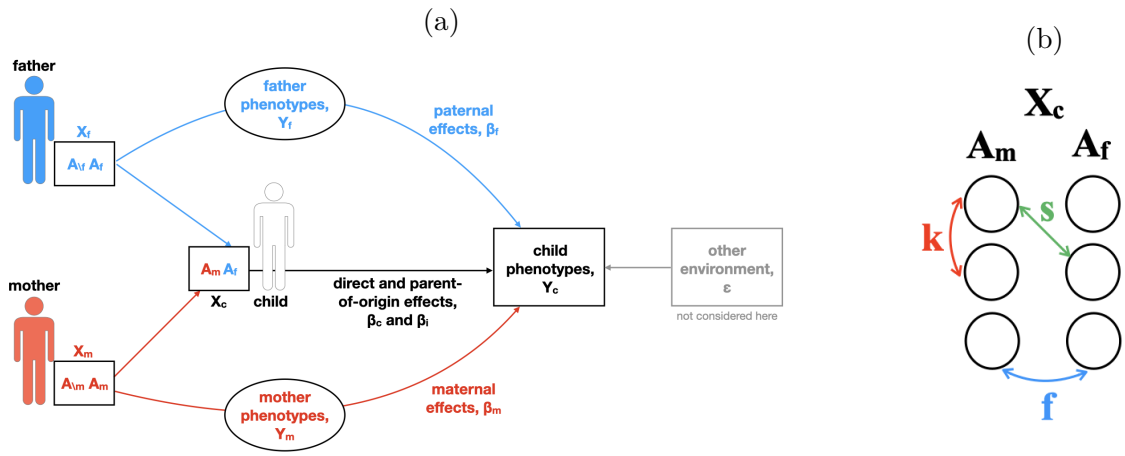

Figure 1: Definition of variables. (a) Direct ($\boldsymbol{\beta_c}$), parent-of-origin ($\boldsymbol{\beta_i}$) and indirect maternal ($\boldsymbol{\beta_m}$) and paternal ($\boldsymbol{\beta_f}$) genetic effects on the phenotypes of children ($\boldsymbol{Y_c}$). The parental genotypes (X_m, X_f) are split into transmitted alleles (A_m, A_f), which constitute the genotype of the child (X_c), and untransmitted ones ($A_{\backslash m}$, $A_{\backslash f}$). Parental effects can be mediated through any parental phenotype (Y_m, Y_f). The variable $\boldsymbol{\epsilon}$ represents variations created by other environmental effects experienced by the children. (b) Correlations within and between loci introduced through AM. The child's DNA (X_c) consists of maternally inherited alleles (A_m) and paternally inherited ones (A_f). The correlation between alleles at a locus is given by f, the one between alleles at different loci within the same gamete by k and the one between pairs of A_m and A_f alleles at different loci by s.

point out that our calculations are not limited to humans, but generally applicable to diploid sexually reproducing species that rear their offspring.

Single-locus model under random mating

We first consider a simple single autosomal locus with two alleles (A_1 and A_2) with frequencies q_1 and q_2 that can have the following four effects on the child's phenotype: direct (β_c), indirect maternal (β_m), indirect paternal (β_f)and parent-of-origin (imprinting, β_i). The alleles of the child at the locus are ordered, referring first to the maternally inherited allele and second to the paternally inherited one, i.e. A_mA_f. The genotypes X_c, X_m, X_f, are coded as 0, 1 and 2 for A_1A_1 homozygotes, A_1A_2 or A_2A_1 heterozygotes and A_2A_2 homozygotes, respectively. We use subscripts m, f, c, and i to denote maternal, paternal, direct (child) and parent-of-origin (imprinting) effects, genotypes and genotypic

values. For heterozygous children, the parent-of-origin effect is positive for maternally inherited A_2 and negative for paternally inherited A_2. From these definitions, the possible genotypes, X_c, and phenotypes, Y_c, of the child given the genotype of the parents and imprinting, are shown in Table 1. The phenotypic values represent the deviations from the case where both parents and child have A_1A_1 genotypes.

Table 1 shows that indirect parental genetic effect genotypic values, $X_m\beta_m$ and $X_f\beta_f$, and parent-of-origin effect genotypic values, $X_i\beta_i$, are confounded due to the fact that not every child genotype can be produced by all parental genotypes. For example, parents who are both the same type of homozygote (both either A_1A_1 or A_2A_2) will only have A_1A_1 and A_2A_2 children. Each of the two types of homozygous parents (A_1A_1 and A_2A_2) can produce only one of the two types of reciprocal heterozygote (A_1A_2 and A_2A_1). The children of A_1A_1 fathers all have $+\beta_i$, with the exception of A_1A_1 children whose parents are both A_1A_1. The children of A_2A_2 fathers all have $-\beta_i$, with the exception of A_2A_2 children whose parents are both A_2A_2. The reverse is true for mothers, where the children of A_1A_1 mothers all have $-\beta_i$, with the exception of A_1A_1 children whose parents are both A_1A_1, and the children of A_2A_2 mothers all have $+\beta_i$, with the exception of A_2A_2 children whose parents are both A_2A_2. This clearly shows that when there are indirect parental genetic effects, the phenotypes of the reciprocal heterozygotes will differ, creating the appearance of parent-of-origin effects. Likewise, differences between reciprocal heterozygotes due to parent-of-origin effects will appear as parental genetic effects.

Table 1 also provides the expected probability of the different parent-child combinations within the population (mating frequency), z, under random mating and Hardy-Weinberg equilibrium, where $(q_1 + q_2) = 1 = (q_1 + q_2)^2$, as well as the probabilities under assortative mating, z_{AM}, which will be discussed in the next section. The phenotypic values, Y_c, and expected probabilities, z, in Table 1 can be used to calculate the mean and variance contributed by direct, indirect parental genetic and parent-of-origin effects within the human population under random mating. The population mean is given by

$$\mu_{RM} = \sum_j Y_{cj} z_j = 2q_2(\beta_c + \beta_m + \beta_f), \tag{1}$$

Table 1: A simple single-locus model with indirect parental and parent-of-origin genetic effects. Maternal (X_m), paternal (X_f) and child's genotypes (X_c), phenotypes of the child (Y_c), mating frequencies under random mating (z) and under assortative mating (z_{AM}) taken from Ref.[15] where $S = 2q_1q_2(1+\rho)$. Parental heterozygous genotypes are not ordered and are used interchangeably. Refer to text for more details.

X_m	X_f	X_c (A_mA_f)	Y_c				z	z_{AM}
A_1A_1	A_1A_1	A_1A_1					q_1^4	$[q_1^4 + 2\rho q_1^3 q_2 + \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S} 4q_2^2]$
	A_1A_2	A_1A_1			$+\beta_f$		$q_1^3 q_2$	$[q_1^3 q_2 + \rho q_1^2 q_2(q_2 - q_1) - \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S} 2q_2(q_2 - q_1)]$
		A_1A_2	$+\beta_c$		$+\beta_f$	$-\beta_i$	$q_1^3 q_2$	$[q_1^3 q_2 + \rho q_1^2 q_2(q_2 - q_1) - \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S} 2q_2(q_2 - q_1)]$
	A_2A_2	A_1A_2	$+\beta_c$		$+2\beta_f$	$-\beta_i$	$q_1^2 q_2^2$	$[q_1^2 q_2^2 + \rho q_1 q_2(q_1^2 + q_2^2) + \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S}(-4q_1q_2)]$
A_1A_2	A_1A_1	A_1A_1		$+\beta_m$			$q_1^3 q_2$	$[q_1^3 q_2 + \rho q_1^2 q_2(q_2 - q_1) - \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S} 2q_2(q_2 - q_1)]$
		A_2A_1	$+\beta_c$	$+\beta_m$		$+\beta_i$	$q_1^3 q_2$	$[q_1^3 q_2 + \rho q_1^2 q_2(q_2 - q_1) - \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S} 2q_2(q_2 - q_1)]$
	A_1A_2	A_1A_1		$+\beta_m$	$+\beta_f$		$q_1^2 q_2^2$	$[q_1^2 q_2^2 - 2\rho q_1^2 q_2^2 + \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S}(q_1 - q_2)^2]$
		A_2A_1	$+\beta_c$	$+\beta_m$	$+\beta_f$	$+\beta_i$	$q_1^2 q_2^2$	$[q_1^2 q_2^2 - 2\rho q_1^2 q_2^2 + \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S}(q_1 - q_2)^2]$
		A_1A_2	$+\beta_c$	$+\beta_m$	$+\beta_f$	$-\beta_i$	$q_1^2 q_2^2$	$[q_1^2 q_2^2 - 2\rho q_1^2 q_2^2 + \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S}(q_1 - q_2)^2]$
		A_2A_2	$+2\beta_c$	$+\beta_m$	$+\beta_f$		$q_1^2 q_2^2$	$[q_1^2 q_2^2 - 2\rho q_1^2 q_2^2 + \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S}(q_1 - q_2)^2]$
	A_2A_2	A_1A_2	$+\beta_c$	$+\beta_m$	$+2\beta_f$	$-\beta_i$	$q_1 q_2^3$	$[q_1 q_2^3 + \rho q_1 q_2^2(q_1 - q_2) - \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S} 2q_1(q_1 - q_2)]$
		A_2A_2	$+2\beta_c$	$+\beta_m$	$+2\beta_f$		$q_1 q_2^3$	$[q_1 q_2^3 + \rho q_1 q_2^2(q_1 - q_2) - \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S} 2q_1(q_1 - q_2)]$
A_2A_2	A_1A_1	A_2A_1	$+\beta_c$	$+2\beta_m$		$+\beta_i$	$q_1^2 q_2^2$	$[q_1^2 q_2^2 + \rho q_1 q_2(q_1^2 + q_2^2) + \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S}(-4q_1q_2)]$
	A_1A_2	A_2A_1	$+\beta_c$	$+2\beta_m$	$+\beta_f$	$+\beta_i$	$q_1 q_2^3$	$[q_1 q_2^3 + \rho q_1 q_2^2(q_1 - q_2) - \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S} 2q_1(q_1 - q_2)]$
		A_2A_2	$+2\beta_c$	$+2\beta_m$	$+\beta_f$		$q_1 q_2^3$	$[q_1 q_2^3 + \rho q_1 q_2^2(q_1 - q_2) - \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S} 2q_1(q_1 - q_2)]$
	A_2A_2	A_2A_2	$+2\beta_c$	$+2\beta_m$	$+2\beta_f$		q_2^4	$[q_2^4 + 2\rho q_1 q_2^3 + \rho^2 q_1^2 q_2^2][1 + \frac{\rho}{S} 4q_1^2]$

where j loops over all possible genotypes of the child. The mean is not influenced by the parent-of-origin effects, as they are on average zero. The genetic variance for a single locus is:

$$V_{RM} = \sum_j Y_{cj}^2 z_j - \mu_{RM}^2 = 2q_1q_2(\beta_c^2 + \beta_m^2 + \beta_f^2 + \beta_i^2 + \beta_c\beta_m + \beta_c\beta_f + \beta_m\beta_i - \beta_f\beta_i). \quad (2)$$

More details on the calculations are provided in the Supporting Information.

Equation 2 shows that direct genetic effect genotypic values covary with both maternal and paternal genetic effect genotypic values, which in turn covary with parent-of-origin effect genotypic values, creating a greater degree of potential confounding between all four components than has been previously considered. The additive genetic variance among child's genotypes, apart from the expected terms due to direct effects (β_c^2), also contains components contributed by parental genetic effects (β_m^2, β_f^2), direct-parental genetic covariance ($\beta_c\beta_m$, $\beta_c\beta_f$), parent-of-origin effects (β_i^2), and parental-parent-of-origin

genetic covariance ($\beta_i\beta_m$, $\beta_i\beta_f$). Parent-of-origin effects contribute not only directly, but also dependent on their covariance with parental genetic effects, i.e. $\beta_i^2 + \beta_i\beta_m - \beta_i\beta_f$, which represents the degree to which loci with parental genetic effects are also imprinted loci in the child. Note the negative sign of $\beta_i\beta_f$, which is due to the definition of parent-of-origin effects (positive for the maternally inherited allele and negative for the paternal one).

The implications of Equation 2 are that phenotypic variation can be attributable to the single nucleotide polymorphism (SNP) markers of an individual (i.e., SNP heritability $h^2_{SNP} > 0$) and marker-phenotype associations can be discovered in population studies, but the underlying pattern of causality can be anything from: (i) the locus does not directly affect the expression of the trait ($\beta_c^2 = 0$), but instead only has an effect in one or both of the parents; to (ii) both direct and parental genetic effects may be present at a locus, but the parental effects may cancel out if they covary negatively; or (iii) any combination of terms and their covariance.

Differences in siblings

The model presented in the previous section can be extended to include genetic effects of a sibling on the child's phenotype, additionally to direct, indirect parental and parent-of-origin genetic effects. The derivations of population mean and variance including sibling effects (β_s) are given in the Supporting Information. However, it is more interesting to look at the relationship of the phenotypic and genotypic differences in siblings as the difference in siblings provides an estimate of direct genetics effects unbiased of indirect parental effects[16,17,18]. Typically, sibling analyses do not consider parent-of-origin effects, nor the indirect genetic effects from siblings. Using the phenotypic differences between child and sibling, ($Y_c - Y_s$), and the expected probabilities of the genotypes for the different sibling combinations, which are given in Table 2, we can calculate the expected mean and variance at population level. The total mean is 0, while the variance is

$$V_{s_{RM}} = \sum_j (Y_{cj} - Y_{sj})^2 z_j = 2q_1q_2(\beta_c^2 + \beta_i^2 + \beta_s^2 - 2\beta_c\beta_s), \tag{3}$$

which shows that components due to indirect parental effects do indeed cancel in the difference, but those due to indirect sibling and imprinting effects remain.

Importantly, if there is an indirect effect from the sibling on an individual's phenotype, the variance increases with the squared sibling's effects, but decreases with the covariance between the individual and its sibling. This results in a confounding of direct and indirect sibling effect genotypic values, which always oppose each other in sign, as shown in Table 2. Thus, any indirect sibling genetic effect will act to alter the variance due to the direct genetic effects. But it does appear possible to separate parent-of-origin effects within this experimental design under random mating.

Single-locus model under assortative mating

Assortative mating (AM) acts to alter the mating frequencies in the population as it increases parental similarity at trait-associated loci, which is reflected in the correlation, ρ, among the parents in their alleles. As a consequence of AM, the genotype frequencies are changed, and thus also the probabilities of parents producing offspring of a certain genotype, given as z_{AM} in Table 1. But the minor allele frequencies in the population remain invariant. The altered offspring frequencies, z_{AM}, are taken from Ref.[15]. This modification leaves the population mean unchanged under AM, but it alters the variance.

Under assortative mating, we find that a single locus with direct, indirect parental and parent-of-origin effects makes the following contribution to the genetic variance:

$$\begin{aligned} V_{AM} = {} & 2q_1q_2(\beta_c^2 + \beta_i^2 + \beta_m^2 + \beta_f^2 + \beta_c\beta_m + \beta_c\beta_f + \beta_m\beta_i - \beta_f\beta_i) \\ & + q_1q_2\rho(\beta_c^2 + 2\beta_m^2 + 2\beta_f^2 - \beta_i^2 + 4\beta_c\beta_m + 4\beta_c\beta_f + 4\beta_m\beta_f) \\ & + q_1q_2\rho^2(\beta_c^2 - \beta_i^2 + 2\beta_c\beta_m + 2\beta_c\beta_f - 2\beta_m\beta_i + 2\beta_f\beta_i + 4\beta_m\beta_f) \end{aligned} \tag{4}$$

The ρ term reflects a correlation among mates at causal SNP variants (or those correlated with an underlying causal variant) and is equivalent to the inbreeding coefficient, F. Equation 4 shows that the increase in the variance contributed by a locus under AM is dependent upon the relationships among the direct, parental genetic and parent-of-origin effects.

Table 2: Extension of Table 1 to include siblings and their differences. Mating frequencies are given in Table 1 which have to be multiplied with the factor f_z to account for the different sibling combinations. Parental heterozygous genotypes are not ordered and are used interchangeably.

X_m	X_f	X_c (A_mA_f)	X_s (A_mA_f)	Y_c					Y_s					$Y_c - Y_s$			f_z
A_1A_1	A_1A_1	A_1A_1	A_1A_1														1
	A_1A_2	A_1A_1	A_1A_1			$+\beta_f$					$+\beta_f$						$\frac{1}{2}$
			A_1A_2			$+\beta_f$		$+\beta_s$	$+\beta_c$		$+\beta_f$	$-\beta_i$		$-\beta_c$	$+\beta_i$	$+\beta_s$	$\frac{1}{2}$
		A_1A_2	A_1A_1	$+\beta_c$		$+\beta_f$	$-\beta_i$				$+\beta_f$		$+\beta_s$	$+\beta_c$	$-\beta_i$	$-\beta_s$	$\frac{1}{2}$
			A_1A_2	$+\beta_c$		$+\beta_f$	$-\beta_i$	$+\beta_s$	$+\beta_c$		$+\beta_f$	$-\beta_i$	$+\beta_s$				$\frac{1}{2}$
	A_2A_2	A_1A_2	A_1A_2	$+\beta_c$		$+2\beta_f$	$-\beta_i$	$+\beta_s$	$+\beta_c$		$+2\beta_f$	$-\beta_i$	$+\beta_s$				1
A_1A_2	A_1A_1	A_1A_1	A_1A_1		$+\beta_m$					$+\beta_m$							$\frac{1}{2}$
			A_2A_1		$+\beta_m$			$+\beta_s$	$+\beta_c$	$+\beta_m$		$+\beta_i$		$-\beta_c$	$-\beta_i$	$+\beta_s$	$\frac{1}{2}$
		A_2A_1	A_1A_1	$+\beta_c$	$+\beta_m$		$+\beta_i$			$+\beta_m$			$+\beta_s$	$+\beta_c$	$+\beta_i$	$-\beta_s$	$\frac{1}{2}$
			A_2A_1	$+\beta_c$	$+\beta_m$		$+\beta_i$	$+\beta_s$	$+\beta_c$	$+\beta_m$		$+\beta_i$	$+\beta_s$				$\frac{1}{2}$
	A_1A_2	A_1A_1	A_1A_1		$+\beta_m$	$+\beta_f$				$+\beta_m$	$+\beta_f$						$\frac{1}{4}$
			A_1A_2		$+\beta_m$	$+\beta_f$		$+\beta_s$	$+\beta_c$	$+\beta_m$	$+\beta_f$	$-\beta_i$		$-\beta_c$	$+\beta_i$	$+\beta_s$	$\frac{1}{4}$
			A_2A_1		$+\beta_m$	$+\beta_f$		$+\beta_s$	$+\beta_c$	$+\beta_m$	$+\beta_f$	$+\beta_i$		$-\beta_c$	$-\beta_i$	$+\beta_s$	$\frac{1}{4}$
			A_2A_2		$+\beta_m$	$+\beta_f$		$+2\beta_s$	$+2\beta_c$	$+\beta_m$	$+\beta_f$			$-2\beta_c$		$+2\beta_s$	$\frac{1}{4}$
		A_1A_2	A_1A_1	$+\beta_c$	$+\beta_m$	$+\beta_f$	$-\beta_i$			$+\beta_m$	$+\beta_f$		$+\beta_s$	$+\beta_c$	$-\beta_i$	$-\beta_s$	$\frac{1}{4}$
			A_1A_2	$+\beta_c$	$+\beta_m$	$+\beta_f$	$-\beta_i$	$+\beta_s$	$+\beta_c$	$+\beta_m$	$+\beta_f$	$-\beta_i$	$+\beta_s$				$\frac{1}{4}$
			A_2A_1	$+\beta_c$	$+\beta_m$	$+\beta_f$	$-\beta_i$	$+\beta_s$	$+\beta_c$	$+\beta_m$	$+\beta_f$	$+\beta_i$	$+\beta_s$		$-2\beta_i$		$\frac{1}{4}$
			A_2A_2	$+\beta_c$	$+\beta_m$	$+\beta_f$	$-\beta_i$	$+2\beta_s$	$+2\beta_c$	$+\beta_m$	$+\beta_f$		$+\beta_s$	$-\beta_c$	$-\beta_i$	$+\beta_s$	$\frac{1}{4}$
		A_2A_1	A_1A_1	$+\beta_c$	$+\beta_m$	$+\beta_f$	$+\beta_i$			$+\beta_m$	$+\beta_f$		$+\beta_s$	$+\beta_c$	$+\beta_i$	$-\beta_s$	$\frac{1}{4}$
			A_1A_2	$+\beta_c$	$+\beta_m$	$+\beta_f$	$+\beta_i$	$+\beta_s$	$+\beta_c$	$+\beta_m$	$+\beta_f$	$-\beta_i$	$+\beta_s$		$+2\beta_i$		$\frac{1}{4}$
			A_2A_1	$+\beta_c$	$+\beta_m$	$+\beta_f$	$+\beta_i$	$+\beta_s$	$+\beta_c$	$+\beta_m$	$+\beta_f$	$+\beta_i$	$+\beta_s$				$\frac{1}{4}$
			A_2A_2	$+\beta_c$	$+\beta_m$	$+\beta_f$	$+\beta_i$	$+2\beta_s$	$+2\beta_c$	$+\beta_m$	$+\beta_f$		$+\beta_s$	$-\beta_c$	$+\beta_i$	$+\beta_s$	$\frac{1}{4}$
		A_2A_2	A_1A_1	$+2\beta_c$	$+\beta_m$	$+\beta_f$				$+\beta_m$	$+\beta_f$		$+2\beta_s$	$+2\beta_c$		$-2\beta_s$	$\frac{1}{4}$
			A_1A_2	$+2\beta_c$	$+\beta_m$	$+\beta_f$		$+\beta_s$	$+\beta_c$	$+\beta_m$	$+\beta_f$	$-\beta_i$	$+2\beta_s$	$+\beta_c$	$+\beta_i$	$-\beta_s$	$\frac{1}{4}$
			A_2A_1	$+2\beta_c$	$+\beta_m$	$+\beta_f$		$+\beta_s$	$+\beta_c$	$+\beta_m$	$+\beta_f$	$+\beta_i$	$+2\beta_s$	$+\beta_c$	$-\beta_i$	$-\beta_s$	$\frac{1}{4}$
			A_2A_2	$+2\beta_c$	$+\beta_m$	$+\beta_f$		$+2\beta_s$	$+2\beta_c$	$+\beta_m$	$+\beta_f$		$+2\beta_s$				$\frac{1}{4}$
A_2A_2	A_1A_1	A_2A_1	A_2A_1	$+\beta_c$	$+2\beta_m$		$+\beta_i$	$+\beta_s$	$+\beta_c$	$+2\beta_m$		$+\beta_i$	$+\beta_s$				1
	A_1A_2	A_2A_1	A_2A_1	$+\beta_c$	$+2\beta_m$	$+\beta_f$	$+\beta_i$	$+\beta_s$	$+\beta_c$	$+2\beta_m$	$+\beta_f$	$+\beta_i$	$+\beta_s$				$\frac{1}{2}$
			A_2A_2	$+\beta_c$	$+2\beta_m$	$+\beta_f$	$+\beta_i$	$+2\beta_s$	$+2\beta_c$	$+2\beta_m$	$+\beta_f$		$+\beta_s$	$-\beta_c$	$+\beta_i$	$+\beta_s$	$\frac{1}{2}$
		A_2A_2	A_2A_1	$+2\beta_c$	$+2\beta_m$	$+\beta_f$		$+\beta_s$	$+\beta_c$	$+2\beta_m$	$+\beta_f$	$+\beta_i$	$+2\beta_s$	$+\beta_c$	$-\beta_i$	$-\beta_s$	$\frac{1}{2}$
			A_2A_2	$+2\beta_c$	$+2\beta_m$	$+\beta_f$		$+2\beta_s$	$+2\beta_c$	$+2\beta_m$	$+\beta_f$		$+2\beta_s$				$\frac{1}{2}$
	A_2A_2	A_2A_2	A_2A_2	$+2\beta_c$	$+2\beta_m$	$+2\beta_f$		$+2\beta_s$	$+2\beta_c$	$+2\beta_m$	$+2\beta_f$		$+2\beta_s$				1

In contrast, we find that the single-locus genetic variance under assortative mating for the sibling difference is reduced by the parental correlation, ρ:

$$V_{s_{AM}} = 2q_1q_2(1-\rho)(\beta_c^2 + \beta_i^2 + \beta_s^2 - 2\beta_c\beta_s). \tag{5}$$

More details can be found in the Supporting Information.

Multi-locus model

We now extend our model from a single locus to multiple loci (assuming a polygenic additive genetic model) by summing over all l causal variants and taking into account correlations, k_{ij}, between loci. The total genetic variance under random mating is

$$V_{tot_{RM}} = \sum_{j=1}^{l} V_{RM_j} + \sum_{i \neq j} k_{ij}\sqrt{V_{RM_i}V_{RM_j}}. \tag{6}$$

The covariances between loci, given in the second term of Equation 6, are dependent upon the correlations among the causal variants and the concordance in their effects size.

Extending the single-locus model for differences in siblings to multiple loci can be done through Equation 6, substituting V_{RM} with Equation 3. We can also easily adapt Equation 6 to include AM by substituting V_{RM} with V_{AM}, or in case of the sibling difference V_{sAM}, where AM-induced changes within a single locus are already taken into account. AM-induced correlations between loci (both cis and trans), on the other hand, will alter the value of the correlation term, k_{ij}.

The change in variance due to AM induced correlations between alleles of mates at a locus and also between loci has been shown by previous theory[19] and empirical research[20]. An equilibrium is reached under AM, where there is inflated genetic variance, a larger correlation between relatives for traits driving AM and increased homozygosity at causal loci. To put our work in the context of previous theory[19], we use a similar approach to Ref.[19] to understand the effect of AM on the variance taking into account multiple loci and parental influences. We follow the approach outlined in Chapter 4.7 of Ref.[19], where all alleles of l causal loci are assumed to have the same frequencies, q_1 and $q_2 = 1 - q_1$, and the same effect α on the trait under study. In our work, we consider the parental

genetic influence as well as direct effects, and thus α needs to be replaced, depending upon whether the allele is inherited from the mother or the father. The variance of a single allele from the mother is determined to be $q_1 q_2 (\beta_c + \beta_m)^2$, assuming that the effect of the allele is either $(\beta_c + \beta_m)$ or 0, with frequency q_1 and q_2. We temporarily assume that parent-of-origin effects are absent, as the procedure of Ref.[19] is based on the variance of single alleles, but we will introduce these effects again later. The mean of the maternal allele is then given by $q_1(\beta_c + \beta_m) + q_2 \cdot 0 = q_1(\beta_c + \beta_m)$. The variance is calculated as

$$\begin{aligned} V(A_m) &= q_1[(\beta_c + \beta_m) - q_1(\beta_c + \beta_m)]^2 + q_2[0 - q_1(\beta_c + \beta_m)]^2 \\ &= q_1 q_2 (\beta_c + \beta_m)^2. \end{aligned} \tag{7}$$

Similarly, the variance for one paternal allele is

$$V(A_f) = q_1 q_2 (\beta_c + \beta_f)^2 \tag{8}$$

Without any correlations between loci or alleles, the sum of the total variance of an individual, $V(X)$, is just the sum of the variances over all loci, where one allele at each locus is inherited from the mother and one from the father

$$\begin{aligned} V(X) =& \sum_{i=1}^{l} (V(A_{m_i}) + V(A_{f_i})) \\ =& l q_1 q_2 [(\beta_c + \beta_m)^2 + (\beta_c + \beta_f)^2] \\ =& 2 l q_1 q_2 (\beta_c^2 + \frac{1}{2}\beta_m^2 + \frac{1}{2}\beta_f^2 + \beta_c \beta_m + \beta_c \beta_f), \end{aligned} \tag{9}$$

which is the expectation for the additive genetic variance in a randomly mating population in the presence of direct and indirect parental effects.

Through AM, correlations between alleles of mates and between loci are introduced, as shown in Chapter 4.7 of Ref.[19]. Relevant correlations present in the child are correlations between the alleles at the same locus, denoted f; at different loci, denoted s; as well as at different loci, but within the same gamete, denoted k. These are also displayed in Figure 1. Due to the correlations, the total variance then changes to

$$V(X) = \sum_{i=1}^{l} (V(A_{m_i}) + V(A_{f_i})) + \sum_{i \neq j} cov_{ij}, \tag{10}$$

where the covariance term cov_{ij} consists of the following three parts, adapted from Ref.[19]:

1. The covariance of a pair of A_m and A_f alleles at the same locus for all l pairs is given by $lfq_1q_2(\beta_c + \beta_m)(\beta_c + \beta_f)$, where f represents the correlation between A_m and A_f.

2. The covariance between alleles at different loci, but within the same gamete, is $l(l-1)/2 \cdot kq_1q_2(\beta_c + \beta_{m/f})$, where the correlation is denoted with k. The covariance differs for maternal and paternal alleles in the parental effect $\beta_{m/f}$.

3. There are $l(l-1)$ pairs of A_m and A_f alleles at different loci with correlation s and covariance $l(l-1)sq_1q_2(\beta_c + \beta_m)(\beta_c + \beta_f)$.

These three covariance terms all act to inflate the genetic variance, as is expected from AM. However, AM also induces a change in the mating frequencies, which the first covariance term does not explicitly take into account. Rather, it is assumed that the correlation f acts in the same way on all genotypes, which is akin to assuming that the mating frequencies are not changed. Thus, while previous theory has established that an equilibrium is reached under AM, where there is inflated genetic variance, a larger correlation between relatives for traits driving AM and increased homozygosity at causal loci[19], the increased homozygosity is not directly accounted for when calculating the expected increase in genetic variance.

Our single-locus model under AM presented above incorporates changes in the mating frequency under AM, which accounts for the increase in variability at one locus that is induced by f (which we denote as ρ above). To extend our model to multiple loci, we therefore only need to consider the covariances across loci. Instead of studying the correlations between a pair of alleles, we look at the correlation k between a pair of genotypes at different loci, i.e. $(A_mA_f)_i$ and $(A_mA_f)_j$. This also enables us to include imprinting effects.

For $k = corr((A_mA_f)_i, (A_mA_f)_j)$, the covariance for $l(l-1)$ possible pairs is given by $l(l-1)k\sqrt{V_{AMi}V_{AMj}} = l(l-1)kV_{AM}$, using the variance under AM defined in either Equation 4, or Equation 5 for differences in siblings. Correlations induced by AM between loci will thus inflate the variance by the factor $l(l-1)k$. Extending this to the case where the correlations between the loci are not equal and effects and allele frequencies are not

the same for each locus, the covariance changes to $\sum_{i \neq j} k_{ij} \sqrt{V_{Ai} V_{Aj}}$. Depending on the sign of the covariance term, it will either increase or reduce the genetic variance, and is thus important to take into account.

In human studies, another common family design is using transmitted and untransmitted parental alleles[3]. We show how our work can be placed in the context of transmitted and untransmitted alleles in the Supplementary Information.

Multi-locus simulations

We conducted a simple forward-in-time simulation study using realistic LD to confirm our theoretical calculations for multiple loci. We create genotypes and phenotypes of 32,000 unrelated individuals based on about 50,000 single nucleotide polymorphism (SNPs) randomly selected from the 1000 genomes project. These individuals are the parents for the next generation, creating two offspring in each of the 10 generations. We create mate assortment by ordering the phenotypes according to a specified phenotypic correlation, ρ_Y, which is a simplified and not realistic process. However, our aim was to simply confirm our theory and to demonstrate the directions of the (co)variances and the degree of confounding under AM. More details on the simulation are given in the Supporting Information.

The left top panel of Figure 2 shows that the expected theoretical variance matches the estimated variance spread over 1000 causal loci for a wide range of variance scenarios: (i) only direct effects (V1), (ii) direct and imprinting effects (V2), (iii) direct, maternal and paternal effects (V3), (iv) direct, maternal, paternal and imprinting effects (V4) and (v) direct, maternal, paternal and imprinting effects including correlations between them (V5) under random mating. As already demonstrated numerically in Equation 2, the population level variance increases when indirect genetic and imprinting effects are present. If there are no parental influences apart from parent-of-origin effects and no sibling effects, then the variance simply increases by β_i^2, with no covariance between imprinting and direct effects. If parental genetic effects are also present, covariances then emerge and contribute substantially to the population-level variance. These covariances will be dependent upon the direction and magnitude of the maternal, paternal and imprinting

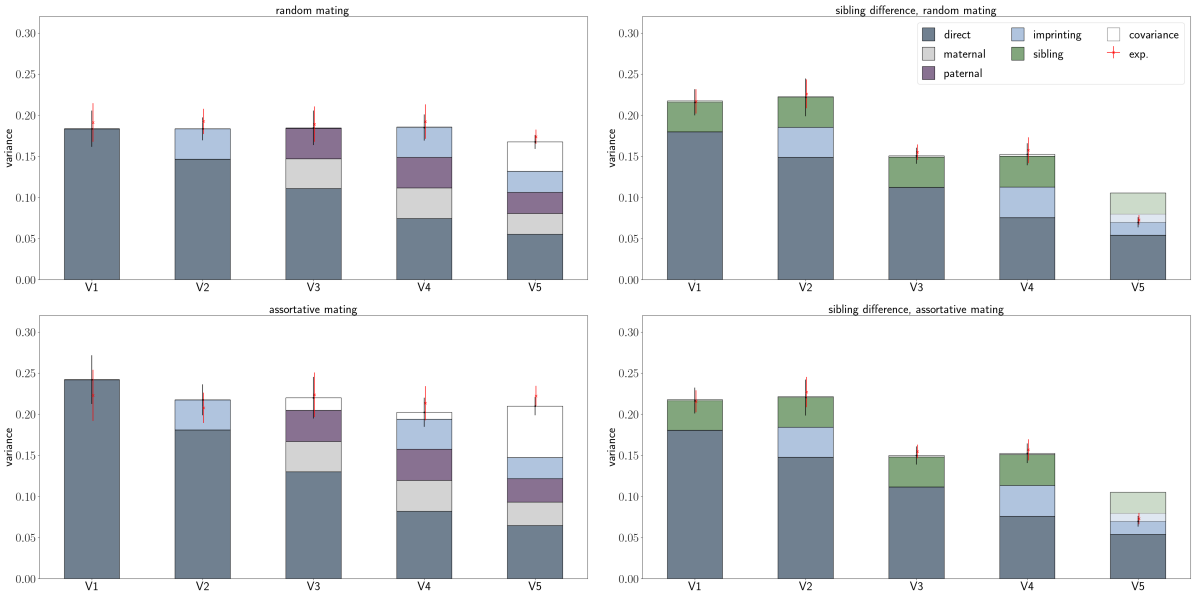

Figure 2: Results of simulation study. Expected theoretical total variance (exp.), and estimated variance (bars) for children (left) and sibling differences differences (right) under random mating (top) and assortative mating ($\rho_Y = 0.25$, bottom) are shown for various variance-covariance scenarios (V1-V5) for multiple loci. Bars and points represent the mean across 10 simulations, while the uncertainties indicate 2x standard deviations. Note that the direct-sibling covariance is negative and is thus shown opaque, covering parts of the variance.

effects of a given locus. The results given in Figure 2 clearly show that components of variance cannot be simply separated into independent direct and indirect sources that then simply sum. This highlights again the confounded nature of direct, indirect genetic and parent-of-origin effects.

Also under assortative mating for a correlation between the parental phenotypes of $\rho_Y = 0.25$, the estimated and expected theoretical variances match for the different variance scenarios, as shown in the left bottom panel of Figure 2. If we consider loci with only direct effects (V1), the variance is enlarged only due to the increased homozygosity of causal loci that is expected under AM.

As soon as indirect parental effects exist, the correlation between parental phenotypes generates a genotypic correlation between the parents which in turn creates a covariance term between the maternal and paternal effect genotypic values, $X_m\beta_m X_f\beta_f$, as can be

clearly seen in V3 comparing the top and bottom left panels of Figure 2. This term is an important component of variance and can be interpreted as a measure of the similarity of the indirect genetic effects among parents on their children. This suggests that assortative mating on a trait may also create assortment for parental characteristics, where biological parents create similar environments for their children, creating covariance in the indirect genetic effects that may have a large role to play in shaping the population-level variation. Until now this covariance has been generally overlooked. We propose that shared parental effects across traits can also create the cross-trait assortative mating patterns that are observed in the human population[21].

The right panels of Figure 2 shows the expected theoretical and estimated variance when modelling the difference between siblings and including indirect sibling effects under random (top) and assortative mating (bottom). As demonstrated numerically in Equation 3, parental effects do not contribute to the variance as they cancel, although we note that with parental genetic effects the correlation within and among loci under AM alters, which changes the variance as compared to the case when parental genetic effects are absent. However, sibling effects and their covariance with the direct effects do contribute and are indistinguishable from the direct effects here, as we are calculating the difference between siblings. Interestingly, the decrease in variance under AM in case of the sibling difference is more substantial than the increase of variance in trios, as is already expected when comparing Equations 4 and 5.

Discussion

Here, we provide a quantitative genetics model that includes the influence of indirect maternal and paternal, and parent-of-origin genetic effects on the additive genetic variance. We explicitly derive the influence of correlation of direct genetic, indirect genetic and epigenetic effects on how AM shapes the population-level genetic variance. Previous theory has established that an equilibrium is reached under AM, where there is inflated genetic variance, a larger correlation between relatives for traits driving AM and increased homozygosity at causal loci[19,22]. But imprinting and indirect genetic effects, alongside

their covariances, and the change in mating frequency, have not been considered, when calculating the expected change in genetic variance.

When modelling sibling differences instead of child-mother-father trios, we additionally considered indirect genetic effects from siblings. This design removes indirect parental effects from the variance, but alters the variance of direct genetic effects in the presence of indirect sibling genetic effects. This design seems to be able to separate parent-of-origin effects under random mating. Under assortative mating, the variances in the trio study design increase, but the variances in the sibling design decrease. Therefore, we suggest that contrasting the estimates across within-family studies may give the upper and lower bounds of the degree to which the components influence the phenotypic variance within the human population.

Our results demonstrate that it is important to control for indirect parental effects when estimating the genetic variance. But even single-locus, marginal estimates common to genome-wide association studies (GWAS) from within-family studies are difficult to interpret causally[12,13] due to the highly correlated nature of markers and effects. Controlling for the covariances among loci across the genome (both cis- and trans-correlations) when estimating genetic effects would require fitting all variants and all forms of genetic effect (direct, maternal, paternal, parent-of-origin) jointly. We show how covariances depend upon the direction and magnitude of the direct, maternal, paternal and imprinting effects within and across loci. These covariances contribute substantially to the population-level phenotypic variance, if parental genetic effects are present. In particular, covariance of the parental indirect genetic effects, a measure of the similarity of the indirect genetic effects among parents on their children, may be an important component of variance. We propose that assortative mating on a trait may also create assortment for parental characteristics, where biological parents create similar environments for their children, creating shared parental effects across traits and cross-trait assortative mating[21].

The separation of imprinting and indirect genetic effects becomes more complex when one considers the underlying potential causal mechanism. Genomic imprinting has been considered as a form of parental effect in the sense that there is parental influence on the genotype-phenotype relationship in the offspring; while others suggest that because the

offspring genotype accounts for all of the genetic variance in the offspring trait, imprinting is a form of direct effect as there is no variation explained by the parental genotype. Differentiating the genes in the parental genome that cause imprinting from parental genetic effects and from the genes in the offspring genome that become imprinted is difficult - if not impossible - especially if there are parental loci that control the imprinting state of other loci in the offspring. This complex phenomenon creates variation in offspring traits that is dependent on the combination of parental and offspring genotypes and it is important to recognize that: (i) patterns interpreted as indirect genetic effects can result from genomic imprinting (and vice versa), and (ii) it is not straightforward to distinguish parental effects on imprinting from imprinted genes showing parental expression; and (iii) under assortative mating, which is common for many human phenotypes, obtaining accurate estimates will be an even more difficult task.

We have used the terms parent-of-origin and the epigenetic phenomenon of imprinting interchangeably, assuming that parent-of-origin effects are underlain by epigenetic mechanisms, where a parent influences the genotype-phenotype relationship of their children. However, we appreciate that other factors may create parent-of-origin effects in the population[23]. Additionally, the ρ term used throughout this work reflects a correlation among mates at causal SNP variants (or those correlated with an underlying causal variant). A non-zero ρ term can be induced through actively choosing mates similar (or dissimilar) to oneself at specific phenotypic characters, termed direct assortative (disassortative) mating, which creates a phenotypic correlation among parents that translates to a correlation in underlying causal variants. Correlations can also appear through cultural, environmental, or geographical stratification in mating patterns that correlate with trait and trait-associated allele frequency stratification in the population. We do not discriminate between these here as they are likely inseparable in population-level data. Rather, we focus on demonstrating how correlations between mates at causal loci influence the phenotypic variance attributable to the DNA.

Here, we ignore dominance, genotype-environment interaction, genotype-environment correlation, epistasis, etc. Future work can incorporate these into our model. Moreover, the framework used here could be combined with mating frequencies expected under

different forms of mate choice and thus be used to explore how indirect and epigenetic effects shape phenotypic variation in other systems.

In summary, population-level variance in human phenotypes is shaped by potentially complex relationships between direct, indirect genetic and imprinting effects that render their separation and quantification almost impossible, especially when assortative mating occurs within the population. Statistical modelling developments are required if we are to understand the genetic basis of human complex traits in the presence of indirect parental and imprinting effects.

Acknowledgements

We thank members of the Medical Genomics group at ISTA for their comments, which improved this manuscript. This work was funded by an SNSF Eccellenza Grant to MRR (PCEGP3-181181), and by core funding from the Institute of Science and Technology Austria.

Author contributions

IK and MRR conceived and designed the study, derived the theory, conducted the analyses and wrote the paper.

Author competing interests

MRR receives research funding from Boehringer Ingelheim for work unrelated to that presented here. IK declares no competing interests.

Code availability

The code used to simulate data can be found at https://github.com/medical-genomics-group/familyMC

References

[1] Gluckman, P. D., Hanson, M. A., Cooper, C. & Thornburg, K. L. Effect of in utero and early-life conditions on adult health and disease. *New England Journal of Medicine* **359**, 61–73 (2008). PMID: 18596274.

[2] Lynch, M. & Walsh, B. *Genetics and analysis of quantitative traits.* (Sinauer, Sunderland, Mass, 1998).

[3] Kong, A. *et al.* The nature of nurture: Effects of parental genotypes. *Science* **359**, 424–428 (2018).

[4] Shen, H. & Feldman, M. W. Genetic nurturing, missing heritability, and causal analysis in genetic statistics. *Proceedings of the National Academy of Sciences* **117**, 25646–25654 (2020).

[5] Eilertsen, E. M. *et al.* Direct and indirect effects of maternal, paternal, and offspring genotypes: Trio-gcta. *Behavior Genetics* **51**, 154–161 (2021).

[6] Wolf, J. B. Gene interactions from maternal effects. *Evolution* **54**, 1882–1898 (2000).

[7] Hager, R., Cheverud, J. M. & Wolf, J. B. Maternal Effects as the Cause of Parent-of-Origin Effects That Mimic Genomic Imprinting. *Genetics* **178**, 1755–1762 (2008).

[8] Wolf, J. B. & Wade, M. J. Evolutionary genetics of maternal effects. *Evolution* **70**, 827–839 (2016).

[9] Spencer, H. G. The correlation between relatives on the supposition of genomic imprinting. *Genetics* **161**, 411–417 (2002).

[10] Santure, A. W. & Spencer, H. G. Influence of Mom and Dad: Quantitative Genetic Models for Maternal Effects and Genomic Imprinting. *Genetics* **173**, 2297–2316 (2006).

[11] Wolf, J. B. & Hager, R. A maternal–offspring coadaptation theory for the evolution of genomic imprinting. *PLOS Biology* **4**, 1–6 (2006).

[12] Veller, C., Przeworski, M. & Coop, G. Causal interpretations of family gwas in the presence of heterogeneous effects. *Proceedings of the National Academy of Sciences* **121** (2024).

[13] Veller, C. & Coop, G. M. Interpreting population- and family-based genome-wide association studies in the presence of confounding. *PLOS Biology* **22**, 1–35 (2024).

[14] Young, A. S. Estimation of indirect genetic effects and heritability under assortative mating (2023). URL `http://dx.doi.org/10.1101/2023.07.10.548458`.

[15] Stark, A. E. A model of assortative mating. *Annals of Human Genetics* **39**, 455–464 (1976).

[16] Morris, T. T., Davies, N. M., Hemani, G. & Smith, G. D. Population phenomena inflate genetic associations of complex social traits. *Science Advances* **6** (2020).

[17] Brumpton, B. *et al.* Avoiding dynastic, assortative mating, and population stratification biases in mendelian randomization through within-family analyses. *Nature Communications* **11** (2020).

[18] Young, A. I. *et al.* Mendelian imputation of parental genotypes improves estimates of direct genetic effects. *Nature Genetics* **54**, 897–905 (2022).

[19] Crow, J. & Kimura, M. *An Introduction to Population Genetics Theory* (Blackburn Press, 1970).

[20] Yengo, L. *et al.* Imprint of assortative mating on the human genome. *Nature Human Behaviour* **2**, 948–954 (2018).

[21] Border, R. *et al.* Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science* **378**, 754–761 (2022).

[22] Fisher, R. A. Statistical methods in genetics. *Heredity* **6**, 1–12 (1952).

[23] Lawson, H. A., Cheverud, J. M. & Wolf, J. B. Genomic imprinting and parent-of-origin effects on complex traits. *Nature Reviews Genetics* **14**, 609–617 (2013).

[24] Dimitromanolakis, A., Xu, J., Krol, A. & Briollais, L. sim1000g: a user-friendly genetic variant simulator in r for unrelated individuals and family-based designs. *BMC Bioinformatics* **20** (2019).

Supplementary Material

A quantitative genetic model for indirect genetic effects and genomic imprinting under random and assortative mating

Ilse Krätschmer, Matthew R. Robinson

Single-locus model under random mating

We first consider a single autosomal locus with two alleles (A_1 and A_2) with frequencies q_1 and q_2 that can have the following four effects on the child's phenotype: direct (β_c), indirect maternal (β_m), indirect paternal (β_f) and parent-of-origin (imprinting, β_i). The alleles at the locus are ordered, referring first to the maternally inherited allele and second to the paternally inherited one, i.e. $A_m A_f$. The genotypes, X_c, X_m and X_f, are coded as 0, 1 and 2 for A_1A_1 homozygotes, A_1A_2 or A_2A_1 heterozygotes and A_2A_2 homozygotes, respectively. For heterozygous children, the parent-of-origin effect is positive for maternally inherited A_2 and negative for paternally inherited A_2. The possible genotypic and phenotypic values of the child, Y_c, given the genotype of the parents and imprinting are shown in Table 1, where the phenotypic values represent the deviations from the case where both parents and child have A_1A_1 genotypes.

We begin by assuming random mating and Hardy-Weinberg equilibrium, where $(q_1 + q_2) = 1 = (q_1 + q_2)^2$. The mean phenotypic value over the population is calculated as

$$\mu = \sum_j Y_{cj} z_j, \tag{11}$$

where j loops over all possible genotypes of the child, Y_c is the expected phenotype produced by a given genotype and z represents the corresponding mating frequency. The means for the different genotype classes of the child and the total mean are determined to be

$$\mu(A_1A_1) = q_1^2 q_2(\beta_m + \beta_f)\,, \tag{12}$$

$$\mu(A_1A_2) = q_1 q_2(\beta_c - \beta_i + q_2\beta_m + (1 + q_2)\beta_f)\,, \tag{13}$$

$$\mu(A_2A_1) = q_1 q_2(\beta_c + \beta_i + (1 + q_2)\beta_m + q_2\beta_f)\,, \tag{14}$$

$$\mu(A_2A_2) = q_2^2[2\beta_c + (1 + q_2)(\beta_m + \beta_f)]\,, \tag{15}$$

$$\mu = 2q_2(\beta_c + \beta_m + \beta_f)\,. \tag{16}$$

The variance is calculated as

$$V_{RM} = \sum_j Y_{cj}^2 z_j - \mu^2, \tag{17}$$

where

$$\sum_j Y_{cj}^2 z_j = 2q_2(q_2+1)(\beta_c^2+\beta_m^2+\beta_f^2)+2q_1q_2(\beta_i^2+\beta_i\beta_m-\beta_i\beta_f)+2q_2(3q_2+1)(\beta_c\beta_m+\beta_c\beta_f)+8q_2^2\beta_m\beta_f, \tag{18}$$

and results in

$$V_{RM} = 2q_1q_2(\beta_c^2 + \beta_m^2 + \beta_f^2 + \beta_i^2 + \beta_c\beta_m + \beta_c\beta_f + \beta_m\beta_i - \beta_f\beta_i). \tag{19}$$

Including indirect sibling effects

We first extend our single autosomal locus model to the case of siblings where there are the following five effects on the child's phenotype: direct (β_c), indirect maternal (β_m), indirect paternal (β_f), parent-of-origin (imprinting, β_i) and sibling (β_s). The possible genotypic and phenotypic values of the child, Y_c, given the genotype of the parents, sibling and imprinting are shown in Table 2, where the phenotypic values represent the deviations from the case where all parents and children have A_1A_1 genotypes.

We begin by assuming random mating and Hardy-Weinberg equilibrium, where the means for the different genotype classes of the child and the total mean are determined to be

$$\mu(A_1A_1) = q_1^2q_2(\beta_m + \beta_f + \beta_s)\,, \tag{20}$$

$$\mu(A_1A_2) = q_1q_2[\beta_c - \beta_i + q_2\beta_m + (1+q_2)\beta_f + \frac{1}{2}(1+2q_2)\beta_s]\,, \tag{21}$$

$$\mu(A_2A_1) = q_1q_2[\beta_c + \beta_i + (1+q_2)\beta_m + q_2\beta_f + \frac{1}{2}(1+2q_2)\beta_s]\,, \tag{22}$$

$$\mu(A_2A_2) = q_2^2[2\beta_c + (1+q_2)(\beta_m + \beta_f + \beta_s)]\,, \tag{23}$$

$$\mu = 2q_2(\beta_c + \beta_m + \beta_f + \beta_s)\,. \tag{24}$$

The variance is calculated as

$$V_{RM} = \sum_j Y_{cj}^2 z_j - \mu^2, \tag{25}$$

where

$$\begin{aligned}\sum_j Y_{cj}^2 z_j =& 2q_2(q_2+1)(\beta_c^2 + \beta_m^2 + \beta_f^2 + \beta_s^2) + 2q_1q_2(\beta_i^2 + \beta_m\beta_i - \beta_f\beta_i) \\ &+ 8q_2^2\beta_m\beta_f + 2q_2(3q_2+1)(\beta_c\beta_m + \beta_c\beta_f + \beta_c\beta_s + \beta_m\beta_s + \beta_f\beta_s),\end{aligned} \tag{26}$$

and results in

$$V_{RM} = 2q_1q_2(\beta_c^2+\beta_i^2+\beta_m^2+\beta_f^2+\beta_s^2+\beta_c\beta_m+\beta_c\beta_f+\beta_m\beta_i-\beta_f\beta_i+\beta_c\beta_s+\beta_m\beta_s+\beta_f\beta_s). \quad (27)$$

The mean and population-level variance depend on the mating frequency, z, of the parents. We again introduce a correlation ρ between the maternal and paternal alleles which impacts the mating frequencies, z_{AM}, as can be seen in Table 1. The means for the different genotype classes of the child and the total mean are now determined to be

$$\mu(A_1A_1) = q_1q_2(1-\rho)(\beta_m+\beta_f+\beta_s)[q_1+\rho(\frac{1}{2}-q_1)]\,, \quad (28)$$

$$\begin{aligned}\mu(A_1A_2) =& q_1q_2[\beta_c-\beta_i+q_2(1-\rho)\beta_m+(1+q_2+\rho q_1)\beta_f+\frac{1}{2}(1+2q_2+\rho q_1-\rho q_2)\beta_s] \\ &+\frac{1}{2}\rho q_1q_2[(1+\rho)(-\beta_c+\beta_i)+(q_1-q_2)(1-\rho)\beta_m \\ &\qquad -(3+\rho-2q_1+2\rho q_1))\beta_f-2(q_2+\rho q_1)\beta_s],\end{aligned} \quad (29)$$

$$\begin{aligned}\mu(A_2A_1) =& q_1q_2[\beta_c+\beta_i+(1+q_2+\rho q_1)\beta_m+q_2(1-\rho)\beta_f+\frac{1}{2}(1+2q_2+\rho q_1-\rho q_2)\beta_s] \\ &+\frac{1}{2}\rho q_1q_2[(1+\rho)(-\beta_c-\beta_i)-(3+\rho-2q_1+2\rho q_1))\beta_m \\ &\qquad +(q_1-q_2)(1-\rho)\beta_f-2(q_2+\rho q_1)\beta_s],\end{aligned} \quad (30)$$

$$\begin{aligned}\mu(A_2A_2) =& q_2^2[2\beta_c+(1+q_2+\rho q_1)(\beta_m+\beta_f+\beta_s)] \\ &+\rho q_1q_2[(1+\rho)\beta_c+\frac{1}{2}(3+\rho-2q_1+2\rho q_1)(\beta_m+\beta_f+\beta_s)]\,,\end{aligned} \quad (31)$$

$$\mu = 2q_2(\beta_c+\beta_m+\beta_f+\beta_s)\,. \quad (32)$$

The genotypic mean is the same as the one assuming that there are no correlations between the parental genotypes, which is expected as the allele frequencies do not change. The variance is then

$$\begin{aligned}V_{AM} =& 2q_1q_2(\beta_c^2+\beta_i^2+\beta_s^2+\beta_m^2+\beta_f^2+\beta_c\beta_m+\beta_c\beta_f+\beta_m\beta_i-\beta_f\beta_i+\beta_c\beta_s+\beta_m\beta_s+\beta_f\beta_s) \\ &+q_1q_2\rho(\beta_c^2-\beta_i^2+\beta_s^2+2\beta_m^2+2\beta_f^2+4\beta_c\beta_m+4\beta_c\beta_f+4\beta_c\beta_s+4\beta_m\beta_s+4\beta_f\beta_s+4\beta_m\beta_f) \\ &+q_1q_2\rho^2(\beta_c^2-\beta_i^2+\beta_s^2+2\beta_c\beta_m+2\beta_c\beta_f+2\beta_c\beta_s+2\beta_m\beta_s+2\beta_f\beta_s-2\beta_m\beta_i+2\beta_f\beta_i+4\beta_m\beta_f),\end{aligned} \quad (33)$$

using

$$\sum_j Y_j^2 z_j = 2q_2(q_2+1)(\beta_c^2+\beta_m^2+\beta_f^2+\beta_s^2) + 2q_1q_2(\beta_i^2+\beta_m\beta_i-\beta_f\beta_i)$$
$$+ 8q_2^2\beta_m\beta_f + 2q_2(3q_2+1)(\beta_c\beta_m+\beta_c\beta_f+\beta_c\beta_s+\beta_m\beta_s+\beta_f\beta_s)$$
$$+ q_1q_2\rho(\beta_c^2-\beta_i^2+\beta_s^2+2\beta_m^2+2\beta_f^2+4\beta_c\beta_m+4\beta_c\beta_f+4\beta_c\beta_s+4\beta_m\beta_s+4\beta_f\beta_s+4\beta_m\beta_f)$$
$$+ q_1q_2\rho^2(\beta_c^2-\beta_i^2+\beta_s^2+2\beta_c\beta_m+2\beta_c\beta_f+2\beta_c\beta_s+2\beta_m\beta_s+2\beta_f\beta_s-2\beta_m\beta_i+2\beta_f\beta_i+4\beta_m\beta_f). \tag{34}$$

This shows that indirect effects of siblings are also confounded with the direct effects, the parental genetic effects and the imprinting effects. Assortative mating acts to inflate the covariance of the indirect sibling genetic effects and all other components, making a partitioning of variance components intractable.

Modelling the phenotypic difference of siblings

Previous studies[16] have proposed to control for indirect effects by looking at the phenotypic and genetic difference in siblings. Table 2 shows the possible genotypic and phenotypic values of the child, Y_c, and sibling, Y_s, given the genotype of the parents, the respective sibling and imprinting, as well as the phenotypic difference between the offspring, $Y_c - Y_s$.

The means in phenotypic difference for the various genotype classes of the child and the total mean are determined to be

$$\mu(A_1A_1) = q_1^2q_2(-\beta_c+\beta_s)\,, \tag{35}$$
$$\mu(A_1A_2) = \frac{1}{2}q_1q_2[(1-2q_2)(\beta_c-\beta_s)-\beta_i]\,, \tag{36}$$
$$\mu(A_2A_1) = \frac{1}{2}q_1q_2[(1-2q_2)(\beta_c-\beta_s)+\beta_i]\,, \tag{37}$$
$$\mu(A_2A_2) = q_1q_2^2(\beta_c-\beta_s)\,, \tag{38}$$
$$\mu = 0\,. \tag{39}$$

The total mean is 0 as expected when calculating the difference between siblings. Thus, the variance is

$$V_{sRM} = \sum_j (Y_{cj}-Y_{sj})^2 z_j = 2q_1q_2(\beta_c^2+\beta_i^2+\beta_s^2-2\beta_c\beta_s), \tag{40}$$

which shows that indirect parental effects do indeed cancel in the difference, but indirect sibling and imprinting effects remain.

Single-locus model under assortative mating

The expected phenotype of an offspring of a given genotype depends on the probability that it was produced by each of the different possible parental genotypes. Assortative mating (AM) acts to alter the mating frequencies in the population as it increases parental similarity at trait-associated loci which is reflected in the correlation, ρ, among the parents in their alleles. This changes the genotype frequencies, but does not affect the minor allele frequencies in the population. The genotype frequencies under AM are given in Ref.[15]. The means for the different genotype classes of the child and the total mean are determined to be

$$\mu(A_1A_1) = q_1q_2(1-\rho)(\beta_m+\beta_f)[q_1+\rho(\frac{1}{2}-q_1)]\,, \tag{41}$$

$$\begin{aligned}\mu(A_1A_2) =& q_1q_2[\beta_c-\beta_i+q_2(1-\rho)\beta_m+(1+q_2+\rho q_1)\beta_f]\\ &+\frac{1}{2}\rho q_1q_2[(1+\rho)(-\beta_c+\beta_i)+(q_1-q_2)(1-\rho)\beta_m-(3+\rho-2q_1+2\rho q_1))\beta_f],\end{aligned} \tag{42}$$

$$\begin{aligned}\mu(A_2A_1) =& q_1q_2[\beta_c+\beta_i+(1+q_2+\rho q_1)\beta_m+q_2(1-\rho)\beta_f]\\ &+\frac{1}{2}\rho q_1q_2[(1+\rho)(-\beta_c-\beta_i)-(3+\rho-2q_1+2\rho q_1))\beta_m+(q_1-q_2)(1-\rho)\beta_f],\end{aligned} \tag{43}$$

$$\mu(A_2A_2) = q_2^2[2\beta_c+(1+q_2+\rho q_1)(\beta_m+\beta_f)] \tag{44}$$

$$+\rho q_1q_2[(1+\rho)\beta_c+\frac{1}{2}(3+\rho-2q_1+2\rho q_1)(\beta_m+\beta_f)]\,, \tag{45}$$

$$\mu = 2q_2(\beta_c+\beta_m+\beta_f)\,. \tag{46}$$

The genotypic mean is the same as the one assuming that there are no correlations between the parental genotypes, which is expected as the allele frequencies do not change. The variance is then

$$\begin{aligned}V_{AM} =& 2q_1q_2(\beta_c^2+\beta_i^2+\beta_m^2+\beta_f^2+\beta_c\beta_m+\beta_c\beta_f+\beta_m\beta_i-\beta_f\beta_i)\\ &+q_1q_2\rho(\beta_c^2+2\beta_m^2+2\beta_f^2-\beta_i^2+4\beta_c\beta_m+4\beta_c\beta_f+4\beta_m\beta_f)\\ &+q_1q_2\rho^2(\beta_c^2-\beta_i^2+2\beta_c\beta_m+2\beta_c\beta_f-2\beta_m\beta_i+2\beta_f\beta_i+4\beta_m\beta_f),\end{aligned} \tag{47}$$

using

$$\sum_j Y_j^2 z_j =$$
$$2q_2[(q_2+1)(\beta_c^2+\beta_m^2+\beta_f^2)+q_1(\beta_i^2+\beta_m\beta_i-\beta_f\beta_i)+(3q_2+1)(\beta_c\beta_m+\beta_c\beta_f)+4q_2\beta_m\beta_f]$$
$$+q_1q_2\rho(\beta_c^2-\beta_i^2+2\beta_m^2+2\beta_f^2+4\beta_c\beta_m+4\beta_c\beta_f+4\beta_m\beta_f)$$
$$+q_1q_2\rho^2(\beta_c^2-\beta_i^2+2\beta_c\beta_m+2\beta_c\beta_f-2\beta_m\beta_i+2\beta_f\beta_i+4\beta_m\beta_f). \tag{48}$$

All terms are increased under assortative mating with the exception of imprinting which is reduced. Note that there is now an additional correlation term between the parental genetic effects, $\beta_m\beta_f$, that depends on ρ.

Modelling the phenotypic difference in siblings

Extending the model to also include the change in genotype frequencies due to assortative mating, we find

$$\mu(A_1A_1) = q_1q_2(1-\rho)(-\beta_c+\beta_s)[q_1+\rho(\frac{1}{2}-q_1)]\,, \tag{49}$$
$$\mu(A_1A_2) = \frac{1}{2}q_1q_2(1-\rho)[(1-2q_2+\rho-2\rho q_1)(\beta_c-\beta_s)-\beta_i]\,, \tag{50}$$
$$\mu(A_2A_1) = \frac{1}{2}q_1q_2(1-\rho)[(1-2q_2+\rho-2\rho q_1)(\beta_c-\beta_s)+\beta_i]\,, \tag{51}$$
$$\mu(A_2A_2) = q_1q_2(1-\rho)(\beta_c-\beta_s)[q_2-\rho(\frac{1}{2}-q_1)]\,, \tag{52}$$
$$\mu = 0\,. \tag{53}$$

The variance is reduced by the parental correlation, ρ,

$$V_{sAM} = \sum_j (Y_{cj}-Y_{sj})^2 z_{AMj} = 2q_1q_2(1-\rho)(\beta_c^2+\beta_i^2+\beta_s^2-2\beta_c\beta_s) \tag{54}$$

Transmitted and untransmitted parental alleles

We sought to place our work in the wider context of previous research, which has focussed on transmitted and untransmitted alleles[3]. The phenotype of a child, Y_c, that is influenced by direct, indirect maternal, indirect paternal and imprinting effects across l

loci, can be written in terms of transmitted (A_m, A_f) and untransmitted maternal and paternal alleles ($A_{\backslash m}$, $A_{\backslash f}$):

$$\begin{aligned} Y_c &= \sum_{j=1}^{l}(X_{c_j}\beta_{c_j} + X_{m_j}\beta_{m_j} + X_{f_j}\beta_{f_j} + X_{i_j}\beta_{i_j}) + \epsilon \\ &= \sum_{j=1}^{l}[(A_{m_j} + A_{f_j})\beta_{c_j} + (A_{m_j} + A_{\backslash m_j})\beta_{m_j} + (A_{f_j} + A_{\backslash f_j})\beta_{f_j} + (A_{m_j} - A_{f_j})\beta_{i_j}] + \epsilon \\ &= \sum_{j=1}^{l}[A_{m_j}(\beta_{c_j} + \beta_{m_j} + \beta_{i_j}) + A_{f_j}(\beta_{c_j} + \beta_{f_j} - \beta_{i_j}) + A_{\backslash m_j}\beta_{m_j} + A_{\backslash f_j}\beta_{f_j}] + \epsilon, \end{aligned} \tag{55}$$

where ϵ represents the residual error due to other environmental influences. Note that the imprinting effect will have a positive sign if only A_m is transmitted, a negative one if only A_f is transmitted, and will be equal to 0 for all the other cases. Thus, a transmitted allele has direct, indirect and imprinting effects, while an untransmitted allele only has indirect effects on a child's phenotype.

Assuming Hardy-Weinberg equilibrium and random mating, thus that the four alleles are independent, the total variance that each transmitted and untransmitted allele contributes to the trait of a child per locus j is

$$\begin{aligned} V_j &= q_{1_j}q_{2_j}[(\beta_{c_j} + \beta_{m_j} + \beta_{i_j})^2 + (\beta_{c_j} + \beta_{f_j} - \beta_{i_j})^2 + \beta_{m_j}^2 + \beta_{f_j}^2] \\ &= 2q_{1_j}q_{2_j}(\beta_{c_j}^2 + \beta_{i_j}^2 + \beta_{m_j}^2 + \beta_{f_j}^2 + \beta_{c_j}\beta_{m_j} + \beta_{c_j}\beta_{f_j} + \beta_{m_j}\beta_{i_j} - \beta_{f_j}\beta_{i_j}), \end{aligned} \tag{56}$$

where q_{1_j} and q_{2_j} are the allele frequencies. Equation 56 is equivalent to Equation 2.

When effect sizes are estimated in GWAS, a linear relationship between an individual's phenotype and their genotype, i.e. the two transmitted alleles, is assumed, ignoring imprinting effects. Thus, the SNP heritability, defined as the proportion of phenotypic variance among children attributable to the SNP genotypes of the children, would correspond to

$$h_{SNP}^2 = \sum_{j=1}^{l} 2q_{1_j}q_{2_j}(\beta_{c_j}^2 + 0.5\beta_{m_j}^2 + 0.5\beta_{f_j}^2 + \beta_{c_j}\beta_{m_j} + \beta_{c_j}\beta_{f_j}) \tag{57}$$

assuming that all loci are independent. Our results suggest that in reality, it is not known how much the indirect effects contribute to the GWAS estimate unless all the direct and

indirect terms are taken into account, as we do not know the size and magnitude of the covariances between direct and paternal effects, nor do we know if indirect components are actually totally estimated or estimated as part of the residual variance, nor is it clear how the abundant covariances across loci will influence the effect estimates (both cis- and trans-LD). It is also important to note that the variance due to the untransmitted alleles does not equal the total variance of the indirect effects, but only half (at most). Thus, the effects attributed to the untransmitted alleles are in fact not a good estimate of indirect parental effects, as has already been noted by Ref.[13].

Previous work[3] used polygenic risk scores computed from transmitted alleles, $\hat{\boldsymbol{g}}_T = \mathbf{A}_m\hat{\boldsymbol{\beta}}_{GWAS} + \mathbf{A}_f\hat{\boldsymbol{\beta}}_{GWAS}$, and untransmitted ones $\hat{\boldsymbol{g}}_{NT} = \mathbf{A}_{\backslash m}\hat{\boldsymbol{\beta}}_{GWAS} + \mathbf{A}_{\backslash f}\hat{\boldsymbol{\beta}}_{GWAS}$, to separate direct from indirect effects. The effects of transmitted and untransmitted alleles, θ_T and θ_{NT}, are estimated from the linear regression,

$$\mathbf{y} = \theta_T\hat{\boldsymbol{g}}_T + \theta_{NT}\hat{\boldsymbol{g}}_{NT} + \boldsymbol{\epsilon}, \tag{58}$$

where $\mathbf{y}$ is a vector of phenotypic observations and $\mathbf{y}$, $\hat{\boldsymbol{g}}_T$ and $\hat{\boldsymbol{g}}_{NT}$ are all standardized to have mean zero and variance 1. The direct effects of the transmitted polygenic risk score, δ, are then determined by calculating the difference between transmitted and untransmitted effects, $\delta = \theta_T - \theta_{NT}$. The estimated variance accounted for by the direct effect alone is then given by

$$\begin{aligned} R^2_\delta &= (\theta_T - \theta_{NT})^2/\theta^2_T \\ &= \sigma^2_C/(\sigma^2_C + 0.5 * \sigma^2_M + 0.5 * \sigma^2_F), \end{aligned} \tag{59}$$

where σ^2_C is the variance contributed by the direct effects $\sigma^2_C = \sum_{j=1}^{l} 2q_{1_j}q_{2_j}\beta^2_{c_j}$, σ^2_M is the variance contributed by the indirect maternal effects $\sigma^2_M = \sum_{j=1}^{l} 2q_{1_j}q_{2_j}\beta^2_{m_j}$, and σ^2_F is the variance contributed by the direct effects $\sigma^2_F = \sum_{j=1}^{l} 2q_{1_j}q_{2_j}\beta^2_{f_j}$.

Equation 59 makes the following assumptions: (i) $\theta_T = (\sigma^2_C + 0.5 * \sigma^2_M + 0.5 * \sigma^2_F)$ and $\theta_{NT} = (0.5 * \sigma^2_M + 0.5 * \sigma^2_F)$; (ii) loci are independent; (iii) there are no covariances among direct and indirect effects; (iv) imprinting effects are absent; (v) transmitted and untransmitted alleles have the same indirect effects; and (vi) estimation error can be ignored. Note that the estimated variance for direct effects, R^2_δ, in Equation 59 is defined

differently than the direct variance, β_c^2, in Equations 2-5. R_δ^2 is a relative estimate of the direct effects in the transmitted predictor, while β_c^2 is an absolute value and can only be transformed to something similar to R_δ^2 by making an assumption on the heritability of the transmitted alleles.

Forward-in-time simulation

A simple, but realistic simulation (based loosely on Ref.[24]) for parents and children is set up and moved forward in time for a number of generations, $ngen = 10$, to simulate assortative mating. Genotypes of 32,000 unrelated individuals in generation 0 are created with a realistic LD based on 52,310 SNPs. The SNPs were randomly selected across all chromosomes from the 1000 genomes project (`http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/`) using vcfrandomsample (`https://github.com/vcflib/vcflib#vcflib`) to downsample the data to make it more manageable. A pair of individuals produce two offspring each by recombining the parents' haplotypes based on a genetic map (release b37, for example from `https://github.com/odelaneau/shapeit5/tree/main/resources/maps/b37`). The offspring of generation 0 become the parents for the next generation. Parents after generation 0 mate based on their ordered phenotpyes to create assortative mating. The phenotypes for ordering are calculated as:

$$\boldsymbol{Y} = \boldsymbol{X_c}\boldsymbol{\beta_c} + \boldsymbol{X_f}\boldsymbol{\beta_f} + \boldsymbol{X_m}\boldsymbol{\beta_m} + \boldsymbol{X_h}\boldsymbol{\beta_h} + \boldsymbol{\epsilon} \tag{60}$$

where ϵ is a value drawn from a Normal distribution with mean 0 and standard deviation of 0.01 to facilitate ordering. The phenotypes are ordered to introduce a correlation, $\rho = 0.25$, between the parents. This procedure is a simplified version of the unknown realistic process. For the implementation of the AM procedure, see Code availability.

Simulation studies were carried out 10 times each for the following five different variance-covariance matrices of effects,

$$\boldsymbol{V} = \begin{pmatrix} V_c & V_{cf} & V_{cm} & V_{ci} & V_{cs} \\ V_{fc} & V_f & V_{fm} & V_{fi} & V_{fs} \\ V_{mc} & V_{mf} & V_m & V_{mi} & V_{ms} \\ V_{ic} & V_{if} & V_{im} & V_i & V_{is} \\ V_{sc} & V_{sf} & V_{sm} & V_{si} & V_s \end{pmatrix}.$$

When the indirect effects of siblings are not modelled, the last column and the last row are removed.

1. Scenario 1 represents a covariance matrix with only direct effects and, in case of the sibling difference, sibling effects:

$$\boldsymbol{V_1} = \begin{pmatrix} 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.1 \end{pmatrix}. \tag{61}$$

2. In the second scenario, direct as well as parent-of-origin effects and, in case of the sibling difference, sibling effects are contributing to the variance:

$$\boldsymbol{V_2} = \begin{pmatrix} 0.4 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.1 \end{pmatrix}. \tag{62}$$

3. Scenario 3 assumes direct effects as well as smaller indirect maternal and paternal effects and, in case of the sibling difference, sibling effects:

$$\boldsymbol{V_3} = \begin{pmatrix} 0.3 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.1 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.1 \end{pmatrix}. \tag{63}$$

4. In scenario 4, the total variance of 0.5 is spread over all four genetic components. In the case of the sibling difference, there are additionally sibling effects. There are no correlations.

$$\boldsymbol{V_4} = \begin{pmatrix} 0.2 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.1 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.1 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.1 \end{pmatrix}. \tag{64}$$

5. The last scenario represents a scenario where all effects are correlated:

$$\boldsymbol{V_5} = \begin{pmatrix} 0.15 & 0.05 & 0.05 & -0.02 & 0.05 \\ 0.05 & 0.07 & 0.05 & -0.02 & 0.05 \\ 0.05 & 0.05 & 0.07 & -0.02 & 0.05 \\ -0.02 & -0.02 & -0.02 & 0.07 & 0.00 \\ 0.05 & 0.05 & 0.05 & 0.00 & 0.07 \end{pmatrix} \tag{65}$$

Effects were generated for 1000 markers with a multivariate normal distribution with mean 0 and variance $\mathbf{V}$ scaled by the number of causal markers and randomly assigned to 1000 markers for each simulation. The effects were assumed to be constant across generations.