Background

Assortative mating

Assortative mating is the observation of increased phenotypic similarity between couples compared to random pairs. This is most prominent in anthropometric traits, such as BMI and height, measures of socioeconomic status and various behavioral and lifestyle measures (diet, interests, hobbies, etc.). This can be explained by the fact that people tend to choose partners more similar to themselves (i.e. opposites do not actually attract in reality). Additionally, there is also genetic evidence of assortative mating, meaning couples are not only more phenotypically similar but also more genetically similar (see [Genetic evidence of assortative mating in humans](https://www.nature.com/articles/s41562-016-0016) [Matt Robinson] and [Genetic determination of height-mediated mate choice](https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0833-8) [Albert Tenesa]).

There are many factors that strongly influence couple choice, with geographical location being potentially one the strongest indicators. Not only do couples tend to choose partners that are phenotypically similar to them (because of common interests, political views, etc.), but also because of location and convenience. Geographical similarity induces genetic and in turn phenotypic similarity.

Couple convergence

Beside assortative mating, couples living together can also become even more similar over the years. However, it is currently Unknown to what extent couples converge and influence each other over time. Evaluating this is made difficult even further by the fact that people tend to choose partners that are also geographically closer to them, which induces both genetic and phenotypic similarity.

Mendelian Randomization

Give a brief background…

Objectives

* Identify traits which are directly assorted for. In other words, which traits are responsible for the phenotypic similarity we observe amongst couples. In the case of similarity observed with respect to both diet and BMI, perhaps one phenotype is driving the similarity between couples and the other is either indirect or a confounder (as diet and BMI are themselves correlated).
* Do these patterns differ amongst sex (i.e. are females or males more particular with respect to certain traits)?
* Is there evidence for these patterns changing with age?
* Can we find evidence for convergence overtime?
* Are there additional traits that are responsible for the path from an exposure in an index case impacting an outcome in their partner, and if so what are these traits?

Methods

Sample selection and couple definition

This study used the UK Biobank (UKBB) cohort, a prospective population-based study with over 500,000 adult participants. Couples were identified and selected within the UKBB according to the following procedure. The initial UKBB sample comprised 502,616 individuals. First, participants were filtered to only genetotyped, white, unrelated individuals according to the genetic QC file (specifically participants were retained if they had the following values in the QC file: “excess.relatives” = 0, “putative.sex.chromosome.aneuploidy” = 0, “in.white.british.ancestry.subset” = 1 and “used.in.pca.calculation” = 1). Redacted samples and participants that removed consent were also excluded. After filtering, 337,138 participants remained. Within this sample, we included individuals coming from households with exactly two unrelated, opposite-sex participants, leaving 108, 898 participants. Finally, using the data at data-field 6141, “How are people in household related to participant” pairs were filtered to only include couples who had both responded “Husband, wife, or partner”, leaving 103,328 participants, comprising 51,664 couples for downstream analyses (Supplementary Figure 1).

Diagram

Description automatically generated

**Supplementary Figure 1:** Flow chart shows a summary of the couple determination and selection in the UKBB.

Phenotype selection and processing

We performed an agnostic, phenome-wide approach for selecting phenotypes in this study. Specifically, we first selected phenotypes which were analyzed by the Neale group and which had both male, female and joint summary statistics available. Next, after intersecting this list of phenotypes with our internal database (application number #16389), we had 1,278 phenotypes available for analysis. Phenotypes were processed in the filtered QC-data set (N = 337,138) according to a slightly modified version of the PHESANT pipeline to accommodate the phenotypes that we had available in our database1. Continuous variables were transformed to a normal distribution using a rank-preserving inverse normal quantile transformation (INQT), while ordinal and binary traits were re-categorized according to PHESANT documentation (for e.g. categories with less than 10 participants were removed). We then filtered these phenotypes as follows. First, we computed the Pearson phenotypic correlation amongst couples and removed phenotypes with a correlation < 0.1, in order to focus on traits with some indication of assortment. To ensure INQT was not significantly impacting the correlation results, we also calculated the Spearman correlation for each trait and found consistent findings (Supplementary Figure X). Next, we removed phenotypes which had less than 5 valid instrumental variables (IVs) for MR. IVs were defined based on an association *p* < 5 x 10-8 in the joint Neale summary statistics, after pruning for independence (based on a clumping procedure performed in PLINK with the options --clump-kb 10000 and --clump-r2 0.001 using the 1000 Genomes European samples as a reference). Third, using the sex-specific summary statistics, the IV heterogeneity between sexes was calculated. IVs that showed (Bonferroni corrected) significant evidence of heterogeneity between sexes were excluded (*p* < 0.05/[number of IVs]). After filtering IVs with significant sex-heterogeneity, phenotypes were again filtered to only those with at least five valid IVs remaining. Fourth, dietary phenotypes were removed due to high correlation amongst these phenotypes, insufficient power, problems with reverse causation and difficult interpretation2. Finally, we removed several duplicated and redundant phenotypes. Specifically, all left-side body traits (highly correlated with right-side) were removed and we also retained only one of the duplicated phenotypes for BMI and weight (retaining data fields 21001 and 21002, respectively). Additionally, all “qualifications” data was removed (corresponding to field [6138](https://biobank.ctsu.ox.ac.uk/showcase/field.cgi?id=6138)) due to the availability of finer-scale correlated variables, such as “age completed full time education” (data field 845). After this process, 118 phenotypes remained for analysis (see Supplementary Figure 2).

Chart, scatter chart

Description automatically generated

Comparison of spearman vs Pearson correlation

Diagram

Description automatically generated

**Supplementary Figure 2:** Flow chart shows a summary of the phenotype selection included in the pipeline, resulting in 118 phenotypes for analysis. SGG refers to the Statistical Genetics Group, data was used from the corresponding internal UKBB database under application number #16389.

Assessing the role of confounders on trait correlation in couples

We sought to explore the impact of geography on mate-choice by calculating the trait correlations between partners that are due to confounding (captured by genetic principal components (PCs) or geographic location). Specifically, we first tested the correlation of PC values between couples (i.e. *),* with *i* and *p* referring to index and partner). Second, for each trait *X*, we tested the correlation with each PC in the entire UKBB (n = 313XXX). Finally, we estimated the correlation due to confounding as ). We then summed the correlation due to confounding and PC correlation within couples across 40 PCs, as they are orthogonal by definition. These confounding estimates were finally contrasted to the actual couple correlation values (*)*) to explore the extent ancestry may confound couple correlations. The above process was repeated with North and East birth co-ordinates (data fields [129](https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=129) and [130](https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=130), respectively) to further assess the impact of geography on couple choice. We expanded the analysis to other potential confounders of couple correlations, including household income, age completed full time education, Townsend deprivation index and fluid intelligence score (data fields 738, 845, 189, and 20016, respectively).

Estimation of single-trait causal effects in couples

To investigate the causal effect of a trait in one individual on the same trait of their partner, we performed a couple-specific MR analysis. Specifically, the trait in the index case was used as the exposure, and the same trait in the partner was used as the outcome trait. The effect of genetic variants on the exposure were obtained from the Neale lab summary statistics, using the full UK Biobank sample. Instruments for each trait were selected as described above, i.e. being both GW-significant (p < 5 x 10-8) and pruned for independence. Next, we estimated the effects of SNPs on the outcomes of interest by testing the association between each genetic instrument measured in the index individual with the phenotype measured in the partner using the UKBB partner data set described above. In other words, for each phenotype, the corresponding genetic data for the IVs were obtained from the index case while the phenotypes (dependent variable) were taken from the corresponding partner. All SNP-trait estimates were estimated in males and females separately (i.e. using the sex-specific Neale results or two separate models in the couple data), adjusting for age and the first 40 genetic PCs of both the index and partner. We performed linear regression of SNP effects on phenotypes, regardless of data type (including binary), to mimic the Neale models as closely as possible. Continuous phenotypes were scaled to have mean 0 and SD of 1 before regression, while ordinal and binary phenotypes were left as processed by PHESANT.

To estimate the causal effect of a trait from an index case to a partner, we combined the effects of genetic instruments on the exposure (from Neale) with effects on the outcomes (measured among couples) in a MR framework using the inverse-variance weighted (IVW) method (Supplementary Figure 3)3. To estimate the causal effects in both sexes combined, SNP-effects were first meta-analyzed across sexes using fixed effects models prior to performing MR (rather than meta-analyzing the MR estimates directly) to minimize weak instrument bias4. Effects of the genetic estimates on both the exposure and outcome were first standardized to allow for seamless comparison across traits and to the raw phenotype correlation. Significance was determined by adjusting for the number of effective tests based on the correlation matrix of phenotypes tested5. The significance threshold was adapted accordingly as *p* < 0.05/[number of effective tests].

Diagram

Description automatically generated

**Supplementary Figure 3:** Mendelian randomization schematic to assess the causal effect among couples with a single trait (, where represents genetic variant(s), represents a single trait (in an index and a partner), and represents confounding factors which are not associated with genetic variance owning to the random distribution of alleles at conception.

Investigating the effect of time and age on causal relationship in couples

Trait similarity in couples can be driven by both mate choice and/or trait convergence over time spent together. To tease out the contribution of these different sources, we explored whether the cross-partner causal effects change as a function of the length of the relationship and age. The length of relationship for each couple was proxied by the minimum value in the two partners of “length of time at current address” (data field 699). To estimate the effect of age, we took the median age of couples. For each of the two derived variables for each couple (length of relationship and median age), we split the couples into five roughly equal sized bins (using the “smart\_cut” function from the cutr R package). Next, for each single-trait MR described above, analyses were run in the full sample as well as in the different bins. Of the significant results identified in the sex-combined analysis above, we tested to see if there was any significant difference in MR estimates amongst the two sets of bins. Binned MR-estimates were computed using SNP-outcome effect estimated in each bin separately, and the SNP-outcome effects used the same SNP-exposure effects from Neale. Analyses were run in each sex separately and combined (meta-analyzed at the SNP level). As above, SNP effects were standardized prior to calculating MR estimates. To assess for the presence of a trend across bins, we tested the significance of the slope of a linear model of bin-specific MR estimate, inversely weighted by the SE of each MR estimate, versus the  bin center (more precisely, the median  age or time spent together for the given bin). We also tested for evidence of heterogeneity between sexes of the slope estimate. Multiple testing was, as described above, adapted based on the effective number of tests, restricted to traits which showed significant causal effects in the joint (both sexes combined), non-binned MR.

Estimation of two-trait causal effects in couples

Using the same process described above involving a single trait, we also sought to investigate causal effects within couples involving two traits. In other words, two different traits were used as exposure (in the index individual) and outcome in the partner to determine the causal effects in couples involving different traits (for example, effect of education in index case on BMI in the partner). Here, the number of tests could be as high as the number of traits squared, but we only considered  trait combinations with phenotypic correlation < 0.8 (estimated in the entire UKBB, N = XX), in order to avoid too closely related traits. Before performing each two-trait MR, genetic variants were first filtered for evidence of reverse causality at a threshold of *p* < 0.001 (Steiger filter)6. Using meta-analyzed Neale summary statistics, SNPs were removed if the standardized SNP effect with the outcome was stronger than the effect with the exposure based on a one-tailed t-test at a significance level of p < 0.001. As in the single trait MR, SNP-exposure effects were obtained from the Neale summary statistics and SNP-outcome effects were estimated in the couple derived dataset. MR models were run in both sexes separately and jointly (meta-analyzing the SNP effects before performing MR analyses). After removing highly correlated traits, significance was determined based on a standard Bonferroni (BF) correction of p < 0.05/number of tests.

Comparison of paths from index to partner

There are several independent paths through which a trait in an index case could exert a causal effect on a trait in the partner, and we wanted to explore if one path was more dominant, in general, and if there was evidence for the presence of other traits involved. Restricting to only BF-signficant trait pairs (with phenotypic correlation < 0.8) from the couple MR, we sought to explore the various paths from a phenotype in an index case () to a phenotype in the partner () as illustrated in Figure 1. Logically, cannot directly have an effect on another , with the exception of exposure traits that directly alter the environment of their partner, such as smoking creating the presence of second-hand smoke. For instance, increased BMI in an index case could not directly increase cardiovascular disease risk in their partner without acting first on BMI itself in the partner. Specifically, the presence of a causal effect from to () can be summarized by three possible (non-independent) scenarios. First, could exert a causal effect on , followed by having a causal effect on in the partner alone (). Secondly, the reverse could occur whereby has a causal effect on in the index alone, followed by a causal effect of case on (). Third, there could be other mechanisms, either acting directly or through other unmeasured/considered variables. These three scenarios could also act in some combination. In this way, the estimate would capture the paths of , and other mechanisms combined.

Same-individual MR estimates were calculated using meta-analyzed sex-specific Neale estimates for both the SNP-exposure and SNP-outcome effects using the IVW-method. The same set of SNPs were used as in the two-trait couple MR (i.e. first filtered for the presence of reverse causality). SNP-effects were standardized prior to calculating MR effects. Next, we estimated and representing the various paths from to . To quantify , the single-trait couple causal effect estimate (i.e. from the regression ) were multiplied by the same-individual causal estimate (i.e. ). To quantify , we first estimated the causal effect of on in multivariable MR (MVMR), to exclude any residual effect of on phenotype from index to partner. Specifically, was used as the independent variable with both and as independent variables (i.e. the MVMR was ). We included both IVs from and , pruned for independence (performed in PLINK with the options --clump-kb 10000 and --clump-r2 0.001 using the 1000 Genomes European samples as a reference). We took the coefficient of as the direct causal effect from to . Finally, we estimated directly from our two-trait couple MR framework. We compared the estimates of , , and using a Z-test to assess their difference and assessed their relationship using linear regression with the intercept forced through the origin. Finally, we assessed the proportion of that could not be explained merely by the and paths. As and are not perfectly independent, potentially due to correlation between and or pleiotropic limitations of MR, we accounted for this non-independence by estimating the proportion of variance explained () of on a residualized sum of and . We residualized for by running a linear regression as follows: and then summing to .

Diagram

Description automatically generated

**Figure 1:** Directed acyclic graph (DAG) representing the various estimated causal paths from an index case () to a phenotype in the partner () given by , , and . Causal effects from to () can be summarized by three possible (non-independent) scenarios: (1) could exert a causal effect on , followed by having a causal effect on in the partner alone (); (2) the reverse could occur whereby has a causal effect on in the index alone, followed by a causal effect of case on (); (3) there could be other mechanisms, either acting directly or through other unmeasured/considered variables. To quantify , we first estimated the causal effect of on in multivariable MR (MVMR), to exclude any residual effect of on phenotype from index to partner. These three scenarios could also act in some combination. Therefore, the estimate would capture the paths of , and other mechanisms combined.

Results

Relationship between causal effects and raw phenotypic correlation in couples

A total of 118 phenotypes were selected for analysis (based on their elevated correlation between partners and sufficient [> 5] valid IVs suitable for MR analysis) and subsequently tested for a causal effect from index to partner within couples, using MR. We then compared the causal effects obtained using MR with the raw phenotypic correlation observed among couples to identify any traits where the correlation was larger than the MR-estimate, using a one-tailed z-test to test for a significant difference between absolute value of the estimates. In other words, we were particularly interested in identifying traits where the phenotypic correlation in couples was larger than the causal effect from index to partner, which would be indicative of the presence of confounders driving the observed phenotypic correlation. After adjusting for the effective number of tested traits (p < 0.05/66), we identified 38 traits which showed larger absolute phenotypic correlation compared to MR estimates (see Figure 2). Among these included place of birth, North-coordinate (NC); height, and various blood cell counts, including red blood cell (RBC) count (place of birth NC: correlation () = 0.58, MR-estimate () = 0.33, one tailed p-value for difference (*p*) < 1 x 10-8; BFP: = XX, = XX, *p* = XX; height: = XX, = XX, *p* = XX; RBC count: = XX, = XX, *p* = XX).

Chart, line chart

Description automatically generated

**Figure 2:** Most of the significant effects have stronger MR effects compared to correlation.

Of these 38 traits where couple correlation was significantly greater than MR causal effects, we sought to identify potential confounders which may, in part, explain the discrepant estimates. For each of the remaining 117 phenotypes in our pipeline (), we tested the causal effect of each on each trait of the 38 traits of interest () using MR (). For each , we then explored those with a significant impact on (p < 0.05/66). For each of the 38 traits, we found many potential confounders. Namely, the mean number of potential confounders from our set of 109 candidates was 28 (with maximum of 49 and minimum of 9). For instance, for standing height, we identified 32 potential confounders which may explain the larger phenotypic correlation as compared to MR effect. These potential confounders included age completed full time education, BMI, overall health rating, household income, blood cell counts, time spent watching TV, time spent using the computer, medication use, among many others.

After identifying potential confounder traits, we combined these with the within couple causal effect (), and calculated the correlation due to confounding as to determine the contribution each trait () confounds the within couple correlation for trait (). We subsequently calculated the ratio of this correlation () and the correlation of in partners a .  Finally, we sought to determine the overall contribution of all potential confounders impact the correlation for each trait . To do so, we considered a MVMR model with each as an independent variable and as the dependent variable. Potential confounders to include in the MVMR were first pruned to only those that had correlation with < 0.8 and also correlation with any other < 0.8 (prioritized by largest ). Instruments for the MVMR were selected as the union of IVs for each as described above for the two-trait MR (i.e. *p* < 5x10-8 in the joint Neale summary stats without significant evidence of reverse causality with ). Using only unique SNPs, IVs were then pruned for independence. SNPs were prioritized according to minimum ranked p-value with any . Specifically, given a matrix of p-values representing SNP x effects, each column was ranked according to descending p-values (where lower p-value’s received higher ranks) and subsequently the p-value corresponding to the minimum rank for each SNP was used to order SNPs by importance in the clumping procedure. MVMR based was then calculated as , where were obtained from the MVMR and represent the within couple causal effect () as shown in Figure 3. Finally, we compared the difference between the absolute phenotypic correlation and the absolute MR-estimate to the MVMR C as shown in Figure 4.

Diagram

Description automatically generated

**Figure 3:** DAG illustrates the impact a confounder or multiple confounders (trait ) could have on the phenotypic correlation between partners for a given trait (). Correlation due to confounding can be calculated as as , where were obtained from a MVMR with each potential as an independent variable and as the dependent variable, and represent the within couple causal effect ().

Chart, scatter chart

Description automatically generated

**Figure 4**

Effect of sex, age and time together causal effects in couples

Among the 118 phenotypes tested, we identified 64 significant causal effects in partners after adjusting for the effective number of tests (*p* < 0.05/66). We also examined Cochran’s heterogeneity Q-stat to identify traits with high heterogeneity and found no evidence of heterogeneity in the MR-estimates (all *p* > 0.05/66). Of the 64 significant results, we then tested to see if there was any difference between sexes. After adjusting for the effective number of tests among the remaining 60 traits based on the correlation matrix (*p* < 0.05/29), no traits showed significant difference between sexes. However, 15 traits showed a nominally significant difference between sexes (*p* < 0.05), as shown in Table X, which is 4-times higher than expected (*pbinomial* = 7.45x10-8) .

Next, we explored the impact of age and time-spent-together among the 64 significant results in both males and females separately and both sexes combined using linear regression to assess the trend across different bins (measured using a linear regression of MR-estimates versus median bin). There were no significant results in the sex-combined results after adjustment for number of effective tests (p < 0.05/29) in both the results binned by age and time-spent-together (proxied by the amount of time at same address). However, we did identify two phenotypes which showed a significant difference between sexes in the effect estimates across bins (z-test for difference, *p* < 0.05/29). Specifically, when we binned couples by median age, the age effect estimate was significantly different between males and females for paracetamol use (*pdifference* = 0.0016), where the slope was positive for female-to-male effects (MR estimates were stronger as median age increased) and negative for male-to-female effects. Secondly, we identified a significant difference in time-together effect estimate for the variable “job involves mainly walking or standing” (*pdifference* = 0.00043), where the female-to-male effects increased as time-spent-together increased, and the male-to-female effects decreased (Figures to illustrate are below – not sure we will include them though?).

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Finally, we examined Pearson’s phenotypic correlation within the different bins and assessed for the presence of a trend, again using linear models (phenotypic correlation versus median bin). After adjusting for number of tests (*p* < 0.05/29), we identified two traits which showed a significant trend across the bins according to time-spent-together, namely body fat percentage and hand grip strength (right). In both cases, the correlation decreased as time-spent-together increased. We found another two traits which were showed a significant trend across the bins by median age, smoking status: previous and aspirin use. In this case, for both phenotypes, the slope increased as age median-age increased. We found consistent findings using Spearman correlation rather than Pearson.

Impact of potential confounders on trait correlation in couples

We also wanted to assess the impact of potential confounders on trait correlation in couples. Our first observation was that geographical location (using both genetic PCs and North/East coordinates) has a negligible impact on phenotypic correlations. However, we found a significant impact of household income on raw phenotypic correlation among couples (Figure 5) . We also calculated the ratio of correlation due to confounding over the raw phenotypic correlation among couples and took the average across all traits tested (Table X).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Trait | Mean ratio | Median ratio | SD | Number of traits |
| Age completed full time education | 0.0192137 | 0.0093429 | 0.0282624 | 115 |
| Average total household income before tax | 0.0954284 | 0.0543736 | 0.1102624 | 115 |
| Fluid intelligence score | 0.0173814 | 0.0045452 | 0.0643602 | 115 |
| Household coordinates | 0.0059022 | 0.0041529 | 0.0072719 | 116 |
| PCs | 0.0033361 | 0.0021729 | 0.0030056 | 116 |
| Townsend deprivation index at recruitment | 0.0227744 | 0.0106241 | 0.0321605 | 116 |

A close-up of a document

Description automatically generated with low confidence

**Figure 5:** Figure displays couple correlation due to confounding versus the phenotypic trait correlation among couples for selected potential confounder traits. For each trait in the pipeline, we tested the contribution six confounder traits (household income, age completed education, Townsend deprivation index, fluid intelligence score, PCs, and place of birth coordinates) could impact the phenotypic couple correlation. The couple correlation due to confounding for each trait was calculated for each confounder as ) (in the case of PCs and birth coordinates, correlation due to confounding was summed across PCs and coordinates, respectively).

Identification of dominant paths and indirect effects

We sought to compare the three estimated paths from a phenotype in the index case () to another trait in its partner () as illustrated in Figure 1. The total causal effect between and (denoted by ) can be split up into three components: 1) assortative mating through (i.e. ) and then a causal effect between and in the partner (i.e. ), their product is denoted by ; 2) causal effect between and in the index individual (i.e. ), followed by assortative mating through (i.e. ), their product is denoted by ; 3) any remaining effect of on .  We computed within-couple cross-trait causal effect estimates (i.e. ) for all combinations of trait pairs () after removing pairs with phenotypic correlation > 0.8 (number of trait combinations tested = 13,568). Of these, we identified 1006 significant MR effects ( < 0.05/13,568) among couples. We then estimated the contribution of the first two components contributing to these significant cross-trait effects, and , and compared their contribution to the total effect using standard linear regression (Figure 6). The linear regression between and revealed that (assortative mating through ) is stronger in general compared to (assortative mating through ). [Alternatively we could say: paired t-test shows that gamma > rho - which is significant p = 0.03]. When we summed up the effects of and , we found that the sum was significantly larger than . However, these two effects seemed to be correlated, carrying potentially shared signals. Hence, we first residualized for the effects of (), to ensure complete independence between the two estimates, and then added to (). We found no significant difference between and the sum of in this analysis and points in general were near the identity line.

Chart

Description automatically generatedGraphical user interface

Description automatically generated with low confidence

Discussion

Phenotypic similarity among couples is a well-establish phenomena in the field of epidemiology, most notably among behavioural, anthropometric, and cultural/religious traits. Such similarity is predominantly driven by two main factors. First, people tend to choose “mates” with those that are similar with themselves. known as assortative mating (AM). Secondly, and less well-studied, couples may influence each other overtime which induces further phenotypic similarity. However, phenotypic correlations may be subject to confounding and therefore not representative of true causal relationships.

In this report, we sought to investigate causal relationships among couples within the UKBB using MR. We analyzed 118 traits, representing a wide range of anthropometric, behavioural, and disease-related traits. We identified 38 traits which showed significant differences between the MR causal estimates and raw phenotypic correlation. These results suggest that raw phenotypic correlations among couples are likely subject to confounding due to correlation with other traits. For instance, the phenotypic correlation of height among couples was 0.38, while the MR estimate within couples was 0.21 (95%CI: XX, p-val = XX), suggestive that various confounding factors may be playing a role the large phenotypic correlation as compared to causal effect. Indeed, we identified many potential confounders of this relationship by examining traits with a causal effect on height and for each potential confounder, calculated the corresponding correlation due to confounding. Among the largest effects included comparative height at age 10, household income and lung capacity measures. Together these results suggest that while indeed AM exist at both the phenotypic and causal level for height, the phenotypic correlation among couples may be capturing other important aspects of AM including measures of socioeconomic status and various measures of health and fitness. When we investigated the impact of common confounders to our entire panel of phenotypes, we found that household income is indeed a strong confounder of observational phenotypic correlations among couples, explaining on average 10% of the couple correlation. These results also suggest that phenotypic correlations are significantly confounded by few traits and point to a relatively few key traits which are driving AM observations.

Among the 118 phenotypes tested, we found widespread evidence of causal effects among partners. In particular, we identified 64 same-trait causal effects within partners (out of 118 traits), and we also found no evidence of heterogeneity among same-trait couple MR estimates (). This suggests that associations between the index genotype and partner’s phenotype are primarily acting indirectly through the causal relationship between the traits, rather than the presence of a direct effect for index genotype to the partner’s phenotype. We found an enrichment for sex-differences amongst the 64 causal estimates, suggestive that there exists both a sex- and trait-dependent relationship on causal effects among couples. However, in general, female-to-male estimates were stronger than male-to-female estimates. These results point to AM being stronger among females compared to males. When we investigated phenotypic correlations across different ages and amount of time-spent-together, our results suggest that fitness and anthropometric measures are important initially, however the correlation decreases as time increases. On the other hand, we found that smoking and medication use (aspirin, specifically) become more concordant among couples as age increases. As age and time-spent-together are highly correlated variables, it is difficult to distinguish whether this is an effect of convergence or suggestive of an age-dependent AM effect.

Our findings investigating two-trait relationships within couples suggest that causal effects from to are primarily driven by assortative mating through (i.e. ) followed by a causal effect within the partner from to in the partner (i.e. ). In other words, among trait pairs which show a significant causal effect in couples, our results suggest that the exposure is passed from the index to the partner before a causal effect from to in the partner. In contrast, the less likely path would be the inverse, whereby the presence of a causal effect from to in an index case is then followed by being passed directly from index to partner. This result was expected, as it is more reasonable for couples to influence each other at the exposure level rather than the outcome level. Furthermore, our results revealed that the majority of the effect from to goes through assortative mating through and ( and , primarily ) rather than directly from to or through another (third) variable.

We found 1006 significant two-trait causal effects within couples. In particular, we found a positive causal effect of time spent watching television on BMI, with a dominant path through . In other words, these results indicate that partners causally influence each other with respect time spent watching television which in turn has an impact on BMI at the individual level. We also found a positive causal relationship from height to education, with a stronger path through , representing a path whereby height increases education status (found previously) within a single individual, and AM then occurs via education status. We also found a few cases where the causal effect from to (given by ) was not capture by or suggestive of either a direct effect to or the presence of a confounder variable. For example, we found a negative causal effect from smoking status: never on white blood cell leucocyte count within partners, in other words leucocyte count was higher among individuals with partners who smoked. While we also identified a significant effect through (AM through smoking), the effect was much strong through . These findings suggest that there could be a direct effect from index partner by way of second-hand smoke.

This study has limitations which should be taken into account. First, with the current data, we were not able to perfectly disentangle assortative mating from convergence. We did make use of both age and time-together data (proxied by time at the same address) to help shed light on this question, and were able to show that certain traits indeed appear to converge as a couple spends more time together while other traits appear to be more important in the selection process (i.e. true assortative mating). However, to properly assess the question, longitudinal data including measures before couples were together would be best suited to disentangle the complex relationship between assortative mating and convergence. Secondly, While gamma and rho represent independent paths from X\_i to Y\_p (assortative mating through the exposure and outcome, respectively), our results suggest that the computed effects using MR estimates are not perfectly independent. This could potentially be due to overlap in genetic instruments or phenotypic similarity. To the best of our ability, we tried to mitigate this bias. In the calculation of rho, we used a MVMR approach to remove effects of X on Y. Also, when summing gamma and rho we first residualized gamma for effects of rho to ensure independence. Finally, in an effort to increase statistical power, we limited our traits to those with significant correlation amongst couples and >5 valid IVs (among a few other filtering criteria). As a result, anthropometric traits constituted a larger proportion of our traits under study and represent a large percentage of our significant findings. Other phenotypes, such as behavioural and lifestyle traits, were included where possible but, in general, with less statistical power due to both lower couple correlation and less IVs. Additionally, we were limited to the traits in the UKBB.

 In conclusion, we …

References

1. Millard, L. A. C., Davies, N. M., Gaunt, T. R., Smith, G. D. & Tilling, K. Software application profile: PHESANT: A tool for performing automated phenome scans in UK Biobank. *Int. J. Epidemiol.* **47**, 29–35 (2018).

2. Pirastu, N. *et al.* Using genetic variation to disentangle the complex relationship between food intake and health outcomes. *bioRxiv* 829952 (2020) doi:10.1101/829952.

3. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).

4. Burgess, S., Small, D. S. & Thompson, S. G. A review of instrumental variable estimators for Mendelian randomization. *Stat. Methods Med. Res.* **26**, 2333–2355 (2017).

5. Gao, X., Starmer, J. & Martin, E. R. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* **32**, 361–369 (2008).

6. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLOS Genet.* **13**, e1007081 (2017).