Intelligence 59 (2016) 103–108

Contents lists available at ScienceDirect

Intelligence

Assortative mating on educational attainment leads to genetic spousal resemblance for polygenic scores

David Hugh-Jones [a,*,1], Karin J.H. Verweij [b,1], Beate St. Pourcain [c], Abdel Abdellaoui [b,*]

[a] Department of Economics, University of East Anglia, Research Park, Courtyard B, Norwich NR4 7TJ Norwich, England, United Kingdom
[b] Department of Biological Psychology, VU University, van der Boechorstraat 1, 1081 BT Amsterdam, The Netherlands
[c] Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 7 June 2016
Received in revised form 25 July 2016
Accepted 14 August 2016
Available online 29 August 2016

Keywords:
Assortative mating
Educational attainment
Polygenic scores

ABSTRACT

We examined whether assortative mating for educational attainment ("like marries like") can be detected in the genomes of ~1600 UK spouse pairs of European descent. Assortative mating on heritable traits like educational attainment increases the genetic variance and heritability of the trait in the population, which may increase social inequalities. We test for genetic assortative mating in the UK on educational attainment, a phenotype that is indicative of socio-economic status and has shown substantial levels of assortative mating. We use genome-wide allelic effect sizes from a large genome-wide association study on educational attainment (N ~ 300 k) to create polygenic scores that are predictive of educational attainment in our independent sample ($r = 0.23$, $p < 2 \times 10^{-16}$). The polygenic scores significantly predict partners' educational outcome ($r = 0.14$, $p = 4 \times 10^{-8}$ and $r = 0.19$, $p = 2 \times 10^{-14}$, for prediction from males to females and vice versa, respectively), and are themselves significantly correlated between spouses ($r = 0.11$, $p = 7 \times 10^{-6}$). Our findings provide molecular genetic evidence for genetic assortative mating on education in the UK.

1. Introduction

Humans generally do not choose their mates randomly. In search for a suitable mate, among the highest-ranking qualities people look for in a potential partner are intelligence and educational attainment (Buss and Barnes, 1986; Zietsch, Verweij, and Burri, 2012). Previous work consistently shows substantial assortative mating for intelligence and educational attainment, with spousal correlations for intelligence ranging between 0.33 and 0.72 (Bouchard and McGue, 1981; Gualtieri, 2013; Mascie-Taylor and Vandenberg, 1988; Watson et al., 2004) and for educational attainment between 0.45 and 0.66 (Abdellaoui et al., 2015; Conley et al., 2016; Watson et al., 2004; Zietsch, Verweij, Heath, and Martin, 2011). Assortative mating can occur via different mechanisms (which are not always mutually exclusive). Partners can become more similar to each other over the course of their relationship (i.e., convergence); however, there is no evidence for convergence for cognitive abilities and educational attainment (Mascie-Taylor and Vandenberg, 1988; Watson et al., 2004; Zietsch et al., 2011). This suggests that assortative mating for educational attainment is due to initial partner choice. This can happen because of social homogamy, where similar people find themselves in similar social environments because of their social background, and/or because of phenotypic matching, where people select their partner based on similarity in characteristics.

The consequences of assortative mating on education and cognitive abilities are relevant for society and for the genetic make-up and therefore the evolutionary development of subsequent generations (Thiessen and Gregg, 1980). Assortative mating increases the variance of characteristics in the population, and may increase social inequality with respect to education or income (Schwartz, 2013). Greenwood, Guner, Kocharkov, and Santos (2014) for instance reported a rise in assortative mating for educational attainment in the United States between 1960 and 2005 and showed that this clustering of academic success may have caused an increase in income inequality. It is a priori very plausible that phenotypic similarity between partners on heritable traits is reflected in their genomic similarities, and thus in the genetic composition of their offspring. Assortative mating on a heritable trait increases the additive genetic variance for genetic loci associated with that trait, as well as for other traits that are genetically correlated with it (Crow and Felsenstein, 1968; Fisher, 1918; Lande, 1977), as assortative mating generates phenotypes with more extreme genetic values. The increase in assortment for educational attainment (Greenwood et al., 2014; Schwartz, 2013) may explain why heritability estimates for educational attainment have risen over time (Branigan, McCallum, and Freese, 2013), although there may also be other explanations for this increase, such as the recently increased equality in educational

* Corresponding authors.
E-mail addresses: davidhughjones@gmail.com (D. Hugh-Jones), a.abdellaoui@vu.nl (A. Abdellaoui).
[1] Joint first authors.

http://dx.doi.org/10.1016/j.intell.2016.08.005
0160-2896/

opportunities (Colodro-Conde, Rijsdijk, Tornero-Gómez, Sánchez-Romera, and Ordoñana, 2015). Another genetic consequence of assortative mating on education is the influence on genome-wide ancestral variation and homozygosity. Abdellaoui et al. (2015) showed that more educated individuals are more likely to migrate, which increases their chance of meeting a spouse with a different ancestral background. Accordingly, assortment on educational attainment can result in greater ancestral variation and lower levels of genome-wide homozygosity (a genetic signature used to study effects of inbreeding) in the offspring of higher educated spouse pairs.

Several studies have tried to detect assortative mating on a molecular genetic level by estimating spousal resemblance on genome-wide single nucleotide polymorphisms (SNPs) (Domingue, Fletcher, Conley, and Boardman, 2014; Guo, Wang, Liu, and Randall, 2014; Sebro, Hoffman, Lange, Rogus, and Risch, 2010). These studies report spouses to be more similar on genome-wide SNPs than expected under random mating. However, these reported spousal resemblances are more likely to be explained by population stratification, i.e., spouse pairs sharing more ancestry than random male-female pairs (Abdellaoui, Verweij, and Zietsch, 2014; Sebro et al., 2010), than by phenotypic assortative mating. Assortative mating on complex phenotypes, such as education, intelligence, personality, psychiatry, or height, is expected to lead to genetic spousal resemblance. However, these traits are influenced by many genetic variants throughout the genome with very small individual effects that require exceptionally large sample sizes to detect. The largest patterns of genome-wide variation, which can be captured with principal component analyses (PCA) in much smaller datasets, reflect ancestry differences (Price et al., 2006), correlate strongly with geography, and show significant spouse correlations (Abdellaoui et al., 2013b). Geographic proximity is a strong predictor of shared ancestry and a major determinant of potential spouse pairs meeting, especially in the presence of social catalysts that narrow mate choice and correlate with geography, such as religion (Abdellaoui et al., 2013a; Haber et al., 2013). We therefore expect spousal resemblance on a genome-wide level to be dominated by shared ancestry, and indeed the above studies do not show a significant genetic spousal resemblance once ancestry is appropriately accounted for. A trait-based approach is more powerful, less susceptible to population stratification, and thus more informative in detecting genetic assortative mating than estimating allelic spousal resemblance in a hypothesis-free manner. With the advent of large-scale genome-wide association studies (GWASs), we can now quantify significant portions of a person's genetic predisposition for a wide range of traits with polygenic scores by summing their individual alleles weighted by their estimated effect sizes. Polygenic scores can have significant predictive power and generally improve for complex traits when adding SNPs that individually did not reach genome-wide significance (Dudbridge, 2013).

The highly polygenic trait educational attainment is well suited for a study on genetic assortative mating because the phenotype itself is subject to high levels of assortment and genome-wide estimates of allelic effect sizes are available from large GWASs. Conley et al. (2016) show that polygenic scores based on results from a GWAS on educational attainment of ~126,000 participants (Rietveld et al., 2013) significantly correlate between spouse pairs born between 1920 and 1950 in the US. We use genome-wide effect sizes from a GWAS on educational attainment of ~300,000 participants (Okbay et al., 2016) to create polygenic scores for couples born between 1919 and 1994 from the UK Household Longitudinal Study (UKHLS), a survey that aims to be representative of the UK population. Given similar levels of phenotypic assortative mating in the US and the UK, we expect to replicate that there is genetic assortative mating for educational attainment and to find higher levels of genetic assortative mating than Conley et al. (2016) given the more accurate summary statistics and a novel and more powerful polygenic score approach (Vilhjálmsson et al., 2015). We test whether individuals' polygenic risk scores for educational attainment can predict their partners' educational attainment, and their partners' polygenic scores. We control for similarities in ancestral background by taking into account ancestry-informative principal components (PCs).

2. Materials and methods

2.1. Phenotypes

The sample is derived from the UK Household Longitudinal Study: *Understanding Society* (UKHLS) (Buck and McFall, 2011), a representative sample of the UK population. 9944 individuals were genotyped, including 1699 pairs who were living together either as husband and wife or as a couple. Individuals under 25 years of age were removed from the analyses, because they are likely to not have reached their final education level; this resulted in an N of 8989. For the cross-spouse analyses we also removed all pairs where either partner was under 25, resulting in a sample of 1616 spouse pairs.

We derived a variable for individuals' educational attainment as follows: 0 = no educational qualifications; 1 = GCSE (national exams taken at age 16) or "other qualifications"; 2 = A-level or equivalent (national exams taken at age 18, roughly equivalent to French Baccalaureate or US High School Diploma); 3 = University degree or equivalent. Educational attainment was standardized to have a mean of 0 and a standard deviation of 1.

The UKHLS is a stratified probability sample of the UK population. The dataset for the nurse visit sample (from which the SNP data are derived) includes response weights which are meant to account for ascertainment bias and non-response, including non-participation in the nurse visit and not donating blood. We used the cross-sectional weights, i.e., the reciprocal of the probability of blood measures to be present for a particular individual, predicted from a variety of socio-economic characteristics. Further details are given in Benzeval, Davillas, Kumari, and Lynn (2014). For analyses where each case represents a pair of partners, such as the main regressions on partner characteristics, we used the arithmetic mean of male and female partner's weight.

2.2. Genotyping, quality control (QC), and principal component analysis (PCA)

Genotyping was done on the Illumina HumanCoreExome chip for White/European participants of Waves 2 and 3 of the *Understanding Society* study. QC was performed on the entire set of 9944 participants in PLINK (Purcell et al., 2007), and only autosomal SNPs were included. SNPs were excluded if they: 1) had a missing rate >5%; 2) showed a minor allele frequency (MAF) smaller than 5%; 3) deviated from Hardy–Weinberg equilibrium (HWE) with a p-value smaller than 10^{-8}. The QC resulted in 261,965 SNPs with a mean individual genotyping rate of >99.9% (ranging from 97.2% to 99.99%, with only 15 individuals having >1% missingness). There were no individuals detected with a non-European or non-British ancestry by projecting principal components (PCs) from the 1000 Genomes dataset (procedure described in more detail in the supplementary material of Abdellaoui et al. (2013b)). To control for ancestry differences within the UK, we conducted a PCA on the genotype data in EIGENSTRAT (Price et al., 2006). In order to detect the relatively small ancestry differences within the UK, we pruned for linkage disequilibrium (LD) (window size = 50, number of SNPs to shift after each step = 5, based on a variance inflation factor [VIF] of 2) and removed long-range LD regions, since LD can result in larger patterns of variation than ancestry differences within relatively homogeneous populations (Abdellaoui et al., 2013b). After minimizing LD, 91,708 autosomal SNPs remained. The PCA was conducted on unrelated individuals (9091 out of 9944 participants) and projected onto the rest. Unrelated individuals were chosen using GCTA (Yang, Lee, Goddard, and Visscher, 2011), by excluding one of each pair of individuals with an estimated genetic relationship of >0.025 (i.e., closer related than third or fourth cousin).