

1 Impact of cross-ancestry genetic architecture on

2 GWAS in admixed populations

3

4 Rachel Mester,^{1,*} Kangcheng Hou,² Yi Ding,² Gillian Meeks,³ Kathryn S. Burch,² Arjun

5 Bhattacharya,⁴ Brenna M. Henn,⁵ Bogdan Pasaniuc^{1,2,4,6,7,*}

6

7 ¹Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los

8 Angeles, CA, 90095 USA.

9 ²Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, 90095 USA.

10 ³Integrative Genetics and Genomics Graduate Group, University of California, Davis, Davis, CA, 95616 USA.

11 ⁴Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los

12 Angeles, Los Angeles, CA, 90095 USA.

13 ⁵Department of Anthropology, Center for Population Biology and the Genome Center, University of California, Davis,

14 Davis, CA, 95616 USA.

15 ⁶Institute of Precision Health, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA,

16 90095 USA.

17 ⁷Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles,

18 CA, 90095 USA.

19 *Correspondence: rmaster@ucla.edu, pasaniuc@ucla.edu

20

21 Abstract

22 Genome-wide association studies (GWAS) have identified thousands of variants for disease risk.

23 These studies have predominantly been conducted in individuals of European ancestries, which

24 raises questions about their transferability to individuals of other ancestries. Of particular interest

25 are admixed populations, usually defined as populations with recent ancestry from two or more

26 continental sources. Admixed genomes contain segments of distinct ancestries that vary in

27 composition across individuals in the population, allowing for the same allele to induce risk for
28 disease on different ancestral backgrounds. This mosaicism raises unique challenges for GWAS
29 in admixed populations, such as the need to correctly adjust for population stratification to balance
30 type I error with statistical power. In this work we quantify the impact of differences in estimated
31 allelic effect sizes for risk variants between ancestry backgrounds on association statistics.
32 Specifically, while the possibility of estimated allelic effect-size heterogeneity by ancestry
33 (HetLanc) can be modeled when performing GWAS in admixed populations, the extent of HetLanc
34 needed to overcome the penalty from an additional degree of freedom in the association statistic
35 has not been thoroughly quantified. Using extensive simulations of admixed genotypes and
36 phenotypes we find that modeling HetLanc in its absence reduces statistical power by up to
37 72%. This finding is especially pronounced in the presence of allele frequency differentiation. We
38 replicate simulation results using 4,327 African-European admixed genomes from the UK Biobank
39 for 12 traits to find that for most significant SNPs HetLanc is not large enough for GWAS to benefit
40 from modeling heterogeneity.

41

42 **Introduction**

43 The success of genomics in disease studies depends on our ability to incorporate diverse
44 populations into large-scale genome-wide association studies (GWAS)¹⁻⁴. Cohort and biobank
45 studies are growing to reflect this diversity⁵⁻⁷, and a variety of techniques exist which incorporate
46 populations of different continental ancestries into GWAS⁸. However, while admixture has been
47 an important factor in other steps in the disease mapping process, such as fine-mapping⁹ and
48 estimating heritability^{10,11}, individuals of mixed ancestries (admixed individuals) have largely been
49 left out of traditional association studies. GWAS performed in admixed populations have greater
50 power for discovery compared to similar sized GWAS in homogeneous populations^{12,13}. Thus,
51 excluding admixed individuals from association studies will not only increase health disparities,

52 but will also disadvantage other populations. To prevent this exclusion, approaches to association
53 studies have been developed specifically for admixed populations¹⁴⁻¹⁷. However, the impact of
54 HetLanc (differences in estimated allelic effect sizes for risk variants between ancestry
55 backgrounds) on GWAS methods remains underexplored. Of particular interest are recently
56 admixed populations, defined as less than 20 generations of mixture between two ancestrally
57 distinct populations. In such populations, the admixture process creates mosaic genomes
58 comprised of chromosomal segments originating from each of the ancestral populations (i.e., local
59 ancestry segments). Local ancestry segments are much larger than linkage disequilibrium (LD)
60 blocks¹⁸; thus, LD patterns within each local ancestry block of an admixed genome reflect LD
61 patterns of the ancestral population. Similarly, allele frequency estimates from segments of a
62 particular local ancestry are expected to reflect allele frequencies of the ancestral population.
63 Variation in local ancestry across the genome leads to variability in global ancestry (the average
64 of all local ancestries within a given individual). Such variability in local and global ancestries could
65 pose a problem to GWAS in admixed populations as genetic ancestries are often correlated with
66 socio-economic factors that also impact disease risk, thus yielding false positives in studies that
67 do not properly correct for genetic ancestries. Because local and global ancestry are only weakly
68 correlated¹⁹, complete control of confounding due to admixture requires conditioning on both local
69 and global ancestry²⁰. However, the success of admixture mapping indicates that the possibility
70 of losing power due to over-correction for local ancestry stratification is serious^{17,21}.

71
72 GWAS in admixed populations is typically performed either using a statistical test that ignores
73 local ancestry altogether (e.g., the Armitage trend test, ATT) or using a test that explicitly allows
74 for HetLanc (e.g., Tractor). The former provides superior power in the absence of HetLanc with
75 the latter having great potential for discovery in its presence. However, these methods' relative
76 statistical power for discovery depends on the cross-ancestry genetic architecture of the trait: i.e.,
77 which variants are causal and what are those variants' ancestry-specific frequencies, causal

78 effects, and linkage disequilibrium patterns. For example, existing studies have found that ATT
79 can yield a 25% increase in power over Tractor³ in the absence of HetLanc while Tractor has
80 higher power when causal effects are different by more than 60%¹⁵. However, the full impact of
81 cross-ancestry genetic architecture on GWAS power in admixed populations remains under-
82 explored.

83

84 In this work, we use simulations to perform a comprehensive evaluation quantifying the impact of
85 these factors on the power of GWAS approaches in admixed populations. We provide guidelines
86 for when to use each test as a function of cross-ancestry genetic architecture. Elements of cross-
87 ancestry genetic architecture such as allele frequencies, global ancestry ratios, and LD are known
88 or can be calculated in advance of a GWAS to determine which of our simulation results apply in
89 each case. Using extensive simulations, we find that ATT should be preferred when HetLanc is
90 small or non-existent. We quantify the extent of HetLanc and the ancestry-specific allele
91 frequency differences required for Tractor to overcome the extra degree of freedom penalty. We
92 further validate our results using the African-European admixed population in the UK Biobank
93 (UKBB). By examining the HetLanc of significant SNPs in the UKBB, we can understand how
94 often it rises to a level that impacts the power of traditional GWAS.

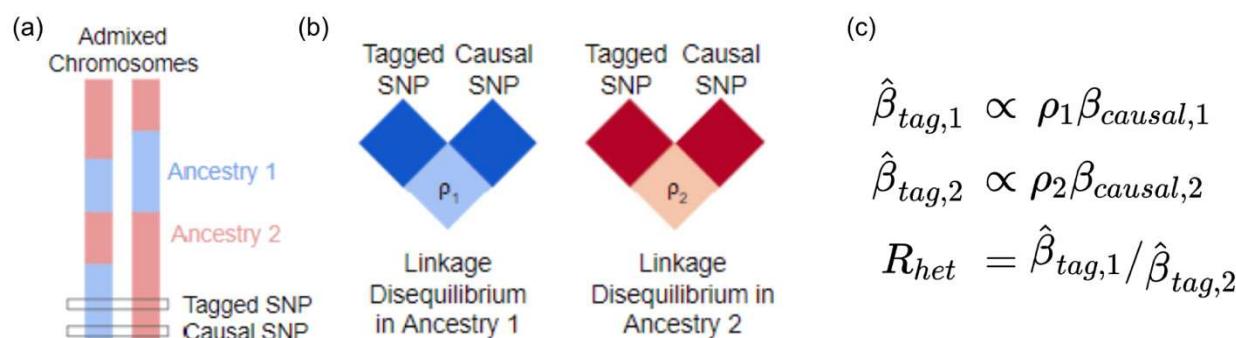
95

96 **Results**

97 **Heterogeneity by Local Ancestry Impacts Association Statistics in Admixed 98 Populations**

99 HetLanc occurs when a SNP exhibits different estimated allelic effect sizes depending on its local
100 ancestry background. HetLanc can manifest itself at causal SNPs due to genetic interactions
101 between multiple causal variants or differential environments, although recent work suggests that
102 the magnitude and frequency of these types of epistatic effects between causal variants is

103 limited²². A more common form of HetLanc is observed at non-causal SNPs that tag the causal
104 effect in a differential manner across ancestries. Differential linkage disequilibrium by local
105 ancestry at these non-causal SNPs (tagged SNPs) can cause HetLanc even when allele
106 frequencies and causal effect sizes are the same across ancestries. The extent to which HetLanc
107 exists and the magnitude of these differences in effect sizes are yet uncertain²²⁻³⁸. However, the
108 existence of HetLanc plays an important role in the power of GWAS methods to detect
109 associations. Consider the example in Figure 1 in which the allelic effect size for a tagged SNP is
110 estimated for a phenotype in an admixed population. In this population, both the tagged SNP and
111 the true causal SNP may exist in regions attributed to both local ancestries present in the
112 population (Figure 1a). Since LD patterns differ by local ancestry, the correlation between the
113 tagged and causal SNPs will also depend on local ancestry (Figure 1b). This differential
114 correlation between tagged and causal SNPs will cause the estimated allelic effect size for the
115 tagged SNP $\hat{\beta}_{tag,i}$ to depend on local ancestry i (Figure 1c). Thus, even for cases in which true
116 causal effect sizes are the same across ancestries, allelic effect sizes estimated for the tagged
117 SNP may be heterogeneous. Since GWAS cannot determine true causal effect sizes, we
118 introduce R_{het} , a measure of HetLanc which allows for both true causal effect-size heterogeneity
119 and LD- and allele frequency-induced estimated allelic effect-size heterogeneity.



120
121

122 **Figure 1: Toy example of how differential LD by local ancestry can induce HetLanc.** (a)
123 Admixed populations contain haplotypes with different local ancestry at the causal or tagged SNP.

124 (b) The correlation between tagged and causal SNPs depends on their local ancestry due to
125 differential LD by local ancestry. (c) In a GWAS, the estimated marginal SNP effect size is
126 proportional to the true causal effect size and the correlation between the tagged and causal
127 SNPs ($\hat{\beta}_{tag,i} \propto \rho_i \beta_{causal,i}$, where i refers to the i_{th} ancestry).

128

129 **Methods for association testing in admixed populations**

130 We start with a formal definition for a full model relating genotype, phenotype, and ancestry for a
131 single causal SNP:

132

133
$$y = \beta_1 g_1 + \beta_2 g_2 + e_l l + e_\alpha^T \alpha + \epsilon \quad (1)$$

134

135 where y is a phenotype, g_1 and g_2 are vectors that represent the number of alternate alleles with
136 local ancestry 1 and 2 (such that $g_1 + g_2 = g$, the genotype in standard form), β_1 and β_2 are
137 ancestry-specific marginal effect sizes of the SNP, l is the vector of local ancestry counts at the
138 locus, e_l is the effect size of l , α is a matrix of other covariates such as global ancestry, e_α^T is the
139 vector of the effect sizes of α , and ϵ is random environmental noise.

140

141 Variability across local and global ancestries has been leveraged in various statistical approaches
142 for disease mapping in admixed populations. One of the first methods developed for association
143 was admixture mapping (ADM)^{17,29}. ADM tests for association between local ancestry and disease
144 status in cases and controls or in a case-only fashion. This association is achieved by contrasting
145 local ancestry deviation with expectations from per-individual global ancestry proportions.
146 Therefore, ADM is often underpowered especially in situations in which allele frequency at the
147 causal variant is similar across ancestral populations³⁰. Genotype association testing is
148 traditionally performed using an Armitage trend test (ATT). ATT tests for association between

149 genotypes and disease status while correcting for global ancestry to account for stratification^{17,31}.
150 However, neither ADM nor ATT take advantage of the full disease association signal in admixed
151 individuals. SNP1, SUM, and MIX are examples of association tests that combine local ancestry
152 and genotype information. SNP1 regresses out local ancestry in addition to global ancestry to
153 control for fine-scale population structure. This approach helps control for fine-scale population
154 stratification but may remove the signal contained in local ancestry information³². SUM³³
155 combines the SNP1¹⁴ and ADM statistics into a 2 degree of freedom test. MIX¹⁴ is a case-control
156 test that incorporates SNP and local ancestry information into a single degree of freedom test.
157 Most recently Tractor¹⁵ conditions the effect size of each SNP on its local ancestry followed by a
158 joint test allowing for different effects on different ancestral backgrounds. This step builds the
159 possibility of HetLanc explicitly into the model, which may be particularly important when SNPs
160 are negatively correlated across ancestries³⁴. Other varieties of tests have also been developed
161 using different types of frameworks, most notably BMIX³⁴ which leverages a Bayesian approach
162 to reduce multiple testing burden. These statistics have been compared at length^{3,14,17,35}.
163 However, existing comparisons do not consider HetLanc, nor do they thoroughly discuss allele
164 frequency differences across ancestries.

165

Association Statistic	Statistical Test (H_0)	Assumptions on β	Covariates	Degrees of Freedom
ADM	$e_l = 0$	--	α	1
ATT	$\beta = 0$	$\beta = \beta_1 = \beta_2$	α	1
SNP1	$\beta = 0$	$\beta = \beta_1 = \beta_2$	l, α	1
MIX	$e_l \circ \beta = 0$	$\beta = \beta_1 = \beta_2$	α	1

SUM	$\beta = 0$ and $e_l = 0$	$\beta = \beta_1 = \beta_2$	l, α	2
Tractor	$\beta_1 = 0$ and $\beta_2 = 0$	--	l, α	2

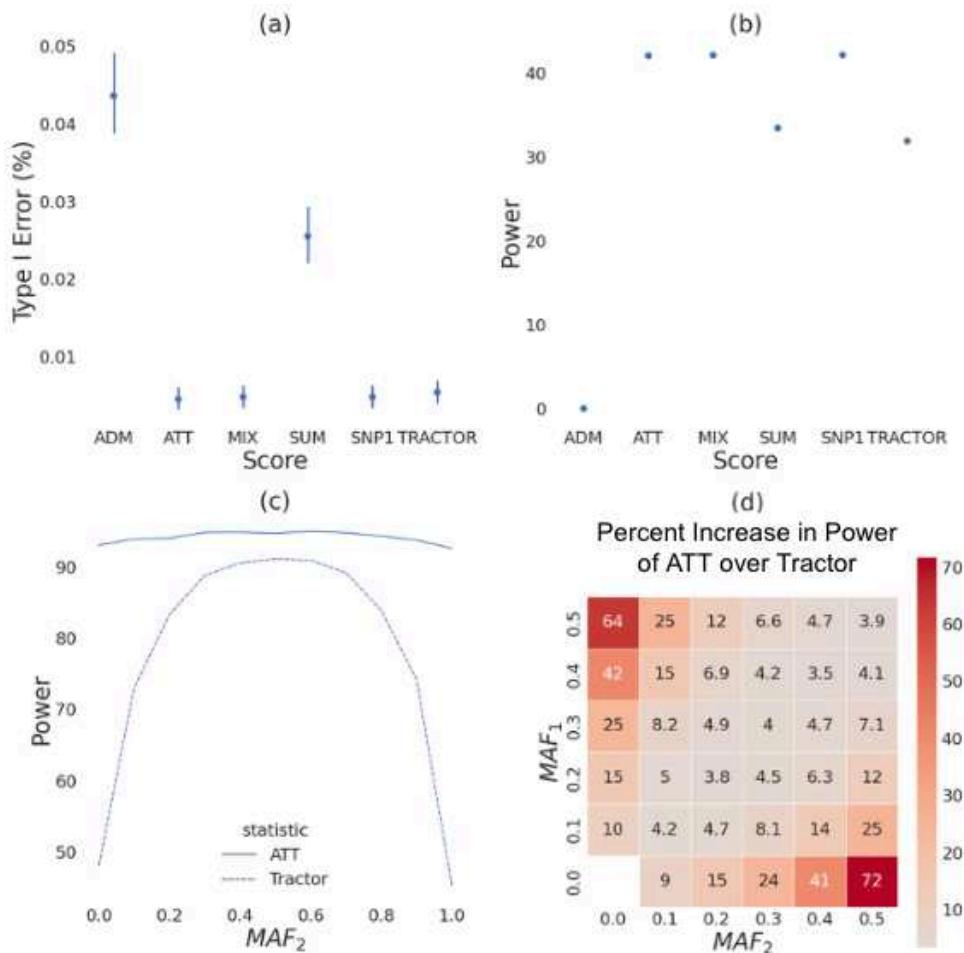
166

167 **Table 1: Summary of GWAS association statistics.** All tests adjust for global ancestry and can
168 be used on binary traits, and all tests except MIX can be implemented with adjustment for
169 additional covariates and use on quantitative traits. For more information on the comparison of
170 ATT, ADM, SUM, and MIX see³². We note that while additional methods exist³⁵⁻³⁹ we do not focus
171 on them in this work because they do not directly relate to equation 1.

172

173 **ATT has more power than Tractor in the absence of heterogeneity by ancestry**

174 First, we use simulations to compare type I error and power for each association statistic in Table
175 1. Starting with 10,000 simulated admixed individuals based on a 50/50 admixture proportion, we
176 simulate 1,000 case-control phenotypes with a single causal SNP (see Methods). We calculate
177 type I error as the probability of each method to detect significant associations in non-causal SNPs
178 (see Methods). Type I error is well controlled for every association test, well under the 5%
179 threshold expected by the chosen p-value (Figure 2a). The mean type I error was $\leq 4.36 \times 10^{-2}\%$
180 for every association test. The maximum value was $\leq 0.6\%$ for every association test. We next
181 calculate power to detect SNPs with an odds ratio of $OR_1 = OR_2 = 1.2$ (see Methods). We find
182 that SNP1 had the highest power at 42.14%. However, SNP1 was not significantly more powerful
183 than either MIX (power 42.12%, p-value 0.878) or ATT (power 42.05%, p-value 0.325, Figure 2b).
184 The power of all three of these tests was significantly higher (p-value $\leq 1 \times 10^{-16}$) than for SUM
185 (power = 33.44%), ADM (power = 0.039%), or Tractor (power = 31.89%). Thus, we find that while
186 these association statistics are all well controlled, power does substantially differ between them.
187 In the absence of both HetLanc and allele frequency difference, 1 degree of freedom SNP
188 association tests outperform 2 degree of freedom tests.



189

190 **Figure 2: Association statistics in the absence of HetLanc.** (a) Type I error for association
191 statistics. Type I error calculated as the probability for detecting significant association for a
192 null SNP. 95% confidence interval shown. (b) Power for association statistics. Power
193 calculated at odds ratios $OR_1 = OR_2 = 1.2$. 95% confidence interval too narrow for display. (c)
194 Power for ATT and Tractor as MAF_2 is varied between 0.0 and 1.0 and MAF_1 is fixed at 0.5.
195 Power for both methods varies as MAF difference varies. 95% confidence interval too narrow
196 for display. (d) Heatmap of percent increase in power of ATT over Tractor when $\beta_1 = \beta_2 = 1.0$.
197 Minor allele frequencies MAF_1 and MAF_2 varied from 0.0 to 0.5 in increments of 0.1. All
198 simulations are for a case-control (a-b) or quantitative (c-d) traits simulated 1,000 times for a
199 population of 10,000 individuals with global ancestry proportion 50/50. Power calculated using
200 a Bonferroni-corrected threshold of standard threshold p-value $< 1 \times 10^{-5}$ (a-b) or standard

201 threshold p-value $< 5 \times 10^{-8}$ (**c-d**). Case-control traits (**a-b**) have case-control ratio 1:1 and 10%
202 case prevalence, quantitative traits (**c-d**) have heritability $h^2 = 0.005$. Heritability, global
203 ancestry, causal effect size β and overall MAF do not qualitatively impact these results (Figures
204 S1, S2 and S3).

205
206 We next investigate how differences in minor allele frequency (MAF) impact the power of ATT
207 and Tractor in the case where true causal effect sizes are the same. We investigate the impact
208 of varying MAF in each ancestry independently. Using our 10,000 simulated admixed individuals
209 from the previous experiment, we simulate 1,000 quantitative phenotypes with a single causal
210 SNP (see Methods). First, we let $MAF_1 = 0.5$ and MAF_2 range from 0.0 to 1.0 with a 0.1 increment
211 and plot power over MAF_2 (Figure 2c). We find that ATT has higher power than Tractor at all levels
212 of MAF difference. Since Tractor has an extra degree of freedom compared to ATT, Tractor is
213 disadvantaged when $\beta_1 = \beta_2$. When $MAF_1 = MAF_2$, ATT has 94.7% power, with Tractor at 91.1%
214 power. However, as MAF_2 becomes more different from MAF_1 , ATT maintains its power at 93.0%.
215 By contrast, Tractor loses much of its power, with only 45.3% power when the causal allele is
216 fixed at 100% in population 2 and only 48.1% power when the causal allele is absent in population
217 2. ATT maintains higher power than Tractor even at varying levels of heritability (Figures S1, S2,
218 S3), MAF_1 (Figure S1), global ancestry (Figure S2), and effect size β (Figure S3). However, the
219 difference in power has a large range depending on the MAF difference between local ancestries.
220

221 Next we introduce percent difference in power, a one-dimensional metric to compare between
222 these association statistics (see Methods). We use this metric to visualize how varying MAF_1 and
223 MAF_2 independently impacts the power of ATT and Tractor (Figure 2d). The percent increase in
224 power when using ATT over Tractor when the causal SNP is absent in population 2 is 64%. The
225 power difference between ATT and Tractor increases as MAF difference increases. Furthermore,

226 the lower the MAF starts out in population 1, the larger the power difference between these two
227 statistics. Specifically, when $MAF_1 = 0.5$ and $MAF_2 = 0.1$, the difference in MAF is 0.4 and ATT
228 has a 25% power increase over Tractor. However, when $MAF_1 = 0.4$ and $MAF_2 = 0.0$, the
229 difference in MAF is still 0.4 but ATT has a 42% increase in power over Tractor.

230

231 While this result corroborates previous studies⁴⁰⁻⁴², the relationship between Tractor and
232 admixture mapping provides insight into the mechanism behind this dynamic. Mainly, as allele
233 frequency differentiation by local ancestry increases, so does the power of the admixture mapping
234 test statistic. In fact, ADM has no power when minor allele frequencies do not differ by ancestry
235 but achieves up to 6.7% power when $MAF_1 = 0.0$ and $MAF_2 = 0.5$ (Figure S4a). However, the
236 Tractor method uses the admixture mapping statistic as its null hypothesis. A stronger null
237 hypothesis will be rejected less often than a weaker one even when the alternative hypothesis is
238 the same, causing any test utilizing a strong null hypothesis to have less power. Thus, Tractor will
239 have less power when its null hypothesis (ADM) has more power, which occurs in situations with
240 high allele frequency differentiation. When allele frequencies do not differ by ancestry, Tractor
241 achieves 91% power in our simulations. However, when $MAF_1 = 0.0$ and $MAF_2 = 0.5$ Tractor
242 power plummets to 44% (Figure S4b).

243

244 While high levels of allele frequency differentiation drastically decrease the power of Tractor, ATT
245 also has a smaller decrease in power at high levels of allele frequency differentiation, from 95%
246 at equal allele frequencies to 93% when $MAF_1 = 0.0$ and $MAF_2 = 0.5$ (Figure S4c). This decrease
247 in power is not as large as that suffered by Tractor, but it is also due to increased power of the
248 null hypothesis at higher frequency differentiation across populations. The null hypothesis of the
249 ATT test statistic only includes global ancestry, but the power of global ancestry alone to predict
250 a trait increases as allele frequency differentiation increases³². The idea that including global
251 ancestry as a covariate in these analyses reduces power for SNPs with large MAF differences

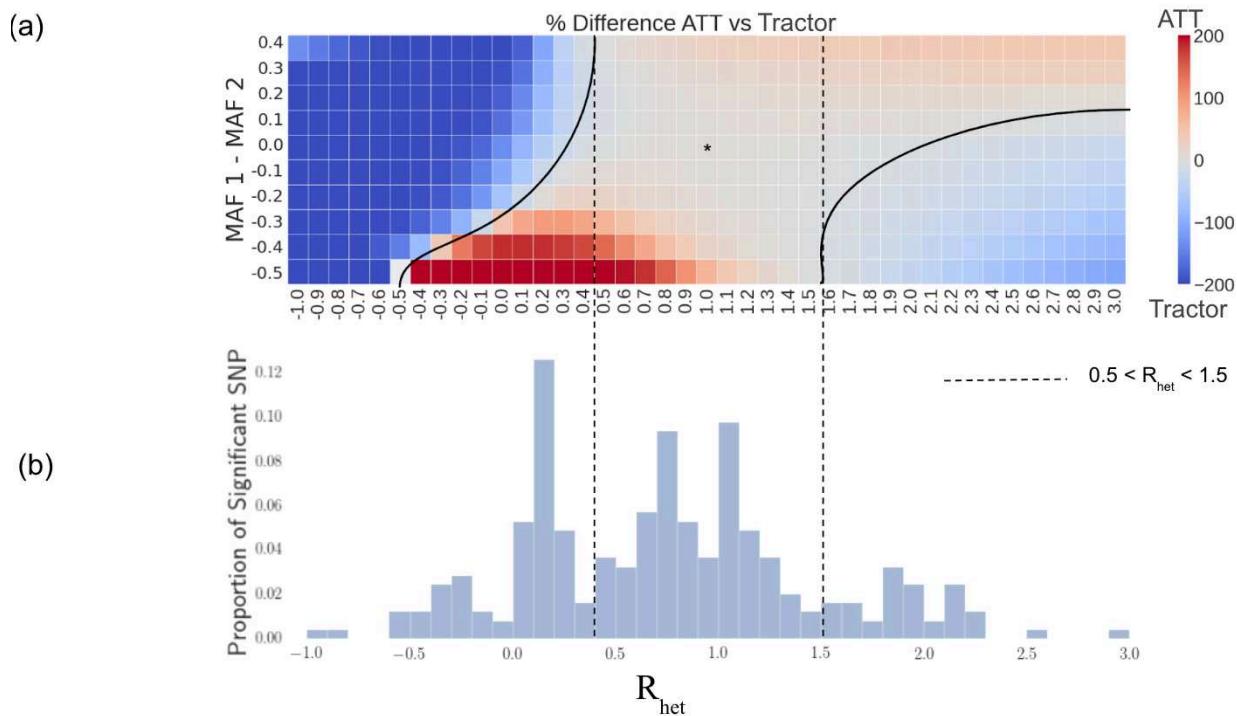
252 raises the question of how much attenuation can be expected when more exact measures of
253 global ancestry (such as principal components) are included in the analysis. However, the overall
254 power attenuation due to the inclusion of global ancestry is small compared to that due to local
255 ancestry; thus, we shift our focus back to considering local ancestry-specific effects on power.

256

257 **Impact of HetLanc on Power Depends on Allele Frequency Differences**

258 Next, we investigate the impact of MAF differences and HetLanc on power differences between
259 ATT and Tractor. The exact relationship between HetLanc (measured as R_{het}), MAF difference,
260 and percent difference in power is complex (Figure 3a). First, there is a window when $0.5 < R_{het} < 1.5$ in which, regardless of MAF difference, HetLanc is not enough to empower Tractor
261 over ATT. Thus, at these “low” levels of HetLanc, ATT will reliably have more power than Tractor
262 across the allele frequency spectrum. Similarly, when $R_{het} < -0.5$, there is no allele frequency
263 difference which would empower ATT over Tractor. This corroborates our findings that when
264 effect sizes are in opposite directions, Tractor is expected to have improved power over ATT
265 regardless of MAF difference. We can see that it is characteristics of both ATT and Tractor that
266 drive this trend (Figure S8). The power of ATT depends most strongly on the magnitude of R_{het}
267 and is diminished the most when effect sizes are in opposite directions. By contrast, the power
268 of Tractor depends strongly on both MAF difference and R_{het} . These two factors combine to
269 create an asymmetric shape for the percent difference in power (Figure 3a). This asymmetry in
270 power observed for the Tractor method is likely due to correlations between effective sample size,
271 allele frequency, global ancestry, and local ancestry that can occur in an asymmetric manner
272 when causal effect sizes and minor allele frequencies differ between local ancestries³². We
273 additionally investigate similar scenarios with varied global ancestry proportions (Figure S5),
274 heritability (Figure S6), and population-level MAF (Figure S7). While the exact boundaries of
275 these regions do differ, the overall shape of this heatmap and the conclusions mentioned above
276 do not qualitatively change.

278



279

280

281 **Figure 3: Impact of HetLanc on percent difference in power depends on MAF difference.**

282 **(a)** Heatmap of percent difference in power for ATT vs Tractor. The “*” indicates the center with
283 no HetLanc or MAF difference. Quantitative trait simulated 1,000 times for a population of 10,000
284 individuals on a trait with effect size β_1 ranging from -1.0 to 3.0 in increments of 0.1, and effect
285 size $\beta_2 = 1.0$. Global ancestry proportion 50/50, heritability at $h^2 = 0.005$, and minor allele
286 frequencies $MAF_1 = 0.5$ and MAF_2 ranging from 0.1 to 1.0 in increments of 0.1. Power calculated
287 using a standard threshold p-value $< 5 \times 10^{-8}$. **(b)** Histogram of empirical $R_{het} \frac{\beta_1}{\beta_2}$ for significant
288 SNPs found for 12 phenotypes in the UKBB. $\widehat{\beta}_1, \widehat{\beta}_2$ estimated using Tractor.

289

290 **ATT Finds More Significant Loci Across 12 Traits in the UK Biobank**

291 We next seek to understand the impact of correcting for local ancestry in genetic analyses in real
292 data. We investigate both Tractor and ATT in individuals with African-European admixture in the

293 UK Biobank. These individuals have on average 58.9% African and 41.1% European ancestry
294 over the population of 4,327 individuals. First, we investigate MAF differences between segments
295 of African and European local ancestry over 16,584,433 imputed SNPs. We find that 72.8% of
296 them have an absolute allele frequency difference of < 0.115 across local ancestry (Figure S10).

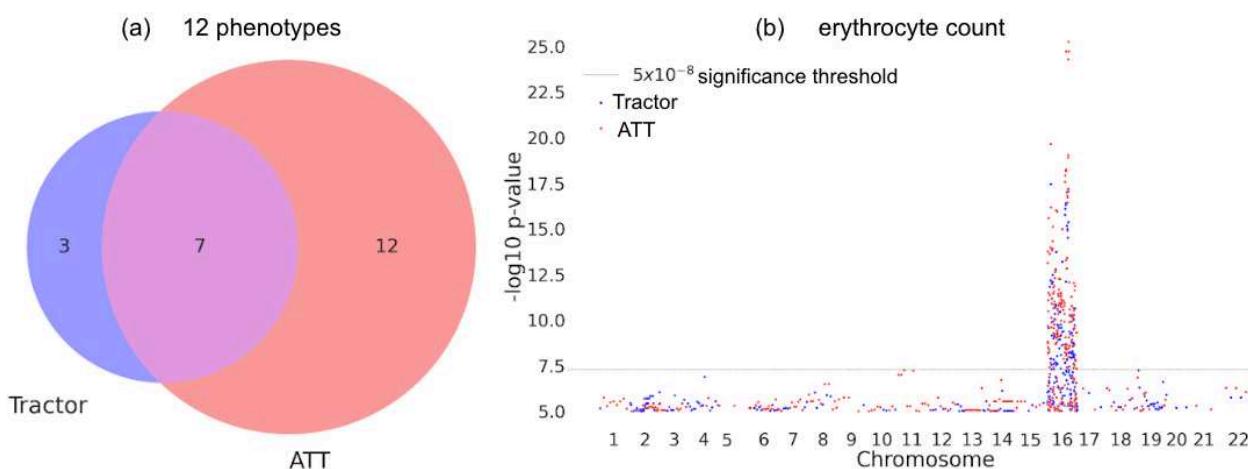
297

298 Next, we investigate empirically derived values of R_{het} to determine in which region of the heatmap
299 estimated effect sizes are likely to be found in real data (Figure 3b). We ran the Tractor method
300 on 12 quantitative traits to find the actual values of R_{het} for the estimated effect sizes β_{AFR} and
301 β_{EUR} . These traits were aspartate transferase enzyme (AST), BMI, cholesterol, erythrocyte count,
302 HDL, height, LDL, leukocyte count, lymphocyte count, monocyte count, platelet count, and
303 triglycerides. Then, we line up the histogram of these empirically derived values of R_{het} with the
304 heatmap. We find that for 69.3% of all significant SNPs, the empirical value for R_{het} is within this
305 [-0.5, 1.5] window. While this is an estimate, we predict the true difference between estimated
306 marginal effect sizes might be smaller than indicated by these empirical values because Tractor
307 is more powerful in identifying SNPs with heterogenous effect sizes. This result reflects previous
308 findings that causal effects are similar across ancestries within admixed populations²². Due to
309 this similarity in effect size, most of the significant SNPs sit in the center of the heatmap. This
310 region of this heatmap predicts ATT will have more power than Tractor. We can compare the
311 mean adjusted chi-square statistic of the SNPs found to be significant in this case. We find that
312 this statistic is significantly larger for the ATT method than the Tractor method (Figure S9). For
313 significant SNPs, the mean ATT χ^2 is 42.9, the mean adjusted Tractor χ^2 is 37.5, and the p-value
314 for the difference is 2.11×10^{-4} .

315

316 In addition to assessing HetLanc directly, we can also compare the number of independent
317 significant SNPs found by ATT and Tractor for these phenotypes. We find that while the number

318 of independent significant SNPs varies across all traits (Table S1), overall ATT finds more
319 significant independent signals than Tractor (Figure 4a). We find 22 independent significant loci,
320 with 19 loci found in ATT and 10 found in Tractor. This trend is most pronounced in HDL, in which
321 5 independent loci were determined to be significant by ATT compared to none for Tractor.
322 Similarly, BMI, leukocyte count, and monocyte count also only had independent significant loci
323 when testing using ATT as opposed to Tractor. Cholesterol and LDL had significant loci found by
324 both ATT and Tractor, with a larger number found by ATT. Height is the only trait for which Tractor
325 identified one significant locus but not ATT. Unfortunately, our sample sizes were not large
326 enough to detect any significant loci for platelet count, triglycerides, or lymphocyte count. All
327 significant loci for these 12 phenotypes are detailed in Table S1.



328 **Figure 4: Comparing significant SNPs found with ATT and Tractor. (a)** Venn diagram of
329 independent significant loci found using ATT and Tractor in the UKBB across 12 quantitative
330 traits. **(b)** Manhattan plot of chromosome for erythrocyte count in the UKBB. Significant SNPs
331 found with ATT shown in red and significant SNPs found with Tractor shown in blue.
332
333 Additionally, we find that while ATT often finds more significant independent loci than Tractor, the
335 two methods do not always find the same loci. Erythrocyte count is one phenotype in which we
336 find an equal number of independent significant loci using both ATT and Tractor. However, not all

337 loci overlap. Investigating the Manhattan plot of erythrocyte count specifically (Figure 4b) we see
338 that loci on chromosome 16 are found by both ATT and Tractor. But outside of the main locus,
339 both ATT and Tractor find separate additional significant regions. At the main locus, this
340 Manhattan plot clearly shows that ATT has significantly smaller p-values for the same locus.
341 Thus, in a smaller sample size only ATT would have found this important region. This example
342 highlights the importance of choosing the most highly powered association statistic for any given
343 situation. Manhattan plots for other phenotypes can be found in Figure S11.

344

345 **Discussion**

346 In this work, we seek to understand the impact that estimated allelic effect-size heterogeneity by
347 ancestry (HetLanc) has on the power of GWAS in admixed populations. Our main goal is to find
348 whether conditioning disease mapping on local ancestry leads to an increase or decrease in
349 power. We find that HetLanc and MAF differences are the two most important factors when
350 considering various methods for disease mapping in admixed populations. We consider two
351 association statistics - ATT, which ignores local ancestry, and Tractor, which conditions effect
352 sizes on local ancestry. We find that in cases with small or absent levels of HetLanc, ATT is more
353 powerful than Tractor in simulations of quantitative traits. This conclusion holds across a variety
354 of global ancestry proportions and SNP heritabilities. We find that as MAF differentiation between
355 ancestries increases, so does the improvement of power of ATT compared to Tractor. At high
356 HetLanc ($R_{het} > 1.5$) or when effect sizes are in opposing directions ($R_{het} < -0.5$), we find that
357 Tractor out-performs ATT. For African-European admixed individuals in the UKBB, most
358 significant loci have both small measured HetLanc and MAF differences. We find that across 12
359 quantitative traits, ATT finds more significant independent loci than Tractor. Furthermore, ATT
360 has smaller p-values for the loci that it shares with Tractor. This suggests that on smaller datasets
361 more of the shared loci would be found by ATT than by Tractor.

362

363 This work has several implications for GWAS in admixed populations. Our results suggest that
364 usually, ATT adjusted for global ancestry is the most powerful way to perform GWAS in an
365 admixed population. However, it may be possible to predict the comparative power of ATT and
366 Tractor using the allele frequencies and linkage disequilibria of a specific sample. Additionally,
367 since in real analyses ATT and Tractor often find different loci, it is important to keep both methods
368 in mind when performing analyses. These methods prioritize different types of loci, with ATT likely
369 prioritizing loci with higher MAF differences and Tractor prioritizing loci with higher levels of
370 HetLanc. From both scientific and social perspectives, it is important that admixed populations
371 are incorporated more effectively into genetic studies. By providing insight into the strengths and
372 limitations of these methods, we hope to enable studies to maximize their power in admixed
373 populations.

374

375 We conclude with caveats and limitations of our work. When hoping to understand these patterns
376 of power for association statistics, there are many combinations of different elements of genetic
377 architecture to consider. These include phenotypic factors such as environmental variance and
378 polygenicity, as well as elements of admixture such as the number of generations of admixture
379 and the strength of linkage disequilibrium. We could not consider them all, and thus it is likely
380 that additional nuances to our findings exist when other factors are considered. One major
381 element not considered in this work is case-control traits. While we chose to focus on quantitative
382 traits in this analysis due to their importance in simplicity and ubiquity, case-control traits are also
383 important in medicine. It is possible that the behavior of these phenotypes will vary compared to
384 the quantitative traits that we analyze here, both in simulations and real data. We suggest case-
385 control traits as an interesting avenue of research for future works. Lastly, we chose to focus our
386 analyses on ATT and Tractor due to their popularity and ease of use. We compare how these
387 methods work “out of the box” to provide simple and usable guidance for others. However, as

388 discussed in the introduction to this work, a variety of other association tests exist. It is likely that
389 in certain circumstances one of these existing methods would outperform both ATT and Tractor.

390

391 **Declaration of Interests**

392 The authors declare no competing interests.

393

394 **Acknowledgements**

395 The authors would like to acknowledge Ella Petter, Ruth Johnson, and Vidhya Venkateswaran for
396 their insightful feedback. RM supported in part by National Institutes for Health (NIH) award no.
397 T32HG002536 and BMH and GLM supported in part by NIH grant R35GM133531 to BMH. The
398 content is solely the responsibility of the authors and does not necessarily represent the official
399 views of the NIH.

400

401 **Data and Code Availability**

402 Code for this project, including simulation experiments, data processing pipeline, are available at
403 <https://github.com/rachelmester/AdmixedAssociation>. An application for UK Biobank individual-
404 level genotype and phenotype data can be made at <http://www.ukbiobank.ac.uk>.

405

406 **References**

- 407 1. Tian C, Gregersen PK, Seldin MF. Accounting for ancestry: population substructure and genome-wide association
408 studies. Human molecular genetics. 2008 Oct 15;17(R2): R143-50.
- 409 2. Mills MC, Rahal C. The GWAS Diversity Monitor tracks diversity by disease in real time. Nature genetics. 2020
410 Mar;52(3):242-3.

411 3. Hou K, Bhattacharya A, Mester R, Burch KS, Pasaniuc B. On powerful GWAS in admixed populations. *Nature*
412 *genetics*. 2021 Dec;53(12):1631-3.

413 4. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE. Human
414 demographic history impacts genetic risk prediction across diverse populations. *The American Journal of Human*
415 *Genetics*. 2017 Apr 6;100(4):635-49.

416 5. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J,
417 Cortes A. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018 Oct;562(7726):203-
418 9.

419 6. Ramirez AH, Sulieman L, Schlueter DJ, Halvorson A, Qian J, Ratsimbazafy F, Loperena R, Mayo K, Basford M,
420 Deflaux N, Muthuraman KN. The All of Us Research Program: data quality, utility, and diversity. *Patterns*. 2022
421 Aug 12;3(8):100570.

422 7. Zhou W, Kanai M, Wu KH, Rasheed H, Tsuo K, Hirbo JB, Wang Y, Bhattacharya A, Zhao H, Namba S, Surakka I.
423 Global Biobank Meta-Analysis Initiative: Powering genetic discovery across human disease. *Cell Genomics*. 2022
424 Oct 12;2(10):100192.

425 8. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in
426 diverse populations. *Nature Reviews Genetics*. 2010 May;11(5):356-66.

427 9. Qin H, Morris N, Kang SJ, Li M, Tayo B, Lyon H, Hirschhorn J, Cooper RS, Zhu X. Interrogating local population
428 structure for fine mapping in genome-wide association studies. *Bioinformatics*. 2010 Dec 1;26(23):2961-8.

429 10. Zaitlen N, Pasaniuc B, Sankararaman S, Bhatia G, Zhang J, Gusev A, Young T, Tandon A, Pollack S, Vilhjálmsson
430 BJ, Assimes TL. Leveraging population admixture to characterize the heritability of complex traits. *Nature genetics*.
431 2014 Dec;46(12):1356-62.

432 11. Zhong Y, Perera MA, Gamazon ER. On using local ancestry to characterize the genetic architecture of human
433 traits: genetic regulation of gene expression in multiethnic or admixed populations. *The American Journal of Human*
434 *Genetics*. 2019 Jun 6;104(6):1097-115.

435 12. Lin M, Park DS, Zaitlen NA, Henn BM, Gignoux CR. Admixed populations improve power for variant discovery and
436 portability in genome-wide association studies. *Frontiers in genetics*. 2021 May 24; 12:673167.

437 13. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, Highland HM, Patel YM, Sorokin EP, Avery
438 CL, Belbin GM. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*. 2019
439 Jun;570(7762):514-8.

440 14. Pasaniuc B, Zaitlen N, Lettre G, Chen GK, Tandon A, Kao WL, Ruczinski I, Fornage M, Siscovick DS, Zhu X,
441 Larkin E. Enhanced statistical tests for GWAS in admixed populations: assessment using African Americans from
442 CARE and a Breast Cancer Consortium. *PLoS genetics*. 2011 Apr 21;7(4): e1001371.

443 15. Atkinson EG, Maihofer AX, Kanai M, Martin AR, Karczewski KJ, Santoro ML, Ulirsch JC, Kamatani Y, Okada Y,
444 Finucane HK, Koenen KC. Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and
445 to boost power. *Nature genetics*. 2021 Feb;53(2):195-204.

446 16. Smith MW, O'Brien SJ. Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nature*
447 *Reviews Genetics*. 2005 Aug;6(8):623-32.

448 17. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide
449 association studies. *Nature reviews genetics*. 2010 Jul;11(7):459-63.

450 18. Korunes KL, Goldberg A. Human genetic admixture. *PLoS Genetics*. 2021 Mar 11;17(3):e1009374.

451 19. Kang SJ, Larkin EK, Song Y, Barnholtz-Sloan J, Baechle D, Feng T, Zhu X. Assessing the impact of global versus
452 local ancestry in association studies. In *BMC proceedings* 2009 Dec (Vol. 3, No. 7, pp. 1-6). BioMed Central.

453 20. Shriner D, Adeyemo A, Ramos E, Chen G, Rotimi CN. Mapping of disease-associated variants in admixed
454 populations. *Genome biology*. 2011 May;12(5):1-8.

455 21. Peterson RE, Kuchenbaecker K, Walters RK, Chen CY, Popejoy AB, Periyasamy S, Lam M, Iyegbe C, Strawbridge
456 RJ, Brick L, Carey CE. Genome-wide association studies in ancestrally diverse populations: opportunities,
457 methods, pitfalls, and recommendations. *Cell*. 2019 Oct 17;179(3):589-603.

458 22. Hou K, Ding Y, Xu Z, Wu Y, Bhattacharya A, Mester R, Belbin G, Conti D, Darst BF, Fornage M, Gignoux C. Causal
459 effects on complex traits are similar across segments of different continental ancestries within admixed individuals.
460 medRxiv. 2022 Jan 1.

461 23. Patel RA, Musharoff SA, Spence JP, Pimentel H, Tcheandjieu C, Mostafavi H, Sinnott-Armstrong N, Clarke SL,
462 Smith CJ, Durda PP, Taylor KD. Effect sizes of causal variants for gene expression and complex traits differ
463 between populations. *bioRxiv*. 2021 Jan 1.

464 24. Marigorta UM, Navarro A. High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS*
465 *genetics*. 2013 Jun 13;9(6): e1003566.

466 25. Shi H, Gazal S, Kanai M, Koch EM, Schoech AP, Siewert KM, Kim SS, Luo Y, Amariuta T, Huang H, Okada Y.
467 Population-specific causal disease effect sizes in functionally important regions impacted by selection. *Nature*
468 *communications*. 2021 Feb 17;12(1):1-5.

469 26. Brown BC, Ye CJ, Price AL, Zaitlen N, Asian Genetic Epidemiology Network Type 2 Diabetes Consortium.
470 Transethnic genetic-correlation estimates from summary statistics. *The American Journal of Human Genetics*.
471 2016 Jul 7;99(1):76-88

472 27. Galinsky KJ, Reshef YA, Finucane HK, Loh PR, Zaitlen N, Patterson NJ, Brown BC, Price AL. Estimating cross-
473 population genetic correlations of causal effect sizes. *Genetic epidemiology*. 2019 Mar;43(2):180-8.

474 28. Shi H, Burch KS, Johnson R, Freund MK, Kichaev G, Mancuso N, Manuel AM, Dong N, Pasaniuc B. Localizing
475 components of shared transethnic genetic architecture of complex traits from GWAS summary data. *The American*
476 *Journal of Human Genetics*. 2020 Jun 4;106(6):805-17.

477 29. McKeigue PM. Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in
478 admixed populations, by conditioning on parental admixture. *The American Journal of Human Genetics*. 1998 Jul
479 1;63(1):241-51.

480 30. Mani A. Local ancestry association, admixture mapping, and ongoing challenges. *Circulation: Cardiovascular*
481 *Genetics*. 2017 Apr;10(2): e001747.

482 31. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects
483 for stratification in genome-wide association studies. *Nature genetics*. 2006 Aug;38(8):904-9.

484 32. Liu J, Lewinger JP, Gilliland FD, Gauderman WJ, Conti DV. Confounding and heterogeneity in genetic association
485 studies with admixed populations. *American journal of epidemiology*. 2013 Feb 15;177(4):351-60.

486 33. Tang H, Siegmund DO, Johnson NA, Romieu I, London SJ. Joint testing of genotype and ancestry association in
487 admixed families. *Genetic epidemiology*. 2010 Dec;34(8):783-91.

488 34. Shriner D, Adeyemo A, Rotimi CN. Joint ancestry and association testing in admixed individuals. *PLoS*
489 *computational biology*. 2011 Dec 22;7(12): e1002325.

490 35. Seldin MF, Pasaniuc B, Price AL. New approaches to disease mapping in admixed populations. *Nature Reviews*
491 *Genetics*. 2011 Aug;12(8):523-8.

492 36. Wang X, Zhu X, Qin H, Cooper RS, Ewens WJ, Li C, Li M. Adjustment for local ancestry in genetic association
493 analysis of admixed populations. *Bioinformatics*. 2011 Mar 1;27(5):670-7.

494 37. Zhang J, Stram DO. The role of local ancestry adjustment in association studies using admixed populations.
495 *Genetic epidemiology*. 2014 Sep;38(6):502-15.

496 38. Duan Q, Xu Z, Raffield LM, Chang S, Wu D, Lange EM, Reiner AP, Li Y. A robust and powerful two-step testing
497 procedure for local ancestry adjusted allelic association analysis in admixed populations. *Genetic epidemiology*.
498 2018 Apr;42(3):288-302.

499 39. Chen W, Ren C, Qin H, Archer KJ, Ouyang W, Liu N, Chen X, Luo X, Zhu X, Sun S, Gao G. A generalized
500 sequential Bonferroni procedure for GWAS in admixed populations incorporating admixture mapping information
501 into association tests. *Human heredity*. 2015;79(2):80-92.

502 40. Simonin-Wilmer I, Orozco-Del-Pino P, Bishop T, Iles MM, Robles-Espinoza CD. An overview of strategies for
503 detecting genotype-phenotype associations across ancestrally diverse populations. *Frontiers in genetics*. 2021
504 Nov 5:2141.

505 41. Martin ER, Tunc I, Liu Z, Slifer SH, Beecham AH, Beecham GW. Properties of global-and local-ancestry
506 adjustments in genetic association tests in admixed populations. *Genetic epidemiology*. 2018 Mar;42(2):214-29.

507 42. Qin H, Zhu X. Power comparison of admixture mapping and direct association analysis in genome-wide association
508 studies. *Genetic epidemiology*. 2012 Apr;36(3):235-43.

509

510

511 **Methods**

512 **Simulated Genotypes and Phenotypes**

513 We simulate genotypes using the following procedure:

514 1. Draw global ancestry proportions $\alpha \sim N(\theta, \sigma^2)$ for 10,000 individuals where θ is the
515 expected global ancestry proportion (either 0.5, 0.6, or 0.8) of ancestry 2, and σ^2 is the
516 variance of global ancestry in the population ($\sigma^2 = 0.125$). We use $\sigma^2 = 0.125$ to reflect
517 the variance of global ancestry found in the UK Biobank admixed population. α is coerced
518 to 0 if it is negative and 1 if it's larger than 1.

519 2. For each individual, draw a local ancestry count $l \sim Binomial(\alpha, 2)$, where l represents
520 the local ancestry count of ancestry 2.

521 3. For each local ancestry, draw a genotype $g_i \sim Binomial(l, f_i)$, where f_i represents the
522 minor allele frequency at local ancestry i .

523 We simulate phenotypes using the following procedure:

524 1. Standardize genotypes so that they have a mean 0 and variance 1.

525 2. Given some effect sizes β_1, β_2 calculate $Var_g = (\beta_1 g_1 + \beta_2 g_2)^2$, where Var_g is the genetic
526 variance component of the phenotypes.

527 3. Given some heritability h^2 , calculate $Var_e = Var_g \frac{1-h^2}{h^2}$, where Var_e is the environmental
528 variance component of the phenotypes. This comes from the equation $h^2 = \frac{Var_g}{Var_g + Var_e}$.

529 4. For each individual, draw $\epsilon \sim N(0, Var_e)$ where ϵ is the random noise to add to the
530 phenotype to represent environmental variables.

531 5. Repeat for 1,000 replicates.

532

533 **Real Genotypes and Phenotypes**

534 For our real data analysis, we used genotypes from the UK Biobank. We limited our study to
535 participants with admixed African-European ancestry. Overall, we had 4,327 individuals with an
536 average of 58.9% African and 41.1% European ancestry. We used the imputed genotypes for
537 these individuals with a total of 16,584,433 SNPs. We calculated the top 10 PCs for these
538 genotypes and added these PCs as covariates to all analyses as our global ancestry component.
539 The phenotypes we used are also from the UK Biobank, and include aspartate transferase
540 enzyme (AST), BMI, cholesterol, erythrocyte count, HDL, height, LDL, leukocyte count,
541 lymphocyte count, monocyte count, platelet count, and triglycerides. We log transformed AST,
542 BMI, HDL, leukocyte count, lymphocyte count, monocyte count, platelet count, and triglycerides
543 to analyze all 12 traits as quantitative, continuous traits. We standardized all genotypes and
544 phenotypes to be mean centered at 0.0 and have a standardized variance of 1.

545

546 **Association Testing**

547 Simulated Data

548 We calculate the ATT and Tractor association tests on simulated data using scripts that can be
549 found on <https://github.com/rachelmester/AdmixedAssociation>. ATT is a 1 degree of freedom
550 association test that uses the model $y = \beta g + e_\alpha^T \alpha + \epsilon$ to test for $\beta = 0$ against a null
551 hypothesis that includes global ancestry (α). Tractor is a two degree of freedom association test
552 that uses the model $y = \beta_1 g_1 + \beta_2 g_2 + e_l l + e_\alpha^T \alpha + \epsilon$ to test for $\beta_1 = 0$ and $\beta_2 = 0$ against a null
553 hypothesis that includes local ancestry (l) and global ancestry (α). They can both be adapted to
554 be used on case-control phenotypes or to adjust for additional covariates such as age and sex.
555 For our simulations, we used global ancestry proportions as our measure of global ancestry (α)
556 and did not need to adjust for any additional covariates such as age and sex as we did not model

557 those factors in our simulations. For power calculations, we use a standard significance threshold
558 of p-value $< 5 \times 10^{-8}$.

559

560 **Real Data**

561 We used admix-kit (<https://kangchenghou.github.io/admix-kit/index.html>) to perform the ATT and
562 Tractor association tests on this data and extracted the p-values. In order to determine significant
563 SNPs, we filtered for SNPs with a standard p-value of $< 5 \times 10^{-8}$. For the Manhattan plots, we
564 plot all SNPs with a p-value $< 1 \times 10^{-5}$ for computational plotting purposes. For the Venn
565 diagrams, in order to determine whether SNPs were part of the same loci, we grouped SNPs
566 within a 500kB radius, and kept the most significant SNP from each test (ATT and Tractor) in that
567 locus.

568

569 **Measures Used to Compare Our Results**

570 In this work, we introduce several key measures that we use to compare our results. The formal
571 definitions of these are the following:

572 Percent difference in power:
$$\frac{2(Power_{ATT} - Power_{Tractor})}{Power_{ATT} + Power_{Tractor}}$$

573 Adjusted chi square: We take the p-value from a χ^2 statistic and convert it back to a χ^2_1 statistic,
574 regardless of the original degrees of freedom. The adjusted chi square score for a χ^2_1 is itself.