bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Analysis of genetic similarity among friends and schoolmates in the National Longitudinal Study of Adolescent to Adult Health (Add Health)

Benjamin W. Domingue[a*]
Daniel W. Belsky[b]
Jason M. Fletcher[c]
Dalton Conley[d]
Jason D. Boardman[e]
Kathleen Mullan Harris[f]

[a] Stanford University Graduate School of Education
[b] Duke University School of Medicine & Social Science Research Institute
[c] La Follette School of Public Affairs, Department of Sociology and Center for Demography and Ecology, University of Wisconsin-Madison
[d] Sociology Department, Princeton University
[e] Institute of Behavioral Science & Sociology Department, University of Colorado Boulder
[f] Sociology Department and Carolina Population Center, University of North Carolina

* bdomingue@stanford.edu

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genetic similarity among friends and schoolmates

Abstract

Humans tend to form social relationships with others who resemble them. Whether this sorting of like with like, homophily, may arise from historical patterns of migration, macro-level social structures in modern society, or individual-level peer selection remains unsettled. We analyzed data from 9,500 adolescents from the National Longitudinal Study of Adolescent to Adult Health to examine genetic similarities among pairs of friends. While there is some evidence that friends have correlated genotypes, both at the whole-genome level as well as at trait-associated loci (via polygenic scores), further analysis suggests that macro-level forces, such as school assignment, are a principal source of genetic similarity between friends. We also observe associations of an individual's educational attainment and the polygenic scores of those in their broader social environment (e.g., school) and of their friends (net of their own score). In contrast, individual BMI and height are largely unassociated with the genetics of peers.

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genetic similarity among friends and schoolmates

Introduction

The degree to which genetics are implicated in the formation and consequences of social relationships is of growing interest to the new field of sociogenomics (1,2). Analysis of the genotypes of spouses suggests that spouses are more genetically similar to one another as compared to random pairs of individuals in the population (3–8). The degree of this genetic "homogamy" is modest. In previous analyses, we estimated that genetic homogamy was about one-third the magnitude of educational homogamy (3), even when looking specifically at education-associated genotypes (6). But even modest genetic homogamy can have implications for statistical and medical genetic models of inheritance and social models of spousal effects (9–11).

Marriage is not the only social grouping which may demonstrate genetic homophily. It has been suggested elsewhere that adult friends are, on average, more genetically similar than random pairs from the population (12). Genetic similarity among friendship networks is important for at least two reasons. First, social friendship networks are often the source of mates in modern society (if not our friends themselves, then 2nd degree connections) where arranged marriages are no longer the norm. Second, there may exist social genetic effects—the effects of alter's genotype on ego's phenotype (1,13,14)—which would further suggest that social sorting on genotype may have consequences on the distribution of phenotypes in a population beyond its effect on subsequent generations through assortative mating.

Adolescence is not only an important development time in the lifecourse for a number of social and behavioral phenotypes ranging from health behaviors and mental health (15) to socioeconomic attainment (16,17), it is also a time of heightened salience for peer networks and influence (18–20). For these reasons, in the present study we characterize genetic homophily within adolescent social networks in the United States. Specifically, we analyze data from the National Longitudinal Study of Adolescent to Adult Health (Add Health). Add Health surveyed 90,118 US adolescents aged 12-18 in 1994-1995 using a school-based sampling frame. As part of the survey, students were asked to list the names of their friends. Responses were collated within schools to identify social ties between individuals and their friends (21). Of the adolescents surveyed, 20,745 were enrolled in a longitudinal study that included in-home interviews with the adolescents and their parents and that followed the adolescents prospectively across four assessments spanning 14 years. At the most recent assessment in 2008, ~12,000 Add Health participants provided DNA for genotyping and genome-wide single-nucleotide polymorphism (SNP) data were assayed. We linked these genetic data with social network information from the original school-based surveys along with information about personal characteristics and social environments accumulated across Add Health follow-up. Complete details on the data are in Section 1 of the Supplemental Information (SI).

Our analysis proceeds in three steps. In Step 1, we tested if friends were more genetically similar to one another than to randomly selected peers. In Step 2, we sought to disentangle potential sources of any genetic similarity between friends. In Step 3, we evaluated a potential

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

implication of genetic similarity among friends: social genetic effects, or the association between the genotypes of one's social peers and one's own phenotype (net of own genotype).

Results

1. Are friends more genetically similar to one another than they are to randomly selected peers?

We tested if friends were more genetically similar to one another as compared to random pairs of individuals using two approaches. First, we tested if genomes of friends were more similar as compared to genomes of random pairs of individuals. We used KING (22) and REAP (23) to compute genetic kinships between all pairs of Add Health participants. We then compared kinships among friends to kinships among random pairs of individuals to estimate the degree of genetic similarity among friends (3). To address potential confounding of analysis by ancestry, we focus first on a group of genetically homogeneous respondents of European ancestry. We then consider a similar result using the full set of respondent based on kinship estimates from REAP, which are more robust to population stratification.

Estimates of genetic similarity among friends were positive (**Table 1**). Among non-Hispanic whites in Add Health, KING estimates of genetic homophily were about 2/3 the magnitude of our previous KING estimates of genetic homogamy in the US Health and Retirement Study (friend similarity=0.031, CI=0.021-0.038, as compared to spousal similarity =0.045 from (3)). REAP estimates of genetic similarity were slightly smaller.

Our second analysis tested if friends were more similar to one another on specific trait-related genetic dimensions. We analyzed polygenic scores for three phenotypes: height, body-mass index, and educational attainment. Polygenic scores are genome-wide summaries of genetic influence. They are computed by weighting alleles at loci across the genome according to their association with a phenotype of interest and then summing weighted allele counts across loci. We computed polygenic scores based on weights from published genome-wide association studies (GWAS) (24–26) using established methods (27). Because the original GWAS were performed on non-Hispanic whites, we restricted polygenic score analysis to this population as population stratification may dilute genetic associations (28). To correct for any residual population stratification, we adjusted polygenic score analyses for the first 10 principal components computed based on the genetically homogeneous set of European-ancestry respondents (29). Further details are reported in the SI.

We first confirmed polygenic scores were associated with their respective phenotypes (all polygenic score-phenotype correlations were ~0.25; see Section 1 of SI). Next, for each score, we computed associations of an Add Health participant's polygenic score with the average polygenic score among their friends. Polygenic scores for body-mass index (r=0.05, p=0.03; **Table 2**) and educational attainment (r=0.06, p=0.002) were positively correlated among friends. A weaker correlation was observed for height (r=0.02, p=0.3). In contrast, phenotypes were more strongly correlated between friends, especially for educational attainment (r=0.4 for educational

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

attainment, r=0.09 for BMI, r=0.13 for height). Correlations among polygenic scores between friends were generally smaller than published correlations between polygenic scores among spouses (6), especially for height. Associations amongst same-sex friends were generally stronger in females (SI Section 2).

2. Why are friends more genetically similar to one another than they are to randomly selected peers?

We considered two hypotheses. One hypothesis is that friends are more genetically similar to one another because they form their friendships partly because they share similar phenotypes and family backgrounds (e.g. being short or tall, heavy or slim, well-educated families or poorly-educated families). This process is called "social homophily" (30–33). Because phenotypes that influence human social tie formation tend to be heritable (34), the first hypothesis suggests that social homophily can generate genetic similarity between friends. A second hypothesis is that friends are more genetically similar because they form their friendships within socially stratified environments (e.g. living in the same community, attending the same school). We refer to this process as social structuring (35). Because genetic factors influence selection into social environments (36), the second hypothesis suggests that social structuring can generate genetic similarity between friends even without explicit selection on phenotypic similarity. Social structuring and social homophily are not mutually exclusive and may indeed be complementary processes. We tested for evidence of social structuring processes by estimating genetic similarity at the level of the social environment. Then, in cases where we found evidence of social structuring, we tested for complementary homophily by re-estimating genetic similarity between friends within social environments.

To test the social structuring hypothesis, we asked about the degree to which schoolmates were genetically similar, when measured at the whole-genome level, using the same technique as above with friends (but replacing being in the same school with being friends as the relevant pair-level indicator). Add Health participants who were schoolmates were genetically more similar—over half the genetic similarity of friends, in fact—to one another as compared to random pairs of individuals (e.g., for European school-mates=0.022, CI=0.021-0.023; see **Table 1**).

We therefore repeated our analysis of genetic similarity among friends, this time comparing friend-level similarity to random pairs of individuals selected from within their social environment. Within-school analysis tested if genetic similarity among friends observed in our original analysis was fully explained by the social structuring hypothesis. Friend-level genetic similarity was substantially reduced when the analysis compared friends to random pairs of individuals from within the same school (KING-estimated within-school friend similarity=0.016, CI=0.004-0.026; **Figure 1A**). Despite our attempts to control for population stratification, the possibility of our results being inflated due to population structure remains (8,37). However, we think that inclusion of these results is illuminating as it helps to illustrate the role of school-based social proximity.

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genetic similarity among friends and schoolmates

We next repeated polygenic score analysis. Within-school polygenic score analysis tested correlations between Add Health participants' polygenic scores and the average polygenic scores of their friends after adjusting for the school-average polygenic score. Friend-mean polygenic scores for height and BMI do not significantly predict an individual's own score either before or after controlling for the school-mean score in these analyses (**Figure 1B**; note that the confidence intervals for these estimates adjust for school-level clustering). While it is the case an individual's friends are predictive of the individual's own polygenic score for educational attainment, this effect is greatly attenuated once we account for school assignment. After adjusting for school-mean polygenic score, friends' polygenic scores for educational attainment were no longer predictive (results were comparable outside of the saturation school, see Section 3 of SI). We also explored the sensitivity of these findings to remaining population structure amongst the respondents of European ancestry that we focus on here. Results are comparable in even more genetically homogeneous groups of respondents (Section 3 of SI).

3. Is the social genome associated with an individual's phenotype?

Social genetic effects refer to influence of genotypes in a focal organism's social environment on the focal organism's phenotype (13,14). Social genetic effects may take several forms (1). We tested (a) social-genetic main effects, that is, whether friend genotypes were associated with a focal individual's phenotype net of that focal individual's own genotype, and (b) social epistasis, if the association between a focal individual's own genotype and phenotype was moderated by the genotypes of their social (i.e. schoolmate) environment.

First, we tested if average friend polygenic scores predicted the focal Add Health participant's height, BMI, and educational attainment measured at the Wave IV Add Health follow-up, 13 years after social network information was collected. Friend polygenic scores for educational attainment, but not height or BMI, were associated with Add Health participants' respective phenotypes at the Wave IV assessment (for educational attainment $b=0.15$, $p<0.001$; Section 4 of SI).

Next, we tested if any social genetic main effects might be accounted for by social genetic correlation, i.e. similar polygenic scores among friends. This analysis included statistical adjustment for the focal participant's own polygenic score. After adjusting for the focal participant's polygenic score, social genetic main effects were reduced but, for educational attainment, remained statistically significant (social genetic main effect $b=0.14$, $p<0.001$ for educational attainment). These social genetic main effects are summarized in **Figure 2**.

We also tested social genetic main effects at the school level after accounting for the individual's own polygenic score. Findings were similar in nature (but stronger in magnitude) to those at the friend-level ($b=0.21$, $p<0.001$ for educational attainment). Add Health participants who attended schools with other students who had higher polygenic scores tended to complete more education over the subsequent 13 years even after accounting for their own polygenic score.

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genetic similarity among friends and schoolmates

To test if school-level social genetic main effects might account for social genetic main effects among friends, we re-estimated friend-level social genetic main effects using a within-school design. This analysis included statistical adjustment for school-average polygenic scores using school fixed effects. For educational attainment, friend-level social genetic main effects remained statistically significant even after accounting for school-level social genetic main effects (within-school social genetic main effect b=0.08, p=0.004). Motivated by recent work (8), we considered the robustness of our findings to the inclusion of information about overall genetic relatedness. Results are similar to those reported here (section 4 of SI). Results were similar when we considered estimates outside of the saturation school and with only mutual friendship nominations. Finally, we considered results based on only those students attending schools with more than 80% self-identified white respondents. Estimates were slightly attenuated but qualitatively similar.

We additionally tested for "social epistasis", i.e. variability in genetic main effects according to the genetic composition of a focal individual's social environment. We tested social epistasis as interaction between Add Health participant's own polygenic scores and average polygenic scores for their friends and schoolmates. We found no evidence of social epistasis in these tests (results are in Section 4 of SI).

Discussion

We present novel evidence for positive genetic similarity amongst friends from a nationally representative sample of adolescents in the US. Our findings echo and extend earlier work suggesting that adult friends exhibit overall genetic similarity (12) and that adolescents exhibit similarity on specific candidate genes (38,39). This conclusion is supported by evidence of genetic similarity as measured both at the genome-wide level although we may be underpowered to detect genome-wide similarities (8) and by polygenic scores. We also observe the clustering of specific genetic polymorphisms across schools. We estimate that a relatively small proportion (<3%, **Table 2**) of the variation in the educational attainment polygenic score is between schools. However, even this small amount of clustering has implications for our understanding of genetic stratifications across human relationship. While it is possible that genetically similar individuals actively select into friendships because of common observable traits with genetic origins; the social structuring hypothesis also seems to be relevant. Similar results have been shown for demographic characteristics of romantic partners (40). Even within a school, further social structuring may be possible. Educational tracks, for example, are not assigned randomly (41), have long-lasting consequences, and are potential sources for friend formation. If educational tracks are a source of further environmental concentration within grades, they may be a mechanism leading to the observed association between friends' genotypes.

Social, or indirect, genetic effects are accounted for in evolutionary theory (42,43) and have been observed among animals (44–46) and siblings (13,47). In mice, evidence of social

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genetic effects has been found for a broad variety of phenotypes, ranging from psychological phenotypes (e.g., anxiety) to biomedical phenotypes such as wound healing (14). If the effect of a certain allele may extend beyond the organism that carries it, estimates of direct genetic effects may be biased when social genetic effects are unaccounted for (14). Thus, evidence for the existence of social genetic effects would raise important questions about the meaning and interpretation of genetic "effects." However, it is difficult to identify causes for effects propagated via social networks. Animal studies can use aspects of randomization (14), but this is not generally possible in most studies of human social interaction, although there are exceptions (48). Previous attempts to identify causal effects (49) have been met with skepticism (50). In observational studies of friends, the fundamental problem is that the friendship bond is endogenous (with respect to the characteristics of the individuals in the friendship), so it is challenging to disentangle the effects of the friendship from effects associated with the individual characteristics of the individuals in said friendship (51).

Results from our analyses suggest the possibility of social genetic effects in specific domains. Of the traits we consider (educational attainment, height, and BMI), the clearest null finding related to social genetic effects comes from the phenotype—height—that seems least likely to be influenced by broader social processes, at least in contemporary US society. Interpretation of the findings related to BMI and educational attainment is challenging given the difficulties of causal inference in network settings. The genetics of one's broad set of peers could be associated with a historical set of processes regarding school funding, for example, and/or the set of policies and social forces that have led to high levels of residential segregation by ancestry in the US (52). Given that students are not randomly assigned to schools and that, once in schools, students are not randomly assigned to friendships—indeed, earlier research has demonstrated the "heritability" of friend GPA (53)—we are unable to fully distinguish between these mechanisms. In order to be more definitive, future research may require exogenous mechanisms underlying social contact (48).

This study contributes to what is currently known about the genetic ecology that exists in human societies. We have provided specific evidence about both gene-environment correlations and the potential for social genetic effects, and noted that the joint existence of these two genetic features could produce important feedback loops. If the genetics of one's social environment matter *and* the relevant genetics are stratified across environments, then being in certain environments might substantially modulate genetic signals. More work is needed to document both features in humans before such modulations can be unambiguously documented, but their possible existence challenges simple notions of genetic "effects."

Methods & Materials

Data

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genetic similarity among friends and schoolmates

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a nationally representative cohort drawn from a probability sample of 80 US high schools and 52 US middle schools (in roughly 90 US communities), representative of US schools in 1994–95 with respect to region, urban setting, school size, school type, and race or ethnic background. About 15,000 respondents (or 96%) consented to genotyping during the Wave 4 interview in 2008–09 for purposes of approved Add Health Wave 4 research. Of those who consented to genotyping, ~12,000 (or 80%) agreed to have their DNA archived for future testing (see SI for a comparison of genotyped and non-genotyped respondents). DNA extraction and genotyping on this archived sample using two platforms yielded a sample of 9,696 Add Health members with GWAS data consisting of 631,990 SNPs. The SI contains additional details on the genotyping process.

In the in-school survey (which was administered to every student in the participating schools, not only the Add Health study members who are prospectively followed into adulthood), as well as in the in-home surveys at waves 1 and 2, students were asked to nominate up to 5 of their male friends and 5 female friends. We accept a nomination in either direction (i.e. "undirected" friendships) as evidence of a friendship between two individuals. Of those with genetic data, only 5,128 people were in a friendship pair with another genotyped respondent. We focus largely on friendship nominations within race/ethnicity. Of the 7,259 friendship pairs between genotyped respondents, ~90% were within self-reported race/ethnicity. We emphasize two additional caveats. First, related respondents (as identified by measures of genetic similarity) were not included in the analyses. Second, one school—a so called "saturated school" in the Add Health data—contributed a disproportionate number of friend pairs to the sample of non-Hispanic white respondents. Results reported are robust to the removal of this school.

Measures

To measure genetic similarity, we use a kinship measures (KING; 11) that has been the focus of earlier research (3,5). We also consider an alternative measure, REAP (23), that is less sensitive to population stratification. We construct principal components using all genotyped respondents. We construct polygenic scores for anthropometric traits (BMI, height) and educational attainment. To construct these scores, we utilize publicly available genome-wide association study results (24–26) derived from large consortia studies (that did not include Add Health) of populations of European descent. Polygenic scores do not typically generalize across racial groups (28), so we focus these analyses on a group of genetically homogeneous European-ancestry respondents.

We use information from Wave 4 (when respondents were 24–32 years old) on years of education completed, BMI, and height (see additional details on all measures in SI).

Methods

Overall Genetic Similarity

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Genetic similarity among friends and schoolmates

We first consider a measure of genetic similarity amongst friends used previously to study genetic similarity amongst spouses (3,5). We compute the area between the 45 degree line and the P-P plot (given two CDFs F and G, the P-P plot is the set of points (F(x),G(x)) for all x) comparing the density of genetic similarity between friends with the density of genetic similarity for all dyads.

Targeted Genetic Similarity

We first consider a baseline model for some polygenic score G_i of the form

$$G_i=a+b\bullet\mu_F(G_i)+e_i. \text{ (M1)}$$

where $\mu_F(G_i)$ is the mean of G_i for all an individual's friends. The coefficient b captures the degree to which being friends is associated with levels of overall genetic similarity. Motivated by previous studies of the Add Health social networks (54,55), we next consider the role of genetic clustering into schools in the observed degree of friend genetic similarity via

$$G_i=a+b\bullet\mu_F(y_i)+c\bullet\mu_S(G_i)+e_i. \text{ (M2)}$$

where $\mu_S(G_i)$ is the mean of G_i for all other individual's at the school of individual i. We interpret attenuation in estimates of b going from M1 to M2 as evidence for the importance of social structure, specifically school assignment, in observed genetic similarity amongst friends. We adjust all standard errors for clustering of students into schools (56).

Turning to our analysis of social genetic effects, we first consider models of the (potentially confounded) effect of the social genome on an individual's phenotype (P_i) net of the individual's own polygenic score (PGS_i). The social genome will be characterized via $\mu_F(G_i)$ and $\mu_S(G_i)$ as above. We first consider

$$P_i=a+b\bullet PGS_i+c\bullet\mu_F(G_i)+e. \text{ (M3)}$$

We construe this as a test of "narrow" social genetic effects of friends. In contrast, we also consider measures of "broad" social genetic effects:

$$P_i=a+b\bullet PGS_i+c\bullet\mu_S(G_i)+e, \text{ (M4)}$$

Again, all standard errors are adjusted for clustering into schools.

Acknowledgments: This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). We gratefully acknowledge funding from NICHD to Boardman and Harris through grants R01HD073342 and R01HD060726. Belsky is supported by a Jacobs Foundation Early Career Research Fellowship and by NIA grants R01AG032282 and P30AG028716. Conley is supported by a Russell Sage Foundation grant on "GxE and Health Inequality across the Lifecourse" (83-15-29). This research benefitted from GWAS results made publicly available by the SSGAC and GIANT consortia. We would like to acknowledge

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Matthew Robinson for helpful comments and Robbee Wedow, Joyce Tabor, and Christy Avery for help with the Add Health genetic sample.

References

1. Domingue BW, Belsky DW. The social genome: Current findings and implications for the study of human genetics. PLoS Genet. 2017;13(3):e1006615.

2. Conley D, Fletcher J. The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future. Princeton: Princeton University Press; 2017.

3. Domingue BW, Fletcher J, Conley D, Boardman JD. Genetic and educational assortative mating among US adults. Proc Natl Acad Sci. 2014;111(22):7996–8000.

4. Guo G, Wang L, Liu H, Randall T. Genomic Assortative Mating in Marriages in the United States. 2014 [cited 2015 Sep 12]; Available from: http://dx.plos.org/10.1371/journal.pone.0112322

5. Zou JY, Park DS, Burchard EG, Torgerson DG, Pino-Yanes M, Song YS, et al. Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. Proc Natl Acad Sci. 2015;112(44):13621–13626.

6. Conley D, Laidley T, Belsky D, Fletcher J, Boardman J, Domingue B. Assortative mating and differential fertility by phenotype and genotype across the 20th century. Proc Natl Acad Sci. 2016;

7. Hugh-Jones D, Verweij KJ, Pourcain BS, Abdellaoui A. Assortative mating on educational attainment leads to genetic spousal resemblance for causal alleles. Intelligence. 2016;

8. Robinson MR, Kleinman A, Graff M, Vinkhuyzen AAE, Couper D, Miller MB, et al. Genetic evidence of assortative mating in humans. Nat Hum Behav. 2017 Jan 9;1:0016.

9. Trikalinos TA, Salanti G, Khoury MJ, Ioannidis JP. Impact of violations and deviations in Hardy-Weinberg equilibrium on postulated gene-disease associations. Am J Epidemiol. 2006;163(4):300–309.

10. Peyrot WJ, Robinson MR, Penninx BWJH, Wray NR. Exploring Boundaries for the Genetic Consequences of Assortative Mating for Psychiatric Traits. JAMA Psychiatry. 2016 Oct 12;

11. Richmond-Baker L, Belsky D. Swedish Register Analysis of Divorce and Alcohol Use Disorder Highlights Social Relationships as a Target for Preventive Psychiatry and Genetic Research. Am J Psychiatry. In press;

12. Christakis NA, Fowler JH. Friendship and natural selection. Proc Natl Acad Sci. 2014 Jul 22;111(Supplement 3):10796–801.

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

13. Rauscher E, Conley D, Siegal ML. Sibling genes as environment: Sibling dopamine genotypes and adolescent health support frequency dependent selection. Soc Sci Res. 2015;54:209–220.

14. Baud A, Mulligan M, Casale F, Ingels J, Bohl C, Callebert J, et al. Genetic variation in the social environment contributes to health and disease. PLOS Genet. 2017;

15. Resnick MD, Bearman PS, Blum RW, Bauman KE, Harris KM, Jones J, et al. Protecting adolescents from harm: findings from the National Longitudinal Study on Adolescent Health. Jama. 1997;278(10):823–832.

16. Crosnoe R, Mistry RS, Elder GH. Economic disadvantage, family dynamics, and adolescent enrollment in higher education. J Marriage Fam. 2002;64(3):690–702.

17. Leventhal T, Brooks-Gunn J. The neighborhoods they live in: the effects of neighborhood residence on child and adolescent outcomes. Psychol Bull. 2000;126(2):309.

18. Mollborn S, Domingue BW, Boardman JD. Understanding multiple levels of norms about teen pregnancy and their relationships to teens' sexual behaviors. Adv Life Course Res. 2014;20:1–15.

19. Maxwell KA. Friends: The role of peer influence across adolescent risk behaviors. J Youth Adolesc. 2002;31(4):267–277.

20. Prinstein MJ, Boergers J, Spirito A. Adolescents' and their friends' health-risk behavior: Factors that alter or add to peer influence. J Pediatr Psychol. 2001;26(5):287–298.

21. Bearman P, Moody J, Stovel K. The Add Health network variable codebook. Univ N C. 1997;

22. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010;26(22):2867–2873.

23. Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N. Estimating kinship in admixed populations. Am J Hum Genet. 2012;91(1):122–138.

24. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat Genet. 2014 Nov;46(11):1173–86.

25. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. Nature. 2015 Feb 12;518(7538):197–206.

26. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. Nature. 2016;533(7604):539–542.

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

27. Conley D, Laidley TM, Boardman JD, Domingue BW. Changing Polygenic Penetrance on Phenotypes in the 20th Century Among Adults in the US Population. Sci Rep. 2016;6:30348.

28. Carlson CS, Matise TC, North KE, Haiman CA, Fesinmeyer MD, Buyske S, et al. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. PLoS Biol. 2013;11(9):e1001661.

29. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904–909.

30. Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, Mataix-Cols D. Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. JAMA Psychiatry. 2016 Apr 1;73(4):354–61.

31. Goodreau SM, Kitts JA, Morris M. Birds of a feather, or friend of a friend?: Using exponential random graph models to investigate adolescent social networks. Demograhpy. 2009;46(1):103–125.

32. Schwartz CR, Mare RD. The proximate determinants of educational homogamy: The effects of first marriage, marital dissolution, remarriage, and educational upgrading. Demography. 2012;49(2):629–650.

33. Youyou W, Schwartz HA, Stillwell D, Kosinski M. Birds of a Feather Do Flock Together: Behavior-Based Personality-Assessment Method Reveals Personality Similarity Among Couples and Friends. Psychol Sci. 2017;0956797616678187.

34. Polderman TJ, Benyamin B, De Leeuw CA, Sullivan PF, Van Bochoven A, Visscher PM, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet. 2015;47(7):702–709.

35. Zeng Z, Xie Y. A preference-opportunity-choice framework with applications to intergroup friendship. AJS. 2008;114(3):615.

36. Belsky DW, Moffitt T, Corcoran D, Domingue B, Harrington H, Hogan S, et al. The genetics of success: How SNPs associated with educational attainment relate to life course development. Psychol Sci. 2016;

37. Abdellaoui A, Verweij KJ, Zietsch BP. No evidence for genetic assortative mating beyond that due to population stratification. Proc Natl Acad Sci. 2014;111(40):E4137–E4137.

38. Fowler JH, Settle JE, Christakis NA. Correlated genotypes in friendship networks. Proc Natl Acad Sci. 2011 Feb 1;108(5):1993–7.

39. Boardman JD, Domingue BW, Fletcher JM. How social and genetic factors predict friendship networks. Proc Natl Acad Sci. 2012 Oct 23;109(43):17377–81.

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

40. Eastwick PW, Harden KP, Shukusky JA, Morgan TA, Joel S. Consistency and Inconsistency Among Romantic Partners Over Time. 2017;

41. Gamoran A, Mare RD. Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? Am J Sociol. 1989;94(5):1146–1183.

42. Moore AJ, Brodie III ED, Wolf JB. Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. Evolution. 1997;51(5):1352–1362.

43. Wolf JB, Brodie III ED, Cheverud JM, Moore AJ, Wade MJ. Evolutionary consequences of indirect genetic effects. Trends Ecol Evol. 1998;13(2):64–69.

44. Bergsma R, Kanis E, Knol EF, Bijma P. The contribution of social effects to heritable variation in finishing traits of domestic pigs (Sus scrofa). Genetics. 2008;178(3):1559–1570.

45. Benton CH, Delahay RJ, Robertson A, McDonald RA, Wilson AJ, Burke TA, et al. Blood thicker than water: kinship, disease prevalence and group size drive divergent patterns of infection risk in a social mammal. In: Proc R Soc B. The Royal Society; 2016. p. 20160798.

46. Petfield D, Chenoweth SF, Rundle HD, Blows MW. Genetic variance in female condition predicts indirect genetic variance in male sexual display traits. Proc Natl Acad Sci. 2005;102(17):6045–6050.

47. Conley D, Rauscher E, Siegal ML. Beyond orchids and dandelions: testing the 5-HTT "risky" allele for evidence of phenotypic capacitance and frequency-dependent selection. Biodemography Soc Biol. 2013;59(1):37–56.

48. Guo G, Li Y, Wang H, Cai T, Duncan GJ. Peer Influence, Genetic Propensity, and Binge Drinking: A Natural Experiment and a Replication. Am J Sociol. 2015;121(3):914–954.

49. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med. 2007;357(4):370–379.

50. Cohen-Cole E, Fletcher JM. Detecting implausible social network effects in acne, height, and headaches: Longitudinal analysis. Br Med J. 2008;337:a2533.

51. Shalizi CR, Thomas AC. Homophily and contagion are generically confounded in observational social network studies. Sociol Methods Res. 2011;40(2):211–239.

52. Boustan LP. Racial residential segregation in American cities. National Bureau of Economic Research; 2013.

53. Barnes J, Beaver KM, Young JT, TenEyck M. A behavior genetic analysis of the tendency for youth to associate according to GPA. Soc Netw. 2014;38:41–49.

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

54. Daw J, Margolis R, Verdery AM. Siblings, friends, course-mates, club-mates: how adolescent health behavior homophily varies by race, class, gender, and health status. Soc Sci Med. 2015;125:32–39.

55. Balbo N, Barban N. Does Fertility Behavior Spread among Friends? Am Sociol Rev. 2014 Jun 1;79(3):412–31.

56. Zeileis A. Econometric computing with HC and HAC covariance matrix estimators. J Stat Softw. 2004;11(10):1–17.

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Similarity estimates based on overall estimates of genetic similarity (KING) and when using schoolmate status (attend same school?) in place of friendship status in calculation of similarity.

Social Proximity	Group	Relatedness Measure	N people	N friend pairs	Similarity	95% CI
Friends	European	KING	2918	4636	0.031	0.021-0.038
	European	REAP	2922	4626	0.024	0.016-0.031
	All	REAP	4997	7016	0.028	
Schoolmates	European	KING	5672	265056	0.022	0.021-0.023
	European	REAP	5649	264945	0.021	0.020-0.022
	All	REAP	9455	666119	0.018	

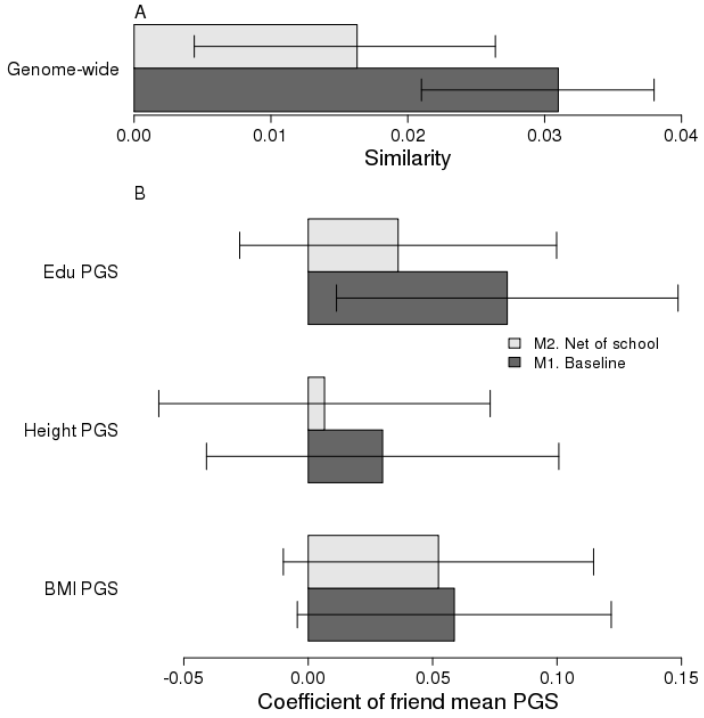
bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2. Correlations between social genotype (columns) and individual genotypes/phenotypes (rows) and ICC of individual genotypes within schools among non-Hispanic white respondents.

	Correlations, Individual and		ICCs
	Mean friend PGS	Mean schoolmate PGS	School
Edu PGS	0.063	0.139	0.023
BMI PGS	0.045	0.038	0.003
Height PGS	0.023	0.077	0.016
Education	0.155	0.242	0.177
BMI	0.044	0.065	0.020
Height	0.034	0.052	0.024

bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1. A. Genetic similarity between friends before and after accounting for genetic similarity of schoolmates. B. Comparison of genetic prediction between friends before (M1) and after (M2) adjusting for the genetic makeup of their school (via mean PGS).



bioRxiv preprint doi: https://doi.org/10.1101/107045; this version posted April 10, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

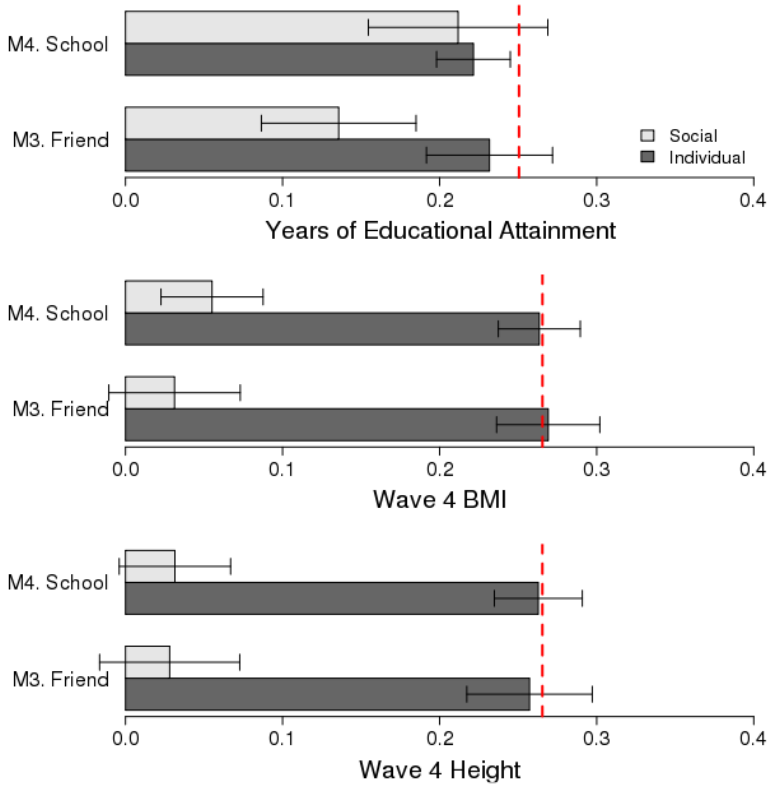



Figure 2. Social Genetic Effects: Effect of friend and school mean PGS net of one's own PGS for educational attainment, BMI, and height (all measured at Wave 4) on associated outcome. Outcomes are standardized as are social genotypes. The red line is the baseline effect of own PGS on the outcome in a null model with no other predictors. Estimates are based on a sample of unrelated respondents. SEs robust to school clustering.