nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-020-00991-9

Check for updates

Evidence of horizontal indirect genetic effects in humans

Charley Xia[1], Oriol Canela-Xandri[1,2], Konrad Rawlik[1] and Albert Tenesa[1,2,3]✉

Indirect genetic effects, the effects of the genotype of one individual on the phenotype of other individuals, are environmental factors associated with human disease and complex trait variation that could help to expand our understanding of the environment linked to complex traits. Here, we study indirect genetic effects in 80,889 human couples of European ancestry for 105 complex traits. Using a linear mixed model approach, we estimate partner indirect heritability and find evidence of partner heritability on ~50% of the analysed traits. Follow-up analysis suggests that in at least ~25% of these traits, the partner heritability is consistent with the existence of indirect genetic effects including a wide variety of traits such as dietary traits, mental health and disease. This shows that the environment linked to complex traits is partially explained by the genotype of other individuals and motivates the need to find new ways of studying the environment.

Indirect genetic effects (IGEs), the effect of the genotype of one individual on the phenotype of other individuals[1,2], have previously been described in a limited number of traits in plants, farm animals, model organisms and humans[3–12]. IGEs are part of the environment of an individual that are contributed through the genotype of other socially or biologically related individuals. Unlike many other environmental contributors to phenotypic variation, genotypes are measured consistently and accurately, and are fixed from birth. As such, IGEs could potentially be important environmental factors associated with human disease and complex trait variation. This class of environmental factors offer the opportunity to expand the repertoire of environmental factors measured in epidemiological studies and improve disease risk prediction and stratification.

In humans, vertical IGEs (that is, across-generation IGEs) such as maternal genetic effects, where the maternal genotypes affect their offspring phenotypes, have been described for traits such as birth weight[9]. Similarly, non-transmitted parental genotypes linked to cognitive ability have been reported to influence educational attainment of children[10]. These studies show that vertical IGEs can be mediated through biological (for example, birth weight) and social or behavioural factors that modify the environment in which a person lives (for example, in the case of cognition). Horizontal IGEs (that is, within-generation IGEs) have also been reported for school friends in relation to educational attainment[11]. However, it is as yet unclear to what degree IGEs are spread across a broad range of phenotypes, and whether within-generation effects exist among non-blood relatives (for example, partners) and among blood relatives (for example, siblings).

Here, we search for evidence of horizontal IGEs for 105 complex traits using 80,889 couples of European ancestry present in UK Biobank[13]. We replicate our findings in 8,144 sibling pairs of European ancestry and 3,752 additional couples of mixed ancestries (one individual of European and the other of non-European ancestry). We use a linear mixed model approach to estimate what we define as partner indirect heritability (h^2_{Partner}). This is similar to the single-nucleotide polymorphism (SNP) heritability (we use the term h^2_{Own} to distinguish it from h^2_{Partner}), and it is defined as the proportion of phenotypic variance observed for individuals that is explained by the additive genetic effects of their social partners. We estimate h^2_{Partner} for a wide variety of traits that include anthropometric traits, lifestyle choices, mental health and late-onset diseases and demonstrate the utility of h^2_{Partner} through extensive simulations based on UK Biobank data. In particular, we demonstrate that our h^2_{Partner} estimates observed for real phenotypes capture IGEs besides assortative mating, and furthermore that the estimates are not consistent with the presence of assortative mating alone. Overall, about 50% of traits investigated have significant h^2_{Partner} estimates ($P<0.05$) and about 25% of those have corroborating evidence of IGEs.

Results

Evidence for partner indirect heritability and its difference between sexes. We analysed over 80,889 couples of European ancestry from UK Biobank (Methods) and 105 complex traits that represent a wide range of human phenotypic variation (Supplementary Table 1). We first estimated h^2_{Partner} for 91 traits that are expressed in women and men. To avoid possible problems of confounding due to the individual's own genotype, each trait was modelled as a function of the genotype of the individual and the genotype of their partner (that is, we estimated h^2_{Own} and h^2_{Partner} jointly; Methods). For binary traits, heritability estimates on the observed scale were transformed to the liability scale[14,15] using the prevalences in both the UK Biobank and in the couples (Supplementary Table 2).

Among the 91 traits tested, 28 (31%) were found to have h^2_{Partner} estimates significantly different from zero after Bonferroni correction ($P<0.05/105=4.8\times10^{-4}$; Table 1). Anthropometric traits (6 out of 6) and dietary traits (11 out of 18) were the categories most enriched, while educational attainment (in years), habits such as time spent watching television and smoking status, diseases and mental-health-related traits such as astigmatism and mood swings were also associated with the partner genotype (Table 1). However, Bonferroni correction is too stringent for our set of correlated traits. Among these 91 traits, 47 are significant at a FDR of 5% and 51 are significant at a marginal level of $P<0.05$ (Table 1 and Fig. 1). As the aim of the work is to evaluate the evidence of IGEs across a range of traits rather than to identify a few traits affected by IGEs, we believe

[1]The Roslin Institute, The Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, UK. [2]MRC IGMM, Western General Hospital, University of Edinburgh, Edinburgh, UK. [3]Usher Institute, Edinburgh bioQuarter, University of Edinburgh, Edinburgh, UK. ✉e-mail: albert.tenesa@ed.ac.uk

Table 1 | GREML results (in %) for significant non-sex-specific traits and couple-shared traits

Trait	Heritability and *P* values			
Non-sex-specific trait (significant at $P < 0.05/91$)	h^2_{Own} (s.e.)	*P*	$h^2_{Partner}$ (s.e.)	*P*
Standing height	51.9 (0.4)	$<1.0\times10^{-300}$	2.4 (0.2)	7.9×10^{-43}
Body mass index	24.9 (0.4)	$<1.0\times10^{-300}$	1.7 (0.2)	8.8×10^{-16}
Weight	28.3 (0.4)	$<1.0\times10^{-300}$	1.0 (0.2)	4.4×10^{-8}
Body fat percentage	24.7 (0.4)	$<1.0\times10^{-300}$	1.7 (0.2)	3.9×10^{-16}
Waist-to-hip ratio	18.6 (0.4)	$<1.0\times10^{-300}$	1.3 (0.2)	3.5×10^{-8}
Basal metabolic rate	32.0 (0.4)	$<1.0\times10^{-300}$	1.1 (0.2)	1.9×10^{-8}
Ease of skin tanning	23.4 (0.4)	$<1.0\times10^{-300}$	0.9 (0.2)	2.6×10^{-5}
Cooked vegetable intake	4.8 (0.3)	3.9×10^{-76}	1.1 (0.2)	8.4×10^{-7}
Salad/raw vegetable intake	5.5 (0.3)	3.0×10^{-98}	0.9 (0.3)	3.1×10^{-4}
Dried fruit intake	4.6 (0.3)	3.2×10^{-69}	1.2 (0.3)	1.1×10^{-7}
Cereal intake	7.1 (0.3)	2.9×10^{-167}	1.0 (0.2)	2.3×10^{-6}
Tea intake	7.0 (0.3)	3.0×10^{-160}	1.2 (0.3)	5.0×10^{-7}
Alcohol intake frequency	8.7 (0.3)	2.1×10^{-256}	2.5 (0.3)	3.9×10^{-27}
Oily fish intake frequency	7.1 (0.3)	6.3×10^{-180}	1.4 (0.2)	2.1×10^{-10}
Poultry intake frequency	3.0 (0.3)	1.8×10^{-33}	1.4 (0.2)	6.2×10^{-11}
Beef intake frequency	3.8 (0.3)	2.5×10^{-53}	1.5 (0.2)	1.8×10^{-12}
Lamb/mutton intake frequency	4.2 (0.3)	7.5×10^{-62}	1.5 (0.3)	2.5×10^{-11}
Pork intake frequency	2.9 (0.3)	1.2×10^{-30}	1.1 (0.2)	4.0×10^{-6}
Time spent watching television	14.7 (0.3)	$<1.0\times10^{-300}$	5.7 (0.3)	3.1×10^{-123}
Smoking status	10.6 (0.3)	$<1.0\times10^{-300}$	2.5 (0.3)	2.6×10^{-26}
Number of siblings	4.7 (0.3)	3.3×10^{-64}	1.3 (0.3)	1.2×10^{-9}
White blood cell count	23.2 (0.4)	$<1.0\times10^{-300}$	0.9 (0.2)	2.7×10^{-5}
Neutrophil count	20.2 (0.4)	$<1.0\times10^{-300}$	0.9 (0.2)	1.0×10^{-4}
Number of treatments taken	8.6 (0.3)	6.2×10^{-252}	0.9 (0.2)	4.5×10^{-5}
Self-reported astigmatism[a]	10.8 (1.8)	2.9×10^{-10}	6.3 (1.8)	1.0×10^{-4}
Educational attainment	15.2 (0.3)	$<1.0\times10^{-300}$	7.3 (0.3)	1.4×10^{-197}
Mood swings[a]	12.3 (0.5)	4.7×10^{-185}	1.5 (0.4)	8.8×10^{-5}
Fed-up feelings[a]	11.7 (0.5)	4.8×10^{-159}	1.4 (0.4)	3.5×10^{-4}
Non-sex-specific trait (significant at FDR < 0.05)	h^2_{Own} (s.e)	*P*	$h^2_{Partner}$ (s.e)	*P*
Skin colour	23.5 (0.4)	$<1.0\times10^{-300}$	0.5 (0.2)	9.0×10^{-3}
Bread intake	5.7 (0.3)	3.2×10^{-104}	0.5 (0.2)	2.5×10^{-2}
Coffee intake	5.9 (0.3)	9.3×10^{-128}	0.5 (0.2)	1.2×10^{-2}
Water intake	8.5 (0.3)	7.8×10^{-215}	0.5 (0.2)	2.5×10^{-2}
Non-oily fish intake frequency	3.3 (0.3)	1.8×10^{-43}	0.6 (0.2)	2.9×10^{-3}
Processed meat intake	4.9 (0.3)	2.9×10^{-84}	0.6 (0.2)	4.2×10^{-3}
Sleep duration	8.2 (0.3)	2.4×10^{-207}	0.4 (0.2)	1.6×10^{-2}
Lymphocyte count	24.7 (0.4)	$<1.0\times10^{-300}$	0.7 (0.2)	3.5×10^{-3}
Monocyte count	27.3 (0.4)	$<1.0\times10^{-300}$	0.5 (0.2)	1.5×10^{-2}
Reticulocyte count	24.3 (0.4)	$<1.0\times10^{-300}$	0.5 (0.2)	5.6×10^{-3}
High-light-scatter reticulocytes	24.3 (0.4)	$<1.0\times10^{-300}$	0.6 (0.2)	3.5×10^{-3}
Number of non-cancer illnesses	8.4 (0.3)	2.9×10^{-230}	0.6 (0.2)	3.0×10^{-3}
Self-reported hypertension[a]	27.8 (0.6)	$<1.0\times10^{-300}$	0.8 (0.4)	7.2×10^{-3}
Arthrosis[a]	12.6 (0.8)	2.2×10^{-76}	1.6 (0.6)	5.4×10^{-3}
Substance use disorder[a]	9.7 (2.0)	2.6×10^{-7}	5.1 (1.9)	2.4×10^{-3}
Irritability[a]	12.7 (0.6)	6.8×10^{-148}	1.0 (0.4)	1.0×10^{-2}
Worrier/anxious feelings[a]	13.4 (0.5)	1.9×10^{-215}	0.9 (0.4)	9.4×10^{-3}
Tense/'highly strung'[a]	13.5 (0.7)	3.0×10^{-110}	1.5 (0.6)	4.2×10^{-3}

Continued

Table 1 | GREML results (in %) for significant non-sex-specific traits and couple-shared traits (Continued)

Trait	Heritability and *P* values			
Non-sex-specific trait (significant at $P < 0.05/91$)	**h^2_{Own} (s.e.)**	***P***	**$h^2_{Partner}$ (s.e.)**	***P***
Loneliness[a]	12.6 (0.9)	1.8×10^{-54}	1.6 (0.7)	5.8×10^{-3}
Non-sex-specific trait (significant at $P < 0.05$)	**h^2_{Own} (s.e)**	***P***	**$h^2_{Partner}$ (s.e)**	***P***
Pack years of smoking	10.3 (2.1)	2.8×10^{-7}	2.9 (1.9)	4.9×10^{-2}
Parents have bowel cancer	1.8 (0.3)	2.1×10^{-11}	0.5 (0.3)	5.0×10^{-2}
Metabolic disorders[a]	10.3 (0.7)	1.6×10^{-56}	1.1 (0.6)	4.9×10^{-2}
Self-reported heart problem[a]	9.6 (0.9)	1.2×10^{-30}	1.2 (0.8)	4.0×10^{-2}
Couple-shared trait	**h^2_{Male} (s.e)**	***P***	**h^2_{Female} (s.e)**	***P***
Number of children	2.6 (0.6)	9.0×10^{-6}	5.5 (0.6)	1.1×10^{-20}
Townsend deprivation index	2.7 (0.5)	4.1×10^{-8}	1.9 (0.5)	5.3×10^{-5}
Number of vehicles	3.7 (0.5)	1.1×10^{-12}	2.5 (0.5)	2.8×10^{-7}

[a]For binary traits, liability-scale heritability is shown.

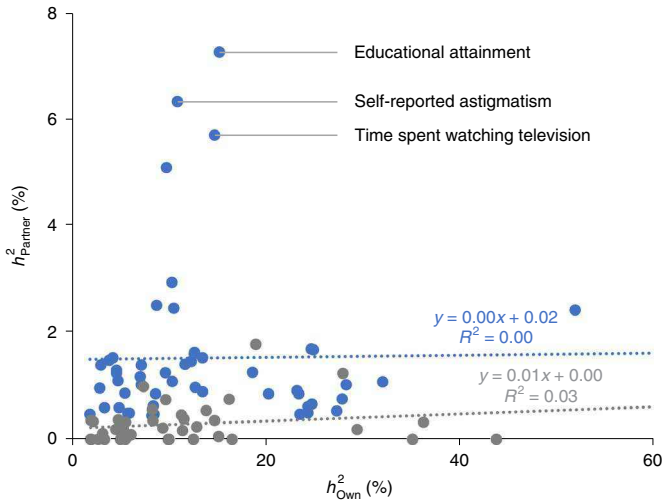

Fig. 1 | Relationship between direct and indirect genetic effects. Estimates of direct (h^2_{Own}) and indirect ($h^2_{Partner}$) genetic effects in 80,889 couples for 91 traits expressed in females and males. Each dot represents a trait. The blue dots represent traits with a significant estimate of $h^2_{Partner}$ and the blue dotted line shows the associated linear trend line. The grey dots represent traits with a nonsignificant estimate of $h^2_{Partner}$ and the grey dotted line shows the associated linear trend line. Significance level, $\alpha = 0.05$. Liability-scale heritability is shown for binary traits.

working with as many traits as possible in downstream analyses reduces the possibility of generalizing the conclusions drawn from a few traits that happen to be significant because they have slightly larger effects. For completeness, all 51 marginally significant traits were summarized and followed up in further analysis. Unless specified, a marginal significance level of $P < 0.05$ was applied.

On average, the partner genotype explained 1.5% of the phenotypic variance among those 51 traits (Supplementary Fig. 1). The variance explained by the partner genotype mainly came from the environment of the individual (that is, from the environmental residual variance), whereas the estimate of individuals' own genetic variance was largely unchanged compared to the base model containing only the person's genotype (Supplementary Fig. 2). Although the amount of variance captured by the partner genotype was small on average, for moderately heritable traits ($h^2_{Own} < 0.15$) it was comparatively large (Supplementary Fig. 1). For instance, the effect of the partner genotype was half that of the individual's own genotype for the trait 'mental and behavioural disorder due to psychoactive substance' and between 10 and 50% of the h^2_{Own} for dietary traits, mental health and educational attainment (column K, Supplementary Table 3). We also found that, generally, including the partner's genotype in a prediction model (Methods) improved prediction accuracy (for example, 4% increase for educational attainment; $P = 8.25\times10^{-3}$). For non-binary traits, the mean increase in prediction accuracy was 0.004 ± 0.001 and was significantly different from zero (two-sided paired *t*-test, d.f. = 38, $P = 6.89\times10^{-3}$; column C versus F, Supplementary Table 4).

We then tested whether the partner effects are different between sexes. For the 51 traits under consideration, we conducted sex-stratified analyses (that is, we estimated the partner indirect heritability of female's genotypes on their male partners ($h^2_{Partner(Female\rightarrow Male)}$) in male-stratified analysis and vice versa in female-stratified analysis ($h^2_{Partner(Male\rightarrow Female)}$); Methods). For binary traits, heritability estimates on the observed scale were transformed to the liability scale[14,15] using sex-specific prevalences in both the UK Biobank and in the couple subset (Supplementary Table 2). We found that $h^2_{Partner(Female\rightarrow Male)}$ was significantly larger than $h^2_{Partner(Male\rightarrow Female)}$ for six traits at a marginal level ($P < 0.05$; Fig. 2a and column W, Supplementary Table 5). These traits are mainly dietary- and obesity-related traits such as cereal intake ($P = 3.89\times10^{-3}$), beef intake frequency ($P = 9.45\times10^{-3}$) or waist-to-hip ratio ($P = 5.53\times10^{-5}$). On average, the $h^2_{Partner(Male\rightarrow Female)}$ for dietary traits was only 40% of that of $h^2_{Partner(Female\rightarrow Male)}$ (Fig. 2b). We further looked at partner effects of 11 sex-specific diseases related to the genitourinary system (Supplementary Table 1 and Methods), but found no evidence of contributions of the partner genotype to these sex-specific diseases (rows 56–66, Supplementary Table 5).

Finally, we selected three traits for which the phenotype is the same for both partners (that is, couple-shared traits). We modelled the couples' number of children, number of vehicles in household and socioeconomic status (Townsend deprivation index) as a function of the female and male genotype (that is, estimated h^2_{Male} and h^2_{Female} jointly; Methods). This can be interpreted as the direct effects of the female and male genotypes to a shared trait. Both h^2_{Male} and h^2_{Female} were significantly greater than zero for all three traits (Table 1). Furthermore, for number of children, estimates of h^2_{Male} ($2.60 \pm 0.61\%$; $P = 8.95\times10^{-6}$) and h^2_{Female} ($5.50 \pm 0.64\%$; $P = 1.14\times10^{-20}$) were significantly different from each other (two-sided *z*-test, $P = 1.05\times10^{-3}$). This suggests that both female and male biological and social factors, mediated

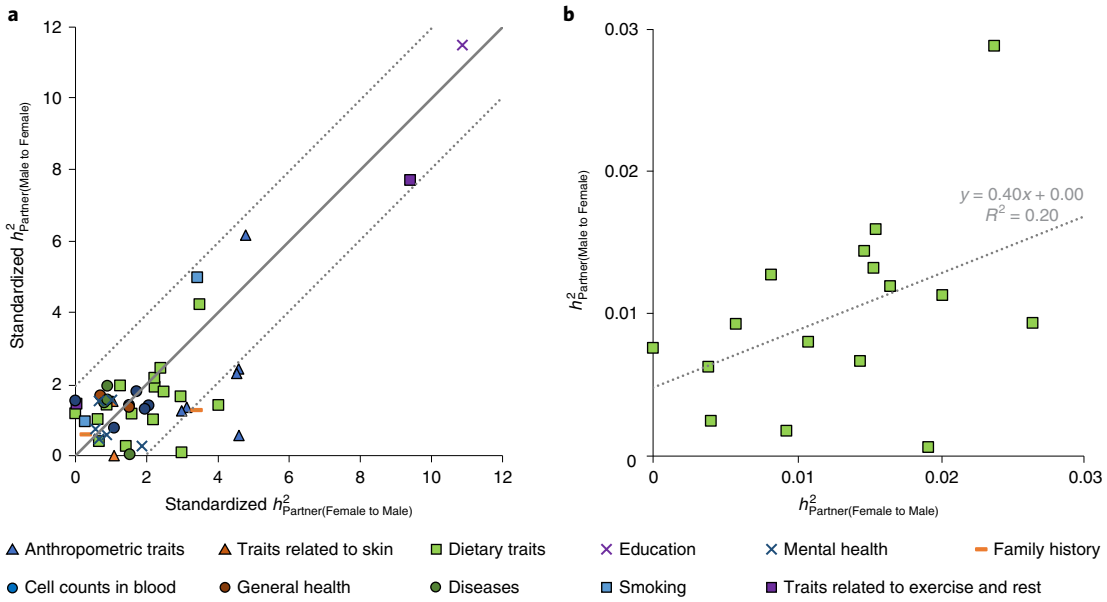

Fig. 2 | Difference in IGEs between sexes. a,**b**, Sex-specific IGEs in 80,889 couples for 51 traits with significant partner indirect heritability ($P<0.05$). The effect of males on their female partners ($h^2_{Partner(Male\rightarrow Female)}$) is shown on the *y* axis; $h^2_{Partner(Female\rightarrow Male)}$ is shown on the *x* axis. In **a**, the lines show the diagonal (solid line) and its 95% confidence interval (dotted lines). For comparison, we standardized the partner heritability by dividing by the pooled standard error (that is, $\sqrt{s.e._1^2 + s.e._2^2}$). Hence, dots above the top dotted line indicate that $h^2_{Partner(Male\rightarrow Female)}$ is significantly larger than $h^2_{Partner(Female\rightarrow Male)}$ for that trait at $P<0.05$. Similarly, dots below the bottom dotted line indicate the opposite. In **b**, as an example we took dietary traits and regressed $h^2_{Partner(Male\rightarrow Female)}$ on $h^2_{Partner(Female\rightarrow Male)}$ (grey dotted line) and found that the regression coefficient is 0.40 (±0.22), significantly lower than 1 at $P=0.006$.

through a combination of direct and indirect genetic effects, affect the number of children a couple has.

IGEs under assortative mating. Owing to the inherent limitations associated with the cross-sectional nature of the data, direct assortative mating (that is, assortment based on the trait of interest) and indirect assortative mating (that is, assortment based on traits other than the trait of interest) can be important confounders when testing for the presence of IGEs. It is therefore important to assess to what degree the observed $h^2_{Partner}$ could be explained by a model that does not need to invoke the existence of IGEs. To address this question, we carried out three further analyses for the 51 traits with marginally significant $h^2_{Partner}$ estimates ($P<0.05$). First, we assessed the fit of the data to theoretical predictions under a pure model of assortative mating where $h^2_{Partner}$ is generated exclusively through direct assortment. Second, we used permutations to estimate the expected $h^2_{Partner}$ in the European couples under the same strength of assortment (r_{CP}; Supplementary Table 6). Third, we examined whether the IGEs learned in our European couples also explain variation in couples of other ancestries and siblings. We then performed a meta-analysis combing all of these results. A marginal significance threshold ($P<0.05$) was used for the first three follow-up analyses, and Bonferroni correction ($P<0.05/51=0.001$) was applied for the final meta-analysis.

Theoretical expectation of the partner heritability under direct assortative mating. Under a model where $h^2_{Partner}$ is generated exclusively through direct assortative mating on the focal trait, the expectation of $h^2_{Partner}$ is $r^2_{CP}h^2_{Own}$ (ref. [16] and Supplementary Methods). Inflated values of $h^2_{Partner}$ compared to the theoretical expectation would imply that the partner genotypes explain more variation than expected under direct assortative mating. Among the 51 traits under consideration, we found 29 that deviated from this theoretical expectation at a marginal level ($P<0.05$; Fig. 3a and Supplementary Table 7). Height, for which $h^2_{Partner}$ is probably driven mainly by direct assortative mating, closely follows the theoretical expectation. The most differentiated traits were educational attainment ($P=1.84\times10^{-85}$), time spent watching television ($P=4.48\times10^{-30}$) and smoking status ($P=6.16\times10^{-13}$). The other 26 traits were mainly mental health traits, diseases and dietary traits (Fig. 3a and Supplementary Table 7).

Extensive simulations under different models (direct and indirect assortment, IGEs and their combinations) and mathematically derived expectations suggest that the patterns observed in Fig. 3a are better explained by models that involve both direct assortative mating and IGEs that are mediated through a trait that is not strongly correlated (for example, the phenotypic correlation between two traits <0.1) with the focal trait of assortment (Supplementary Results). Simulations under realistic spectra of parameters (heritability: 0.05 to 0.5; strength of assortment: −0.05 to 0.5; genetic and phenotypic correlation between two traits: −0.6 to 0.9) showed that neither a stand-alone direct assortment or indirect assortment model could recapitulate the results from the UK Biobank data (Supplementary Results). To recapitulate similar results observed in the UK Biobank data, we had to add IGEs to the simulation model.

Partner heritability in randomized couples with similar strength of assortment. If the estimated partner indirect heritability were due exclusively to direct assortative mating on the focal trait, then randomizing couples while keeping similar r_{CP} for that trait would recapitulate the estimates of $h^2_{Partner}$ we observed (Supplementary Fig. 3). For the 51 traits under consideration, we permuted couples on the basis of the r_{CP} of each trait (Supplementary Table 6) and re-estimated $h^2_{Partner}$ (Methods). We found that the within-trait difference in $h^2_{Partner}$ estimates was marginally significant for 14 traits ($P<0.05$; column H, Supplementary Table 8), the most differentiated traits being time spent watching television ($P=1.36\times10^{-9}$), smoking status ($P=6.25\times10^{-9}$) and educational attainment

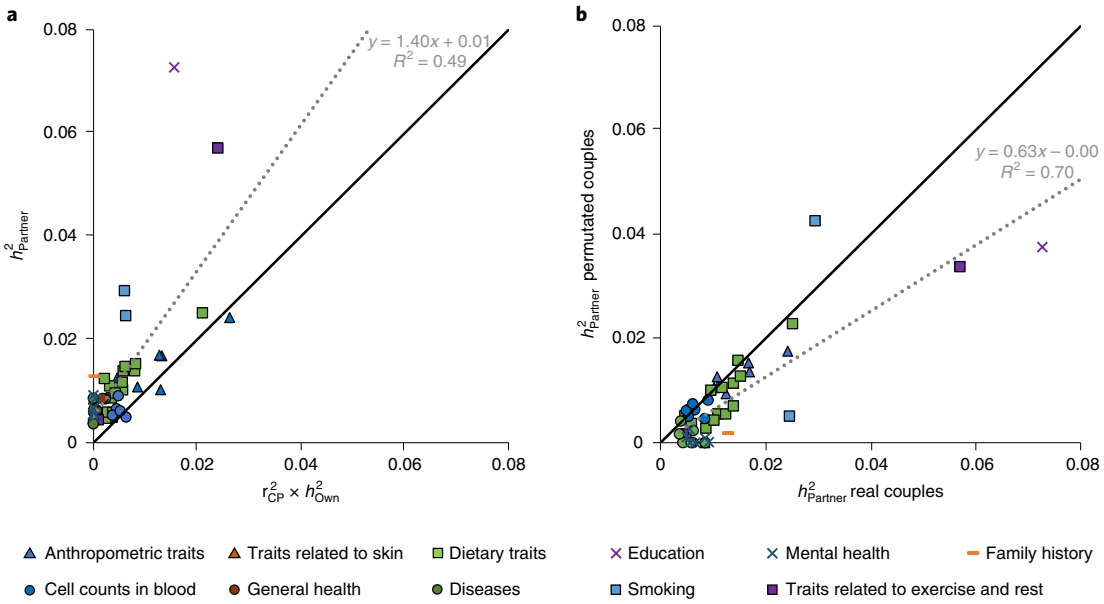

Fig. 3 | Evidence of IGEs over the expectation under assortative mating. a,**b**, Plots showing that the partner indirect heritability is capturing IGEs above that expected under assortative mating. The black line represents $y = x$. The grey dotted line represents a fitted regression line for all traits. In **a**, we compared the estimate of partner indirect heritability in 80,889 couples (h^2_{Partner}, y axis) with its expected value under direct assortative mating theory ($r^2_{\text{CP}} \times h^2_{\text{Own}}$, x axis). Deviation from the diagonal line indicates that the data do not fit a simple model of direct assortative mating. In **b**, we permutated our 80,889 couples by replacing both partners with other individuals with similar phenotypes to keep the same size of couple correlation for that trait. We compared the estimates of partner indirect heritability of the permutation (y axis) with those of real couples (x axis). We found that, on average, the estimate of partner indirect heritability is 37% smaller for the permutated couples.

($P = 2.06 \times 10^{-19}$). On average across the 51 traits, the h^2_{Partner} was ~40% smaller in the randomized couples than in the real couples (Fig. 3b) and the mean decrease in h^2_{Partner} was $0.45 \pm 0.10\%$, which was significantly different from zero (two-sided paired t-test, d.f. = 50, $P = 2.16 \times 10^{-5}$; column B versus E, Supplementary Table 8). This confirms the previous theoretical results and suggests again that a model that involves only direct assortment on the focal trait cannot explain our data and that our partner heritability estimates in real couples capture IGEs. This permutation is biased towards discovering cases where the focal traits are mediated through an uncorrelated trait. When the mediator of IGEs is correlated with the focal trait (for example, the same trait), the permutation is unable to break down the correlation structure made up by assortment and IGEs, thus making it hard to detect those traits and to separate them from direct assortment (Supplementary Fig. 3). As a consequence, the permutation results are probably conservative and underestimate the number of traits affected by IGEs.

Furthermore, to test whether direct assortment on height and educational attainment has an influence on other traits (that is, indirect assortative mating), we permuted couples on the basis of the phenotype of either educational attainment or height, two well-known traits under direct assortment[17,18], and re-estimated h^2_{Partner} for the other 50 traits. Permutating couples on the basis of educational attainment, the average h^2_{Partner} of the other 50 traits was only one-sixth of their original values observed from real couples (Supplementary Fig. 4). The estimates of h^2_{Partner} were significantly greater than zero at a marginal level of $P < 0.05$ only for four traits (columns I–K, Supplementary Table 8). The most significant one was time spent watching television ($h^2_{\text{Partner}} = 0.72 \pm 0.23\%$; $P = 2.85 \times 10^{-4}$), which is much less significant and about eight times smaller than that from real couples ($h^2_{\text{Partner}} = 5.71 \pm 0.28\%$; $P = 3.05 \times 10^{-123}$). Regarding height-based permutation, the average h^2_{Partner} is close to zero and the individual estimates are not related to their original values obtained in real couples (Supplementary Fig. 4 and columns M–O, Supplementary Table 8). Therefore, indirect assortative mating on these traits due to direct assortment on height or educational attainment alone is unlikely to explain the observed h^2_{Partner} for the majority of the considered traits.

Replication in couples and evidence of IGEs in siblings. For the 51 traits under consideration, we estimated best linear unbiased predictions (BLUPs) of the marker effects for both the direct and indirect genetic effects using DISSECT[19] and 80,889 couples. These BLUPs were used to predict direct and indirect polygenic scores (PGSs) of another 3,752 independent couples and 8,144 independent sibling pairs (Methods).

In the couple replication set, we first examined whether one's indirect PGS can explain phenotypic variations in his/her partner's phenotype using linear models where the partner's direct PGS was fitted jointly (Methods). We found evidence that the partner indirect PGS was marginally significantly associated with the phenotype of 19 out 51 traits ($P < 0.05$; column D, Supplementary Table 9). To further explore whether IGEs contributed to the association, we re-examined the association after adjusting both the phenotype and the indirect PGS for assortment (Methods). By doing so, three traits (including height) were no longer significant (column J, Supplementary Table 9). The most significant signals retained belong to time spent watching television ($P = 3.22 \times 10^{-8}$) and educational attainment ($P = 4.58 \times 10^{-6}$). As in the permutation analysis, this analysis may miss cases where a trait is mediated by IGEs through a correlated trait because of overcorrection.

We then applied the same tests in the sibling pairs to examine whether one's phenotype was associated with his/her siblings' couple-estimated indirect PGS in models where the individual's own direct PGS was fitted jointly (Methods). Since siblings do not choose each other, an association of the couple-estimated indirect PGS in sibling pairs would be consistent with the presence of IGEs for these traits not only among couples but also in siblings (that is,

IGEs are not restricted within couples). We found evidence that the couple-estimated indirect PGS was marginally significantly associated with the phenotype of 8 out of 51 traits at $P<0.05$ (column V, Supplementary Table 9), including 4 traits (body fat percentage, time spent watching television, high-light-scatter reticulocyte count and educational attainment) previously found in the couple replication set. It is possible that these associations arise from vertical IGEs in parents captured by proxy (for example, IGEs on number of siblings detected in the sibling may be caused by IGEs on number of children in the parent) or from horizontal IGEs from parents to offspring captured by proxy (for example, vertical IGEs on educational attainment).

However, when we repeated the above couple and sibling analyses using the BLUPs learned from the randomized couples previously used in permutations, most signals disappeared (column J versus M, column V versus Y, Supplementary Table 9). These results suggest that the indirect PGS learned from real couples were picking up IGEs whereas IGEs were largely removed in randomized couples after permutation. This further confirms our previous conclusion that IGEs contribute to our observed partner IGEs in real couples.

Evidence of IGEs under assortative mating across follow-up analyses. The follow-up analyses presented above used different methods in different samples and had different weaknesses and strengths to discover IGEs in different settings. To examine whether there is consistent evidence of IGEs for a trait across all of these analyses, we performed a meta-analysis (Fisher's method[20]) using the P values from the following four tests: whether the estimate of h^2_{Partner} in real couples deviates from its theoretical expectation (column G, Supplementary Table 7); whether the estimate of h^2_{Partner} in real couples is different from that in permuted couples (column H, Supplementary Table 8); and whether there is evidence of IGEs on top of assortment in mixed couples and in sibling pairs (columns J and V, Supplementary Table 9). In total, we have consistent evidence of IGEs for 13 out of the 51 traits at $P<0.05/51=0.001$ (Supplementary Table 10). These 13 traits include body fat percentage ($P=8.66\times10^{-5}$), ease of skin tanning ($P=1.61\times10^{-7}$), dried fruit intake ($P=3.01\times10^{-5}$), oily fish intake frequency ($P=7.53\times10^{-4}$), beef intake frequency ($P=2.11\times10^{-4}$), lamb/mutton intake frequency ($P=8.77\times10^{-4}$), time spent watching television ($P=1.33\times10^{-42}$), smoking status ($P=2.77\times10^{-17}$), number of siblings ($P=4.68\times10^{-8}$), self-reported astigmatism ($P=5.55\times10^{-4}$), educational attainment ($P=1.75\times10^{-109}$), mood swings ($P=5.36\times10^{-7}$) and fed-up feelings ($P=4.84\times10^{-5}$).

Discussion

This study has investigated IGEs in 105 traits representing a wide range of anthropometric, behavioural and disease traits. We find that IGEs are common in humans and present among non-blood and blood relatives. Extensive simulations under realistic parameter spectra and permutations tests consistently found that the data cannot be explained by direct or indirect assortative mating alone and that the trait mediating IGEs is unlikely to be the same trait as the focal trait. The last argument is also supported by a recent genome-wide association study in mice, where researchers conducted a genome-wide association study of 170 phenotypes to identify loci contributing to direct and indirect genetic effects and found no overlap between direct and indirect genetic loci within the same trait[12]. This is also consistent with the data from Kong et al.[10] and Cheesman et al.[21] where Kong et al. found evidence of IGEs for height, mediated through direct genetic effects for educational attainment from their parents, while Cheesman et al. found no evidence of IGEs for height mediated through direct genetic effects for height. In our study, we found evidence of partner heritability for 51 of the traits we investigated, 13 of them showing consistent evidence across the follow-up analyses that IGEs contributed to the observed estimates (Supplementary Table 10). These mainly consisted of traits related to behaviour, socioeconomics and personality such as dietary traits, watching television, smoking, educational attainment and mood swings. A previous study of horizontal IGEs among schoolmates and friends found IGEs for educational attainment[11]. We also detected IGEs for educational attainment in our couple data, which provides additional evidence that horizontal IGEs for educational attainment are not restricted to schoolmates.

Three big challenges lie ahead. The biggest challenge is to disentangle IGEs from assortative mating. Assortment also creates genetic–phenotypic association between partners, resulting in false-discovery of IGEs. We have implemented several different methods to explore whether there is evidence of IGEs on top of assortment. However, most methods favour discovering cases where the trait mediating IGEs is different from the focal trait. When the trait mediating IGEs and the focal trait under assortment are the same, it is hard to separate them given current data. For this, it will be essential to have large longitudinal studies that recruit participants earlier in life, when mating choice most often happens, and follow participants until later in life. However, it implies that we might be underestimating the number of traits affected by IGEs in couples among those studied here. Although our estimates of IGEs may potentially be affected by a covariance among couples[1], half of the traits still had significant IGE signals after fitting a random shared environment in the model (Supplementary Fig. 5 and Supplementary Table 11). This model is likely to be conservative. The second challenge will be to identify the genetic variants that mediate the IGEs to better understand whether these are the same variants that directly affect the phenotype or not. These studies will also require much larger cohorts than currently available as h^2_{Partner} was generally smaller than h^2_{Own}. Since there is a correlation between the number of identified loci and h^2_{Own} for a fixed sample size[22], indirect SNP effects will probably be more difficult to detect than direct SNP effects. The final challenge will be to understand how IGEs mediate the focal phenotype. We hypothesize that many of the horizontal IGEs we found will be mediated through behavioural changes; this again will require large longitudinal studies with a variety of blood and non-blood relatives. Adopted relatives could potentially help in the study of IGEs as they are not as affected by the issues introduced by assortative mating and relatedness.

We have provided substantial evidence suggesting that a proportion of the environmental component of complex traits is determined by the genotype of social partners, and this might call for new ways of studying the environment that mediates human disease. Our results open many new questions and opportunities that will require further investigation.

Methods

Ethical compliance. The UK Biobank project was approved by the National Research Ethics Service Committee North West-Haydock (REC reference: 11/NW/0382). An electronic signed consent was obtained from the participants.

UK Biobank cohort. Our study uses the UK Biobank cohort, a prospective population study with over 500,000 adult (age at recruitment ≥ 37 years) participants[13].

Genotype and sample. Participants of UK Biobank were genotyped for more than 800,000 genome-wide genetic markers using either the Applied Biosystems UK BiLEVE Axiom or Applied Biosystems UK Biobank Axiom arrays, two similar arrays with over 95% content in common. Details of the array design, blood sample collection, DNA extraction and the initial data quality control have been described comprehensively elsewhere[22,23]. Our discovery study was restricted to a subset of 333,852 autosomal bi-allelic genetic markers in 80,889 heterosexual couples of European ancestry. The retained genetic markers comprised 331,898 SNPs and 1,954 indels, are common (minor allele frequency > 5%), have a call rate >98% and are in Hardy–Weinberg equilibrium (exact-test P value > 10^{-50}) in the subset of 80,889 couples. All retained individuals have a genotype missing rate of <5%. European participants were defined as self-reported European who were within five standard deviations from the mean of the first two principal components of the genomic relationship matrix of all self-reported European participants. Couple

relationships within the UK Biobank cohort were identified following the method described in Tenesa et al.[16]. A total of 138 couple pairs with a genomic coefficient of relationship above 0.025 and 7 related couples (for example, two couples made up of two relatives from one family and two relatives from another family) were removed from analysis, resulting in a final set of 80,889 couples for discovery. In addition, a total of 8,144 pairs of European siblings and 3,752 mixed couples (1 European and 1 non-European) not used in the discovery set were used for follow-up analyses. These samples had a genotype missing rate of less than 5%. The same set of markers were extracted for them with no further quality control performed. Full-sibling relationships were identified as a pair of individuals with IBS0 (proportion of genetic markers for which siblings share zero alleles) > 0.001 and pairwise kinship in the range 0.177–0.354 (ref. [23]). Only sibling pairs with an age difference of less than 15 years were analysed.

Phenotype. We analysed 105 phenotypes in this study, including diseases, continuous traits, a few sex-specific phenotypes and phenotypes shared between partners (for example, number of children). For couple-shared phenotypes, we removed from 1,365 to 16,060 couples that had different phenotypic values and assessment centres. For continuous phenotypes and integer phenotypes with ten or more unique values, outliers (identified as values more than three standard deviations from the mean after correcting for sex, age and assessment centre) were removed. For diseases, a minimum of 1,000 cases were required. Additionally, except for sex-specific phenotypes, we kept only couples where both partners have non-missing values for a given phenotype. A summary table of the phenotypic data and quality control procedures is available in Supplementary Table 1. The same set of phenotypes were extracted for sibling and mixed-couple datasets with no further quality control performed.

Genomic restricted maximum-likelihood analysis. We estimated the proportion of the phenotypic variation that can be explained by an individual's own and by their partner's genotype, respectively. To this end, we performed genomic restricted maximum-likelihood (GREML) analysis in DISSECT using mixed linear models with two genomic variance components[19].

Model. The main model we employed takes the form

$$\mathbf{y} = X\beta + \mathbf{g}_{\mathrm{o}} + \mathbf{g}_{\mathrm{p}} + \boldsymbol{\varepsilon}$$

Here $\mathbf{y}$ is the vector of phenotypes and X is the design matrix of the fixed effects β. The terms $\mathbf{g}_{\mathrm{o}}$, $\mathbf{g}_{\mathrm{p}}$ and $\boldsymbol{\varepsilon}$ are random effects. Specifically, $\mathbf{g}_{\mathrm{o}}$ is the effect of the individual's own genotype, $\mathbf{g}_{\mathrm{p}}$ is the effect of the partner's genotype (partner IGEs) and $\boldsymbol{\varepsilon}$ is a residual effect that accounts for any effects that are not accounted for by the model. These random effects follow $\mathbf{g}_{\mathrm{o}} \sim N(0, GRM_{\mathrm{own}}\sigma^2_{\mathrm{o}})$, $\mathbf{g}_{\mathrm{p}} \sim N(0, GRM_{\mathrm{partner}}\sigma^2_{\mathrm{p}})$ and $\boldsymbol{\varepsilon} \sim N(0, I\sigma^2_{\boldsymbol{\varepsilon}})$, where σ^2_{o}, σ^2_{p} and $\sigma^2_{\boldsymbol{\varepsilon}}$ are the variance component parameters and I is the identity matrix. The covariance matrices $\mathrm{GRM}_{\mathrm{own}}$ and $\mathrm{GRM}_{\mathrm{partner}}$ are genomic relationship matrices computed using an individual and their partner's genotypes, respectively.

Covariates. We fitted sex, genotyping batch, assessment centre, age, age-squared and the first 20 genomic principal components of the focal individuals as fixed effects in the model. The genotyping array was not included as a covariate since it is accounted for by the genotyping batch. The latter is nested within the former.

Model interpretation. We refer to the proportion of phenotypic variance explained by individuals' own genotype as 'own heritability', $h^2_{\mathrm{Own}} = \frac{\sigma^2_{\mathrm{o}}}{\sigma^2_{\mathrm{o}}+\sigma^2_{\mathrm{p}}+\sigma^2_{\boldsymbol{\varepsilon}}}$ (which is also known as SNP or GREML heritability[24]), and the proportion of phenotypic variance explained by the genotype of the partner as the 'partner indirect heritability', $h^2_{\mathrm{Partner}} = \frac{\sigma^2_{\mathrm{p}}}{\sigma^2_{\mathrm{o}}+\sigma^2_{\mathrm{p}}+\sigma^2_{\boldsymbol{\varepsilon}}}$.

Statistical test. For each trait, we tested against the two null hypotheses $\sigma^2_{\mathrm{o}} = 0$ and $\sigma^2_{\mathrm{p}} = 0$ by performing likelihood ratio tests against the appropriate reduced model. We used a two-component mixture, a delta spike at 0 and $\chi^2_{\mathrm{d.f.}=1}$, each component with a probability of 0.5 as the null distribution of the test statistic[25] using DISSECT[19].

Sex-specific analysis. For sex-specific phenotypes, we fitted sex-specific models (that is, models including only phenotypes of the individuals of the appropriate sex). We also fitted sex-stratified models (that is, separate models including only male or female phenotypes) for non-sex-specific phenotypes with estimates of h^2_{Partner} significantly greater than 0 ($P < 0.05$). Since these models were sex-specific, sex was not included as a covariate. All other covariates were retained. Note, in male-specific/stratified analysis, we redefined male's direct effects on themselves as $h^2_{\mathrm{Own(Male)}}$ and the female indirect effect on the male as $h^2_{\mathrm{Partner(Female\rightarrow Male)}}$. Accordingly, their counterparts in female-specific/stratified analysis were defined as $h^2_{\mathrm{Own(Female)}}$ and $h^2_{\mathrm{Partner(Male\rightarrow Female)}}$.

Couple-shared phenotype. We defined h^2_{Male} and h^2_{Female} meaning the SNP heritability contributed by male genotype and female genotype, respectively. We included covariates of both sexes in the model. The exceptions were sex (which was not included as a covariate) and assessment centre, which is shared between couples and hence was fitted only once.

Prediction and replication. *Prediction of own and partner IGEs.* We learned the marker effects in 80,889 European couples and predicted the direct PGS ($\widehat{\mathbf{g}_{\boldsymbol{o}}}$) and indirect PGS ($\widehat{\mathbf{g}_{\boldsymbol{p}}}$) of sibling and mixed-couple samples in DISSECT[19].

Non-binary traits. For non-binary traits, the prediction accuracy was defined as the Pearson correlation between the phenotype and the predictions. We compared $\mathrm{Cor}(\mathbf{y}, \widehat{\mathbf{g}_{\mathrm{o}}})$ with $\mathrm{Cor}(\mathbf{y}, \widehat{\mathbf{g}_{\mathrm{o}}} + \widehat{\mathbf{g}_{\mathrm{p}}})$ to see whether including partner IGEs could increase the correlation, where $\mathbf{y}$ is the phenotype. Standard errors of correlations were calculated as $\sqrt{\frac{1-r^2}{N-2}}$, where r is the estimate of correlation and N is the sample size.

Binary traits. For binary traits, the prediction accuracy was defined as the area under the receiver operating characteristic curve (AUC), which was calculated in R using the ROCR package[26]. We compared the AUC of a model including only $\widehat{\mathbf{g}_{\mathrm{o}}}$ with that of a model that also included $\widehat{\mathbf{g}_{\mathrm{p}}}$ to see whether including partner indirect effects could increase the AUC. The standard error of the AUC was calculated following ref. [27].

Couple and sibling replication analysis. We estimated whether the IGEs learned from our European couples could explain phenotypic variation in other datasets (3,752 mixed couples and 8,144 European sibling pairs). We tested whether $\beta_{\mathrm{p}} \neq 0$ in the following linear regression model:

$$\mathbf{y} = X\beta + \beta_{\mathrm{o}}\mathbf{g}_{\mathrm{o}} + \beta_{\mathrm{p}}\mathbf{g}_{\mathrm{p}} + \boldsymbol{\varepsilon}.$$

Here $\mathbf{y}$ is the vector of own phenotypes, X is the design matrix including all own covariates used in the discovery model (sex, age and so on), $\mathbf{g}_{\mathrm{o}}$ is the vector of own direct genetic effects (that is, own $\widehat{\mathbf{g}_{\mathrm{o}}}$), $\mathbf{g}_{\mathrm{p}}$ is the vector of partner IGEs (that is, partner's $\widehat{\mathbf{g}_{\mathrm{p}}}$), $\boldsymbol{\varepsilon}$ is the vector of residuals and the β terms are the corresponding regression coefficients.

In a second method, we examined whether the association between the phenotype and the predicted partner IGEs observed in other datasets resulted from IGEs. To this end, we removed the effects of assortative mating out of both sides of the equation. This was accomplished in three steps. We first performed the following regression:

$$\mathbf{y} = X\beta + \beta_{\mathbf{y}_{\mathrm{p}}}\mathbf{y}_{\mathrm{p}} + \beta_{\mathrm{o}}\mathbf{g}_{\mathrm{o}} + \beta_{\mathrm{o}_{\mathrm{p}}}\mathbf{g}_{\mathrm{o}_{\mathrm{p}}} + \boldsymbol{\varepsilon}_1.$$

Here $\mathbf{y}$ is the vector of own phenotypes, X is the design matrix including all own covariates used in the discovery model (sex, age and so on), $\mathbf{y}_{\mathrm{p}}$ is the vector of partner phenotypes, $\mathbf{g}_{\mathrm{o}}$ is the vector of own direct genetic effects (that is, own $\widehat{\mathbf{g}_{\mathrm{o}}}$), $\mathbf{g}_{\mathrm{o}_{\mathrm{p}}}$ is the vector of partner direct genetic effects (that is, partner's $\widehat{\mathbf{g}_{\mathrm{o}}}$), $\boldsymbol{\varepsilon}_1$ is the vector of residuals and the β terms are the corresponding regression coefficients.

In the second step, we performed a similar approach using the following regression model:

$$\mathbf{g}_{\mathrm{p}} = \beta'_{\mathrm{o}}\mathbf{g}_{\mathrm{o}} + \beta'_{\mathrm{o}_{\mathrm{p}}}\mathbf{g}_{\mathrm{o}_{\mathrm{p}}} + \beta'_{\mathbf{y}_{\mathrm{p}}}\mathbf{y}_{\mathrm{p}} + \boldsymbol{\varepsilon}_2.$$

Here, $\mathbf{g}_{\mathrm{p}}$ is partner IGEs (that is, partner's $\widehat{\mathbf{g}_{\mathrm{p}}}$), $\boldsymbol{\varepsilon}_2$ is the vector of residuals, the β terms are the corresponding regression coefficients and the remaining terms are the same as defined above.

In the final step, we tested whether $\beta_* \neq 0$ in the following regression

$$\boldsymbol{\varepsilon}_1 = \beta_*\boldsymbol{\varepsilon}_2 + \boldsymbol{\varepsilon}$$

Here $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ are the residuals from the first and the second step, $\boldsymbol{\varepsilon}$ is the vector of residuals of this model and β_* is the regression coefficient to be tested. A significant association supports the existence of IGEs mediated through a phenotype different to the focal phenotype. This is because we have removed correlations due to own and partner's direct genetic effects on the focal phenotype as well as partner's focal phenotype. This is a conservative approach, since it would also remove IGEs if they are mediated through the same trait.

Permutation. Two different permutation strategies were implemented, one for direct assortment and one for indirect assortment. In the first method, for each trait, we first ordered the couples on the basis of the phenotype of the female partner, with the order for couples with equal phenotypes being randomly chosen. We then took blocks of ten couples and permuted the female partners within each block, ensuring that no female was assigned to her true partner. We repeated the same process for male partners and then re-estimated h^2_{Partner} in randomized couples using our general GREML model. Such permutation could break down the original couple relationship while keeping the observed couple correlations (column F versus J, Supplementary Table 6). This permutation strategy was tested using simulated data (Supplementary Fig. 3). In the second method, we performed the above permutation strategy for both male and female partners. However, the permutation was based on the phenotype of either height or educational attainment, while the estimation of h^2_{Partner} was for other traits (that is, to explore

the impact of direct assortment on height and educational attainment on other traits).

Theoretical and simulation study. We mathematically derived the expectation of h^2_{Partner}, h^2_{Own} and r_{CP} under different scenarios, which cover direct assortment, indirect assortment, IGEs and different combinations of assortative mating and IGEs. We conducted simulation studies to check whether our derivations were accurate (Supplementary Methods and Supplementary Results).

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

Data availability

The data that support the results presented in this paper can be accessed from UK Biobank after publication. UK Biobank will link the dataset returned to the publication through their application system.

Code availability

The main results of this work were obtained using DISSECT[19], which can be freely downloaded from http://www.dissect.ed.ac.uk.

Received: 24 August 2019; Accepted: 29 September 2020;
Published online: 14 December 2020

References

1. Bijma, P. The quantitative genetics of indirect genetic effects: a selective review of modelling issues. *Heredity* **112**, 61–69 (2014).
2. Wolf, J. B., Brodie, E. D. III, Cheverud, J. M., Moore, A. J. & Wade, M. J. Evolutionary consequences of indirect genetic effects. *Trends Ecol. Evol.* **13**, 64–69 (1998).
3. Moore, A. J., Brodie, E. D. III & Wolf, J. B. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. *Evolution* **51**, 1352–1362 (1997).
4. Mousseau, T. A. & Fox, C. W. The adaptive significance of maternal effects. *Trends Ecol. Evol.* **13**, 403–407 (1998).
5. Willham, R. L. The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* **19**, 18–27 (1963).
6. Santostefano, F., Wilson, A. J., Niemelä, P. T. & Dingemanse, N. J. Indirect genetic effects: a key component of the genetic architecture of behaviour. *Sci. Rep.* **7**, 10235 (2017).
7. Brotherstone, S. et al. Competition effects in a young Sitka spruce (*Picea sitchensis*, Bong. Carr) clonal trial. *Silvae Genet.* **60**, 149–155 (2011).
8. Camerlink, I., Ursinus, W. W., Bijma, P., Kemp, B. & Bolhuis, J. E. Indirect genetic effects for growth rate in domestic pigs alter aggressive and manipulative biting behaviour. *Behav. Genet.* **45**, 117–126 (2015).
9. Warrington, N. M. et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat. Genet.* **51**, 804–814 (2019).
10. Kong, A. et al. The nature of nurture: effects of parental genotypes. *Science* **359**, 424–428 (2018).
11. Domingue, B. W. et al. The social genome of friends and schoolmates in the national longitudinal study of adolescent to adult health. *Proc. Natl Acad. Sci. USA* **115**, 702–707 (2018).
12. Baud, A., Casale, F. P., Nicod, J. & Stegle, O. Comparative architectures of direct and social genetic effects from the genome-wide association study of 170 phenotypes in laboratory mice. Preprint at *bioRxiv* https://doi.org/10.1101/302349 (2019).
13. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
14. Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* **88**, 294–305 (2011).
15. Dempster, E. R. & Lerner, I. M. Heritability of threshold characters. *Genetics* **35**, 212–236 (1950).
16. Tenesa, A., Rawlik, K., Navarro, P. & Canela-Xandri, O. Genetic determination of height-mediated mate choice. *Genome Biol.* **16**, 269 (2016).
17. Hugh-Jones, D., Verweij, K. J. H., St Pourcain, B. & Abdellaoui, A. Assortative mating on educational attainment leads to genetic spousal resemblance for polygenic scores. *Intelligence* **59**, 103–108 (2016).
18. Stulp, G., Simons, M. J. P., Grasman, S. & Pollet, T. V. Assortative mating for human height: a meta-analysis. *Am. J. Hum. Biol.* **29**, e22917 (2017).
19. Canela-Xandri, O., Law, A., Gray, A., Woolliams, J. A. & Tenesa, A. A new tool called DISSECT for analysing large genomic data sets using a big data approach. *Nat. Commun.* **6**, 10162 (2015).
20. Fisher, R. A. *Statistical Methods for Research Workers* 4th edn (Oliver and Boyd, 1932).
21. Cheesman, R. et al. Comparison of adopted and non-adopted individuals reveals gene-environment interplay for education in the UK Biobank. *Psychol. Sci.* **31**, 582–591 (2019).
22. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).
23. Bycroft, C. et al. Genome-wide genetic data on ~500,000 UK Biobank participants. Preprint at *bioRxiv* https://doi.org/10.1101/166298 (2017).
24. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
25. Visscher, P. M. A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Res. Hum. Genet.* **9**, 490–495 (2006).
26. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: visualizing classifier performance in R. *Bioinformatics* **21**, 3940–3941 (2005).
27. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).

Acknowledgements

This research has been conducted using the UK Biobank Resource under project 788. The work was funded by Roslin Institute Strategic Programme Grants from the BBSRC (BBS/E/D/10002070 and BBS/E/D/30002275) and an MRC grant (MR/P015514/1). A.T. also acknowledges funding from the Medical Research Council Human Genetic Unit and Health Data Research UK (references HDR-9004 and HDR-9003). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Analyses were performed using the ARCHER UK National Supercomputing Service.

Author contributions

C.X. contributed to the analysis. A.T. designed the study. C.X., O.C.-X., K.R. and A.T. contributed to the interpretation of data and the writing of manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-00991-9.

Correspondence and requests for materials should be addressed to A.T.

Peer review information *Nature Human Behaviour* thanks Loic Yengo, Piter Bijmaand and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Primary Handling Editor: Stavroula Kousta.

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

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

nature research

Corresponding author(s): Albert Tenesa

Last updated by author(s): Sep 25, 2020

Reporting Summary

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

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	UK biobank data showcase downloading website
Data analysis	DISSECT v1.6.1 & R v3.5.3

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

Data

Policy information about availability of data

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

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

UK biobank is free dataset accessed by application.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

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

Study description	We used mixed-linear model to estimate social genetic effects (i.e. whether an individual's partner's genotype is associated with the phenotypic variation) for 100 complex traits and diseases.
Research sample	80,899 pairs of European couples, 3,752 pairs of mixed-ethnic couples and 8,144 pairs of European siblings in UK biobank.
Sampling strategy	How we identified European ancestry individuals within UK Biobank, couples and siblings is described in other published paper and is cited accurately.
Data collection	Data collection information is available on UK biobank website.
Timing	Recruitment happened between 2006-2010, based on UK biobank website
Data exclusions	Co-habitant pairs of non-European ancestry or same-sex pairs within UK Biobank were excluded
Non-participation	11 participants were removed from the study as they withdraw the consent during the time between we received the data and we started the analysis, although none of them are in the couple and sibling subsets we used.
Randomization	NA

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	From UK biobank website "UK Biobank is a major national and international health resource, and a registered charity in its own right, with the aim of improving the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses – including cancer, heart diseases, stroke, diabetes, arthritis, osteoporosis, eye disorders, depression and forms of dementia. UK Biobank recruited 500,000 people aged between 40-69 years in 2006-2010 from across the country to take part in this project. They have undergone measures, provided blood, urine and saliva samples for future analysis, detailed information about themselves and agreed to have their health followed. Over many years this will build into a powerful resource to help scientists discover why some people develop particular diseases and others do not.". More details please go to UK biobank website.
Recruitment	"UK Biobank recruited 500,000 people aged between 40-69 years in 2006-2010 from across the country to take part in this project. "More details see UK biobank website.
Ethics oversight	NHS REC Ref 11/NW/0382.

Note that full information on the approval of the study protocol must also be provided in the manuscript.