bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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.

Biogeographic Ancestry and Socioeconomic Outcomes in the Americas: A Meta-Analysis

Emil O. W. Kirkegaard[1]
Ulster Institute for Social Research, London, UK

Mingrui Wang
Department of Biology, Beijing University of Agriculture, China

John Fuerst
Ulster Institute for Social Research, London, UK

Narrative reports suggest that socioeconomic status (SES) is associated with biogeographic ancestry (BGA) in the Americas. If so, SES potentially acts as a confound that needs to be taken into account when evaluating the relation between medical outcomes and BGA. To explore how systematic BGA-SES associations are, a meta-analysis of American studies was conducted. 40 studies were identified, yielding a total of 64 independent samples with directions of associations, including 48 independent samples with effect sizes. An analysis of association directions found a high degree of consistency. The square root *n*-weighted directions were 0.83 (*K* = 36), -0.81 (*K* = 41) and -0.82 (*K* = 39) for European, Amerindian and African BGA, respectively. An analysis of effect size magnitudes found that European BGA was positively associated with SES, with a meta-analytic effect size of *r* = .18 [95% CI: .13 to .24, *K* = 28, *n* = 35,476.5], while both Amerindian and African BGA were negatively associated with SES, having meta-analytic effect sizes of -.14 [-.18 to -.10, *K* = 31, *n* = 28,937.5] and -.11 [-.15 to -.07, *K* = 28, *n* = 32,710.5], respectively. There was considerable cross-sample variation in effect sizes (mean I^2 = 92%), but the sample size was not enough for performing credible moderator analysis. Implications for future studies are discussed.

Keywords: Biogeographic ancestry, admixture, self-identified race/ethnicity, SES, medical outcomes

Admixture analysis is a potent tool for the exploration of the etiology of traits and trait differences in admixed populations. Admixture mapping is a form of admixture analysis that allows for the detection of specific disease- and trait-associated genes (Shriner, 2013; Winkler, Nelson & Smith, 2010). When biogeographic ancestry (BGA) groups differ in the frequency of disease- or trait-causing genetic variants, in admixed populations the phenotype of interest will be correlated with BGA in genomic regions near the causal genetic variants, allowing for the identification of associated loci. Such associations will occur *within* self-identified race/ethnicity (SIRE) groups. When the genetic architecture of a trait is complex, with hundreds of loci assumed to contribute to the phenotype, an analysis of global admixture is an appropriate first step. This strategy seeks to identify associations of global admixture proportions with the measured trait, without attempting to identify individual loci. Unlike admixture mapping, which requires sample sizes of many thousands, global admixture analysis can produce meaningful results with smaller sample sizes.

[1] Corresponding author, email: emil@emilkirkegaard.dk

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

As used here, *population* refers to geographically delineated population. In contrast, *BGA* refers to genetic ancestry as defined by the presence of ancestrally informative molecular markers in a person's genome, which allow the calculation of an individual's resemblance to the average genotypes in the ancestral populations. These ancestral populations have also been called clusters or ancestral groups (e.g., Frudakis, 2010; Frudakis & Shriver, 2003; Shriver & Kittles, 2004). With respect to studies of American populations, the ancestral populations are typically Europeans, Sub-Saharan Africans, and Amerindians (Salzano & Sans, 2014).

Socioeconomic status (SES) inequalities between and within SIRE groups can lead to spurious associations of BGA with medical outcomes when both are associated with SES (Marden et al., 2016). Thus, genetic traits associated with SES whose allele frequencies differ among ancestral groups can be misidentified as being associated with a specific medical outcome in admixture mapping. For this reason, controls for SES are not infrequently incorporated into such analyses. Narrative reports (e.g., González Burchard et al., 2005) have suggested that SES covaries with admixture such that, relative to European BGA, Amerindian and African BGA are associated with lower SES. If this is typically the case, it would be prudent for researchers to include reliable measures of SES as covariates in analyses to provide lower-bound estimates of the BGA-medical outcome associations. Moreover, it would be advisable to investigate the causal pathways mediating the BGA-SES associations, to identify possible unobserved non-genetic mediators of BGA-medical outcomes associations. However, no meta-analysis has been conducted to date to establish whether SES outcomes are associated with BGA in any consistent way.

1. Methods

1.1. Collection of studies and data exclusion

Papers were sought which reported quantitative or qualitative associations between European, African, or Amerindian BGA and socioeconomic outcomes. In 2014/15, a literature search was conducted using the PubMed and BIOSIS previews databases. Searches such as the following were used: '(admixture) AND (socioeconomic or education or income or SES or poverty) AND (African or European or Amerindian)'. A search, limited to the years 2003 to early 2016, was also conducted using Google Scholar. For the latter, phrases such as '(ses, OR income, OR education) AND (genomic OR biogeographical, OR ancestry, OR admixture) AND (African, OR European, OR Amerindian)' were used. The Google Scholar search yielded over twenty thousand hits in descending order of relevance to the search terms; the first 1,500 abstracts were skimmed and approximately 250 papers were identified as potential sources and read. The references of the identified papers, from the PubMed, BIOSIS, and Google Scholar searches, were inspected for additional sources. In total, we identified 54 papers (51 from the database search and three from the reference search). Of these, six used redundant samples and thus were excluded. For eight of the papers, while SES was a covariate in an admixture analysis, no usable information was recoverable and authors did not reply to emails sent to their listed emails. The remaining 40 papers either contained directions of associations or these were provided by the authors when contacted; for 27 of the 40 papers, either effect sizes (or metrics that could be converted into effect sizes) were reported or provided by the authors. Figure 1 shows a flowchart of the procedure.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

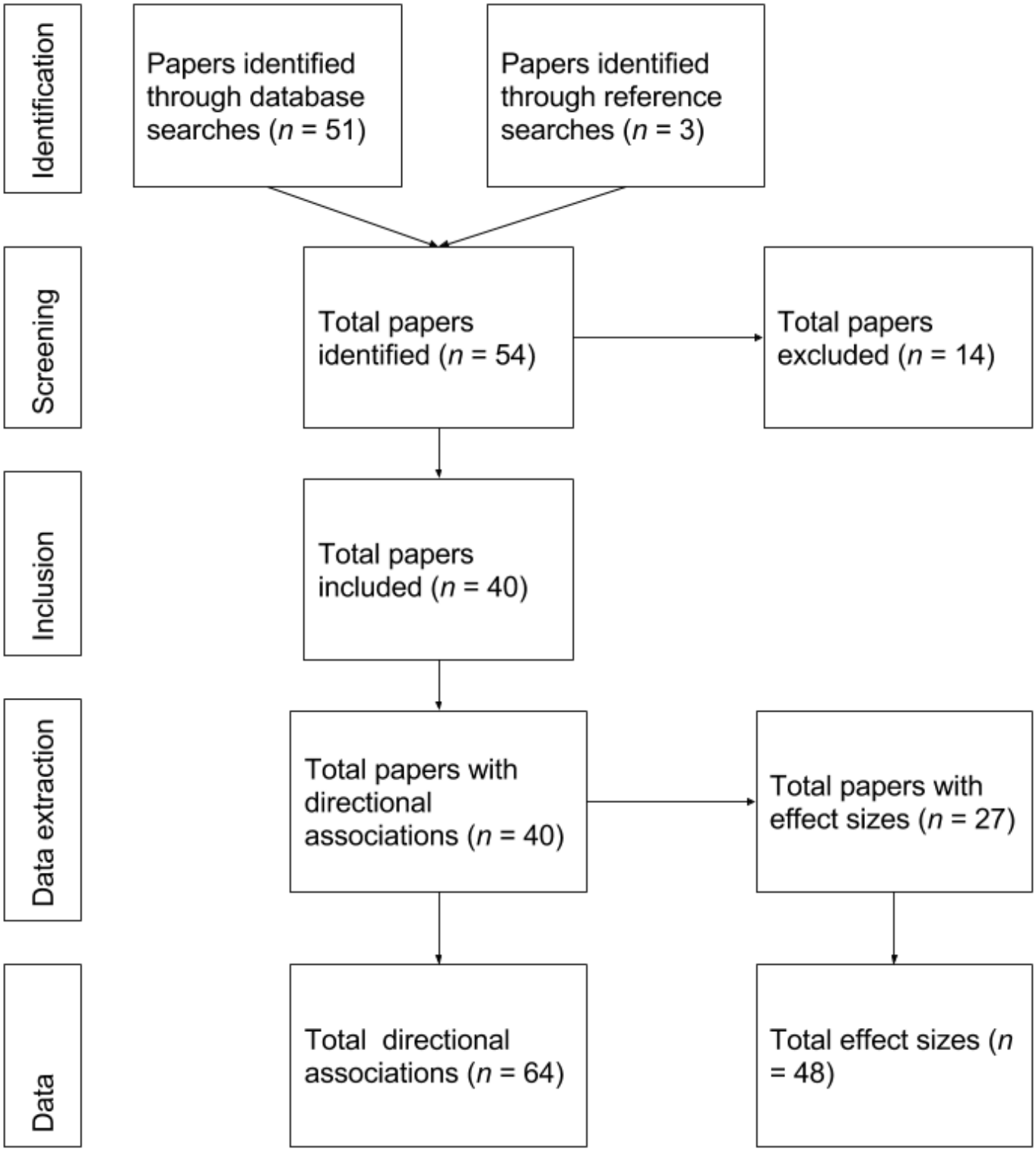

Figure 1. Flowchart of study inclusion.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

The identified studies and reasons for exclusions are listed in Supplementary File 2. Relevant information from each study and sample was recorded and is reported in Supplementary File 1. With respect to coding, two of this paper's authors reviewed studies and reached agreement on ambiguous cases. Many studies did not report effect sizes or statistics convertible to effect sizes; some studies failed to clearly report association directions. For studies that did not report effect sizes or directions of associations, authors were contacted. Contact dates, details, and notes are reported in Supplementary File 1 along with annotated replies when forthcoming. Seventy-three percent of our requests for data elicited a reply. For 38 percent of requests the authors provided results not reported in the original papers.

Most papers were published in the last few years, with the median year of publication being 2015 (range: 1988 to 2016). Many studies had medical themes and often included both case and control samples. Table 1 shows the independent samples broken down by sample type. Samples were coded as 'medical' versus 'non-medical' based on the authors' research goal. As a result, for some samples, this classification was somewhat arbitrary. For example, Zou et al. (2015) explored genetic assortative mating using the Genes-Environments and Admixture in Latino Americans (GALA-II) community survey. Since the outcome of interest was 'non-medical', the samples were classified as 'non-medical'. However, if the same authors had instead explored the relation between BGA and hypertension, using the same samples, and if results were undifferentiated by clinical conditions, the samples would have been classified as 'case and control' (meaning samples which include both cases and controls).

Table 1. Independent samples by sample type

Sample type	Direction of associations available	Effect sizes available
Case	10	8
Case and control	24	19
Control	8	4
Non-medical	22	17
Total	64	48

For the meta-analysis, we report results based on associations for which the relevant BGA proportions are above 5%. For comparison, we also note the results with low-admixture samples included. Associations in cases of low admixture were excluded because when means and thus variances of admixture are low, effect sizes will be attenuated. Regarding unreliability, one concern is that the present meta-analysis is based on correlation coefficients. As such, the statistical effects of other BGA components are not held constant and covariance between components can confound associations, leading to spurious results. Possible confounding is more of a concern when admixture fractions are low. With this exclusion, the total number of independent samples (64 for directions of associations and 48 for effect sizes) remained the same, but the number of samples when decomposed by specific BGAs was reduced. The number of samples is shown in Table 2.

Table 2. Number of independent associations by BGA without (w/o) and with (w) low-admixture samples.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

	Independent samples	Associations (African)	Associations (Amerindian)	Associations (European)
Directional associations (w/o low admix)	64	39	41	36
Directional associations (w low admix)	64	43	43	36
Effect sizes (w/o low admix)	48	28	31	28
Effect sizes (w low admix)	48	31	33	28

Only a subset of the independent samples reported associations for all three BGAs. As a result, the number of associations for a particular BGA is less than the number of independent samples. When studies reported associations between outcomes and one or two BGA components, one could attempt to estimate association directions and effect sizes for the remaining BGAs since the three BGAs are directly related to one another, with ancestries adding up to unity. For instance, Bonilla (2015, extra information provided by the author) reported that European and Amerindian BGA correlated at -.82 and -.85 in two samples (*n*s = 148, 164). The correlations between European BGA and SES were .10 and .14, respectively; thus, one might infer that the Amerindian × SES correlations were -.10 and -.14. In this case, however, they were -.01 and -.13; the departure from expectation resulted from the association between African BGA and SES. To avoid estimation error, unreported associations and effect sizes were not estimated.

1.2. Descriptive statistics of studies

In some instances, associations with multiple outcome measures were reported for a single sample and the same BGA. Given this situation, treating BGA × outcome associations as independent data points would lead to double counting. Two recent meta-analyses encountered an analogous problem (Tucker-Drob & Bates, 2015; Tate & McDaniel, 2008). The method employed by the first was to use a complex weighting approach to avoid double counting, while Tate and McDaniel (2008) used a simpler approach of averaging results within samples before aggregating. An approach similar to the second was implemented in the present study. Median values, which are more robust to outliers than are means, were taken across outcome values for the same BGA within each sample before meta-analyzing the sample associations. Different classes of outcomes (e.g., education vs. income) were not weighted differently. As Tate and McDaniel (2008) point out, this method slightly throws off the standard errors, but this problem was judged to be minor. Henceforth, all results are reported for independent samples (with the exception of Table 3), with BGA × outcome associations having been first averaged within each sample.

For the 64 independent samples, there were 179 BGA × outcome associations, since some samples included data for multiple BGAs and multiple SES indicators. Studies reported a large variety of specific outcome variables (Table 3). As shown in Supplementary File 1, these variables were coded into broad categories. Supplementary Table 1 shows the breakdown of all outcomes by broad category. Most measures were of individual education, income or occupation. A few BGA × SES associations were based on the socioeconomic level of the participant's neighborhood. Depending on the model of the proposed covariance, this could be a questionable index. For example, a recent study based on the UK Biobank (N ≈ 112k) found a modest phenotypic correlation between individual and neighborhood-level SES (.24) but a

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

strong genetic one (.87) (Hill et al., 2016). While neighborhood-level measures are included in the meta-analysis, it is advisable that future investigators conduct moderator analyses to estimate the impact of using neighborhood versus individual indicators. Additionally, several studies reported parental SES. Because most children are the biological offspring of both of their parents, their admixture will index the average of their parents, and these results can then be seen as showing the correlation between the parents' SES and the parents' BGA. Roughly sixty percent of the 64 independent samples came from the US. Studies often divided their samples into SIRE groups. Tables 4a and 4b show a breakdown by country and SIRE group.

Table 3. Broad outcomes (specific outcomes) for all BGA × outcome associations with/without low admixture excluded

Broad outcome (specific outcomes)	Frequency	Broad outcome (specific outcomes)	Frequency
Education (education, years of schooling)	69/73	Parental SES (parental Hollingshead index, running water, had a car)	6/6
Income (income, wealth index, household income per capita)	49/53	Parental education	2/2
SES (SES, SES + education, occupation, general factor of SES, Hollingshead index, Yost index, home property band, household asset index)	34/35	Parental income	1/1
Neighborhood SES	18/19	Total	179/189

Table 4a. Number of independent samples and sample sizes for directional association analysis by SIRE group and country. Low admixture excluded.

SIRE group	*K*	*N*	SIRE group	*K*	*N*	SIRE group	*K*	*N*
African American (US)	10	20581	Multi-ethnic (Chile)	2	1967	African-descent (T&T)	1	107
Hispanic (US)	10	9983	Multi-ethnic (Mexico)	2	1991	Multi-ethnic (Argentina)	1	220
Mexican (US)	5	1050	Multi-ethnic (US)	2	2031	Multi-ethnic (Colombia)	1	2092
Multi-ethnic (Brazil)	6	6696	Mestizo (Chile)	2	2381	Not stated (Costa Rica)	1	1998
Puerto Rican (PR)	7	2368	Native American (US)	2	741	Puerto Rican (US)	1	122
Not stated (Mexico)	3	1831	Not stated (Uruguay)	2	312	White American (US)	1	73
Multi-ethnic (Peru)	3	1979	Not stated (Colombia)	2	1971	Total	64	

Note: *N* = participant sample size; *K* = number of independent samples. PR = Puerto Rico; T&T = Trinidad and Tobago.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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 4b. Number of independent samples and sample sizes for effect size analysis by SIRE group and country. Low admixture excluded.

SIRE group	*K*	*N*	SIRE group	*K*	*N*	SIRE group	*K*	*N*
African American (US)	7	13589	Multi-ethnic (Chile)	2	1967	African-descent (T&T)	0	
Hispanic (US)	6	8041	Multi-ethnic (Mexico)	2	1991	Multi-ethnic (Argentina)	1	220
Mexican (US)	5	1050	Multi-ethnic (US)	2	2031	Multi-ethnic (Colombia)	1	2092
Multi-ethnic (Brazil)	5	6518.5	Mestizo (Chile)	0		Not stated (Costa Rica)	0	
Puerto Rican (PR)	3	1943	Native American (US)	2	741	Puerto Rican (US)	1	122
Not stated (Mexico)	3	1831	Not stated (Uruguay)	2	312	White American (US)	1	73
Multi-ethnic (Peru)	3	1979	Not stated (Colombia)	2	1971	Total	48	

1.3. Directions of associations

As shown in Table 2, directions of associations were more frequently available than were actual effect sizes. Association directions can provide some indication as to whether the findings are in line with a null hypothesis. Other meta-analyses have included analyses of association directions (e.g., Van der Meer & Tolsma, 2014).

1.4. Statistical analysis

For each sample, directions (negative, null, positive) were assigned to associations. With respect to this assignment, neither *p*-values nor the magnitude of effects were taken into account. Thus, for example, the correlation between African BGA and education for Bonilla et al.'s (2015) sample 1 was -.16 (n = 164, p = .08), while the correlation between African BGA and education for Bonilla et al.'s (2015) sample 2 was .04 (n = 148, p = .59); the first association was coded as 'negative' and the second as 'positive' despite the difference in effect size. Those samples coded as 'null' were ones where the effect size was zero. After directions were assigned, they were coded as negative = –1, null = 0, and positive = 1). If a sample had consistent directions (across multiple SES indicators), it received a score of either -1 or 1. If a sample had mixed results, it received an intermediate score equivalent to the mean of the different associations (e.g., a sample reporting three positive and one negative associations received a score of (3-1)/4=0.5.) The results are reported unweighted; weighted by the square root of the sample sizes; and weighted by the sample sizes. We consider square root *n*-weighted results to be preferable as they strike a good balance between taking into account the effect of sample size and not obscuring results from smaller samples; other researchers have employed this weighting strategy (for example, Kan et al., 2013).

2. Results

2.1. Directions of associations

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

The aggregated within-sample directions are shown in Table 5. In general, the directions are positive for European BGA and negative for the other two ancestries. With respect to African and Amerindian BGA, the weighted results are substantially stronger when low-admixture samples are excluded. There were no samples with low European BGA.

Table 5. Directions of associations by BGA

BGA	*N*	*K*	Mean direction (unweighted)	Mean direction (SQRT(*n*)-weighted)	Mean direction (*n*-weighted)
African (w/o low admix)	40931.5	39	-0.67	-0.82	-0.94
African (w low admix)	47742.5	43	-0.75	-0.64	-0.56
Amerindian (w/o low admix)	35577.5	41	-0.76	-0.81	-0.85
Amerindian (w low admix)	47107.5	43	-0.72	-0.65	-0.41
European (w/o low admix)	39548.5	36	0.72	0.83	0.89
European (w low admix)	39548.5	36	0.72	0.83	0.89

2.2. Effect sizes

For 92% of the samples with effect sizes, Pearson correlations or other effect sizes that could be directly converted into these (i.e., r^2, Spearman's rho, beta coefficients, and odds ratio) were available. For other samples, effect sizes needed to be computed from *t*-tests (N = 1), *F*-tests (N = 4), and frequency tables (N = 5). Conversions were made using the formulas and methods noted in Supplementary File 1. To provide an overview of the effect sizes, a density histogram plot was made of the correlations for each BGA component. Additional statistics for the effect sizes were also computed. For the meta-analysis, because samples were from different populations in different countries, a random effects model was appropriate (Schmidt & Hunter, 2014). We conducted the analysis using the *metafor* package for R (Viechtbauer, 2015).

The studies varied substantially in sample size. Figure 2 shows the distribution of the sample sizes for the effect size analysis with low-admixture samples excluded. The density histogram plot of the effect sizes is shown in Figure 3. Descriptive statistics for the effect sizes with and without low admixture samples excluded are shown in Table 6.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

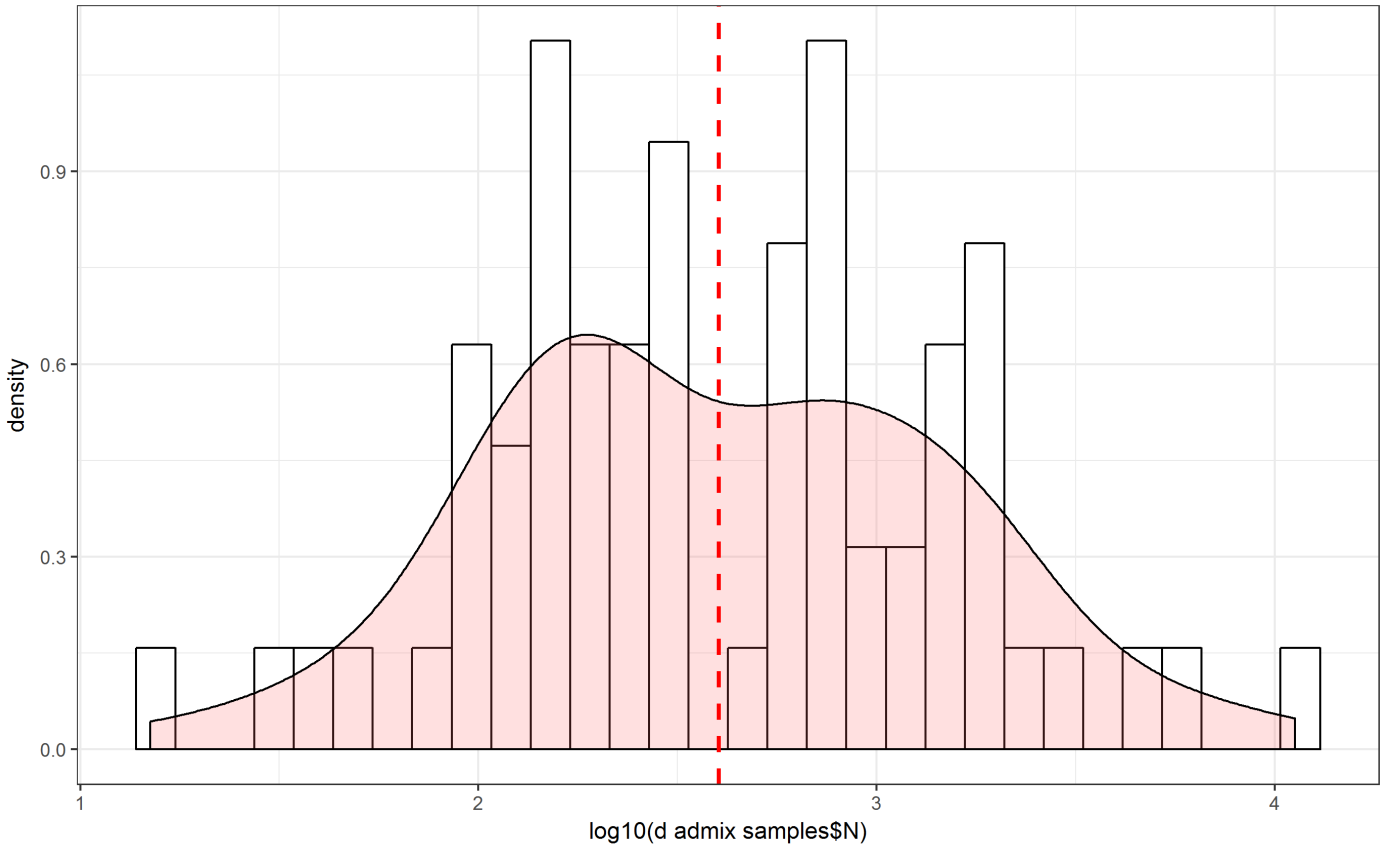

Figure 2. Density histogram of sample sizes, with a stippled line at the mean.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

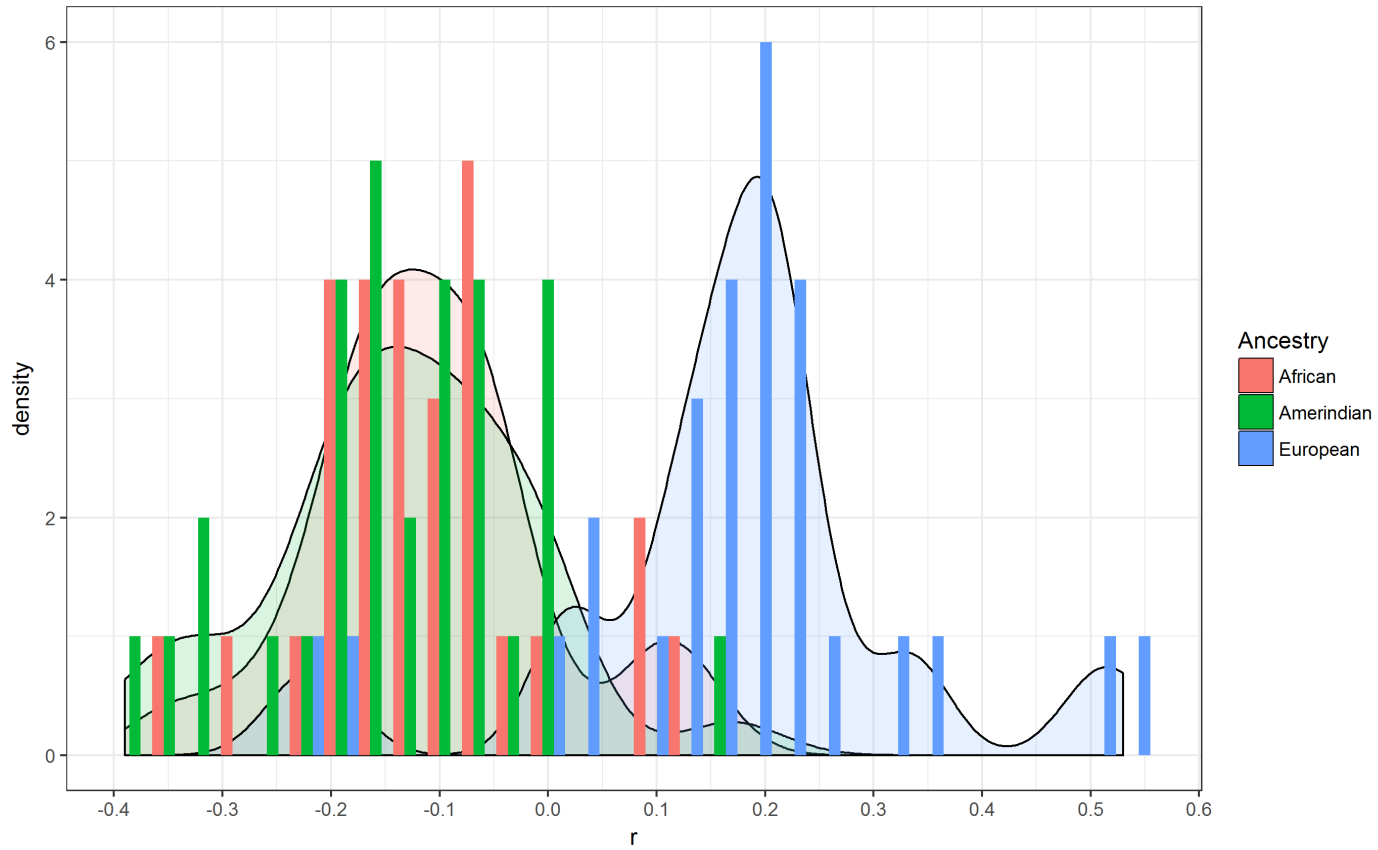

Figure 3. Density histogram plot for effect sizes by BGA.

Table 6. Descriptive statistics for the effect sizes

BGA	*N*	*K*	Mean	Median	Max	Min	*SD*	10th centile	90th centile
African (w/o low admix)	32710.5	28	-.11	-.12	.14	-.35	.11	-.21	.03
African (w low admix)	37523.5	31	-.09	-.10	.14	-.35	.11	-.20	.08
Amerindian (w/o low admix)	28937.5	31	-.13	-.13	.17	-.39	.12	-.31	.00
Amerindian (w low admix)	40467.5	33	-.13	-.12	.17	-.39	.12	-.30	.00
European (w/o low admix)	35476.5	28	.17	.19	.53	-.24	.16	.02	.33

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

European (w low admix)	35476.5	28	.17	.19	.53	-.24	.16	.02	.33

Note: n = participant sample size; *K* = number of independent samples. With/without BGA <5%.

There are two clear outliers for European BGA at -.24 and -.20. The first one was based on an extremely small sample (*n* = 15). The second one was both an outlier for European BGA (Zou et al. sample 6, *N* = 158) and for Amerindian BGA.[2] When these outliers are excluded, the mean correlations changed little (*r*s = .16, -.14, and -.11 for European, Amerindian and African BGA respectively). Figures 4, 5 and 6 show the forest plots for the random effects models. Tables 7a to 7c show study details and effect sizes.

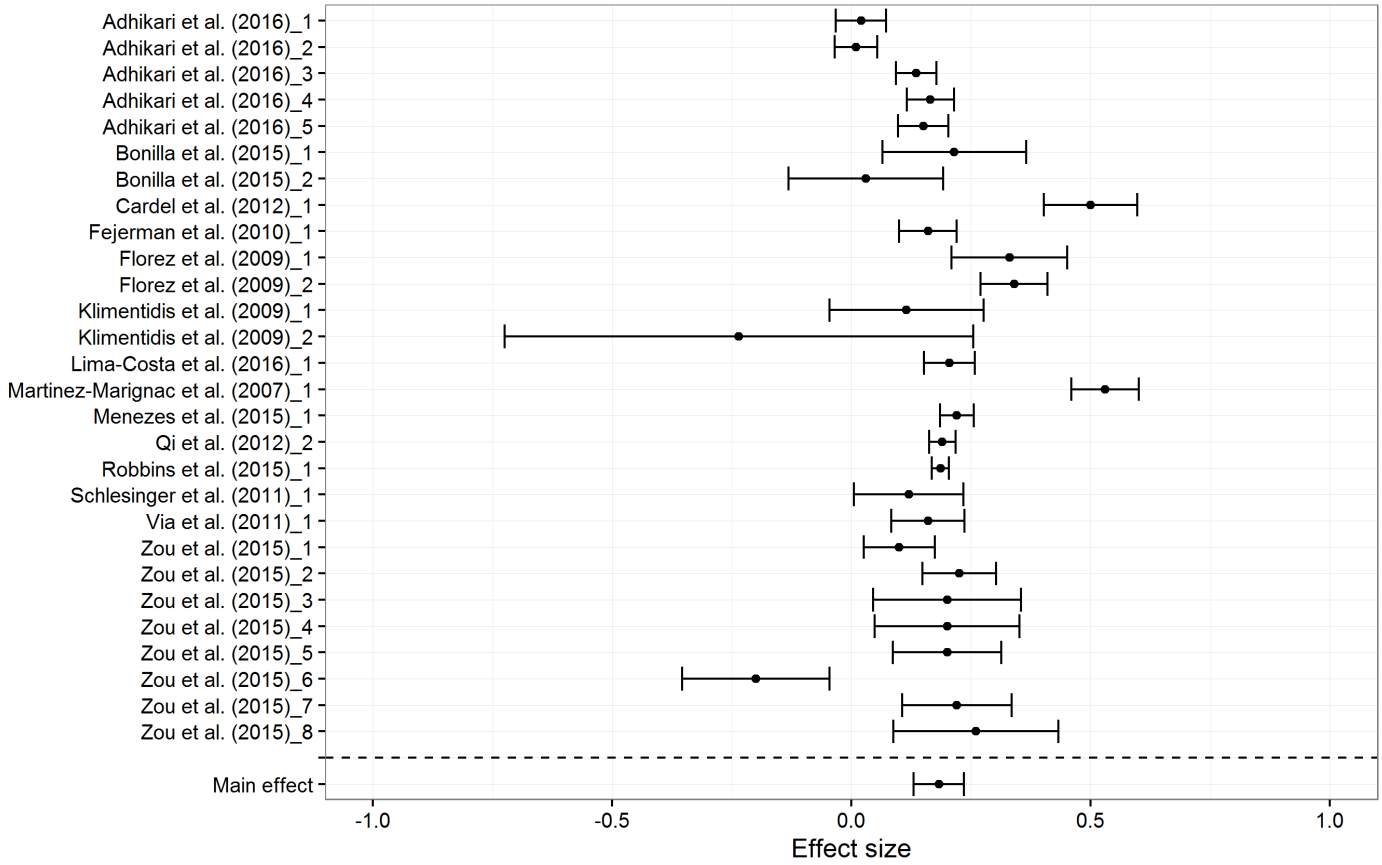

Figure 4. Forest plot for European BGA results, based on random effects model

[2] The lead author (James Zou) was contacted to verify the data. He checked the code and data and reported that there were no errors.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

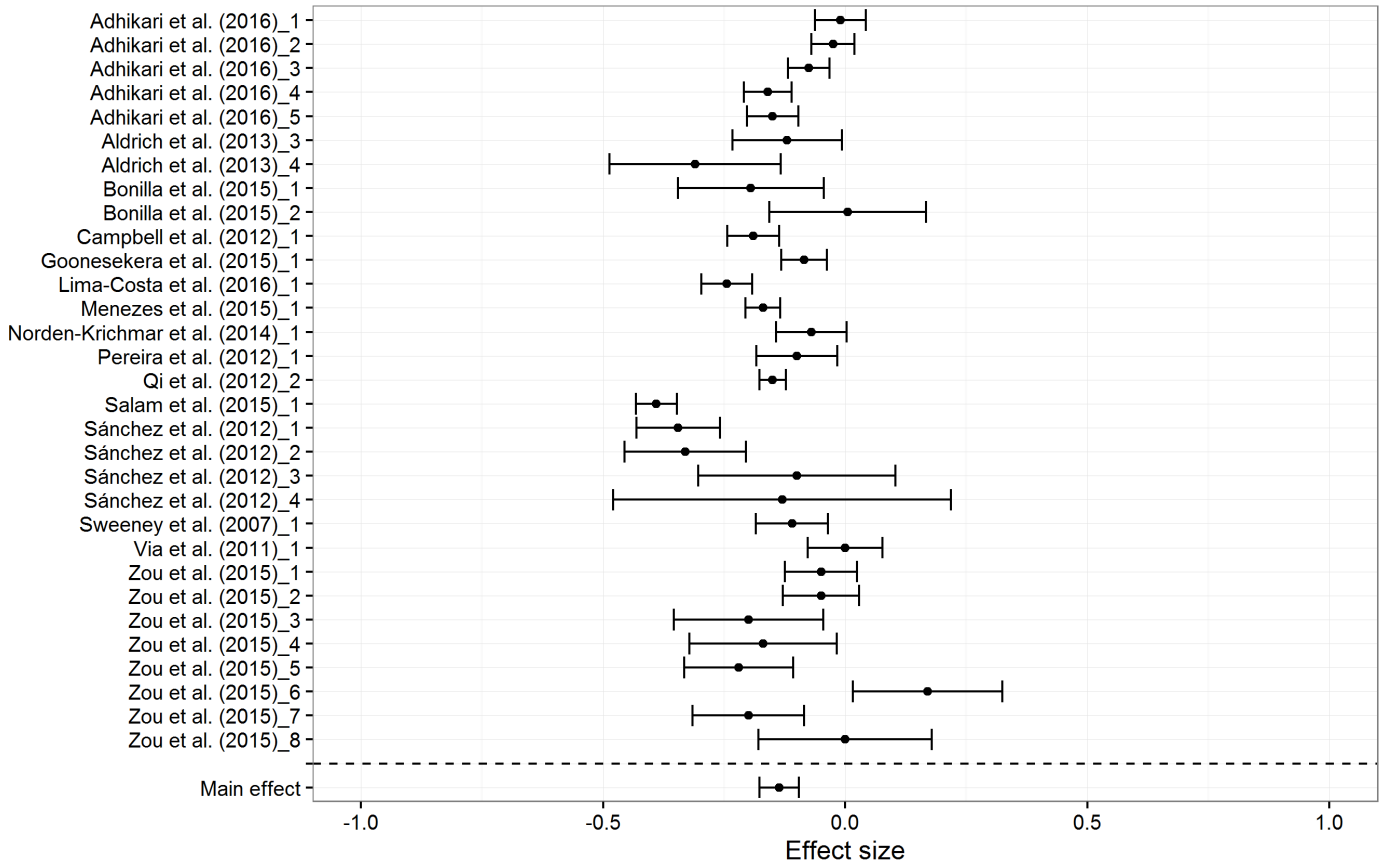

Figure 5. Forest plot for Amerindian BGA results, based on random effects model

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

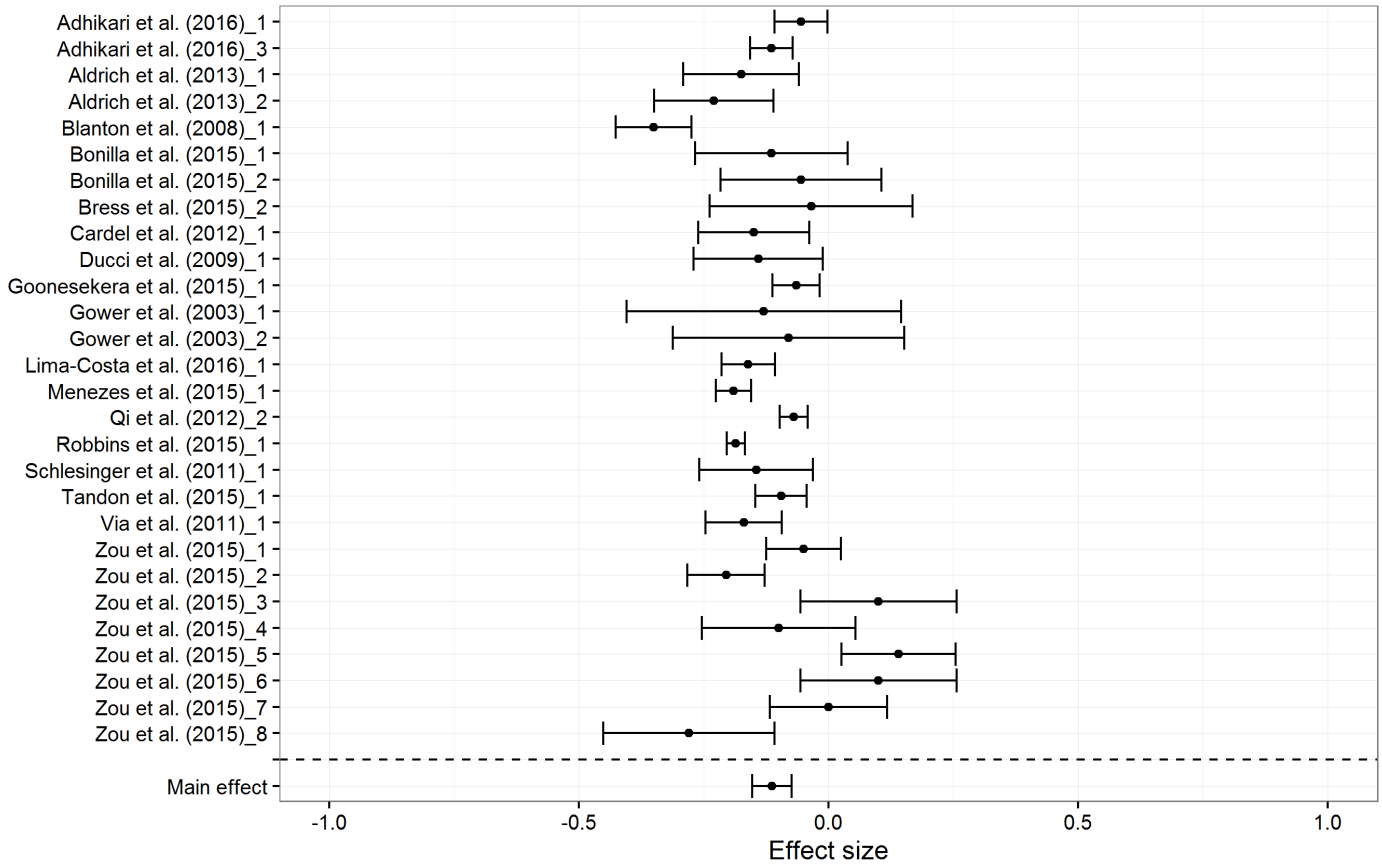

Figure 6. Forest plot for African BGA results, based on random effects model

Table 7a. Study details and effect sizes (with lower and upper 95% confidence intervals) for the meta-analyses: European BGA.

Author sample	Country	Ethnic group	N	*r*	Lower	Upper
Adhikari et al. (2016)_1	Brazil	Multi-ethnic	1375	0.02	-0.03	0.07
Adhikari et al. (2016)_2	Chile	Multi-ethnic	1934	0.01	-0.03	0.05
Adhikari et al. (2016)_3	Colombia	Multi-ethnic	2092	0.14	0.09	0.18
Adhikari et al. (2016)_4	Mexico	Multi-ethnic	1535	0.16	0.12	0.21
Adhikari et al. (2016)_5	Peru	Multi-ethnic	1344	0.15	0.10	0.20
Bonilla et al. (2015)_1	Uruguay	Not stated	164	0.22	0.07	0.36
Bonilla et al. (2015)_2	Uruguay	Not stated	148	0.03	-0.13	0.19
Cardel et al. (2012)_1	US	Multi-ethnic	305	0.50	0.40	0.60
Fejerman et al. (2010)_1	Mexico	Not stated	1035	0.16	0.10	0.22
Florez et al. (2009)_1	Mexico	Not stated	235	0.33	0.21	0.45

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

Florez et al. (2009)_2	Colombia	Not stated	696	0.34	0.27	0.41
Klimentidis et al. (2009)_1	US	Hispanic	147	0.12	-0.05	0.28
Klimentidis et al. (2009)_2	US	Native American	15	-0.24	-0.72	0.25
Lima-Costa et al. (2016)_1	Brazil	Not stated	1305.5	0.20	0.15	0.26
Martinez-Marignac et al. (2007)_1	Mexico	Not stated	561	0.53	0.46	0.60
Menezes et al. (2015)_1	Brazil	Multi-ethnic	2960	0.22	0.18	0.26
Qi et al. (2012)_2	US	Hispanic	4980	0.19	0.16	0.22
Robbins et al. (2015)_1	US	African-American	11239	0.19	0.17	0.20
Schlesinger et al. (2011)_1	Brazil	Not stated	291	0.12	0.01	0.23
Via et al. (2011)_1	US	Puerto-Rican	642	0.16	0.08	0.24
Zou et al. (2015)_1	US	Puerto-Rican	686	0.10	0.03	0.17
Zou et al. (2015)_2	US	Puerto-Rican	615	0.23	0.15	0.30
Zou et al. (2015)_3	US	Mexican	157	0.20	0.05	0.35
Zou et al. (2015)_4	US	Mexican	163	0.20	0.05	0.35
Zou et al. (2015)_5	US	Mexican	291	0.20	0.09	0.31
Zou et al. (2015)_6	US	Mexican	158	-0.20	-0.35	-0.05
Zou et al. (2015)_7	US	Mexican	281	0.22	0.11	0.33
Zou et al. (2015)_8	US	Puerto-Rican	122	0.26	0.09	0.43
RE Model			**35476.5**	**0.18**	**0.13**	**0.24**

Table 7b. Study details and effect sizes (with lower and upper 95% confidence intervals) for the meta-analyses: Amerindian BGA.

Author sample	Country	Ethnic group	N	*r*	Lower	Upper
Adhikari et al. (2016)_1	Brazil	Multi-ethnic	1375	-0.01	-0.06	0.04
Adhikari et al. (2016)_2	Chile	Multi-ethnic	1934	-0.02	-0.07	0.02
Adhikari et al. (2016)_3	Colombia	Multi-ethnic	2092	-0.08	-0.12	-0.03
Adhikari et al. (2016)_4	Mexico	Multi-ethnic	1535	-0.16	-0.21	-0.11
Adhikari et al. (2016)_5	Peru	Multi-ethnic	1344	-0.15	-0.20	-0.10
Aldrich et al. (2013)_3	US	Hispanic	299	-0.12	-0.23	-0.01
Aldrich et al. (2013)_4	US	Hispanic	113	-0.31	-0.49	-0.13
Bonilla et al. (2015)_1	Uruguay	Not stated	164	-0.20	-0.35	-0.04

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

Bonilla et al. (2015)_2	Uruguay	Not stated	148	0.00	-0.16	0.17
Campbell et al. (2012)_1	Colombia	Not stated	1275	-0.19	-0.24	-0.14
Goonesekera et al. (2015)_1	US	Multi-ethnic	1726	-0.08	-0.13	-0.04
Lima-Costa et al. (2016)_1	Brazil	Not stated	1305.5	-0.24	-0.30	-0.19
Menezes et al. (2015)_1	Brazil	Multi-ethnic	2960	-0.17	-0.21	-0.13
Norden-Krichmar et al. (2014)_1	US	Native American	726	-0.07	-0.14	0.00
Pereira et al. (2012)_1	Peru	Multi-ethnic	541	-0.10	-0.18	-0.02
Qi et al. (2012)_2	US	Hispanic	4980	-0.15	-0.18	-0.12
Salam et al. (2015)_1	US	Hispanic	1818	-0.39	-0.43	-0.35
Sánchez et al. (2012)_1	Mexico	Not stated	456	-0.34	-0.43	-0.26
Sánchez et al. (2012)_2	Argentina	Not stated	220	-0.33	-0.46	-0.20
Sánchez et al. (2012)_3	Peru	Not stated	94	-0.10	-0.30	0.10
Sánchez et al. (2012)_4	Chile	Not stated	33	-0.13	-0.48	0.22
Sweeney et al. (2007)_1	US	Hispanic	684	-0.11	-0.18	-0.04
Via et al. (2011)_1	US	Puerto-Rican	642	0.00	-0.08	0.08
Zou et al. (2015)_1	US	Puerto-Rican	686	-0.05	-0.12	0.02
Zou et al. (2015)_2	US	Puerto-Rican	615	-0.05	-0.13	0.03
Zou et al. (2015)_3	US	Mexican	157	-0.20	-0.35	-0.05
Zou et al. (2015)_4	US	Mexican	163	-0.17	-0.32	-0.02
Zou et al. (2015)_5	US	Mexican	291	-0.22	-0.33	-0.11
Zou et al. (2015)_6	US	Mexican	158	0.17	0.02	0.32
Zou et al. (2015)_7	US	Mexican	281	-0.20	-0.31	-0.09
Zou et al. (2015)_8	US	Puerto-Rican	122	0.00	-0.18	0.18
RE Model			**28937.5**	**-0.14**	**-0.18**	**-0.10**

Table 7c. Study details and effect sizes (with lower and upper 95% confidence intervals) for the meta-analyses: African BGA.

Author sample	Country	Ethnic group	n	r	Lower	Upper
Adhikari et al. (2016)_1	Brazil	Multi-ethnic	1375	-0.06	-0.11	0.00
Adhikari et al. (2016)_3	Colombia	Multi-ethnic	2092	-0.12	-0.16	-0.07
Aldrich et al. (2013)_1	US	African-American	280	-0.18	-0.29	-0.06
Aldrich et al. (2013)_2	US	African-American	255	-0.23	-0.35	-0.11

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

Blanton et al. (2008)_1	Brazil	Multi-ethnic	587	-0.35	-0.43	-0.27
Bonilla et al. (2015)_1	Uruguay	Not stated	164	-0.12	-0.27	0.04
Bonilla et al. (2015)_2	Uruguay	Not stated	148	-0.05	-0.22	0.11
Bress et al. (2015)_2	US	African-American	95	-0.04	-0.24	0.17
Cardel et al. (2012)_1	US	Multi-ethnic	305	-0.15	-0.26	-0.04
Ducci et al. (2009)_1	US	African-American	228	-0.14	-0.27	-0.01
Goonesekera et al. (2015)_1	US	Multi-ethnic	1726	-0.06	-0.11	-0.02
Gower et al. (2003)_1	US	African-American	52	-0.13	-0.40	0.14
Gower et al. (2003)_2	US	White-American	73	-0.08	-0.31	0.15
Lima-Costa et al. (2016)_1	Brazil	Not stated	1305.5	-0.16	-0.21	-0.11
Menezes et al. (2015)_1	Brazil	Multi-ethnic	2960	-0.19	-0.23	-0.15
Qi et al. (2012)_2	US	Hispanic	4980	-0.07	-0.10	-0.04
Robbins et al. (2015)_1	US	African-American	11239	-0.19	-0.20	-0.17
Schlesinger et al. (2011)_1	Brazil	Not stated	291	-0.15	-0.26	-0.03
Tandon et al. (2015)_1	US	African-American	1440	-0.10	-0.15	-0.04
Via et al. (2011)_1	US	Puerto-Rican	642	-0.17	-0.25	-0.09
Zou et al. (2015)_1	US	Puerto-Rican	686	-0.05	-0.12	0.02
Zou et al. (2015)_2	US	Puerto-Rican	615	-0.20	-0.28	-0.13
Zou et al. (2015)_3	US	Mexican	157	0.10	-0.06	0.26
Zou et al. (2015)_4	US	Mexican	163	-0.10	-0.25	0.05
Zou et al. (2015)_5	US	Mexican	291	0.14	0.03	0.25
Zou et al. (2015)_6	US	Mexican	158	0.10	-0.06	0.26
Zou et al. (2015)_7	US	Mexican	281	0.00	-0.12	0.12
Zou et al. (2015)_8	US	Puerto-Rican	122	-0.28	-0.45	-0.11
RE Model			**32710.5**	**-0.11**	**-0.15**	**-0.07**

The random effect meta-analytic mean effect sizes are .18, -.14, and -.11 for European, Amerindian and African BGA, respectively. European BGA shows a stronger association with SES than do Amerindian and African BGA, presumably because it is the largest ancestry component (mean European BGA = 54.41%). The difference could be due to the higher variation in the proportion of European BGA in the subjects. The mean standard deviation of admixture (18.01, 11.18, and 14.50 for European, Amerindian and African, respectively) was lower for the non-European ancestries, and less variation is expected to lead to smaller effect sizes owing to restriction of range (Schmidt & Hunter, 2014).

The exclusion of samples with a mean BGA below 5% had only a slight impact on the results. When all samples are included, the meta-analytic effect sizes are *r* = .18 [.13, .24], *r* = -.13 [-.17, -.09], and r = -.09 [-.13, -.05] for European,

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

Amerindian and African BGA, respectively. Results for African BGA are impacted the most owing to the presence of a positive association between African BGA and outcomes for several samples with very low African admixture (e.g., Adhikari et al., 2016: Chile, African admixture 2.61%, *r* BGA × SES = .08). It is notable that for two of the same countries, namely Chile and Mexico, Fuerst and Kirkegaard (2016a, 2016b) found that regional African BGA is associated with higher SES.

The funnel plots did not show notable asymmetry, suggesting no publication bias. Figures 7, 8, and 9 show, respectively, the funnel plots for European, Amerindian, and African BGA.

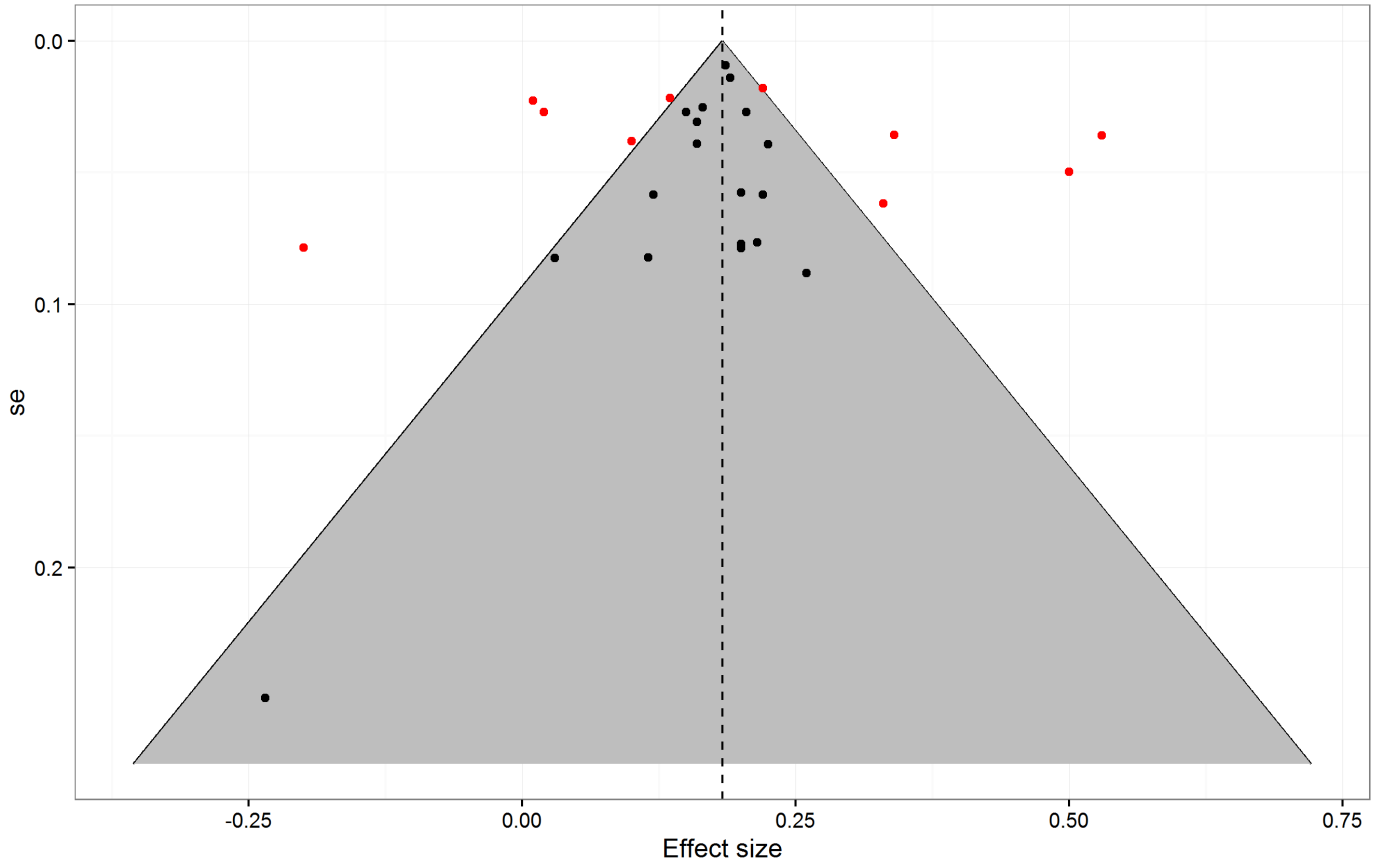

Figure 7. Funnel plot for European BGA, with standard error on the y-axis and effect size on the x-axis

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

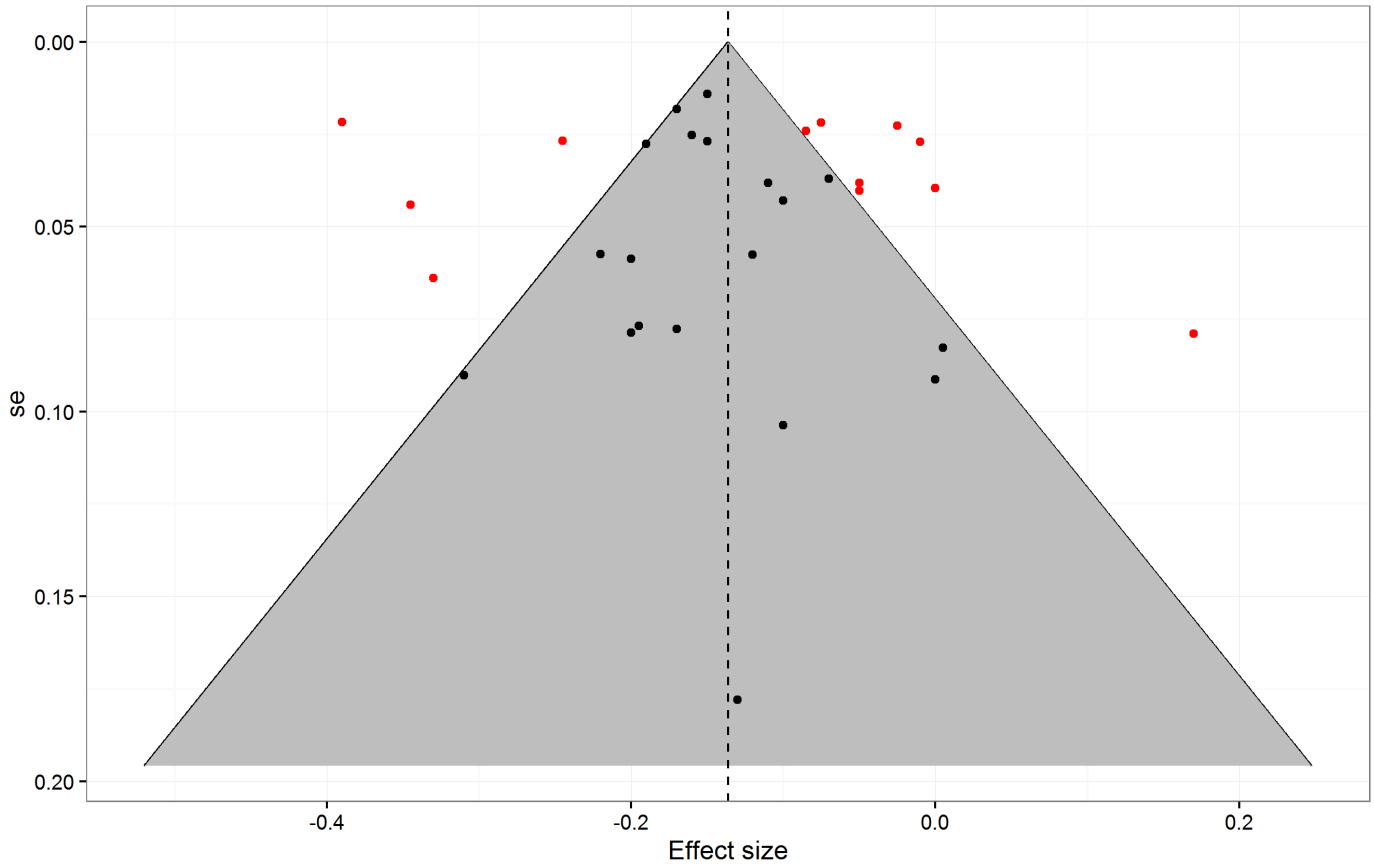

Figure 8. Funnel plot for Amerindian BGA, with standard error on the y-axis and effect size on the x-axis

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

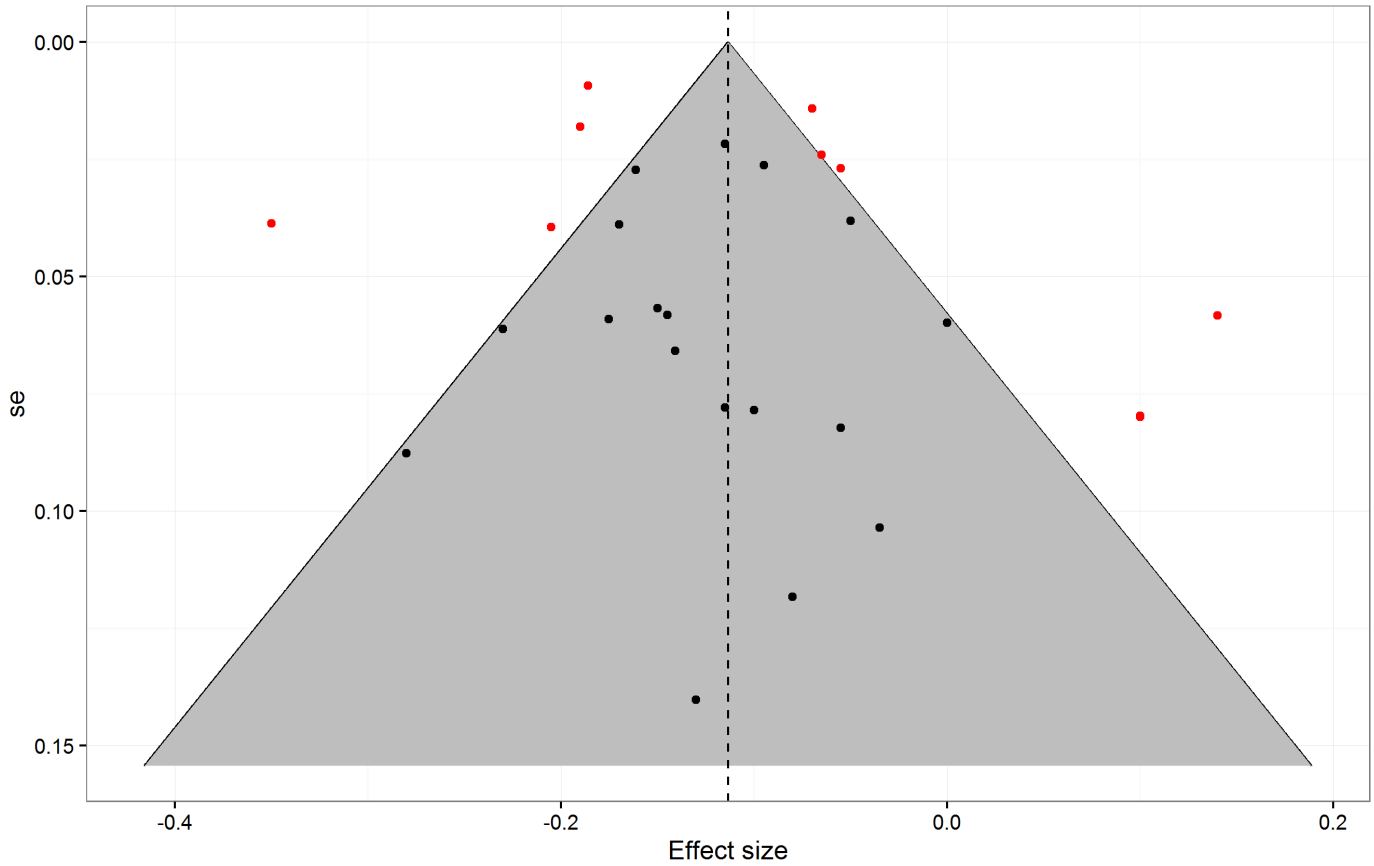

Figure 9. Funnel plot for African BGA, with standard error on the y-axis and effect size on the x-axis

3. Discussion

3.1. Main results

The results of the meta-analysis are consistent with those reported in earlier narrative reviews in that European BGA was statistically associated with more favorable socioeconomic outcomes relative to Amerindian and African BGA. When low-admixture samples were excluded, the random effects meta-analysis gave a mean SES × BGA correlation of $r = .18$, $r = -.14$, and $r = -.11$ for European, Amerindian and African BGA, respectively; furthermore, the analysis of SES × BGA association directions gave square root *n*-weighted means of $M = 0.83$, $M = -0.81$, and $M = -0.82$ for European, Amerindian and African BGA, respectively. For the random effects meta-analysis, the heterogeneity, or the percentage of variance due to variance between samples as opposed to sampling error, was considerable ($I^2_{mean} = 92\%$, $I^2_{European} = 95\%$, $I^2_{Amerindian} = 91\%$, $I^2_{African} = 89\%$) by conventional standards (Higgins & Green, 2008). Due to the limited number of samples, it is difficult to evaluate the cause of this pattern of results. Some possibilities are as follows:

1. Discretization: Many outcome variables were ordinal instead of continuous, even when continuous values were available (e.g., income). Pearson correlations assume that the data are normally and continuously distributed, so the use of non-continuous data induces a downward bias in the results. Corrections for discretization were not attempted, although one could attempt to apply these (Schmidt & Hunter, 2014).

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

2. Number of BGA-informative markers: There were large differences in the number of genomic markers used to estimate individual BGA. Figure 10 shows a density histogram of the distribution for the effect size analysis with low-admixture samples removed. Using fewer markers results in more measurement error with respect to true BGA (Russo et al., 2016). Ruiz-Linares et al. (2014) reported correlations between admixture estimates using different numbers of markers. They noted that a recent study (Scharf et al., 2013) had found that using 15 markers resulted in correlations of about .60 with estimates derived from 50k markers. Furthermore, using 30 and 152 markers resulted in r = .70 and r = .85, respectively, with respect to estimated admixture based on 50k markers. It is clear that there are diminishing returns to using more markers, but also that using more reduces measurement error with respect to true BGA. It would be difficult to correct for measurement error with respect to BGA. One option would be to acquire a sufficient number of data points from one study to allow for the estimation of a predictive model. One could then use that model to estimate the measurement error in other studies based on the number of reported markers used. However, we failed to find a study which had sufficient data points, so a correction was not attempted.

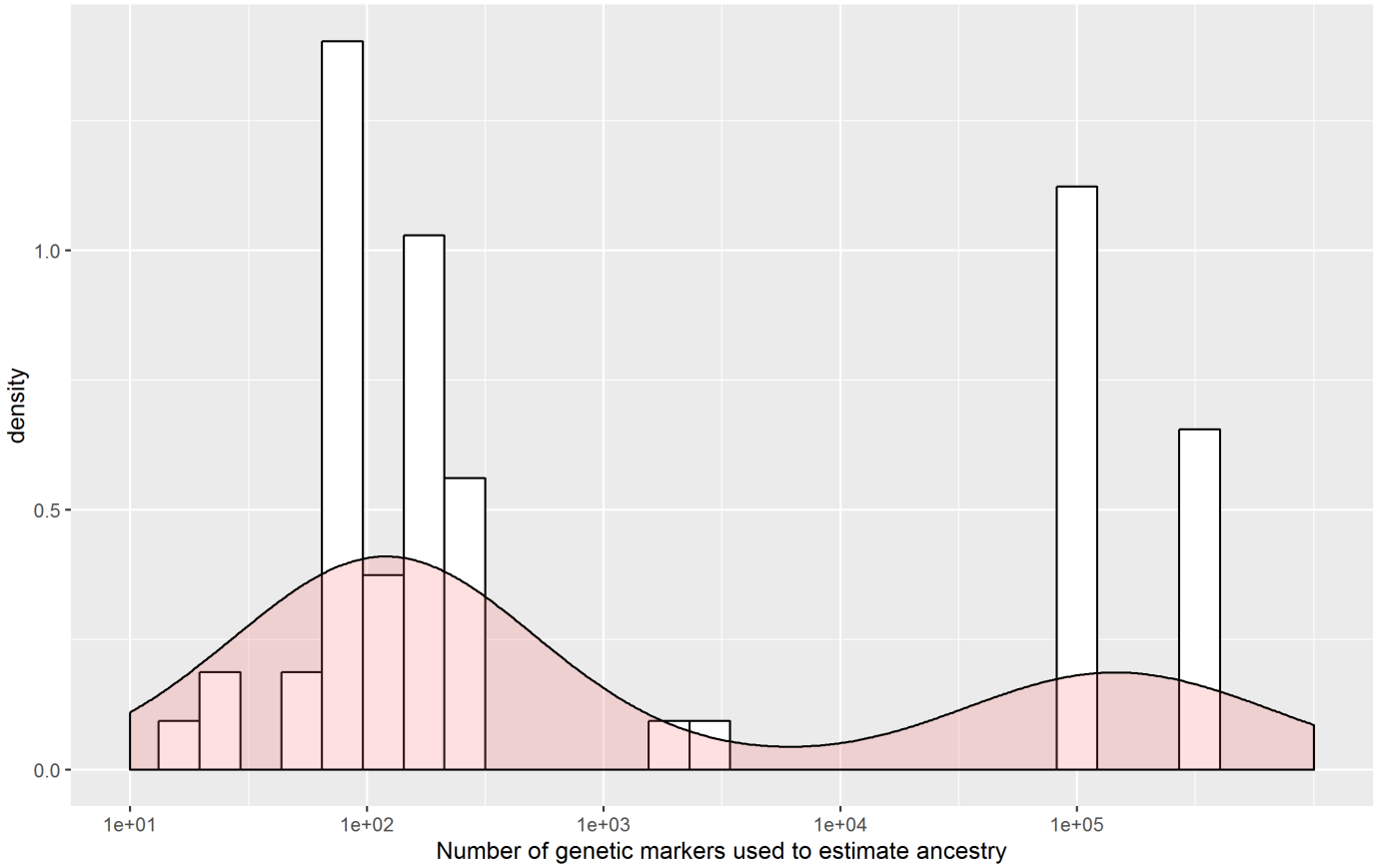

Figure 10. Density histogram of the number of genetic markers in each sample. Average frequency density on the y-axis and number of genetic markers on the x-axis.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

3. Variation in admixture: Based on the samples for which standard deviations of BGA were reported, variability in BGA differed considerably between samples. For instance, for Uruguayans, Bonilla et al. (2015) found a standard deviation of African BGA of 7.52; in contrast, for Brazilians, Schlesinger et al. (2011) found one of 24.08. As mentioned earlier, the magnitude of associations between outcomes and BGA will be impacted by variance in BGA. Thus, between-sample heterogeneity in BGA variance is expected to contribute to heterogeneity in BGA × SES associations.

4. Heterogeneous origin: The studies in the meta-analysis came from many countries. It is probable that the statistical effect of one's BGA depends on local cultural norms or practices that differ between countries or even between regions within countries.

5. Sampling: Most of the studies used convenience samples, an approach that can have an impact on associations. Adhikari et al.'s (2016) samples, for example, were recruited mostly at universities, meaning that individuals were selected with respect to educational status. Thus, the correlations they obtained are expected to be attenuated relative to those derived from more representative samples.

3.2. Implications for epidemiological studies

Fairly robust associations between BGA and SES were found. Given this, to avoid spurious associations in regression analyses due to omitted variable bias, it is potentially important to include reliable measures of SES in studies of medical outcomes. Studies reviewed in this analysis employed a wide variety of specific measures (see Table 3), some of which may be better indices of medically relevant SES than others. It would be worthwhile to investigate which measures (neighborhood SES, education, income, etc.) are generally associated to a greater degree with BGA and with medical outcomes, so that future studies could attempt to incorporate these measures into analyses.

Further, it is also important to identify the factors mediating the BGA-SES associations, as these could incrementally explain the BGA-medical outcome associations. Figure 11 shows one possible model. In this model, local ancestry is a subset of global ancestry. Both local and global ancestry correlate with disease loci which, in turn, are causally related to health outcomes. Global ancestry is also related both to an individual's SIRE and to cultural practices associated with SIRE, and to an individual's race/ethnicity-associated phenotype, and social reactions, perhaps in the form of "colorism" (e.g., Hunter, 2007; Telles, 2014), associated with that phenotype. In this model, both direct and indirect pathways run from cultural practices and social reactions to health outcomes. An example of a direct pathway would be through observed race/ethnicity-related differential treatments by providers (as discussed by, e.g., Van Ryn, 2002). Indirectly, these factors potentially impact both individuals' human capital and SES, which are reciprocally related to one another and to health outcomes.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

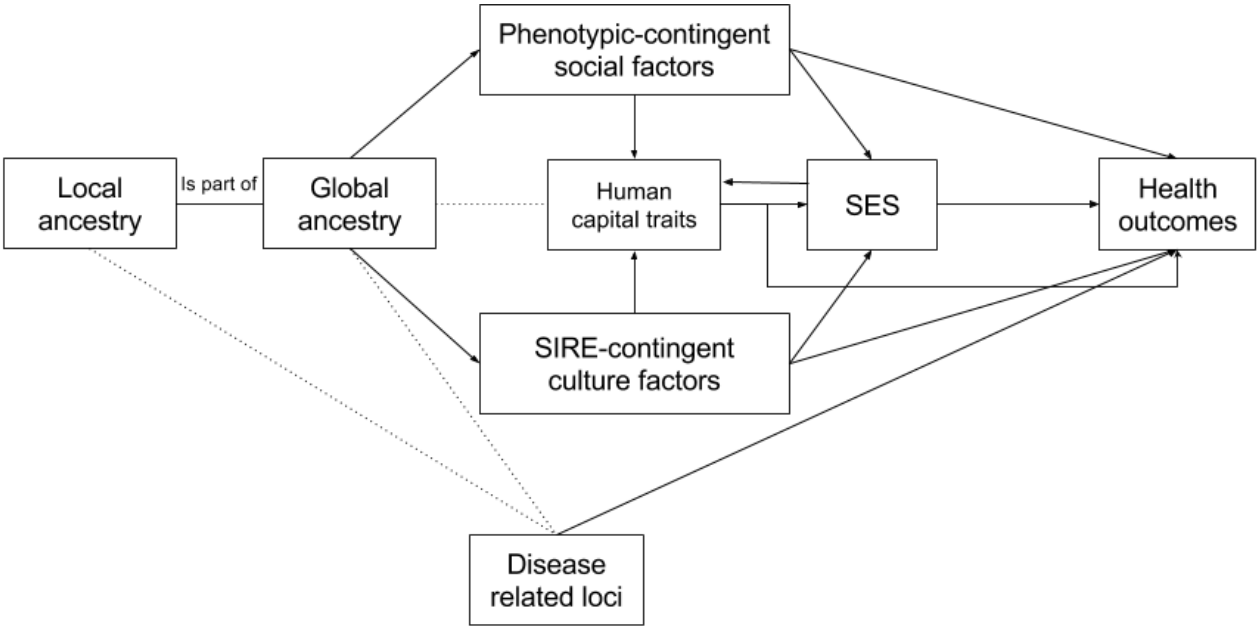

Figure 11. Model of possible relations between BGA and SES. The stippled lines represent correlations and the arrowed lines represent causal associations.

This model is similar to that proposed by Marden et al. (2016). A notable exception is that we include human capital as a mediating variable. Cognitive epidemiology studies have shown that measures of cognitive ability (one component of human capital) can explain a portion of medical-related outcomes, independent of SES (Deary, 2009; Der, Batty & Deary, 2009; Wraw et al., 2015). Because, throughout the Americas, SIRE groups are known to differ in mean levels of phenotypic cognitive ability (Fuerst & Kirkegaard, 2016b; Roth et al., 2001), it would be reasonable to include measures of cognitive ability in admixture analyses to see if such measures explain some of the health outcome differences independent of SES. However, we could not find any study that did this. On the other hand, we did locate one recent admixture paper (Akshoomoff et al., 2014) which reported that BGA was associated with cognitive ability. On inquiry, as expected given the SIRE differences, European BGA was positively associated with cognitive ability, while African and Amerindian BGA were negatively associated (Akshoomoff, personal communications, November 9, 2014). If cognitive ability mediates the relationship between BGA and SES, not including measures of it constitutes omitted variable bias.

3.3. Untangling the statistical effects of BGA and SIRE

In many of the samples included in this meta-analysis, individual SES outcomes are associated with BGA within SIRE groups. Hence, SIRE membership does not mediate the relationship. In other cases, particularly in Latin America, the issue is less clear and BGA may be confounded with SIRE. For instance, Leite et al. (2011) found that European BGA was positively correlated with SES in Brasília. This could be because SES was positively correlated with BGA independently of SIRE, or because it was positively correlated with SIRE but not with BGA independent of SIRE. The analysis by Ruiz-Linares et al. (2014) has helped to clarify the issue. The authors looked at the association between BGA, skin pigmentation and SIRE in a multi-country sample from Brazil, Chile, Colombia, Mexico and Peru (mean age 20 to 25, depending on country). The authors found modest correlations between BGA and SIRE (e.g., 0.48 in the case

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

of both European and Amerindian BGA/SIRE). They reported that wealth and educational attainment correlated with European BGA (*r* = .12 for the full sample). However, when the statistical effect of BGA was controlled for, education was not significantly associated with SIRE, and wealth was only related to SIRE within the European/White color group. Similar results in this regard were found with Menezes et al.'s (2015) sample from the 1982 Pelotas Birth Cohort study (F. Hartwig, personal communication, March 4, 2016). Details pertaining to this sample can be found in Menezes et al. (2015). The results pertaining to self-identified color and interviewer-rated color are shown in Supplementary File 3. For this sample, the positive association between European BGA and household assets, schooling, and income was robust to controls for interviewer-observed race/ethnicity and SIRE. Contrary to a simple colorism model, Black (*preta)* and Brown (*parda*) interviewer-rated race/ethnicity were generally positively related to outcomes, controlling for European BGA. Taken together, the results suggest that the association between BGA and SES is substantially independent of SIRE and thus SIRE-related cultural factors. Quite possibly, phenotype-based social reactions, which would differentially impact individuals within SIRE groups, affect outcomes. For example, consistent with this possibility, Telles (2014) found, in a large multinational sample, that interviewer-rated skin brightness was related to SES controlling for SIRE.

3.4. Limitations and suggestions for future research

The number of effect sizes used in the present meta-analysis is limited. This is because many studies did not report effect sizes and, in some cases, the authors either did not reply to emails or were unable to provide results or data. Meta-analyses would be more reliable and easier to perform if scientists were more willing to publish their data and report effect sizes. The results significantly varied across studies (mean effect size heterogeneity = 92%), which means that there are effect size moderators. A moderator analysis was not conducted due to the relatively small number of studies in the dataset. We suggest that moderator analyses be conducted as relevant data accumulates. Detailed analyses for publication bias were also not conducted because the number of effect sizes was too small for a reliable analysis. This concern should be investigated in future meta-analyses.

Supplementary material

Supplementary File 1-3, high-quality figures and R analysis code are available at the repository at *Open Science Framework* https://osf.io/ydc3f/files/.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

* indicates that the study included data used in the meta-analysis.

*Adhikari, K., Fontanil, T., Cal, S., Mendoza-Revilla, J., Fuentes-Guajardo, M., Chacón-Duque, J.C., ... & Jaramillo, C. (2016). A genome-wide association scan in admixed Latin Americans identifies loci influencing facial and scalp hair features. *Nature Communications*, 7: 10815.

Akshoomoff, N., Newman, E., Thompson, W.K., McCabe, C., Bloss, C.S., Chang, L., Amaral, D.G. et al. (2014). The NIH Toolbox Cognition Battery: Results from a large normative developmental sample (PING). *Neuropsychology* 28(1): 1.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

*Aldrich, M.C., Selvin, S., Wrensch, M.R., Sison, J.D., Hansen, H.M., Quesenberry Jr, C.P. ... & Wiencke, J.K. (2013). Socioeconomic status and lung cancer: Unraveling the contribution of genetic admixture. *American Journal of Public Health* 103(10): e73-e80.

*Allison, M.A., Peralta, C.A., Wassel, C.L., Aboyans, V., Arnett, D.K., Cushman, M. ... & Criqui, M.H. (2010). Genetic ancestry and lower extremity peripheral artery disease in the Multi-Ethnic Study of Atherosclerosis. *Vascular Medicine* 15(5): 351-359.

*Blanton, R.E., Silva, L.K., Morato, V.G., Parrado, A.R., Dias, J P., Melo, P.R. ... & Teixeira, M.G. (2008). Genetic ancestry and income are associated with dengue hemorrhagic fever in a highly admixed population. *European Journal of Human Genetics* 16(6): 762-765.

*Bonilla, C., Bertoni, B., Hidalgo, P.C., Artagaveytia, N., Ackermann, E., Barreto, I., Cancela, P. et al. (2015). Breast cancer risk and genetic ancestry: A case–control study in Uruguay. *BMC Women's Health* 15(1): 1.

*Bress, A., Kittles, R., Wing, C., Hooker Jr, S.E. & King, A. (2015). Genetic ancestry as an effect modifier of naltrexone in smoking cessation among African Americans: An analysis of a randomized controlled trial. *Pharmacogenetics and Genomics* 25(6): 305-312.

*Campbell, D.D., Parra, M.V., Duque, C., Gallego, N., Franco, L., Tandon, A. ... & McCarthy, M.I. (2012). Amerind ancestry, socioeconomic status and the genetics of type 2 diabetes in a Colombian population. *PLoS One* 7(4): e33570.

*Cardel, M., Willig, A.L., Dulin-Keita, A., Casazza, K., Beasley, T.M. & Fernández, J.R. (2012). Parental feeding practices and socioeconomic status are associated with child adiposity in a multi-ethnic sample of children. *Appetite* 58(1): 347-353.

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

*Choudhry, S., Burchard, E.G., Borrell, L.N., Tang, H., Gomez, I., Naqvi, M. ... & Ziv, E. (2006). Ancestry–environment interactions and asthma risk among Puerto Ricans. *American Journal of Respiratory and Critical Care Medicine* 174(10): 1088-1093.

*Cifuentes, L., Morales, R., Sepulveda, D., Jorquera, H. & Acuna, M. (2004). DYS19 and DYS199 loci in a Chilean population of mixed ancestry. *American Journal of Physical Anthropology* 125(1): 85-89.

Deary, I.J. (2009). Introduction to the special issue on cognitive epidemiology. *Intelligence* 37: 517-519.

Der, G., Batty, G.D. & Deary, I.J. (2009). The association between IQ in adolescence and a range of health outcomes at 40 in the 1979 US National Longitudinal Study of Youth. *Intelligence* 37: 573-580.

*Ducci, F., Roy, A., Shen, P.H., Yuan, Q., Yuan, N.P., Hodgkinson, C.A. ... & Goldman, D. (2009). Association of substance use disorders with childhood trauma but not African genetic heritage in an African American cohort. *American Journal of Psychiatry* 166: 1031-1040.

*Fejerman, L., Hu, D., Huntsman, S., John, E.M., Stern, M.C., Haiman, C.A. ... & Ziv, E. (2013). Genetic ancestry and risk of mortality among US Latinas with breast cancer. *Cancer Research* 73: 7243-7253.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

*Fejerman, L., Romieu, I., John, E.M., Lazcano-Ponce, E., Huntsman, S., Beckman, K.B. ... & Torres-Mejía, G. (2010). European ancestry is positively associated with breast cancer risk in Mexican women. *Cancer Epidemiology Biomarkers & Prevention* 19: 1074-1082.

*Florez, J.C., Price, A.L., Campbell, D., Riba, L., Parra, M.V., Yu, F. ... & Franco, L. (2009). Strong association of socioeconomic status with genetic ancestry in Latinos: Implications for admixture studies of type 2 diabetes. *Diabetologia* 52: 1528-1536.

Frudakis, T. (2010). *Molecular Photofitting: Predicting Ancestry and Phenotype Using DNA*. Elsevier.

Frudakis, T. & Shriver, M. (2003). Compositions and methods for inferring ancestry. U.S. Patent Application 10/644,594, filed August 19, 2003.

Fuerst, J. & Kirkegaard, E.O.W. (2016a). Admixture in the Americas: Regional and national differences. *Mankind Quarterly* 56: 255-373.

Fuerst, J. & Kirkegaard, E.O.W. (2016b). The genealogy of differences in the Americas. *Mankind Quarterly* 56: 425-481.

*González Burchard, E., Borrell, L.N., Choudhry, S., Naqvi, M., Tsai, H.-J., Rodriguez-Santana, J.R., Chapela, R. et al. (2005). Latino populations: A unique opportunity for the study of race, genetics, and social environment in epidemiological research. *American Journal of Public Health* 95: 2161-2168.

*Goonesekera, S.D., Fang, S.C., Piccolo, R.S., Florez, J.C. & McKinlay, J.B. (2015). Biogeographic ancestry is associated with higher total body adiposity among African-American females: The Boston area community health survey. *PloS One* 10(4): e0122808.

*Gower, B.A., Fernández, J.R., Beasley, T.M., Shriver, M.D. & Goran, M.I. (2003). Using genetic admixture to explain racial differences in insulin-related phenotypes. *Diabetes* 52: 1047-1051.

*Gravlee, C.C., Non, A.L. & Mulligan, C.J. (2009). Genetic ancestry, social classification, and racial inequalities in blood pressure in Southeastern Puerto Rico. *PLoS One* 4(9): e6821.

Higgins, J.P.T & Green, S., eds. (2008). *Cochrane Handbook for Systematic Reviews of Interventions*, Vol. 5. Chichester: Wiley-Blackwell.

Hill, W.D., Hagenaars, S.P., Marioni, R.E., Harris, S.E., Liewald, D.C.M., Davies, G., McIntosh, A.M., Gale, C.R., Deary, I.J. & International Consortium for Blood Pressure (2016). Molecular genetic contributions to social deprivation and household income in UK Biobank (n= 112,151). *bioRxiv*: 043000.

Hunter, M. (2007). The persistent problem of colorism: Skin tone, status, and inequality. *Sociology Compass* 1(1): 237-254.

Kan, K.-J., Wicherts, J.M., Dolan, C.V. & van der Maas, H.L.J. (2013). On the nature and nurture of intelligence and specific cognitive abilities. The more heritable, the more culture dependent. *Psychological Science* 24: 2420-2428.

*Klimentidis, Y.C., Miller, G.F. & Shriver, M.D. (2009). The relationship between European genetic admixture and body composition among Hispanics and Native Americans. *American Journal of Human Biology* 21: 377-382.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

*Leite, T.K.M., Fonseca, R.M.C., De França, N.M., Parra, E.J. & Pereira, R.W. (2011). Genomic ancestry, self-reported "color" and quantitative measures of skin pigmentation in Brazilian admixed siblings. *PLoS One* 6(11): e27162.

*Lima-Costa, M.F., de Mello Mambrini, J.V., Leite, M.L.C., Peixoto, S.V., Firmo, J.O.A., de Loyola Filho, A.I. ... & Tarazona-Santos, E. (2016). Socioeconomic position, but not African genomic ancestry, is associated with blood pressure in the Bambui-Epigen (Brazil) Cohort Study of Aging. *Hypertension* 67: 349-355.

Marden, J.R., Walter, S, Kaufman, J.S. & Glymour, M.M. (2016). African ancestry, social factors, and hypertension among non-Hispanic Blacks in the Health and Retirement Study. *Biodemography and Social Biology* 62: 19-35.

*Martinez-Marignac, V.L., Valladares, A., Cameron, E., Chan, A., Perera, A., Globus-Goldberg, R. ... & Shriver, M.D. (2007). Admixture in Mexico City: Implications for admixture mapping of type 2 diabetes genetic risk factors. *Human Genetics* 120: 807-819.

*Menezes, A.M.B., Wehrmeister F.C., Hartwig, F.P., Perez-Padilla, R., Gigante, D.P., Barros, F.C., Oliveira, I.O., Ferreira, G.D. & Horta, B.L. (2015). African ancestry, lung function and the effect of genetics. *European Respiratory Journal*: ERJ-01121.

*Molokhia, M., Hoggart, C., Patrick, A.L., Shriver, M., Parra, E., Ye, J. & McKeigue, P.M. (2003). Relation of risk of systemic lupus erythematosus to West African admixture in a Caribbean population. *Human Genetics* 112: 310-318.

*Norden-Krichmar, T.M., Gizer, I.R., Libiger, O., Wilhelmsen, K.C., Ehlers, C.L. & Schork, N.J. (2014). Correlation analysis of genetic admixture and social identification with body mass index in a Native American community. *American Journal of Human Biology* 26: 347-360.

*Pereira, L., Zamudio, R., Soares-Souza, G., Herrera, P., Cabrera, L., Hooper, C.C. ... & Schneider, S. (2012). Socioeconomic and nutritional factors account for the association of gastric cancer with Amerindian ancestry in a Latin American admixed population. *PLoS One* 7(8): e41200.

*Qi, L., Nassir, R., Kosoy, R., Garcia, L., Curb, J.D., Tinker, L. ... & Seldin, M.F. (2012). Relationship between diabetes risk and admixture in postmenopausal African-American and Hispanic-American women. *Diabetologia* 55: 1329-1337.

*Reiner, A.P., Ziv, E., Lind, D.L., Nievergelt, C.M., Schork, N.J., Cummings, S.R. ... & Kwok, P.Y. (2005). Population structure, admixture, and aging-related phenotypes in African American adults: The Cardiovascular Health Study. *American Journal of Human Genetics* 76: 463-477.

*Risch, N., Choudhry, S., Via, M., Basu, A., Sebro, R., Eng, C. ... & Burchard, E.G. (2009). Ancestry-related assortative mating in Latino populations. *Genome Biology* 10(11): R132.

*Robbins, J.A., Qi, L., Garcia, L., Younger, J.W. & Seldin, M.F. (2015). Relationship of pain and ancestry in African American women. *European Journal of Pain* 19: 601-610.

Roth, P.L., Bevier, C.A., Bobko, P., Switzer, F.S. & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology* 54: 297-330.

Ruiz-Linares, A., Adhikari, K., Acuña-Alonzo, V., Quinto-Sanchez, M., Jaramillo, C., Arias, W., Fuentes, M. et al. (2014). Admixture in Latin America: Geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genetics* 10(9): e1004572.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

*Ruiz-Narváez, E.A., Bare, L., Arellano, A., Catanese, J. & Campos, H. (2010). West African and Amerindian ancestry and risk of myocardial infarction and metabolic syndrome in the Central Valley population of Costa Rica. *Human Genetics* 127: 629-638.

Russo, M.G., Di Fabio Rocca, F., Doldán, P., Gonzalo Cardozo, D., Dejean, C., Seldes, V. & Avena, S. (2016). Evaluation of the minimum number of markers for individual ancestry estimation in an Argentinean population sample. *Revista del Museo de Antropología* 9(1): 49-56.

*Salam, M.T., Avoundjian, T., Knight, W.M. & Gilliland, F.D. (2015). Genetic ancestry and asthma and rhinitis occurrence in Hispanic children: Findings from the Southern California Children's Health Study. *PloS One* 10(8): e0135384.

Salzano, F.M. & Sans, M. (2014). Interethnic admixture and the evolution of Latin American populations. *Genetics and Molecular Biology* 37: 151-170.

*Sánchez, E., Rasmussen, A., Riba, L., Acevedo-Vasquez, E., Kelly, J.A., Langefeld, C.D. ... & Maradiaga-Ceceña, M.A. (2012). Impact of genetic ancestry and sociodemographic status on the clinical expression of systemic lupus erythematosus in American Indian–European populations. *Arthritis & Rheumatism* 64: 3687-3694.

Scharf, J.M., Yu, D., Mathews, C.A., Neale, B.M., Stewart, S.E., Fagerness, J.A., Evans, P. et al. (2013). Genome-wide association study of Tourette's syndrome. *Molecular Psychiatry* 18: 721-728.

*Schlesinger, D., Grinberg, L.T., Alba, J.G., Naslavsky, M.S., Licinio, L., Farfel, J.M., ... & Dos Santos, A.C.F. (2011). African ancestry protects against Alzheimer's disease-related neuropathology. *Epub 2011 Nov 8*. [(*Molecular Psychiatry* 18: 79-85 (2013)].

Schmidt, F.L. & Hunter, J.E. (2014). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Sage.

Shriner, D. (2013). Overview of admixture mapping. *Current Protocols in Human Genetics*: 1-23.

Shriver, M.D. & Kittles, R.A. (2004). Genetic ancestry and the search for personalized genetic histories. *Nature Reviews Genetics* 5: 611-618.

*Sweeney, C., Wolff, R.K., Byers, T., Baumgartner, K.B., Giuliano, A.R., Herrick, J.S. ... & Slattery, M.L. (2007). Genetic admixture among Hispanics and candidate gene polymorphisms: Potential for confounding in a breast cancer study? *Cancer Epidemiology Biomarkers & Prevention* 16: 142-150.

*Tandon, A., Chen, C.J., Penman, A., Hancock, H., James, M., Husain, D. ... & Riche, D. (2015). African ancestry analysis and admixture genetic mapping for proliferative diabetic retinopathy in African Americans. *Investigative Ophthalmology & Visual Science* 56: 3999-4005.

Tate, B.W. & McDaniel, M.A. (2008, August). Race differences in personality: An evaluation of moderators and publication bias. Paper presented at the Annual Meeting of the Academy of Management, Anaheim CA.

Telles, E. (2014). *Pigmentocracies: Ethnicity, Race, and Color in Latin America*. UNC Press Books.

Tucker-Drob, E.M. & Bates, T.C. (2015). Large cross-national differences in gene × socioeconomic status interaction on intelligence. *Psychological Science* 27: 138-149.

bioRxiv preprint doi: https://doi.org/10.1101/055681; this version posted March 27, 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.

*Valenzuela, C.Y. (1988). On sociogenetic clines. *Ethology and Sociobiology* 9: 259-268.

Van der Meer, T. & Tolsma, J. (2014). Ethnic diversity and its effects on social cohesion. *Annual Review of Sociology* 40: 459-478.

Van Ryn, M. (2002). Research on the provider contribution to race/ethnicity disparities in medical care. *Medical Care* 40(1): 1-140.

*Via, M., Gignoux, C.R., Roth, L.A., Fejerman, L., Galanter, J., Choudhry, S. ... & Martínez-Cruzado, J.C. (2011). History shaped the geographic distribution of genomic admixture on the island of Puerto Rico. *PLoS One* 6(1): e16513.

Viechtbauer, W. (2015). metafor: Meta-Analysis Package for R (Version 1.9-8). Retrieved from https://cran.r-project.org/web/packages/metafor/index.html

Winkler, C.A., Nelson, G.W. & Smith, M.W. (2010). Admixture mapping comes of age. *Annual Review of Genomics and Human Genetics* 11: 65-89.

Wraw, C., Deary, I.J., Gale, C.R. & Der, G. (2015). Intelligence in youth and health at age 50. *Intelligence* 53: 23-32.

*Zou, J.Y., Park, D.S., Burchard, E.G., Torgerson, D.G., Pino-Yanes, M., Song, Y.S., Sankararaman, S., Halperin, E. & Zaitlen, N. (2015). Genetic and socioeconomic study of mate choice in Latinos reveals novel assortment patterns. *Proceedings of the National Academy of Sciences* 112: 13621-13626.