MANKIND QUARTERLY 2021 62:1 186-216

Genetic Ancestry and General Cognitive Ability in a Sample of American Youths

John G.R. Fuerst*
Cleveland State University, USA
Ulster Institute for Social Research, London, UK

Meng Hu
Independent Researcher, France

Gregory Connor
Maynooth University, Ireland

* Corresponding author: email j122177@hotmail.com

Black and Hispanic children in the United States have lower mean cognitive test scores than White children. The reasons for this are contested. The test score gap may be caused by socio-cultural factors, but the high heritability of *g* suggests that genetic variance might play a role. Differences between self-identified race or ethnicity (SIRE) groups could be the product of ancestral genetic differences. This genetic hypothesis predicts that genetic ancestry will predict *g* within these admixed groups. To investigate this hypothesis, we performed admixture-regression analyses with data from the Adolescent Brain Cognitive Development Cohort. Consistent with predictions from the genetic hypothesis, African and Amerindian ancestry were both found to be negatively associated with *g*. The association was robust to controls for multiple cultural, socioeconomic, and phenotypic factors. In the models with all controls the effects were as follows: (a) Blacks, African ancestry: b = -0.89, *N* = 1690; (b) Hispanics, African ancestry: b = -0.58, Amerindian ancestry: b = -0.86, *N* = 2021), and (c) a largely African-European mixed Other group, African ancestry: b = -1.08, *N* = 748). These coefficients indicate how many standard deviations *g* is predicted to change when an individual's African or Amerindian ancestry proportion changes from 0% to 100%. Genetic ancestry statistically explained the self-identified race and ethnicity (SIRE) differences found in the full sample. Lastly, within all

samples, the relation between genetic ancestry and *g* was partially accounted for by cognitive ability and educational polygenic scores (eduPGS). These eduPGS were found to be significantly predictive of *g* within all SIRE groups, even when controlling for ancestry. The results are supportive of the genetic model.

Keywords: Ancestry, Admixture, education, Intelligence, ABCD cohort

There are substantial differences in mean cognitive test scores between self-identified racial and ethnic (SIRE) groups such as Whites, Hispanics, Blacks, and Asians in the United States (Murray, 2021; Roth et al., 2017). These differences are not attributable to psychometric bias, since cognitive test batteries typically exhibit measurement invariance across American SIRE groups (Scheiber, 2016a,b; Warne, 2020a). As such, they represent real differences in latent cognitive ability.

Although cognitive ability researchers disagree as to the causes of these differences (Rindermann, Becker & Coyle, 2020), a number of explanatory factors have been proposed. Most of these factors appeal to either cultural differences between groups or differing social and economic circumstances such as differences in poverty levels. Many researchers studying cognitive ability attribute some part of the differences to genetics (Rindermann, Becker & Coyle, 2020).

Analyses of both nationally representative samples and cognitive battery standardization data (Magnuson & Duncan, 2006; Weiss & Saklofske, 2020) indicate that socioeconomic status (SES) can statistically explain a substantial percentage of test score variance across SIRE groups. However, it is not clear to what extent SES captures predominantly environmental or genetic causes, since SES is related to both genetic and environmental differences within groups (Belsky et al., 2018; Krapohl & Plomin, 2016; Rowe, Vesterdal & Rodgers, 1998). Regardless, the data suggest that even after fully accounting for average SES differences, there remains unexplained variance in cognitive test scores between groups.

The genetic hypothesis is that differences in genes inherited from ancestors play a significant role in causing the SIRE related variation in cognitive test scores. Twin studies show non-trivial heritability for individual differences in cognitive ability within ethnic groups (Pesta et al., 2020). SIRE groups differ in European ancestry-based polygenic scores, and these scores are predictive of cognitive ability within ethnic groups (Lasker et al., 2019). Taken together, these results suggest that variation in cognitive ability among SIRE groups may be due in part to allele frequency differences at trait-associated gene loci.

Most studies that investigate the source of group differences categorize individuals by SIRE. However, SIRE reflects both genetic and sociocultural factors. This makes interpretation of simple SIRE-based results ambiguous. Admixture regression analyses, by including both SIRE and admixture variables simultaneously, can quantify the association between genetic ancestry and phenotype in admixed populations (Halder et al., 2015). These analyses test the extent to which the genetic differences between SIRE groups are responsible for observed trait differences.

If continental populations differ polygenically in trait-related phenotypes, these continental differences are expected to contribute to individual differences in admixed populations. In aggregate, alleles that vary between ancestral populations will be associated with phenotype in admixed populations. Individuals who have a larger proportion of their global ancestry from the ancestral group with higher average frequencies of trait-enhancing alleles are likely to also have higher polygenic scores and higher average values in the phenotype. As such, associations between genetic ancestry and phenotype in admixed populations would suggest that phenotypic differences between continental populations have a genetic basis (Halder et al., 2015). This admixture regression methodology can be extended by including polygenic scores (PGS) to see if PGS mediate the association between *g* and ancestry.

In this way, admixture-regression analysis allows researchers to separate the genetic element of SIRE from its cultural, behavioral, and psychosocial aspects, which may alternatively be responsible for the observed phenotypic differences. As Fang et al. (2019, p. 764) note: SIRE "acts as a surrogate to an array of social, cultural, behavioral, and environmental variables" and so "stratifying on SIRE has the potential benefits of reducing heterogeneity of these non-genetic variables and decoupling the correlation between genetic and non-genetic factors."

These analyses require a substantial degree of admixture in populations, and are more robust when that admixture has taken place in the course of the last seven to ten generations (Halder et al., 2015). In the USA, African and Hispanic Americans meet these requirements. They exhibit a wide range of European, African, and/or Amerindian ancestry due to admixture over the course of several generations.

Thus we apply admixture analysis to examine if SIRE differences in general cognitive ability (*g)* can be accounted for by genetic variation related to continental ancestry. We hypothesize that *g* will be lower in African and Hispanic American samples relative to European American samples. These differences will be associated with African and Amerindian genetic ancestry within the SIRE groups. Additionally, we hypothesize that the association between genetic ancestry and *g* will be robust to controls for possible socio-environmental

confounds and that genetic ancestry will also statistically account for the SIRE differences found in the full sample. Finally, we hypothesize that current polygenic scores for cognitive ability and education (eduPGS) will both predict individual differences within SIRE groups and explain a portion of the effect of ancestry on *g*.

Methods

1. Dataset

The Adolescent Brain Cognitive Development Study (ABCD) is a collaborative longitudinal project involving 21 sites across the USA. ABCD is the largest, longitudinal study of brain development and child health ever conducted in the USA. It was created to research the psychological and neurobiological bases of human development. At baseline, around 11,000 children aged 9-10 years were sampled, mostly from public and private elementary schools. A probabilistic sampling strategy was used, with the goal of creating a broadly representative sample of US children in that age range. Children with severe neurological, psychiatric, or medical conditions were excluded. Children were also excluded if they were not fluent in English or if their parents were not fluent in either English or Spanish. Parents provided informed consent. For this study, we utilized the baseline ABCD 3.01 data. We excluded individuals missing either cognitive or admixture scores. We also excluded any individual identified as being either Asian or Pacific Islander in order to focus on groups who were primarily of African, European, and Amerindian ancestry. This left 10,370 children.

For the admixture-regression analyses, SIRE groups were delineated using the ABCD race_ethnicity variable. This was a summary variable computed from 18 separate multiple choice questions asking about the child's race ("What race do you consider the child to be? Please check all that apply") and one question asking if the child is of Hispanic ethnicity ("Do you consider yourself Hispanic/Latino/Latina?"). Children were classified into one of five mutually exclusive groups: non-Hispanic White (*White*), non-Hispanic Black (*Black*), Hispanic of any race (*Hispanic*), non-Hispanic Asian (*Asian*) or any other (*Other*). The Other category included any non-Hispanic children who were reported to be two or more racial groups. Because we dropped the Asian category, we were left with four mutually exclusive SIRE groups.

2. Variables for admixture regression analyses

The following variables were used for the admixture regression analyses:

1. g scores

ABCD baseline data contain the following cognitive subtests, the first seven of which are from the NIH Toolbox® cognitive battery: Picture Vocabulary, Flanker, List Sorting, Card Sorting, Pattern Comparison, Picture Sequence Memory, Oral Reading Recognition, Wechsler Intelligence Scale for Children's Matrix Reasoning, The Little Man Test (efficiency score), The Rey Auditory Verbal Learning Test (RAVLT), immediate recall, and RAVLT delayed recall. For details about these measures, see Thompson et al. (2019).

We conducted multi-group confirmatory factor analysis (MGCFA) on these subtests, as detailed in Supplementary File 1. Briefly, we first checked whether outliers and missing data had any impact, and whether our results remained strong after correction. We then conducted exploratory factor analysis and multi-group confirmatory factor analysis on the aforementioned set of subtests as a check for bias. After adjustment for age, we did not find any non-linear effects of age. Adjustment for sex did not reveal any evidence of meaningful differences in fit between the competing models, the *g*-model and the correlated factors model. We find that a three broad factor model (memory, complex cognition, and executive function) with *g* at the apex fits the data well. Moreover, strict measurement invariance holds between SIRE groups. The best fitting model (M6A, Table S2 of Supplementary File 1; CFI = .954, RMSEA = .044) was one in which *g* alone explains SIRE group differences. We output the *g*-factor scores from this model for use in the analyses. These score magnitudes are approximately the same as those derived from exploratory factor analysis.

2. Socioeconomic status (SES)

We identified seven indicators of SES: financial adversity, area deprivation index, neighborhood safety protocol, parental education, parental income, parental marital status, and parental employment status. These are detailed in Supplementary File 2. We submitted the seven SES indicators to Principal Components Analysis (PCA). We used the R package PCAmixdata, which handles mixed categorical and continuous data (Chavent et al., 2014). The first unrotated component explained 42% of the variance. The PCA_1 loadings for the seven SES indicators were as follows: financial adversity (.31), area deprivation index (.49), neighborhood safety protocol (.31), parental education (.53), parental income (.66), parental marital status (.42), and parental employment status (0.21). More details and the correlation matrix for the SES indicators is provided in Supplemental File 2.

3. Child US-born

Parents were asked about the country of the child's birth. We recoded this variable as 1 for "United States" and 0 for all other responses.

4. Immigrant family

Parents were asked if anyone in the child's family, including maternal or paternal grandparents, was born outside of the United States. This variable was coded as 1 for "Yes" and 0 for all other responses.

5. Nationality (Puerto Rican, Mexican, and Cuban)

If a child was reported to be Hispanic, parents were additionally asked about the specific Latin American nation of origin ("Please choose the group that best represents the child's Hispanic origin or ancestry"). Seventy percent of the Hispanic children were reported as being either Mexican, Mexican American, or Chicano ($N = 1028$), Puerto Rican ($N = 210$), or Cuban or Cuban American ($N = 174$). Dummy variables were created for these three nationality groups, with "1" indicating "Yes" and "0" indicating "No".

6. Frac_SIRE

Four dummy SIRE variables (Black, White, Native American, and Not Otherwise Classified (NOC) were computed from the 18 questions asking about the child's specific race. The NOC SIRE group included those who were marked as: "Other Race," "Refused to answer," or "Don't Know." These were then recoded into interval variables in which individuals are assigned a SIRE fraction ranging from 0 to 1 (Liebler & Halpern-Manners, 2008). These were calculated as the value selected for each of the four groups (0 or 1) over the total number of responses (0 to 4) chosen. For example, someone marked as only Black and White would be assigned scores of (Black: ½; White: ½; Native_American: 0; NOC = 0). This SIRE coding was used as it was previously found to be the most predictive in models which also included genetic ancestry (Kirkegaard et al., 2019).

7. Hispanic

For the admixture-regression analysis conducted on the full sample, we additionally included a dummy variable for Hispanic ethnicity. This was coded as "1" for "Hispanic" and "0" for not Hispanic. As the subsamples were either Hispanic or non-Hispanic, this variable was not used in the subsample analyses.

8. *Ethnic attachment*

Parents were given the Multigroup Ethnic Identity Measure-Revised (MEIM-R) Survey. In this they were asked six Likert-scaled (1 = strongly agree; 5 = strongly disagree) questions regarding their ethnic group: "I have spent time trying to find out more about my ethnic group, such as its history, traditions, and customs", "I have a strong sense of belonging to my own ethnic group," "I understand pretty well what my ethnic group membership means to me," "I have often done things that will help me understand my ethnic background better,", "I have often talked to other people in order to learn more about my ethnic group,", "I feel a strong attachment towards my own ethnic group." ABCD computed MEIM-R summary scores, which we standardized. We treat this as a measure of family ethnic-related culture. We only included this variable in the subsample analyses. In these the members belonged to the same broad ethnic group (e.g., Black or Hispanic).

9. State racism

ABCD calculated state-level indicators of both racism and immigrant bias. These were based on both implicit bias measures and state-level structural variables. The two indicators correlated at $r = .41$ ($N = 9386$). We standardized both measures ($M = 0$, $SD = 1$) and then averaged them and standardized the resulting average.

10. Discrimination factor

In Year 1 follow-up, the children were asked 6 questions regarding perceived ethnic, racial, national, or color based discrimination. The questions were as follows: "In the past 12 months, have you felt discriminated against: because of your race, ethnicity, or color?", "In the past 12 months, have you felt discriminated against: because you are (or your family is) from another country?", "How often do the following people treat you unfairly or negatively because of your ethnic background?" (Teachers? Other adults outside school? Other students?), "I feel that others behave in an unfair or negative way toward my ethnic group." We imputed missing data using the mice package (df, m = 5, maxit = 50, method = 'pmm', seed = 500). We used the mirt package in R to perform factor analysis on the six questions. We then standardized and saved the factor scores.

11. Skin_color, P_Brown_Eye, P_Intermediate_Eye, P_Blue_Eye, P_Black_Hair, P_Brown Hair, P_Red_or_Blond_Hair).

Conley and Fletcher (2017) have suggested that phenotypic-based discrimination might mediate the association between cognitive ability and genetic ancestry. This is called the colorism model (Hu et al., 2019). It can be

tested by including indices of race-related phenotype into the regression models to see if these capture the association between ancestry and cognitive ability. As such, we include measures of eye, hair, and skin color. Skin, Hair, Eye color were calculated based on the publicly available, "Hirisplex Eye, Hair and Skin Colour DNA Phenotyping Webtool." This tool and score calculations have been detailed by Lasker et al. (2019). We additionally combined (summed) the red and blond hair probabilities. Skin color was scaled as detailed in Lasker et al. (2019), with higher scores representing darker color, and then standardized. The eye and hair color variables represent the percent in the full sample with the specific color and were left unstandardized to retain interpretability.

12. Admixture estimates

Imputing and genotyping was done by the ABCD Research Consortium using Illumina XX. 516,598 variants survived the quality control. Before global admixture estimation, we applied quality control using PLINK 1.9. We used only directly genotyped, bi-allelic, autosomal SNP variants (494,433 before, 493,196 after lifting). We pruned variants for linkage disequilibrium at the 0.1 R^2 level using PLINK 1.9 (--indep-pairwise 10000 100 0.1). This variant filtering was done in the reference population dataset to reduce bias from sample non-representativeness. 99,642 variants were left after pruning. We merged the target samples from ABCD with reference population data for the populations of interest. A k=5 solution with European, Amerindian, African, East Asian and South Asian components provides the most comprehensive but parsimonious model of the US population, capturing all the predominant ancestral backgrounds in the US population. We merged our sample with relevant samples from 1000 Genomes and from the HGDP to perform the cluster analysis and identify these k=5 components. The following populations from 1000 Genomes and from the HGDP reference populations were excluded: Adygei, Balochi, Bedouin, Bougainville, Brahui, Burusho, Druze, Hazara, Makrani, Mozabite, Palestinian, Papuan, San, Sindhi, Uygur, Yakut. We excluded these populations because they were overly admixed or because the individuals in the ABCD sample lacked significant portions of these ancestries (e.g., Melanesians and San). We split the ABCD target samples into 50 random subsets (222 persons each) and merged them sequentially with the reference data. Admixture at k = 5 was run on each of the 50 merged subsets. This repeated subsetting was done to avoid skewing the admixture algorithm to European ancestry which was predominant in the ABCD sample.

13. First 20 Principal components

For the analysis of PGS predictivity within SIRE groups we controlled for the first 20 ancestry principal components to take into account population structure related effects. These components were generated by PLINK v1.90b6.8 when computing polygenic scores.

14. eduPGS

For polygenic scores (PGS), we scored the genomes using PLINK v1.90b6.8. For background, a polygenic score (PGS) "is an estimate of an individual's genetic liability to a trait or disease, calculated according to their genotype profile and relevant genome-wide association study (GWAS) data" (Choi, Mak & O'Reilly, 2020). We used the genome-wide association study (GWAS) results from Lee et al. (2018). Specifically, we used the multi-trait analysis of genome-wide association study (MTAG) eduPGS SNPs (*N* = 8,898 variants in this sample) to compute eduPGS. The MTAG eduPGS were computed using MTAG, a method for analyzing statistics from genome-wide association studies (GWAS) on different but genetically correlated traits (e.g., education and intelligence). These scores were based on cognitive ability (*n* = 257,841), hardest math class taken (*n* = 430,445), and mathematical ability (*n* = 564,698) (Lee et al., 2018). We use these PGS because previous research has shown them to have trans-ethnic predictive validity in European, Hispanic, and African American populations (Fuerst, Kirkegaard & Piffer, 2021; Lasker et al., 2019). Moreover, common forms of bias were found not to account for the ancestry-related eduPGS differences (Fuerst et al., 2021). Thus, we can say that this PGS *plausibly* captures genetic effects between ancestry groups.

15. The NIH Toolbox® (NIHTBX) neuropsychological battery

For one validation analysis of the eduPGS which included Asians, we used the NIHTBX summary scores. This was because we did not run MGCFA on the small Asian samples and so did not have *g* scores for these groups. This battery has been shown to be measurement invariant across American Black, Hispanic, and White SIRE groups (Lasker et al., 2019). The effects of age and sex were controlled for. We standardized the residuals.

3. Methods (Analyses)

We first present the descriptive statistics for the sample and the subsamples. We then explore the bivariate relation between European admixture and *g*-scores. We include both linear regression lines and loess lines in the regression plots (based on the gg_scatter package in R). These analyses are descriptive and do not take into account the complex structure of the data.

After, we run a series of within-SIRE (Black, Hispanic, and Other) admixture-regression analyses to control for potential environmental confounds. For these analyses, we set European ancestry as a reference value with a value of zero. Following Heeringa and Berglund's (2021) recommendations, we use a multi-level mixed effects three-level (site, family, individual) model. In this model, recruitment site and family common factors are treated as random effects (i.e., as random samples from a population). We further report the dense numeric matrix results for the regression models in Supplementary File 3.

The pooled data with both the regular ABCD baseline sample and the pooled twin samples were used. As Heeringa and Berglund (2021) note, the specification replicates that used by the ABCD Data Exploration and Analysis Portal (DEAP). Thus, the use of this multilevel model also aids in replication. For the regression analyses on the SIRE subsamples, we ran four models. The first model includes genetic ancestry and controls for both child and family immigrant status. The second model adds a term for SIRE and ethnic attachment to capture SIRE specific cultural effects. The third model adds terms to capture possible discrimination related effects: state-level racism, child reported experiences of discrimination, and race-related phenotype. The fourth model adds our general SES variable. Geographic effects are controlled for by including study site as a random effect in the model.

For the regression analyses, general cognitive ability scores (*g*-scores) are used as the dependent variable. This variable was standardized (M = 0.00; SD = 1.00) in the full sample. As for the independent variables, both the ancestry and fractional SIRE variables were left unstandardized. This allows the unstandardized beta coefficients for these variables to be interpreted as the effect of a change in 100 percent ancestry/SIRE identity on one standardized unit of cognitive ability. The rationale for this method has been detailed elsewhere (Lasker et al., 2019). The Child_USA_Born and Immigrant_Family dummy variables were also left unstandardized to retain interpretability. The three eye color and the three hair color variables, which represent probabilities that sum to one in the full samples, are also not standardized to retain interpretability. The remaining variables — ethnic attachment, state racism, discrimination factor, skin color, and SES — are all standardized in the full sample. Thus, the unstandardized B coefficient for these variables represents the change in *g* induced by a change of one standard deviation in the independent variable.

Results

1. Descriptive statistics

The descriptive statistics for the total sample and the four SIRE subsamples are shown in Table 1. Cohen's *d* for the difference in *g* between Black and White

Americans comes to 1.02 *d*. This represents a large effect by conventional standards (Cohen, 1988) and is typically sized for measured g differences across SIRE groups (Roth et al., 2017). The difference between Hispanic and White Americans is 0.38 *d*, while that between Others and White Americans is 0.37 *d*. These latter two differences represent small to medium sized effects (Cohen, 1988). The Hispanic-White difference is smaller than usually reported (e.g., Roth et al., 2017). This could be due to the exclusion of children who were not fluent in English.

Table 1. *Total sample and subsample characteristics.*

	Total sample M ± SD	Black M ± SD	Hispanic M ± SD	Other M ± SD	White M ± SD
Age (in Months)	119.0 ± 7.49	118.9 ± 7.28	118.6 ± 7.58	118.7 ± 7.40	119.2 ± 7.52
g	0.00 ± 1.00	-0.69 ± 1.07	-0.10 ± 0.99	-0.09 ± 1.09	0.24 ± 0.86
SES	0.00 ± 1.00	-0.98 ± 0.93	-0.36 ± 0.91	-0.40 ± 1.00	0.45 ± 0.75
frac_White_SIRE	0.73 ± 0.43	0.00 ± 0.00	0.67 ± 0.45	0.39 ± 0.25	1.00 ± 0.03
frac_Black_SIRE	0.20 ± 0.38	1.00 ± 0.04	0.07 ± 0.23	0.29 ± 0.26	0.00 ± 0.00
frac_Native_Amer._SIRE	0.02 ± 0.11	0.00 ± 0.00	0.03 ± 0.13	0.18 ± 0.28	0.00 ± 0.00
frac_Other_SIRE	0.06 ± 0.23	0.00 ± 0.04	0.23 ± 0.42	0.13 ± 0.34	0.00 ± 0.03
European_ancestry	0.75 ± 0.33	0.16 ± 0.11	0.60 ± 0.21	0.62 ± 0.25	0.98 ± 0.05
African_ancestry	0.18 ± 0.31	0.82 ± 0.11	0.10 ± 0.15	0.32 ± 0.26	0.01 ± 0.02
Amerindian_ancestry	0.06 ± 0.14	0.01 ± 0.02	0.28 ± 0.19	0.04 ± 0.09	0.01 ± 0.03
South_Asian_ancestry	0.00 ± 0.02	0.00 ± 0.01	0.01 ± 0.01	0.01 ± 0.05	0.00 ± 0.01
East_Asian_ancestry	0.01 ± 0.03	0.01 ± 0.02	0.01 ± 0.02	0.01 ± 0.07	0.00 ± 0.02
State_racism	0.00 ± 1.00	0.44 ± 0.93	-0.35 ± 0.93	0.26 ± 0.99	-0.04 ± 0.99
Discrim_fact	0.00 ± 1.00	0.47 ± 1.24	0.09 ± 1.04	0.19 ± 1.05	-0.19 ± 0.83
Ethnic_attachment	0.00 ± 1.00	0.38 ± 1.03	0.23 ± 1.01	0.05 ± 1.00	-0.19 ± 0.94
Skin_color	0.00 ± 1.00	1.32 ± 0.42	0.58 ± 0.80	0.36 ± 0.91	-0.62 ± 0.60
P_Brown_Eye	0.56 ± 0.41	0.97 ± 0.10	0.83 ± 0.28	0.73 ± 0.35	0.33 ± 0.35
P_Intermediate_Eye	0.08 ± 0.07	0.02 ± 0.03	0.06 ± 0.06	0.07 ± 0.06	0.10 ± 0.07
P_Blue_Eye	0.36 ± 0.41	0.02 ± 0.08	0.11 ± 0.25	0.20 ± 0.33	0.57 ± 0.39
P_Black_Hair	0.23 ± 0.24	0.55 ± 0.16	0.38 ± 0.23	0.28 ± 0.20	0.09 ± 0.10
P_Brown Hair	0.46 ± 0.18	0.43 ± 0.13	0.49 ± 0.16	0.54 ± 0.15	0.45 ± 0.20
P_Red_or_Blond_Hair	0.30 ± 0.29	0.02 ± 0.05	0.13 ± 0.18	0.18 ± 0.21	0.46 ± 0.26
Child_USA_Born	0.98 ± 0.15	0.98 ± 0.15	0.94 ± 0.24	0.98 ± 0.14	0.99 ± 0.12
Immigrant_family	0.28 ± 0.45	0.14 ± 0.35	0.73 ± 0.44	0.20 ± 0.40	0.18 ± 0.38
Puerto_Rican			0.10 ± 0.31		
Mexican			0.51 ± 0.50		
Cuban			0.09 ± 0.28		
eduPGS	0.00 ± 1.00	-1.33 ± 0.57	-0.21 ± 0.76	-0.36 ± 0.85	0.50 ± 0.77

Note: Nationality variables (Mexico, Cuba, Puerto Rico) were only computed for Hispanics.

In this sample, parent-identified Whites are 98% European in ancestry (1% African; 1% Amerindian). Since this group has little admixture, we relegate the within SIRE admixture-regression analyses to the supplementary file. Both the Black (82% African, 16% European, 1% Amerindian) and the Other (62% European, 32% African, 4% Amerindian) groups are African-European admixed groups. Hispanics additionally have a substantial Amerindian component (60% European, 28% Amerindian, 10% African). Figure 1 shows the distribution of ancestry by SIRE groups. These admixture percentages correspond with those typically reported in the literature (e.g., Bryc et al., 2015).

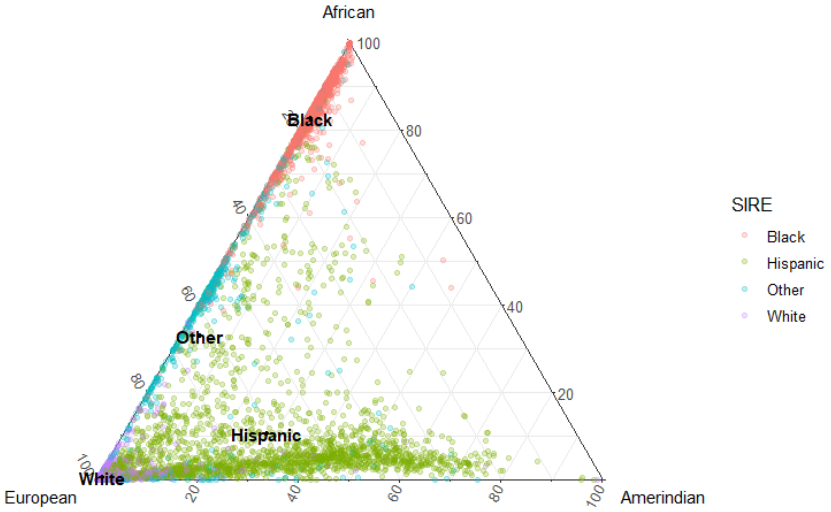

Figure 1. *Admixture triangle plot for SIRE groups in the ABCD sample.*

2. Regression plots and admixture regression analyses

1. Black Americans

Figure 2 shows the regression plot for European ancestry and *g*-scores among Black children. European ancestry is significantly (r = .10, N = 1690) associated with *g* scores. The R boxplot function indicated 13 outliers. However, removing these had no effect on the bivariate correlation (r = .10, N = 1677). Additionally, the Loess regression line indicated a possible curvilinear relation with a slight uptick in scores at the lowest European ancestry decile. Further analysis showed that this was due to relatively high scores of individuals from African immigrant families ($M_{African\_immigrant}$ = -.28, N = 60). Limiting our scope to African Americans within US-born families raises the correlation to r = .13 (N = 1475); for African Americans with 2% to 80% European admixture, this correlation

is r = .11 (*N* = 1635). These results are shown in Table S4 of Supplementary file 3 along with scores by African American subgroups. The full correlation matrices are also provided in Supplementary File 3.

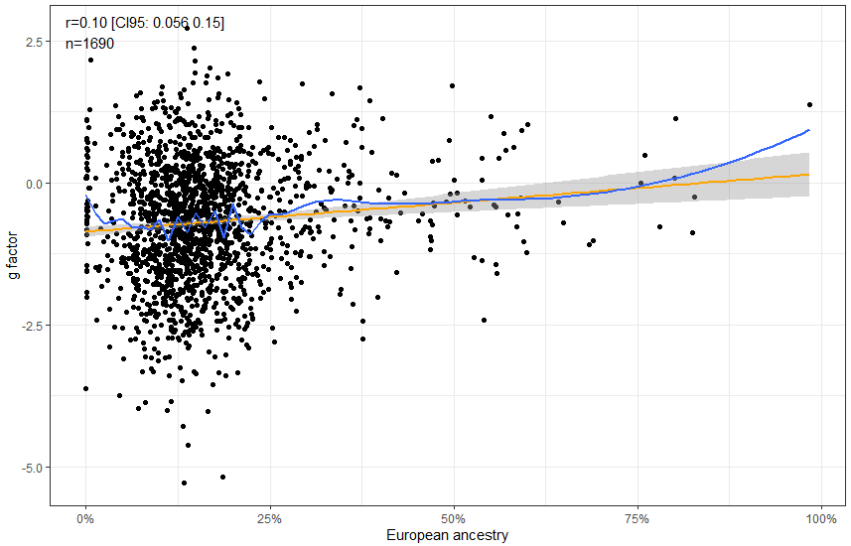

Figure 2. *Regression plot of European ancestry and g in the Black American subsample (N = 1690).*

We next proceed to the admixture-regression analyses. Because the Black SIRE category excludes multi-racial individuals, we do not include a term for fraction SIRE in these models. As seen in Table 2, African ancestry is strongly and significantly negatively related to cognitive ability in all four models. Amerindian ancestry is also negatively related to *g*-scores; however, owing to the low Amerindian admixture among non-Hispanic Blacks — and consequently the high standard errors — these estimates are not reliable. Adding ethnic attachment scores in Model 2 does not change the relationship with Amerindian and African ancestry. As seen in Model 3, measures of racial discrimination do not mediate the relation between *g* and ancestry. Of these variables added to Model 3, only experiences of discrimination had a significant independent effect. Finally, as seen in Model 4, while SES was significantly related to *g*, it did not substantially attenuate the association between African ancestry and *g* ($b_{\text{African ancestry}}$ = -1.08 to -0.89).

Table 2. *Regression results for the effect of genetic ancestry on g Among Black Americans (N = 1690). Shown are the beta coefficients (b) and p-values (p) from the mixed effects models with recruitment site and family common factors treated as random effects. The values in parentheses are standard errors. The marginal and conditional R^2 are provided at the bottom.*

	Model 1		Model 2		Model 3		Model 4	
Predictors	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
(Intercept)	0.00 (0.29)	0.997	0.02 (0.29)	0.934	0.69 (0.74)	0.351	0.31 (0.72)	0.669
Amerindian_ancestry	-3.80 (1.22)	**0.002**	-3.95 (1.22)	**0.001**	-3.83 (1.23)	**0.002**	-3.46 (1.20)	**0.004**
African_ancestry	-1.07 (0.26)	**<0.001**	-1.11 (0.26)	**<0.001**	-1.08 (0.29)	**<0.001**	-0.89 (0.29)	**0.002**
East_Asian_ancestry	-0.03 (1.57)	0.982	-0.07 (1.56)	0.965	0.16 (1.56)	0.920	-0.29 (1.50)	0.850
South_Asian_ancestry	-5.67 (3.42)	0.098	-5.83 (3.40)	0.087	-5.97 (3.42)	0.080	-5.46 (3.32)	0.100
Child_US_Born	0.19 (0.18)	0.287	0.16 (0.18)	0.365	0.17 (0.18)	0.354	0.32 (0.18)	0.067
Immigrant_Family	0.29 (0.08)	**<0.001**	0.27 (0.08)	**0.001**	0.27 (0.08)	**0.001**	0.15 (0.08)	0.058
Ethnic_attachment			0.11 (0.03)	**<0.001**	0.11 (0.03)	**<0.001**	0.08 (0.03)	**0.001**
State_racism					0.01 (0.05)	0.866	0.04 (0.05)	0.456
discrim_fact					-0.05 (0.02)	**0.014**	-0.04 (0.02)	**0.027**
Skin_color					0.08 (0.07)	0.290	0.11 (0.07)	0.124
P_Brown_Eye					-0.32 (0.42)	0.455	-0.33 (0.41)	0.430
P_Intermediate_Eye					-0.39 (1.76)	0.825	0.61 (1.72)	0.721
P_Black_Hair					-0.42 (0.77)	0.589	-0.12 (0.75)	0.874
P_Brown_Hair					-0.56 (0.82)	0.496	-0.26 (0.80)	0.747
SES							0.29 (0.03)	**<0.001**
Random Effects								
σ^2	0.56		0.56		0.57		0.56	
τ_{00}	$0.55_{\text{site\_id\_1:rel\_family\_id}}$		$0.53_{\text{site\_id\_1:rel\_family\_id}}$		$0.52_{\text{site\_id\_1:rel\_family\_id}}$		$0.46_{\text{site\_id\_1:rel\_family\_id}}$	
	$0.03_{\text{site\_id\_1}}$		$0.03_{\text{site\_id\_1}}$		$0.03_{\text{site\_id\_1}}$		$0.04_{\text{site\_id\_1}}$	
ICC	0.51		0.50		0.49		0.47	
N	$22_{\text{site\_id\_1}}$		$22_{\text{site\_id\_1}}$		$22_{\text{site\_id\_1}}$		$22_{\text{site\_id\_1}}$	
	$1436_{\text{rel\_family\_id}}$		$1436_{\text{rel\_family\_id}}$		$1436_{\text{rel\_family\_id}}$		$1436_{\text{rel\_family\_id}}$	
Observations	1690		1690		1690		1690	
Marginal R^2 / Conditional R^2	0.022 / 0.517		0.033 / 0.515		0.037 / 0.512		0.096 / 0.520	

2. Hispanic Americans

Figure 3 shows the regression plot for European ancestry and *g* scores among Hispanic children. As seen, European ancestry is significantly (*r* = .23, *N* = 2021) associated with *g* scores. While the R boxplot function indicates that there are 23 outliers, removing these had little effect on the bivariate correlation (*r* = .22, *N* = 1998). The Loess regression line suggests a possible slight uptick in scores at the lowest European ancestry decile. However, the 95% confidence intervals of this line (not shown) overlapped with the linear regression line.

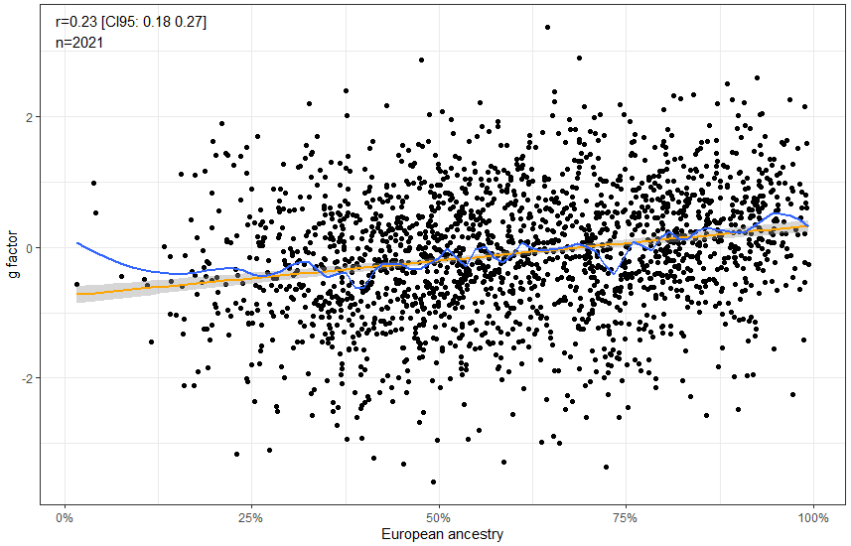

Figure 3. *Regression plot of European ancestry and g in the Hispanic American subsample (N = 2021).*

For the Hispanic admixture-regression analyses, we include a term for race because the Hispanic ethnic category is inclusive of all self-identified racial groups. As shown in Table 3, both Amerindian and African ancestry are strongly negatively associated with *g* in the first three models. Adding SIRE ethnic identity and the ethnic attachment variable in Model 2 had little effect on the beta for Amerindian ancestry. Doing so increases the effect of African ancestry. In Model 3, both skin color and experiences of discrimination have significant independent effects on *g*, but these variables only slightly attenuated the relation between *g* and Amerindian and African ancestry. However, as seen in Model 4, SES attenuated the effect of Amerindian and African ancestry (Model 3: $b_{\text{african ancestry}}$ = -0.96 →Model 4: $b_{\text{african ancestry}}$ = -0.58; Model 3: $b_{\text{amerindian ancestry}}$ = -1.37 →Model

4: $b_{amerindian\ ancestry}$ = -0.86). Nonetheless, the magnitudes of the Amerindian and African ancestry effects remained medium to large in size and statistically significant.

Table 3. *Regression results for the effect of genetic ancestry on g among Hispanic American children (N = 2021). Shown are the beta coefficients (b) and p-values (p) from the mixed effects models with recruitment site and family common factors treated as random effects. The values in parentheses are standard errors. The marginal and conditional R^2 are provided at the bottom.*

	Model 1		**Model 2**		**Model 3**		**Model 4**	
Predictors	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
(Intercept)	0.31 (0.12)	**0.009**	0.34 (0.12)	**0.005**	0.09 (0.17)	0.589	-0.07 (0.17)	0.694
Amerindian_ancestry	-1.58 (0.15)	**<0.001**	-1.50 (0.15)	**<0.001**	-1.37 (0.18)	**<0.001**	-0.86 (0.18)	**<0.001**
African_ancestry	-0.76 (0.16)	**<0.001**	-1.08 (0.25)	**<0.001**	-0.96 (0.26)	**<0.001**	-0.58 (0.26)	**0.027**
East_Asian_ancestry	-1.73 (0.93)	0.062	-1.59 (0.93)	0.088	-1.57 (0.92)	0.087	-1.43 (0.90)	0.112
South_Asian_ancestry	4.23 (3.46)	0.222	4.10 (3.45)	0.235	3.83 (3.41)	0.262	3.55 (3.33)	0.287
Child_US_Born	0.09 (0.09)	0.322	0.08 (0.09)	0.408	0.08 (0.09)	0.383	0.08 (0.09)	0.364
Immigrant_Family	0.17 (0.06)	**0.003**	0.16 (0.06)	**0.005**	0.15 (0.06)	**0.009**	0.12 (0.06)	**0.026**
Mexican	-0.14 (0.06)	**0.014**	-0.16 (0.06)	**0.009**	-0.14 (0.06)	**0.017**	-0.10 (0.06)	0.090
Cuban	-0.24 (0.09)	**0.009**	-0.25 (0.09)	**0.007**	-0.23 (0.09)	**0.011**	-0.15 (0.09)	0.104
Puerto_Rican	-0.40 (0.09)	**<0.001**	-0.39 (0.09)	**<0.001**	-0.39 (0.09)	**<0.001**	-0.29 (0.09)	**0.001**
frac_Black_SIRE			0.25 (0.15)	0.107	0.28 (0.15)	0.060	0.22 (0.15)	0.144
frac_Native_American_SIRE			-0.09 (0.17)	0.599	-0.04 (0.17)	0.812	0.03 (0.16)	0.835
frac_NOC_SIRE			-0.09 (0.06)	0.138	-0.06 (0.06)	0.272	-0.03 (0.06)	0.589
Ethnic_attachment			0.06 (0.02)	**0.008**	0.06 (0.02)	**0.008**	0.04 (0.02)	**0.045**
State_racism					0.03 (0.05)	0.533	0.06 (0.05)	0.236
discrim_fact					-0.14 (0.02)	**<0.001**	-0.13 (0.02)	**<0.001**
Skin_color					-0.08 (0.04)	**0.032**	-0.07 (0.04)	**0.044**
P_Brown_Eye					0.20 (0.11)	0.072	0.26 (0.11)	**0.017**
P_Intermediate_Eye					0.20 (0.44)	0.653	0.26 (0.43)	0.548
P_Black_Hair					0.03 (0.19)	0.869	-0.03 (0.18)	0.865
P_Brown_Hair					0.12 (0.19)	0.541	0.12 (0.19)	0.535
SES							0.25 (0.03)	**<0.001**
Random Effects								
σ^2	0.40		0.39		0.39		0.39	
τ_{00}	0.49 $_{site\_id\_1:rel\_family\_id}$		0.48 $_{site\_id\_1:rel\_family\_id}$		0.46 $_{site\_id\_1:rel\_family\_id}$		0.42 $_{site\_id\_1:rel\_family\_id}$	
	0.02 $_{site\_id\_1}$		0.02 $_{site\_id\_1}$		0.03 $_{site\_id\_1}$		0.03 $_{site\_id\_1}$	
ICC	0.56		0.56		0.56		0.54	
N	22 $_{site\_id\_1}$		22 $_{site\_id\_1}$		22 $_{site\_id\_1}$		22 $_{site\_id\_1}$	
	1794 $_{rel\_family\_id}$		1794 $_{rel\_family\_id}$		1794 $_{rel\_family\_id}$		1794 $_{rel\_family\_id}$	
Observations	2021		2021		2021		2021	
Marginal R^2 / Conditional R^2	0.087 / 0.600		0.094 / 0.603		0.119 / 0.611		0.160 / 0.611	

3. Other Americans

Figure 4 shows the regression plot for European ancestry and *g*-scores among the Other group. European ancestry is significantly (r = .19, N = 748) associated with *g* scores. Eight outliers were identified using the R boxplot function. Removing these had little effect on the correlation (r = .16, N = 740). The Loess regression line show a slight uptick in scores at the lowest European ancestry decile. However, the 95% confidence intervals of this line overlapped with the linear regression line. The correlation matrix is provided in the Supplementary File.

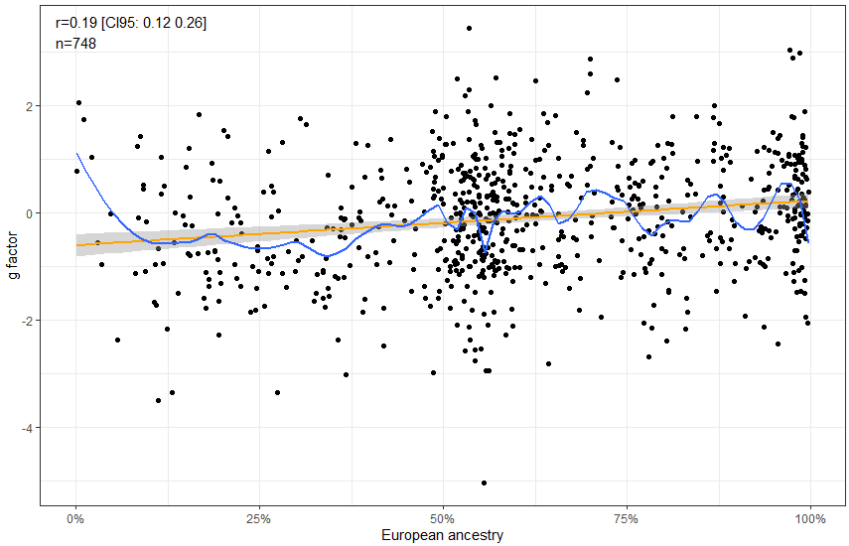

Figure 4. *Regression plot of European ancestry and g in Other American subsample (N = 748).*

As for Hispanics, we include a term for race because the Other American ethnic category is inclusive of all self-identified racial groups. As seen in Table 4, both coefficients for Amerindian and African ancestry show a strong negative association with *g* from Model 1 through Model 4. Adding SIRE ethnic identity and the ethnic attachment variable in Model 2 has little effect on the beta for Amerindian ancestry. Doing so increases the effect of African ancestry. In Model 3, only experiences of discrimination has a significant independent effect on *g*. The discrimination variables did not attenuate the relation between *g* and Amerindian and African ancestry. As seen in Model 4, SES moderately attenuated the effects of Amerindian and African ancestry (Model 3: $b_{\text{african ancestry}}$

= -1.38 →Model 4: $b_{african\ ancestry}$ = -1.08; Model 3: $b_{amerindian\ ancestry}$ = -1.51 →Model 4: $b_{amerindian\ ancestry}$ = -1.09). Nonetheless, the magnitudes of these ancestry effects remained large in magnitude and statistically significant.

Table 4. *Regression results for the effect of genetic ancestry on g among Other Americans children (N = 748). Shown are the beta coefficients (b) and p-values (p) from the mixed effects models with recruitment site and family common factors treated as random effects. The values in parentheses are standard errors. The marginal and conditional R^2 are provided at the bottom.*

	Model 1		**Model 2**		**Model 3**		**Model 4**	
Predictors	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
(Intercept)	-0.14 (0.32)	0.671	0.03 (0.34)	0.923	0.14 (0.39)	0.716	0.01 (0.38)	0.977
Amerindian_ancestry	-1.55 (0.45)	**0.001**	-1.40 (0.48)	**0.003**	-1.51 (0.51)	**0.003**	-1.09 (0.50)	**0.031**
African_ancestry	-1.07 (0.17)	**<0.001**	-1.27 (0.24)	**<0.001**	-1.38 (0.29)	**<0.001**	-1.08 (0.29)	**<0.001**
East_Asian_ancestry	0.95 (0.56)	0.088	0.96 (0.57)	0.093	0.59 (0.60)	0.328	0.57 (0.58)	0.330
South_Asian_ancestry	0.61 (0.79)	0.436	0.74 (0.80)	0.354	0.65 (0.81)	0.425	1.11 (0.79)	0.160
Child_US_Born	0.43 (0.30)	0.155	0.32 (0.31)	0.299	0.27 (0.31)	0.371	0.27 (0.30)	0.362
Immigrant_Family	0.14 (0.11)	0.215	0.09 (0.11)	0.433	0.07 (0.12)	0.564	-0.04 (0.11)	0.718
frac_Black_SIRE			0.16 (0.33)	0.625	0.18 (0.33)	0.584	0.22 (0.32)	0.504
frac_Native_American_SIRE			-0.22 (0.23)	0.346	-0.23 (0.24)	0.337	-0.21 (0.23)	0.362
frac_NOC_SIRE			-0.13 (0.20)	0.528	-0.15 (0.20)	0.435	-0.06 (0.19)	0.757
Ethnic_attachment			0.07 (0.04)	0.086	0.08 (0.04)	0.057	0.06 (0.04)	0.095
State_racism					0.03 (0.07)	0.683	0.07 (0.07)	0.322
discrim_fact					-0.11 (0.04)	**0.002**	-0.10 (0.04)	**0.007**
Skin_color					0.01 (0.06)	0.832	0.01 (0.06)	0.861
P_Brown_Eye					0.11 (0.17)	0.514	0.09 (0.17)	0.606
P_Intermediate_Eye					0.72 (0.75)	0.343	0.66 (0.73)	0.370
P_Black_Hair					0.22 (0.33)	0.506	0.33 (0.32)	0.315
P_Brown_Hair					-0.36 (0.31)	0.243	-0.21 (0.31)	0.493
SES							0.27 (0.05)	**<0.001**
Random Effects								
σ^2	0.62		0.61		0.62		0.65	
τ_{00}	0.45 $_{site\_id\_1:rel\_family\_id}$		0.46 $_{site\_id\_1:rel\_family\_id}$		0.44 $_{site\_id\_1:rel\_family\_id}$		0.35 $_{site\_id\_1:rel\_family\_id}$	
	0.06 $_{site\_id\_1}$		0.06 $_{site\_id\_1}$		0.06 $_{site\_id\_1}$		0.07 $_{site\_id\_1}$	
ICC	0.45		0.46		0.45		0.39	
N	22 $_{site\_id\_1}$		22 $_{site\_id\_1}$		22 $_{site\_id\_1}$		22 $_{site\_id\_1}$	
	644 $_{rel\_family\_id}$		644 $_{rel\_family\_id}$		644 $_{rel\_family\_id}$		644 $_{rel\_family\_id}$	
Observations	748		748		748		748	
Marginal R^2 / Conditional R^2	0.076 / 0.493		0.083 / 0.503		0.098 / 0.503		0.142 / 0.481	

4. White Americans

We do not report the admixture regression results for the 5911 non-Hispanic White Americans. These results are unreliable owing to the low dispersion in African and Amerindian ancestry within this SIRE group (see Table 1). Thus, we relegate these results to the supplemental material. Briefly, though, in Model 4 for this subsample, both African ancestry ($b_{African\ ancestry}$ = -.85, p = .110) and Amerindian ancestry ($b_{Amerindian\ ancestry}$ = -.96) also have large negative effects. However, this effect is only statistically significant for Amerindian ancestry (p = .012).

5. Full sample

The results above indicate that factors associated with genetic ancestry are related to *g within* SIRE groups. These findings also suggest that these same factors explain differences *between* SIRE groups (Halder et al., 2015). Using the full sample, we examine this implication. The relation between European ancestry and *g* for the full sample is shown in Figure 5. As expected, there is a strong positive association for the SIRE groups between ancestry and *g* (r = .36; N = 10370). Because the range of ancestry is not restricted — restriction of range attenuates correlations — the correlation is high. In this plot, we again see the uptick at the lowest decile of European admixture. This is due to the relatively high scores of children of recent African immigrants.

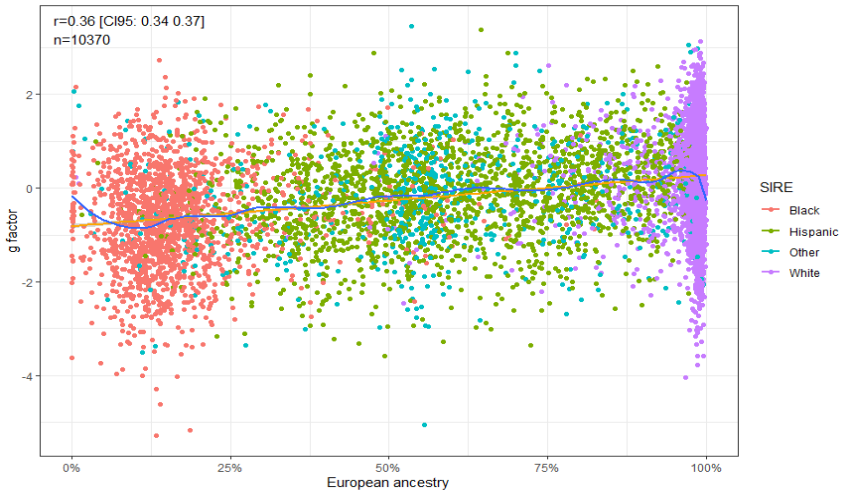

Figure 5. *Regression plot of European ancestry and g in the full sample (N = 10370).*

Table 5. *Regression results for the effect of ancestry on cognitive ability in the full sample (N = 10,370). Shown are the beta coefficients (b) and p-values (p) from the mixed effects models with recruitment site and family common factors treated as random effects. The values in parentheses are standard errors. The marginal and conditional R^2 are provided at the bottom.*

	Model 1a		Model 1b		Model 2		Model 3	
Predictors	*b*	*p*	*b*	*p*	*b*	*p*	*b*	*p*
(Intercept)	0.18 (0.07)	**0.015**	0.16 (0.07)	**0.024**	0.18 (0.07)	**0.013**	-0.02 (0.08)	0.855
Amerindian_ancestry	-1.60 (0.08)	**<0.001**			-1.57 (0.12)	**<0.001**	-0.86 (0.12)	**<0.001**
African_ancestry	-1.30 (0.03)	**<0.001**			-1.31 (0.11)	**<0.001**	-0.79 (0.11)	**<0.001**
East_Asian_ancestry	0.06 (0.33)	0.863			0.12 (0.33)	0.711	0.21 (0.32)	0.509
South_Asian_ancestry	0.35 (0.56)	0.533			0.44 (0.56)	0.437	0.84 (0.54)	0.122
Child_US_Born	0.14 (0.06)	**0.027**	0.12 (0.06)	0.059	0.13 (0.06)	**0.031**	0.17 (0.06)	**0.005**
Immigrant_Family	0.15 (0.03)	**<0.001**	0.10 (0.03)	**<0.001**	0.15 (0.03)	**<0.001**	0.09 (0.02)	**<0.001**
frac_Black_SIRE			-0.99 (0.03)	**<0.001**	0.01 (0.09)	0.914	0.04 (0.08)	0.652
frac_Native_American_SIRE			-0.38 (0.09)	**<0.001**	-0.09 (0.09)	0.333	0.00 (0.09)	0.978
frac_NOC_SIRE			-0.42 (0.05)	**<0.001**	-0.08 (0.05)	0.097	-0.01 (0.05)	0.807
Hispanic			-0.29 (0.03)	**<0.001**	0.02 (0.04)	0.623	0.06 (0.04)	0.136
State_racism							0.08 (0.04)	0.050
discrim_fact							-0.09 (0.01)	**<0.001**
Skin_color							-0.01 (0.02)	0.525
P_Brown_Eye							0.03 (0.04)	0.376
P_Intermediate_Eye							-0.07 (0.15)	0.654
P_Black_Hair							-0.05 (0.08)	0.541
P_Brown_Hair							0.03 (0.06)	0.638
SES							0.31 (0.01)	**<0.001**
Random Effects								
σ^2	0.42		0.42		0.42		0.42	
τ_{00}	0.43 $_{site\_id\_1:rel\_family\_id}$		0.46 $_{site\_id\_1:rel\_family\_id}$		0.43 $_{site\_id\_1:rel\_family\_id}$		0.35 $_{site\_id\_1:rel\_family\_id}$	
	0.03 $_{site\_id\_1}$		0.02 $_{site\_id\_1}$		0.03 $_{site\_id\_1}$		0.05 $_{site\_id\_1}$	
ICC	0.53		0.53		0.52		0.49	
N	22 $_{site\_id\_1}$		22 $_{site\_id\_1}$		22 $_{site\_id\_1}$		22 $_{site\_id\_1}$	
	8672 $_{rel\_family\_id}$		8672 $_{rel\_family\_id}$		8672 $_{rel\_family\_id}$		8672 $_{rel\_family\_id}$	
Observations	10370		10370		10370		10370	
Marginal R^2 / Conditional R^2	0.178 / 0.609		0.145 / 0.600		0.178 / 0.609		0.247 / 0.615	

To examine if SIRE differences can be accounted for by genetic ancestry, we construct a new set of regression models using the full sample. As seen in Table 5 in the first two models, Model 1a and Model 1b, we include only genetic ancestry variables or alternatively SIRE variables along with controls for migrant status. As seen in Model 2, none of the SIRE values remain significant after adding genetic ancestry to the model. These results indicate that ancestry-associated factors account for the SIRE differences in *g*. We additionally include a Model 3, which adds the cultural, socioeconomic, and phenotypic indices. As seen in Model 3, these variables attenuated the effect of African and Amerindian ancestry (Model 2: $b_{\text{african ancestry}} = -1.31 \rightarrow$Model 3: $b_{\text{african ancestry}} = -0.80$; Model 2: $b_{\text{amerindian ancestry}} = -1.57 \rightarrow$Model 3: $b_{\text{amerindian ancestry}} = -0.86$), but the ancestry effects remain large. Note that the effects of East Asian and South Asian ancestry are insignificant because there is little variance in these ancestry components. This is because we excluded everyone identified as Asian and Pacific Islander.

Finally, we can check the extent to which eduPGS can explain ancestry effects. Before doing so, we verify that eduPGS are associated with *g* within each of the SIRE groups. In doing so, we include controls for the first 20 genetic principal components or, alternatively, continental ancestry (with European ancestry left as the reference). Moreover, we run the analysis both using all families and using only singleton families (i.e., families with only one child). The full results are provided in the supplementary material. The results are summarized in Table 6. As previously found, the eduPGS by *g* associations are attenuated among African Americans, but not among Hispanic and Other Americans (Fuerst et al., 2021). Nonetheless, eduPGS are significantly associated with *g* within all SIRE groups.

Table 6. *Validities (b) of eduPGS by American SIRE groups from multilevel regression models with g as the dependent variable and PGS as a predictor.*

Controls	Sample	Black	Hispanic	Other	White
20 PCs	Full Sample	0.17	0.27	0.32	0.26
Ancestry	Full Sample	0.16	0.27	0.32	0.27
20 PCs	Singletons	0.19	0.29	0.26	0.27
Ancestry	Singletons	0.18	0.29	0.28	0.26

Note: Sample sizes for the full samples are: Black ($N = 1690$), Hispanic ($N = 2021$), Other ($N = 748$) and White ($N = 5911$). The sample sizes for the singletons subsamples are: Black ($N = 1159$), Hispanic ($N = 1516$), Other ($N = 505$), and White ($N = 3674$). All betas are statistically significant at the $p < .01$ level. Singletons = single child families.

For further validation of the PGS, we correlated the eduPGS with the *NIHTBX* summary scores which we had for all groups, including Asians. We computed

mean scores for all ABCD SIRE subgroups and combinations with $N \geq 50$. There were 17 such groups. We then correlated the eduPGS with the mean subgroup test scores. This correlation came to $r = .93$. Thus, we conclude, in line with Chande et al. (2020), that "the general concordance seen between genetically inferred (predicted) phenotypic differences and the observed differences for anthropometric traits, or known prevalence differences in the case of disease traits, supports the approach taken here" (*p*. 1525-6), despite concerns raised in the literature. The regression plot is shown in Figure 6. The number of individuals in each SIRE group is represented by the size of the associated data point.

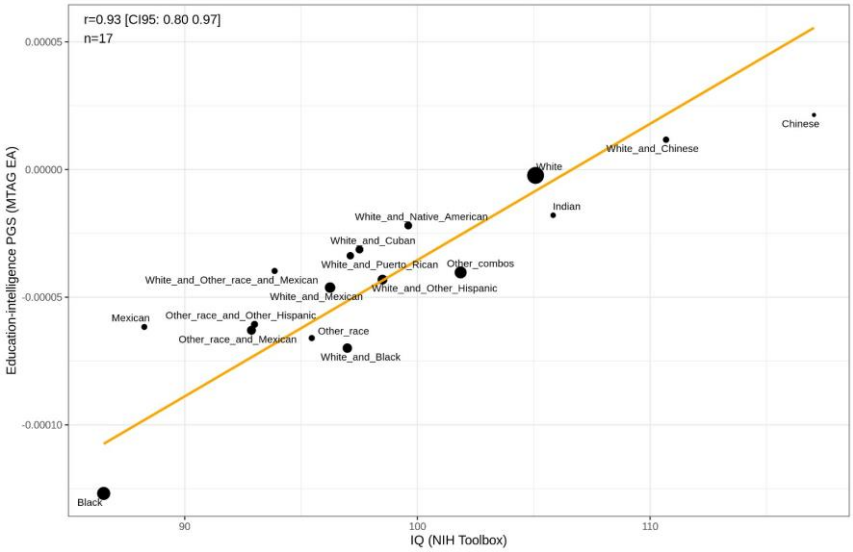

Figure 6. *Regression plot of eduPGS and NIHTBX scores for the 17 largest SIRE groups in the ABCD sample, with SIRE group sample sizes represented by the size of the data points.*

Next, we include the PGS in the model, starting with Model 3 from Table 5. Comparing Model 3 and Model 4 (which adds eduPGS) of Table 7, we see that eduPGS explains a substantial portion of the residual effect of African and Amerindian ancestry after controls for SES.

Table 7. *Regression results for the effect of eduPGS and ancestry on cognitive ability in the full sample. Shown are the beta coefficients (b) and p-values (p) from the mixed effects models with recruitment site and family common factors treated*

as random effects. The values in parentheses are standard errors. The marginal and conditional R^2 are provided at the bottom.

	Model 3		Model 4	
Predictors	*b*	*p*	*b*	*p*
(Intercept)	-0.01 (0.08)	0.934	-0.10 (0.08)	0.243
frac_Black_SIRE	0.04 (0.08)	0.629	0.05 (0.08)	0.582
frac_Native_American_SIRE	0.00 (0.09)	1.000	0.03 (0.09)	0.775
frac_NOC_SIRE	-0.01 (0.05)	0.813	-0.01 (0.05)	0.842
Hispanic	0.06 (0.04)	0.124	0.07 (0.04)	0.082
Child_US_Born	0.16 (0.06)	**0.007**	0.16 (0.06)	**0.005**
Immigrant_Family	0.09 (0.02)	**<0.001**	0.08 (0.02)	**0.001**
Amerindian_ancestry	-0.86 (0.12)	**<0.001**	-0.58 (0.12)	**<0.001**
African_ancestry	-0.79 (0.11)	**<0.001**	-0.40 (0.11)	**<0.001**
East_Asian_ancestry	0.21 (0.32)	0.522	0.34 (0.32)	0.284
South_Asian_ancestry	0.72 (0.54)	0.180	0.87 (0.53)	0.101
State_racism	0.08 (0.04)	0.051	0.09 (0.04)	**0.035**
discrim_fact	-0.09 (0.01)	**<0.001**	-0.08 (0.01)	**<0.001**
Skin_color	-0.01 (0.02)	0.568	-0.00 (0.02)	0.855
P_Brown_Eye	0.03 (0.04)	0.381	0.03 (0.04)	0.506
P_Intermediate_Eye	-0.07 (0.15)	0.639	-0.03 (0.15)	0.847
P_Black_Hair	-0.06 (0.08)	0.493	-0.08 (0.08)	0.326
P_Brown_Hair	0.03 (0.06)	0.619	0.04 (0.06)	0.527
SES	0.31 (0.01)	**<0.001**	0.28 (0.01)	**<0.001**
eduPGS			0.20 (0.01)	**<0.001**
Random Effects				
σ^2	0.42		0.42	
τ_{00}	$0.35_{\text{site\_id\_1:rel\_family\_id}}$		$0.34_{\text{site\_id\_1:rel\_family\_id}}$	
	$0.05_{\text{site\_id\_1}}$		$0.05_{\text{site\_id\_1}}$	
ICC	0.49		0.48	
N	$22_{\text{site\_id\_1}}$		$22_{\text{site\_id\_1}}$	
	$8672_{\text{rel\_family\_id}}$		$8672_{\text{rel\_family\_id}}$	
Observations	10370		10370	
Marginal R^2 / Conditional R^2	0.247 / 0.615		0.265 / 0.619	

We also examine the individual SIRE subsample results for eduPGS. The model adds eduPGS to the respective Model 4s for each SIRE group (i.e., the model with potential environmental factors included). The full results are provided in the Supplementary File. These results are summarized in Table 8. Specifically, Table 8 shows the effects for Amerindian and African ancestry on *g* with possible environmental controls. These come from the fourth models of Tables 2, 3, 4, and S8 and the third model from Table 5. It next shows the effects when eduPGS is added. As seen, eduPGS accounts for a portion of the ancestry by *g* association in all SIRE subsamples.

Table 8. *Effects (b) of Amerindian and African ancestry on g in multi-level models with environmental controls (Model 4/3), and multi-level models with environmental controls and eduPGS (Model 5).*

		Amerindian	African
Black	Model 4	-3.46	-0.89
	Model 5	-3.15	-0.69
Hispanic	Model 4	-0.86	-0.58
	Model 5	-0.55	-0.16
Other	Model 4	-1.09	-1.08
	Model 5	-0.55	-0.48
White	Model 4	-0.96	-0.85
	Model 5	-0.65	-0.32
Full	Model 4	-0.86	-0.79
	Model 5	-0.58	-0.40

It is conceptually possible that our eduPGS are just capturing global ancestry effects. Our ancestry components are based on more SNPs. Moreover, they are not weighted by trait-associations which will attenuate the association with ancestry. As such this is unlikely. However, to test this possibility we created pseudoPGS. To do so, we used PLINK v1.90b6.8 to select random sets of 8,898 variants to match the eduPGS. Then we randomly assigned the eduPGS beta weights (from Lee et al., 2018) to the respective sets of SNPs.

Following this procedure, we create 10 pseudo eduPGS scores. This procedure produced PGS with the same set of SNPs as the SNPs used to calculate genetic ancestry, but randomized trait-association information. The full results are provided in Supplementary File 3, Table S17. Unlike the real PGS, these pseudoPGS had no validity independent of genetic ancestry. This is because of the random assignment of eduPGS betas to the SNP frequencies resulting in poor indices of ancestry. Generally, we conclude that PGS will not

necessarily capture effects of global ancestry. This finding suggests that our eduPGS are in fact capturing causal genetic effects on *g* both within and between ancestries.

Discussion

Genetic ancestry measures provide very powerful scientific value in studying SIRE differences in *g*. Using ancestry allows one to examine how the trait varies by genetic ancestry within self-identified racial and ethnic groups. Doing so offers a potential solution to the problem of decomposing genetic and environmental variance (Halder et al., 2015). Admixture regression has been widely applied to medical and behavioral traits. This includes type 2 diabetes (Cheng et al., 2013), asthma (Salari et al., 2005), blood pressure (Klimentidis et al., 2012), and sleep depth (Halder et al., 2015). Admixture regression has a natural application to studying *g*.

Here we apply this technique to examine SIRE differences in *g*. We find that African and Amerindian ancestry are strongly negatively associated with general cognitive ability among African, Hispanic, and other American subsamples. This replicates previous research which showed that genetic ancestry predicts cognitive ability, independent of socioeconomic status and phenotypic discrimination variables which are the usual suspects (Kirkegaard et al., 2019; Lasker et al., 2019; Warne, 2020). The importance of such analyses *within* SIRE groups is that they shed light on the cause of *g* differences *between* SIRE groups with respect to similarities in developmental processes (Rowe, Vazsonyi & Flannery, 1994).

The ancestry effects are consistent in direction across subsamples and hold after controlling for a wide array of economic and social factors, including migrant status, SIRE, ethnic attachment, measures of discrimination, phenotypic indices of race, and general SES. These results suggest that African, Hispanic, and other groups have inherited alleles from their African and Amerindian ancestors which make them liable to lower levels of *g*. In fact, as seen in Table 5 (Model 2), 100%, 76%, 81%, and 100% of the respective Black, Native American, Other, and Hispanic SIRE effects were explained by genetic ancestry. This association between genetic ancestry and *g* suggests a partial genetic basis for observed SIRE differences.

This inference is supported by additional findings based on the eduPGS analyses. These polygenic scores were found to be predictive of *g* within SIRE groups controlling for the first 20 principal components and for ancestry. Moreover, they explain a substantial portion of the ancestry effects both in the full sample and all subsamples. Also, they were almost perfectly correlated with SIRE group means in cognitive ability ($r = .93$). The most parsimonious explanation for

this, given the apparent absence of obvious forms of confounding (Fuerst et al., 2021), would seem to be that eduPGS are capturing causal effects of genes on *g* both within and between ancestry groups and thus also SIRE groups. Firm conclusions, though, will require a better understanding of the relation between polygenic scores and ancestry (Lawson et al., 2020; Fuerst et al., 2021).

It is worth emphasizing that our *g* scores were from a confirmatory factor model in which strict factorial invariance (SFI) held between SIRE groups. SFI entails that the differences between SIRE groups have the same psychometric meaning as the differences between individuals within these groups (i.e., the scores are psychometrically unbiased). Moreover, SFI implies that the causes of group differences are a subset of the causes of the individual differences within groups (Lubke et al., 2003; Dalliard, 2014). In this sample of children, individual differences in general cognitive ability are largely due to genes (Freis et al., 2020).

It should be noted that the polygenic scores represent genetic variation that is caused by common alleles, not genetic variation that is caused by rare alleles under mutation-selection balance. The causal alleles that are tapped by polygenic scores are ancient. Most were already polymorphic 60,000 years ago when people left Africa and spread all over Eurasia. Today's racial allele frequency differences are the cumulative effects of selection and genetic drift acting over more than 2,000 generations, while rare variants under mutation-selection balance are much younger, no more than one or two millennia or even less. Therefore it is predictable that genetic race differences that evolved over a long time are differences in polygenic scores but not necessarily differences in mutational load. The latter are the result of strength of selection during the last centuries.

Overall, the results suggest that genetic variants related to general cognitive ability vary between source genetic populations and have a causal effect on intelligence. Because individuals within SIRE groups differ in their proportion of African, European and Amerindian ancestors, general cognitive ability varies by genetic ancestry within SIRE groups.

Limitations

This study advances over previous studies in that we used a diverse national sample, a good measure of *g*, multiple indices of racial discrimination including multiple race-associated phenotypes, and a composite index of SES based on seven different indices. Moreover, our multilevel model controlled for the effects of geography. Unfortunately, our index of skin color was imperfect. However, it seems unlikely that skin color discrimination is a significant *immediate* cause of *g* differences among 9-10 year old children. Such color discrimination explanations usually propose labor market based discrimination (Hersch, 2011), which would

be captured by our index of SES. Regardless, admixture-regression results can only provide indirect evidence for a genetic hypothesis because there could be unmeasured environmental factors that are related to both ancestry and cognitive ability.

While the results also show that educational and intelligence-related polygenic scores can account for some of the effects of ancestry on *g*, these results are only tentative. It is not certain that these PGS are capturing genetic effects, at least between ancestries (Fuerst et al., 2021). Thus these results do not provide definitive evidence for a genetic hypothesis. However, following the methodology of genetic epidemiology, admixture regression analyses are just a first step in elucidating the genetic and environmental causes of group differences.

Author contributions

All analyses were conducted by JGRF. MH and GC helped prepare the manuscript.

Acknowledgments

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children aged 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data, but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the research results and interpretations of the authors alone and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from Version 3.01. The raw data are available at https://nda.nih.gov/edit_collection.html?id=2573. Instructions on how to create an

NDA study are available at https://nda.nih.gov/training/modules/study.html). Additional support for this work was made possible from supplements to U24DA041123 and U24DA041147, the National Science Foundation (NSF 2028680), and Children and Screens: Institute of Digital Media and Child Development Inc.

Additionally, we thank Jordan Lasker for guidance with the MGCFA analysis code.

All supplemental files are available online at:
http://www.mankindquarterly.org/files/papers/62/1/11_sup_x0CHZFC.pdf

References

Belsky, D.W., Domingue, B.W., Wedow, R., Arseneault, L., Boardman, J.D., Caspi, A., ... & Harris, K.M. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences* 115: E7275-E7284.

Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D. & Mountain, J.L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics* 96: 37-53.

Chande, A.T., Rishishwar, L., Ban, D., Nagar, S.D., Conley, A.B., Rowell, J., ... & Jordan, I.K. (2020). The phenotypic consequences of genetic divergence between admixed Latin American populations: Antioquia and Chocó, Colombia. *Genome Biology and Evolution* 12: 1516-1527.

Cheng, C.Y., Reich, D., Haiman, C.A., Tandon, A., Patterson, N., Elizabeth, S., ... & Kao, W.L. (2012). African ancestry and its correlation to type 2 diabetes in African Americans: A genetic admixture analysis in three U.S. population cohorts. *PLoS One* 7(3): e32840.

Choi, S.W., Mak, T.S.H. & O'Reilly, P.F. (2020). Tutorial: A guide to performing polygenic risk score analyses. *Nature Protocols* 15: 2759-2772.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Abingdon, UK: Routledge.

Conley, D. & Fletcher, J. (2017). *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future*. Princeton Univ. Press: Princeton, NJ.

Chavent, M., Kuentz-Simonet, V., Labenne, A. & Saracco, J. (2014). Multivariate analysis of mixed data: The R Package PCAmixdata. *arXiv preprint arXiv:1411.4911*. https://arxiv.org/pdf/1411.4911.pdf

Dalliard, M. (2014). The elusive x-factor: A critique of J.M. Kaplan's model of race and IQ. *Open Differential Psychology*: 1-44.

Fang, H., Hui, Q., Lynch, J., Honerlaw, J., Assimes, T.L., Huang, J., ... & DuVall, S.L. (2019). Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. *American Journal of Human Genetics* 105: 763-772.

Freis, S., Morrison, C., Lessem, J.M., Hewitt, J.K. & Friedman, N.P. (2020). Genetic and environmental influences on executive functions and intelligence in the ABCD study. PsyArXiv. https://psyarxiv.com/xpba4/

Fuerst, J.G.R., Kirkegaard, E.O., Piffer, D. (2021). More research needed: There is a robust causal vs. confounding problem for intelligence-associated polygenic scores in context to admixed American populations. *Mankind Quarterly*, this issue.

Halder, I., Matthews, K.A., Buysse, D.J., Strollo, P.J., Causer, V., Reis, S.E. & Hall, M.H. (2015). African genetic ancestry is associated with sleep depth in older African Americans. *Sleep* 38: 1185-1193.

Heeringa, S.G. & Berglund, P.A. (2020). A guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) Study baseline data. *BioRxiv*.

Hersch, J. (2011). The persistence of skin color discrimination for immigrants. *Social Science Research* 40: 1337-1349.

Hu, M., Lasker, J., Kirkegaard, E.O. & Fuerst, J.G. (2019). Filling in the gaps: The association between intelligence and both color and parent-reported ancestry in the National Longitudinal Survey of Youth 1997. *Psych* 1(1): 240-261.

Kim, M.S., Patel, K.P., Teng, A.K., Berens, A.J. & Lachance, J. (2018). Genetic disease risks can be misestimated across global populations. *Genome Biology* 19(1): 1-14.

Kirkegaard, E.O., Williams, R.L., Fuerst, J. & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability and socioeconomic outcomes. *Psych* 1(1): 1-25.

Klimentidis, Y.C., Dulin-Keita, A., Casazza, K., Willig, A.L., Allison, D.B. & Fernandez, J.R. (2012). Genetic admixture, social-behavioural factors and body composition are associated with blood pressure differently by racial-ethnic group among children. *Journal of Human Hypertension* 26: 98-107.

Krapohl, E. & Plomin, R. (2016). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Molecular Psychiatry* 21: 437-443.

Lasker, J., Pesta, B.J., Fuerst, J.G. & Kirkegaard, E.O. (2019). Global ancestry and cognitive ability. *Psych* 1(1): 431-459.

Lawson, D.J., Davies, N.M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., ... & Timpson, N.J. (2020). Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* 139: 23-41.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & 23andMe Research Team (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nature Genetics* 50: 1112-1121.

Liebler, C.A. & Halpern-Manners, A. (2008). A practical approach to using multiple-race response data: A bridging method for public use microdata. *Demography* 45: 143-155.

Lubke, G.H., Dolan, C.V., Kelderman, H. & Mellenbergh, G.J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence* 31: 543-566.

Magnuson, K.A. & Duncan, G.J. (2006). The role of family socioeconomic resources in the black–white test score gap among young children. *Developmental Review* 26: 365-399.

Murray, C. (2021). *Facing Reality: Two Truths about Race in America.* Encounter Books.

Pesta, B.J., Kirkegaard, E.O., te Nijenhuis, J., Lasker, J. & Fuerst, J.G. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence* 78: 101408.

Piffer, D. (2019). Evidence for recent polygenic selection on educational attainment and intelligence inferred from GWAS hits: A replication of previous findings using recent data *Psych* 1(1): 55-75.

Rindermann, H., Becker, D. & Coyle, T.R. (2020). Survey of expert opinion on intelligence: Intelligence research, experts' background, controversial issues, and the media. *Intelligence* 78: 101406.

Roth, P.L., Van Iddekinge, C.H., DeOrtentiis, P.S., Hackney, K.J., Zhang, L. & Buster, M.A. (2017). Hispanic and Asian performance on selection procedures: A narrative and meta-analytic review of 12 common predictors. *Journal of Applied Psychology* 102: 1178-1202.

Rowe, D.C., Vazsonyi, A.T. & Flannery, D.J. (1994). No more than skin deep: Ethnic and racial similarity in developmental processes. *Psychological Review* 101: 396-413.

Rowe, D.C., Vesterdal, W.J. & Rodgers, J.L. (1998). Herrnstein's syllogism: Genetic and shared environmental influences on IQ, education, and income. *Intelligence* 26: 405-423.

Salari, K., Choudhry, S., Tang, H., Naqvi, M., Lind, D., Avila, P.C., ... & Ziv, E. (2005). Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. *Genetic Epidemiology* 29: 76-86.

Savage, J.E., Jansen, P.R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C.A., Nagel, M., Awasthi, S., Barr, P.B. & Coleman, J.R. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics* 50: 912-919.

Scheiber, C. (2016a). Do the Kaufman tests of cognitive ability and academic achievement display construct bias across a representative sample of Black, Hispanic, and Caucasian school-age children in grades 1 through 12? *Psychological Assessment* 28: 942-952.

Scheiber, C. (2016b). Is the Cattell–Horn–Carroll-based factor structure of the Wechsler Intelligence Scale for Children—Fifth Edition (WISC-V) construct invariant for a representative sample of African-American, Hispanic, and Caucasian male and female students ages 6 to 16 years? *Journal of Pediatric Neuropsychology* 2(3): 79-88.

Thompson, W.K., Barch, D.M., Bjork, J.M., Gonzalez, R., Nagel, B.J., Nixon, S.J. & Luciana, M. (2019). The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: Findings from the ABCD study's baseline neurocognitive battery. *Developmental Cognitive Neuroscience* 36: 100606.

Warne, R.T. (2020a). *In the Know: Debunking 35 Myths about Human Intelligence*. Cambridge Univ. Press.

Warne, R.T. (2020b). Continental genetic ancestry source correlates with global cognitive ability score. *Mankind Quarterly* 60: 400-422.

Weiss, L.G. & Saklofske, D.H. (2020). Mediators of IQ test score differences across racial and ethnic groups: The case for environmental and social justice. *Personality and Individual Differences* 161: 109962.