Recent fluctuations in Mexican American genomes have altered the genetic architecture of biomedical traits

Authors: Melissa L. Spear[1,2], Alex Diaz-Papkovich[3,4], Elad Ziv[5,6,7,8], Simon Gravel[4,9], Dara G. Torgerson[4,9], Ryan D. Hernandez[2,4,7,9,10,11*]

1. Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA, USA
2. Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA
3. Quantitative Life Sciences Program, McGill University, Montreal, QC, Canada
4. McGill University and Genome Quebec Innovation Center, McGill University, Montreal, QC, Canada
5. Division of General Internal Medicine, University of California, San Francisco, San Francisco, CA, USA
6. Department of Medicine, University of California, San Francisco, San Francisco, CA, USA
7. Institute of Human Genetics, University of California, San Francisco, San Francisco, CA, USA
8. Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, USA
9. Department of Human Genetics, McGill University, Montreal, QC, Canada
10. Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, CA, USA
11. Institute for Human Genetics, University of California, San Francisco, CA, USA

*Corresponding author: ryan.hernandez@me.com

Abstract

Hispanics/Latinos are a diverse group of admixed populations with African, European, and Native American ancestries. They remain understudied, and thus little is known about the genetic architecture of phenotypic variation in these populations. Using genome-wide genotype data from the Hispanic Community Health Study/Study of Latinos, we find that Native American ancestry has increased over time across Hispanic/Latino populations, particularly in Mexican Americans where Native American ancestry increased by an average of ~20% over the 50-year period spanning 1940s-1990s. We find similar patterns across American cities, and replicate our observations in an independent sample of Mexican Americans. These dynamic ancestry patterns are a result of a complex interaction of several population and cultural factors, including strong ancestry-related assortative mating and subtle shifts in migration with differences in subcontinental Native American ancestry over time. These factors have shaped patterns of genetic variation, including an increase in runs of homozygosity in Native American ancestral tracts, and also influenced the genetic architecture of complex traits within the Mexican American population. We show for height, a trait correlated with ancestry, polygenic risk scores based on summary statistics from a European-based genome-wide association study perform poorly in Mexican Americans. Our findings reveal temporal changes in population structure within Hispanics/Latinos that may influence biomedical traits, demonstrating a crucial need to improve our understanding of the genetic diversity of admixed populations.

Introduction

Hispanics/Latinos living in the United States are culturally, phenotypically, and genetically diverse populations. Individuals who identify as Hispanic/Latino have varying proportions of Native American, African, and European genetic ancestry, each with its own unique continental demographic history. Demographic forces such as population bottlenecks and expansions, migration and adaptation to novel environments resulted in observable differences in continental

53 patterns of genetic variation (1-3). These differing patterns were shaped by many historical events
54 of migration which included the founding of the Americas by Native American populations, the
55 colonization by Europeans, and the African slave trade (4-8). These large-scale migrations and
56 additional demographic events shaped the genetic diversity of individuals living today within the
57 United States (9-13).

58 Demographic history has shaped the genetic architecture of modern human phenotypic
59 variation (14-19), and is critical to consider in the search for the genetic basis of complex diseases.
60 The demography of the United States has changed drastically over the 20th century, and by 2044
61 is predicted to become a 'minority-majority' country whereby no one racial/ethnic group comprises
62 more than 50% of the population (20). By 2060 Hispanics/Latinos are projected to make up the
63 largest of that share at 29% or 119 million individuals (20). However, to date, population-based
64 medical genomics research [and its subsequent benefits, including polygenic risk score (PRS)
65 profiling] have been disproportionately focused on individuals of European descent, with the
66 findings primarily benefiting European populations (21, 22). Despite the increases in sample
67 sizes, rates of discovery, and traits studied, Hispanic or Latin American participation in genome-
68 wide association studies (GWAS) has continued to hover around 1% (23, 24). This shows a
69 continuing trend of leaving this population particularly vulnerable to falling behind in receiving the
70 benefits of the precision medicine revolution (22, 23).

71 In this study we utilize the largest genetic study of Hispanics/Latinos in the U.S. to date --
72 the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) (10) -- to understand how
73 patterns of genetic variation in Hispanic/Latino populations in the United States have changed
74 over the last century, and evaluate the impact such changes may be having on complex traits.

75

76 **Results**

77 **Global ancestry proportions among HCHS/SOL Hispanic/Latino Populations**

Using the subset of sites that overlapped with our African, European, and Native American reference panels, we called 3-way global ancestry estimates for 10,268 unrelated HCHS/SOL individuals (see methods). Figure 1A summarizes the global ancestry proportions shaded by admixture estimates in a ternary plot, recapitulating the original HCHS/SOL analysis of continental ancestry (10). However, while several population groups appear to have overlapping ancestry proportions, this analysis masks more subtle structure in subcontinental ancestry. To investigate subtle population structure across these self-identified population groups, we performed UMAP on the top 3 principal components (see methods), and find substantial structure across self-identified groups (Figure 1B and Supplementary Figure 1B). We find that Dominicans, who have the highest average proportions of African ancestry, are in the middle, with Puerto Ricans and Cubans, diverging in opposite directions (Supplementary Figure 1B) with clines of increasing European ancestry proportions (Figure 1B). Further, while self-identified Mexican, Central, and South American groups appear to have overlapping ancestry proportions in Figure 1A, UMAP represents the Mexican Americans and Central/South American groups as large, separate wings that diverge from self-identified Cubans and Dominicans, with both clusters diverging with clines of increasing ancestry toward different Native American populations (Figure 1B and Supplementary Figure 1B and 1C).

Dynamic Global Ancestry Proportions in Mexican Americans

For each of the HCHS/SOL Latino populations, we evaluated differences in global ancestry estimates over time while accounting for the sampling method (referred to as “sampling weight”, see methods) used for the design of the HCHS/SOL study (25). We found that in all populations, the effect size for Native American ancestry on birth year is positive, though only statistically significant after multiple testing in the Mexican American (β =0.0023; P=3.58E-22; Figure 1C) and Central American (β=0.0013; P=0.0013) cohorts (Supplementary Table 1). Due

to the larger sample size, magnitude of the effect, and statistical significance, we shift our focus to Mexican Americans. In Mexican Americans, the increase in Native American global ancestry over time was consistent across multiple data stratifications including recruitment region, US born or not US born, educational attainment, and sex (Table 1), and was robust to alternative methods for estimating global ancestry proportions (e.g. based on the summation of RFMix local ancestry estimates; Supplementary Figures 2 and 3). We performed bootstrap resampling (n=1000) of global Native American ancestry for the Mexican Americans and observed a consistent increase in Native American ancestry with fitted LOESS smoothing (Figure 1D) and when individuals were binned by birth year decades (Supplementary Figure 4). On average, global Native American ancestry has increased ~20% over the last 50 years in Mexican Americans.

We replicated our original findings of the increase in global Native American ancestry over time in a smaller, independent cohort of self-identified Mexican Americans (n=705) from the Health and Retirement Study (HRS) [34]. The HRS Mexican Americans in this study are older compared to the HCHS/SOL Mexican Americans (birth year distribution: 1915-1981; mean=1943,median:1942) and have lower levels of global Native American ancestry on average (mean=0.29), but we still observed an increase in global Native American ancestry over time (β=0.00082; P=0.02; SE=0.0003673) (Supplementary Figure 5A). We performed 1000 bootstrap resampling iterations of the linear regression model (global Native American ancestry ~ birth year) fitted to the data. From these resampling iterations, 98.2% of the tests had a slope > 0 (average β=0.00083) and 61.5% of the regression p-values were less than 0.05 as illustrated in (Supplementary Figures 5B-5D).

A previous study (12) identified ancestry biased migration in African Americans where individuals with higher proportions of European ancestry migrated first out of the South during the Great Migration followed by individuals with higher proportions African ancestry. We hypothesized that earlier immigrants to the US had higher proportions of European ancestry followed by recent

immigrants having higher proportions of global Native American ancestry. In our non-US born individuals (N=2987), we evaluated differences in ancestry estimates over time while accounting for years in the US and sampling weight and identified a significant effect of years in the US (β=-0.0009; P=0.0006; SE=0.0003). However, this did not change the effect of birth year on the proportion of global Native American ancestry (β =0.0028; P<2e-16, SE=0.0003).

For US born individuals we assessed whether parental birth place could explain the increases in global Native American ancestry. Of the 634 US born individuals, 385 had parents both born outside of the US, 149 had one parent born outside of the US. 97 had both parents born within the US. For the 385 individuals, we identified a significant association between birth year and global Native American ancestry (β=0.004, P=2.34e-10, SE=0.0006). For the remaining individuals we were unable to identify a significant effect of birth year on global Native American ancestry, possibly due to a small sample size.

Individual loci are not driving global ancestry proportions

We used local ancestry estimates generated across the genome to perform admixture mapping in HCHS/SOL Mexican Americans to determine if younger individuals harbored excess Native American ancestry in certain regions of the genome. Although we tried two different models (see methods), we did not find any loci to be significantly associated with birth year across the genome (Supplementary Figure 6).

Little evidence for subcontinental population structure

It is possible that the increase in global Native American ancestry over time could be biased by changes in the specific subcontinental Native American ancestries over time (though such an effect is not visible in our UMAP analysis, Figure 1B). If it were the case, then we would expect subtle signals of genetic divergence in Native American ancestry tracts over time. To

153 investigate this, we calculated F_{ST} between all pairs of birth-decades (see methods). Figure 2A
154 shows all pairwise comparisons among birth-decades, and demonstrates that while the estimates
155 of F_{ST} are negligible (with many estimates below 0), there is a subtle trend of increasing F_{ST} as
156 birth-decade differences increase (though individuals born in the 80s and 90s show a conflicting
157 pattern). We further investigated this pattern using genetic diversity, π, in Native American
158 ancestry tracts for each birth-decade (see methods). We hypothesized that if there were
159 increased migration from multiple Native American source populations (coupled with rapid
160 population growth in Mexican American communities), then genetic diversity should be increasing
161 over time. We found the opposite: Supplementary Figure 7 shows a subtle decrease in genetic
162 diversity (π) over time from the 1930s to the 1980s in non-US born Mexican Americans, and a
163 subtle decrease in US born Mexican Americans from the 70s to the 90s (while remaining roughly
164 constant from the 30s to the 70s).
165
166 **Little evidence that Native American ancestry tract lengths have changed**
167 We next sought to test whether differences at the local ancestry level could explain the
168 shift in global Native American ancestry over time in the Mexican Americans. We calculated the
169 length of each RFMix inferred local ancestry tract in each Mexican American individual, and tested
170 for differences in the distribution of tract lengths across birth-decades using a multiple linear
171 regression model (see methods). We found no significant associations between the decade bin
172 and the proportion of Native American ancestral tracts at various lengths (Figure 2B), even when
173 testing for violations of model assumptions (e.g. normalizing the tracts per bin by the number of
174 individuals, or excluding the 1930s and/or 1990s individuals due to the small sample size in each
175 bin).
176
177

178 **Increased runs of homozygosity over time**

179 Since genetic diversity has decreased over time in the Native American ancestry tracts of
180 Mexican Americans (despite rapid growth of the census population size), it is possible that this
181 population has also experienced increased haplotype homozygosity over time. We investigated
182 this possibility by exploring runs of homozygosity (ROH) in Native American ancestry tracts in
183 each of the 3622 Mexican Americans. We classified ROH into three categories: short, medium,
184 and long, based on the length distribution in the population. Generally, short ROH are tens of
185 kilobases in length and likely reflect the homozygosity of old haplotypes; medium ROH are
186 hundreds of kilobases in length and likely reflect background relatedness in the population; and
187 long ROH are hundreds of kilobases to several megabases in length and are likely the result of
188 recent parental relatedness. Figure 2C shows a fitted loess curve to the log of the total length of
189 ROH summed across each Mexican American's genome as a function of their birth year, broken
190 down by ROH size class (as well as the total of each size class that overlaps all ancestry tracts
191 (Supplementary Figure 8A) and Native American ancestry tracts (Supplementary Figure 8B)).
192 Overall, we find a significant positive correlation between birth year and the total summed ROH
193 across size classes (τ=0.0449, P=6.12e-5, Kendall's rank correlation), but this becomes more
194 significant when we restrict our analysis to ROH calls that overlap Native American ancestry tracts
195 (τ=0.0873, P=9.46e-15). When stratified by size class, the associations (all Kendall's rank
196 correlation) in ROH were primarily driven by the short (τ=0.0833, P=9.45e-14), and medium
197 (τ=0.0718, P=1.46e-10) size classes, and are again strongest when ROH overlap Native
198 American ancestry tracts (short τ=0.107, P<2.2e-16, medium τ=0.1003, P<2.2e-16). The long
199 ROH had a negative correlation with birth year, but was insignificant after multiple testing (τ=-
200 0.0291, P=0.01499; note that 1694 individuals did not have any long ROH calls in their genome).
201
202

203 **Strong ancestry-related assortative mating in HCHS/SOL Mexicans**

204 Given that short and medium length ROH have increased over time, it appears that background
205 relatedness within Native American ancestry in Mexican Americans has increased over time (but
206 not an increase in recent parental relatedness). One way for this to occur is if individuals with
207 similar ancestry patterns tend to mate with one another more often than expected under a model
208 of random mating (i.e. assortative mating). To measure assortative mating, we estimated the
209 ancestral proportions of the biological parents of each HCHS/SOL Mexican American (see
210 methods). With individuals from all decades pooled together, we found the inferred biological
211 parental Native American ancestries to be significantly correlated (Figure 2D, r=0.708, 95%
212 CI:0.69-0.72, P<2.2e-16, Pearson correlation). When stratified by decade, the correlation in
213 inferred parental Native American global ancestry ranged from 0.65 to 0.74 (Supplementary
214 Figure 9), but were not statistically different from each other. This shows that there was a strong
215 parental ancestry correlation among Mexican Americans over different generations. This
216 signature of assortative mating is not due to recent parental relatedness, because there is no
217 trend in long ROH with birth year (and an overall low rate of long ROH among Mexican
218 Americans).

219

220 **Genetic correlation of global Native American ancestry with biomedical traits**

221 We have shown that genetic variation in the Mexican American population is dynamic, with Native
222 American ancestry increasing over a short period of time (combined with decreased genetic
223 diversity and increased short and medium length ROH within Native American ancestry tracts).
224 These features may have implications for the genetic architecture of complex traits within Mexican
225 Americans, a topic that is understudied and poorly understood. To further our understanding of
226 the genetic architecture of complex traits in Mexican Americans, we investigated the relationship
227 between Native American ancestry and various complex traits that may be relevant to biomedical
228 phenotypes. Specifically, we tested for a correlation between 66 complex traits from the

229 HCHS/SOL phenotypic dataset and global Native American ancestry (Kendall's tau). As illustrated
230 in Figure 3, many of these traits are significantly correlated after Bonferroni correction
231 (P<0.00076). We found that the effect of global Native American ancestry on many of these
232 phenotypes persisted when using multiple regression to account for age, sex, center, and the
233 sampling weight (Supplementary Table 2), highlighting the need for increased investigation into
234 the role of Native American genetic ancestry in admixed populations such as Mexican Americans.
235

236 **Assessing the genetic contribution of Native American ancestry to height**

237 Among the traits we tested for a correlation with global Native American ancestry, height
238 had the strongest negative correlation, and our regression model indicated that height also had a
239 strong positive relationship with birth year (Figure 4A and Supplementary Table 3). Globally,
240 populations have grown taller over time due to a variety of non-genetic, environmental factors
241 (26). We find a similar trend in the HCHS/SOL Mexican Americans (Figure 4A). Indeed, when we
242 stratified individuals by quartiles of global Native American ancestry, we see that all quartiles have
243 increased in height by a similar amount over the period investigated (though individuals with lower
244 Native American ancestry were taller on average).

245 Height is one of the most highly studied complex traits, with GWAS sample sizes
246 numbering in the hundreds of thousands (27). Results for many of these studies have been made
247 readily available on public databases as summary association statistics that can be leveraged to
248 build genetic predictions through polygenic risk scores (PRS) (28). In Europeans, PRS have been
249 shown to have great predictive power for several traits, including breast cancer, prostate cancer,
250 and type 1 diabetes (22, 29-31). PRS are most effective in populations of European descent as
251 GWAS studies have been primarily performed in these populations (21-23) and are expected to
252 be biased when applied to other populations due to differences in the genetic architecture of traits
253 across diverse populations (32). Since Mexican Americans have some fraction of European

ancestry, we sought to determine whether PRS calculated utilizing GWAS summary statistics from European populations could still provide useful insight.

To evaluate the effectiveness of PRS for height (i.e. the polygenic height score, or PHS, see methods), we first tested whether there was an association between the observed height and the predicted height estimates while controlling for sampling weight, sex, and recruitment center (see methods). We identified a significant association between observed height and predicted height for the population as a whole (β=0.0044881, P=2.19e-12; Figure 4B, Supplementary Table 4). However, when we stratified by quartiles of Native American global ancestry, the association only remained for the individuals in the lower two quartiles of global Native American ancestry (NAM<0.37: β=0.004, P=0.0008 and 0.36<NAM<0.46: β=0.004, P=0.003, Supplementary Table 4). The association between predicted height and observed height was no longer significant for individuals in the upper two quartiles of global Native American ancestry proportions (0.46<NAM<0.58: β=0.0011, P=0.39 and 0.58<NAM: β=0.0022, P=0.08, Supplementary Table 4).

As we had found global Native American ancestry to be increasing over time, we hypothesized that there would be a change in PHS over time as well. However, we find little evidence supporting this hypothesis. While individuals born earlier than 1950 or in the 1950s have a stronger correlation between their PHS and observed height (β=0.034 and 0.039; p=5.6e-4 and 2.7e-7 respectively) than individuals born in the 1960s, 1970s, or 1980s (β=0.016, 0.029, and 0.029; p=0.044, 0.0066, and 7.8e-5 respectively), there is no clear trend and we did not find a significant effect of birth year on PHS (P=0.09) even when we stratified by the quartiles of global Native American ancestry.

Discussion

The United States is a dynamic, rapidly changing population, and this will continue to occur as the population size grows (20). Hispanics/Latinos are the largest and fastest growing minority group, and are projected to comprise over 25% of the US population by 2060. They are a genetically and phenotypically diverse population as a result of extensive admixture between Native Americans and immigrants from multiple geographic locations around the world. In this study, we identified additional population substructure complexities that may contribute to phenotypic variation within Hispanics/Latinos.

Specifically, we demonstrated how the admixture dynamics of Mexican Americans have changed over time, resulting in an increase of ~20% Native American ancestry on average over the 50-year period studied. This change in ancestry is equivalent to a mean increase in Native American ancestry of ~0.4% per year. While the effect sizes vary to some extent, we replicate the underlying pattern across multiple data stratifications (two metropolitan cities, US born and non-US born) and also replicate this feature in an independent cohort of Mexican Americans. Further, we find that a similar trend holds across multiple self-identified Hispanic/Latino populations in the US (and is statistically significant in Central Americans). This effect does not appear to have a simple explanation: we do not see any statistically significant increases at individual loci, we do not see more than a negligible degree of population differentiation over time, and this increase cannot be entirely explained by very recent migration. We do, however, find that as Native American ancestry has increased, genetic diversity within Native American ancestry tracts across Mexican Americans has decreased over time, and is associated with increased short and medium length ROH over time. This suggested that there could be increased relatedness within Native American ancestries within Mexican Americans, and we confirmed that there is a very high degree of ancestry-based assortative mating within the Mexican American population.

What could be driving the increased Native American ancestry in Mexican Americans (12)? Population genetic theory suggests that while assortative mating could result in increased

304 ROH and decreased genetic diversity, ancestry-based assortative mating alone should not result
305 in mean changes in global ancestry proportions. Regardless of the underlying mechanisms driving
306 increased Native American ancestry in Mexican Americans, this additional source of temporal
307 substructure within this population has substantial consequences for phenotypic variation in
308 biomedical traits. We identify several biomedical traits that are correlated with Native American
309 ancestry, and show that in the case of height, there are both ancestry and temporal effects.
310 Further study is necessary to understand whether other biomedical traits are also changing over
311 time as the genomic ancestry proportions change in this population.

312 In our study, we bring specific attention to the biases that continue to exist with using
313 European GWAS summary statistics to calculate polygenic risk scores in admixed populations
314 such as Mexican Americans that are comprised of European, Native American, and African
315 genetic ancestries. In particular, in the case of height, we found that the polygenic height score
316 (PHS) correlated with observed height only in the subset of individuals with the lowest levels of
317 Native American ancestry (i.e. the subset of individuals with highest European ancestry). As the
318 population dynamics of the US continue to change, it is imperative that we study diverse
319 populations, or we risk exacerbating the health disparities that currently exist. To date, population-
320 based medical genomics research (and its subsequent benefits) have been disproportionately
321 focused on populations of European ancestry. In order to improve the design and implementation
322 of medical genetics studies for the ethnically diverse U.S. population, we need detailed insights
323 into the population history of diverse U.S. populations. This includes characterizing the admixture
324 dynamics of Hispanic/Latino populations, as well as the evolutionary forces that shaped patterns
325 of genetic variation of the ancestral populations that contributed to modern day Hispanic/Latino
326 populations.

327 The genetic variation of the Hispanic community in the United States belies
328 categorization under a single label (10). The events that have shaped and continue to shape
329 this genetic diversity are complex, numerous, and nuanced, and the social history of such a

330 diverse population is intrinsic to any genetic study. Mexico's society was largely defined by an
331 established social caste system based on ancestry, which disappeared after Mexico's
332 independence in 1821 (33). Even so, social inequalities persist today with skin colour having a
333 significant effect on wealth and education (34). A multitude of factors within and outside Mexico
334 — whether related to trade, immigration policies, or armed conflicts — acted to influence who
335 immigrated to the United States, and the impact of each of these fluctuates over time (35-37).
336 These changes shift the demographics of immigration, which is inherently related to the genetic
337 ancestry of the population.

338 Consequently, this shapes the genetic architecture of complex traits. Diverse
339 populations are at risk not only from underrepresentation in research, but because of poor
340 understanding of the temporal and spatial dynamics at play in genetic variation. The promise of
341 equitable precision medicine — one of the ultimate goals of medical genomics — cannot be
342 kept without understanding this interplay. Health disparities in the United States are fed by
343 structural inequalities. For example, studies that use modern Artificial Intelligence (AI)
344 techniques have already been shown to inflate existing disparities between Black Americans
345 and White Americans (38). Such biases, whether from algorithms, study designs, or
346 misunderstandings of subtleties in data, feed into the larger systemic pressures faced by
347 minority populations in the United States.

348 While we have shown a dramatic shift in ancestry proportions in US Hispanic/Latinos, one
349 of the caveats of this study is that the HCHS/SOL cohort is not representative of all US
350 Hispanics/Latinos. HCHS/SOL participants were recruited at four primary centers: Bronx,
351 Chicago, Miami, and San Diego. There may be additional genetic diversity that has not been
352 captured by this dataset and trends exhibited in this dataset may not translate to Hispanic/Latino
353 populations living in other regions of the US (though the temporal increase in Native American
354 ancestry was replicated in an independent sample of Mexican Americans). With better population
355 genetic modeling and a deeper understanding of the social and historical aspects of

356 Hispanic/Latino populations, we will be able to improve our understanding of the genetic and
357 phenotypic diversity across these populations, and subsequently improve our ability to understand
358 genetic contributions to complex traits and disease. These insights will lead to optimization of
359 population sampling for the design of future medical genetic studies, the identification of disease
360 risk variants, and ultimately, precision medicine for all.
361

362 **Methods**

363 **Study dataset and initial quality control**

364 The HCHS/SOL study is a community-based cohort study of self-identified Hispanic/Latino
365 individuals from four US metropolitan areas with the general goal of identifying risk and protective
366 factors for various medical conditions including cardiovascular disease, diabetes, pulmonary
367 disease, and sleep disorders (25). 12,434 participants with birth year estimates between 1934-
368 1993 who self-identified as being of Cuban, Dominican, Puerto Rican, Mexican, Central American,
369 or South American background consented to genetics studies and posting of their genetic and
370 phenotype data on the publicly available Database of Genotypes and Phenotypes (dbGaP)
371 through Study Accession phs000810.v1.p1. Samples were genotyped on an Illumina custom
372 array, SoL HCHS Custom 15041502 array (annotation B3, genome build 37), consisting of the
373 Illumina Omni 2.5M array and 148,353 custom single nucleotide polymorphisms (SNPs) (10).
374 Data posted to dbGaP had passed initial sample quality control filters, including removing samples
375 with differences in reported vs. genetic sex, call rates > 95%, and evidence for sample
376 contamination (e.g. heterozygosity and sample call rates). For initial SNP quality control, we
377 filtered out SNPs that were monomorphic, positional duplicates, or Illumina technical failures, as
378 well as SNPs that had cluster separation <= 0.3, call rate <=2%, >2 disconcordant calls in 291
379 duplicate samples, >3 Mendelian errors in parent-offspring pairs/trios, Hardy-Weinberg
380 Equilibrium combined P-value $<10^{-5}$, and sex differences in allele frequency ≥0.2. Our filtering
381 resulted in 1,763,935 genotyped SNPs with minor allele frequency (MAF) >0.01.

382 Additional sample quality control performed in the HCHS/SOL dataset included filtering
383 out samples with 1) large chromosomal anomalies, 2) substantial Asian ancestry as previously
384 identified in HCHS/SOL (*12*) and 3) individuals with up to third degree genetic relatedness in the
385 dataset as inferred by REAP (39). For genetic relatedness filtering, individuals from pairs were
386 kept to maximize representation of the birth year distribution, which resulted in 10,268 unrelated
387 remaining individuals.
388 From the original HCHS/SOL analysis, individuals were classified into genetic-analysis
389 groups, similar to self-identified background groups in that they share cultural and environmental
390 characteristics, but are also more genetically homogenous (10).
391 Birth year for all individuals was estimated by subtracting the difference between date of
392 first clinic visit for the baseline examination (25) and age. Year of arrival was estimated by
393 subtracting the difference between date of first clinic visit for the baseline examination and years
394 in the US.
395
396 **Global, local, and parental ancestry inference**
397 All ancestry analyses were restricted to the 211,152 autosomal SNP markers that overlapped
398 between the study and reference panel genotyping array. For the HCHS/SOL dataset, global
399 African, European, and Native American ancestries were inferred with ADMIXTURE, in an
400 unsupervised manner, with K=3. For some analyses, HCHS/SOL individuals with greater than
401 95% of either African, European, or Native American ancestry were filtered out resulting in 9,913
402 individuals: 1,099 Central American, 1,536 Cuban, 954 Dominican, 3,622 Mexican, 1,783 Puerto
403 Rican, 652 South American and 267 “Other” individuals.
404 Ancestral tracts, known as ‘local’ ancestry, along the genome for all HCHS/SOL
405 individuals were inferred using RFMix (40) and a three population reference panel, comprised of
406 315 individuals: 104 HapMap phase 3 CEU (European) and 107 YRI (African) individuals (41)
407 and 112 Native American individuals from throughout Latin America (8). The reference panel was

limited to individuals with 99% continental ancestry as inferred by unsupervised ADMIXTURE (42). Prior to local ancestry inference, HCHS/SOL individuals were merged with the reference panels and then phased using SHAPEIT2 (43). For all HCHS/SOL Mexican American individuals, parental genomic ancestry was inferred with ANCESTOR (44) using the local ancestry estimates generated by RFMix.

Uniform Manifold Approximation and Projection (UMAP)

Principal components for HCHS/SOL and the reference panel were computed using smartPCA (45). UMAP (version 0.3.8) was run using the Python script freely available at https://github.com/diazale/gt-dimred with parameter specification set at 15 nearest neighbours and a minimum distance between points of 0.5.

Admixture mapping

Local ancestry estimates for 211,151 SNPs across the genome were used to perform admixture mapping in HCHS/SOL Mexican Americans to determine if younger individuals harbored excess Native American ancestry in certain regions of the genome. Admixture mapping was performed applying two different models: 1) a linear regression model with age as the dependent variable adjusting for global Native American ancestry, sampling weight and center and 2) a logistic regression model dividing the HCHS/SOL Mexican cohort in to an older vs younger generation with 1965 set as the dividing point while also adjusting for global Native American ancestry, sampling weight, and center. The threshold for genome-wide significance, 1.38×10^{-4} was calculated using the empirical autoregression framework with the package *coda* in R to estimate the total number of ancestral blocks (46, 47).

433 **Tract Lengths**

434 The multiple regression model: $\log(f) = \beta_0 + \beta_1\ T + \beta_2\ A + \beta_3\ TA + \varepsilon$, where f is a matrix
435 containing the proportion of lengths of all ancestral tracts across the genome for all 3622 Mexican
436 American individuals, T the tract length bin and A decade of birth year bin, was used to test for an
437 effect of birth decade on the proportion of Native American ancestral tract lengths. For
438 assessment between the fraction of ancestry tracts in an individual's genome and birth year, long
439 tract cutoffs were chosen based on tract separation between the birth year decades in Figure 2B.
440

441 **Diversity Calculations**

442 Subcontinental ancestry was assessed using the diversity measurements π and F_{ST}. π was
443 calculated as the average number of pairwise genetic differences among all pairs of overlapping
444 Native American ancestry tracts across individuals. F_{ST} was calculated as:

445 $$F_{ST} = (H_T - H_S) / H_T$$

446 where H_T is the average heterozygosity when all individuals are pooled across decades and H_S is
447 the average heterozygosity within each decade of individuals.
448

449 **Inference of Runs of Homozygosity**

450 Runs of homozygosity (ROH) were called using the program GARLIC v1.1.4 (48) on 211,152
451 sites for the Mexican American individuals. An analysis window size of 50 SNPs and an
452 overlap fraction of 0.25 were both chosen using GARLIC's rule of thumb parameter
453 estimation. GARLIC chose a LOD score cutoff of 0. Using a three-component Gaussian
454 mixture, GARLIC determined class A/B (short/medium) and class B/C (medium/long) size
455 boundaries as 845,097 bp and 2,501,750 bp, respectively.
456
457

458 **Imputation**

459 Imputation for HCHS/SOL was performed locally using IMPUTE2 with the 1000 Genomes Project
460 Phase 3 haplotypes used as a reference panel. After filtering on an info score cutoff of 0.3, this
461 resulted in 33,041,084 SNPs.

462

463 **Polygenic Risk Score Calculations**

464 Polygenic risk scores for height were calculated using the publicly available UK Biobank
465 (UKBB) GWAS Round 2 Summary Statistics retrieved from http://www.nealelab.is/uk-biobank.
466 Briefly, for sample quality control, sample inclusion was limited to unrelated samples who passed
467 the sex chromosome aneuploidy filter. British ancestry was determined using the 1st 6 PCs;
468 individuals more than 7 standard deviations away from the 1st 6 PCs were excluded. Further
469 filtering included limiting to self -reported 'white-British' / 'Irish' / 'White' resulting in a QCed sample
470 count of 361,194 individuals https://github.com/Nealelab/UK_Biobank_GWAS#imputed-v3-
471 sample-qc. An imputation panel of ~90 million SNPs from HRC, UK10K and 1KG were used to
472 impute genotypes. 13.7 million autosomal and X-chromosome SNPs passed quality control
473 thresholds including Info score>0.8, MAF>0.0001, and HWE p-value>1e-10. For the phenotype,
474 a linear regression model in Hail was run for all individuals (both sexes) adjusting by the first 20
475 PCs + sex + age + age^2 + (sex*age) + (sex*age)^2. For height, there was complete phenotype
476 information for 360,388 individuals.

477 Risk scores were calculated by extracting the overlapping genome-wide significant hits
478 initially discovered in the UKBB GWASs of height and selecting SNPs with the lowest p-value in
479 each 1Mb window across the genome. For height this resulted in a dataset of 1,103 overlapping
480 SNPs that were present in our dataset of genotyped and imputed SNPs.

481

482

483 **Health and Retirement Study (HRS)**

484 For replication, we used genotype data from 705 self-identified Mexican-Americans from the
485 Health and Retirement Study (HRS) (49), genotyped on the Illumina Human Omni 2.5M platform.
486 HRS data was made available under IRB Study No. A11-E91-13B - The apportionment of genetic
487 diversity within the United States. Estimated global ancestry proportions for the Mexican American
488 population in the HRS were calculated as in Baharian et al. (12), which used an alternative
489 reference panel and alternative ancestry inference approach. Briefly, RFMix was used to infer
490 local ancestry estimates across the genome utilizing CHS, YRI, and CEU individuals from the
491 1000 Genomes Project as reference populations for Native American/Asian, African, and
492 European ancestries, respectively. Global ancestry estimates were calculated using the summed
493 RFMix calls.

494

495 **Statistical Analyses and Plots**

496 Statistical analyses and plot generation were performed within Rstudio using Version 1.1.463 and
497 R version 3.5.3. ternary and ggridges/ggplot2 packages were used to create the simplex and
498 ridgeline plots.

499

500 **Acknowledgements**

501 MLS was supported through the National Human Genome Research Institute (NHGRI) of the
502 National Institutes of Health (NIH) under Award Number F31HG010104. ADP and SG were
503 supported, in part, thanks to funding from the Canada Research Chairs program and CIHR
504 grant MOP-136855. RDH was supported, in part, by NHGRI grant R01HG007644 and the
505 Canadian Research Chairs program.

506 **Declaration of Interests**

507 The authors declare no competing interests.

508

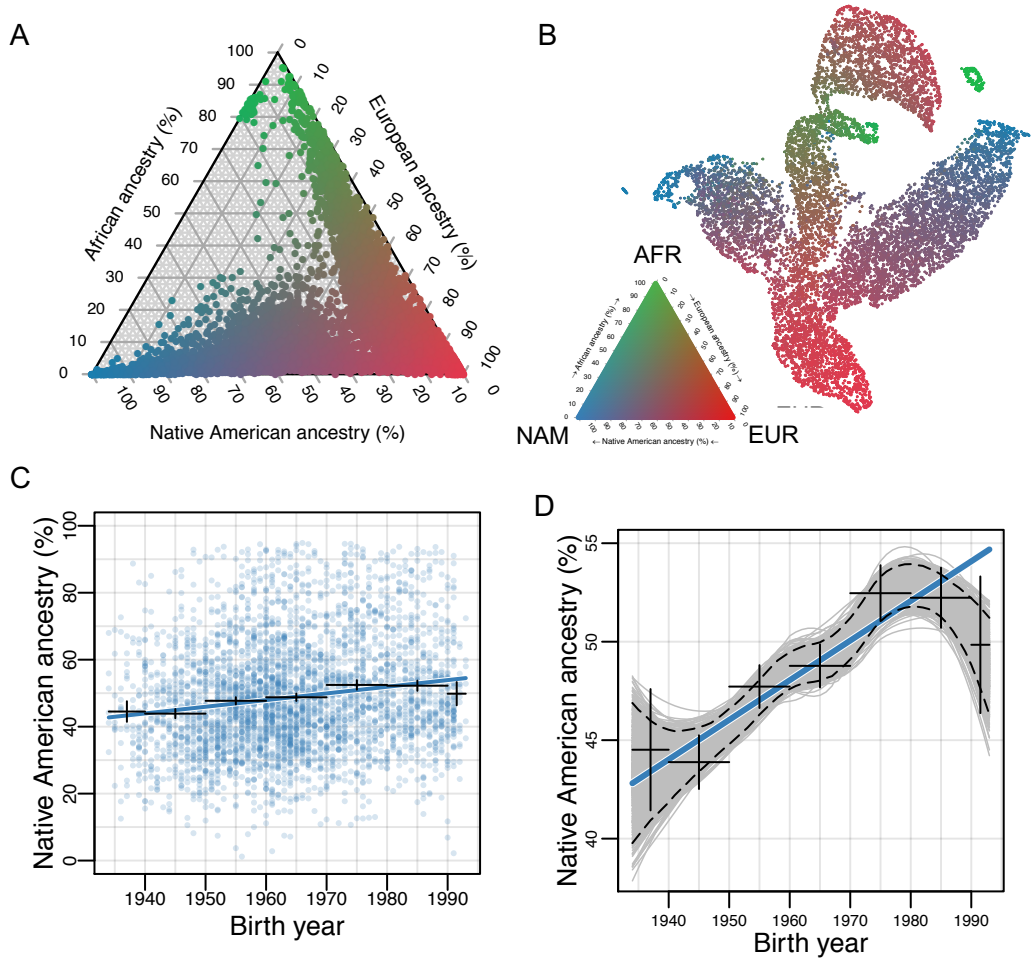


Figure 1. Recent dynamics continually shape the continuum of continental ancestry Hispanic/Latino populations. A. Ternary plot of HCHS/SOL (n=10,268) colored by admixture proportions. B. Uniform Manifold Approximation and Projection (UMAP) plot depicting the genetic diversity of HCHS/SOL and the reference panel (n=10,591) using 3 principal components, colored by admixture proportions (see Supplemental Fig 1 for population labels). Within the legend, AFR, EUR, and NAM refer to African, European, and Native American global ancestries, respectively. C. Global Native American ancestry proportions plotted by birth year for Mexican Americans (n=3,622). Fitted line is multiple regression of Native American ~ birth year + sampling weight. Bars represent 95% confidence intervals for individuals grouped by decade. D. Bootstrap resampling (n=1000 iterations) of Native American global ancestry for the Mexican American individuals with a fitted LOESS regression line for each iteration. Dashed lines represent the 95% confidence interval and the blue line represents the fitted regression line from Figure 1C.

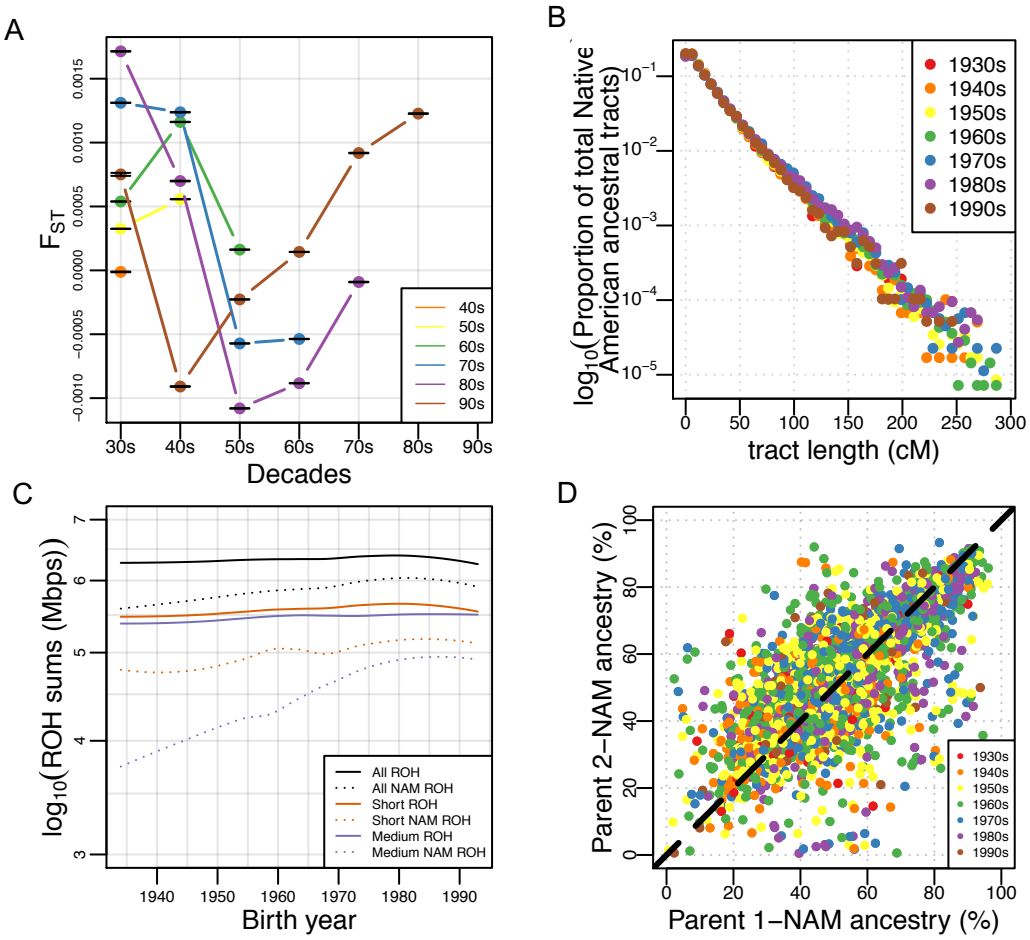


Figure 2. Diversity of and within Native American ancestral tracts. A) F_{ST} estimates calculated between each decade group. Bars represent the 95% CI. B) Proportion of total Native American ancestral tracts in the HCHS/SOL Mexican American population by decade. C) Loess regression of the log of the sum of total ROH and ROH overlapping Native American ancestral tracts separated by ROH class. Total long ROH is not represented as an individual group due to the high number of individuals missing long ROH (1694 for long ROH across ancestries and 1987 for long NAM ROH) but was included in the sum of "All ROH" and "All NAM ROH". D) Correlation of parent's inferred global Native American ancestries using ANCESTOR.

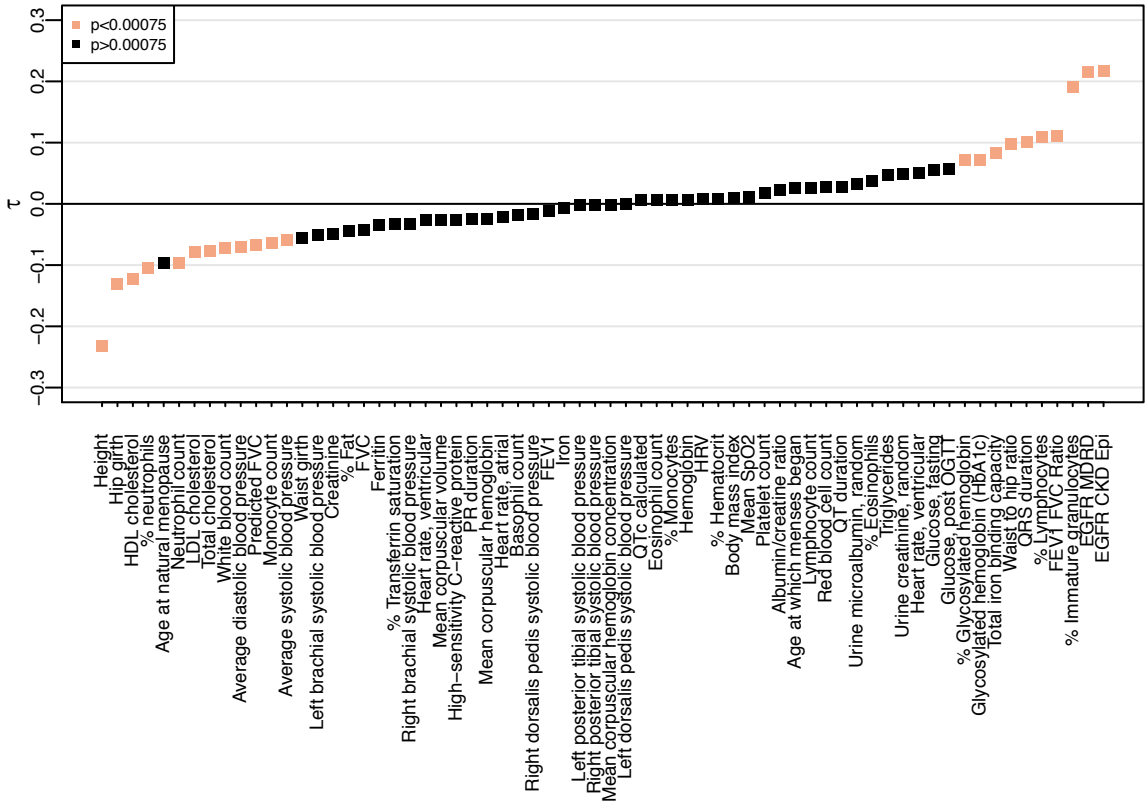


Figure 3. Correlation of 66 quantitative traits with global Native American ancestry. Significance level was determined using Bonferroni correction adjusting by the number of quantitative traits tested (0.05/66=0.00075).

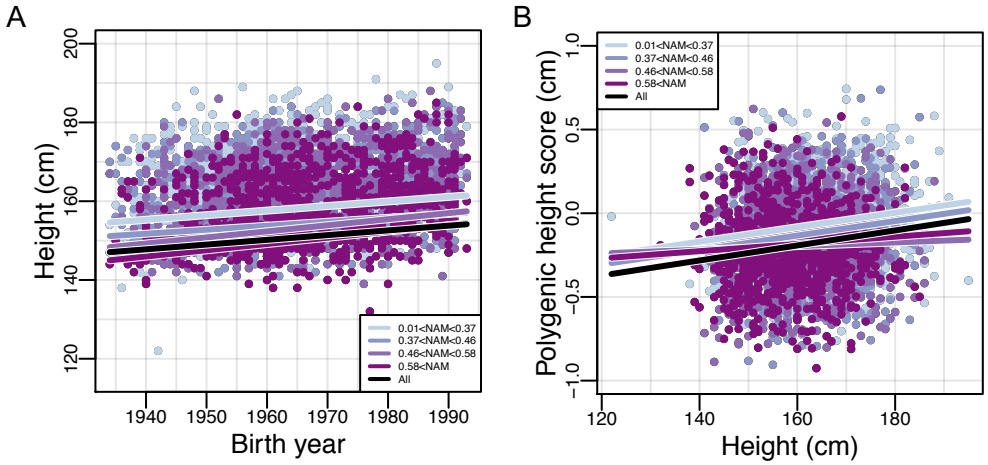
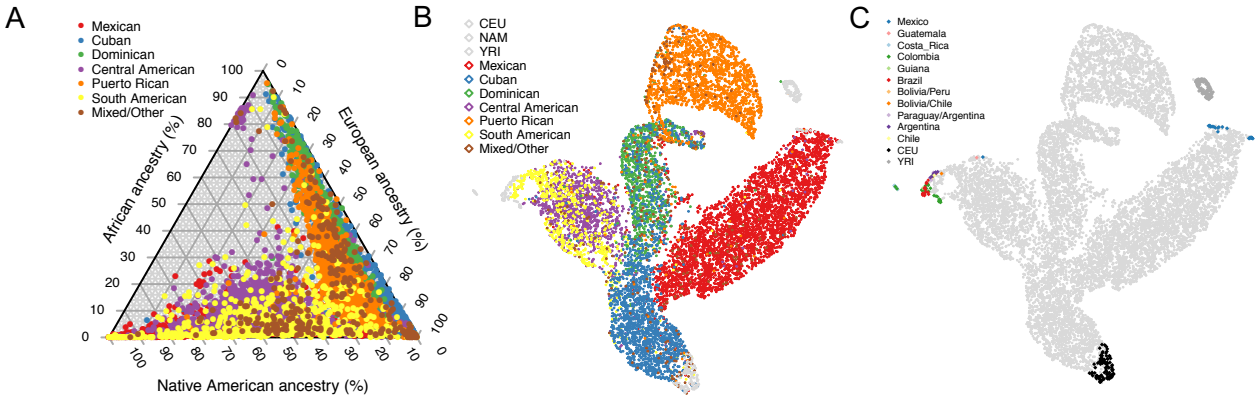


Figure 4: Height and global Native American ancestry in HCHS/SOL Mexican Americans. Each plot illustrates the relationship between A) Birth year and height B) Height and polygenic height score (PHS). The black line indicates the fitted linear model for all individuals. Each color represents a different quartile of Native American global ancestry. Polygenic height scores were assessed utilizing UKBB summary statistics for 1,128 SNPs.

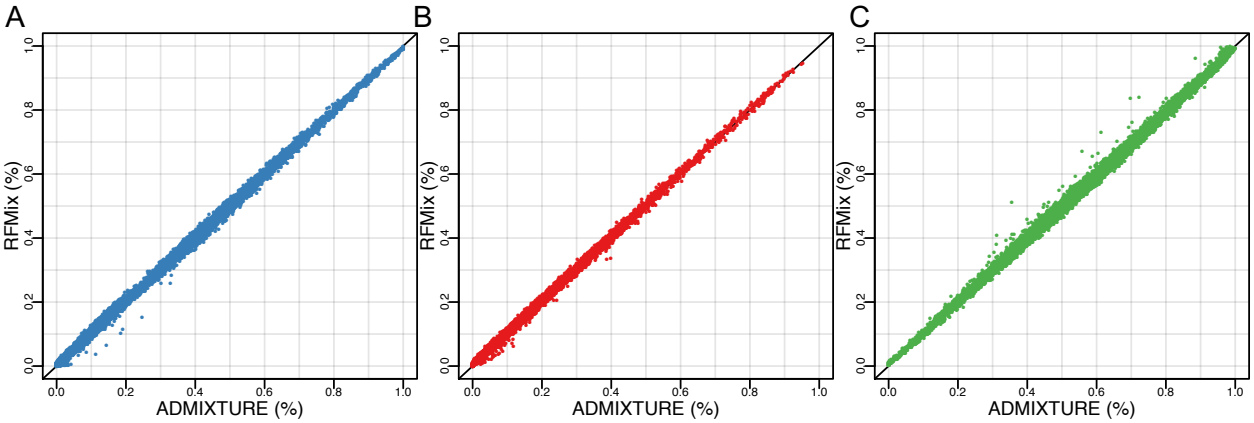
Table 1: Relationship of Native American global ancestry and birth year for Mexican Americans stratified by recruitment region, US born vs non-US born status, sex and educational attainment. For recruitment region, data stratification was limited to Chicago and San Diego as sample size for the Bronx and Miami was limited: 124 and 25 individuals, respectively. Education attainment was categorized as either less than a high school diploma or equivalent degree (<HS), equal to a high school diploma or equivalent degree (=HS), or post-secondary education (>HS).

Category	N	Mean	Median	R2	Effect	Std.Err	P
All	3622	0.489	0.468	0.027	0.0023	0.0002	3.58E-22
Chicago	1310	0.562	0.550	0.017	0.0016	0.0005	0.0006
San Diego	2163	0.428	0.422	0.012	0.0012	0.0002	4.29E-07
US born	634	0.427	0.418	0.063	0.0027	0.0004	1.77E-10
Not US born	2987	0.502	0.481	0.050	0.0032	0.0003	1.38E-30
Male	1500	0.494	0.475	0.038	0.0028	0.0004	3.83E-14
Female	2122	0.485	0.462	0.022	0.0019	0.0003	3.07E-10
<HS	1518	0.520	0.500	0.045	0.0026	0.0004	1.39E-12
=HS	960	0.501	0.479	0.022	0.0018	0.0005	0.0003
>HS	1140	0.436	0.422	0.045	0.0027	0.0004	6.53E-13

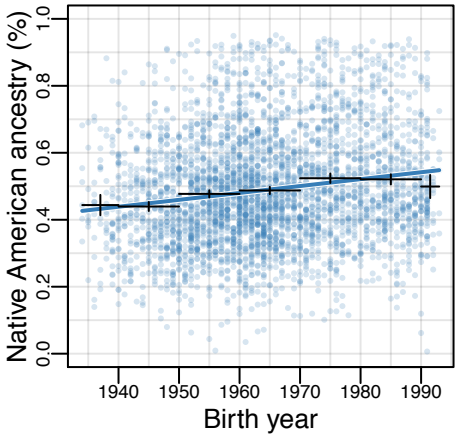
Supplementary Material



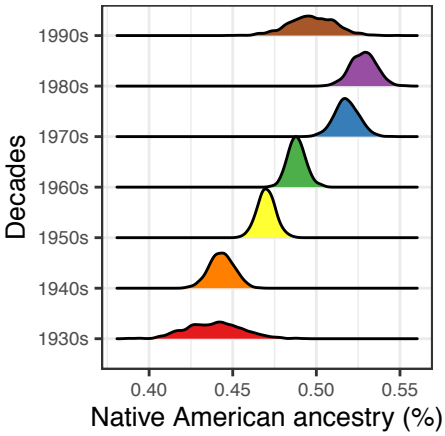
Supplementary Figure 1. Continental ancestral diversity of HCHS/SOL A) Ternary plot of global ancestry proportions colored by population for 10,268 HCHS/SOL individuals B) Uniform Manifold Approximation and Projection (UMAP) plot of HCHS/SOL and the reference panel (n=10,591) using 3 principal components, colored by HCHS/SOL population. C) UMAP plot of HCHS/SOL and the reference panel (n=10,591) using 3 principal components, colored by reference population.



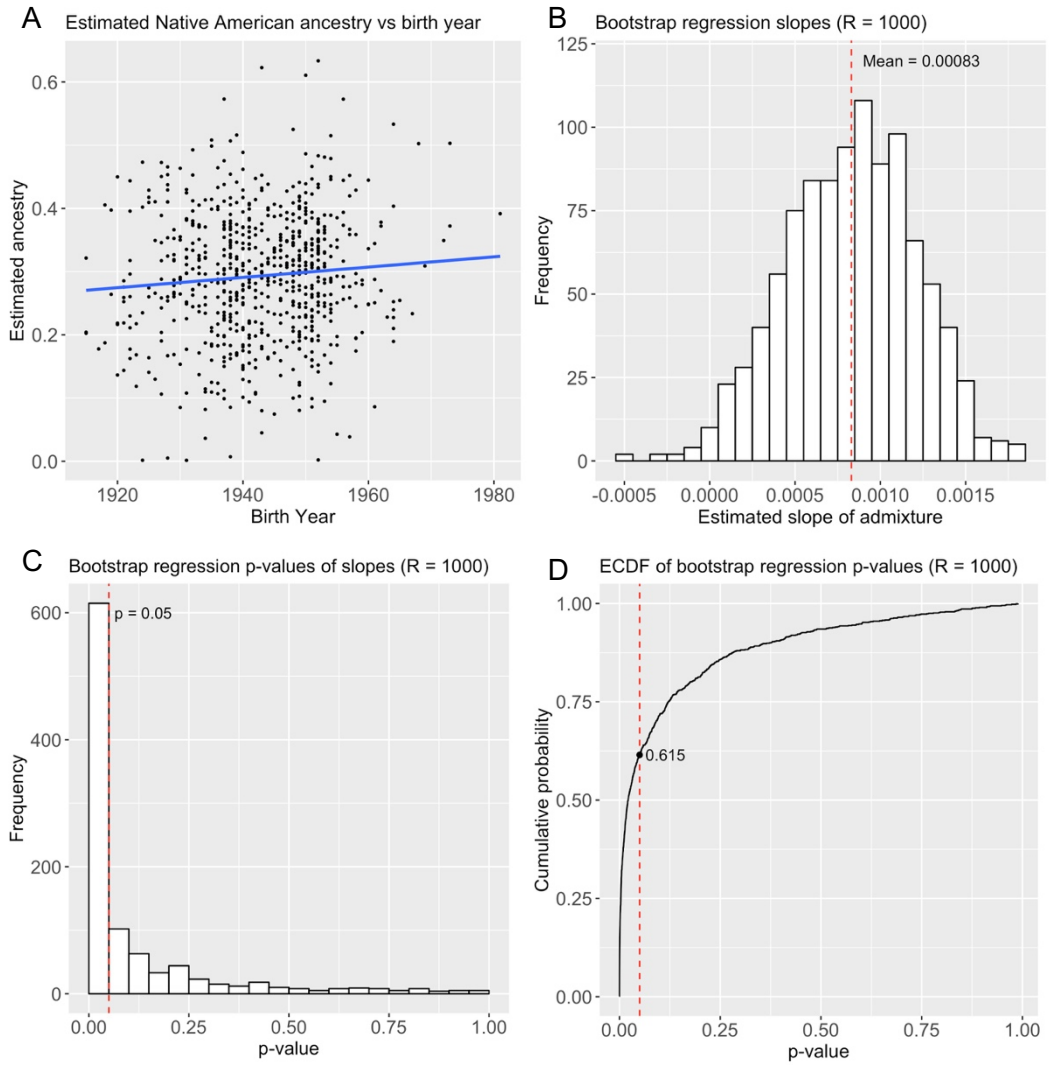
Supplementary Figure 2. Concordance of ADMIXTURE and RFMix global ancestry estimates. A) Native American ancestry B) African ancestry and C) European ancestry.



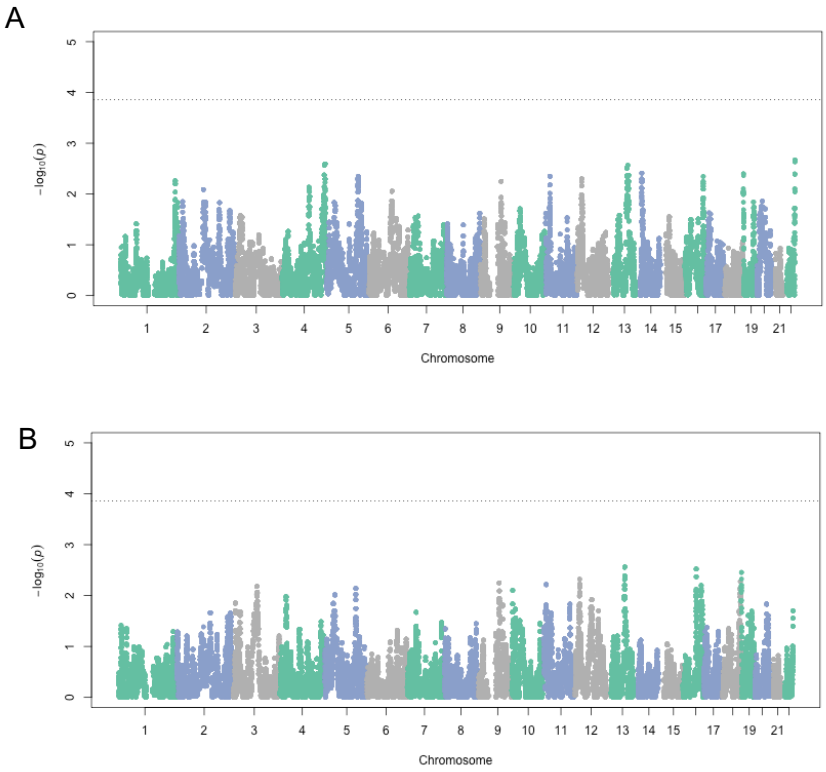
Supplementary Figure 3. RFMix inferred Native American global ancestry proportions plotted over time for HCHS/SOL Mexican Americans (n=3622).



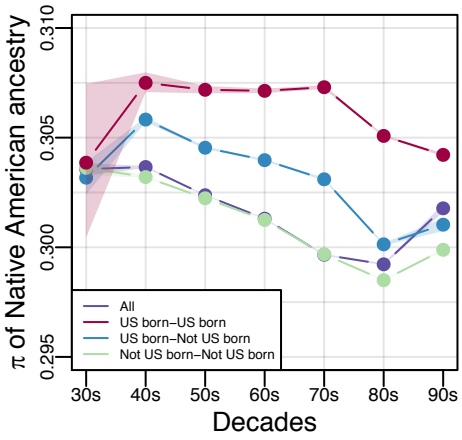
Supplementary Figure 4: Distributions of Native American global ancestry means for HCHS/SOL Mexican Americans (n=3622) generated by 1000 bootstrap resampling iterations within each decade of binned birth years.



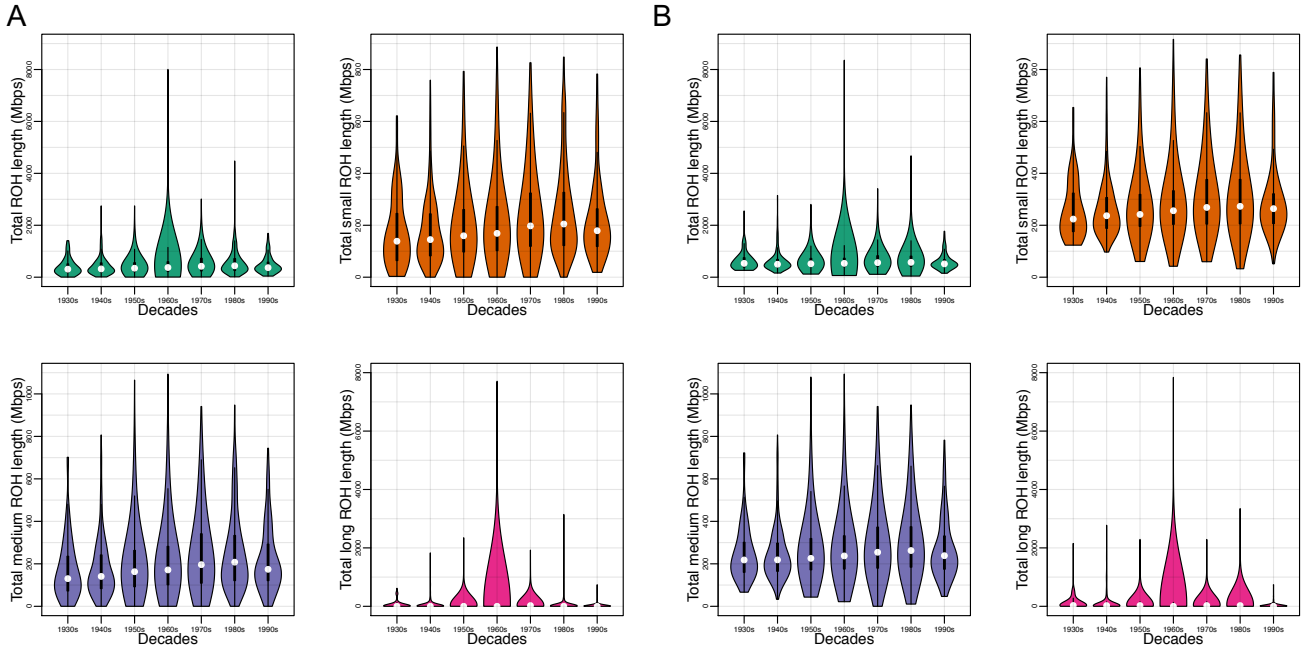
Supplementary Figure 5. Replication in the Health and Retirement Study for 705 self-identified Mexican Americans. A) Ancestry over time B) Distribution of regression slopes after 1000 bootstrap resampling iterations C) Distribution of bootstrap regression p-values D) ECDF of bootstrap regression p-values.



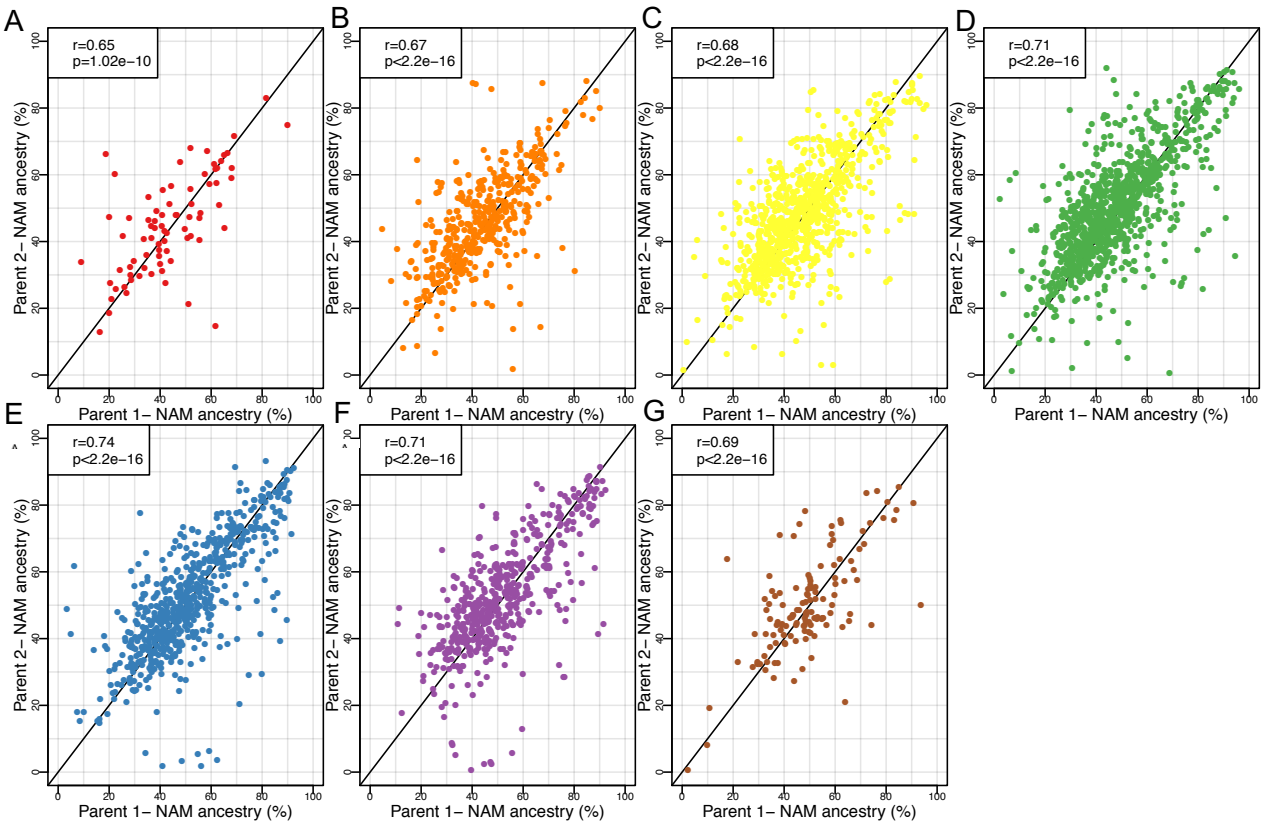
Supplementary Figure 6. Admixture mapping in HCHS/SOL Mexicans (n=3622) for Native American ancestry and A) birth year and B) generation. Ancestry association testing was performed at 211,151 markers using A) linear regression and B) logistic regression, both including global Native American ancestry, sampling weight and center as covariates.



Supplementary Figure 7. Diversity of and within Native American ancestral tracts. Diversity (π) of subcontinental Native American ancestry stratified by US born/not US born status. π was calculated between pairs within each decade of birth years. 95% confidence intervals are highlighted by the shaded regions for each group.



Supplementary Figure 8. Runs of homozygosity (ROH) in HCHS/SOL Mexican Americans. A) ROH across all ancestries separated by ROH class B) ROH overlapping Native American haplotypes separated by ROH class.



Supplementary Figure 9. Ancestry-related assortative mating in HCHS/SOL Mexican Americans separated by decade. Each plot represents the correlation of parent's inferred Native American ancestries using ANCESTOR by decade beginning with the 1930s (A) and ending with the 1990s (G). Each point corresponds to one Mexican American couple and the axes correspond to the inferred Native American ancestry of each partner.

Supplementary Table 1: Association of global ancestries and birth year for all HCHS/SOL individuals. For each population, we tested for an association between global ancestry and birth year while accounting for the sampling design. NAM, AFR, and EUR refer to Native American, African, and European ancestry respectively.

Population	Ancestry	N	R2	Effect	Std.Err	P
Central American	NAM	1099	0.0139	0.0013	0.0004	0.0013
Central American	AFR	1099	0.0136	0.0002	0.0004	0.6561
Central American	EUR	1099	0.0138	-0.0015	0.0004	0.0002
Cuban	NAM	1536	0.0023	0.0002	0.0001	0.0938
Cuban	AFR	1536	0.0014	-0.0005	0.0004	0.1490
Cuban	EUR	1536	0.0005	0.0003	0.0004	0.3879
Dominican	NAM	954	0.0035	0.0002	0.0001	0.0663
Dominican	AFR	954	0.0030	-0.0007	0.0004	0.1287
Dominican	EUR	954	0.0022	0.0005	0.0004	0.2374
Mexican	NAM	3622	0.0268	0.0023	0.0002	0.0000
Mexican	AFR	3622	0.0008	0.0000	0.0000	0.4189
Mexican	EUR	3622	0.0285	-0.0023	0.0002	0.0000
Puerto Rican	NAM	1783	0.0014	0.0001	0.0001	0.1533
Puerto Rican	AFR	1783	0.0014	0.0003	0.0002	0.1743
Puerto Rican	EUR	1783	0.0027	-0.0005	0.0002	0.0355
South American	NAM	652	0.0110	0.0016	0.0007	0.0211
South American	AFR	652	0.0027	-0.0002	0.0004	0.5053
South American	EUR	652	0.0080	-0.0014	0.0006	0.0335

791 **Supplementary Table 2 - Multiple regression table with traits that were significantly**

792 **correlated with global Native American ancestry.**

793

Trait	N	R2	Effect	Std.Err	P
Height	3615	0.588	-13.637	0.676	1.01E-85
Predicted FVC	3522	0.851	-695.376	36.867	1.13E-75
EGFR MDRD	3308	0.196	23.398	2.435	1.39E-21
EGFR CKD Epi	3308	0.441	14.752	1.581	1.90E-20
Waist to hip ratio	3617	0.195	0.043	0.007	6.96E-10
Glycosylated hemoglobin (HbA1c)	3609	0.081	9.184	1.590	8.30E-09
% Glycosylated hemoglobin	3609	0.080	0.837	0.146	9.88E-09
HDL cholesterol	3621	0.095	-6.740	1.372	9.39E-07
Total iron binding capacity	3620	0.066	23.540	5.467	1.71E-05
FEV1 FVC Ratio	3505	0.176	2.560	0.639	6.37E-05
% Lymphocytes	3442	0.021	3.768	1.023	2.35E-04
Hip girth	3617	0.077	-4.452	1.273	4.78E-04
% Neutrophils	3442	0.039	-3.649	1.178	1.96E-03
Monocyte count	3443	0.024	-0.048	0.019	1.27E-02
Neutrophil count	3442	0.043	-0.399	0.165	1.59E-02
LDL cholesterol	3529	0.047	-8.269	3.963	3.70E-02
Average diastolic blood pressure	3616	0.072	-2.067	1.136	6.88E-02
Total cholesterol	3622	0.065	-6.248	4.614	1.76E-01
White blood cell count	3442	0.032	-0.260	0.208	2.12E-01
Average systolic blood pressure	3619	0.211	1.590	1.712	3.53E-01
% Immature granulocytes	555	0.261	-0.090	0.110	4.14E-01
QRS duration	3596	0.168	0.062	1.268	9.61E-01

794

795

796

797

798

Supplementary Table 3: Height over time: Height (cm) as a function of birth year adjusting by sex, center, and sampling weight for 3614 Mexican Americans stratified by the quartiles of global Native American ancestry.

Group	N	R2	Effect	Std.Err	P
All	3614	0.54186	0.1200379	0.00905366	3.28E-39
NAM>0.58	929	0.5745454	0.158895	0.01783895	2.73E-18
0.46<=NAM<=0.58	955	0.5784636	0.1542762	0.01683919	3.07E-19
0.37<=NAM<0.46	842	0.5472275	0.1008054	0.01739819	9.73E-09
NAM<0.37	888	0.5369993	0.1161775	0.01815834	2.55E-10

Supplementary Table 4: Predicted height vs. observed height. Predicted height (cm) as a function of observed height (cm) adjusting by sex, center, and sampling weight for 3614 Mexican Americans.

Group	N	R2	Effect	Std.Err	P
All	3614	0.0249	0.0045	0.0006	2.19E-12
NAM>0.58	929	0.0072	0.0022	0.0012	7.79E-02
0.46<=NAM<=0.58	955	0.0058	0.0011	0.0013	3.90E-01
0.37<=NAM<0.46	842	0.0144	0.0043	0.0015	3.22E-03
NAM<0.37	888	0.0164	0.0043	0.0013	7.91E-04

815 **References:**

816 1. Nelson MR, Bryc K, King KS, Indap A, Boyko AR, Novembre J, et al. The Population
817 Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics
818 research. Am J Hum Genet. 2008;83(3):347-58.
819 2. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al.
820 An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-
821 65.
822 3. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A
823 global reference for human genetic variation. Nature. 2015;526(7571):68-74.
824 4. Gravel S, Zakharia F, Moreno-Estrada A, Byrnes JK, Muzzio M, Rodriguez-Flores JL, et
825 al. Reconstructing Native American migrations from whole-genome and whole-exome data.
826 PLoS Genet. 2013;9(12):e1004023.
827 5. Homburger JR, Moreno-Estrada A, Gignoux CR, Nelson D, Sanchez E, Ortiz-Tello P, et
828 al. Genomic Insights into the Ancestry and Demographic History of South America. PLoS
829 Genet. 2015;11(12):e1005602.
830 6. Moreno-Estrada A, Gignoux CR, Fernandez-Lopez JC, Zakharia F, Sikora M, Contreras
831 AV, et al. Human genetics. The genetics of Mexico recapitulates Native American substructure
832 and affects biomedical traits. Science. 2014;344(6189):1280-5.
833 7. Moreno-Estrada A, Gravel S, Zakharia F, McCauley JL, Byrnes JK, Gignoux CR, et al.
834 Reconstructing the population genetic history of the Caribbean. PLoS Genet.
835 2013;9(11):e1003925.
836 8. Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, Ray N, et al. Reconstructing
837 Native American population history. Nature. 2012;488(7411):370-4.
838 9. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of
839 African Americans, Latinos, and European Americans across the United States. Am J Hum
840 Genet. 2015;96(1):37-53.

841 10. Conomos MP, Laurie CA, Stilp AM, Gogarten SM, McHugh CP, Nelson SC, et al.
842 Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the
843 Hispanic Community Health Study/Study of Latinos. Am J Hum Genet. 2016;98(1):165-84.
844 11. Han E, Carbonetto P, Curtis RE, Wang Y, Granka JM, Byrnes J, et al. Clustering of
845 770,000 genomes reveals post-colonial population structure of North America. Nat Commun.
846 2017;8:14238.
847 12. Baharian S, Barakatt M, Gignoux CR, Shringarpure S, Errington J, Blot WJ, et al. The
848 Great Migration and African-American Genomic Diversity. PLoS Genet. 2016;12(5):e1006059.
849 13. Jordan IK, Rishishwar L, Conley AB. Native American admixture recapitulates
850 population-specific migration and settlement of the continental United States. PLoS Genet.
851 2019;15(9):e1008225.
852 14. Agarwala V, Flannick J, Sunyaev S, Go TDC, Altshuler D. Evaluating empirical bounds
853 on complex disease genetic architecture. Nat Genet. 2013;45(12):1418-27.
854 15. Eyre-Walker A. Evolution in health and medicine Sackler colloquium: Genetic
855 architecture of a complex trait and its implications for fitness and genome-wide association
856 studies. Proc Natl Acad Sci U S A. 2010;107 Suppl 1:1752-6.
857 16. Maher MC, Uricchio LH, Torgerson DG, Hernandez RD. Population genetics of rare
858 variants and complex diseases. Hum Hered. 2012;74(3-4):118-28.
859 17. Simons YB, Turchin MC, Pritchard JK, Sella G. The deleterious mutation load is
860 insensitive to recent population history. Nat Genet. 2014;46(3):220-4.
861 18. Uricchio LH, Zaitlen NA, Ye CJ, Witte JS, Hernandez RD. Selection and explosive
862 growth alter genetic architecture and hamper the detection of causal rare variants. Genome
863 Res. 2016;26(7):863-73.
864 19. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AA, Lee SH, et al. Genetic variance
865 estimation with imputed variants finds negligible missing heritability for human height and body
866 mass index. Nat Genet. 2015;47(10):1114-20.

867 20. Colby SLO, J.M. Projections of the Size and Compositon of the U.S. Population: 2014 to
868 2060. U.S. Census Bureau, Commerce USDo; 2015.
869 21. Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. Nature.
870 2011;475(7355):163-5.
871 22. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current
872 polygenic risk scores may exacerbate health disparities. Nat Genet. 2019;51(4):584-91.
873 23. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016;538(7624):161-
874 4.
875 24. Mills MC, Rahal C. A scientometric review of genome-wide association studies.
876 Commun Biol. 2019;2:9.
877 25. Sorlie PD, Aviles-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglus ML, Giachello
878 AL, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos.
879 Ann Epidemiol. 2010;20(8):629-41.
880 26. Collaboration NCDRF. A century of trends in adult human height. Elife. 2016;5.
881 27. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-
882 analysis of genome-wide association studies for height and body mass index in approximately
883 700000 individuals of European ancestry. Hum Mol Genet. 2018;27(20):3641-9.
884 28. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary
885 association statistics. Nat Rev Genet. 2017;18(2):117-27.
886 29. Sharp SA, Rich SS, Wood AR, Jones SE, Beaumont RN, Harrison JW, et al.
887 Development and Standardization of an Improved Type 1 Diabetes Genetic Risk Score for Use
888 in Newborn Screening and Incident Diagnosis. Diabetes Care. 2019;42(2):200-7.
889 30. Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al.
890 Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility
891 loci. Nat Genet. 2018;50(7):928-36.

892 31. Maas P, Barrdahl M, Joshi AD, Auer PL, Gaudet MM, Milne RL, et al. Breast Cancer
893 Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United
894 States. JAMA Oncol. 2016;2(10):1295-302.
895 32. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human
896 Demographic History Impacts Genetic Risk Prediction across Diverse Populations. Am J Hum
897 Genet. 2017;100(4):635-49.
898 33. Lisker R, Ramirez E, Briceno RP, Granados J, Babinsky V. Gene frequencies and
899 admixture estimates in four Mexican urban centers. Hum Biol. 1990;62(6):791-801.
900 34. Martinez DZ-CIF. Is Mexico a Post-Racial Country? Inequality and Skin Tone across the
901 Americas. AmericasBarometer Insights Series by the Latin American Public Opinion Project
902 (LAPOP), wwwLapopSurveysorg [Internet]. 2017. Available from:
903 https://www.vanderbilt.edu/lapop/insights/ITB031en.pdf.
904 35. Fernández-Kelly P, Massey DS. Borders for Whom? The Role of NAFTA in Mexico-U.S.
905 Migration. The Annals of the American Academy of Political and Social Science. 2007;610:98-
906 118.
907 36. Verea M. Immigration Trends After 20 Years of NAFTA. 2014. 2014;9(2).
908 37. Contreras VR. THE ROLE OF DRUG-RELATED VIOLENCE AND EXTORTION IN
909 PROMOTING MEXICAN MIGRATION: Unexpected Consequences of a Drug War. Latin
910 American Research Review. 2014;49(3):199-217.
911 38. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm
912 used to manage the health of populations. Science. 2019;366(6464):447-53.
913 39. Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N. Estimating
914 kinship in admixed populations. Am J Hum Genet. 2012;91(1):122-38.
915 40. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling
916 approach for rapid and robust local-ancestry inference. Am J Hum Genet. 2013;93(2):278-88.

917 41. International HapMap C, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA,
918 et al. Integrating common and rare genetic variation in diverse human populations. Nature.
919 2010;467(7311):52-8.
920 42. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in
921 unrelated individuals. Genome Res. 2009;19(9):1655-64.
922 43. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease
923 and population genetic studies. Nat Methods. 2013;10(1):5-6.
924 44. Zou JY, Halperin E, Burchard E, Sankararaman S. Inferring parental genomic ancestries
925 using pooled semi-Markov processes. Bioinformatics. 2015;31(12):i190-6.
926 45. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet.
927 2006;2(12):e190.
928 46. PLUMMER M BN, COWLES K, VINES K. CODA: Convergence Diagnosis and Output
929 Analysis for MCMC. R News. 2012(6):7-11.
930 47. Sobota RS, Shriner D, Kodaman N, Goodloe R, Zheng W, Gao YT, et al. Addressing
931 population-specific multiple testing burdens in genetic association studies. Ann Hum Genet.
932 2015;79(2):136-47.
933 48. Szpiech ZA, Blant A, Pemberton TJ. GARLIC: Genomic Autozygosity Regions
934 Likelihood-based Inference and Classification. Bioinformatics. 2017;33(13):2059-62.
935 49. Fisher GG, Ryan LH. Overview of the Health and Retirement Study and Introduction to
936 the Special Issue. Work Aging Retire. 2018;4(1):1-9.
937