Genetic similarity between relatives provides evidence on the presence and history of assortative mating

Authors:

Hans Fredrik Sunde[1,2*], Nikolai Haahjem Eftedal[3], Rosa Cheesman[3], Elizabeth C. Corfield[4,5], Thomas H. Kleppesto[1,6], Anne Caroline Seierstad[3], Eivind Ystrom[3,5], Espen Moen Eilertsen[3], Fartein Ask Torvik[1,3]

[1] Centre for Fertility and Health, Norwegian Institute of Public Health, Oslo, Norway

[2] Department of Psychology, University of Oslo, Oslo, Norway

[3] PROMENTA Research Centre, Department of Psychology, University of Oslo, Oslo, Norway

[4] Nic Waals Institute, Lovisenberg Diakonale Hospital, Oslo, Norway

[5] Centre for Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway

[6] Department of Psychology, Norwegian University of Science and Technology, Trondheim, Norway

* Corresponding author: hansfredrik.sunde@fhi.no

Word Count: 4879 (excl. methods)

Abstract

Assortative mating – the non-random mating of individuals with similar traits – is known to increase trait-specific genetic variance and genetic similarity between relatives. However, empirical evidence is limited for many traits, and the implications hinge on whether assortative mating has started recently or many generations ago. Here we show theoretically and empirically that genetic similarity between relatives can provide evidence on the presence and history of assortative mating. First, we employed path analysis to understand how assortative mating affects genetic similarity between family members across generations, finding that similarity between distant relatives is more affected than close relatives. Next, we correlated polygenic indices of 47,135 co-parents from the Norwegian Mother, Father, and Child Cohort Study (MoBa) and found genetic evidence of assortative mating in nine out of sixteen examined traits. The same traits showed elevated similarity between relatives, especially distant relatives. Six of the nine traits, including educational attainment, showed greater genetic variance among offspring, which is inconsistent with stable assortative mating over many generations. These results suggest an ongoing increase in familial similarity for these traits. The implications of this research extend to genetic methodology and the understanding of social and economic disparities.

Keywords: Assortative mating, Homogamy, Genetics, Polygenic index, MoBa

Introduction

Assortative mating – the non-random pairing of individuals with similar traits – has long been a challenging topic of interest across various fields, including genetics[1-9], sociology[10-12], and economics[13,14]. Consequences of assortative mating are wide-ranging, affecting topics such as genetic research methods[15,16], relationship quality[10,17,18], and the perpetuation of social and economic inequalities[10,13,14]. Although partner similarity have been documented for numerous characteristics[15,16,19], it remains uncertain to what extent these similarities result from assortative mating or other processes, such as convergence over time[18,20]. Hence, the genetic consequences are unknown. Recent advances in data availability have enabled empirical investigation into the genetic consequences of assortative mating, wherein two are of key interest: First, partners should exhibit genetic similarity for assorted traits; and second, genetic similarity between relatives should increase for the assorted traits in subsequent generations[1-3]. In this paper, we aim to: 1) clarify the theoretical consequences of assortative mating on genetic similarity in extended families; 2) use polygenic indices to assess trait-specific genetic similarity between partners for a range of psychosocial, anthropometric, and health-related traits; 3) investigate whether these traits also exhibit increased genetic similarity among relatives; and 4) use the observed genetic similarity in mother-father-child trios to investigate the stability of assortative mating over many generations.

According to a recent meta-analysis, phenotypic correlations between partners exist for many traits[16]. The correlations are particularly high for cognitive and social traits like educational attainment (0.53) and political values (0.58), but moderate correlations exist for many diverse traits such as height (0.23), depression (0.14), and personality (0.08–0.21). Positive correlations between partners can arise from numerous processes, including convergence (partners becoming more alike over time due to mutual influence), common environments (partners originating from similar environments that affect their traits, but without influencing partner formation), and assortative mating (individuals tending to form partnerships with those having similar traits)[18]. If partner similarity arises because of assortative mating, then this will induce cross-partner correlations between factors that are associated with the trait. If the trait is heritable – which most traits are[21,22] – then partners will tend to carry genetic variants with similar effects on the trait. Genetic similarity between partners has been documented for some traits, including height and educational attainment[6,19,23-25]. For example, Yengo, et al. [25] investigated genetic similarity in partners from the UK Biobank across 32 complex traits, but lack of statistical power left the question unresolved for most traits. Here, we remedy this by investigating partners in the Norwegian Mother, Father, and Child Cohort Study (MoBa)[26,27], the largest cohort of confirmed partners with available genetic data (n=47,135).

If assortative mating leads to genetic similarity between partners, then any resulting offspring are likely to inherit trait-specific genetic variants with similar effects from both parents. This has two important consequences: First, the trait-specific genetic variance in the population will increase because genetic variants with similar effects will tend to co-occur in the same individuals (i.e., variants will be in linkage disequilibrium)[3,8,28,29]. Second, trait-specific genetic similarity between relatives will increase because other family members are more likely to inherit genetic variants with similar effects[1-4,8]. With no assortative mating, genotypic correlations between family members for a trait should equal the coefficient of relationship. For example, full siblings (not including monozygotic twins) and parent-offspring pairs are first-degree relatives, with a coefficient of 0.50; aunt/uncle-niece/nephew and grandparent-grandchild pairs are second-degree relatives, with a coefficient of 0.25; and first

cousins are third-degree relatives, with a coefficient of 0.125. Under assortative mating, however, the trait-specific genotypic correlations will be higher than the corresponding coefficients of relationship. Importantly, assortative mating only induces correlations between trait-associated loci and should not be confused with inbreeding, which induces correlations between all loci[30]. With successive generations of stable assortative mating, trait-specific genetic variance and genotypic correlations between relatives increase asymptotically towards an equilibrium, at which point they become constant across generations[3,8,28,29]. (See also Supplementary Information, section 2).

In this paper, we study the extent of assortative mating on a range of phenotypes and its historical consequences by using genetic data from extended family members. Our first aim is to derive the expected genotypic correlations between family members under various assumptions using path analysis. There are earlier theoretical papers that lays out the consequences of assortative mating on familial resemblance[1-5,31]. However: 1) they often focus on phenotypic rather than genotypic resemblance; 2) they don't consider imperfectly measured genetic factors (i.e., polygenic indices); 3) they often do not consider gene-environment correlations; and 4) they either don't consider disequilibrium or do so only under simplistic assumptions. We use path analysis because it offers a ready way to relax assumptions while making the theory accessible for non-specialists. In doing so, we describe a general formula for finding such correlations between any two extended family members under assortative mating at equilibrium. Our results imply that genetic similarity between distant relatives should be more affected by assortative mating than similarity between close relatives[1,3,32]. Our second aim is to document polygenic index correlations for various traits among partners in MoBa[26,27]. We find genetic evidence of assortative mating for nine out of sixteen investigated traits. Our third aim is to investigate whether genetic similarity between relatives was increased as predicted for these traits. We find that polygenic index correlations among relatives was increased in a way that broadly corresponded to the theoretical expectations. Trait-specific genetic similarity between partners and elevated genetic similarity between relatives indicate that many of the previously observed phenotypic correlations are partly attributable to assortative mating. Our fourth aim is to use mother-father-child trios to test whether the observations were consistent with equilibrium. Although some traits did not significantly deviate from equilibrium expectations, psychosocial traits like education attainment did. This would imply that that the genetic variance and genetic similarity between relatives for these traits are still increasing across generations.

Results

Fig. 1 shows a theoretical model of similarity in extended families in the presence of assortative mating at intergenerational equilibrium. The model includes eight individuals (i) in three generations (t): two partners in the first generation, their two children in the second generation (who are each other's full sibling) along with their respective partners, and two children in the third generation (who are each other's first cousin). The phenotype that is assorted on is denoted with P_{it}, whereas trait-associated additive genetic factors and unique environmental factors are denoted with A_{it} and E_{it}, respectively. The genotypic correlation between any two individuals is the sum of all valid chains of paths between their respective additive genetic factors and the value of a single chain is the product of its path coefficients[33,34]. Valid chains always begin by tracing backward (←) in relation to the direction of arrows, incorporating exactly one double-headed arrow (↔), after which tracing continues in a forward direction (→). Because the variables in Fig. 1 have unit variances, all valid chains connecting a variable to itself will sum to 1, allowing us to immediately trace in a forward direction (i.e., change direction at once). Copaths (—), which are arrowless paths representing associations arising from assortment[35], link together valid chains per the rules above, forming longer, valid chains. For a more detailed description of path tracing rules involving copaths, see Balbona, Kim and Keller [36] or Keller, et al. [37]. Path diagrams with relaxed assumptions (e.g., gene-environment correlations) are presented and discussed in the Supplementary Information, where we also provide simulations validating our theoretical expectations.

Expected genotypic correlations in the nuclear family

In Fig. 1, there is only one valid chain between partners' additive genetic factors (e.g., A_{11}↔A_{21}): $h \times \mu \times h$. The genotypic correlation between partners (denoted ρ_g) is thus the phenotypic correlation attributable to assortative mating, μ, weighted by the trait's heritability, h^2:

$$\rho_g = \mu h^2 \tag{1}$$

Similarly, we can trace the valid chains between the additive genetic factors of a parent and their offspring (e.g., A_{11}↔A_{22}). There are two valid chains: one directly from parental genetic factors to offspring genetic factors, $\frac{1}{2}$, and one through the other parent via the assorted phenotype: $h \times \mu \times h \times \frac{1}{2}$. The genotypic correlation between parent and offspring is therefore $\frac{1}{2} + \frac{h\mu h}{2}$. With no assortative mating ($\mu = 0$), this reduces to $\frac{1}{2}$. For siblings (A_{22}↔A_{32}), there are four valid chains: $\frac{1}{4} + \frac{1}{4} + \frac{h\mu h}{4} + \frac{h\mu h}{4}$, which can be rearranged so that it equals the genotypic parent-offspring correlation. Because they are equal, we can define a common denotation (r_{g_1}) for first-degree relatives. We can also substitute $h \times \mu \times h$ with ρ_g giving us:

$$r_{g_1} = \frac{1 + \rho_g}{2} \tag{2}$$

In other words, the genotypic correlation between first-degree relatives, r_{g_1}, is increased by half the genotypic correlation between partners at equilibrium. (Note that the phenotypic correlation will not be the same for siblings and parent-offspring despite the same genotypic correlation[3]). An advantage of using path analysis is how easy

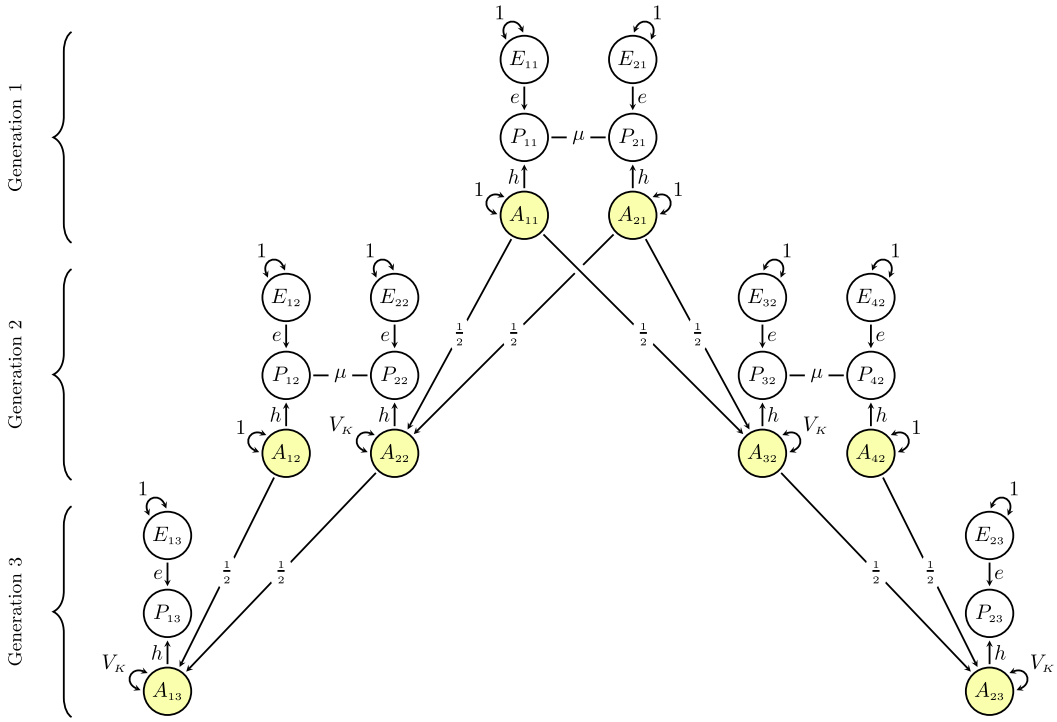

Fig. 1. Path diagram for a model of genetic similarity in extended families under phenotypic assortative mating at intergenerational equilibrium (i.e., equal variance across generations). The partner correlation attributable to assortment is denoted by μ, the recombination variance is denoted by V_K, and h and e denote the effect of additive genetic (A_{it}) and environmental factors (E_{it}), respectively, on the phenotype (P_{it}) of individual i in generation t. All variables have unit variance, meaning $e = \sqrt{1-h^2}$ and $V_K = \frac{1-\mu h^2}{2}$. See the Supplementary Information for path diagrams with relaxed assumptions.

path diagrams are to expand. In the Supplementary Information, we detail how relaxing the assumption of equilibrium (section 2) and including polygenic indices (section 3) changes the correlations. During disequilibrium, the genotypic correlation between partners will still conform to Equation (1), but the correlation between relatives will be less than what Equation (2) would predict. For polygenic index correlations, one must include a term representing the imperfect correlation between the polygenic index and the true genetic factor. The polygenic index correlation between partners should therefore be:

$$\rho_{pgi} = \mu h^2 s^2 \tag{3}$$

where s^2 is the shared variance between the polygenic index and the true additive genetic factor (i.e., the genetic signal[19]). Assortative mating will induce covariance between different loci (i.e., linkage disequilibrium), which is included in the genetic signal. This means that s may be larger than the correlation between the true direct effects and the polygenic index weights, and as such do not represent the accuracy of the polygenic index weights (see Supplementary Information, section 3 and section 4.4). If the genetic signal is low, the polygenic index correlation between partners will be biased towards zero compared to the true genotypic correlation[19]. For first-degree relatives, the equation becomes similarly altered, but because the error terms in the polygenic indices are correlated between relatives, the polygenic index correlation will be biased towards the coefficient of relatedness rather than zero:

$$r_{pgi_1} = \frac{1 + \rho_{pgi}}{2} \tag{4}$$

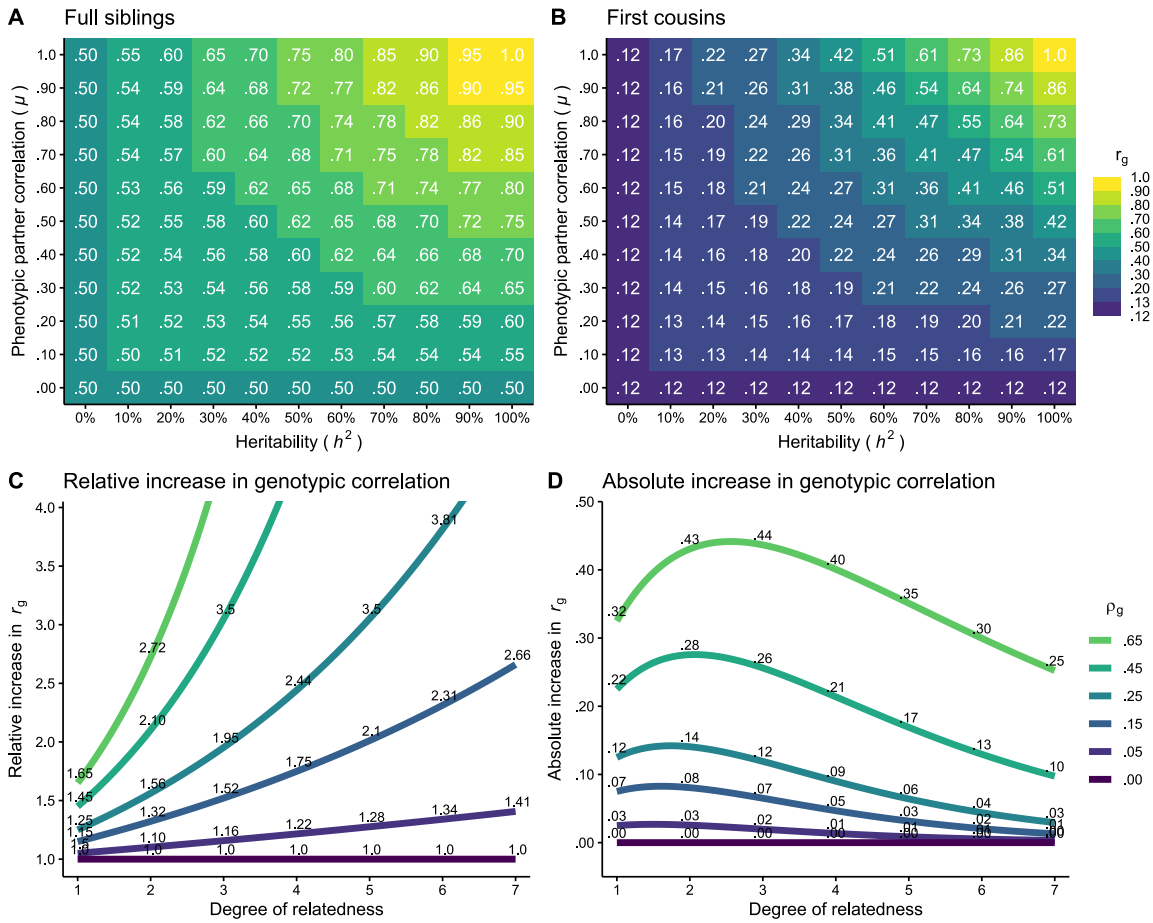

Fig. 2. Assortative mating's effect on genotypic correlations between various relatives. **Top**: The expected genotypic correlation (r_g) at equilibrium between (**A**) full siblings (i.e., first-degree relatives) and (**B**) first cousins (i.e., third-degree relatives) under different combinations of assortment strengths (μ) and heritabilities (h^2). **Bottom**: The relative (**C**) and absolute (**D**) increase in genotypic correlation at equilibrium for various relatives and genotypic correlations between partners (p_g). The reproducible code for this figure is available at https://osf.io/dgw4r/.

Expected genotypic correlations in the extended family

The model in Fig. 1 has two properties that allow a general algorithm to find the expected genotypic correlation between any two members in extended families. First, all the chains that connect the genotypes of first-degree relatives can readily be continued without breaking path tracing rules. Second, all chains between the genotypes of any two related individuals are mediated sequentially through the genotypes of first-degree relatives. The genotypic correlation between k^{th}-degree relatives, denoted r_{g_k}, can thus be attained by raising the genotypic correlation between first-degree relatives to the degree of relatedness:

$$r_{g_k} = \left(\frac{1+\rho_g}{2}\right)^k \tag{5}$$

For example, the expected genotypic correlation between third-degree relatives like first cousins is $\left(\frac{1+\rho_g}{2}\right)^3$, which can be verified by manually tracing all valid chains between A_{13} and A_{23} in Fig. 1. The genotypic correlation between non-blood relatives like in-laws, which will be non-zero under assortative mating, can be attained by linking together chains of r_{g_k} and ρ_g (for example, $Corr(A_{12}, A_{42}) = r_{g_1}\rho_g^2$). As for polygenic index correlations,

they can be approximated by replacing ρ_g with ρ_{pgi} in Equation (5), although depending on the genetic signal, the true correlation between polygenic indices may be slightly higher (see Supplementary Information, section 3).

Fig. 2 shows how assortative mating changes genotypic correlations between relatives at equilibrium. In Panels A and B, it is evident that assortative mating has a much larger effect on first cousins than full siblings. For example, for a trait where $\mu = .50$ and $h^2 = 50\%$ (meaning $\rho_g = .25$), siblings (Panel A) will have a correlation of $r_{g_1} = .625$ whereas cousins (Panel B) will have a correlation of $r_{g_3} = .244$, reflecting increases of 25% and 95%, respectively, compared to random mating. Panel C shows how this pattern extends to more distant relatives, with the genotypic correlation between second cousins 3.5 times higher than normal if $\rho_g = .25$ ($r_{g_5} = .095$ *vs.* $.031$). The larger relative increase is not merely because the correlations are smaller to begin with: Panel D shows that the largest absolute increase typically occurs in second-degree relatives like uncles/aunts and nephews/nieces.

The relatively greater increase in correlation between cousins is because third-degree relatives are affected by three assortment processes: Mother-father, uncle-aunt, and grandfather-grandmother partnerships are all correlated under assortative mating and contribute to the increased correlation (Fig. 1). For each additional degree of relatedness, there is an additional assortment process opening pathways for relatives to correlate. This pattern extends to unrelated individuals like siblings-in-laws, who would have a genotypic correlation of $\rho_g r_{g_1} = .157$ if $\rho_g = .25$. It is evident that assortative mating has a relatively larger impact on the genotypic correlation between distant relatives compared to close relatives, and that heritable traits subject to strong assortment can produce significant genotypic correlations between family members who would otherwise be virtually uncorrelated.

Gene-environment correlations, shared environment, and dominance effects

One limitation with most earlier work, such as Fisher[1], is that they assume a simplistic model where genetic similarity is the only cause of familial resemblance. In the Supplementary Information, we detail how genetic similarity between relatives are affected by dominance effects, shared environmental effects, and various forms of environmental transmission. If genetic and environmental transmission occur simultaneously, assortative mating will induce (and greatly increase) correlations between genetic and environmental factors. Such gene-environment correlations will, in this context, mimic higher heritability, leading to higher genotypic correlations between partners and thereby exacerbated genetic consequences of assortative mating. However, the relationship between the genotypic correlation between partners and the genotypic correlation between first-degree relatives will stay the same, meaning Equation (2) and Equation (4) can be used without making assumptions about gene-environment correlations or other sources of familial resemblance.

This is not the case for distant relatives. If there are substantial shared environmental effects, gene-environment correlations, or other sources of familial resemblance, the properties of Fig. 1 that allow the general algorithm in Equation (5) are no longer present. This is because non-genetic causes of familial resemblance result in pathways between distant relatives that bypass the genotypes of intermediate relatives, thus increasing the true genotypic correlation to beyond what Equation (5) would predict. Equation (5) still serves as a rough approximation, although any statistical model that relies on it could be biased if such extra pathways exist.

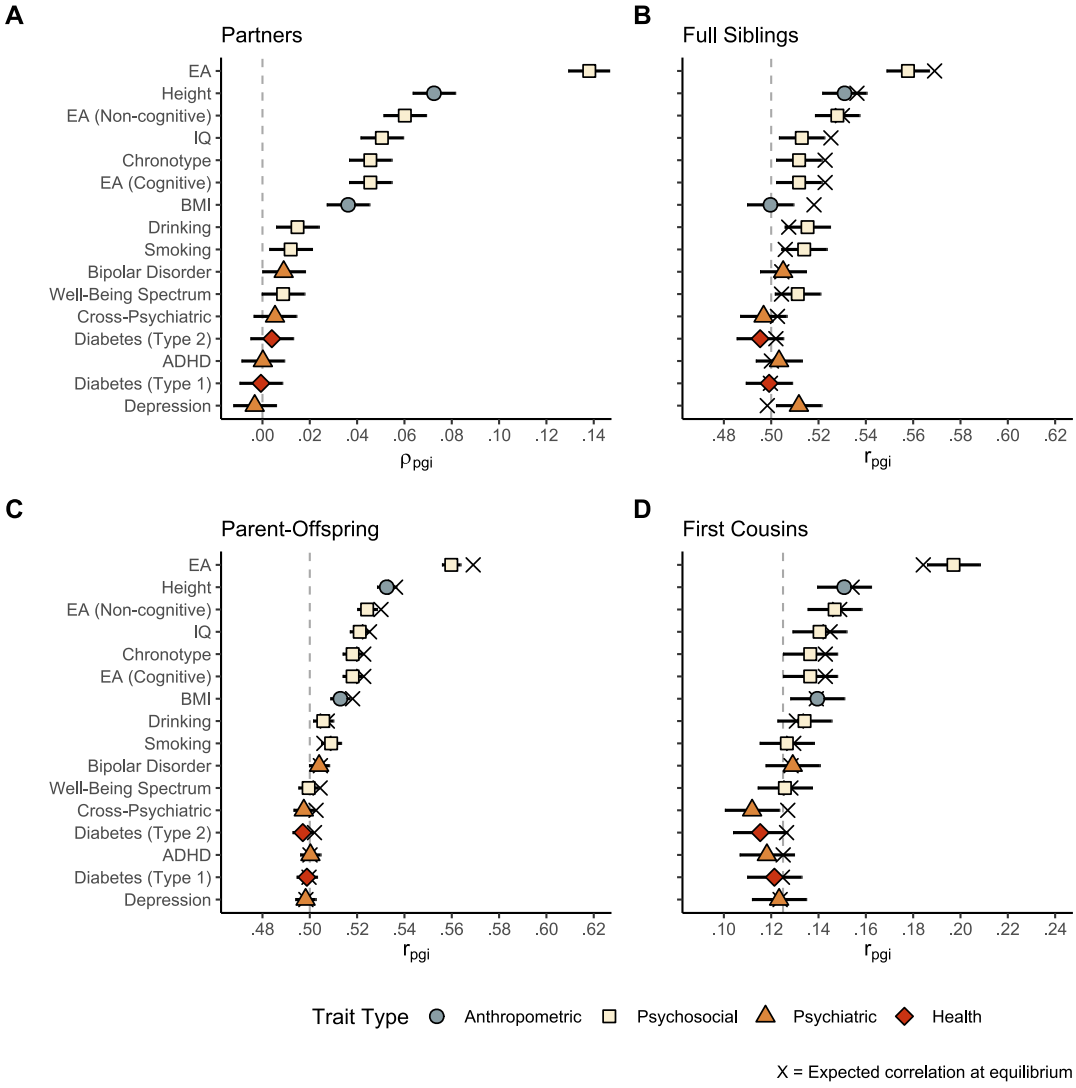

Fig. 3: Polygenic index correlations for various traits between various family members: (**A**) partners, (**B**) full siblings, (**C**) parent-offspring, and (**D**) first cousins. The vertical dashed lines are the expected correlation under random mating (i.e., the coefficient of relatedness), and the black crosses are the expected correlation at equilibrium given Equation (5). Abbreviations: EA = educational attainment; BMI = body mass index; IQ = intelligence; ADHD = attention-deficit hyperactivity disorder. Error bars are 95% confidence intervals. Correlations are also reported in Supplementary Table S17. The summary statistics and reproducible code for this figure is available at https://osf.io/dgw4r/.

Empirical polygenic index correlations between partners and relatives

Fig. 3 shows polygenic index correlations between family members for a range of traits. Nine out of sixteen traits were significantly correlated between partners (Panel A), including height (.07), body mass index (.04), intelligence (.04), and educational attainment (.14). When educational attainment was split into cognitive and non-cognitive factors (GWAS-by-subtraction[38]), we find roughly equal partner correlations for both components. Psychiatric traits like ADHD, depression, cross-psychiatric disorder, and bipolar disorder exhibited no significant correlations between partners. Keep in mind that the correlations will be biased downwards to the extent the genetic signal is poor (ref. Equation (3)).

Panels B, C, and D show polygenic index correlations between full siblings, parents and offspring, and first cousins, respectively (see Supplementary Fig. S28 for other family members). The vertical dashed lines are the expected correlations under random mating and the black crosses are the expected correlations at equilibrium given the partner correlation and Equation (5). All traits with significant correlations between partners had significantly

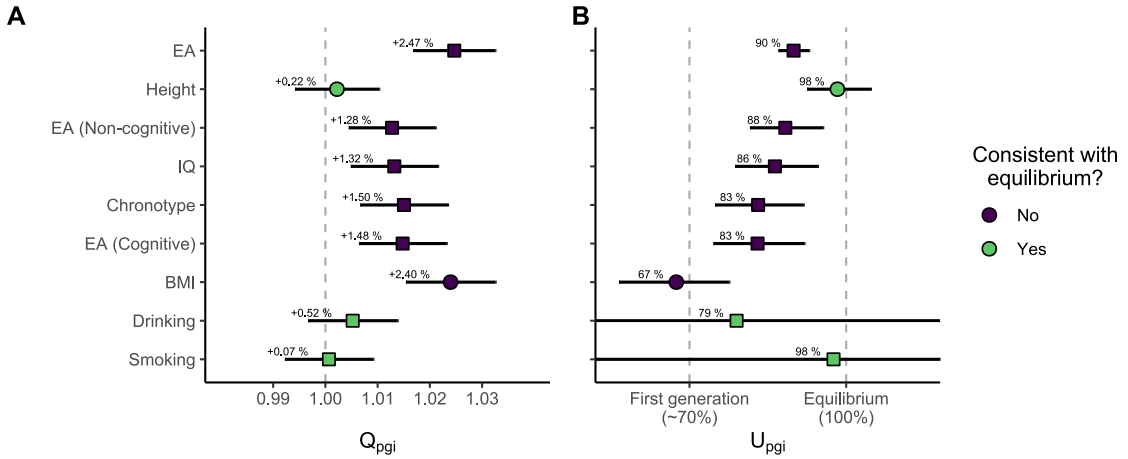

Fig. 4: (**A**) Ratio of offspring polygenic index variance to parental polygenic index variance (Q_{pgi}, see Supplementary Information, section 2.1). A value above 1 would indicate that the variance is greater in the offspring generation compared to the parental generation, as expected during disequilibrium. (**B**) Observed increase in parent-offspring correlation compared to expected increase at equilibrium (U_{pgi}, see Supplementary Information, Section 2.3). A value of about 70% would indicate that the parent generation was the first generation to assort on this trait, whereas 100% would indicate that the trait is in intergenerational equilibrium. Only traits with significant correlations between partners are shown. Abbreviations: EA = educational attainment; BMI = body mass index; IQ = intelligence. Error bars are 95% confidence intervals. The summary statistics and reproducible code for this figure is available at https://osf.io/dgw4r/.

higher parent-offspring correlations than would be expected under random mating, and we observed similar patterns for other relatives. For example, the polygenic index correlation for educational attainment was 0.56 (instead of 0.50) between full siblings and 0.20 (instead of 0.125) between first cousins.

Testing intergenerational equilibrium

We fitted structural equation models using mother-father-child trios to see if a model constrained to equal variance across generations (i.e., equilibrium) resulted in significantly worse fit (see Supplementary Information, section 6). Six out of nine traits were significantly different from equilibrium. We also investigated two consequences of disequilibrium, namely greater variance in the offspring generation (Fig. 4A) and smaller-than-expected parent-offspring correlations (Fig. 4B). During disequilibrium, the ratio of offspring polygenic index variance to parental polygenic index variance should be positive: $Q_{pgi} = \frac{Offspring\ Variance}{Parental\ Variance}$ (see Supplementary Information, section 2.1 and 3.3). However, this ratio is quite sensitive to the genetic signal of the polygenic index, and therefore provides limited information about the history of assortative mating beyond demonstrating disequilibrium. An alternative measure that is less sensitive to the genetic signal is the observed increase in polygenic index correlation as a percentage of the expected increase[19]: $U_{pgi} = \frac{Observed\ Increase}{Expected\ Increase}$ (see Supplementary Information, section 2.3 and 3.4). This provides a measure of how close the trait is to equilibrium. By comparing U_{pgi} to reference values under various heritabilities and assortment strengths, it is possible to infer the equivalent number of generations of stable assortative mating. If the parental generation was the first generation to mate assortatively, we would expect $U_{pgi} \approx 70\%$, while we would expect $U_{pgi} = 100\%$ if the trait was in equilibrium.

Height did not deviate from equilibrium: There was no significant difference between the parental and offspring variance nor between the observed and expected correlations. The results for drinking and smoking behavior were also consistent with equilibrium, although the observed partner correlation was too small to make this test informative. Body mass index and other psychosocial traits, on the other hand, did deviate from equilibrium: For

example, the polygenic index variance for educational attainment was 2.4% greater in the offspring generation compared to the parental generation. The true genetic variance ratio is likely much larger: For example, if the polygenic index captures one third of the true genetic factor ($s^2 = 1/3$), then the true variance ratio would be approximately 7.2% (see Supplementary Information, section 3.3). The parent-offspring polygenic index correlation was also slightly but significantly lower than expected at equilibrium ($U_{pgi} = 90\%$, 95% CIs: $87-93\%$), with similar results for the other psychosocial traits. When we compared this to calculations of what the observed increase would have been after successive generations of stable assortment, we found that $U_{pgi} = 90\%$ is equivalent to approximately three generations of stable assortative mating (see Supplementary Information, section 2.3). Body mass index, on the other hand, had a parent-offspring polygenic index correlation that would imply that the parent generation was the first to mate assortatively ($U_{pgi} = 67\%$, 95% CIs: $57-78\%$). This would also explain why the sibling correlation – most of whom are in the parent generation – was not higher than expected under random mating.

Discussion

In this study, our goal was to clarify the theoretical consequences of assortative mating on genetic similarity in extended families and assess empirical measures of genetic similarity to provide insights into the presence and history of assortative mating. We first employed path analysis to deduce the expected polygenic index correlations between relatives under assortative mating. We then presented empirical evidence that assortative mating is present for many traits, leading to significantly increased genetic similarity among relatives for those traits. Finally, we showed that – while assortative mating does not appear to be a recent phenomenon for most traits – genetic similarity is still increasing across generations for psychosocial traits. Here, we discuss the implications of our findings.

Our first aim was to clarify the theoretical consequences of assortative mating. One key finding is the stronger impact of assortative mating on genotypic correlations between more distant relatives. Although not a novel discovery – even Fisher mentioned it offhandedly in his seminal paper[1] – this effect has been largely overlooked in the literature (cf. [32]). This is despite important implications. A Swedish economics paper reported that nearly one-third of persistence in inequality across generations – traditionally attributable to parent-offspring relationships – is attributable to the extended family[39]. Assortative mating's effects on similarity in extended families may be key to understanding these issues. Similar logic may also apply to environmentally mediated sources of similarity[40]. We also described how assortative mating can induce and increase gene-environment correlations, which mimic higher heritability and thereby exacerbate the genetic consequences of assortative mating – especially correlations between distant relatives.

The second aim of this study was to investigate which traits show genetic evidence of assortative mating. One key challenge when evaluating the pervasiveness of assortative mating is that phenotypic partner similarity can come about from multiple processes. Genotypic similarity, on the other hand, can more confidently be attributed to assortative mating. Most anthropometric traits and psychosocial traits had significant polygenic index correlations between partners. The largest correlation was for educational attainment (.14), which adds to the growing list of evidence that variants associated with educational attainment are undergoing assortative mating[6,19,23,25,41].

Somewhat surprisingly, psychiatric traits did not show evidence of assortative mating despite pervasive phenotypic partner correlations[15,16]. The findings are in line with a recent study that found no genetic partner similarity on general risk for psychopathology (i.e., the "p-factor")[42]. Our findings seemingly contradict Torvik, et al. [19], who reported evidence of assortative mating on depression using a smaller subset of the same cohort. However, that paper used a structural equation model requiring both genetic and phenotypic data, and the polygenic index correlations reported in that paper match those we report here. This could indicate that phenotypic partner similarity in mental health is caused by processes other than assortative mating, such as convergence[20] (which was not modelled in Torvik, et al. [19]). On the other hand, they could also be false negatives resulting from inadequate polygenic indices. As highlighted in Equation (3), the polygenic index correlation between partners should be the product of the phenotypic correlation attributable to assortative mating (μ), the heritability (h^2), and the genetic signal (s^2). If the polygenic index fails to adequately measure the relevant genetic factors (meaning $s^2 \approx 0$), for example due to lack of statistical power or other measurement issues[43] in the underlying genome-wide associations

study (GWAS), then the polygenic index correlation will be biased towards zero. The highest observed correlations were for educational attainment and height, which are among the traits with the largest sample sizes in the underlying GWAS. A corollary is that the correlations reported here do not quantify the exact degree of assortative mating because it is confounded by the genetic signal of the polygenic index. Complicating inference further is that the genetic signal is itself increased under assortative mating.

Our third aim was to investigate whether relatives were more genetically similar for traits that exhibit evidence of assortative mating. Our findings broadly correspond to theoretical expectations: Traits with significant polygenic index correlations between partners showed increased similarity between relatives, whereas traits with no correlations between partners broadly exhibit patterns as expected under random mating. These empirical patterns demonstrate the theoretical expectations derived earlier, meaning we should expect distant relatives to be highly correlated for traits under strong assortment. The correlations reported here are underestimated by the quality of the polygenic index, meaning the true genotypic correlations between relatives are likely much larger. Our findings have at least two implications. First, genetic variants associated with traits undergoing assortment, such as educational attainment, cluster in extended families, thus increasing or maintaining societal stratification by families[39] (i.e. between-family variation); and second, genetic studies that unknowingly involve numerous distantly related individuals may be biased if the genotypic correlations between them are not negligible.

For educational attainment, the polygenic index correlations between first-degree relatives are lower than expected (indicating disequilibrium, see below) while correlations between third- and fourth-degree relatives are higher than expected. This is consistent with substantial gene-environment correlations for educational attainment[44-46]. In the Supplementary Information, we showed theoretically and with simulations that correlations between higher-degree relatives (but not first-degree relatives) will be higher than expected given Equation (5) if such gene-environment correlations are present.

Our fourth aim was to investigate the history of assortative mating. Our findings differed across traits: Height did not deviate from equilibrium expectations, whereas psychosocial traits such as educational attainment did. This was evident in both lower-than-expected parent-offspring polygenic index correlations and greater variance in the offspring generation. Whether or not a trait is in intergenerational equilibrium has important implications for the consequences of assortative mating because it decides whether differences are increasing across generations or merely maintained. We found that polygenic index variance was stable across generations for height (as well as for traits not undergoing assortative mating). However, psychosocial traits have greater variance in the offspring generation, implying that the traits are in disequilibrium and that assortative mating is currently leading to *increased* genetic differences in these traits. Although the non-genetic consequences may differ, assortative mating may therefore play a key role in explaining recent increases in inequality[10,13].

Despite being in disequilibrium, the evidence does not suggest that the parental generation was the first to assort on educational attainment. Instead, it appears that the trait is quite near equilibrium. This would also explain the discrepancy between our conclusion and that in Torvik, et al. [19], who found no significant deviation from equilibrium using an earlier version of data from the same cohort. We primarily used variance differences across generations whereas Torvik, et al. [19] compared the predicted and expected correlations between siblings and partners. Considering that the sibling correlation in Fig. 3B is significantly lower than expected given equilibrium,

the change in result likely stem from an increase in statistical power, owing to more genotyped individuals available in the current sample, and further aided by the use of parent-offspring dyads instead of sibling dyads. In this paper, we estimate that the evidence for educational attainment corresponds to approximately three generations of stable assortative mating, but the exact history of assortment may be longer if the strength of assortment has varied over time or if the genotype-phenotype correlation increased for other reasons[47].

Many genetic research methods assume random mating, but our findings suggest that such assumptions are unwarranted for many traits. Accounting for assortative mating poses its own challenges, as the genetic consequences and corresponding methods needed depend on whether assortative mating started recently or has reached intergenerational equilibrium. Studies on the genetics of educational attainment especially – or the many traits that correlate with educational attainment[48] – may therefore be biased unless this is properly accounted for. Twin and family studies that account for assortative mating typically assume equilibrium[37]. For example, Clark [49] uses equations that assume equilibrium when he claims that familial correlations in social class in the United Kingdom can be explained by genetic similarity alone. Conversely, Kong, et al. [50], who investigated genetic nurture effects of educational attainment in an Icelandic sample, assumed no assortment prior to their parental generation. Our findings imply that, for some traits, neither of these assumptions are valid. Although the patterns and history of assortment may be different across populations, future research should investigate how the conclusions from Kong, et al. [50] and related papers depend on these assumptions[51,52].

Newer genetic methods that can account for disequilibrium are being developed[36,53]. When these methods are impractical, the potential biases induced by different assumptions must be considered on a case-by-case and method-by-method basis. Different methods will be biased in different ways. For example, assortative mating leads to underestimated heritability in classical twin designs[37,54] and overestimated heritability in molecular designs[55], with the corollary that the missing heritability problem may be larger than previously assumed[56,57]. Overall, researchers must carefully consider what impacts the presence and history of assortative mating would have on their results.

Despite our large sample size, our results are limited by low-quality polygenic indices, which results in lower partner correlations and consequently less power to detect assortative mating. This is amplified for the tests of equilibrium, where smaller polygenic index correlations between partners result in less statistical power to detect deviations from equilibrium. Our tests for equilibrium are therefore less conclusive for traits with small polygenic index correlations, such as drinking and smoking behavior.

Another concern is that our results may be confounded by population stratification[58], where (1) the trait in question happens to be more common within certain strata (e.g., subcultures or geographical areas), (2) some genetic variants are randomly present at higher frequencies in these strata, and (3) individuals are more likely to mate within these strata. The combination of the first two phenomena would result in a spurious correlation between those genetic variants and the trait, and when coupled with the third phenomenon, similar spurious correlations could emerge between partners. While we controlled for 20 principal components in our analysis, which is the standard method for addressing stratification[59], this approach may not fully account for this phenomenon[60]. However, the evidence we present aligns well with predictions given assortative mating. It is also not obvious how population stratification could explain increased variance in the offspring generation. Consequently, our results

should be considered indicative of assortative mating until a more compelling alternative explanation is offered. Future theoretical work should investigate how the consequences of assortative mating and population stratification differ so that they can better be distinguished in future research.

There are several interesting research avenues that could follow from this work. First, there may be some selection bias in the cohort study our results are based on. Future work using population-wide phenotypic data might provide insights into how much this matters. Second, patterns of assortative mating are likely to vary between populations[61,62], meaning that our empirical findings is not universally generalizable. Replicating these results in other populations will therefore be beneficial. Third, the approach we use here is agnostic as to which trait(s) the polygenic indices actually measure, and which phenotype(s) are being assorted upon. Future research may want to investigate what set of phenotypes mediate the polygenic index correlations between partners, as it may not always be attributable to the phenotype that the polygenic index supposedly measures.

Methods

Sample

We used data from the Norwegian Mother, Father and Child Cohort Study (MoBa)[26]. MoBa is a population-based pregnancy cohort study conducted by the Norwegian Institute of Public Health. Participants were recruited from all over Norway from 1999-2008. The women consented to participation in 41% of the pregnancies. Blood samples were obtained from both parents during pregnancy and from mothers and children (umbilical cord) at birth[63]. The cohort includes approximately 114,500 children, 95,200 mothers and 75,200 fathers. The current study is based on version 12 of the quality-assured data files released for research in January 2019. The establishment of MoBa and initial data collection was based on a license from the Norwegian Data Protection Agency and approval from The Regional Committees for Medical and Health Research Ethics. The MoBa cohort is currently regulated by the Norwegian Health Registry Act. The current study was approved by The Regional Committees for Medical and Health Research Ethics (2017/2205).

For the correlations, the sample included 47,135 unique mother-father dyads (i.e., partners) where both had been genotyped and passed quality control [27]. As described in Corfield, et al. [27], relatedness relationships in MoBa were inferred from genetic data by applying KING programs[64] to a subset of single nucleotide polymorphisms (SNPs) with call rate < 98% and minor allele frequency (MAF) < 5%. KING accurately infers monozygotic twin or duplicate pairs (kinship coefficient > 0.3540), first-degree (parent-offspring, full siblings, dizygotic twin pairs; kinship coefficient range 0.1770 - 0.3540), second-degree (half siblings, grandparent-offspring, avuncular relationships; kinship coefficient range 0.0884 – 0.1770), and third-degree (first cousins; kinship coefficient range 0.0442 – 0.0884) relationships. This method identified 117,041 parent-offspring dyads, 22,575 full sibling dyads, 35,923 second-degree dyads (e.g., uncle-nephew), 28,330 third-degree dyads (e.g., first cousins), 9,392 fourth-degree dyads, and 235,209 dyads of unrelated family members (e.g., in-laws, nephews–uncles' spouses, partners, etc.,) where both members of the dyads had been genotyped and passed quality control.

To test equilibrium, we used all available mother-father-child trios from MoBa. We relied on trios to test equilibrium for the following reasons: 1) It allowed estimating the partner correlation and parent-offspring correlations in the same model; 2) it allowed us to include both the mother-offspring and father-offspring simultaneously thus increasing statistical power; 3) it allowed us to estimate variances separately for the two generations; 4) there was no need to distinguish between correlations between relatives in the parent generation and in the offspring generation, as this is inherent in the design; 5) focusing on the nuclear family removes the need to make assumptions about the genetic signal or gene-environment correlations; and finally, 6) the sample in MoBa is inherently selected on parent-offspring dyads whereas the availability of other relatives is coincidental. Using other relatives, such as siblings, could therefore lead to stronger ascertainment bias. After randomly selecting one offspring from each nuclear family, we were able to construct a sample of 81,145 genotyped mothers, 54,550 genotyped fathers, and 71,544 genotyped offspring, resulting in a total of 93,767 incomplete and complete trios. Of these, 36,764 were complete trios, whereas 10,371 included only partners, 24,659 included only mother-offspring dyads, and 4,914 included only father-offspring dyads.

Measures

We used beta weights from large, publicly available up-to-date genome-wide association studies listed the Supplementary Information, section 8. None of the used genome-wide association studies used data from MoBa. Polygenic indices were calculated using LDPred[65], a Bayesian approach that uses a prior on the expected polygenicity of a trait (assumed fraction of non-zero effect markers) and adjusts for linkage disequilibrium (LD) based on a reference panel to compute SNPs weights. Genotypes were coordinated with the summary statistics, with the number of overlapping SNPs reported in Supplementary Information, section 8. LD adjustment was performed using the European subsample of the 1000 Genomes genotype data as LD reference panel[66]. The weights were estimated based on the heritability explained by the markers in the GWAS summary statistics and the assumed fraction of markers with non-zero effects. For each GWAS trait we created LDpred PGI with the –score command in plink2[67]. Prior to calculating correlations between partners and relatives, we residualised the polygenic indices by regressing out the first 20 principal components of genetic ancestry, as well as chip, imputation, and batch number.

Statistics

The polygenic index correlations (and 95% confidence intervals) were attained by correlating the residualised polygenic indices between partners and relatives using *cor.test* in R. 4.0.3[68]. We tested whether the observed correlations were consistent with equilibrium by fitting structural equation models to data on mother-father-child trios, and testing whether a model constrained to equilibrium via equal variance across generations resulted in significantly worse fit. We describe this procedure in more detail in the Supplementary Information, section 6.

Data Availability

Data from the Norwegian Mother, Father and Child Cohort Study (MoBa) used in this study are managed by the national health register holders in Norway (Norwegian Institute of Public Health) and can be made available to researchers, provided approval from the Regional Committees for Medical and Health Research Ethics (REC), compliance with the EU General Data Protection Regulation (GDPR) and approval from the data owners. The consent given by the participants does not open for storage of data on an individual level in repositories or journals. Researchers who want access to data sets for replication should apply through helsedata.no. Access to data sets requires approval from The Regional Committee for Medical and Health Research Ethics in Norway and an agreement with MoBa.

The summary statistics and reproducible code for the figures in this manuscript is available at https://osf.io/dgw4r/.

Code Availability

Scripts used for simulations are provided in Supplementary Software 1 and at https://osf.io/dgw4r/. Scripts for reproducing figures (including relevant summary statistics) are available at https://osf.io/dgw4r/

References

1 Fisher, R. A. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399-433 (1918). https://doi.org/10.1017/s0080456800012163

2 Wright, S. Systems of mating. III. Assortative Mating based on Somatic resemblance. *Genetics* **6**, 144-161 (1921). https://doi.org/10.1093/genetics/6.2.144

3 Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits*. (Sinauer Associates, Inc., 1998).

4 Nagylaki, T. The correlation between relatives with assortative mating. *Annals of Human Genetics* **42**, 131-137 (1978). https://doi.org/10.1111/j.1469-1809.1978.tb00938.x

5 Gimelfarb, A. A general linear model for the genotypic covariance between relatives under assortative mating. *Journal of Mathematical Biology* **13**, 209-226 (1981). https://doi.org/10.1007/bf00275215

6 Robinson, M. R. *et al.* Genetic evidence of assortative mating in humans. *Nature Human Behaviour* **1**, 0016 (2017). https://doi.org/10.1038/s41562-016-0016

7 Robert Plomin, Eva Krapohl & F. O'Reilly, P. Assortative Mating—A Missing Piece in the Jigsaw of Psychiatric Genetics. *JAMA Psychiatry* **73**, 323 (2016). https://doi.org/10.1001/jamapsychiatry.2015.3204

8 Bulmer, M. G. *The Mathematical Theory of Quantitative Genetics*. (Oxford University Press, 1980).

9 Yengo, L. Mate choice through a genomic lens. *Nature Reviews Genetics* (2023). https://doi.org/10.1038/s41576-023-00605-w

10 Schwartz, C. R. Trends and Variation in Assortative Mating: Causes and Consequences. *Annual Review of Sociology* **39**, 451-470 (2013). https://doi.org/10.1146/annurev-soc-071312-145544

11 Mare, R. D. Five Decades of Educational Assortative Mating. *American Sociological Review* **56**, 15-32 (1991). https://doi.org/10.2307/2095670

12 Birkelund, G. E. & Heldal, J. Who marries whom? Educational homogamy in Norway. *Demographic Research* **8**, 1-30 (2003). https://doi.org/10.4054/DemRes.2003.8.1

13 Greenwood, J., Guner, N., Kocharkov, G. & Santos, C. Marry Your Like: Assortative Mating and Income Inequality. *American Economic Review* **104**, 348-353 (2014). https://doi.org/10.1257/aer.104.5.348

14 Eika, L., Mogstad, M. & Zafar, B. Educational Assortative Mating and Household Income Inequality. *Journal of Political Economy* **127**, 2795-2835 (2019). https://doi.org/10.1086/702018

15 Nordsletten, A. E., Henrik Larsson, J. Crowley, J., Catarina Almqvist, Paul Lichtenstein & David Mataix-Cols. Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. *JAMA Psychiatry* **73**, 354 (2016). https://doi.org/10.1001/jamapsychiatry.2015.3192

16 Horwitz, T. B., Balbona, J. V., Paulich, K. N. & Keller, M. C. Evidence of correlations between human partners based on systematic reviews and meta-analyses of 22 traits and UK Biobank analysis of 133 traits. *Nature Human Behaviour* (2023). https://doi.org/10.1038/s41562-023-01672-z

17 Luo, S. & Klohnen, E. C. Assortative Mating and Marital Quality in Newlyweds: A Couple-Centered Approach. *Journal of Personality and Social Psychology* **88**, 304-326 (2005). https://doi.org/10.1037/0022-3514.88.2.304

18 Luo, S. Assortative mating and couple similarity: Patterns, mechanisms, and consequences. *Social and Personality Psychology Compass* **11**, e12337 (2017). https://doi.org/10.1111/spc3.12337

19 Torvik, F. A. *et al.* Modeling assortative mating and genetic similarities between partners, siblings, and in-laws. *Nature Communications* **13** (2022). https://doi.org/10.1038/s41467-022-28774-y

20 Sjaarda, J. & Kutalik, Z. Partner choice, confounding and trait convergence all contribute to phenotypic partner similarity. *Nature Human Behaviour* (2023). https://doi.org/10.1038/s41562-022-01500-w

21 Turkheimer, E. Three Laws of Behavior Genetics and What They Mean. *Current Directions in Psychological Science* **9**, 160-164 (2000). https://doi.org/10.1111/1467-8721.00084

22 Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* **47**, 702-709 (2015). https://doi.org/10.1038/ng.3285

23 W. Domingue, B., J. Fletcher, D. Conley & D. Boardman, J. Genetic and educational assortative mating among US adults. *Proceedings of the National Academy of Sciences* **111**, 7996-8000 (2014). https://doi.org/10.1073/pnas.1321426111

24 Okbay, A. *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nature Genetics* **54**, 437-449 (2022). https://doi.org/10.1038/s41588-022-01016-z

25 Yengo, L. *et al.* Imprint of assortative mating on the human genome. *Nature Human Behaviour* **2**, 948-954 (2018). https://doi.org/10.1038/s41562-018-0476-3

26 Magnus, P. *et al.* Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology* **45**, 382-388 (2016). https://doi.org/10.1093/ije/dyw029

27 Corfield, E. C. *et al.* The Norwegian Mother, Father, and Child cohort study (MoBa) genotyping data resource: MoBaPsychGen pipeline v.1. Preprint at https://doi.org/10.1101/2022.06.23.496289 (BioRxiv, 2022).

28 Gimelfarb, A. Quantitative characters under assortative mating: Gametic model. *Theoretical Population Biology* **25**, 312-330 (1984). https://doi.org/10.1016/0040-5809(84)90012-1

29 Hayashi, T. Genetic variance under assortative mating in the infinitesimal model. *Genes & Genetic Systems* **73**, 397-405 (1998). https://doi.org/10.1266/ggs.73.397

30 Wright, S. Systems of mating. II. The effects of inbreeding on the genetic composition of a population. *Genetics* **6**, 124-143 (1921). https://doi.org/10.1093/genetics/6.2.124

31 Wilson, S. R. The correlation between relatives under the multifactorial model with assortative mating. *Annals of Human Genetics* **37**, 189-204 (1973). https://doi.org/10.1111/j.1469-1809.1973.tb01826.x

32 Yengo, L. & Visscher, P. M. Assortative mating on complex traits revisited: Double first cousins and the X-chromosome. *Theoretical Population Biology* **124**, 51-60 (2018). https://doi.org/10.1016/j.tpb.2018.09.002

33 Wright, S. Correlation and Causation. *Journal of Agricultural Research* **20**, 557-585 (1921).

34 Wright, S. The Method of Path Coefficients. *The Annals of Mathematical Statistics* **5**, 161-215 (1934). http://www.jstor.org/stable/2957502

35 Cloninger, C. R. Interpretation of intrinsic and extrinsic structural relations by path analysis: theory and applications to assortative mating. *Genetical Research* **36**, 133-145 (1980). https://doi.org/10.1017/s0016672300019765

36 Balbona, J. V., Kim, Y. & Keller, M. C. Estimation of Parental Effects Using Polygenic Scores. *Behavior Genetics* **51**, 264-278 (2021). https://doi.org/10.1007/s10519-020-10032-w

37 Keller, M. C. *et al.* Modeling Extended Twin Family Data I: Description of the Cascade Model. *Twin Research and Human Genetics* **12**, 8-18 (2009). https://doi.org/10.1375/twin.12.1.8

38 Demange, P. A. *et al.* Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nature Genetics* **53**, 35-44 (2021). https://doi.org/10.1038/s41588-020-00754-2

39 Adermon, A., Lindahl, M. & Palme, M. Dynastic Human Capital, Inequality, and Intergenerational Mobility. *American Economic Review* **111**, 1523-1548 (2021). https://doi.org/10.1257/aer.20190553

40 Rice, J., Cloninger, C. R. & Reich, T. Multifactorial inheritance with cultural transmission and assortative mating. I. Description and basic properties of the unitary models. *American Journal of Human Genetics* **30**, 618-643 (1978).

41 Abdellaoui, A., Borcan, O., Chiappori, P.-A. & Hugh-Jones, D. Trading Social Status for Genetics in Marriage Markets: Evidence from UK Biobank. *HCEO Working Paper Series* (2022).

https://hceconomics.uchicago.edu/research/working-paper/trading-social-status-genetics-marriage-markets-evidence-uk-biobank

42 Ayorech, Z. *et al.* The structure of psychiatric comorbidity without selection and assortative mating. Preprint at https://doi.org/10.31234/osf.io/d8t3q (PsyArXiv, 2023).

43 Fried, E. I., Flake, J. K. & Robinaugh, D. J. Revisiting the theoretical and methodological foundations of depression measurement. *Nature Reviews Psychology* (2022). https://doi.org/10.1038/s44159-022-00050-2

44 Augustine Kong *et al.* The nature of nurture: Effects of parental genotypes. *Science* **359**, 424-428 (2018). https://doi.org/10.1126/science.aan6877

45 Young, A. I., Benonisdottir, S., Przeworski, M. & Kong, A. Deconstructing the sources of genotype-phenotype associations in humans. *Science* **365**, 1396-1400 (2019). https://doi.org/doi:10.1126/science.aax3710

46 Young, A. I. *et al.* Relatedness disequilibrium regression estimates heritability without environmental bias. *Nature Genetics* **50**, 1304-1310 (2018). https://doi.org/10.1038/s41588-018-0178-9

47 Heath, A. C. *et al.* Education policy and the heritability of educational attainment. *Nature* **314**, 734-736 (1985). https://doi.org/10.1038/314734a0

48 Brendan Bulik-Sullivan *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**, 1236-1241 (2015). https://doi.org/10.1038/ng.3406

49 Clark, G. The inheritance of social status: England, 1600 to 2022. *Proceedings of the National Academy of Sciences* **120** (2023). https://doi.org/10.1073/pnas.2300926120

50 Kong, A. *et al.* The nature of nurture: Effects of parental genotypes. *Science* **359**, 424-428 (2018). https://doi.org/10.1126/science.aan6877

51 Nivard, M. G. *et al.* Neither nature nor nurture: Using extended pedigree data to elucidate the origins of indirect genetic effects on offspring educational outcomes. Preprint at https://doi.org/10.31234/osf.io/bhpm5 (PsyArXiv, 2022).

52 Demange, P. A. *et al.* Estimating effects of parents' cognitive and non-cognitive skills on offspring education using polygenic scores. *Nature Communications* **13** (2022). https://doi.org/10.1038/s41467-022-32003-x

53 Bilghese, M. *et al.* A General Approach to Adjusting Genetic Studies for Assortative Mating. Preprint at https://doi.org/10.1101/2023.09.01.555983 (Cold Spring Harbor Laboratory, 2023).

54 Knopik, V. S., Neiderhiser, J. M., DeFries, J. C. & Plomin, R. *Behavioral Genetics*. 7th edn, (Worth Publishers, 2017).

55 Border, R. *et al.* Assortative mating biases marker-based heritability estimators. *Nature Communications* **13** (2022). https://doi.org/10.1038/s41467-022-28294-9

56 Young, A. I. Solving the missing heritability problem. *PLOS Genetics* **15**, e1008222 (2019). https://doi.org/10.1371/journal.pgen.1008222

57 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753 (2009). https://doi.org/10.1038/nature08494

58 Abdellaoui, A., Verweij, K. J. H. & Zietsch, B. P. No evidence for genetic assortative mating beyond that due to population stratification. *Proceedings of the National Academy of Sciences* **111**, E4137-E4137 (2014). https://doi.org/10.1073/pnas.1410781111

59 Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904-909 (2006). https://doi.org/10.1038/ng1847

60 Blanc, J. & Berg, J. J. Testing for differences in polygenic scores in the presence of confounding. Preprint at https://doi.org/10.1101/2023.03.12.532301 (BioRxiv, 2023).

61 Yamamoto, K. *et al.* Genetic footprints of assortative mating in the Japanese population. *Nature Human Behaviour* (2022). https://doi.org/10.1038/s41562-022-01438-z

62 Verweij, K. J. H. & Abdellaoui, A. Partner-choice genetics in Japan. *Nature Human Behaviour* (2022). https://doi.org/10.1038/s41562-022-01439-y

63 Paltiel, L. *et al.* The biobank of the Norwegian Mother and Child Cohort Study – present status. *Norsk Epidemiologi* **24** (2014). https://doi.org/10.5324/nje.v24i1-2.1755

64 Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M. & Chen, W.-M. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-2873 (2010). https://doi.org/10.1093/bioinformatics/btq559

65 Vilhjálmsson, J., Bjarni *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *The American Journal of Human Genetics* **97**, 576-592 (2015). https://doi.org/10.1016/j.ajhg.2015.09.001

66 Chou, W.-C. *et al.* A combined reference panel from the 1000 Genomes and UK10K projects improved rare variant imputation in European and Chinese samples. *Scientific Reports* **6**, 39313 (2016). https://doi.org/10.1038/srep39313

67 Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M. & Lee, J. J. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4** (2015). https://doi.org/10.1186/s13742-015-0047-8

68 R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria, 2022).

Acknowledgements

This work is part of the REMENTA and PARMENT projects and was supported by the Research Council of Norway (#300668 and #334093, respectively). The data acquisition and project management was supported by the Research Council of Norway (#262177, #288083, and #336078). E.Y. is funded by the European Union (Grant agreement #101045526 and #818425). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. The Norwegian Mother, Father and Child Cohort Study is supported by the Norwegian Ministry of Health and Care Services and the Ministry of Education and Research. We are grateful to all the participating families in Norway who take part in this on-going cohort study. We thank the Norwegian Institute of Public Health (NIPH) for generating high-quality genomic data. This research is part of the HARVEST collaboration, supported by the Research Council of Norway (#229624). We also thank the NORMENT Centre for providing genotype data, funded by the Research Council of Norway (#223273), South East Norway Health Authorities and Stiftelsen Kristian Gerhard Jebsen. We further thank the Center for Diabetes Research, the University of Bergen for providing genotype funded by the ERC AdG project SELECTionPREDISPOSED, Stiftelsen Kristian Gerhard Jebsen, Trond Mohn Foundation, the Research Council of Norway, the Novo Nordisk Foundation, the University of Bergen, and the Western Norway Health Authorities. This work was performed on the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT). This work was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 262700.

Author Contributions

H.F.S. conceived of the idea, designed the theoretical models, and derived the relevant equations. E.C.C., R.C., and A.C.S. contributed to sample preparation and quality control of genomic data and polygenic indices with support from E.Y. H.F.S. carried out the analyses with support from N.H.E. and R.C. H.F.S. planned and carried out the simulations with help from E.M.E. and F.A.T. T.H.K contributed to the interpretation of the results. H.F.S. wrote the manuscript (incl. the supplementary information) with input from all authors. E.C.C. wrote parts of the methods. E.Y. and E.C.C contributed to data generation and acquisition. E.M.E and F.A.T supervised the project. All authors provided critical feedback, discussed the results, and helped shape the manuscript.

Competing interests

The authors declare no competing interests.