Published in final edited form as:
Am J Hum Biol. 2014 ; 26(3): 347–360. doi:10.1002/ajhb.22521.

Correlation Analysis of Genetic Admixture and Social Identification with Body Mass Index in a Native American Community

Trina M. Norden-Krichmar[1], **Ian R. Gizer**[2], **Ondrej Libiger**[1], **Kirk C. Wilhelmsen**[3], **Cindy L. Ehlers**[4], and **Nicholas J. Schork**[1,5]

[1]Scripps Translational Science Institute and The Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037 USA

[2]Department of Psychological Sciences, University of Missouri, Columbia, MO 65211 USA

[3]Department of Genetics and Neurology, University of North Carolina, Chapel Hill, NC 27599 USA

[4]Department of Molecular and Cellular Neuroscience, The Scripps Research Institute, La Jolla, CA 92037 USA

[5]J. Craig Venter Institute, La Jolla, CA 92037 USA

Abstract

Objectives—Body mass index (BMI) is a well-known measure of obesity with a multitude of genetic and non-genetic determinants. Identifying the underlying factors associated with BMI is difficult because of its multifactorial etiology that varies as a function of geoethnic background and socioeconomic setting. Thus, we pursued a study exploring the influence of the degree of Native American admixture on BMI (as well as weight and height individually) in a community sample of Native Americans (n=846) while accommodating a variety of socioeconomic and cultural factors.

Methods—Participants' degree of Native American (NA) ancestry was estimated using a genome-wide panel of markers. The participants also completed an extensive survey of cultural and social identity measures: the Indian Culture Scale (ICS) and the Orthogonal Cultural Identification Scale (OCIS). Multiple linear regression was used to examine the relation between these measures and BMI.

Results—Our results suggest that BMI is correlated positively with the proportion of NA ancestry. Age was also significantly associated with BMI, while gender and socioeconomic measures (education and income) were not. For the two cultural identity measures, the ICS showed a positive correlation with BMI, while OCIS was not associated with BMI.

Conclusions—Taken together, these results suggest that genetic and cultural environmental factors, rather than socioeconomic factors, account for a substantial proportion of variation in BMI

Address correspondence to: Cindy L. Ehlers, Ph.D., The Scripps Research Institute, La Jolla, CA 92037 USA, 858-784-7058, cindye@scripps.edu, Nicholas J. Schork, Ph.D., J. Craig Venter Institute, La Jolla, CA 92037 USA, 858-200-1813, nschork@jcvi.org.

in this population. Further, significant correlations between degree of NA ancestry and BMI suggest that admixture mapping may be appropriate to identify loci associated with BMI in this population.

Keywords

Genetic ancestry; admixture; body habitus; obesity; Native Americans

INTRODUCTION

Admixed populations in the Americas, including Native American populations, African American populations, and Hispanic American populations, have emerged as predominant sub-populations in North America over the past several centuries (Seldin et al. 2011). Since there is known to be disease prevalence and incidence rate differences across different populations, the degree to which the genetic background and admixture contribute to disease phenotypes is an important question for genetic analyses. For example, identifying genes that influence phenotypes that may have arisen in a particular population and are propagated through admixture is not trivial. This is due to the fact that the ancestral backgrounds of participants in genetic association studies can influence the detection of specific genetic variants since there may be possible differences in the frequencies of those variants across different populations. In some cases, associated variants might even be specific to a population, while in others the genetic background of an individual could moderate the effect of that variant through subtle or overt interaction effects (Patterson et al. 2004; Williams et al. 2000; Winkler et al. 2010; Yang et al. 2011b). In genome-wide association studies, these phenomena can lead to false positive and false negative results (Lander and Schork 1994; Price et al. 2010). In this light, association studies pursued in the Americas to identify gene variants of relevance to disease must either accommodate the genetic backgrounds of the individuals participating in the study – for example through admixture mapping – or exploit analytical methods that correct for genetic background diversity (Freedman et al. 2004; Patterson et al. 2004; Price et al. 2010). A first step in determining the influence of admixture and genetic background on a phenotype is to explore the relationship between ancestral background and phenotypic expression. However, this is not trivial, given confounding between genetic and environmental (e.g., diet) factors between individuals, and the concern that different methods for assessing genetic background have been developed whose relative merits have not been well-studied.

Historically, and prior to the widespread use of genotyping, many studies exploring the relevance of genetic background and/or ancestry, relied on self-reported ancestry. However, self-reported ancestry has several limitations, the most important being that often an individual may not know their exact ‘percentage’ ancestry from different ethnic groups (Klimentidis et al. 2009a). Genotyping individuals with a set of specific ancestry informative markers (AIMs) and then using statistical methods to quantify proportion of ancestry from a particular ancestral population can also be problematic due to the limited ancestries that are targeted in a panel of AIMs and the small number of markers that are typically used. With the advent of next generation sequencing, more precise admixture estimation methods are possible, since, rather than relying on self-reported ancestry, or a

handful of ancestry informative markers (AIMs), the entire genome can be interrogated for signs of contributions from multiple ancestral populations simultaneously. Given these facts, we sought to explore the relationship between Native American ancestry estimates obtained from whole genome sequencing and a global epidemiologically-meaningful phenotype, obesity, as measured by body mass index (BMI).

Obesity is an important disease entity of great relevance to global population health, is highly heritable (Angeli et al. 2011; Ma et al. 2010; Peterson et al. 2011), and has been shown to be on the rise in a number of populations (Finucane et al. 2011; Stevens et al. 2012). In addition, phenotypic manifestations of obesity are known to vary across populations (Kopelman 2000; Yusuf et al. 2005) with particularly high rates noted among specific Native American tribes (Story et al. 2003). Obesity has also been linked to many medical disorders, such as diabetes, cardiovascular disease, cancer, and hypertension that vary in prevalence across populations (Finucane et al. 2011). Given the variations in prevalence across ethnic groups and substantial evidence that both BMI and height have strong heritable components (Maes et al. 1997; Silventoinen and Kaprio 2009; Yang et al. 2010), many studies have been pursued to determine if there is a unique genetic component to obesity in admixed populations such as Native Americans, African Americans, and Mexican Americans (Angeli et al. 2011; Ehlers and Wilhelmsen 2007; Klimentidis et al. 2009b; Ma et al. 2010; Nassir et al. 2012; Peterson et al. 2011; Williams et al. 2000). In fact, difficulties in assessing associations between specific gene variants and BMI may be due to limitations in the methods used to estimate admixture, as noted above.

In light of the emergence of obesity as a global public health epidemic and the fact that it is highly heritable, it is notable that studies exploring the relations between BMI and degree of admixture have been inconsistent, with some studies reporting positive correlations with BMI and some negative (Klimentidis et al. 2009b; Tang et al. 2006; Williams et al. 2000). In order to account for the more subtle factors that may influence the relationship between degree of admixture and BMI, we considered the use of whole genome sequence data from 697 individuals from a Native American (NA) population. Heritability estimates were generated by several different methods that utilized the pedigree information or the genotype data. We calculated the percentage of NA ancestry using different technologies and strategies (i.e., genotyping with exome chip, and low coverage whole genome sequencing). We also considered cultural and socioeconomic measures in order to uncover other factors that may play a role in BMI and that may confound the detection of an association between degree of genetic admixture and BMI. We believe our study is the most comprehensive to explore relations between BMI and admixture from both genetic and cultural standpoints in a NA population.

MATERIALS AND METHODS

Sample Ascertainment

Participants were recruited from eight geographically contiguous NA reservations, with a total population of about 3,000 individuals, using a combination of a venue-based method for sampling hard-to-reach populations (Kalton and Anderson 1986; Muhib et al. 2001), as well as a respondent-driven procedure (Heckathorn 1997) as previously described (Ehlers et

al. 2004; Gilder et al. 2004). The venues for recruitment included: tribal halls and culture centers, health clinics, tribal libraries, and stores on the reservations. A 10–25% rate of refusal was found depending on venue. Refusal rates were higher at tribal libraries and stores than health clinics and tribal halls/culture centers. Transportation from participants' homes to The Scripps Research Institute (TSRI) was provided by the study.

To be included in the study, participants had to be a Native American Indian indigenous to the catchment area, at least 1/16th Native American Heritage (NAH) based on conventional assessments and documentation of their federal Indian blood quantum, 18 years of age or older, and be mobile enough to be transported from his or her home to TSRI. The protocol for the study was approved by the Institutional Review Board (IRB) at TSRI, and the Indian Health Council, a tribal review group overseeing health issues for the reservations where recruitment was undertaken.

Potential participants first met individually with research staff to have the study explained and give written informed consent. During a screening period, participants had blood pressure and pulse taken, their height and weight measured, and completed a questionnaire that was used to gather information on demographics, personal medical history, ethnicity, and drinking history (Schuckit 1985). Each participant also completed an interview with the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA) (Bucholz et al. 1994; Hesselbrock et al. 1999), which was used to make medical and psychiatric diagnoses. Participants were asked to refrain from alcohol and drug usage for 24 hours prior to the testing. No individuals with detectable breath alcohol levels were included in the study dataset. During the screening period, the study coordinator also noted whether the participant was agitated, tremulous, or diaphoretic and their data were eliminated from subsequent analyses.

Self-reported ancestry

All subjects were interviewed for their self-reported blood degree of NA ancestry, given as a percent NA ancestry. Ambiguous values, such as “>50%”, “<50%”, or “don’t know”, were omitted from the correlation calculations. 527 subjects provided such a self-report-based percent Native American ancestry value.

Cultural identity

All subjects completed two different cultural identity measures: the Orthogonal Cultural Identification Scale (OCIS) and the Indian Culture Scale (ICS). Refer to the Supplemental Materials (Supplemental Tables 1 and 2) for the lists of questions that comprise these two measures. The Orthogonal Cultural Identification Scale (OCIS) was developed by Oetting and Beauvais (Oetting and Beauvais 1990; Oetting and Beauvais 1991). This scale’s internal consistency for subscale scores was high and both concurrent and discriminant validity were demonstrated in this American Indian population (Venner and Feldstein 2006). The Indian Culture scale consists of a list of activities associated with Indian culture that the participants may have engaged in over the last year such as sweat lodges, traditional games, pow-wows, etc. (Westermeyer and Neider 1986). The OCIS score is a fractional number between 0 – 4, with 4 representing maximum identification with Native American culture.

The ICS score is a whole number between 0 – 24, with 24 representing maximum identification with American Indian culture. Both cultural identity values were converted to percentages, so that they could be directly compared to the ancestry estimates. The OCIS score was available for 828 subjects, and the ICS score was available for 823 subjects.

Pedigree Structure

168 pedigrees containing 834 individuals were used in the genetic analyses. 87 families have only a single individual with phenotype data. All these individuals were included within some analyses to the extent that they contribute information about trait means and variance and the impact of covariates. The family sizes for the remaining families ranged between 2 and 379 subjects (average 4.96). 72 families had informative genetic data. The data includes 208 parent-child, 296 sibling, 94 half sibling, 33 grandparent-grandchild, 452 avuncular, and 748 cousin relative pairs. Several pedigrees contained large numbers of individuals resulting in distant relative pairs that were by their nature less informative for genetic analyses. Additionally, several pedigrees included complex loops and were included in the genetic analyses when possible.

Sample Preparation, Genotyping, and Sequencing

Blood samples were obtained by venopuncture and DNA was isolated from whole blood using an automated DNA extraction procedure as previously described (Wilhelmsen et al. 2003). The DNA samples were prepared per Affymetrix protocols and the exome chip genotyping was performed on the Affymetrix Axiom Exome 1A Array according to the Affymetrix Axiom 2.0 Assay Manual Workflow documentation. Variant quality from the exome chip was initially assessed according to Affymetrix best practices (Affymetrix 2011). Duplicates for 56 of the samples were also included in the run, which allowed the removal of any variant that displayed discordant results for more than three of the replicated samples. Plink version 1.07 (Purcell et al. 2007) was used to calculate Hardy-Weinberg (HWE) p-values on a subset of 239 unrelated samples, followed by the removal of variants with an HWE $p < 10^{-10}$.

For whole genome sequencing, the DNA libraries were prepared using the Illumina TruSeq DNA Sample Prep protocol, and paired-end sequencing was performed on Illumina HiSeq2000 sequencers. The whole genome sequencing coverage ranged from 3X to 12X, with an average read depth of 7.9. BWA version 0.5.8c (Li and Durbin 2009) was used to align sequencing reads to the genome. Picard (http://picard.sourceforge.net/) was used for de-duplicating and sorting the BAM files. GATK was used to realign possible indels and recalibrate BAM quality scores. Initial variant calls were calculated independently for each BAM file using the GATK Unified Genotyper and following the best practices for low-coverage samples (DePristo et al. 2011; McKenna et al. 2010). To verify sample identification, we compared these initial variant calls to the results of the exome chip genotyping. To generate the final variant files used in this study, linkage-disequilibrium (LD) aware variant calling was executed for the whole genome sequencing data using samtools-hybrid (http://genome.sph.umich.edu/wiki/Samtools-hybrid) (Li et al. 2009), BEAGLE (Browning and Browning 2007), and Thunder (Li et al. 2011; Li et al. 2010).

Heritability

The heritability (h^2) of each of BMI, weight, and height was calculated using SOLAR (Almasy and Blangero 1998) leveraging the pedigree information for the NA participants in this study as described above. The software packages Efficient Mixed-Model Association eXpedited (EMMAX) (Kang et al. 2010) and Genome-wide Complex Trait Analysis (GCTA) (Yang et al. 2011a) were used to calculate the genetic marker-based heritability of BMI, weight, and height with the marker genotype information extracted from the whole genome sequence (WGS) data obtained on each participant.

SOLAR estimates an h^2 parameter by utilizing the self-reported pedigree structure to partition the trait relative pair covariance for a phenotype into additive genetic and environmental contributions while also correcting for covariates. Participant's age at the time of evaluation and sex were evaluated as potential covariates and retained if they accounted for at least 5% of the total variance. The probability that h^2 was greater than zero was determined using a Student's t-test of regression coefficients for each scale. This test of significance was used to evaluate the potential genetic transmission of the traits examined in the present report.

GCTA estimates heritability by modeling the effects of all genotyped SNPs as random effects in a mixed linear model based on the genomic similarity of the individuals. EMMAX also computes a pairwise relatedness matrix from the genotyped markers. It then uses a variance component model to estimate the pseudo-heritability, which is the fraction of phenotypic variance explained by the relatedness matrix. For both GCTA and EMMAX, a cutoff at MAF >= 0.01 was applied to variants arising from the sequencing data, resulting in 6,358,436 markers used for the heritability estimates. The covariates used in GCTA analyses were gender, age, age squared, and European, African, and Native American ancestry proportions. The covariates used in EMMAX were gender, age, age squared, and Native American ancestry.

Ancestry estimate using ADMIXTURE with Affymetrix exome data

The Affymetrix Exome1A exome chip, containing 247,222 markers, was run with DNA from the Native American subjects. Genotyping calls were quality filtered and converted to PLINK format (Purcell et al. 2007). The quality-filtered PLINK files were input into the unsupervised clustering program ADMIXTURE (Alexander et al. 2009), to cluster the study participants into 3 populations, representing Native American, Caucasian European, and other (collectively) ethnicities. The ethnicity label assigned to each of the 3 clusters generated by ADMIXTURE was determined by comparison to the self-reported ancestry and the results from the Ancestry Estimator ANC4 program (see below). With ADMIXTURE we were able to calculate ancestry estimates for 727 of the subjects who had appropriate data.

Ancestry estimates using the Ancestry Estimator program (ANC4) with whole genome sequencing data

Following variant calling and imputation, the genotyping calls from the whole genome sequencing data were extracted at loci used in the ANC4 program (Libiger and Schork

2012). ANC4 calculates the percentage ancestry for each individual for the following 4 ethnicities: European, African, Native American, and East Asian. ANC4 is a supervised clustering program which uses input from genotype data on 364,470 loci collected on reference individuals from global populations (European, African, Native American, and East Asian), included by permission from a recent Native American population history study (Reich et al. 2012). There were 8,650 markers that overlapped in the Affymetrix Exome 1A array and the ANC4 reference panel. 697 subjects had whole genome sequencing data suitable for estimation with this program.

Correlation between ancestry methods and cultural identity measures

In order to assess simple correlations between individual ancestry/degree of admixture estimates and other variables collected on the study participants, custom R code was written. Scatterplots of the correlations were also generated.

Multiple linear regression

Custom R code was also written to calculate and plot multiple linear regression analyses in which BMI, height, and weight were taken as dependent variables with percentage Native American ancestry from the three estimation methods, as well as the two cultural identity scales, taken as independent variables. The following additional covariates were also included in the regression analyses as independent variables: age, age squared, gender, number of years of school, and gross income. Interaction effects between age and the ancestry measures, and gender and the ancestry measures were also tested in the regressions.

Notably, it is possible that correlations between NA ancestry and BMI could be spuriously induced if significant relations between degree of NA ancestry and the genetic similarity of individuals resulting from the nesting of participants within families were present, because the heritability of a trait and NA ancestry would thus be confounded. To address this possibility, we performed two additional analyses to determine if there was a significant relationship between the percentage of NA ancestry and the genetic relatedness of the subjects due to family membership. First, to estimate the amount of variation in the pedigree relationship that is explained by NA ancestry, we used the GAMOVA program (Nievergelt et al. 2007) with the ANC4 ancestry values as a quantitative trait against the distance matrix derived from the kinship coefficients of the subjects derived from the pedigree information. In the second analysis, we conducted a principal coordinate analysis of the kinship distance matrix using the matlab function ‘cmdscale’. The resulting dimensions were regressed against BMI and ancestry to determine their significance. The significant dimensions were then used as covariates in the multiple linear regression of BMI against percent NA ancestry to observe their effect on the correlation.

RESULTS

Demographics of the Native American population

The demographics for the full sample of individuals (N=846) that were included in the BMI correlations are shown in Table 1. The smaller subsets of individuals which had exome data only (N=727), sequencing data only (N=697), self-reported ancestry only (N=527),

Orthogonal Cultural Identity Scale data only (N=828), or Indian Cultural Scale data only (N=823) are described in Supplemental Table 3, which contains the demographic information for the subsets grouped by available data. In the full (n=846) sample set, the subjects exhibited a mean age of 31 (range 18 – 82) years of age, with 42% of the sample being male. Participants had a mean of 11.6 years of education (SD=1.6), and a mean income of $20,000-$29,000. Using self-reported ancestry, 48% of the participants reported at least 50% Native American heritage based on their federal Indian blood quantum. The mean BMI was 32 (SD=8, range 16 - 71), with 0.4% underweight, 19% normal weight, 26% overweight, and 54% obese, as defined by the WHO guidelines (Keil and Kuulasmaa 1989).

Heritability of BMI, weight, and height in the NA population

Heritability (h^2) of BMI, weight, and height were calculated to be 0.47, 0.47, and 0.58, respectively, using SOLAR and leveraging the pedigree information available on the Native American individual in this study (Ehlers and Wilhelmsen 2007). The genotype data (MAF >= 0.01) extracted from the whole genome sequence information on each study participant was used with the software EMMAX (Kang et al. 2010) to calculate the whole genome genotype-based heritability as well and BMI, weight, and height were estimated to have heritabilities of 0.38, 0.41, and 0.54, respectively. To contrast different methods of estimating heritability further, we used the whole genome genotype data and the marker-based heritability estimation method incorporated into the program GCTA (Yang et al. 2011a) and estimated the heritability of BMI, weight, and height to be 0.39, 0.41, and 0.57, respectively. The p-values and standard error values have been added as Supplemental Table 4. The different methods for calculating heritability exhibited some differences in overall estimates, but not overly so.

Correlation between genotype-based ancestry methods and cultural identity measures

The ancestry estimates for each individual calculated by ANC4 based on the sequencing data were sorted by increasing percentage Native American ancestry over all the individuals, and then plotted as a histogram (Figure 1). The other ancestry and cultural measures were sorted by subject identification number in order to align with the ANC4 histogram (Figure 1). The ancestry measures show the same basic trends when plotted in this way, but the differences between the cultural measurements are noticeable. The results of performing pairwise Pearson’s and Spearman’s correlations between the measurements (Table 2 and Figure 2) show a very high correlation between the ancestry estimates from the unsupervised program ADMIXTURE and a supervised method, ANC4 (r = 0.89). There is also a significant correlation between the calculated ancestry estimates and the self-reported ancestry (r = 0.66). Surprisingly, there was no significant correlation between the ancestry estimates and the cultural identity measures, and low correlation between the two cultural measures. This suggests that the genotype-based ancestry measures and the cultural identity measures are essentially capturing different aspects of the NA community biosocial milieu. Additionally, correlations were performed between the ancestry and cultural measures and the number of years of school and gross income. The number of years of school and income do not correlate to the ancestry and cultural measures, or to each other (Supplemental Figure 1, Supplemental Table 5).

To validate the accuracy of the ancestry estimates, we further examined the results from 14 mother-father-offspring trios in the population. Using the principle that a child should exhibit a Native American ancestry percentage that is approximately equal to the average of the two parents, we used the ancestry estimates generated by the ANC4 program to calculate the expected admixture percentage for the child in each trio. The difference of the expected proportion of NA ancestry based on the parents' admixture percentages and the generated ancestry estimate was calculated across all 14 trios. The average degree of Native American ancestry for the parents in the trios was 0.58 +/− 0.21, and ranged from 0.24–1.00 (i.e., 24% to 100% Native American ancestry). The average deviation from the expected ancestry was found to be 0.02 across all trios. There was no significant difference between the observed and expected ancestry proportions (paired t-test, $p = 0.11$). We note that the ancestry of the parents across the trios covered most of the range of Native American ancestry percentages in the entire sample, strengthening the validity of our findings.

Multiple linear regression analyses

The results for the multiple linear regression analyses treating BMI, height, and weight as dependent measures and estimated percentage Native American ancestry and the cultural identity measures (as well as important covariates discussed in the Methods section) as independent variables can be found in Figures 3–5 and Tables 3–5.

For all ancestry estimates, there was a significant positive correlation with BMI (self-reported ancestry p-value = 5.54E-05, Admixture p-value = 1.46E-04, ANC4 p-value = 1.56E-06) and weight (self-reported ancestry p-value = 2.12E-04, Admixture p-value = 2.0E-02, ANC4 p-value = 1.22E-04), and a significant negative correlation to height (self-reported ancestry p-value = 0.47, Admixture p-value = 5.94E-06, ANC4 p-value = 8.00E-03). The Indian Culture Scale exhibited a significant association with BMI (ICS p-value = 1.25E-03) and weight (ICS p-value = 3.07E-03), but not to height (ICS p-value = 0.45). In contrast to the other measures, the Orthogonal Culture Identity Scale did not show any significant correlation to BMI (OCIS p-value = 0.59), weight (OCIS p-value = 0.49), or height (OCIS p-value = 0.72).

To estimate the effect size of the association of ancestry with BMI, the unscaled estimates from the multiple linear regression models for the self-reported, ADMIXTURE, and ANC4 ancestry multiple linear regressions were averaged together to obtain a value of 5.6 units of BMI increase for a change from 0 to 100% in Native American ancestry. Rescaling the effect size, this corresponds to an increase in BMI of 0.56 units for every 10% increase in Native American ancestry.

From the GAMOVA analysis of ancestry against kinship, relatedness of the subjects was found to be significantly associated with degree of NA ancestry (p-value = 0.001), but it does not explain a substantial amount of the degree of NA ancestry variance ($r^2 = 0.006$). Similarly, the inclusion of the significant dimensions generated from the principal coordinate analysis of the kinship coefficients as covariates into the multiple linear regression analysis of BMI and ANC4 ancestry estimates, showed that the ancestry retained a highly significant correlation to BMI (p-value < 0.001). These results suggest the relation

between BMI and NA ancestry is independent of genetic similarity due to family relatedness in this population sample.

The socioeconomic measures assessing the number of years of school and gross income were found not to be significantly associated with BMI, height, or weight. Age and age squared, however, were significantly associated with BMI, height, and weight, but did not exhibit a significant interaction effect with the ancestry or the cultural identity terms on BMI, height or weight. Gender was also a significant predictor of height and weight, but also did not exhibit a significant interaction with the ancestry or the cultural identity measures on height or weight.

In order to test for collinearity between the variables, we tested for correlations between all pairs of predictors (i.e., age, gender, BMI, years of school, and gross income). In all cases, the correlation between the pairs of predictors was very low ($r < 0.25$), which indicates that we did not find evidence of collinearity in our sample.

To test if there were different constructs at play between self-reported ancestry and genotype-derived ancestry, we constructed two models for each of the genotype-derived ancestries (ADMIXTURE and ANC4). One model was the existing model (without self-reported ancestry), and the other model had self-reported ancestry as a covariate. We performed an ANOVA between the two models for each genotype-derived ancestry. For both ADMIXTURE and ANC4, the addition of self-reported ancestry was not significant (ANOVA results for ADMIXTURE and ANC4 models: $p=0.10$ and $p=0.76$, respectively).

DISCUSSION

Demographics

The prevalence of obesity, as defined by a BMI greater than 30 kg/m^2 (Keil and Kuulasmaa 1989), is 54% in the community we studied, as compared to a prevalence of 29% in North America, and 9.8–13% globally (Finucane et al. 2011). Overall, 80% of the subjects are classified as overweight or obese, with a BMI greater than 25 kg/m^2, with only 19% in the normal weight BMI range of 18–25 kg/m^2. Previous studies have indicated that obesity in Native American communities is widespread, and that there is variation across the different tribes (Story et al. 1999; Story et al. 2003). According to data available from the Indian Health Service Clinical Reporting System for American Indian and Alaska Native adults, the overall prevalence rate for obesity is 54%, and the prevalence for overweight or obese is 81% (Indian Health Service 2011). These values are in agreement with the percentages for obesity (54%) and overweight or obese (80%) which were observed in this study population.

Heritability

Similar to previous studies, we found a high heritability for BMI, height, and weight in the NA community we studied. Using three different methods (SOLAR, EMMAX, and GCTA), heritability for BMI was 0.38–0.47, heritability for weight was 0.41–0.47, and heritability for height was 0.54–0.58. These values are consistent with other heritability estimates for BMI from twin studies that considered ethnicities of all types, which ranged from 0.50–0.90 (Maes et al. 1997; Silventoinen and Kaprio 2009), and from a Brazilian family study

reporting a BMI heritability of 0.51 (de Oliveira et al. 2008). Heritability estimates for height in unrelated individuals have been reported as 0.45 (Yang et al. 2010). We noted that our heritability estimates for height may be inflated due to unknown and undocumented relatedness among the individuals in our study population.

Correlation between ancestry methods and cultural measures

We explored the relationship between genotype-based ancestry estimates and self-reported ancestry and found moderate evidence for such associations, but this is not surprising because of the inherent limitations in granularity and accuracy of self-reported measures (Klimentidis et al. 2009a). We believe that the supervised genotype-based ancestry estimates provided by the ANC4 program are more accurate than those from the ADMIXTURE program. We find that ADMIXTURE overestimates the ancestry at high percentages of NA ancestry, and underestimates the ancestry at the low percentages of NA ancestry. Refer to Figure 1 for a graphical representation of this phenomenon, where there are a larger number of subjects estimated by ADMIXTURE to be 0% or 100% NA ancestry, than are estimated by ANC4.

The cultural identity measures did not correlate well to each other, or to the ancestry measures. Self-reported ethnicity and cultural identity have been found in other studies to be uncorrelated to admixture determined by genotyping (Klimentidis et al. 2009a). The two cultural measures are also measurements of different aspects of cultural identity, so it is not surprising that they are not well correlated to each other.

Correlations between morphological phenotypes and ancestry and cultural measures

All NA ancestry estimate measures (self-report, ADMIXTURE, ANC4) exhibited a significant positive correlation with BMI. Regression of height and weight on percent Native American ancestry identified a negative and positive correlation, respectively. Age was highly significant predictor of BMI, height and weight, as was gender with height and weight, but with a lower level of statistical significance. Socioeconomic measures, such as number of years of school and gross income, were not significantly associated with BMI, height or weight.

The correlations of the morphological measurements to the two cultural identity measurements exhibited a mix of supporting and conflicting correlations to those shown by the ancestry estimates. The Indian Culture Scale (ICS) showed a positive correlation with BMI and weight. The Orthogonal Cultural Identification Scale (OCIS) was not correlated with BMI, weight, or height. These results are not surprising, since the two cultural identity measurements are not well correlated to each other, or to the ancestry estimates (Figures 1 and 2, Table 2). Cultural identity has not been found to be a reliable estimate of genetic admixture (Klimentidis et al. 2009a), so correlations with cultural identity should be viewed with caution. It is interesting, however, that the ICS demonstrated a positive correlation to BMI, which suggests that cultural factors may play a role in body mass.

A number of other studies have shown contrasting results for ancestry-percentage-based correlations with BMI in an admixed population. These conflicting results may be due to differences in the ethnic backgrounds of the participants, the number of subjects, the

covariates included, and the methods of estimating admixture. As an example, in a study of African American and Hispanic American post-menopausal women, there was a positive correlation to BMI for the African American women, but there was no correlation with BMI for the Hispanic American women (Nassir et al. 2012). Further, the authors of that study found a high positive correlation with waist-to-hip ratio (WHR) and Native American ancestry in the Hispanic women, which suggests that WHR may be more diagnostic of adiposity in post-menopausal women. Although the authors pursued a large study group of African American and Hispanic American women, the admixture was estimated with genotyping of only 92 AIMs, and it included post-menopausal women, instead of sampling men and women at a range of ages. Several studies found a positive correlation between Native American ancestry and BMI (Klimentidis et al. 2009b; Williams et al. 2000), while another study found a negative correlation (Tang et al. 2006). The present study attempted to reconcile these discrepant findings by utilizing genetic data obtained from whole genome sequencing to precisely estimate ancestry proportions and examine the relations of NA ancestry to BMI.

We also addressed another potential confounding factor in correlations of ancestry with BMI in admixed populations, which is the influence of the relatedness of the subjects in the study. Using several different methods to estimate the effect of relatedness, we found that the relation between NA ancestry and BMI was independent of familial relatedness.

We treated gross income and number of years of education as a proxy for socioeconomic status (SES). We found that SES factors were not significantly correlated with BMI in our population. These findings are similar to other studies that included these SES covariates, where little or no significant effect was found (Klimentidis et al. 2009b; Nassir et al. 2012). Although SES factors may have an influence on access to health care, studies of SES have found inconsistent relationships between SES and adiposity, for example, in children in developed countries (Shrewsbury and Wardle 2008).

It has been suggested that the increase in BMI in Native Americans is related to the shift from traditional food and activity levels to contemporary occidental high fat, high calorie diet and a more sedentary way of life (Story et al. 2003). However, some studies have found that diet did not have a significant association with the observed correlation of BMI to genetic ancestry (Nassir et al. 2012), and other studies have found no relationship between activity level and BMI (Klimentidis et al. 2009b). Additionally, it has been found that self-reported dietary intake questionnaires may be inaccurate or underestimate the caloric intake (Hebert et al. 2003). Diet and physical activity levels were not accessed in the present study.

While we acknowledge that diet, physical activity, environmental and socioeconomic status factors may all contribute to BMI, we chose to focus this study on the genetic component of BMI. By utilizing an ancestry estimation technique (Libiger and Schork 2012) that employs supervised clustering to a reference panel from 4 global populations using ~300K markers, we were able to confidently infer individual ancestry percentages in our population. We validated the ancestry estimates by comparison to exome chip data with the ADMIXTURE program, and by further examination of the ancestry estimates of the trios in the data set. We found a positive correlation between degree of Native American ancestry and BMI, in which

covariates of socioeconomic factors were not significant. We also found one cultural measurement scale to be positively correlated to BMI, while another was not. Taken together, these results suggest that genetic and cultural environmental factors contribute to BMI in this population, rather than socioeconomic factors. The difficulty of finding a specific gene variant associated with BMI may be due to admixture in a population, confounding effects of genealogy, social identity measures and other biosocial factors, although our study suggests there might be potential for using admixture-mapping approaches to uncover genes associated with BMI. Because excessive BMI is linked to major medical disorders such as diabetes and cardiovascular disease, identifying factors that contribute to elevated BMI may be important in informing prevention and intervention programs in this high-risk population.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding for this study was provided by grants from the National Institutes of Health (NIH); from the National Institute on Alcoholism and Alcohol Abuse (NIAAA) and the National Center on Minority Health and Health Disparities (NCMHD) 5R37 AA010201 (CLE) and the National Institute of Drug Abuse (NIDA) 5 R01 DA030976 (CLE, NJS). NIAAA, NCMHD and NIDA had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. NJS and his lab are supported in part by NIH grants 5 UL1 RR025774, R21 AI085374, 5 U01 DA024417, 5 R01 HL089655, 5 R01 AG035020, 1 R01 MH093500, 2 U19 AI063603, 2 U19 AG023122, 5 P01 AG027734, 1 R21 DA033813. The authors have no conflicts of interest to declare. We would also like to thank and acknowledge the following people for their role in the sequencing effort and upstream processing of the sequence data: Chris Bizon, Scott Chasse, Piotr Mieczkowski, Ewa Patrycja Malc, Joshua Sailsbery, and Phil Owens.

LITERATURE CITED

Affymetrix. Affymetrix: Best Practice Supplement to Axiom Genotyping Solution Data Analysis User Guide Rev. 2011; 1:1–33.

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009; 19(9):1655–1664. [PubMed: 19648217]

Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet. 1998; 62(5):1198–1211. [PubMed: 9545414]

Angeli CB, Kimura L, Auricchio MT, Vicente JP, Mattevi VS, Zembrzuski VM, Hutz MH, Pereira AC, Pereira TV, Mingroni-Netto RC. Multilocus analyses of seven candidate genes suggest interacting pathways for obesity-related traits in Brazilian populations. Obesity (Silver Spring). 2011; 19(6):1244–1251. [PubMed: 21233811]

Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet. 2007; 81(5):1084–1097. [PubMed: 17924348]

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, Reich T, Schmidt I, Schuckit MA. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report on the reliability of the SSAGA. J Stud Alcohol. 1994; 55(2):149–158. [PubMed: 8189735]

de Oliveira CM, Pereira AC, de Andrade M, Soler JM, Krieger JE. Heritability of cardiovascular risk factors in a Brazilian population: Baependi Heart Study. BMC Med Genet. 2008; 9:32. [PubMed: 18430212]

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K,

Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43(5):491–498. [PubMed: 21478889]

Ehlers CL, Wall TL, Betancourt M, Gilder DA. The clinical course of alcoholism in 243 Mission Indians. Am J Psychiatry. 2004; 161(7):1204–1210. [PubMed: 15229052]

Ehlers CL, Wilhelmsen KC. Genomic screen for substance dependence and body mass index in southwest California Indians. Genes Brain Behav. 2007; 6(2):184–191. [PubMed: 16764678]

Finucane MM, Stevens GA, Cowan MJ, Danaei G, Lin JK, Paciorek CJ, Singh GM, Gutierrez HR, Lu Y, Bahalim AN, Farzadfar F, Riley LM, Ezzati M. Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating G. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. Lancet. 2011; 377(9765):557–567. [PubMed: 21295846]

Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D. Assessing the impact of population stratification on genetic association studies. Nat Genet. 2004; 36(4):388–393. [PubMed: 15052270]

Gilder DA, Wall TL, Ehlers CL. Comorbidity of select anxiety and affective disorders with alcohol dependence in southwest California Indians. Alcohol Clin Exp Res. 2004; 28(12):1805–1813. [PubMed: 15608596]

Hebert JR, Patterson RE, Gorfine M, Ebbeling CB, St Jeor ST, Chlebowski RT. Differences between estimated caloric requirements and self-reported caloric intake in the women's health initiative. Ann Epidemiol. 2003; 13(9):629–637. [PubMed: 14732302]

Heckathorn DD. Respondent-driven sampling: A new approach to the study of hidden populations. Social Problems. 1997; 44(2):174–199.

Hesselbrock M, Easton C, Bucholz KK, Schuckit M, Hesselbrock V. A validity study of the SSAGA--a comparison with the SCAN. Addiction. 1999; 94(9):1361–1370. [PubMed: 10615721]

Indian Health Service. Healthy weight for life: A vision for healthy weight across the lifespan of American Indians and Alaska Natives, actions for health care teams and leaders. Rockville, MD: U.S. Department of Health and Human Services, Indian Health Service; 2011.

Kalton G, Anderson DW. Sampling rare populations. Journal of the Royal Statistical Society, Series A (General). 1986; 149:65–82.

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010; 42(4):348–354. [PubMed: 20208533]

Keil U, Kuulasmaa K. WHO MONICA Project: risk factors. Int J Epidemiol. 1989; 18(3 Suppl 1):S46–S55. [PubMed: 2807707]

Klimentidis YC, Miller GF, Shriver MD. Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. Am J Phys Anthropol. 2009a; 138(4):375–383. [PubMed: 18951390]

Klimentidis YC, Miller GF, Shriver MD. The relationship between European genetic admixture and body composition among Hispanics and Native Americans. Am J Hum Biol. 2009b; 21(3):377–382. [PubMed: 19214998]

Kopelman PG. Obesity as a medical problem. Nature. 2000; 404(6778):635–643. [PubMed: 10766250]

Lander ES, Schork NJ. Genetic dissection of complex traits. Science. 1994; 265(5181):2037–2048. [PubMed: 8091226]

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–1760. [PubMed: 19451168]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25(16):2078–2079. [PubMed: 19505943]

Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR. Low-coverage sequencing: implications for design of complex trait association studies. Genome Res. 2011; 21(6):940–951. [PubMed: 21460063]

Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010; 34(8):816–834. [PubMed: 21058334]

Libiger O, Schork NJ. A Method for Inferring an Individual's Genetic Ancestry and Degree of Admixture Associated with Six Major Continental Populations. Front Genet. 2012; 3:322. [PubMed: 23335941]

Ma L, Hanson RL, Traurig MT, Muller YL, Kaur BP, Perez JM, Meyre D, Fu M, Korner A, Franks PW, Kiess W, Kobes S, Knowler WC, Kovacs P, Froguel P, Shuldiner AR, Bogardus C, Baier LJ. Evaluation of A2BP1 as an obesity gene. Diabetes. 2010; 59(11):2837–2845. [PubMed: 20724578]

Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. Behav Genet. 1997; 27(4):325–351. [PubMed: 9519560]

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297–1303. [PubMed: 20644199]

Muhib FB, Lin LS, Stueve A, Miller RL, Ford WL, Johnson WD, Smith PJ. Community Intervention Trial for Youth Study T. A venue-based method for sampling hard-to-reach populations. Public Health Rep. 2001; 116(Suppl 1):216–222. [PubMed: 11889287]

Nassir R, Qi L, Kosoy R, Garcia L, Allison M, Ochs-Balcom HM, Tylavsky F, Manson JE, Shigeta R, Robbins J, Seldin MF. Relationship between adiposity and admixture in African-American and Hispanic-American women. Int J Obes (Lond). 2012; 36(2):304–313. [PubMed: 21487399]

Nievergelt CM, Libiger O, Schork NJ. Generalized analysis of molecular variance. PLoS Genet. 2007; 3(4):e51. [PubMed: 17411342]

Oetting ER, Beauvais F. Orthogonal cultural identification theory: the cultural identification of minority adolescents. Int J Addict. 1990; 25(5A-6A):655–685. [PubMed: 2101397]

Oetting ER, Beauvais F. Critical incidents: failure in prevention. Int J Addict. 1991; 26(7):797–820. [PubMed: 1660043]

Patterson N, Hattangadi N, Lane B, Lohmueller KE, Hafler DA, Oksenberg JR, Hauser SL, Smith MW, O'Brien SJ, Altshuler D, Daly MJ, Reich D. Methods for high-density admixture mapping of disease genes. Am J Hum Genet. 2004; 74(5):979–1000. [PubMed: 15088269]

Peterson RE, Maes HH, Holmans P, Sanders AR, Levinson DF, Shi J, Kendler KS, Gejman PV, Webb BT. Genetic risk sum score comprised of common polygenic variation is associated with body mass index. Hum Genet. 2011; 129(2):221–230. [PubMed: 21104096]

Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet. 2010; 11(7):459–463. [PubMed: 20548291]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81(3):559–575. [PubMed: 17701901]

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, Parra MV, Rojas W, Duque C, Mesa N, Garcia LF, Triana O, Blair S, Maestre A, Dib JC, Bravi CM, Bailliet G, Corach D, Hunemeier T, Bortolini MC, Salzano FM, Petzl-Erler ML, Acuna-Alonzo V, Aguilar-Salinas C, Canizales-Quinteros S, Tusie-Luna T, Riba L, Rodriguez-Cruz M, Lopez-Alarcon M, Coral-Vazquez R, Canto-Cetina T, Silva-Zolezzi I, Fernandez-Lopez JC, Contreras AV, Jimenez-Sanchez G, Gomez-Vazquez MJ, Molina J, Carracedo A, Salas A, Gallo C, Poletti G, Witonsky DB, Alkorta-Aranburu G, Sukernik RI, Osipova L, Fedorova SA, Vasquez R, Villena M, Moreau C, Barrantes R, Pauls D, Excoffier L, Bedoya G, Rothhammer F, Dugoujon JM, Larrouy G, Klitz W, Labuda D, Kidd J, Kidd K, Di Rienzo A, Freimer NB, Price AL, Ruiz-Linares A. Reconstructing Native American population history. Nature. 2012; 488(7411):370–374. [PubMed: 22801491]

Schuckit MA. Genetics and the risk for alcoholism. JAMA. 1985; 254(18):2614–2617. [PubMed: 4057470]

Seldin MF, Pasaniuc B, Price AL. New approaches to disease mapping in admixed populations. Nat Rev Genet. 2011; 12(8):523–528. [PubMed: 21709689]

Shrewsbury V, Wardle J. Socioeconomic status and adiposity in childhood: a systematic review of cross-sectional studies 1990–2005. Obesity (Silver Spring). 2008; 16(2):275–284. [PubMed: 18239633]

Silventoinen K, Kaprio J. Genetics of tracking of body mass index from birth to late middle age: evidence from twin and family studies. Obes Facts. 2009; 2(3):196–202. [PubMed: 20054225]

Stevens GA, Singh GM, Lu Y, Danaei G, Lin JK, Finucane MM, Bahalim AN, McIntire RK, Gutierrez HR, Cowan M, Paciorek CJ, Farzadfar F, Riley L, Ezzati M. Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating G. National, regional, and global trends in adult overweight and obesity prevalences. Popul Health Metr. 2012; 10(1):22. [PubMed: 23167948]

Story M, Evans M, Fabsitz RR, Clay TE, Holy Rock B, Broussard B. The epidemic of obesity in American Indian communities and the need for childhood obesity-prevention programs. Am J Clin Nutr. 1999; 69(4 Suppl):747S–754S. [PubMed: 10195597]

Story M, Stevens J, Himes J, Stone E, Rock BH, Ethelbah B, Davis S. Obesity in American-Indian children: prevalence, consequences, and prevention. Prev Med. 2003; 37(6 Pt 2):S3–S12. [PubMed: 14636804]

Tang H, Jorgenson E, Gadde M, Kardia SL, Rao DC, Zhu X, Schork NJ, Hanis CL, Risch N. Racial admixture and its impact on BMI and blood pressure in African and Mexican Americans. Hum Genet. 2006; 119(6):624–633. [PubMed: 16738946]

Venner KL, Feldstein SW. Natural history of alcohol dependence and remission events for a Native American sample. J Stud Alcohol. 2006; 67(5):675–684. [PubMed: 16847535]

Westermeyer J, Neider J. Cultural affiliation among American Indian alcoholics: correlations and change over a ten-year period. Ann N Y Acad Sci. 1986; 472:179–188. [PubMed: 3467612]

Wilhelmsen KC, Schuckit M, Smith TL, Lee JV, Segall SK, Feiler HS, Kalmijn J. The search for genes related to a low-level response to alcohol determined by alcohol challenges. Alcohol Clin Exp Res. 2003; 27(7):1041–1047. [PubMed: 12878909]

Williams RC, Long JC, Hanson RL, Sievers ML, Knowler WC. Individual estimates of European genetic admixture associated with lower body-mass index, plasma glucose, and prevalence of type 2 diabetes in Pima Indians. Am J Hum Genet. 2000; 66(2):527–538. [PubMed: 10677313]

Winkler CA, Nelson GW, Smith MW. Admixture mapping comes of age. Annu Rev Genomics Hum Genet. 2010; 11:65–89. [PubMed: 20594047]

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010; 42(7):565–569. [PubMed: 20562875]

Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011a; 88(1):76–82. [PubMed: 21167468]

Yang JJ, Cheng C, Devidas M, Cao X, Fan Y, Campana D, Yang W, Neale G, Cox NJ, Scheet P, Borowitz MJ, Winick NJ, Martin PL, Willman CL, Bowman WP, Camitta BM, Carroll A, Reaman GH, Carroll WL, Loh M, Hunger SP, Pui CH, Evans WE, Relling MV. Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. Nat Genet. 2011b; 43(3):237–241. [PubMed: 21297632]

Yusuf S, Hawken S, Ounpuu S, Bautista L, Franzosi MG, Commerford P, Lang CC, Rumboldt Z, Onen CL, Lisheng L, Tanomsup S, Wangai P Jr, Razak F, Sharma AM, Anand SS, Investigators IS. Obesity and the risk of myocardial infarction in 27,000 participants from 52 countries: a case-control study. Lancet. 2005; 366(9497):1640–1649. [PubMed: 16271645]

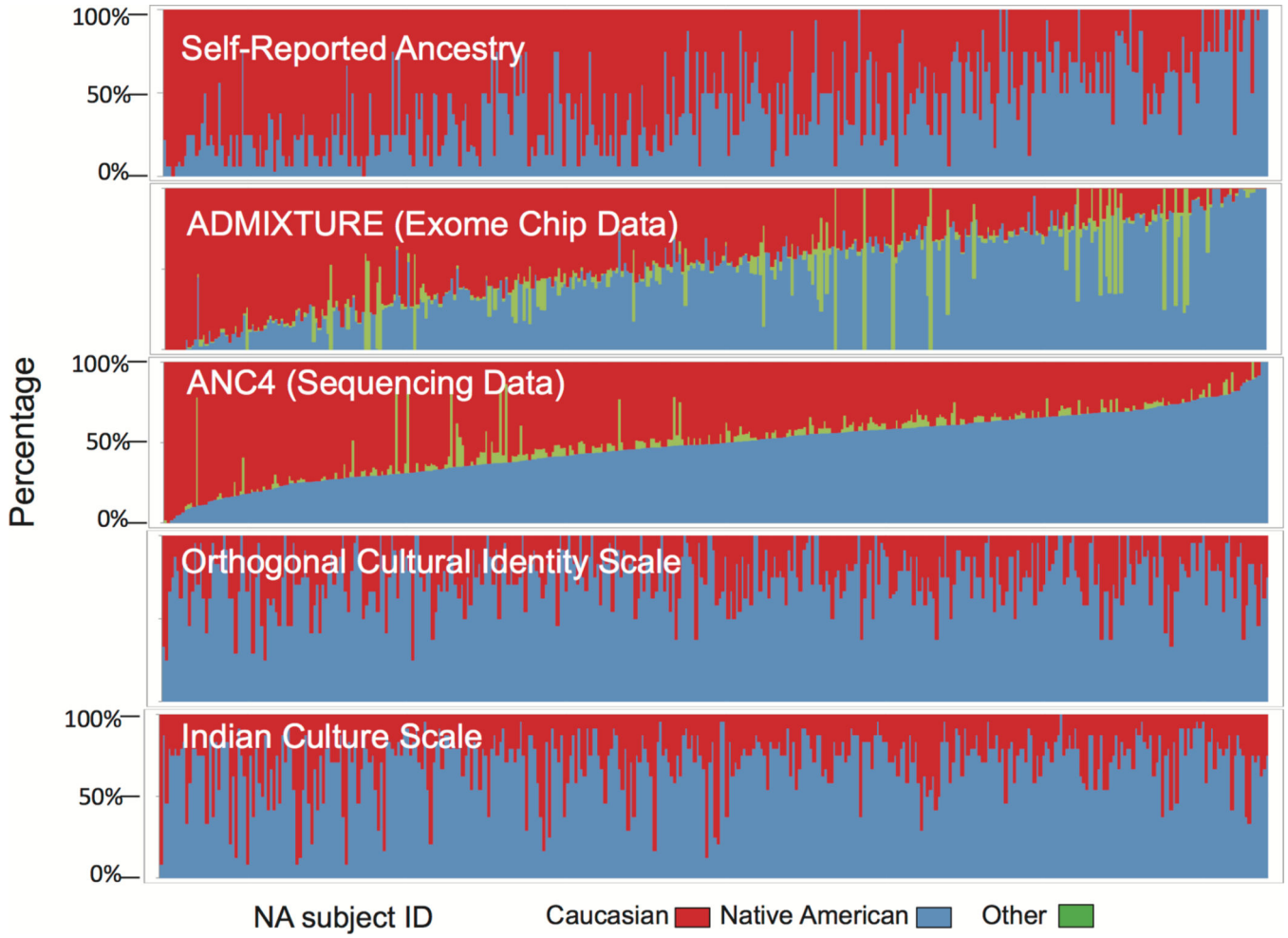

Figure 1. Percentage of Native American ancestry and cultural identity

Ancestry was estimated by 3 methods: self-reported ancestry, the ADMIXTURE program with exome chip data, and the program ANC4 with the whole genome sequencing data. The percentage of Native American identification was estimated using 2 cultural scales: Orthogonal Cultural Identity Scale (OCIS) and Indian Culture Scale (ICS). The ANC4 ancestry estimates were sorted in order of increasing percentage of Native American ancestry. All the other ancestry or cultural measures were sorted by subject identification number, in order to align with the ANC4 plot.

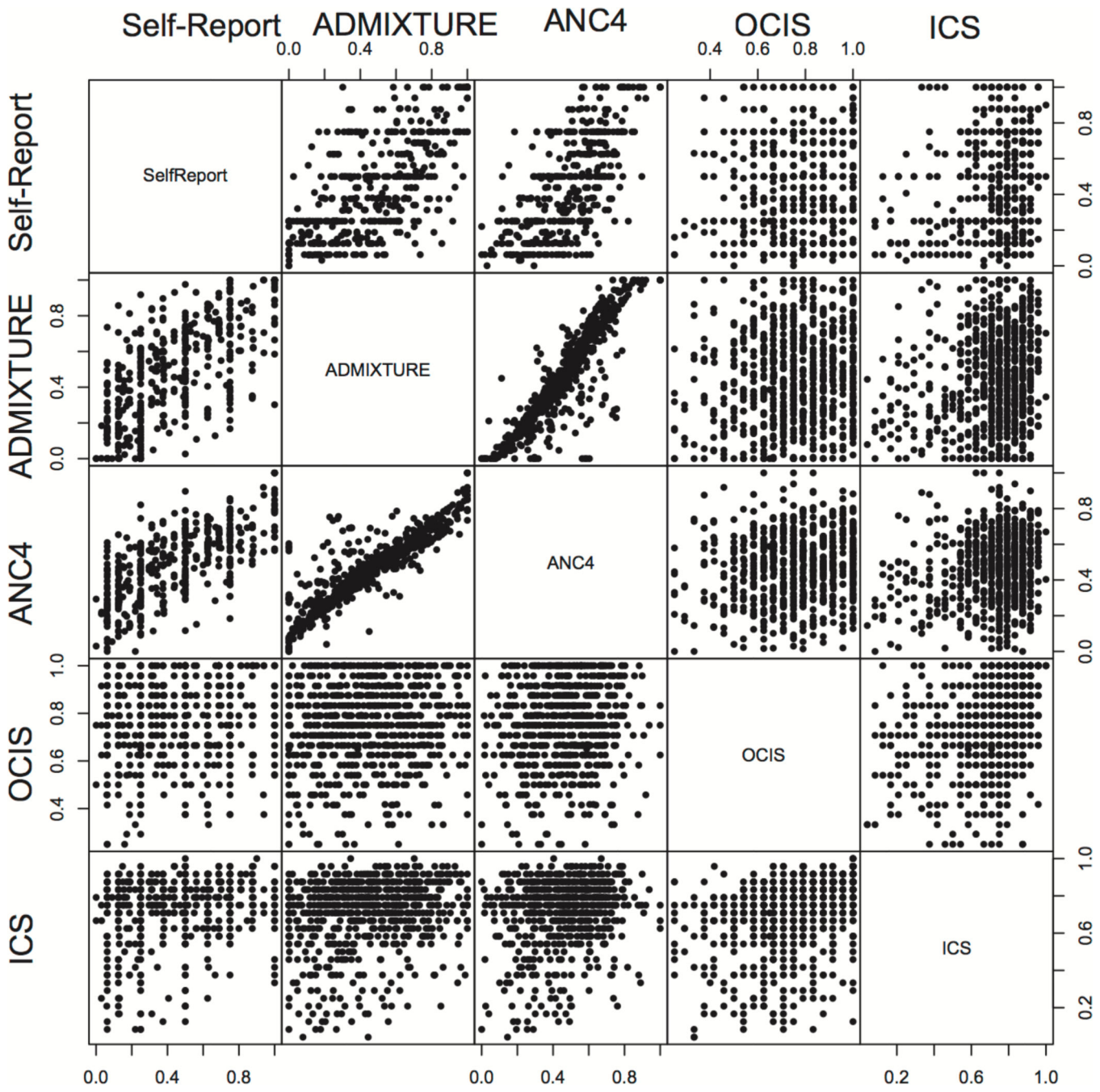

Figure 2. Correlation plot between ancestry estimates and cultural measures

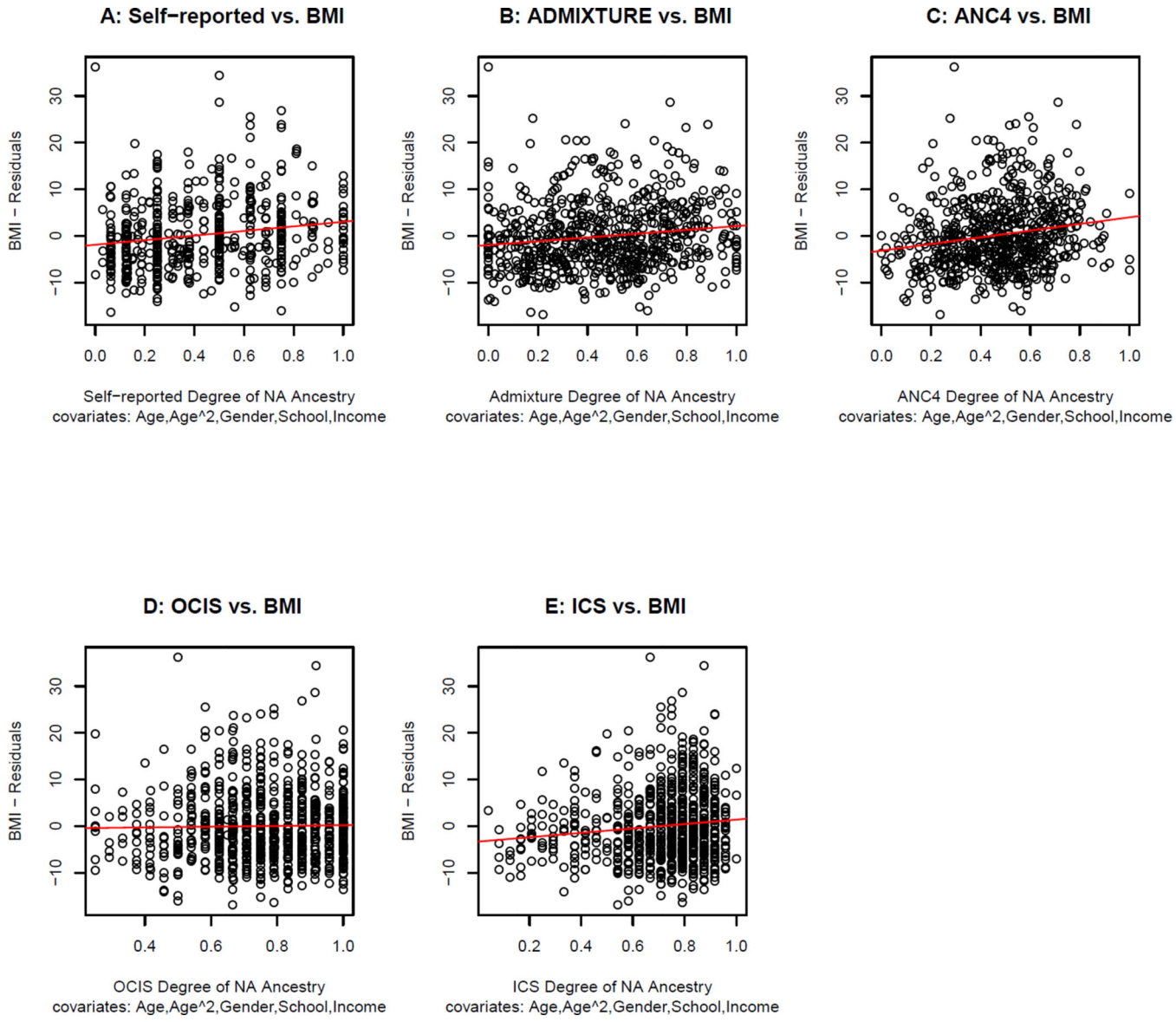

Figure 3. Regression of BMI and %NA ancestry and cultural measures

Covariates: age, age squared, gender, number of years of education, and income. Significant correlations were present for self-reported ancestry (p-value=5.54E-05), ADMIXTURE with the exome chip data (p-value= 1.46E-04), ANC4 with the sequencing data (p-value=1.56E-06), and Indian Culture Scale (ICS) (p-value=1.25E-03).

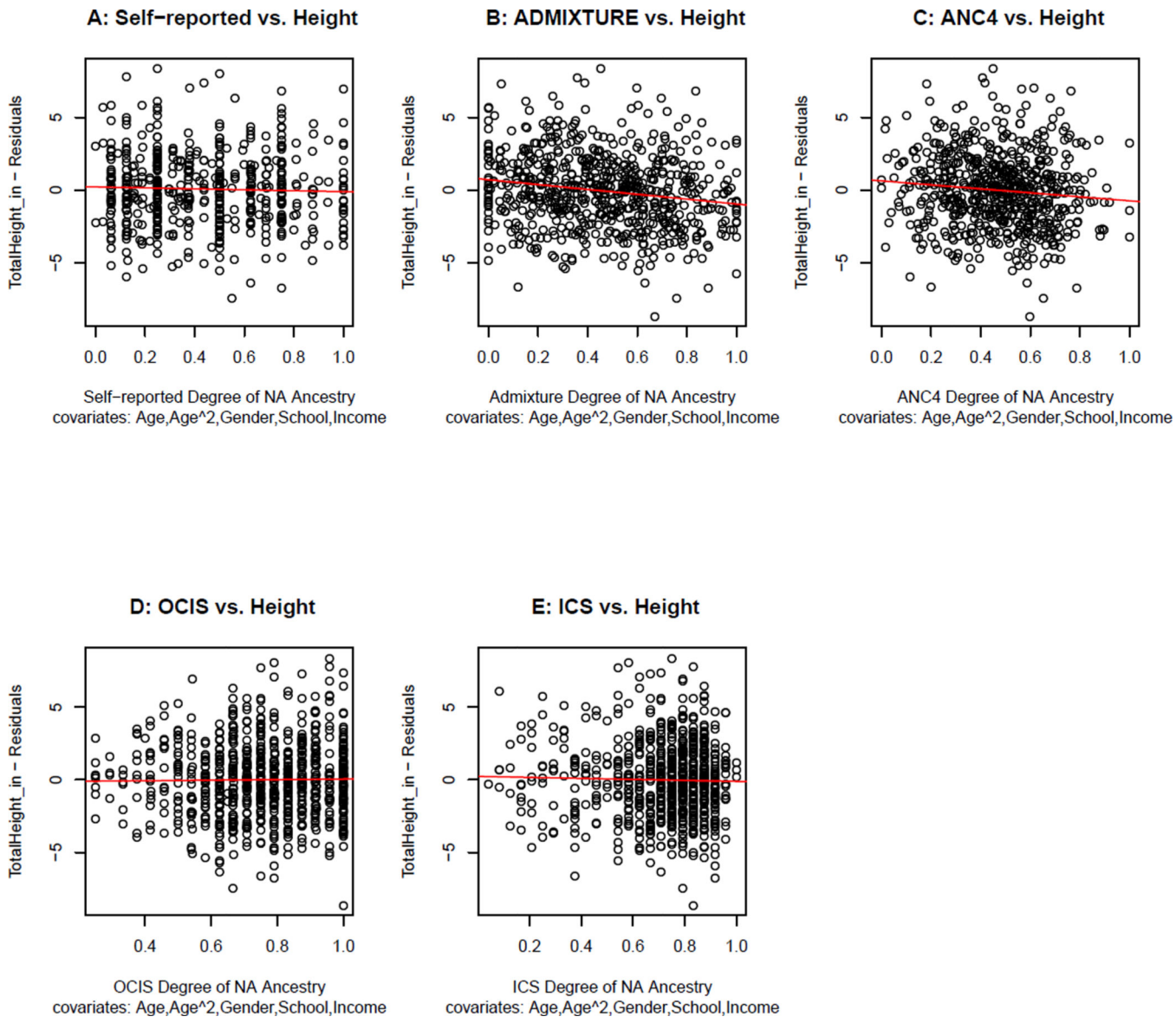

Figure 4. Regression of Height (in inches) and %NA ancestry and cultural measures

Covariates: age, age squared, gender, number of years of education, and income. Significant correlations were present for ADMIXTURE with the exome chip data (p-value=5.94E-06) and ANC4 with the sequencing data (p-value=8.00E-03).

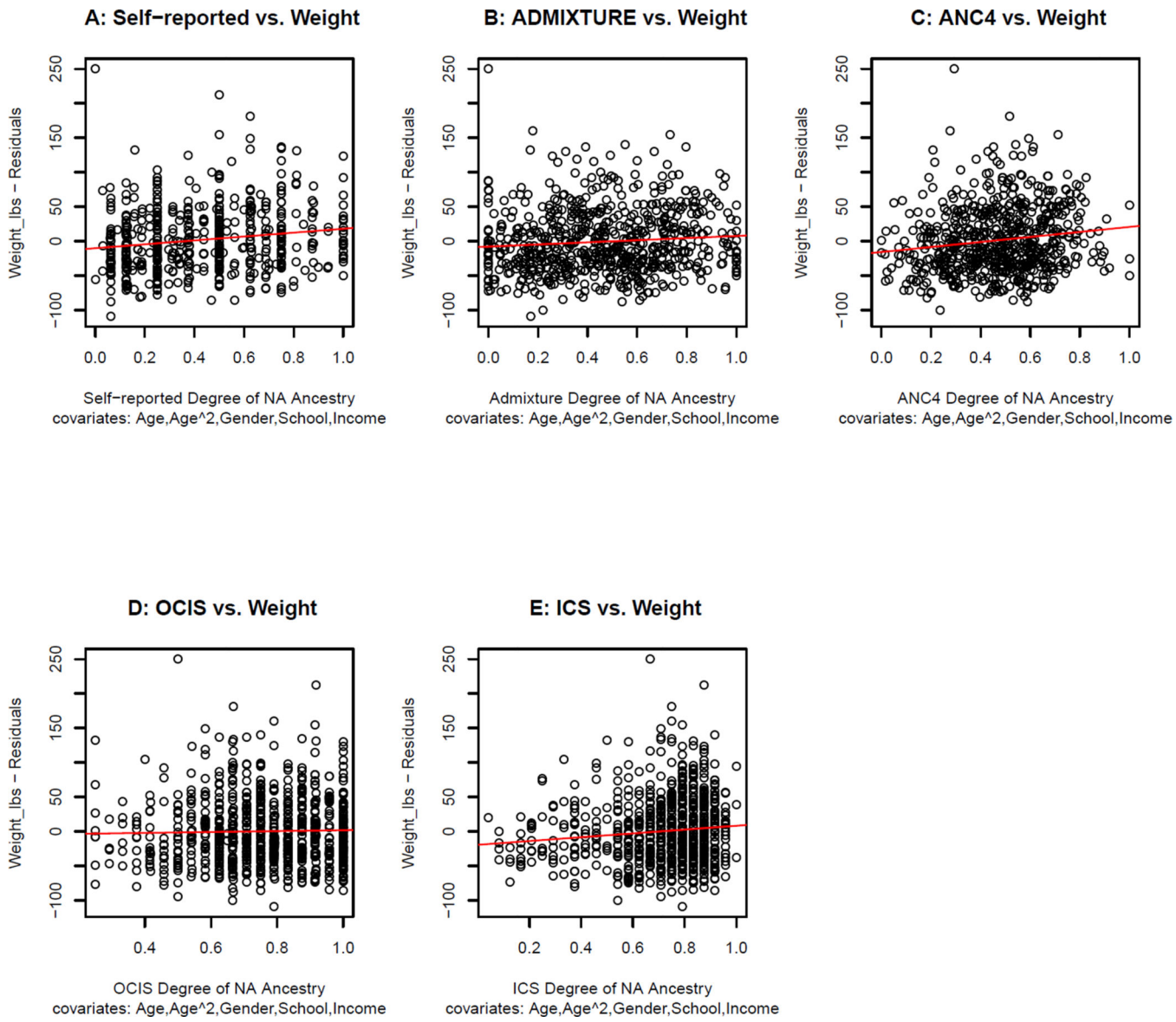

Figure 5. Regression of Weight (in pounds) and %NA ancestry and cultural measures

Covariates: age, age squared, gender, number of years of education, and income. Significant correlations were present for self-reported ancestry (p-value=2.12E-04), ADMIXTURE with the exome chip data (p-value=2.0E-02), ANC4 with the sequencing data (p-value=1.22E-04), and Indian Culture Scale (ICS) (p-value=3.07E-03).

Table 1

Demographics for Native American study participants.

	Full sample (n=846)	Males (n=351) [42%]	Females (n=495) [58%]
Age [range, mean(sd)]	18-82, 31.2 (13.2)	18-80, 29.8 (12.6)	18-82, 32.3 (13.6)
NA heritage >= 50% (self-reported)	48%	50%	46%
Education [range in years, mean(sd)]	3-17, 11.6 (1.6)	6-16, 11.5 (1.4)	3-17, 11.6 (1.7)
Income [mean(sd)]	3.27 (2.1) $20,000-$29,999/yr	3.4 (2.2)	3.2 (2.1)
Proportion in income category			
Unreported	70/846 = 8%	37/351 = 11%	33/495 = 7%
1= $1,000-$9,999/yr	172/846 = 20%	62/351 = 18%	110/495 = 22%
2=$10,000-$19,999/yr	178/846 = 21%	74/351 = 21%	104/495 = 21%
3=$20,000-$29,999/yr	137/846 = 16%	55/351 = 16%	82/495 = 17%
4= $30,000-$49,999/yr	85/846 = 10%	39/351 = 11%	46/495 = 9%
5=$40,000-$49,999/yr	53/846 = 6%	14/351 = 4%	39/495 = 8%
6=$50,000-$74,999/yr	82/846 = 10%	40/351 = 11%	42/495 = 8%
7=$75,000-$99,999/yr	36/846 = 4%	13/351 = 4%	23/495 = 5%
8=$100k-$149,999/yr	19/846 = 2%	11/351 = 3%	8/495 = 2%
9=$150,000 or more/yr	14/846 = 2%	6/351 = 2%	8/495 = 2%
BMI [range, mean(sd)]	15.9-71.0, 31.9 (8.1)	15.9-59.1, 31.0 (7.1)	16.7-71.0, 32.6 (8.7)
% with BMI < 18 (underweight)	4/846 = 0.5%	2/351 = 0.6%	2/495 = 0.4%
% with BMI 18-25 (normal weight)	160/846 = 19%	66/351 = 19%	94/495 = 19%
% with BMI 25-30 (overweight)	221/846 = 26%	109/351 = 31%	112/495 = 23%
% with BMI > 30 (obese)	461/846 = 54%	174/351 = 50%	287/495 = 58%
% with BMI 30-35 (Obesity Class I)	210/846 = 25%	89/351 = 25%	121/495 = 24%
% with BMI 35-40 (Obesity Class II)	126/846 = 15%	47/351 = 13%	79/495 = 16%
% with BMI > 40 (Obesity Class III)	125/846 = 15%	38/351 = 11%	87/495 = 18%

Table 2

Correlation between ancestry estimation methods and cultural identity measures.

		Self-Rep	ADMIX	ANC4	OCIS	ICS
Self-Report	Pearson	1	0.659 (p<2.2e-16)	0.668 (p<2.2e-16)	0.101 (p=0.02)	0.195 (p=8.3e-06)
	Spearman	1	0.653 (p<2.2e-16)	0.666 (p<2.2e-16)	0.102 (p=0.02)	0.168 (p=1.2e-04)
	N	526	447	431	514	511
ADMIXTURE	Pearson		1	0.887 (p<2.2e-16)	0.156 (p=2.9e-05)	0.189 (p=4.3e-07)
	Spearman		1	0.885 (p<2.2e-16)	0.146 (p=9.7e-05)	0.166 (p=9.4e-06)
	N		726	635	709	704
ANC4	Pearson			1	0.133 (p=5.2e-04)	0.188 (p=8.8e-07)
	Spearman			1	0.123 (p=1.3e-03)	0.147 (p=1.2e-04)
	N			695	679	674
OCIS	Pearson				1	0.358 (p<2.2e-16)
	Spearman				1	0.421 (p<2.2e-16)
	N				826	814
ICS	Pearson					1
	Spearman					1
	N					821

Table 3

Multiple regression analyses results for BMI as the outcome variable using all predictors.

Predictor	Unscaled Estimate +/− Std Error	Scaled Estimate +/− Std Error	p-value	
Self-Reported Ancestry Model				
(N=502, r2=0.102, DF=501, p=6.54E-10)				
Self-Reported Ancestry	5.15 +/− 1.27	1.39 +/− 0.34	5.54E-05	*
Gender	1.05 +/− 0.70	0.52 +/− 0.34	0.13	
Age	0.58 +/− 0.13	7.61 +/− 1.70	9.05E-06	*
Age squared	−0.01 +/− 0.00	−6.27 +/− 1.67	1.93E-04	*
#Years of school	0.13 +/− 0.21	0.21 +/− 0.33	0.53	
Gross income	0.26 +/− 0.17	0.56 +/− 0.36	0.12	
ADMIXTURE Ancestry Model				
(N=664, r2=0.114, DF=663, p=3.21E-15)				
ADMIXTURE Ancestry	4.10 +/− 1.07	1.08 +/− 0.28	1.46E-04	*
Gender	0.55 +/− 0.58	0.27 +/− 0.29	0.35	
Age	0.57 +/− 0.11	7.50 +/− 1.44	2.67E-07	*
Age squared	−0.01 +/− 0.00	−5.69 +/− 1.46	1.03E-04	*
#Years of school	0.10 +/− 0.18	0.15 +/− 0.29	0.60	
Gross income	0.04 +/− 0.14	0.08 +/− 0.29	0.79	
ANC4 Ancestry Model				
(N=635, r2=0.122, DF=634, p=9.76E-16)				
ANC4 Ancestry	7.53 +/− 1.55	1.45 +/− 0.30	1.56E-06	*
Gender	0.34 +/− 0.60	0.17 +/− 0.30	0.58	
Age	0.57 +/− 0.11	7.59 +/− 1.41	9.45E-08	*
Age squared	−0.01 +/− 0.00	−5.89 +/− 1.40	2.75E-05	*
#Years of school	0.25 +/− 0.19	0.39 +/− 0.30	0.20	
Gross income	0.20 +/− 0.15	0.43 +/− 0.31	0.16	
Orthogonal Cultural Identity Scale Model				
(N=762, r2=0.088, DF=761, p=3.79E-13)				
Orthogonal Cultural Identity Scale	0.86 +/− 1.59	0.15 +/− 0.28	0.59	
Gender	0.90 +/− 0.57	0.44 +/− 0.28	0.12	
Age	0.61 +/− 0.10	8.02 +/− 1.36	6.20E-09	*
Age squared	−0.01 +/− 0.00	−6.23 +/− 1.36	5.09E-06	*
#Years of school	0.05 +/− 0.18	0.08 +/− 0.29	0.78	
Gross income	0.02 +/− 0.14	0.05 +/− 0.29	0.87	
Indian Culture Scale Model				
(N=756, r2=0.106, DF=755, p=3.77E-16)				
Indian Culture Scale	4.88 +/− 1.50	0.92 +/− 0.28	1.25E-03	*
Gender	0.85 +/− 0.57	0.42 +/− 0.28	0.13	

Predictor	Unscaled Estimate +/− Std Error	Scaled Estimate +/− Std Error	p-value	
Age	0.62 +/− 0.10	8.23 +/− 1.36	2.36E-09	*
Age squared	−0.01 +/− 0.00	−6.24 +/− 1.36	5.02E-06	*
#Years of school	0.04 +/− 0.18	0.06 +/− 0.29	0.83	
Gross income	0.04 +/− 0.14	0.08 +/− 0.29	0.77	

* Denotes p-values less than 0.05. Interaction between ancestry and age was not significant. Interaction between ancestry and gender was not significant.

Table 4

Multiple regression analyses results for Height (in inches) as the outcome variable using all predictors.

Predictor	Unscaled Estimate +/− Std Error	Scaled Estimate +/− Std Error	p-value	
Self-Reported Ancestry Model				
(N=502, r2=0.536, DF=501, p<2.2E-16)				
Self-Reported Ancestry	−0.31 +/− 0.42	−0.08 +/− 0.11	0.47	
Gender	−5.43 +/− 0.23	−2.68 +/− 0.11	<2E-16	*
Age	−0.01 +/− 0.04	−0.14 +/− 0.57	0.80	
Age squared	0.00 +/− 0.00	−0.22 +/− 0.56	0.70	
#Years of school	−0.05 +/− 0.07	−0.07 +/− 0.11	0.50	
Gross income	0.01 +/− 0.06	0.03 +/− 0.12	0.80	
ADMIXTURE Ancestry Model				
(N=664, r2=0.544, DF=663, p<2.2E-16)				
ADMIXTURE Ancestry	−1.68 +/− 0.37	−0.44 +/− 0.10	5.94E-06	*
Gender	−5.30 +/− 0.20	−2.62 +/− 0.10	<2E-16	*
Age	−0.02 +/− 0.04	−0.31 +/− 0.50	0.54	
Age squared	0.00 +/− 0.00	0.02 +/− 0.50	0.97	
#Years of school	0.04 +/− 0.06	0.06 +/− 0.10	0.57	
Gross income	0.03 +/− 0.05	0.06 +/− 0.10	0.53	
ANC4 Ancestry Model				
(N=635, r2=0.543, DF=634, p<2.2E-16)				
ANC4 Ancestry	−1.41 +/− 0.53	−0.27 +/− 0.10	8.00E-03	*
Gender	−5.45 +/− 0.21	−2.69 +/− 0.10	<2E-16	*
Age	−0.01 +/− 0.04	−0.19 +/− 0.48	0.70	
Age squared	0.00 +/− 0.00	−0.02 +/− 0.48	0.96	
#Years of school	0.03 +/− 0.07	0.05 +/− 0.10	0.60	
Gross income	0.03 +/− 0.05	0.06 +/− 0.10	0.59	
Orthogonal Cultural Identity Scale Model				
(N=762, r2=0.529, DF=761, p<2.2E-16)				
Orthogonal Cultural Identity Scale	0.19 +/− 0.53	0.03 +/− 0.09	0.72	
Gender	−5.40 +/− 0.19	−2.66 +/− 0.09	<2E-16	*
Age	−0.01 +/− 0.03	−0.14 +/− 0.45	0.75	
Age squared	0.00 +/− 0.00	−0.16 +/− 0.45	0.73	
#Years of school	0.04 +/− 0.06	0.06 +/− 0.10	0.52	
Gross income	0.04 +/− 0.05	0.08 +/− 0.10	0.41	
Indian Culture Scale Model				
(N=756, r2=0.528, DF=755, p<2.2E-16)				
Indian Culture Scale	−0.38 +/− 0.50	−0.07 +/− 0.09	0.45	
Gender	−5.37 +/− 0.19	−2.65 +/− 0.09	<2E-16	*

Predictor	Unscaled Estimate +/− Std Error	Scaled Estimate +/− Std Error	p-value
Age	−0.01 +/− 0.03	−0.15 +/− 0.46	0.74
Age squared	0.00 +/− 0.00	−0.18 +/− 0.45	0.69
#Years of school	0.05 +/− 0.06	0.09 +/− 0.10	0.37
Gross income	0.05 +/− 0.05	0.10 +/− 0.10	0.31

* Denotes p-values less than 0.05. Interaction between ancestry and age was not significant. Interaction between ancestry and gender was not significant.

Table 5

Multiple regression analyses results for Weight (in pounds) as the outcome variable using all predictors.

Predictor	Unscaled Estimate +/− Std Error	Scaled Estimate +/− Std Error	p-value	
Self-Reported Ancestry Model				
(N=502, r2=0.130, DF=501, p=3.97E-13)				
Self-Reported Ancestry	29.79 +/− 7.98	8.01 +/− 2.15	2.12E-04	*
Gender	−26.96 +/− 4.38	−13.29 +/− 2.16	1.56E-09	*
Age	3.60 +/− 0.81	47.63 +/− 10.70	1.04E-05	*
Age squared	−0.04 +/− 0.01	−41.57 +/− 10.52	8.83E-05	*
#Years of school	0.47 +/− 1.32	0.74 +/− 2.08	0.72	
Gross income	1.92 +/− 1.08	4.04 +/− 2.27	0.08	
ADMIXTURE Ancestry Model				
(N=664, r2=0.132, DF=663, p<2.2E-16)				
ADMIXTURE Ancestry	15.47 +/− 6.88	4.09 +/− 1.82	0.02	*
Gender	−28.42 +/− 3.71	−14.01 +/− 1.83	6.99E-14	*
Age	3.42 +/− 0.70	45.30 +/− 9.25	1.22E-06	*
Age squared	−0.03 +/− 0.01	−36.05 +/− 9.34	1.25E-04	*
#Years of school	0.88 +/− 1.18	1.39 +/− 1.87	0.46	
Gross income	0.37 +/− 0.88	0.79 +/− 1.86	0.67	
ANC4 Ancestry Model				
(N=635, r2=0.156, DF=634, p<2.2E-16)				
ANC4 Ancestry	38.47 +/− 9.95	7.41 +/− 1.92	1.22E-04	*
Gender	−31.03 +/− 3.86	−15.30 +/− 1.90	4.51E-15	*
Age	3.56 +/− 0.68	47.10 +/− 9.02	2.40E-07	*
Age squared	−0.04 +/− 0.01	−37.97 +/− 8.95	2.55E-05	*
#Years of school	1.73 +/− 1.24	2.73 +/− 1.95	0.16	
Gross income	1.42 +/− 0.93	2.99 +/− 1.97	0.13	
Orthogonal Cultural Identity Scale Model				
(N=762, r2=0.116, DF=761, p<2.2E-16)				
Orthogonal Cultural Identity Scale	7.01 +/− 10.10	1.24 +/− 1.79	0.49	
Gender	−26.78 +/− 3.62	−13.21 +/− 1.79	3.83E-13	*
Age	3.76 +/− 0.65	49.74 +/− 8.66	1.36E-08	*
Age squared	−0.04 +/− 0.01	−40.66 +/− 8.62	2.84E-06	*
#Years of school	0.51 +/− 1.16	0.81 +/− 1.83	0.66	
Gross income	0.43 +/− 0.87	0.91 +/− 1.83	0.62	
Indian Culture Scale Model				
(N=756, r2=0.130, DF=755, p<2.2E-16)				
Indian Culture Scale	28.39 +/− 9.56	5.33 +/− 1.80	3.07E-03	*
Gender	−27.01 +/− 3.61	−13.32 +/− 1.78	2.05E-13	*

Predictor	Unscaled Estimate +/− Std Error	Scaled Estimate +/− Std Error	p-value	
Age	3.85 +/− 0.65	50.98 +/− 8.65	5.64E-09	*
Age squared	−0.04 +/− 0.01	−40.83 +/− 8.63	2.65E-06	*
#Years of school	0.48 +/− 1.15	0.76 +/− 1.81	0.68	
Gross income	0.59 +/− 0.86	1.24 +/− 1.81	0.49	

* Denotes p-values less than 0.05. Interaction between ancestry and age was not significant. Interaction between ancestry and gender was not significant.