Indirect assortative mating for human disease and longevity

Title:

Indirect assortative mating for human disease and longevity

Authors:

Konrad Rawlik[1], Oriol Canela-Xandri[1], Albert Tenesa[1,2,3]

Affiliations:

[1]The Roslin Institute, Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Campus, Midlothian, EH25 9RG. Scotland. UK.

[2]MRC HGU at the MRC IGMM, University of Edinburgh, Western General Hospital, Crewe Road South, Edinburgh. EH4 2XU. UK

[3] Corresponding author

Dr Albert Tenesa

The Roslin Institute

The University of Edinburgh

Easter Bush

Roslin, Midlothian

EH25 9RG

Scotland

Tel: 0044 (0)131 651 9100

Fax: 0044 (0)131 651 9220

Email: Albert.Tenesa@ed.ac.uk

Abstract

Phenotypic correlations amongst partners for traits like longevity or late-onset disease have been found to be comparable to phenotypic correlations in first degree relatives. How these correlations arise in late life is poorly understood. Here, we introduce a novel paradigm to establish the presence of indirect assortment on factors correlated across generations, by examining correlations between parents of couples, i.e., in-laws. Using correlations in additive genetic values we further corroborate the presence of indirect assortment on heritable factors. Specifically, using couples from the UK Biobank cohort, we show that longevity and disease history of the parents of white British couples are correlated, with correlations of up to 0.09. The correlations in parental longevity are replicated in the FamiLinx cohort, a larger and geographically more diverse historical ancestry dataset spanning a broader time frame. These correlations in parental longevity significantly (pval < 0.0093 for all pairs of parents) exceed what would be expected due to variations in lifespan based on year and location of birth. For cardiovascular diseases, in particular hypertension, we find significant correlations (r=0.028, pval=0.005) in genetic values among partners, supporting a model where partners assort for risk factors to some extent genetically correlated with cardiovascular disease. Partitioning the relative importance of indirect assortative mating and shared common environment will require large, well characterised longitudinal cohorts aimed at understanding phenotypic correlations among none blood relatives. Identifying the factors that mediate indirect assortment on longevity and human disease risk will help to unravel factors affecting human disease and ultimately longevity.

Introduction

Partner correlations for a variety of phenotypes have been reported when examining environmental and genetic contributions to complex traits (ANONYMOUS 1903; HIPPISLEY-COX *et al.* 2002; SILVENTOINEN *et al.* 2003; ZIETSCH *et al.* 2011; TENESA *et al.* 2015; CONLEY *et al.* 2016; HUGH-JONES *et al.* 2016; MUÑOZ *et al.* 2016; NORDSLETTEN *et al.* 2016; STULP *et al.* 2016; XIA *et al.* 2016). These correlations between nominally unrelated individuals are substantial, with magnitude comparable to correlations between first degree blood relatives, for instance, between parents and children (MUÑOZ *et al.* 2016; XIA *et al.* 2016). Such effects can be interpreted as phenotypic convergence among partners due to the environmental factors that partners share during their co-habitation. In the case of late-onset diseases and longevity, which are not directly observable or present at the time of mate choice, this would arguably be the simpler explanation. Alternatively, partner correlations for late onset disease and longevity could arise due to indirect assortative mating. That is, direct assortative mating for traits, characteristics or social factors that are risk factors of disease and potentially observable at the time partners met (for instance, behavioural risk factors of disease such as smoking) would lead to indirect assortative mating for other focal traits, such as longevity or late-onset disease. Here, we take direct assortative mating to refer in general to non-random mate choice based on expressed phenotypes. In particular, we do not distinguish between mate choice which leads to positive or negative phenotypic correlations, the latter often being referred to as dissortative mating. The distinction between the causes that underpin partner effects has implications for the study of human behaviour, epidemiology and population genetics. It provides information about human mate choice behaviour and informs about the importance of environmental risk factors shared by couples in the household. The importance to population genetics arises because assortative mating for heritable traits induces a correlation of genetic values among partners, whilst assortment on environmental factors (e.g., social homogamy), and environmental effects shared by partner do not. The correlation of the genetic values of the partners in turn affect the amount of genetic variance of the trait

assorted on, as a consequence estimates of heritability reported in the literature which do not account for assortment overestimate the heritability for that trait in a random mating population due to the covariance among alleles at different loci (FALCONER AND MACKAY 1996) (Fig. 1a, Supplementary Methods). Furthermore, assortative mating for a trait would also induce an increase in heritability for genetically correlated traits (GIANOLA 1982) (Fig. 1b) and a change in the genetic correlation between the assortment and focal traits (Fig. 1c). This is the case even if these focal traits do not directly underlie mate choice, or do not manifest at the time of mate choice. For instance, assortment for BMI, would induce an indirect increase in the genetic variance of cardiovascular disease because there is a positive genetic correlation between these two traits (BULIK-SULLIVAN *et al.* 2015), and an increase in their genetic correlation with respect to what would be expected under random mating.

Establishing assortative mating directly requires knowledge of the phenotype at the time of mate choice. Even for phenotypes which are observable at mate choice, like height, such data are rare. For phenotypes like longevity or disease risk, which only manifest long after mate choice, such data can obviously not be collected. Recent work, starting with Tenesa et al. (TENESA *et al.* 2015), has therefore concentrated on using genotype information to establish assortment (ROBINSON *et al.* 2017). As genetic values (i.e. polygenic scores) are fixed at birth, correlations between partners in such values provides direct evidence for assortment. However, this approach is limited by how well genetic values predict phenotype, i.e., the heritability, and the precision with which genetic values can be estimated. The heritabilities of longevity and many late onset diseases are medium to low (CANELA-XANDRI *et al.* 2017), with estimates for SNP heritability of longevity ranging from 0.12 to 0.3 (KAPLANIS *et al.* 2017). Furthermore, numbers of disease cases, for many diseases which are rare in the general population, and individuals with lifespan information are small in large prospectively collected and genotyped cohorts like UK Biobank, limiting the precision of estimates of genetic values.

Here, we propose a related alternative approach. We examine correlations between the parents of partners. That is, for example, between the father of one spouse and the father of the partner. We present data showing that there is indirect assortment for both longevity and

risk of disease. Specifically, we find that humans choose partners with similar parental history of disease and parental longevity. Since partner choice most likely happens before the parental onset of most of these diseases or parental death, these are unlikely to be the traits on which such choice is made. Furthermore, as these traits are correlated across generations indirect assortment present the most parsimonious model. Finally, we demonstrate assortment directly, showing that the genetic values (i.e. GBLUPs) for hypertension are correlated among partners. Given that assortment for hypertension itself is unlikely, we hypothesise that this correlation in genetic values arises through assortment for one or more traits that influence mate choice and which are genetically correlated with hypertension.

Materials and Methods

The general framework of this study is outlined in Figure 2. We investigated partner correlations (ρ_y^{couple}) in longevity (see Partner Correlations for Longevity). To dissect the source of these correlations and in particular to establish whether they arise due to indirect assortment, we followed several approaches. First, we considered correlations in longevity between parents of focal partners ($\rho_y^{♀inlaws}$ and $\rho_y^{♂inlaws}$) (see Parental Correlations of Longevity). That is, for example, $\rho_y^{♂inlaws}$ is the correlation between the two fathers of a husband and wife pair. Then, we considered to what extend potential targets of assortment, like, Body Mass Index or Socio-Economic status, which are correlated across generations explained any observed parental correlations (see Effect of Environmental factors on parental correlations in longevity). Finally, we evaluated correlations between genetic values (GBLUPs) of the focal partners (ρ_g^{couple}) to demonstrate assortment directly (see Partner correlations of genetic values of parental longevity).

We hypothesised that indirect assortative mating for longevity could be driven by assortative mating for disease risk factors. We therefore also examined indirect assortment on disease risk, following the same approaches as for longevity (see Parental Correlations in Disease History).

The majority of analyses were performed using data from the UK Biobank cohort, but where possible results were replicated using the FamiLinx cohort (KAPLANIS *et al.* 2017).

Couples in the UK Biobank cohort

Identification of heterosexual couples in the UK Biobank has been previously reported (TENESA *et al.* 2015). Specifically, using household sharing information we identified a set of 105,380 households with exactly two members in the cohort. Of these 90,297 satisfied all of the following criteria a) individuals reported different ages for one or both parents b) individuals had an age difference of less than 10 years c) individuals were of opposite gender d) both individuals reported to live only with their partner or partner and children. We restricted our analysis to a subset of 79,094 couples for which both partners self-reported to be of White-British ethnicity.

Couples in the FamiLinx cohort

The FamiLinx cohort (KAPLANIS *et al.* 2017), consisting of 86,124,644 individuals, is based on publicly accessible genealogy data ranging back up to the early 15th century and covering individuals born across the world, although individuals of European and North American birth dominate. In our analysis we restricted ourselves to a subset of individuals with full information regarding year of birth and death, latitude and longitude of the birth location. We removed individuals with a birth location along the zero meridian as visual inspection suggested majority of these to be coding errors. We furthermore removed individuals with lifespans below 30 or above 130. Furthermore following previous analysis (KAPLANIS *et al.* 2017) we removed those individuals born before 1600, due to the sparsity and lower reliability of data before that date, and after 1910, due to the bias towards individuals with reduced lifespan after that date. Finally, also following previous analysis (KAPLANIS *et al.* 2017), we removed individuals who died during the American Civil War (year of death 1861 to 1865), the 1st World War (year of death 1914 to 1918) and the 2nd World War (year of death 1939 to 1945) due to the excess

number of early death in these periods. This resulted in a dataset of 3,445,971 individuals. Considering individuals with common offspring, we identified a set of 239,541 couples.

Definition of Birth Location

Both the UK Biobank and FamiLinx contain information about the birth locations of individuals, which we used to adjust for any potential geographical differences between longevity. However, in both cohorts the provided information is at a scale too fine to allow for effective stratification based on birth location. We therefore defined a Birth Location at a coarser scale in both cohorts.

The UK Biobank contains information about the coordinates of the birth location with a resolution of one kilometer (km). We identified a subset of individuals with miscoded coordinates corresponding to birth in the Atlantic Ocean identified through visual inspection and set their Birth Location as missing. We used a 15 km grid to define Birth Location. That is, we assign all individuals who share birth coordinates when divided by 15 km and rounded to an integer to the same Birth Location.

In the FamiLinx cohort we defined a one degree latitude and longitude grid to derive Birth Location.

Genotypes and Estimation of genetic values in UK Biobank

To performed genetic analyses we identified a set of quality controlled, genotypically White-British individuals from the UK Biobank. Using appropriate subsets of these individuals as described for specific analyses, we jointly estimated SNP heritabilities and SNP effects following the mixed model approach using the DISSECT tool (CANELA-XANDRI *et al.* 2015). We used the estimated SNP effects to compute genetic values (i.e. Best Linear Predictors, BLUPs). All models included the leading 20 genomic principal components as fixed effects.

The set of individuals available for genetic analyses was identified as follows. We used the data for the individuals genotyped in phase 1 of the UK Biobank genotyping program. 49,979 individuals were genotyped using the Affymetrix UK BiLEVE Axiom array and 102,750 individuals using the Affymetrix UK Biobank Axiom array. Details regarding genotyping procedure and genotype calling protocols are provided elsewhere

(http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580). We performed quality control using the entire set of genotyped individuals before extracting the White-British cohort used in our analyses. From the overlapping genetic markers between the two arrays, we excluded those which were multi-allelic, their overall missingness rate exceeded 2% or which exhibited a strong platform specific missingness bias (Fisher's exact test, pval < 10^{-100}). We also excluded individuals if they exhibited excess heterozygosity, as identified by UK Biobank internal QC procedures (http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580), if their missingness rate exceeded 5% or if their self-reported sex did not match genetic sex estimated from X chromosome inbreeding coefficients. These criteria resulted in a reduced dataset of 151,532 individuals. To define the genotypically White-British subset, we performed a Principal Components Analysis (PCA) of all individuals passing genotypic QC using a linkage disequilibrium pruned set of 99,101 autosomal markers (http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=149744) that passed our SNP QC protocol. The genotypically White-British individuals were defined as those for whom the projections onto the leading twenty genomic principal components fell within three standard deviations of the mean and who self-reported their ethnicity as White-British. We furthermore pruned the set of genotypically White-British individuals removing one individual from pairs with relatedness above 0.0625 (corresponding to second degree cousins) to obtain a dataset of unrelated genotypically White-British individuals. Finally, in our genetic models we only used genetic variants that had passed QC, that did not exhibit departure from Hardy-Weinberg equilibrium (pval < 10^{-50}) in the unrelated genotypically White-British cohort and which had a minor allele frequency > 5%.

Partner Correlations for Longevity

We estimated partner correlations of longevity, defined as the age in years at death using data from the two cohorts, the UK Biobank and Familinx. We also computed correlations of longevity adjusted for cohort effects. Specifically, we computed adjusted longevity as the difference between an individual's lifespan and the mean lifespan of the stratum defined by

the individual's sex, birth year and birth location (see Definition of Birth Location), excluding all strata with fewer than 10 individuals.

As the majority of UK Biobank participants are alive, we used the biological mothers and fathers of participants. Specifically, we identified self-reported White-British individuals with both parents deceased (using data fields UKBID 21000, 1797 and 1835), and non-missing Birth Location (see Definition of Birth Location). This yielded 252,899 pairs of parents for which we computed Pearson's correlations between longevity extracted from data fields UKBID 1807 and 3526. The UK Biobank does not directly contain information regarding the years or location of birth of parents of participants. As such, we used the participant's place and year of birth (UKBID 34) as proxy measures of the parent's place and year of birth. For a subset of parents, specifically parents who are still alive at recruitment of the participant, we can infer the parents' year of birth from the date of recruitment and the parents' age. The subset of parents who are still alive is relatively small, only 22% of fathers and 39% mothers respectively, and is complementary to the set of parents used in the analysis, who were required to be deceased. While we can therefore not use the data in our analysis, it allows us to evaluate the effect of using a proxy measure. The correlation between the year of birth of the offspring and their parent is relatively high with $\rho=0.78$.

In the FamiLinx cohort we used all 239,541 couples identified as described above (see Couples in the FamiLinx cohort). We computed longevity as the difference of year of death and year of birth.

Parental Correlations of Longevity

We computed Pearson's correlations of longevity and adjusted longevity for parents of partners. That is, we computed, for example, the correlation between the longevity of the two fathers of the male and female partners in a couple. We considered the three combinations of parents, that is, the two fathers or the two mothers of the partners and the father of one partner

and the mother of the other partner, separately. Both longevity and adjusted longevity were computed as for the analysis of partner correlations (see Partner Correlations for Longevity). Of the 79,094 couples identified in the UK Biobank (see Couples in the UK Biobank) 40,504 had both mothers and 60,978 both fathers deceased, while there were 104,922 father-mother pairs. Amongst the 3,445,971 individuals retained for analysis in the FamiLinx cohort (see Couples in the FamiLinx Cohort), we identified 97,223 sets of fathers, 66,077 sets of mothers and 143,896 father-mother pairs.

We computed expected distributions of parental correlations due to geographical and temporal mating structure in the population based on permutations. Specifically, we generated fictitious sets of couples which matched the observed mating structure for birth years and birth locations and computed the parental correlations in longevity for these fictitious couples. To generate the fictitious couples we stratified couples based on the Birth Year and Birth Locations of both partners and permuted male partners within each stratum. To allow for effective permutations we only included couples in strata of size larger than 10 in the analysis. For each permutation we computed Pearson's correlations of parental longevity as a test statistic. Empirical pvalues where then computed as the fraction of statistics exceeding the statistic computed without permutation, based on 10,000 permutations.

Effect of Environmental factors on parental correlations in longevity

We evaluated partner correlations for a range of potential assortment factors and evaluated their contribution to any observed correlations in parental longevity.

Specifically, we extracted Townsend Deprivation Index (UKBID 189), height (UKBID 50), waist to hip ratio (computed from UKBID 48 and 49), BMI (UKBID 21001) and smoking history in Pack Years (UKBID 20161) for all individuals in the 79,094 couples identified in the UK Biobank. The Townsend Deprivation Index is an area measure of socio-economical deprivation. We computed Pearson's correlations between the male and female partners for all pairs of these variables as well as birth year.

We then computed linear regression models, regressing parental longevity on birth year, Birth Location, as well as Townsend Deprivation Index and height, waist to hip ratio, BMI and

smoking history in Pack Years, and the squares of these factors, of their children. Birth Year and Birth Location were coded as categorical variables while all other factors and their squares were included as continuous variables. Using the fitted models, we computed residuals and correlations between couples using these residuals. Comparing these, we quantified the change in correlations due to inclusion of individual covariates in the models.

Partner correlations of genetic values of parental longevity

As the majority of individuals in the UK Biobank are still alive, we cannot estimate genetic values for longevity directly. We therefore again use information about the lifespans of parents of participants and estimate genetic values (GBLUPs) for parental longevity as a proxy for genetic values of individuals longevity.

Of the UK Biobank individuals retained for genetic analysis (see Genotypes and Estimation of genetic values in UK Biobank), subsets of 79,216 and 64,002 had respectively deceased fathers and mothers. Using these individuals, we estimated SNP heritabilities and genetic variant effects for parental longevity based on common variants, i.e., variants with minor allele frequency above 5%. Of the 79,094 couples identified in the UK Biobank (see Couples in the UK Biobank Cohort) a subset of 10,160 couples consisted of individuals retained for genetic analysis. For these couples, using the estimated genetic variant effects, we computed genetic values (CANELA-XANDRI *et al.* 2015; CANELA-XANDRI *et al.* 2016) for parental longevity and computed their Pearson's correlation.

Disease History in the UK Biobank

Participants in the UK Biobank provide information about the family history for twelve diseases for both biological parents (UKBID 20107 and 20110). Considering the 79,094 couples identified in the UK Biobank (see Couples in the UK Biobank Cohort), disease history for both biological parents of each partner was reported by 58,043 couples for Heart Disease, Stroke, Chronic Bronchitis, High Blood Pressure, Diabetes and Alzheimer's Disease and by 57,644 couples in the case of Lung Cancer, Bowel Cancer, Parkinson's Disease and Depression. For

the latter subset, information regarding disease history for the relevant parent for Breast and Prostate Cancer was available for each partner.

The twelve disease for which family history was provided do not directly match disease reported in the self-reported medical history of participants (UKBID 20002). To identify self-reported controls therefore utilized the methodology of Muñoz et al. (MUÑOZ *et al.* 2016) to match diseases to those reported for family history.

Parental Correlations in Disease History

Following the methods for parental correlations for longevity (see Parental Correlations of Longevity), we computed correlations of disease history between the fathers and mothers of couples in the UK Biobank. We also computed correlations for each disease using only couples where both partners are self-reported controls for the relevant disease.

As disease history or status for an individual is a binary trait, Pearson's correlations are not a suitable measure of correlation. Instead we computed polychoric correlations (DRASGOW 1986) using the R package polycor (FOX 2010). In addition we assessed dependence between partner's family histories using a χ^2 test and by computing empirical mutual information (COVER AND THOMAS 2012). For mutual information we computed an empirical pvalue for departure from independence using permutations. That is, we computed empirical mutual information for 1000 datasets in which family history for the male partners had been permuted and compared them to the empirical mutual information on the observed data.

As for longevity we evaluated the expected effect of assortment due to place and year of birth using permutations. Permutations were performed as for longevity, using the χ^2 statistics, rather than Pearson's correlation, as test statistic.

We performed an additional permutation analysis to assess the impact of using the offspring's year of birth as a proxy for the parents' year of birth. Unlike in the analysis of longevity, where all parents are deceased, a subset of parents with family history is still alive. For these parents we can compute the year of birth. On the subset of parents with available year of birth, we permuted UK Biobank couples within the years of birth of their parents. That is, the offspring

within the years of birth of the parents. We did not permute within both Birth Year and Birth Location strata due to the smaller sample size.

Partner correlations of genetic values of disease history

We computed correlations for genetic values of parental disease history and self-reported disease status. For own disease status, we restricted the analysis to diseases with prevalence in the sample above 5% and excluding prostate and breast cancers.

For family disease history traits we fitted models with only genomic principal components, as well as models which also included the participant's Birth Year and Birth Location as categorical and the parents' age as continuous covariates. The parent's age was computed as either the age at death (UKBID 1807 and 3526), if the parent was deceased or age at assessment (UKBID 2946 and 1845) otherwise. Models used to estimate genetic values for self-reported disease also included the participant's Sex, Age and Townsend Deprivation Index as fixed effects.

We fitted models using all individuals available for genetic analysis (see Genotypes and Estimation of genetic values in UK Biobank) who reported family history. We transformed heritabilities which were estimated on the observed scale, i.e., modeling disease status directly, to the liability scale using the sample specific prevalence (LEE *et al.* 2011). Using SNP effects estimated on all individuals, we computed genetic values for the 10,160 couples that comprised individuals retained for genetic analysis (see Genotypes and Estimation of genetic values in UK Biobank) and computed their Pearson's correlations. We combined paternal and maternal estimates using the Olkin-Pratt fixed effect approach (SCHULZE 2004).

Results

Partner Correlations in Longevity

We found that the lifespan of the biological mothers and fathers of all self-reported White-British individuals in the UK Biobank with both parents deceased was correlated and significantly different from zero (ρ_y^{couple} = 0.11, 95% CI 0.107 – 0.114, pval < 10^{-188}). The correlation was only slightly reduced ($\rho_{y\text{-}adj}^{couple}$ =0.10, 95% CI 0.091 – 0.108, pval < 10^{-188}) and

remained significantly different from zero when adjusting for the participants' year of birth as a proxy of the parent's year of birth, which itself was unavailable. This finding reproduced in the FamiLinx cohort. Specifically, although partner correlations for longevity in the FamiLinx cohort were significantly higher (ρ_y^{couple} =0.18, 95% CI 0.176 – 0.183, pval < 10^{-188}), correlations for lifespans adjusted for an individual's year and place of birth were comparable to those in the UK Biobank cohort ($\rho_{y\text{-}adj}^{couple}$ =0.125, 95% CI 0.121 – 0.129, pval < 10^{-188}).

Parental Correlations of Longevity

We found significant correlations for the lifespans of both mothers ($\rho_y^{♀inlaws}$ =0.049, 95% CI 0.038 – 0.062, pval=10^{-15}) and fathers ($\rho_y^{♂inlaws}$ =0.032, 95% CI 0.022-0.042, pval=10^{-10}) of couples in the UK Biobank. This finding reproduced in the FamiLinx cohort. Although we again observed higher correlations in lifespans of mothers ($\rho_y^{♀inlaws}$ =0.061, 95% CI 0.053 – 0.068, pval=10^{-55}) and fathers ($\rho_y^{♂inlaws}$ =0.071, 95% CI 0.064 – 0.077, pval=10^{-107}) of couples compared to the UK Biobank, correlations between adjusted lifespans where again comparable to those in the UK Biobank ($\rho_{y\text{-}adj}^{♀inlaws}$ =0.02, 95% CI 0.012 – 0.030, pval=10^{-7} and $\rho_{y\text{-}adj}^{♂inlaws}$ =0.03, 95% CI 0.023 – 0.038, pval=10^{-17} for mothers and fathers respectively). Considering father-mother pairs, we observed reduced correlations in the UK Biobank ($\rho_y^{♂/♀inlaws}$ =0.014, 95% CI = 0.005 – 0.024, pval=0.003) which however were still significant. In the Familinx cohort on the other hand, correlations for father-mother pairs were comparable to those between fathers and mothers and significant ($\rho_y^{♂/♀inlaws}$ =0.055, 95% CI 0.049 – 0.060, pval=10^{-15} and $\rho_{y\text{-}adj}^{♂/♀inlaws}$ =0.055, 95% CI 0.049 – 0.060, pval=10^{-15} for observed and adjusted lifespan respectively). We did not consider father-mother correlations in the UK Biobank cohort further and discuss the likely reasons for the observed discrepancy below (see Discussion).

We compared the observed parental correlations to the distribution of correlations for fictitious sets of couples with matched mating structure for year and location of birth. The expected correlation due to mating structure, i.e., the mean correlation across fictitious sets of

couples, were small and not significantly different from zero in the UK Biobank (ρ_{mean}= 0.02, s.d. 0.006 and ρ_{mean}= 0.01, s.d. 0.005 for mothers and fathers respectively). Expected correlations where larger and significantly different from zero in the FamiLinx cohort (ρ_{mean}= 0.03, s.e. 0.007, ρ_{mean}= 0.03, s.d. 0.005 and ρ_{mean}= 0.02, s.d. 0.004 for mother, father, and mother-father pairs respectively). The observed correlations lie in the extreme tails of the distributions of correlations between parents' lifespans (Supplemental Figure S1).The empirical pvalues for the observed correlations are 0.0002 and <0.0001 for mothers of couples in UK Biobank and FamiLinx respectively and 0.0093 and <0.0001 for the fathers of couples in UK Biobank and FamiLinx respectively. For father-mother pairs of couples in the FamiLinx cohort the empirical pvalues for observed correlations is <0.0001.

Year and birth place, socioeconomic status (as measured by Townsend Deprivation Index), height, waist to hip ration, body mass index and smoking history measured in Pack Years (as a proxies of a putative behavioural factor associated with disease and longevity), showed significant partner correlations in the UK Biobank (Supplemental Table S1). Adjusting parental lifespans for any of these factors reduced the observed correlations. Birth year and location were the most important factors, reducing the observed correlations for both maternal and paternal longevity by around 55%. Socioeconomic status and the other factors had a lesser but still important effect on the correlation of lifespan of parents, reducing such correlation an additional ~15%.

Significant SNP heritabilities were observed for mother's (h^2=0.03, 95% CI 0.02 – 0.04) and father's (h^2=0.04, 95% CI 0.03 – 0.05) longevity (Supplemental Table S3). These SNP heritabilities for a parental phenotype are under certain assumptions expected to be ½ the SNP heritability of the phenotype measured in the individual. Correlations between partners in genetic values of parental longevity were not found to be significantly different from zero (ρ_g^{couple} = -0.007, 95% CI -0.026 – 0.013, pval = 0.5 and ρ_g^{couple} =0.01, 95% CI -0.009 – 0.030, pval=0.3 for paternal and maternal longevity respectively).

Parental Correlations of Disease History

We found significant (P<0.05) polychoric correlations, which were consistent for both fathers and mothers, for half of the twelve examined diseases: heart disease, stroke, lung cancer, chronic bronchitis, hypertension, and Alzheimer's disease (Table 1, Supplemental Table S4). Only stroke in fathers failed significance after Bonferroni correction (P<0.05/22). Of these, the largest correlation was for paternal hypertension ($\rho_y^{\male\,inlaws}$ =0.09, 95% CI 0.08 – 0.11, pval=10^{-35}) and the smallest for paternal stroke ($\rho_y^{\male\,inlaws}$ =0.02, 95% CI 0.01 – 0.04, pval=0.003). The history of prostate cancer among fathers of couples was also significantly correlated ($\rho_y^{\male\,inlaws}$ =0.04, 95% CI 0.01 – 0.06, pval=0.004). Among mothers, the correlations for lung cancer ($\rho_y^{\female\,inlaws}$, 95% CI 0.04 – 0.11, pval=10^{-5}), hypertension ($\rho_y^{\female\,inlaws}$ =0.08, 95% CI 0.07 – 0.10, pval<10^{-37}) and Alzheimer's ($\rho_y^{\female\,inlaws}$ =0.08, 95% CI 0.06 – 0.10, pval<10^{-12}) were the largest, whilst the correlations for heart disease were only marginally smaller ($\rho_y^{\female\,inlaws}$ =0.07, 95% CI 0.06 – 0.09, pval<10^{-22}). The analysis using only couples of self-reported controls was largely in agreement with the analysis using all couples (Supplemental Table S5).

We compared the observed parental associations to the distribution of associations for fictitious sets of couples with matched mating structure for year and location of birth (Supplemental Table S6). Results using a mating structure based on the parent's year of birth, available in only a subset of parents, were consistent with the results obtained when using the participant's year of birth as a proxy measure (Supplemental Table S7).

We found modest but significant SNP heritabilities for a majority of the considered parental family histories (Supplemental Table S8). Correlations between genetic values of partners were significant (P < 0.05) for maternal and paternal history of hypertension as well as maternal heart disease, stroke and chronic bronchitis (Table 2). However, only maternal chronic bronchitis and hypertension remained significant after Bonferroni correction (P < 0.05/22). Whilst hypertension in fathers did not reached the stringent Bonferroni correction threshold, the size of the correlation was similar to that of maternal hypertension. Furthermore,

hypertension remained significant in the meta-analysis of paternal and maternal correlations (Table 2).

While Correlations between genetic values were reduced, when adjusting for an individual's birth year, birth location and the parent's age, they remained significant ($P < 0.05$) for maternal and paternal hypertension and maternal chronic bronchitis and stroke (Supplemental Table S9).

Despite the smaller numbers of cases, when using own disease status rather then parental disease history, we again found the correlations of genetic value of partners for hypertension to be significant and of similar size to the parental hypertension (ρ_g^{couple} =0.03, 95% CI 0.01 – 0.05, pval = 0.005).

Discussion

Partner correlations for age at death have been demonstrated going back to early work on assortative mating (ANONYMOUS 1903). We were able to reproduce these results in two independent cohorts of unprecedented sample size. The partner correlations we observed were significantly lower than the correlation of 0.23 reported a century ago for a much smaller sample from the UK (ANONYMOUS 1903), but similar to more recent estimates of 0.12 in a Canadian population (PHILIPPE 1978). The sample of partners from the UK Biobank used here was censored, consisting of parents of participants and necessarily excluding all parents who were still alive. However, the close agreement between estimates in the independent FamiLinx cohort and previous estimates does not suggest that this introduced substantial bias. The results suggest that partner correlations for lifespan, after adjusting for mating structure due to year and place of birth, are in the region of 0.1 – 0.12. Estimates of heritability for longevity in the FamiLinx cohort imply a phenotypic correlation between 1st degree relatives of 0.06 (KAPLANIS *et al.* 2017), while previous estimates of heritability suggest higher correlations of 0.13 (HERSKIND *et al.* 1996). Our estimates of SNP heritability for longevity of an individual's parents suggest a phenotypic correlation between 1st degree relatives of 0.03 or 0.04. Unlike previous estimates, our estimates are based on samples of unrelated individuals, largely

precluding inflation due to shared environment which may have affected previous estimates. On the other hand, we only estimate the variance explained by common SNPs and therefore likely underestimate the heritable component of longevity. However, even allowing for the whole range of estimates, we may conclude that partner effects are comparable in magnitude, or even exceed, genetic effects on longevity.

Various possible explanations exist for the observed partner correlations. The year of death of partners could potentially be correlated due to effects directly related to the partner's death (i.e. a partner's death has a causal link with the other partner's death). This together with the assortment by birth year, as we observed in the UK Biobank, would lead to partner correlations for lifespan. More generally, convergence due to shared environmental factors represents in the absence of other data the most plausible explanation for the observed partner correlations. That is, partners share one or more environmental risk factors, such as for example a diet, which affects life expectancy. Such shared environment can be restricted to the partners. More broadly, correlations may reflect mating structure within a broader shared environment. For example, partners may mate preferably in the same socio economical stratum. This may, depending on interpretation, be considered a form of assortative mating. In particular, one's broader environment may have genetic underpinnings. For example, one's socio-economic status may be is influenced by heritable traits like educational attainment (BELSKY *et al.* 2018) and their combined effect reduce social mobility.

By comparison to partner correlations, the estimates of correlations between parental longevity we report are substantially smaller. Indeed, they are arguably small enough to be considered practically insignificant. However, we do not argue for their significance based on their magnitude. As a matter of fact, taking into account the low heritability of longevity, they are expected to be small. Instead, their relevance lies in the information their presence provides about the larger partner correlations. They provide evidence that observed partner correlations arise due to a form of assortment. Specifically, they provide evidence that mating is not random with respect to factors which persist across generations. As the parents of

partners do not share the narrow environment of the couple, our results provide evidence that the observed correlations, at least partly, arise due to mating structure related to factors correlated across generations. Correlations across generations can arise due to several distinct pathways which cannot be differentiated by considering correlations of parents of couples. On the one hand genetic effects lead to across generation correlations. These can take the form of direct effects, i.e., classical heritability, or indirect parent offspring effects as recently described (KONG *et al.* 2018). On the other hand, cross generational correlations can also arise due non genetic transmission, i.e., cultural heritability. For example, low social mobility in a society will lead to parent offspring correlations in socio-economic status.

Like partner correlations, parental correlations are expected to be partly explained by differences in life expectancy across history and geography. We have demonstrated that a mating structure based on these factors alone cannot explain the observed correlations. Identification of the specific factors contributing to the observed partner correlations represents an important question for future research. We have examined the contribution of a small number of baseline factors, each of them heritable (CANELA-XANDRI *et al.* 2017), including known targets of assortment like height and factors reflecting social mating structure, like the Townsend Deprivation Index. All of the examined factors explain parts of the observed correlation and it does not appear a single factor will be able to explain partner correlations in longevity. However, our results suggest that these factors and socioeconomic status are correlated across generations as the children's phenotypes and socioeconomic status explain some of the correlation in longevity of their respective parents.

We were not able to demonstrate correlations in genetic values for longevity. Lack of such correlations would be consistent with environmental assortment, i.e., mating within a broader shared environment or cultural transmission of factors across generations. However, power to detect correlations in genetic values is limited due to the low number of couples available and the low heritability of the trait (Supplemental Table S4). In particular, as a majority of the cohort is still alive it was necessary to use parental longevity to estimate genetic effects. While this

approach has been successful in identifying genetic effects for longevity in a GWAS setting (JOSHI *et al.* 2016), the reduction in heritability due to using a parents phenotype, severely impacts the precision with which genetic values can be estimated. We would therefore suggest that these results do not provide strong evidence against assortment on heritable risk factors.

A majority of the reported estimates were consistent across both cohorts and with previous estimates, where these are available. A notable exception are the reduced correlations for parental longevity for father-mother pairs in the UK Biobank cohort, when compared to the same estimate in the FamiLinx cohort and correlations for same sex parent pairs in both cohorts. We suggest that this is a consequence of the limitations of the UK Biobank data. Specifically, as noted previously, the UK Biobank cohort is censored. Parents who are still alive are excluded. Such censoring will bias observed correlations downwards (BEGIER AND HAMDAN 1971). This is consistent with the lower correlations observed in the UK Biobank compared to the FamiLinx cohort which does not suffer from such censoring. This effect is exacerbated when censoring is stronger on one of the two variables as it is the case for father-mother correlations, due to higher life expectancies for females.

We hypothesised that partner correlations in longevity could be mediated through partner correlations in disease risk. For a majority of the examined disease partner correlations had been previously reported (MUÑOZ *et al.* 2016). Our results for disease risk are in line with those for longevity. That is, the observed partner correlations, at least partly, arise due to assortment on factors correlated across generation. Indeed, for a number of diseases, in particular hypertension, we find direct evidence for assortative mating. As the results for couples of self-reported controls were in line with those using all couples, we can exclude the possibility of direct assortment on disease status. We therefore conclude, that these correlation is likely indirectly generated through genetic correlation between the focal trait (e.g. hypertension) and another, genetically correlated, trait or traits for which assortment happens, e.g., BMI (ROBINSON *et al.* 2017). A consequence of this model is that disease prevalence in the population may potentially be increased through indirect assortment for traits or risk factors

correlated with disease (PEYROT *et al.* 2016). While we find direct evidence for assortment on genetic risk factors for some disease, parental correlations for other disease lack evidence for assortment from correlations of genetic values. Parental correlations for these diseases could arise due to shared broad environment. In the particular case of late onset disease, like for example, Alzheimer's the observed correlations could arise as a consequence of correlations in longevity.

The cohorts used in this study have several limitations. For example, the already mentioned censoring of partners who are still alive in the UK Biobank. Another limitation is the lack of information about the year of birth of a majority of parents in the UK Biobank. However, correlations between the offspring's and parent's year of birth, where both are available as well as, replication of results on the parental disease history using the parents' year of birth, both suggest that adjusting for year of birth of the children is an acceptable, albeit not perfect, proxy for year of birth of the parents. In particular, results did not suggest that using the offspring's year of birth as a proxy introduced a substantial bias. The FamiLinx cohort on the other hand has a genealogical structure, potential biasing observed correlations upwards. However, the close agreement of estimates with those obtained in the UK Biobank does not suggest this is the case.

Taken together the results suggest that the characteristics that influence mate choice lead to detectable assortment for familial disease and longevity. This assortment is only partially explained by birth cohort and the few factors chosen to reflect the social mating structure, suggesting a contribution to assortment for parental disease history and longevity of other traits, lifestyle choices or social factors shared among parents and children. While we have not directly demonstrated that the underlying factors are transferred across generations, that is, that the same behavioural or social factors which drive parental disease risk are also the factors underlying mate choice in the offspring, such a model presents the most canonical explanation. While recent work has highlighted traits which are plausible candidates for direct assortative mating, like for example height (TENESA *et al.* 2015; ROBINSON *et al.* 2017), our

work suggests a network of effects. Whereby direct assortative mating on observable factors, leads to indirect assortment for a multitude of genetically correlated traits. This highlights that assortative mating can have effects far beyond the focal trait and suggests wide-spread levels of pleiotropy. Understanding the contributions that mate choice and cultural transmission of behaviours and environments across generations make to these correlations will present a major but exciting challenge of future research.

Acknowledgements

This work was mainly supported by The Roslin Institute Strategic Grant funding from the BBSRC. AT also acknowledges funding from the Medical Research Council Human Genetics Unit. This work used the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk) and the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/). This research has been conducted using the UK Biobank Resource.

Conflicts of Interest

None

Data Availability Statement

Required data can be accessed through the UK Biobank (http://www.ukbiobank.ac.uk/) and the FamiLinx website (http://www.familinx.org/) respectively. For analyses involving genotypes, we used the individuals genotyped in phase 1 of the UK Biobank genotyping project, which were released by the UK Biobank in June 2015. The genotype data were downloaded on 5 June 2015. The DISSECT software used to perform the analysis based on genetic values is freely available from http://www.dissect.ed.ac.uk/.

582 **References**

583 Anonymous, 1903 Assortative Mating in Man: A Cooperative Study. Biometrika 2**:** 481-
584 498.
585 Begier, M. H., and M. A. Hamdan, 1971 Correlation in a Bivariate Normal Distribution
586 with Truncation in Both Variables. Australian Journal of Statistics 13**:** 77-82.
587 Belsky, D. W., B. W. Domingue, R. Wedow, L. Arseneault, J. D. Boardman *et al.*, 2018
588 Genetic analysis of social-class mobility in five longitudinal studies.
589 Proceedings of the National Academy of Sciences.
590 Bulik-Sullivan, B., H. K. Finucane, V. Anttila, A. Gusev, F. R. Day *et al.*, 2015 An atlas
591 of genetic correlations across human diseases and traits. Nature genetics.
592 Canela-Xandri, O., A. Law, A. Gray, J. A. Woolliams and A. Tenesa, 2015 A new tool
593 called DISSECT for analysing large genomic data sets using a Big Data
594 approach. Nat Commun 6**:** 10162.
595 Canela-Xandri, O., K. Rawlik and A. Tenesa, 2017 An atlas of genetic associations in
596 UK Biobank. bioRxiv.
597 Canela-Xandri, O., K. Rawlik, J. A. Woolliams and A. Tenesa, 2016 Improved Genetic
598 Profiling of Anthropometric Traits Using a Big Data Approach. PloS one 11**:**
599 e0166755.
600 Conley, D., T. Laidley, D. W. Belsky, J. M. Fletcher, J. D. Boardman *et al.*, 2016
601 Assortative mating and differential fertility by phenotype and genotype across
602 the 20th century. Proceedings of the National Academy of Sciences 113**:** 6647-
603 6652.
604 Cover, T. M., and J. A. Thomas, 2012 *Elements of information theory*. John Wiley &
605 Sons.
606 Drasgow, F., 1986 Polychoric and polyserial correlations in *The Encyclopedia of*
607 *Statistics*, edited by S. Kotz and N. Johnson.
608 Falconer, D. S., and T. F. C. Mackay, 1996 *Introduction to quantitative genetics*.
609 Pearson Prentice Hall.
610 Fox, J., 2010 polycor: Polychoric and Polyserial Correlations, pp.
611 Gianola, D., 1982 Assortative mating and the genetic correlation. Theoretical and
612 Applied Genetics 62**:** 225-231.
613 Herskind, A. M., M. McGue, N. V. Holm, T. I. Sörensen, B. Harvald *et al.*, 1996 The
614 heritability of human longevity: a population-based study of 2872 Danish twin
615 pairs born 1870–1900. Human genetics 97**:** 319-323.
616 Hippisley-Cox, J., C. Coupland, M. Pringle, N. Crown and V. Hammersley, 2002
617 Married couples' risk of same disease: cross sectional study. Bmj 325**:** 636.
618 Hugh-Jones, D., K. J. Verweij, B. S. Pourcain and A. Abdellaoui, 2016 Assortative
619 mating on educational attainment leads to genetic spousal resemblance for
620 polygenic scores. Intelligence 59**:** 103-108.
621 Joshi, P. K., K. Fischer, K. E. Schraut, H. Campbell, T. Esko *et al.*, 2016 Variants near
622 CHRNA3/5 and APOE have age-and sex-related effects on human lifespan.
623 Nature communications 7.
624 Kaplanis, J., A. Gordon, M. Wahl, M. Gershovits, B. Markus *et al.*, 2017 Quantitative
625 analysis of population-scale family trees using millions of relatives. bioRxiv.
626 Kong, A., G. Thorleifsson, M. L. Frigge, B. J. Vilhjalmsson, A. I. Young *et al.*, 2018 The
627 nature of nurture: Effects of parental genotypes. Science 359**:** 424-428.
628 Lee, Sang H., Naomi R. Wray, Michael E. Goddard and Peter M. Visscher, 2011
629 Estimating Missing Heritability for Disease from Genome-wide Association
630 Studies. The American Journal of Human Genetics 88**:** 294-305.

631 Muñoz, M., R. Pong-Wong, O. Canela-Xandri, K. Rawlik, C. S. Haley *et al.*, 2016
632 Evaluating the contribution of genetics and familial shared environment to
633 common disease using the UK Biobank. Nature Genetics.
634 Nordsletten, A. E., H. Larsson, J. J. Crowley, C. Almqvist, P. Lichtenstein *et al.*, 2016
635 Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric
636 Disorders. JAMA Psychiatry 73**:** 354-361.
637 Peyrot, W. J., M. R. Robinson, B. W. Penninx and N. R. Wray, 2016 Exploring
638 Boundaries for the Genetic Consequences of Assortative Mating for Psychiatric
639 Traits. JAMA Psychiatry 73**:** 1189-1195.
640 Philippe, P., 1978 Familial correlations of longevity: an isolate-based study. Am J Med
641 Genet 2**:** 121-129.
642 Robinson, M. R., A. Kleinman, M. Graff, A. A. Vinkhuyzen, D. Couper *et al.*, 2017
643 Genetic evidence of assortative mating in humans. Nature Human Behaviour
644 1**:** 0016.
645 Schulze, R., 2004 *Meta-analysis-A comparison of approaches*. Hogrefe Publishing.
646 Silventoinen, K., J. Kaprio, E. Lahelma, R. J. Viken and R. J. Rose, 2003 Assortative
647 mating by body height and BMI: Finnish twins and their spouses. American
648 Journal of Human Biology 15**:** 620-627.
649 Stulp, G., M. J. Simons, S. Grasman and T. V. Pollet, 2016 Assortative mating for
650 human height: A meta-analysis. American journal of human biology.
651 Tenesa, A., K. Rawlik, P. Navarro and O. Canela-Xandri, 2015 Genetic determination
652 of height-mediated mate choice. Genome Biology 16**:** 1-8.
653 Xia, C., C. Amador, J. Huffman, H. Trochet, A. Campbell *et al.*, 2016 Pedigree- and
654 SNP-Associated Genetics and Recent Environment are the Major Contributors
655 to Anthropometric and Cardiometabolic Trait Variation. PLoS Genet 12**:**
656 e1005804.
657 Zietsch, B. P., K. J. Verweij, A. C. Heath and N. G. Martin, 2011 Variation in human
658 mate choice: simultaneously investigating heritability, parental influence, sexual
659 imprinting, and assortative mating. The American Naturalist 177**:** 605-616.

660

Figures and Tables

Figure 1: Effects of indirect assortative mating on heritability and correlations based on the model of (GIANOLA 1982) (see Supplementary Methods). We consider a pair of traits. One trait which is the target of assortment, e.g., BMI, and a genetically correlated focal trait, e.g., hypertension disease liability. Both traits are taken to have heritabilities of 0.3 in a random mating population. We illustrate relative changes in three genetic parameters as functions of the strength of assortative mating (ρ_{couple}) and genetic correlation in a random mating population between the traits (ρ_g). Specifically, (**a**) changes in heritability of the assortment trait, (**b**) changes in heritability of the focal trait and (**c**) changes in genetic correlation between the traits. In all three panels we plot the ratios of the parameter under assortment to random mating. We assume a population at equilibrium after assortative mating (which happens only after a few generations of assortment) relative to a random mating population. In (**b**) and (**c**) colors indicate the ratios of h^2 or ρ_g in the two populations. Specifically, red colors indicate areas where assortative mating leads to increased heritability in the focal trait and increased absolute genetic correlations, i.e., the ratio of h^2 or ρ_g after assortative mating to that in a random mating population is greater than one.

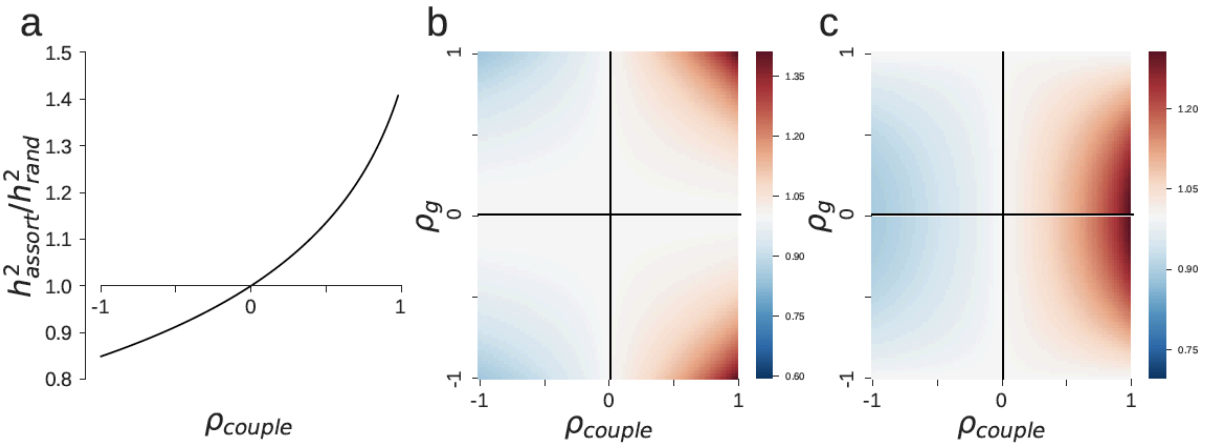
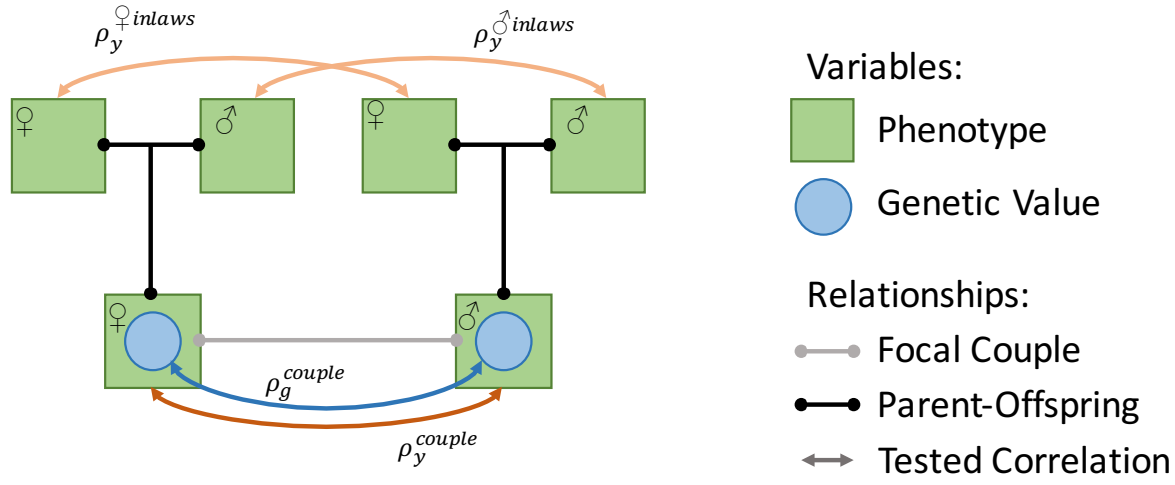


Figure 2: Schematic outline of the study. We consider couples and their parents. We compute phenotypic correlations between couples (ρ_y^{couple}) for longevity and disease status. Such correlations could be explained by the couple sharing a nuclear environment, e.g., shared exposures in the shared home or shared diet. In order to exclude the possibility of convergence based on shared nuclear environment, we examined parental correlations, that is correlations between the fathers ($\rho_y^{♂inlaws}$) and mothers ($\rho_y^{♀inlaws}$) of the partners. Such correlations cannot arise due to the nuclear couple environment, but require non-random mating and across generation correlations. The across generation correlations could arise due to heritable genetic effects or culturally transmitted environmental effects. We therefore also examined correlations in genetic values (ρ_g^{couple}), which provide evidence for non-random mating with respect to heritable factors.



693 **Table 1: Polychroic correlations for family history of fathers and mothers of couples in**

694 **the UK Biobank.**

	Father ($\rho_{y}^{\male\,inlaws}$)			Mother ($\rho_{y}^{\female\,inlaws}$)		
	ρ_{chor}	s.e.	*P*	ρ_{chor}	s.e.	*P*
Heart Disease	0.04	0.006	6×10^{-11}	0.07	0.007	9×10^{-23}
Stroke	0.02	0.009	0.003	0.06	0.009	2×10^{-11}
Lung Cancer	0.04	0.012	1×10^{-4}	0.08	0.018	1×10^{-5}
Bowel Cancer	0.04	0.015	0.009	-0.01	0.017	0.747
Breast Cancer	-	-	-	0.01	0.012	0.325
Chronic Bronchitis	0.06	0.01	2×10^{-9}	0.06	0.015	7×10^{-5}
High Blood Pressure	0.09	0.007	1×10^{-35}	0.08	0.006	7×10^{-38}
Diabetes	0.02	0.012	0.067	0.04	0.011	0.001
Alzheimer's	0.07	0.017	2×10^{-5}	0.08	0.011	3×10^{-13}
Parkinson's	0.02	0.027	0.267	0.04	0.034	0.13
Depression	0.03	0.022	0.103	0.04	0.014	0.005
Prostate Cancer	0.04	0.013	0.004	-	-	-

695 ρ_{chor} = polychoric correlation, s.e. = standard error, *P* = pvalue for $\rho_{chor} = 0$

Table 2: Within couple correlations of genetic values (ρ_g^{couple}) for family history and self-reported disease in genotyped couples in the UK Biobank.

	Parental Family History[1]		Self[2]		
	ρ	***P***	**ρ**	**95% CI**	***P***
Hypertension	0.03	8×10^{-6}	0.028	0.009-0.048	0.005
Chronic Bronchitis	0.019	0.07	0.011	-0.008-0.031	0.26
Heart Disease	0.016	9×10^{-3}	-0.015	-0.034-0.005	0.14
Stroke	0.013	0.12	0.004	-0.016-0.023	0.7
Diabetes	0.009	0.09	0.024	0.004-0.043	0.02
Prostate Cancer	0.009	0.34	-		-
Lung Cancer	0.005	0.32	-		-
Alzheimer's	0.004	0.27	-		-
Severe Depression	0.003	0.41	0.017	-0.002-0.036	0.09
Parkinson's	-0.001	0.42	-		-
Breast Cancer	-0.004	0.68	-		-
Bowel Cancer	-0.008	0.14	-		-

[1]meta-analysis of paternal and maternal results, with the exception of Prostate Cancer and Breast Cancer which are paternal and maternal results respectively, separate results for all disease can be found in Supplementary Table S10, [2]contains only results for self-reported non sex specific disease with UK Biobank prevalence > 5%, ρ = Pearson's correlation between genetic values in couples, *P* = pvalue for ρ=0