Examining the Effect of Social Factors of Health on Human Trait Heritability

Jaclyn L. Liquori[1*†], Oona Shigeno Risse-Adams[1,2†], Rangarajan Bharadwaj[2], Dylan Ong[1], Nasa Sinnott-Armstrong[2,3,4*†], Shaila A. Musharoff[1*†]

[1]Department of Computational Biology, Cornell University, Ithaca, NY, USA.
[2]Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA.
[3]Department of Genome Sciences, University of Washington, Seattle, WA, USA.
[4]Brotman Baty Institute, University of Washington, Seattle, WA, USA.

*Corresponding author(s). E-mail(s): jl3648@cornell.edu; nasa@fredhutch.org; sam442@cornell.edu;
†These authors contributed equally to this work.

Abstract

Social factors of health (SFOH) are critical determinants of human traits, but are rarely incorporated into genetic models. Using individual-level data from the All of Us Research Program, we systematically assess how SFOH survey data, summarized via principal component analysis (PCA), multiple correspondence analysis (MCA), and a novel Social Similarity Index (SSI), influence both heritability estimates and trait variance patterns for 18 traits (6 anthropometric and 12 metabolic) in 85,963 individuals of diverse racial identities. Including SFOH measures as covariates in Haseman-Elston (HE) regression significantly reduced heritability estimates ($p < 0.0001$) relative to models including only demographic factors for body mass index, hip circumference, and waist circumference in self-identified White individuals (n=67,545), female individuals (n=54,967), male individuals (n=30,996), and the full cohort, as well as of HDL cholesterol in the full and White cohorts, with absolute reductions in heritability of up to 0.275. This suggests that unmodeled social structure can be misattributed to genetic effects. Heritability estimates did not change significantly after including SFOH measures as covariates in self-identified Black/African American individuals (n=6,538), potentially due to lower statistical power in this smaller sample. We discovered that specific SFOH components are themselves significantly heritable: PC5, PC6, MC5, SSI have heritabilities between 0.024 (SSI, se 0.004) and 0.050 (MC5, se 0.004), and are generally loaded by questions related to religious attendance, feelings of healthcare discrimination, perceived stress, and social support. The heritability of these SFOH components indicates the entanglement of genetic and social factors. We find that SFOH measures significantly influence trait variance based on double generalized linear models, with anthropometric measures showing the most extensive dispersion effects across all three SFOH measures, even after adjusting for demographic factors. Furthermore, age and sex have extensive effects on trait dispersion, reinforcing that genetic and environmental architecture of these traits varies across demographic factors. Our findings demonstrate that SFOH can influence heritability estimates and trait variance,

underscoring the necessity of careful model specification and appropriate inclusion of social and environmental factors in genetic studies.

Keywords: heritability, social factors of health, human genetics

1 Introduction

Modeling the social drivers of human traits remains a critical challenge in statistical genetics. While genome-wide studies routinely adjust for genetic stratification, trait-relevant social and environmental exposures are typically unmeasured or oversimplified. Social factors of health (SFOH)—such as housing, financial insecurity, education, and perceived stress—shape both disease risk and access to health care. SFOH are increasingly captured in structured survey form in biobank-scale cohorts, including the All of Us Research Program [1]. However, SFOH data are rarely explicitly incorporated in heritability and genetic association models [2, 3], this omission could result in residual trait variation that remains structured by social factors. In practice, this means that social stratification can be misattributed as genetic variance, creating spurious signals that could lead to incorrect conclusions about trait heritability and its underlying genetic architecture. Although the All of Us SFOH data has been previously characterized [4] and connected to health outcomes [5, 6], to our knowledge, it has not been incorporated in heritability models.

Though survey data could contain trait-relevant information, its structure and information content varies from that of genetic data, making its integration with genetic data to model human traits challenging. Survey data has known participation bias [7] as well as completion and non-response rates that are correlated with demographic factors, such as race/ethnicity [8], which complicates the inference of population-specific factors. Furthermore, SFOH survey data contains information that is potentially correlated with genetic data, and understanding the relationship between the two is essential for their joint analysis.

Here, we investigate how high-dimensional social structure captured by the All of Us SFOH survey influences single nucleotide polymorphism based heritability (h_g^2) – the proportion of phenotypic variance in a population explained by additive genetic effects of common SNPs – of 18 anthropometric and metabolic traits. Throughout this paper, we use the terms 'heritability' and h^2 to refer to this SNP-based heritability (h_g^2), which provides a lower-bound estimate for narrow-sense heritability. To summarize the SFOH survey responses, we apply one of three strategies: (i) dimension reduction via principal component analysis (PCA) [9, 10], (ii) dimension reduction via multiple correspondence analysis (MCA), and (iii) a novel environmental relatedness matrix (ERM), from which we computed a novel Social Similarity Index (SSI), an individual-level social environment score. Each strategy yields participant-level covariates that can be used in genetic models to integrate the critical yet often overlooked influences of environmental factors and SFOH. These elements, like genetic variations, are stratified within populations, but can change rapidly and profoundly impact health outcomes.

2 Results

We analyzed data from All of Us (v7) participants who met the following criteria: availability of SFOH survey data, genotyping array data, a selected sex at birth of "female" or "male", and a self-reported racial identity. This resulted in a maximal sample of n=85,963 individuals. Collinearity and structure analyses were performed on this full cohort, while heritability estimation was restricted to the subsets of individuals who also had measurements for the 18 traits of interest.

2.1 Collinearity

To characterize the relationship of SFOH and genetic measures, we assessed collinearity between the three survey-derived SFOH measures (SSI, MCA, PCA) and genetic PCs (Fig. 1). For the SSI, the maximum absolute Pearson correlation with any genetic PC was 0.317 and all variance inflation factors (VIFs), where values above 5 signal moderate multicollinearity, remained below 1.150, indicating an acceptable amount of multicollinearity. For the top forty SFOH MCA axes, the maximum correlation with any genetic PC was 0.011 (VIFs < 1.145), and for the top twenty SFOH PCA axes the maximum correlation with any genetic PC was 0.010 (VIFs < 1.145). Thus, each SFOH measure can safely be modeled alongside genetic PCs without problematic multicollinearity.

However, when analyzing self-identified White (n=67,545) and Black/African American (n=6,538) groups separately, elevated VIFs (3.99) prompted further investigation. Pearson correlation plots of SFOH and genetic PCs within each subgroup revealed modest correlations between genetic components that were not evident in the combined sample (Figs. S5 and S6). Because heritability and SFOH influences vary by sex, we additionally estimated Pearson correlations between SFOH and genetic PCs in a sex-stratified group (Figs. S28 and S29) and observed similar results. These correlations may reflect the expected loss of orthogonality when principal components computed on a full sample are applied to subsets rather than meaningful structure. As such, caution is warranted when interpreting within-group relationships, as restricting analyses to specific demographic groups may amplify spurious associations among otherwise orthogonal PCs and lead to the misattribution of social effects to genetic structure, or vice versa.

2.2 Trait heritability

We estimate heritability using two complementary methods: Haseman–Elston (HE) [11] regression, a computationally efficient method of moments estimator suitable for large datasets, and Genome-based Restricted Maximum Likelihood (GREML) [11], a more resource-intensive approach that potentially provides less biased estimates with smaller standard errors. To assess how SFOH estimates and genetic PCs influence these estimates, we tested six nested covariate models (Table 1) that progress from basic demographic adjustment, with the analysis-relevant subset of age, sex, self-identified race, and self-identified ethnicity (Base, Model 0), to full covariate sets (Model 5). The single-statistic SFOH SSI only appears in two models (Models 2 and 4) while the partitioned multi-axis SFOH MCA and PCA summaries appear in four models (Models 2–5). The top SFOH axes, SFOH PCs 1:3 ($\sim$ 16% of variance explained) (Fig. S3) and SFOH MC axes 1:6 ($\sim$ 15% variance explained) (Fig. S4), appear in Models 2 and 4, while the informative SFOH axes, SFOH PCs

1:20 and SFOH MC axes 1:40, appear in Models 3 and 5. The top genetic PCs 1:3 ($\sim$ 67.5% of variance explained) appear in Models 1, 4, and 5.

Model	Covariates
0	Base: Age + Sex + Race + Ethnicity (or a subset thereof)
1	Model 0 + Genetic PCs 1:3
2	Model 0 + SSI, or SFOH PCs 1:3, or MC axes 1:6
3	Model 0 + SFOH PCs 1:20 or MC axes 1:40
4	Model 2 + Genetic PCs 1:3
5	Model 3 + Genetic PCs 1:3

Table 1: Covariate models used in trait heritability analyses. Models build progressively from minimal to full adjustment.

2.2.1 HE regression is sensitive to unmodeled SFOH

We used HE regression to estimate heritability via cross product (HE-CP). In the full cohort (Fig. S16) (maximal n=85,963), adding genetic PCs 1:3 (Models 1, 4, and 5) consistently reduced heritability estimates relative to the Base (Model 0) and SFOH-only models (Models 2 and 3). These reductions were significant in 11 of 18 traits: BMI, HDL cholesterol, carbon dioxide, creatinine, diastolic blood pressure, glucose, hip circumference, heart rate, sodium, urea nitrogen, and waist circumference. This pattern indicates that genetic PCs capture relevant population structure that inflates heritability estimates when unaccounted for, which is a known property of HE regression. When genetic ancestry is associated with trait values, this method may misattribute this non-causal covariance to genetic factors; including genetic PCs as covariates corrects for this confounding. In the full cohort, adding SFOH MCA or PCA covariates to the Base model (Models 2 and 3) also reduced the heritability estimate of 4 of 18 traits - BMI, hip circumference, waist circumference, and HDL cholesterol - while the SSI had no significant effect on any trait. Specifically, including SFOH MCA axes 1:6 (Model 2) significantly reduced heritability estimates relative to the Base model (Model 0) for BMI (by 0.113 from 0.268 to 0.155), hip circumference (by 0.093 from 0.302 to 0.209), and waist circumference (by 0.107 from 0.339 to 0.232). Including the MCA axes 7:40 (Model 3) further reduced the heritability of these traits, though not significantly. While SFOH PCs 1:3 (Model 2) had little effect on trait heritability, adding SFOH PCs 4:20 (Model 3) significantly reduced heritability relative to both the Base model (Model 0) and SFOH PCs 1:3 alone (Model 2) for BMI (by 0.143 from 0.268 to 0.129), HDL cholesterol (by 0.125 from 0.274 to 0.149), hip circumference (by 0.117 from 0.302 to 0.185), and waist circumference (by 0.127 from 0.339 to 0.212).

2.2.2 Results are recapitulated in White individuals, but not Black/African American individuals

We then used HE regression to estimate heritability via HE-CP in race-stratified groups restricting our analysis to Black/African American (maximal n=6,513) and White individuals (maximal n=67,164). In White individuals (Fig. 4), adding genetic PCs 1:3 (Models 1, 4, and 5) consistently reduced heritability estimates relative to the Base (Model 0) and SFOH-only models (Models 2 and 3).

In White individuals, adding SFOH MCA or PCA covariates to the Base model (Models 2 and 3) also reduced the heritability estimate of 4 of 18 traits - BMI, hip circumference, waist circumference, and HDL cholesterol - while the SSI had no significant effect on any trait, recapitulating the full cohort results. Specifically, including SFOH MCA axes 1:6 (Model 2) significantly reduced heritability estimates relative to the Base model (Model 0) for BMI (by 0.233 from 0.526 to 0.293), hip circumference (by 0.208 from 0.633 to 0.425), and waist circumference (by 0.231 from 0.686 to 0.455). Including the MCA axes 7:40 (Model 3) further reduced the heritability of these traits, though not significantly. While SFOH PCs 1:3 (Model 2) had little effect on trait heritability, adding SFOH PCs 4:20 (Model 3) significantly reduced heritability relative to both the Base model (Model 0) and SFOH PCs 1:3 alone (Model 2) for BMI (by 0.275 from 0.526 to 0.251), HDL cholesterol (by 0.216 from 0.475 to 0.259), hip circumference (by 0.250 from 0.633 to 0.383), and waist circumference (by 0.261 from 0.686 to 0.425).

In Black/African American individuals, (Fig. 5), including only SFOH covariates (Models 2 and 3) did not significantly change heritability relative to the Base model, and adding genetic PCs 1:3 (Models 1, 4, and 5) reduced heritability point estimates, though not significantly so, possibly reflecting reduced statistical power. The smaller sample size in this group limits statistical power, making it difficult to draw strong conclusions about the contribution of SFOH measures. Additionally, these SFOH axes may explain less variation in this group.

2.2.3 SFOH covariates have a larger effect on heritability in females than in males

We then used HE regression to estimate heritability via HE-CP in sex-stratified groups restricting our analysis to individuals who selected "Male" (maximal n=30,996) or "Female" (maximal n=54,967) as their sex at birth. In Female individuals (Fig. S20), adding genetic PCs 1:3 (Models 1, 4, and 5) consistently reduced heritability estimates relative to the Base (Model 0) and SFOH-only models (Models 2 and 3). In female individuals, adding SFOH MCA or PCA covariates to the Base model (Models 2 and 3) also reduced the heritability estimate of 3 of 18 traits - BMI, hip circumference, and waist circumference - while including the SSI had no significant effect on any trait, recapitulating results seen in the whole analysis and White cohorts, which is consistent with females being the predominant group in both of these analyses. Specifically, including SFOH MCA axes 1:6 (Model 2) significantly reduced heritability estimates relative to the Base model (Model 0) for BMI (by 0.116 from 0.293 to 0.177). Including the MCA axes 7:40 (Model 3) further reduced the heritability of these traits, significantly so for hip circumference (by 0.208 from 0.633 to 0.425) and waist circumference (by 0.150 from 0.360 to 0.210). While SFOH PCs 1:3 (Model 2) had little effect on trait heritability, adding SFOH PCs 4:20 (Model 3) significantly reduced heritability relative to both the Base model (Model 0) and SFOH PCs 1:3 alone (Model 2) for BMI (by 0.142 from 0.292 to 0.150), hip circumference (by 0.12 from 0.321 to 0.201), and waist circumference (by 0.124 from 0.360 to 0.236).

In Male individuals (Fig. S24), adding genetic PCs 1:3 (Models 1, 4, and 5) consistently reduced heritability estimates relative to the Base (Model 0) and SFOH-only models (Models 2 and 3). Adding SFOH MCA or PCA covariates to the Base model (Models 2 and 3) also reduced the heritability estimate of 3 of 18 traits - BMI,

hip circumference, and waist circumference - while the SSI had no significant effect on any trait. Specifically, while including SFOH MCA axes 1:6 (Model 2) reduced heritability estimates relative to the Base model (Model 0) for various traits, none of these reductions were significant. However, including MCA Axes 7:40 (Model 3) significantly reduced the heritability estimate for BMI (by 0.144 from 0.223 to 0.079). While SFOH PCs 1:3 (Model 2) had little effect on trait heritability, adding SFOH PCs 4:20 (Model 3) significantly reduced heritability relative to both the Base model (Model 0) and SFOH PCs 1:3 alone (Model 2) for BMI (by 0.133 from 0.223 to 0.090).

One possible explanation for these reductions is that survey-derived SFOH axes capture social and environmental variation that is partially correlated with genetic similarity. Because HE regression attributes any phenotypic covariance aligned with the genetic relatedness matrix (GRM) to additive genetic variance, unmodeled social structure can inflate heritability estimates. Including SFOH covariates removes this shared variance, leading to lower h^2 estimates even without genetic PCs. This suggests that SFOH measures not only reflect non-genetic contributions to traits but can also act as proxies for confounding structure that would otherwise be misattributed to genetic effects.

These analyses demonstrate that survey-derived SFOH data summarized with PCA and MCA can influence heritability estimates obtained via HE regression, sometimes significantly so. Our results suggest that genetic PCs and SFOH measures may capture complementary information relevant to confounding in HE heritability models.

2.2.4 GREML provides robust heritability estimates unaffected by SFOH covariate and genetic PC adjustment

Given the substantial sample size discrepancy between the race-stratified groups, we then repeated the race-stratified analyses in GREML. In both White (Fig. 6) and Black/African American individuals (Fig. 7), heritability estimates obtained via GREML were comparable to the HE models that included genetic PCs, and all GREML estimates remained statistically similar across all models regardless of the inclusion of SFOH covariates or genetic PCs. This stability is a key feature of GREML. Unlike HE regression, GREML uses a joint model that partitions trait variance into fixed and random polygenic effects through likelihood optimization. This full information approach leads to stable heritability estimates even when genetic PCs are added. Due to GREML's computational burden, we only used GREML for anthropometric traits in race-stratified analyses to account for the discrepancy in sample size.

2.3 Specific SFOH components are themselves significantly heritable

To assess whether the SFOH measures are themselves heritable, we applied HE regression and GREML to the three SFOH measures as traits: SFOH PCs, SFOH MC axes, and the SSI. We tested two nested covariate models (Table 2) to assess how basic demographics and genetic PCs influence these estimates.

Model	Covariates
0	Age + Sex + Race + Ethnicity (or a subset thereof)
1	Model 0 + Genetic PCs 1:3

Table 2: Covariate models used in SFOH measure heritability analyses.

2.3.1 HE regression identifies multiple heritable SFOH measures

We first estimate the heritability of SFOH measures in the full cohort. For the SSI (Fig. S15), we find that while including genetic PCs decreases the estimate, the heritability estimate for the SSI measure remains significantly different from zero ($h^2 = 0.029$, $SE = 0.005$). When including genetic PCs in models estimating the heritability of SFOH PCs (Fig. S14), SFOH PCs 1:8 and 12 remained significantly different from zero, with SFOH PC5 ($h^2 = 0.063$, $SE = 0.006$) and SFOH PC6 ($h^2 = 0.040$, $SE = 0.005$) remaining the highest after adjustment. When including genetic PCs in models estimating the heritability of SFOH derived MC axes (Fig. S13), MC axes 1:5, 10, 11, 16, 23, and 24 remained significantly different from zero with MC5 ($h^2 = 0.086$, $SE = 0.007$) remaining the highest after adjustment.

We then repeat these analyses in a sex-stratified cohort, restricting to females and males. In females, we find that the SSI (Fig. S19) heritability estimate remains significantly different from zero after adjusting for genetic PCs ($h^2 = 0.028$, $SE = 0.006$). When including genetic PCs in models estimating the heritability of SFOH PCs (Fig. S18), SFOH PCs 1, 3, 4, 5, and 6 remained significantly different from zero, with SFOH PC5 ($h^2 = 0.057$, $SE = 0.008$) and SFOH PC6 ($h^2 = 0.035$, $SE = 0.007$) remaining the highest after adjustment. When including genetic PCs in models estimating the heritability of SFOH derived MC axes (Fig. S17), MC axes 1, 2, 4, 5, and 16 remained significantly different from zero with MC5 ($h^2 = 0.081$, $SE = 0.009$) remaining the highest after adjustment.

In males, we find that when including genetic PCs in models estimating the heritability of SFOH PCs (Fig. S22), SFOH PCs 3, 5, and 6 remained significantly different from zero, with SFOH PC5 ($h^2 = 0.058$, $SE = 0.014$) and SFOH PC6 ($h^2 = 0.035$, $SE = 0.012$) remaining the highest after adjustment. When including genetic PCs in models estimating the heritability of SFOH derived MC axes (Fig. S21), MC axes 2 and 5 remained significantly different from zero with MC5 ($h^2 = 0.076$, $SE = 0.015$) remaining the highest after adjustment.

We then repeat these analyses in a race-stratified cohort, restricting to Black/African American and White individuals. In Black/African American individuals, we find that no heritability estimates for SFOH measures are significantly different from zero even when excluding genetic PCs.

In White individuals, we find that the heritability estimate for the SSI (Fig. S12) measure remains significantly different from zero when including genetic PCs ($h^2 = 0.042$, $SE = 0.007$). We find that when including genetic PCs in models estimating the heritability of SFOH PCs (Fig. S11), SFOH PCs 1:8 and 12 remained significantly different from zero, with SFOH PC5 ($h^2 = 0.066$, $SE = 0.008$) and SFOH PC6 ($h^2 = 0.055$, $SE = 0.007$) remaining the highest after adjustment. When including genetic PCs in models estimating the heritability of SFOH derived MC axes (Fig. S10), MC axes 1, 2, 4, 5, 10, 11, and 23:25 remained significantly

different from zero with MC5 ($h^2 = 0.112$, $SE = 0.010$) remaining the highest after adjustment.

These heritability analyses reveal several consistent patterns across demographics. The SSI and specific SFOH components, notably SFOH PC5, PC6, and MC5 demonstrate significant heritability in the overall cohort, and across stratified analyses, with estimates remaining stable after adjusting for genetic PCs. The consistent significance of SFOH PC5, PC6, and MC5 across strata suggests they capture SFOH effects with a robustunderlying genetic arcitecture, though this signal varies in stregnth across racial groups, potentially due to differences in social context, ancestry, or statistical power.

2.3.2 Heritability of select SFOH components is validated with GREML

To validate the above findings using a more robust approach we re-analyzed the significantly heritable SFOH measures (SSI, SFOH PCs 5 and 6, and MC axis 5) in the full cohort using GREML. We find that GREML estimates for these traits (Fig. 8) remain significantly different from zero when including genetic PCs, providing stronger evidence for genuine genetic influence on these SFOH measures. They also remain quite substantial. After adjusting for genetic PCs, h^2 for these measures ranged from 0.024 (SSI, $SE = 0.004$) to 0.050 (MC5, $SE = 0.004$). This consistency across estimation methods and stratified cohorts reinforces that these particular SFOH measures may capture social patterns with meaningful genetic influence.

Our analysis identified SFOH PC5, PC6, and MC5 as the most heritable SFOH measures after genetic PC adjustment, indicating they contain substantial genetic signal. We also note that the set of SFOH PCs 4-20 substantially reduces heritability estimates. This is notable as it implies these later PCs may capture residual population structure not accounted for by the first three SFOH PCs. Similarly, SFOH MC5 is the most heritable of the SFOH MC axes, even after genetic PC adjustment. MC5 is in the top set of MC axes 1:6. Including MC axes 1:6 reduces heritability more than SFOH PCs 1:3, even though they explain a similar amount of variance (SFOH PCs 1:3 $\sim$ 16% and MC axes 1:6 $\sim$ 14.9%). This suggests that the stronger effect of MC 1-6 may not be simply due to variance explained, and the high heritability of MC axis 5 is a plausible contributing factor.

The inclusion of SFOH PCs 4-20, MC axes 1-6, and genetic PCs 1-3 all substantially reduce trait heritability estimates for a subset of the traits analyzed here. The highly heritable nature of SFOH PCs 5 and 6, and MC axis 5 within these sets suggests that they are strongly associated with this reduction in heritability estimates.

2.4 SFOH measures and fundamental demographics strongly influence trait variance

We applied double generalized linear models, as implemented in the R package dglm, to investigate how SFOH measures and demographics influence the dispersion of anthropometric and metabolic traits. All models included comprehensive adjustment for age, sex, race, and ethnicity while testing three distinct SFOH measures: 20 SFOH PCs, 20 SFOH MCs, and the SSI.

2.4.1 Comprehensive dispersion effects across SFOH measures

All three SFOH measures showed widespread significant effects on trait dispersion after multiple testing correction. The SFOH PC models (Fig. S30) demonstrate significant dispersion effects across numerous traits. SFOH PC5, in particular shows widespread effects on trait variance in anthropometric measurements, cardiovascular traits, and metabolic markers. SFOH PC 10 also showed broad significant effects, particularly on renal function markers. The pattern of significance tiers revealed that while many effects reached nominal or trait-wise significance, a substantial number survived stringent overall correction indicating robust SFOH influence on trait variance independent of demographic factors.

Multiple MC axes showed widespread significant effects on trait dispersion (Fig. 9). MC5, in particular, demonstrates strong dispersion effects on anthropometric measures and cardiovascular traits. MCs 3 and 4 also show broad significance, particularly on metabolic traits. Like the SFOH PCs, many MC axes reached nominal significance and a substantial amount survived stringent test correction.

The SSI also demonstrated significant dispersion effects on traits (Fig. S31). Anthropometric measures, particularly hip and waist circumference show the strongest and most consistent association with the SSI. The SSI consistently reached overall significance after stringent test correction.

The three SFOH measurements: PCs, MC axes, and SSI suggest that SFOH influence trait variance through both multi-dimensional approaches and a broader single summary. The convergence of all three measures on anthropometric traits reinforces the robust connection between SFOH and body composition variability independent of demographic factors.

2.4.2 Comprehensive dispersion effects of demographics

The demographic covariates in our dispersion models contributed significantly to trait dispersion, independent of SFOH measures. Age demonstrated the most widespread influence across traits, suggesting that variance patterns systematically change with age. Sex, race, and ethnicity showed more targeted, but still significant, dispersion effect, indicating that demographic factors contribute to trait variability beyond their associations with SFOH measures. Further, demographic factors overall showed a greater magnitude of effect sizes, indicating that they are highly influential in trait variability provided they remain significant. The persistence of significant demographic factors after comprehensive SFOH adjustment underscores the importance of including these covariates to isolate true SFOH measure influences. Furthermore, the complex pattern of positive and negative effects across demographic covariates suggests that different groups may experience distinct trait variation, highlighting the multidimensional nature of how demographics and SFOH shape trait variability.

2.5 Structure analysis

We assessed the extent to which SFOH and genetic population structure are associated by examining bidirectional relationships between SFOH PCs and MCs and genetic PCs in the full cohort as well as in the sex- and race-stratified groups separately.

2.5.1 SFOH measures and genetic PCs show bidirectional but limited associations in the full cohort

First, to assess how social factors explain genetic structure, we regressed each of genetic PCs 1:10 on SFOH PCs 1:20 using multiple linear regression in the combined sample of individuals. A linear combination of SFOH PCs explained non-trivial variation in genetic PC1 ($R^2 = 0.184$), likely reflecting alignment between global ancestry and social groupings (Table S7). For genetic PCs 2:10, the variance explained by SFOH PCs was considerably lower (max R^2 for any genetic PC is 0.035), suggesting limited influence beyond the primary ancestry axis, genetic PC 1.

We then followed this up with regressing each of genetic PCs 1:10 on SFOH derived MC axes 1:40. A linear combination of MC axes explained non-trivial variation in genetic PC1 ($R^2 = 0.206$), again, likely reflecting alignment between global ancestry and social groupings (Table S8). As seen in the previous analysis, variance explained by MC axes for genetic PCs 2:10 was considerably lower (max R^2 for any genetic PC is 0.043).

Next, to assess how genetic structure explains social factor patterns, we regressed each of SFOH PCs 1:20 on genetic PCs 1:10. Genetic PCs explained modest amounts of variation in SFOH structure. While all twenty SFOH PCs showed statistically significant associations with genetic PCs, most R^2 values remained under 0.029, with the strongest associations observed for SFOH PCs 2 and 5 ($R^2 = 0.041$ and 0.062, respectively; Table S9). These components (SFOH PCs 2 and 5) likely capture SFOH topics such as healthcare discrimination, social support, neighborhood disorder, religious attendance, religious spirituality, and perceived stress (Fig. 2).

We then followed this up with regressing each of MC axes 1:20 on genetic PCs 1:10. Genetic PCs explained varying amounts of variation in MC structure with substantially stronger associations than observed for SFOH PCs. All twenty MCs showed statistically significant associations with genetic PCs, with R^2 values ranging from 0.001 (MC12) to 0.076 (MC3) (Table S10). Several MCs demonstrated notable strong genetic associations, including the aforementioned MC3, MC4 ($R^2 = 0.047$), MC1 ($R^2 = 0.044$), and MC16 ($R^2 = 0.035$). The consistent statistical significance across all components suggests broad genetic influence on the representation of SFOH measures. These axes likely capture SFOH topics such as healthcare discrimination, loneliness, perceived stress, and neighborhood environment (Fig. 3).

2.5.2 Race-stratified analyses revealed weaker associations, possibly influenced by sample size

When we repeated these analyses separately in Black/African American and White individuals, we observed similar trends as in the combined sample, but the associations were generally weaker. In both Black/African American and White subgroups, genetic PCs explained little of the variation in SFOH PCs ($R^2 < 0.015$), though statistical significance was observed more frequently in the White group (Table S12, S14). Similarly, SFOH PCs explained only a small fraction of variance in genetic PCs, with R^2 values ranging from 0.001 to 0.036, again with more significant associations observed in the White group (Table S13). In the Black/African American group, associations were weaker and more diffuse (Table S11).

Examining how well MC axes predict genetic PCs in the race groups we found modest associations in both the Black/African American and White groups. a linear combination of MC axes explained modest amount of variation of genetic PCs for

both the Black/African American and White groups (genetic PC1 $R^2 = 0.032$ and genetic PC3 $R^2 = 0.042$, respectively). Statistical significance was observed more consistently in the White group where all ten genetic PCs showed significant associations with MC axes (Table S15), compared to only five of ten in the Black/African American group (Table S16). Genetic PCs explained limited variation in MC axes in both groups, with maximum R^2 values of 0.008 in Black/African American individuals (MC2) and 0.017 in White individuals (MC4). Again, statistical significance was observed more frequently in the White group (Table S17), where 19 of 20 MC axes showed significant associations with genetic PCs, compared to only 7 of 20 in the Black/African American group (Table S18).

We note that the sample sizes of the two groups differ substantially, with the Black/African American group being an order of magnitude smaller than the White group (Table S1). This may have contributed to reduced power and greater estimate variability in the stratified analyses.

2.5.3 Sex-stratified analyses show consistent patterns with slightly stronger effects in females

When we examined the sex-stratified groups, we observed generally similar patterns between males and females, but with some differences in strengths of associations. In both male and female groups, genetic PCs explained modest amounts of variation in SFOH PCs, with maximum R^2 values of 0.050 (SFOHPC5) in males (Table S20), and 0.068 (SFOHPC5) in females (Table S19), indicating slightly genetic influence on social structure in females.

Similarly, SFOH PCs explained comparable amounts of variance in genetic PCs in both sexes, with females showing slightly stronger associations for the primary genetic PC (GENPC1 $R^2 = 0.195$ in females vs. 0.163 in males). Both groups showed consistently significant associations across all genetic PCs, though female estimates (Table S21) generally higher than those of males (Table S22).

Examining how well MC axes predict genetic PCs, we found that a linear combination of MC axes explained substantial variation in genetic PC1 for both females ($R^2 = 0.216$) and males ($R^2 = 0.188$). The overall pattern of associations was consistent between sexes, with all ten genetic PCs showing significant associations in both groups (Tables S23, S24).

Genetic PCs explained variation in MC axes with similar patterns across sexes, though females (Table S25) generally showed stronger associations than males (Table S26). The maximum R^2 values were 0.087 (MC3) in females compared to 0.057 (MC3) in males. Both groups showed widespread statistical significance, with all 20 MC axes reaching significance in females and 19 out of 20 in males. The consistent patterns across both sexes suggest that the genetic architecture captured by SFOH measures is largely shared between males and females, though with slightly stronger effects in females. This distinction could be due to difference in statistical power, while not as imbalanced as the race-stratified group, females outnumber males in this specific dataset.

2.5.4 Genetic PCs and SFOH measures constitute largely distinct sources of variation

These findings suggest that while SFOH and genetic PCs are statistically associated—particularly where social and ancestral structures align—the practical overlap

is limited. Most of the variation in each domain is not captured by the other. This underscores the importance of treating social and genetic factors as distinct sources of variation in analyses of complex traits and reinforces the need to measure environmental exposures directly, rather than relying on race as a proxy. Further, this apparent distinction between domains supports the use of both sets of PCs in downstream analyses.

3 Methods

3.1 All of Us Study

The All of Us Research Program is a major longitudinal study designed to enhance biomedical research and improve health outcomes by enrolling at least one million diverse participants across the USA. This initiative aims to address historical gaps in research by including groups that have been underrepresented in biomedical studies. The program recently released clinical-grade genome sequences for 312,924 participants, with a notable emphasis on diversity—77% of these participants come from historically under-represented groups, and 46% are from racial and ethnic minorities. To characterize the genetic ancestry structure of this diverse cohort, we performed PCA and present the population centroids for PC1 and PC2 (Fig. S1). For a detailed visualization of the genetic PC structure of the All of Us dataset, we refer readers to Figures 1 and 2 in [12]. The dataset includes over one billion genetic variants and has facilitated the identification of genetic associations with various diseases. Data is accessible to researchers via the All of Us Researcher Workbench, which allows for broad scientific use while ensuring participant privacy. The program collects a wide range of health data, including electronic health records, physical measurements, and survey responses, in addition to genetic information. This extensive dataset is expected to advance the field of genomic medicine by providing insights into the genetic basis of health and disease across diverse populations.

The Social Factors of Health (SFOH) survey (previously known as the Social Determinants of Health Survey), developed by the All of Us Research Program, aims to gather data on social factors influencing health. The survey was designed to be valid, reliable, and relevant to precision medicine, capturing dimensions of complex social factors such as income, education, and housing quality. Out of 397,732 eligible participants, 117,783 (29.6%) completed the optional survey by June 30, 2022. Of these, 94,030 had array data available. Among those who completed the SFOH survey, 78.4% were from underrepresented populations in biomedical research. The survey was available in English and Spanish, and was designed to be completed in about 10-15 minutes. Participants who responded were generally more likely to be White and female.

3.2 Quality Control

Genotyping was performed using a custom Illumina array built on the design framework of the Global Screening Array. The genome build used is HG38. Please refer to the All of Us Research Program technical documentation and data release notes for further details. Variant filtering was conducted using PLINK 2.0 [13] on the All of Us version 7 controlled-tier array data, restricted to autosomal SNPs. This reassigned variant identifiers and removed all duplicate variants. To mitigate the effects of linkage disequilibrium, we pruned markers using a sliding window approach (200

variants per window, with a step size of 50 variants) and an r^2 threshold of 0.2. Markers with a minor allele frequency below 0.01 were excluded, as were variants with a genotype missingness rate exceeding 1%, markers deviating from Hardy-Weinberg equilibrium ($p < 1x10^{-6}$), and individuals with more than 2% missing genotype calls. We assessed relatedness up to the third degree using KING [14] (kinship coefficient ≥ 0.0625), and removed one member from each related pair. After quality control, we retained 285,123 unrelated individuals and 56,961 high-quality SNPs for genetic PC analysis (Fig. S2).

3.3 Analysis Cohort Selection

We included individuals who had genetic PCs available, completed the SFOH survey, and had self-reported male or female sex at birth, and who had reported a self-identified racial identity. To maximize sample size, we included individuals identifying as "Hispanic or Latino". The final full cohort comprised 85,963 individuals who had genetic data available and who had completed the SFOH survey.

3.4 Trait Processing

We selected 18 clinically relevant anthropometric and metabolic traits known to demonstrate population-level differences across racial groups, enabling us to investigate how additional social factors might contribute to these patterns. These traits include: alanine aminotransferase (AAT), alkaline phosphatase (AlkPhos), aspartate aminotransferase (AspAminTrans), body mass index (BMI), carbon dioxide (CO2), chloride (Cl), cholesterol in HDL (CholHDL), creatinine (Creat), diastolic blood pressure (DBP), glucose, heart rate (HR), hematocrit, hip circumference (HipCirc), protein, sodium, systolic blood pressure (SBP), urea nitrogen (UreaNitro), and waist circumference (WC).

Following trait selection, we performed data cleaning and deduplication procedures prior to downstream analyses. For individuals with multiple measurements, we retained the most recent value. To mitigate the influence of extreme values, we calculated the IQR for each biomarker independently, and observations falling outside these bounds were truncated to the corresponding threshold values. Systolic blood pressure was adjusted for antihypertensive medication using a binary indicator variable. Sample sizes varied by trait and race due to differential measurement availability across participants.

3.5 Estimating Variance Contribution by Genome-wide Complex Trait Analysis (GCTA)

We estimated h_g^2 using both HE regression and GREML as implemented in the Genome-wide Complex Trait Analysis software (GCTA [11]). For our primary inference, we report HE-CP point estimates from HE regression and standard errors (SE) calculated by jackknifing over individuals. 95% confidence intervals (CIs) were constructed as $HE-CP \pm 1.96(SE)$. For comparison and robustness checking, we also computed estimates utilizing GREML, which partitions phenotypic variance (Vp) into additive genetic variance ($V(G)$) and residual variance ($V(e)$) through a linear mixed model, providing direct estimates of $h_g^2 = V(G)/Vp$ with associated SEs. To evaluate whether SFOH covariates alter heritability estimates, and whether SFOH covariates themselves may be heritable, we compared the 95% CIs from models with

and without their inclusion. Non-overlapping CIs were interpreted as evidence of a significant difference in heritability estimates between models, which provides a conservative test.

3.6 Estimating Dispersion Components with Double Generalized Models (dglm)

We employed double generalized linear models (dglms) to investigate the effect of SFOH covariates on both the mean and the variance of traits. Dglms extend traditional generalized linear models by simultaneously modeling both the mean and the dispersion components of the response variable, allowing us to test whether the variance of the trait differs across levels of predictor variables. We implemented the ‘dglm‘ package in R with a Gaussian distribution for the mean model and the default log link for the dispersion model. We fit the following mean and variance models for 18 traits:

$$\text{Mean: } E(Y) = \beta_0 + \beta_1\text{Age} + \beta_2\text{Sex} + \beta_3\text{Race} + \beta_4\text{Ethnicity} + \sum \beta_i\text{SFOH}$$

$$\text{Dispersion: } \log(\text{Var}(Y)) = \gamma_0 + \gamma_1\text{Age} + \gamma_2\text{Sex} + \gamma_3\text{Race} + \gamma_4\text{Ethnicity} + \sum \gamma_i\text{SFOH}$$

Where $SFOH$ represented either the first 20 SFOH PCs, the first 20 MC axes, or the SSI analyzed in separate models using the default maximum likelihood estimation method. To account for multiple comparisons, we applied a hierarchical Bonferroni correction with four tiers of stringency. Each test was assigned a significance level based on the most stringent threshold it passed:

1. **Overall**: $p < 0.05$ / (total number of tests across all traits and SFOH measures)
2. **Family-wise**: $p < 0.05$ / (number of tests within the SFOH measure)
3. **Trait-wise**: $p < 0.05$ / (number of predictors per trait)
4. **Nominal**: $p < 0.05$ (uncorrected)

We then labeled predictors according to the highest significance level achieved (overall, family-wise, trait-wise, or nominal).

3.7 Summarizations of the SFOH Survey

3.7.1 Survey Preparation

Survey items were organized into themes (i.e. conceptual domains grouping related questions, such as “perceived stress,” “social support,” “loneliness,” etc.) based on the SFOH Task Force’s consensus-based conceptual frameworks. Missing responses coded as “PMI: Skip”, “PMI: Don’t Know”, “PMI: Not Sure”, “PMI: Other” or “PMI: Prefer Not To Answer” were excluded. At the question-level, missingness was very low (less than 5%) for nearly all themed items(Table S6), with only two unthemed “orphan” questions exhibiting high skip rates. Theme-level missingness was similar across race groups (2–8%) (Table S4), and at the individual level the median per-participant missingness was approximately 3%, with less than 1% of participants missing more than 20% of items. We used all available survey completers to generate SFOH measure measures (PCA, MCA, and SSI).

3.7.2 PCA & MCA

To capture the multidimensional nature of SFOH, we applied PCA and MCA dimensionality reduction techniques to one-hot encoded survey data. We performed PCA

in R using the prcomp function, and MCA in Python using the prince package, which is specifically optimized for categorical data analysis. We merged the resulting component scores from both analyses into a single analysis dataset using unique subject identifiers. The leading SFOH components (PCs from PCA and dimensions from MCA) capture the major axes of social-factor heterogeneity in our cohort and were included as linear covariates in downstream analyses. SFOH PCs and MCA axes were calculated from all individuals with SFOH survey data (n=117,783).

3.7.3 Environmental Relatedness Matrix (ERM)

We constructed a novel environmental relatedness matrix (ERM) from the same one-hot encoded SFOH survey data. The ERM is conceptually analogous to a GRM, here it is derived entirely from SFOH survey data. In this approach, every individual's binary response vector forms a row of the data matrix

$$X \in \{0,1\}^{n \times p},$$

where n is the number of individuals and p is the total number of response-level columns.

To ensure that each question contributes equally regardless of how many response levels it has, we form a diagonal weight matrix

$$W = \mathrm{diag}(w_1, \ldots, w_p),$$

with

$$w_\ell = \frac{1}{\left|\{\text{levels of question corresponding to column } \ell\}\right|}.$$

We define the weighted design matrix

$$X_w = X\,W,$$

so that

$$(X_w)_{i\ell} = w_\ell\, X_{i\ell}.$$

To account for missing responses, we define a binary mask matrix

$$M \in \{0,1\}^{n \times p},$$

by setting

$$M_{i\ell} = \begin{cases} 1, & \text{if individual } i \text{ answered level column } \ell, \\ 0, & \text{otherwise,} \end{cases}$$

and its weighted version

$$M_w = M\,W.$$

All matrix operations are carried out on the sparse weighted matrices X_w and M_w. Although $\mathrm{ERM}_{ii} = 1$ holds by definition, we never explicitly compute or store diagonal entries.

3.7.4 Social Similarity Index (SSI)

Rather than computing or storing the full $n \times n$ ERM, we summarize each individual's social similarity to the cohort using the ERM row sums. We define each row sum as

$$s_i = \sum_{j=1}^{n} \mathrm{ERM}_{ij}.$$

We compute the row sum in $O(n\,p)$ time by first defining the two vectors

$$c = X_w^\top \mathbf{1}_n, \qquad d = M^\top \mathbf{1}_n,$$

Here, d gives the unweighted sum of answered items per column. We then compute the individual-level weighted sums

$$u = X_w\, c, \qquad v = M_w\, d.$$

Each individual's SSI is then

$$s_i = \frac{u_i}{v_i}, \quad i = 1, \ldots, n,$$

and we set any indeterminate 0/0 to zero (so that individuals with no overlapping responses are treated as unrelated). We standardize the SSI to have mean zero and unit variance. The standardized SSI captures each individual's overall social similarity to the cohort which we include in downstream regression models to account for shared social context.

4 Discussion

Our results demonstrate that SFOH structure as summarized by dimension-reduced axes (via MCA and PCA) reduces heritability estimates obtained from HE regression across multiple traits. However, a novel pairwise similarity measure (the SSI) does not. In the full cohort and White individuals, including SFOH covariates significantly decreased heritability estimates for 4 of 18 traits (BMI, hip circumference, waist circumference, and HDL cholesterol) relative to models with only basic demographic factors, but not for the other traits. In female and male cohorts, including SFOH covariates significantly reduced heritability estimates for 3 of 18 traits (BMI, hip circumference, and waist circumference) relative to models with only basic demographic factors. Heritability estimates in Black/African American individuals did not change significantly after the inclusion of SDOH summaries, possibly due to the smaller sample size as compared to other analyzed groups. Including genetic PCs consistently reduced heritability estimates in all groups, indicating the presence of population structure and underscoring the necessity of including genetic PCs when SFOH measures are included.

The observed reductions in heritability with SFOH covariates in HE regression suggest that these axes capture social variation that may be partially correlated with genetic similarity within the sample. Because HE regression attributes any trait covariance aligned with the GRM to genetic variance, unmodeled social structure can inflate heritability estimates. Adjusting for SFOH measures removes this shared social variance, leading to lower h^2 estimates even in the absence of genetic PCs. In contrast, our GREML analyses demonstrated that heritability estimates were

robust to the inclusion of either SFOH covariates or genetic PCs, highlighting a fundamental methodological difference between the two approaches.

We note that sample size differences and our methodological usage of self-identified race have important potential consequences. The size imbalance between racial groups is a major limitation of this study: the Black/African American cohort (n=6,538) is an order of magnitude smaller than the White cohort (n=67,545). The absence of significant SFOH covariate effects in the Black/African American cohort likely reflects insufficient power rather than a true absence of SFOH influence on heritability. In addition, between-group comparisons are not meaningful because true population differences are indistinguishable from induced statistical artifacts.

While we stratified by self-identified race, we acknowledge that race functions as a social construct that imperfectly captures both genetic ancestry and social experiences. Our stratified analyses risk conflating genetic background with socially mediated SFOH exposures, and our findings should not be interpreted as evidence of biological differences between racial groups.

We found that the sets of SFOH PCs (4-20) and MC axes (1-6) significantly reduced heritability estimates for several traits in HE regression. Within these sets, we identified SFOH PC5, PC6, and MC5 as the most consistently heritable components across demographic strata. In contrast, the SSI, while also heritable, did not decrease heritability estimates. Furthermore, these same heritable components (PC5, PC6, MC5, SSI) exhibited widespread and significant, if small, effects on trait dispersion even when adjusting for demographic factors, indicating that they capture social environment factors that influence trait variance. Specifically, SFOH PC5 appears to capture information on healthcare discrimination, social support, perceived stress, and religious spirituality and attendance. SFOH PC6 also captures religious spirituality and attendance and perceived stress as well as neighborhood environment. MC5 captures primarily information on religious spirituality and social support, while the SSI is a composite score and captures information on all categories of the SFOH survey. The convergence of these findings suggest that the decreases in trait heritability estimates may not be a general effect of social factors, but may be specifically attributable to a subset of SFOH measures that are themselves heritable. This highlights the importance of measuring and adjusting for these specific social environmental factors in genetic studies to avoid missattributing social environmental influences to genetic effects.

Our findings reveal that specific SFOH components are heritable and are contained within the larger set of components associated with changes in heritability estimates. However, our study design cannot disentangle the causal mechanisms underlying this observation. The additive genetic variance captured by these SFOH measures could arise from several, non-exclusive processes, including gene-environment correlation, social stratification, and environmental confounding.

We acknowledge several methodological considerations in interpreting dglm results. The comprehensive dispersion models, while revealing of significant patterns, are complex with numerous parameters that may risk overfitting to our cohort. The inclusion of up to 33 predictors in both mean and dispersion components, while justified by our research question, may capture cohort-specific patterns that do not generalize to other populations. Our primary aim was to comprehensively characterize how these specific SFOH measures influence trait variance in our dataset. These results should be viewed as generating hypotheses about

SFOH-biological interactions in trait variance rather than providing definitive evidence.

While our study employed standard PCA on one-hot encoded indicators as a practical dimensionality reduction technique, we acknowledge that this approach treats binary data as continuous. A more theoretically grounded approach for ordinal survey data might involve polychoric PCA. However, the components derived from our method demonstrably captured meaningful SFOH gradients, validating their use. Furthermore, the use of standard PCA enabled a direct and consistent comparison with MCA, which is specifically designed for categorical data. Future work could explore the sensitivity of these findings to alternative dimensionality reduction techniques tailored for ordinal data.

These results highlight the influence of both genetic and social structures on trait heritability and demonstrate how accounting for social factors can alter trait-specific heritability estimates. This work underscores the need for population-specific heritability estimates and will be a focus of future work to disentangle the roles of genetic structure, exposure to social environments, and trait variation.

5 Data availability

Individual-level genotype and survey data used in this study are available to registered researchers through the All of Us Researcher Workbench (https://workbench.researchallofus.org/) following Data and Research Center (DRC) access procedures. Genomic data (v7 array release) and survey responses, including the Social Factors of Health (SFOH) survey, are part of the All of Us [15] controlled tier dataset. Other information is available upon request.

6 Code availability

All analyses were performed using previously published or developed tools, as indicated in the Methods.

Acknowledgments. We thank the All of Us Research Program and its participants. We thank Lily Hamilton for thoughtful feedback that led to the improvement of this manuscript. This research was supported by NHGRI CEGS RM1: RM1HG010461 (NSA) and NIH NCI U54CA267738 (SAM), The Cornell University Graduate School Deans Excellence Fellowship (OSRA), and The Cornell University Graduate School Walter Schonlenk and Leister Fellowships (JLL).

Declarations. The authors declare no competing interests.

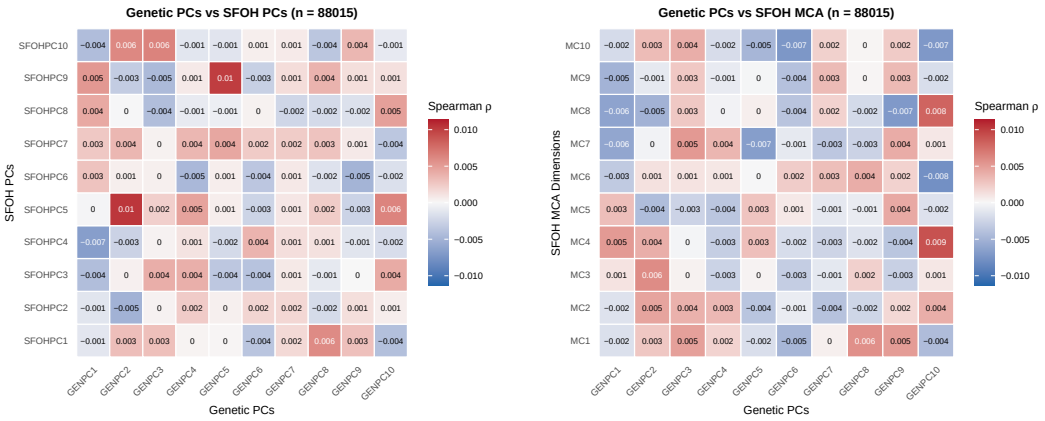

Fig. 1: Relationship between genetic structure and SFOH measures. Panels (a) and (b) display heatmaps showing Spearman correlations between the top 10 genetic PCs and the first 10 SFOH axes derived from PCA and MCA respectively.

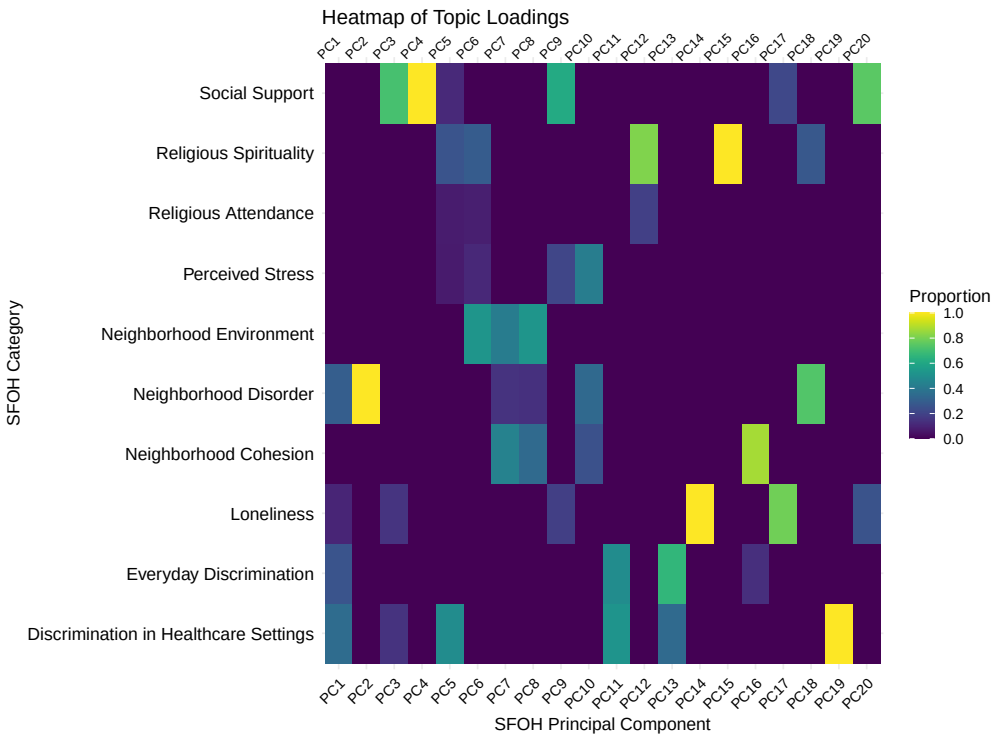

Fig. 2: Heatmap of SFOH topic loadings for SFOH PCs

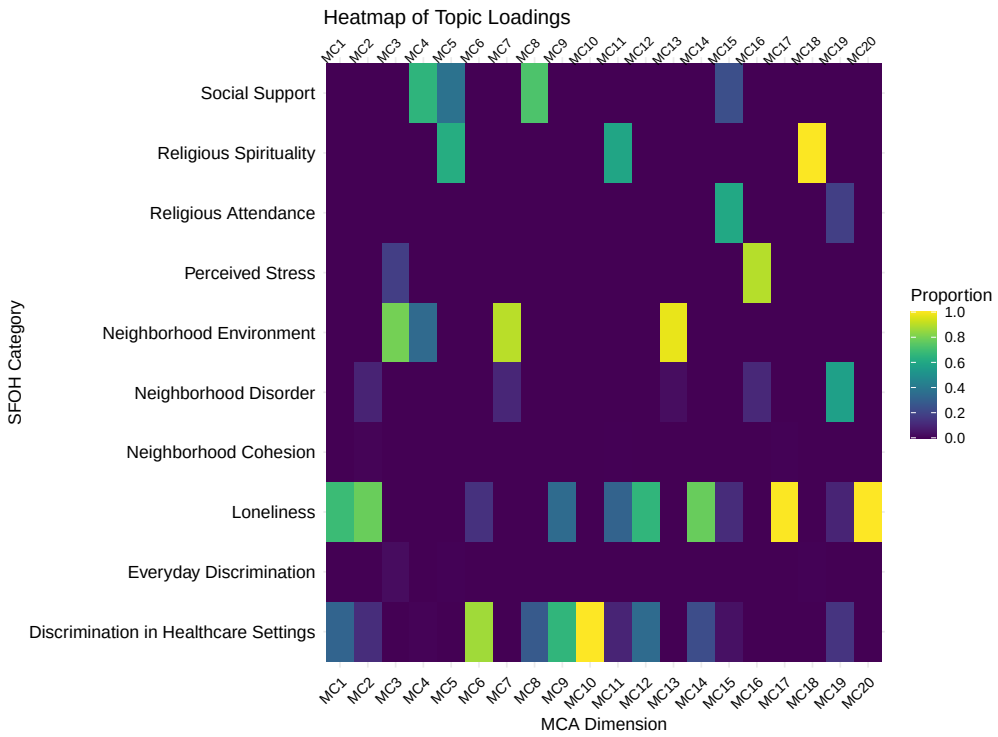

Fig. 3: Heatmap of SFOH topic loadings for SFOH MCA axes.

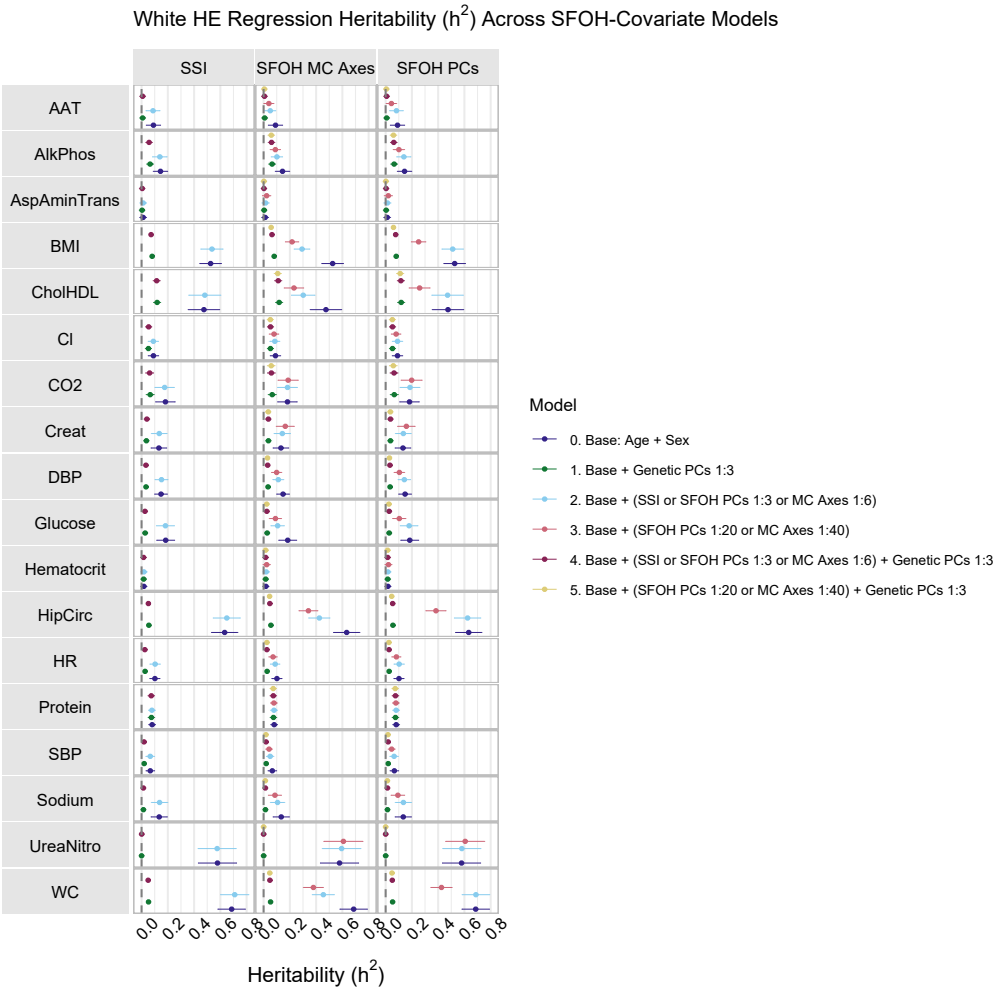

Fig. 4: White individuals: HE regression heritability estimates for 18 traits (gray rows) utilizing one of the following SFOH summarization strategies (gray columns): SSI, which was run with one of four possible models (Models 0, 1, 2, and 4); MCA, run with one of six models (Models 0-5); or PCA, run with one of six models (Models 0-5). Points are HE-CP estimates, intervals are 95% CIs based on HE jackknife standard errors.

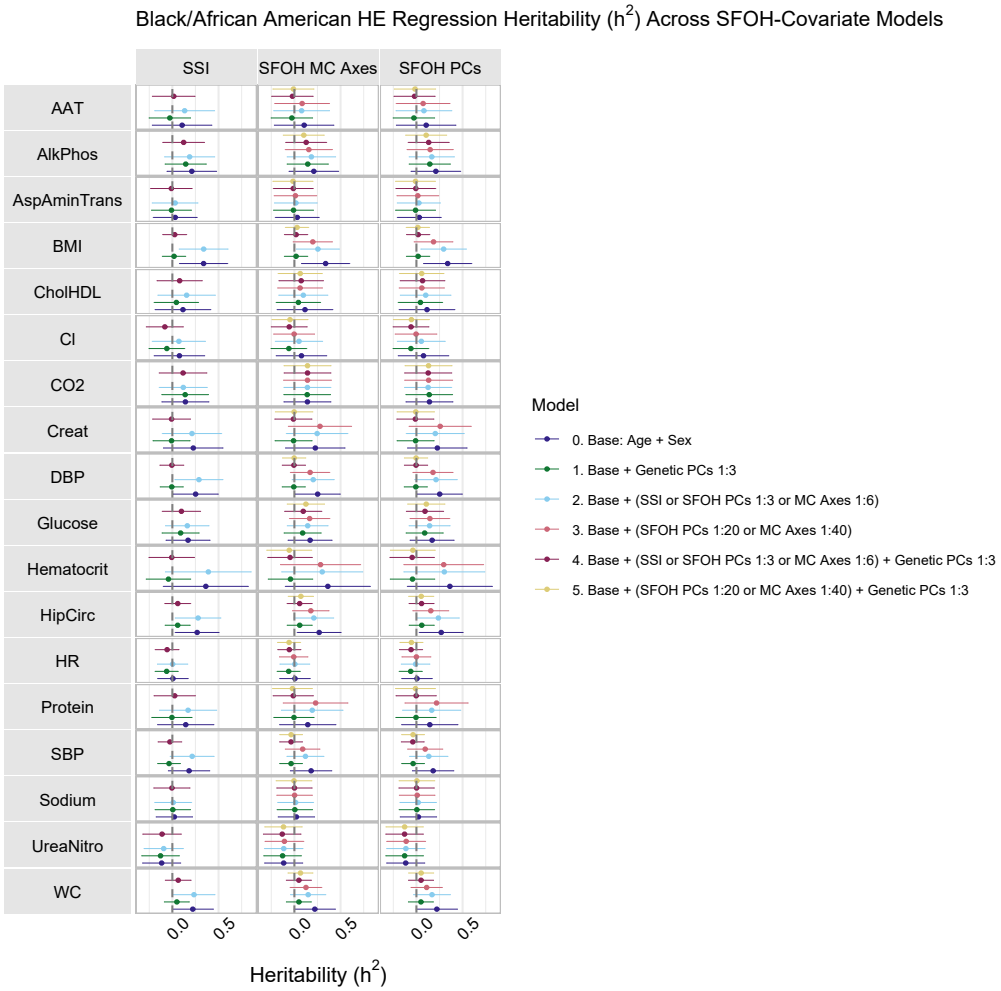

Fig. 5: Black/African American individuals: HE regression heritability estimates for 18 traits (gray rows) utilizing one of the following SFOH summarization strategies (gray columns): SSI, which was run with one of four possible models (Models 0, 1, 2, and 4); MCA, run with one of six models (Models 0-5); or PCA, run with one of six models (Models 0-5). Points are HE-CP estimates, intervals are 95% CIs based on HE jackknife standard errors.

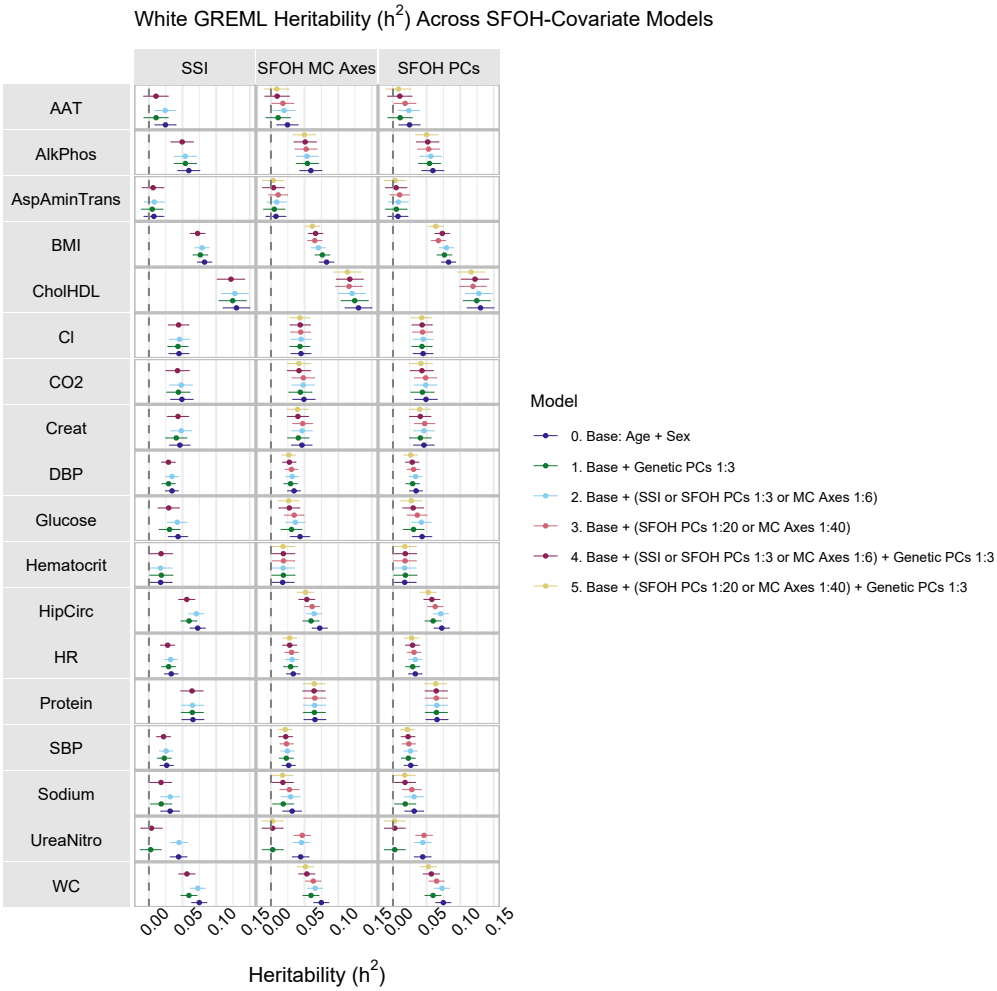

Fig. 6: White individuals: GREML heritability estimates for 18 traits (gray rows) utilizing one of the following SFOH summarization strategies (gray columns): SSI, which was run with one of four possible models (Models 0, 1, 2, and 4); MCA, run with one of six models (Models 0-5); or PCA, run with one of six models (Models 0-5). Points are $V(G)/Vp$ estimates, intervals are 95% CIs based on standard errors.

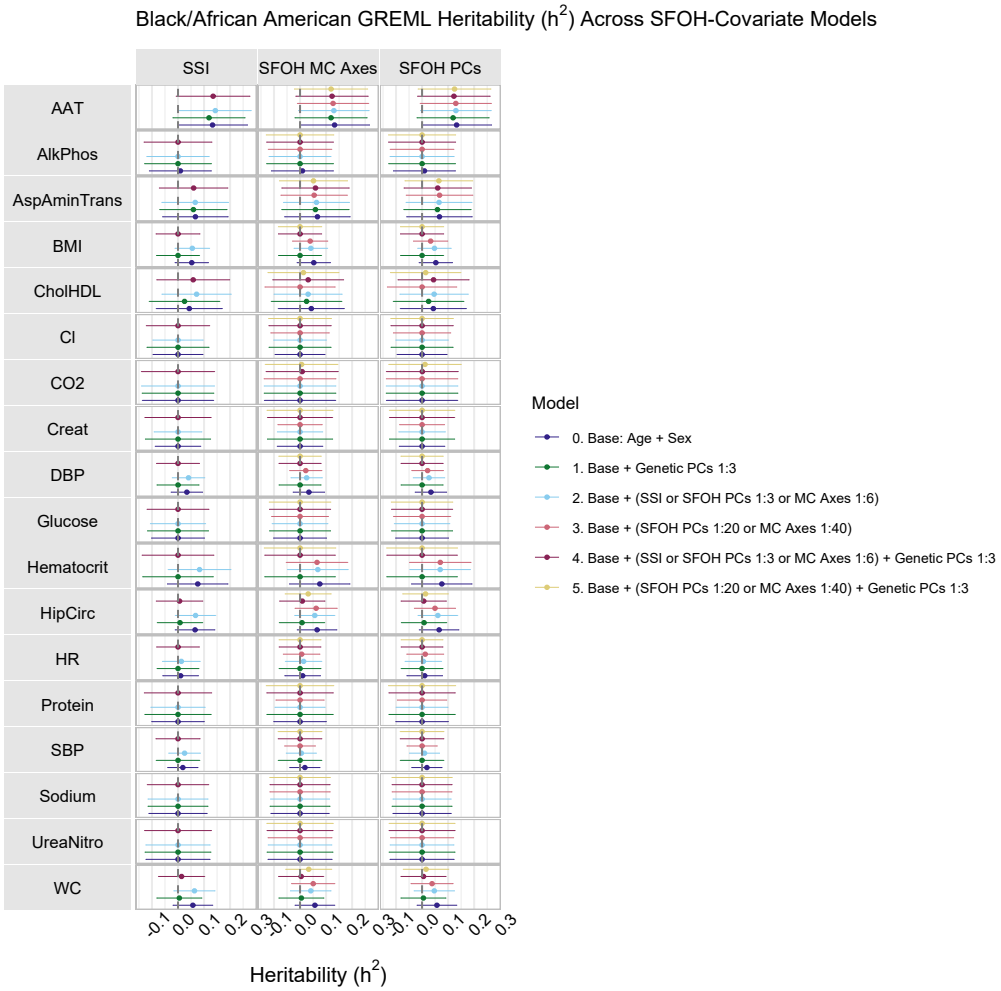

Fig. 7: Black/African American individuals: GREML heritability estimates for 18 traits (gray rows) utilizing one of the following SFOH summarization strategies (gray columns): SSI, which was run with one of four possible models (Models 0, 1, 2, and 4); MCA, run with one of six models (Models 0-5); or PCA, run with one of six models (Models 0-5). Points are $V(G)/Vp$ estimates, intervals are 95% CIs based on standard errors.

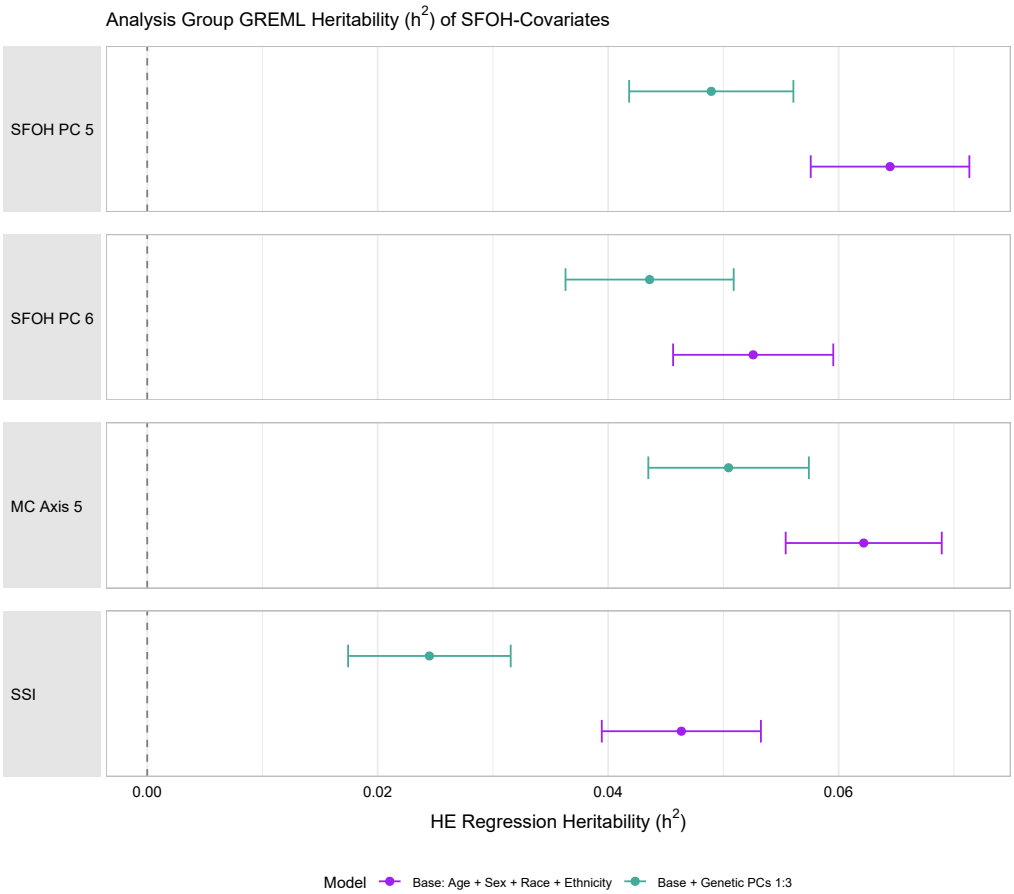

Fig. 8: Group REML Heritability of SFOH measures.

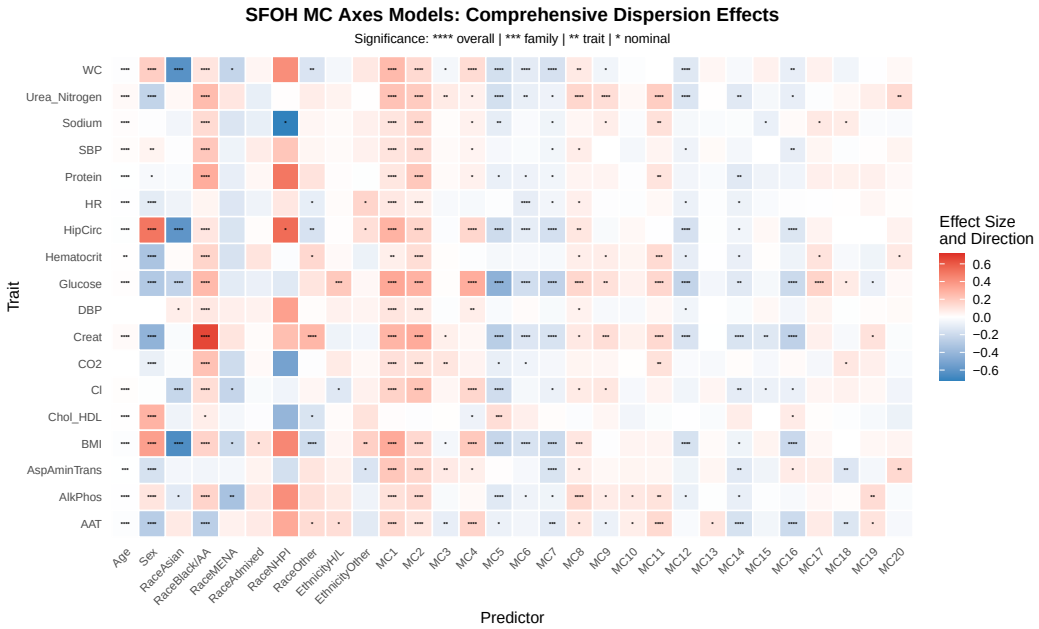

Fig. 9: Heatmap of significant predictors in trait dispersion models with SFOH MC axes.

References

[1] The "All of Us" Research Program | New England Journal of Medicine.

[2] Hailan Liu and Guo-Bo Chen. A novel genomic prediction method combining randomized Haseman-Elston regression with a modified algorithm for Proven and Young for large genomic data. *The Crop Journal*, 10(2):550–554, April 2022. ISSN 2214-5141. doi: 10.1016/j.cj.2021.09.001.

[3] Ali Pazokitoroudi, Zhengtong Liu, Andrew Dahl, Noah Zaitlen, Saharon Rosset, and Sriram Sankararaman. A scalable and robust variance components method reveals insights into the architecture of gene-environment interactions underlying complex traits. *The American Journal of Human Genetics*, 111(7): 1462–1480, July 2024. ISSN 0002-9297, 1537-6605. doi: 10.1016/j.ajhg.2024.05.015. Publisher: Elsevier.

[4] Suresh K. Bhavnani, Weibin Zhang, Daniel Bao, Mukaila Raji, Veronica Ajewole, Rodney Hunter, Yong-Fang Kuo, Susanne Schmidt, Monique R. Pappadis, Elise Smith, Alex Bokov, Timothy Reistetter, Shyam Visweswaran, and Brian Downer. Subtyping Social Determinants of Health in All of Us: Network Analysis and Visualization Approach. *medRxiv*, page 2023.01.27.23285125, August 2023. doi: 10.1101/2023.01.27.23285125.

[5] Sonali Gupta, Vincent Lam, I. King Jordan, and Leonardo Mariño-Ramírez. A composite socioeconomic deprivation index from All of Us survey data: associations with health outcomes and disparities. *medRxiv*, page 2024.10.04.24314904, October 2024. doi: 10.1101/2024.10.04.24314904.

[6] Jane Brown, Abhijith Biji, Kathleen Ferar, Vikas Pejaver, Eimear E. Kenny, Bian Liu, and Samira Asgari. A Scalable Framework to Integrate Social Determinants of Health into Disease Risk Models using Biobank Survey Data, September 2025. ISSN: 3067-2007 Pages: 2025.09.29.25336903.

[7] Stefania Benonisdottir and Augustine Kong. Studying the genetics of participation using footprints left on the ascertained genotypes. *Nature Genetics*, 55(8): 1413–1420, August 2023. ISSN 1546-1718. doi: 10.1038/s41588-023-01439-2. Publisher: Nature Publishing Group.

[8] Gianmarco Mignogna, Caitlin E. Carey, Robbee Wedow, Nikolas Baya, Mattia Cordioli, Nicola Pirastu, Rino Bellocco, Kathryn Fiuza Malerbi, Michel G. Nivard, Benjamin M. Neale, Raymond K. Walters, and Andrea Ganna. Patterns of item nonresponse behaviour to survey questionnaires are systematic and associated with genetic loci. *Nature Human Behaviour*, 7(8):1371–1387, August 2023. ISSN 2397-3374. doi: 10.1038/s41562-023-01632-7.

[9] Cole Brokamp, Andrew F. Beck, Neera K. Goyal, Patrick Ryan, James M. Greenberg, and Eric S. Hall. Material community deprivation and hospital utilization during the first year of life: an urban population–based cohort study. *Annals of Epidemiology*, 30:37–43, February 2019. ISSN 1047-2797. doi: 10.1016/j.annepidem.2018.11.008.

[10] Arun Durvasula and Alkes L. Price. Distinct explanations underlie gene-environment interactions in the UK Biobank. *The American Journal of Human Genetics*, 112(3):644–658, March 2025. ISSN 0002-9297. doi: 10.1016/j.ajhg.2025.01.014. Publisher: Elsevier.

[11] GCTA: A Tool for Genome-wide Complex Trait Analysis - PMC.

[12] Mateus H. Gouveia, Karlijn A. C. Meeks, Victor Borda, Thiago P. Leal, Fernanda S. G. Kehdy, Reagan Mogire, Ayo P. Doumatey, Eduardo Tarazona-Santos, Rick A. Kittles, Ignacio F. Mata, Timothy D. O'Connor, Adebowale A. Adeyemo, Daniel Shriner, and Charles N. Rotimi. Subcontinental genetic variation in the All of Us Research Program: Implications for biomedical research. *American Journal of Human Genetics*, 112(6):1286–1301, June 2025. ISSN 1537-6605. doi: 10.1016/j.ajhg.2025.04.012.

[13] Second-generation PLINK: rising to the challenge of larger and richer datasets | GigaScience | Oxford Academic.

[14] Ani Manichaikul, Josyf C. Mychaleckyj, Stephen S. Rich, Kathy Daly, Michèle Sale, and Wei-Min Chen. Robust relationship inference in genome-wide association studies. *Bioinformatics*, 26(22):2867–2873, November 2010. ISSN 1367-4803. doi: 10.1093/bioinformatics/btq559.

[15] The All of Us Research Program Genomics Investigators. Genomic data in the All of Us Research Program. *Nature*, 627(8003):340–346, March 2024. ISSN 1476-4687. doi: 10.1038/s41586-023-06957-x. Publisher: Nature Publishing Group.

Supplemental Material

Table S1: Counts by Self-Identified Race

Category	Count
Black or African American	6513
Asian	2318
Middle Eastern or North African	380
More than one population	1511
Native Hawaiian or Other Pacific Islander	42
White	67164
Other	8035

Table S2: Missingness by "Theme" and Age

Theme	Age Group	Total Responses	Missing	% Missing
Discrimination in Healthcare Settings	<30	29407	570	1.9
Discrimination in Healthcare Settings	30–50	130004	2184	1.7
Discrimination in Healthcare Settings	50–70	241822	6200	2.6
Discrimination in Healthcare Settings	>70	214879	8129	3.8
Everyday Discrimination	<30	37809	583	1.5
Everyday Discrimination	30–50	167148	2408	1.4
Everyday Discrimination	50–70	310914	7221	2.3
Everyday Discrimination	>70	276273	10419	3.8
Food Insecurity	<30	8402	110	1.3
Food Insecurity	30–50	37144	295	0.8
Food Insecurity	50–70	69092	564	0.8
Food Insecurity	>70	61394	653	1.1
Housing Insecurity	<30	4201	80	1.9
Housing Insecurity	30–50	18572	313	1.7
Housing Insecurity	50–70	34546	1045	3.0
Housing Insecurity	>70	30697	1203	3.9
Housing Instability	<30	4201	93	2.2
Housing Instability	30–50	18572	386	2.1
Housing Instability	50–70	34546	1356	3.9
Housing Instability	>70	30697	2759	9.0
Loneliness	<30	33608	432	1.3
Loneliness	30–50	148576	1674	1.1
Loneliness	50–70	276368	5187	1.9
Loneliness	>70	245576	6914	2.8
Neighborhood Cohesion	<30	16804	144	0.9
Neighborhood Cohesion	30–50	74288	725	1.0
Neighborhood Cohesion	50–70	138184	2415	1.7
Neighborhood Cohesion	>70	122788	3943	3.2
Neighborhood Disorder	<30	54613	635	1.2
Neighborhood Disorder	30–50	241436	3797	1.6
Neighborhood Disorder	50–70	449098	12680	2.8
Neighborhood Disorder	>70	399061	17831	4.5
Neighborhood Environment	<30	33608	2004	6.0
Neighborhood Environment	30–50	148576	8286	5.6
Neighborhood Environment	50–70	276368	19252	7.0
Neighborhood Environment	>70	245576	21933	8.9
Perceived Stress	<30	42010	888	2.1
Perceived Stress	30–50	185720	3684	2.0
Perceived Stress	50–70	345460	12895	3.7
Perceived Stress	>70	306970	19652	6.4
Religious Attendance	<30	4201	308	7.3
Religious Attendance	30–50	18572	1472	7.9
Religious Attendance	50–70	34546	4234	12.3
Religious Attendance	>70	30697	4702	15.3
Religious Spirituality	<30	25206	428	1.7
Religious Spirituality	30–50	111432	1406	1.3
Religious Spirituality	50–70	207276	2933	1.4
Religious Spirituality	>70	184182	3751	2.0
Social Support	<30	33608	431	1.3
Social Support	30–50	148576	1612	1.1
Social Support	50–70	276368	5549	2.0
Social Support	>70	245576	8353	3.4

Table S4: Missingness by "Theme" and Race. Rows containing less than 20 individuals removed as per All of Us data usage agreement.

Theme	Race	Total Responses	Missing	% Missing
Discrimination in Healthcare Settings	Black or African American	46081	2079	4.5
Discrimination in Healthcare Settings	Other	67312	2909	4.3
Discrimination in Healthcare Settings	Middle Eastern or North African	2681	88	3.3
Discrimination in Healthcare Settings	White	472815	11513	2.4
Discrimination in Healthcare Settings	Asian	16261	299	1.8
Discrimination in Healthcare Settings	More than one population	10654	187	1.8
Everyday Discrimination	Black or African American	59247	2621	4.4
Everyday Discrimination	Other	86544	3538	4.1
Everyday Discrimination	White	607905	13792	2.3
Everyday Discrimination	Asian	20907	383	1.8
Everyday Discrimination	More than one population	13698	223	1.6
Everyday Discrimination	Middle Eastern or North African	3447	55	1.6
Food Insecurity	Black or African American	13166	274	2.1
Food Insecurity	Other	19232	339	1.8
Food Insecurity	Asian	4646	52	1.1
Food Insecurity	White	135090	927	0.7
Housing Insecurity	Black or African American	6583	387	5.9
Housing Insecurity	Other	9616	518	5.4
Housing Insecurity	Asian	2323	60	2.6
Housing Insecurity	White	67545	1625	2.4
Housing Insecurity	More than one population	1522	29	1.9
Housing Instability	White	67545	3668	5.4
Housing Instability	Other	9616	506	5.3
Housing Instability	Black or African American	6583	259	3.9
Housing Instability	Asian	2323	90	3.9
Housing Instability	More than one population	1522	48	3.2
Loneliness	Black or African American	52664	1865	3.5
Loneliness	Other	76928	2698	3.5
Loneliness	Middle Eastern or North African	3064	53	1.7
Loneliness	White	540360	9128	1.7
Loneliness	Asian	18584	292	1.6
Loneliness	More than one population	12176	162	1.3
Neighborhood Cohesion	Black or African American	26332	833	3.2
Neighborhood Cohesion	Other	38464	1155	3.0
Neighborhood Cohesion	White	270180	5063	1.9
Neighborhood Cohesion	Middle Eastern or North African	1532	22	1.4
Neighborhood Cohesion	More than one population	6088	65	1.1
Neighborhood Cohesion	Asian	9292	81	0.9
Neighborhood Disorder	Other	125008	5532	4.4
Neighborhood Disorder	Black or African American	85579	3776	4.4
Neighborhood Disorder	White	878085	24622	2.8
Neighborhood Disorder	Middle Eastern or North African	4979	131	2.6
Neighborhood Disorder	Asian	30199	542	1.8
Neighborhood Disorder	More than one population	19786	319	1.6
Neighborhood Environment	Black or African American	52664	4794	9.1
Neighborhood Environment	Other	76928	6401	8.3
Neighborhood Environment	Middle Eastern or North African	3064	220	7.2
Neighborhood Environment	White	540360	38320	7.1
Neighborhood Environment	More than one population	12176	726	6.0
Neighborhood Environment	Asian	18584	982	5.3
Perceived Stress	Black or African American	65830	3945	6.0
Perceived Stress	Other	96160	5316	5.5
Perceived Stress	Middle Eastern or North African	3830	173	4.5
Perceived Stress	White	675450	26568	3.9
Perceived Stress	Asian	23230	728	3.1
Perceived Stress	More than one population	15220	369	2.4
Religious Attendance	Black or African American	6583	942	14.3
Religious Attendance	White	67545	8405	12.4
Religious Attendance	Middle Eastern or North African	383	46	12.0
Religious Attendance	Other	9616	1039	10.8
Religious Attendance	More than one population	1522	114	7.5
Religious Attendance	Asian	2323	165	7.1
Religious Spirituality	Black or African American	39498	1127	2.9
Religious Spirituality	Other	57696	1502	2.6
Religious Spirituality	Middle Eastern or North African	2298	58	2.5
Religious Spirituality	White	405270	5522	1.4
Religious Spirituality	More than one population	9132	120	1.3
Religious Spirituality	Asian	13938	180	1.3
Social Support	Black or African American	52664	1987	3.8
Social Support	Other	76928	2744	3.6
Social Support	Middle Eastern or North African	3064	94	3.1
Social Support	White	540360	10638	2.0
Social Support	Asian	18584	333	1.8

Table S6: Overall Missingness by Theme

Theme	Total Responses	Missing	% Missing
Religious Attendance	88016	10716	12.2
Neighborhood Environment	704128	51475	7.3
Housing Instability	88016	4594	5.2
Perceived Stress	880160	37119	4.2
Neighborhood Disorder	1144208	34943	3.1
Housing Insecurity	88016	2641	3.0
Discrimination in Healthcare Settings	616112	17083	2.8
Everyday Discrimination	792144	20631	2.6
Social Support	704128	15945	2.3
Neighborhood Cohesion	352064	7227	2.1
Loneliness	704128	14207	2.0
Religious Spirituality	528096	8518	1.6
Food Insecurity	176032	1622	0.9

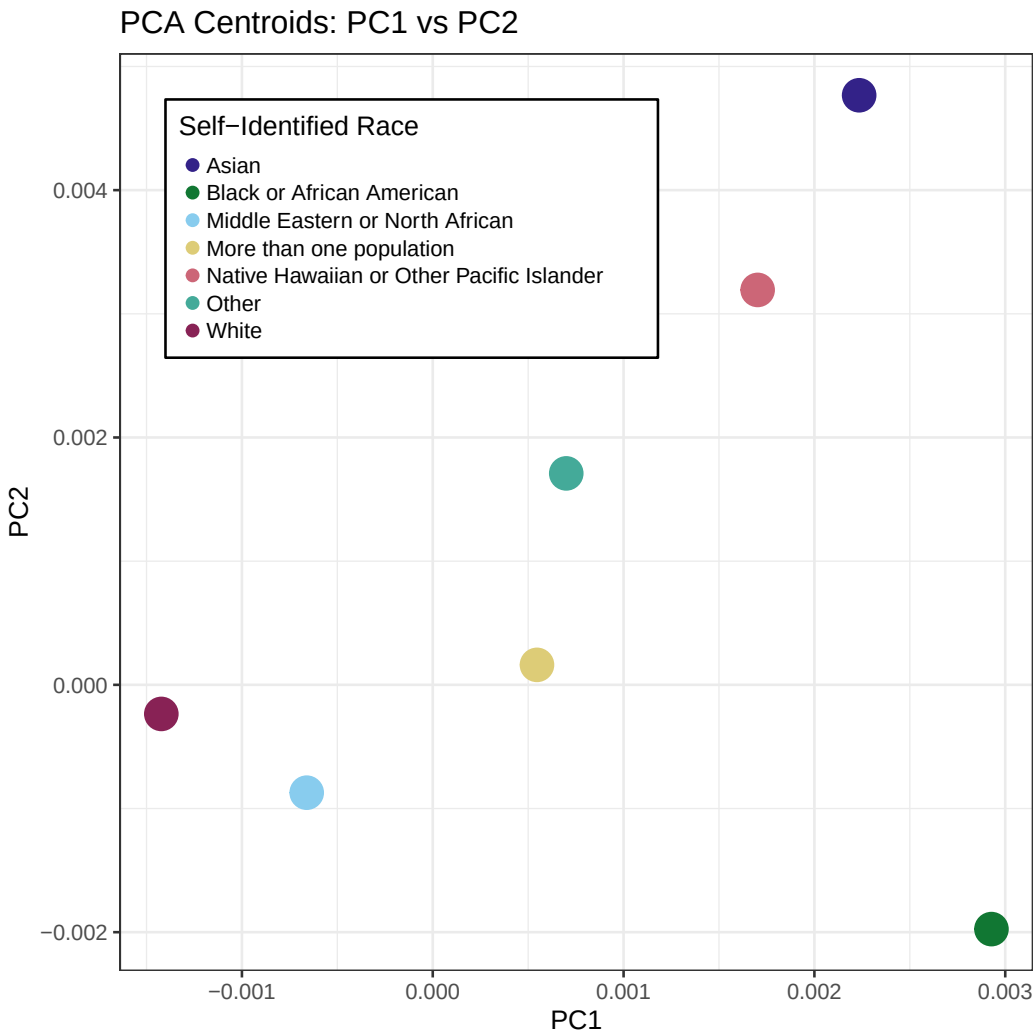

Fig. S1: Population centroids from PCA of the All of Us genetic data. Points represent mean PC1 and PC2 values for each self-identified racial group, summarizing the overall genetic ancestry structure.

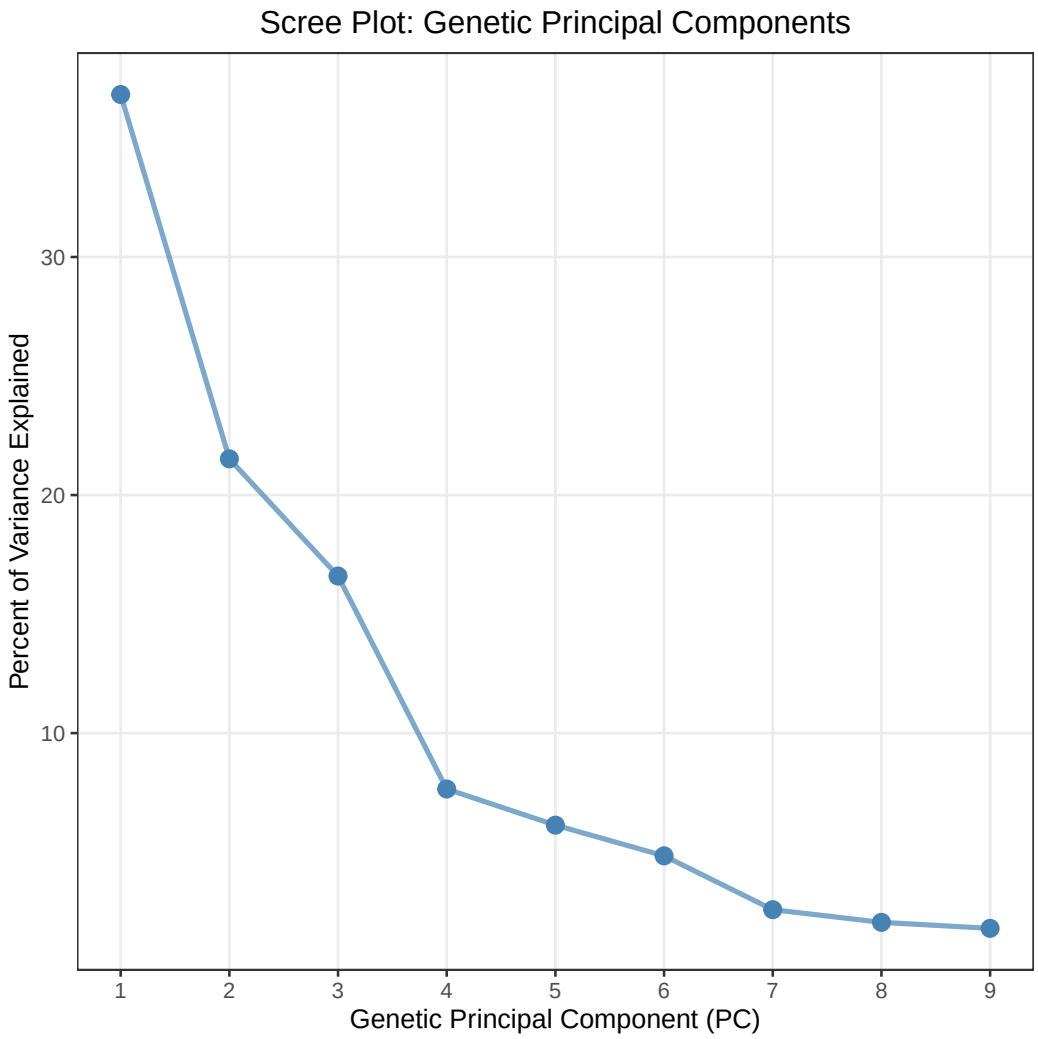

Fig. S2: Genetic PCA scree plot

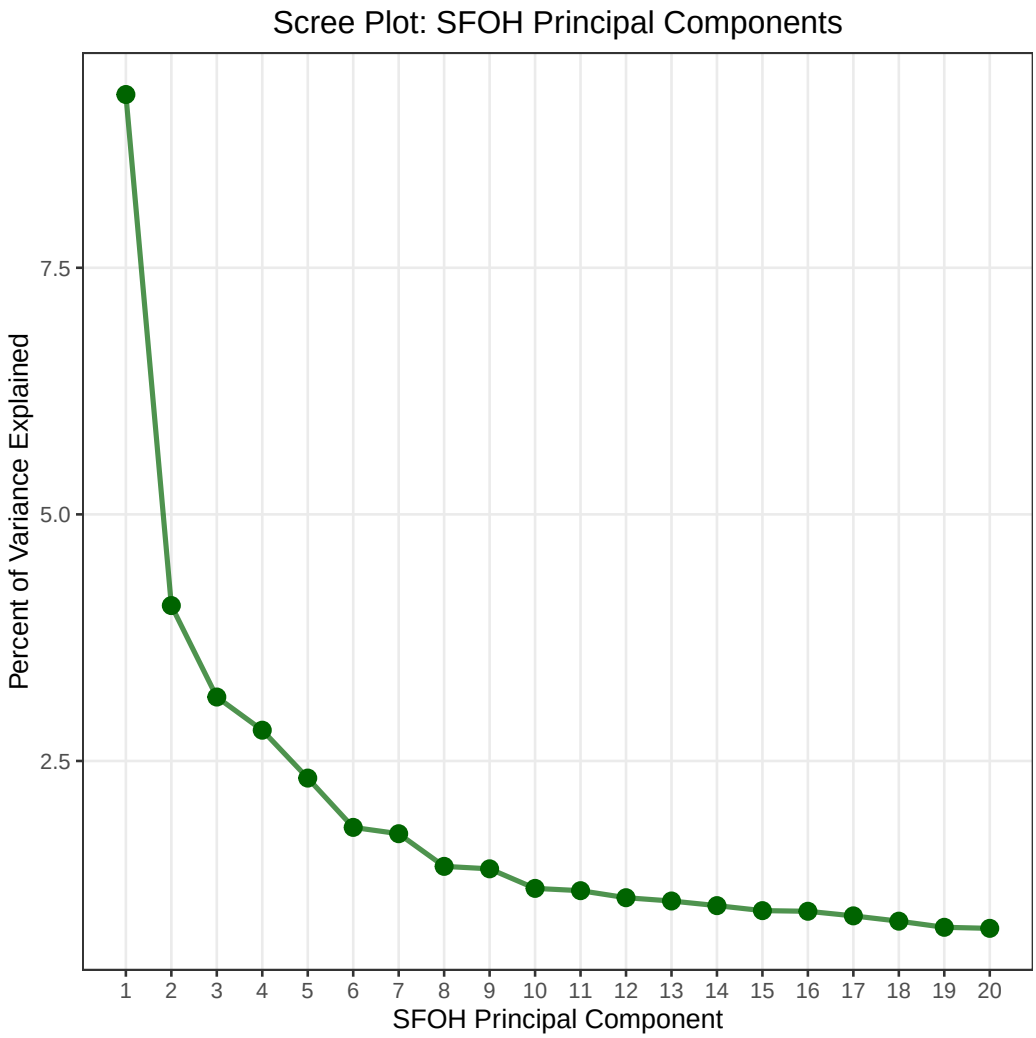

Fig. S3: SFOH PCA scree plot

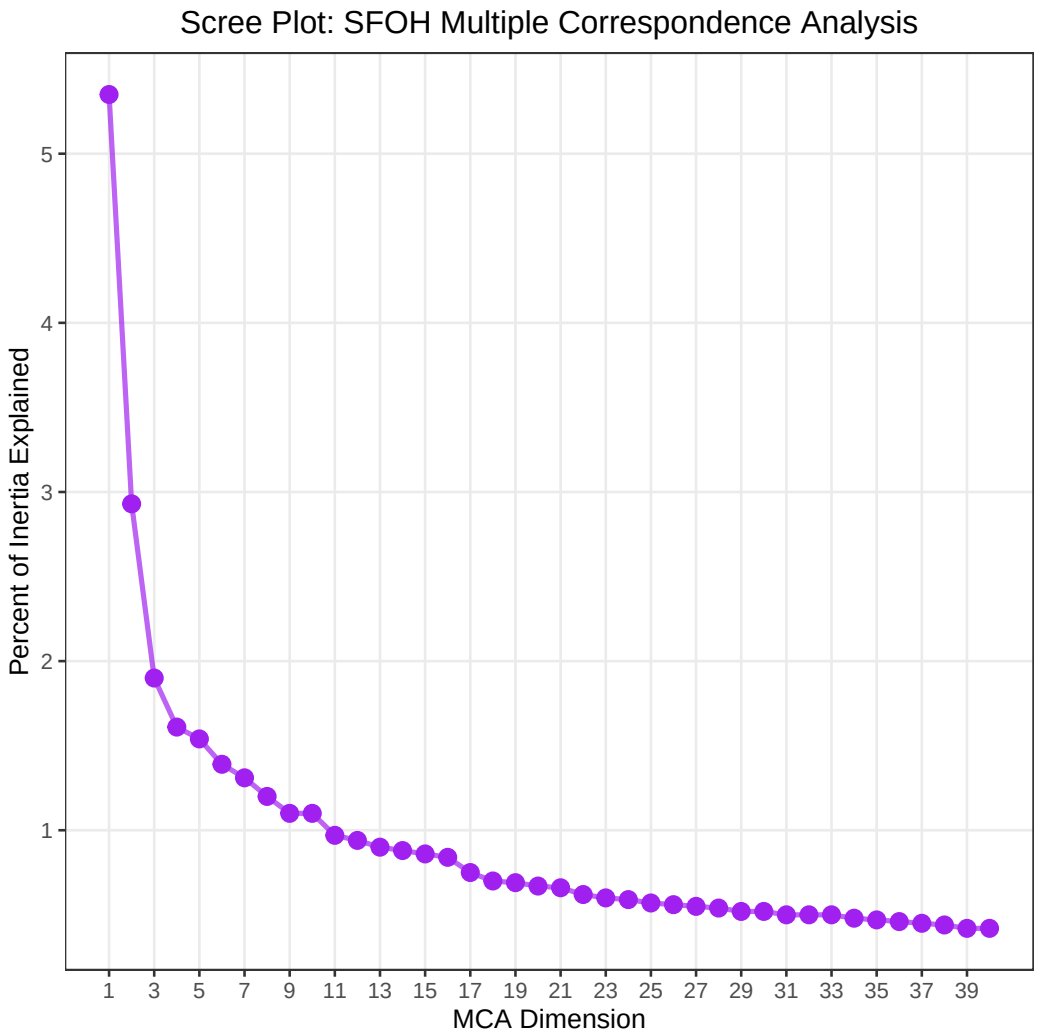

Fig. S4: SFOH MCA scree plot

Table S7: full cohort: Do SFOH PCs explain variance of Genetic PCs

GENPC	R^2	F-stat	F p-value
GENPC1	0.184	992.652	0.000
GENPC2	0.026	118.039	0.000
GENPC3	0.035	159.044	0.000
GENPC4	0.031	140.070	0.000
GENPC5	0.010	43.227	0.000
GENPC6	0.015	65.924	0.000
GENPC7	0.004	19.109	0.000
GENPC8	0.006	25.444	0.000
GENPC9	0.006	26.041	0.000
GENPC10	0.003	12.362	0.000

Table S8: full cohort: Do MC Axes explain variance of Genetic PCs

GENPC	R^2	F-stat	F p-value
GENPC1	0.206	1139.612	0.000
GENPC2	0.039	180.670	0.000
GENPC3	0.043	200.038	0.000
GENPC4	0.031	141.184	0.000
GENPC5	0.011	48.644	0.000
GENPC6	0.017	77.088	0.000
GENPC7	0.004	18.116	0.000
GENPC8	0.008	36.474	0.000
GENPC9	0.004	18.815	0.000
GENPC10	0.003	12.976	0.000

Table S9: full cohort: Do Genetic PCs explain variance of SFOH PCs

SFOHPC	R^2	F-stat	F p-value
SFOHPC1	0.026	236.895	0.000
SFOHPC2	0.041	375.185	0.000
SFOHPC3	0.015	136.914	0.000
SFOHPC4	0.029	258.842	0.000
SFOHPC5	0.062	584.787	0.000
SFOHPC6	0.019	165.936	0.000
SFOHPC7	0.015	131.894	0.000
SFOHPC8	0.021	191.462	0.000
SFOHPC9	0.004	32.169	0.000
SFOHPC10	0.029	259.003	0.000
SFOHPC11	0.003	27.209	0.000
SFOHPC12	0.014	125.033	0.000
SFOHPC13	0.010	86.146	0.000
SFOHPC14	0.006	52.523	0.000
SFOHPC15	0.006	49.701	0.000
SFOHPC16	0.004	33.414	0.000
SFOHPC17	0.005	41.492	0.000
SFOHPC18	0.002	21.156	0.000
SFOHPC19	0.012	111.350	0.000
SFOHPC20	0.002	16.871	0.000

Table S10: full cohort: Do Genetic PCs explain variance of MC Axes

MC	R^2	F-stat	F p-value
MC1	0.044	408.686	0.000
MC2	0.035	319.490	0.000
MC3	0.076	722.006	0.000
MC4	0.047	434.789	0.000
MC5	0.026	236.145	0.000
MC6	0.008	67.450	0.000
MC7	0.009	81.481	0.000
MC8	0.006	50.074	0.000
MC9	0.003	25.713	0.000
MC10	0.003	28.344	0.000
MC11	0.002	16.505	0.000
MC12	0.001	12.127	0.000
MC13	0.002	16.396	0.000
MC14	0.004	32.517	0.000
MC15	0.003	24.711	0.000
MC16	0.035	319.054	0.000
MC17	0.013	112.996	0.000
MC18	0.016	139.778	0.000
MC19	0.029	258.828	0.000
MC20	0.020	183.289	0.000

Table S11: Black/AA Genetic Prediction

GENPC	R^2	F-stat	F p-value
GENPC1	0.031	10.519	0.000
GENPC2	0.013	4.409	0.000
GENPC3	0.008	2.476	0.000
GENPC4	0.002	0.703	0.827
GENPC5	0.004	1.139	0.300
GENPC6	0.009	2.811	0.000
GENPC7	0.007	2.236	0.001
GENPC8	0.003	0.827	0.682
GENPC9	0.002	0.775	0.747
GENPC10	0.004	1.306	0.162

Table S12: Black/AA SFOH Prediction

SFOHPC	R^2	F-stat	F p-value
SFOHPC1	0.003	2.080	0.023
SFOHPC2	0.005	3.428	0.000
SFOHPC3	0.008	5.358	0.000
SFOHPC4	0.003	1.925	0.037
SFOHPC5	0.004	2.859	0.001
SFOHPC6	0.005	3.385	0.000
SFOHPC7	0.001	0.769	0.659
SFOHPC8	0.004	2.522	0.005
SFOHPC9	0.002	1.558	0.113
SFOHPC10	0.003	1.765	0.061
SFOHPC11	0.002	1.186	0.295
SFOHPC12	0.004	2.574	0.004
SFOHPC13	0.002	1.166	0.308
SFOHPC14	0.004	2.486	0.006
SFOHPC15	0.002	1.167	0.308
SFOHPC16	0.004	2.286	0.011
SFOHPC17	0.002	1.345	0.200
SFOHPC18	0.001	0.795	0.633
SFOHPC19	0.002	1.114	0.347
SFOHPC20	0.002	1.289	0.230

Table S13: White Genetic Prediction

GENPC	R^2	F-stat	F p-value
GENPC1	0.021	69.574	0.000
GENPC2	0.009	28.679	0.000
GENPC3	0.036	123.472	0.000
GENPC4	0.005	16.352	0.000
GENPC5	0.021	70.146	0.000
GENPC6	0.008	28.356	0.000
GENPC7	0.006	20.239	0.000
GENPC8	0.003	10.209	0.000
GENPC9	0.008	27.795	0.000
GENPC10	0.001	4.864	0.000

Table S14: White SFOH Prediction

SFOHPC	R^2	F-stat	F p-value
SFOHPC1	0.002	13.339	0.000
SFOHPC2	0.005	31.238	0.000
SFOHPC3	0.001	9.555	0.000
SFOHPC4	0.003	21.540	0.000
SFOHPC5	0.011	72.907	0.000
SFOHPC6	0.013	89.254	0.000
SFOHPC7	0.003	18.598	0.000
SFOHPC8	0.004	28.064	0.000
SFOHPC9	0.003	17.211	0.000
SFOHPC10	0.002	13.380	0.000
SFOHPC11	0.001	7.952	0.000
SFOHPC12	0.006	37.549	0.000
SFOHPC13	0.001	9.072	0.000
SFOHPC14	0.002	13.690	0.000
SFOHPC15	0.001	3.805	0.000
SFOHPC16	0.000	2.429	0.007
SFOHPC17	0.001	9.486	0.000
SFOHPC18	0.002	13.773	0.000
SFOHPC19	0.001	9.650	0.000
SFOHPC20	0.001	8.890	0.000

Table S15: White analysis cohort: Do MC Axes explain variation in Genetic PCs

GENPC	R^2	F-stat	F p-value
GENPC1	0.025	86.610	0.000
GENPC2	0.010	32.289	0.000
GENPC3	0.042	146.962	0.000
GENPC4	0.006	19.333	0.000
GENPC5	0.024	81.329	0.000
GENPC6	0.009	30.923	0.000
GENPC7	0.006	18.972	0.000
GENPC8	0.003	10.655	0.000
GENPC9	0.007	22.426	0.000
GENPC10	0.001	4.625	0.000

Table S16: Black/AA analysis cohort: Do MC Axes explain variation in Genetic PCs

GENPC	R^2	F-stat	F p-value
GENPC1	0.032	10.631	0.000
GENPC2	0.014	4.573	0.000
GENPC3	0.013	4.360	0.000
GENPC4	0.003	1.070	0.375
GENPC5	0.003	0.922	0.559
GENPC6	0.012	3.999	0.000
GENPC7	0.007	2.285	0.001
GENPC8	0.003	1.081	0.362
GENPC9	0.003	0.889	0.602
GENPC10	0.003	0.865	0.634

Table S17: White analysis cohort: Do Genetic PCs explain variation in MC Axes

MC	R^2	F-stat	F p-value
MC1	0.004	28.518	0.000
MC2	0.002	10.648	0.000
MC3	0.003	18.505	0.000
MC4	0.017	119.361	0.000
MC5	0.011	75.259	0.000
MC6	0.001	8.318	0.000
MC7	0.002	14.916	0.000
MC8	0.001	4.838	0.000
MC9	0.001	4.254	0.000
MC10	0.002	10.799	0.000
MC11	0.001	9.454	0.000
MC12	0.000	0.752	0.675
MC13	0.001	8.954	0.000
MC14	0.002	13.543	0.000
MC15	0.002	13.847	0.000
MC16	0.013	88.731	0.000
MC17	0.001	5.255	0.000
MC18	0.002	14.564	0.000
MC19	0.004	26.958	0.000
MC20	0.002	10.943	0.000

Table S18: Black/AA analysis cohort: Genetic PCs explain variation in MC Axes

MC	R^2	F-stat	F p-value
MC1	0.005	3.528	0.000
MC2	0.008	5.177	0.000
MC3	0.004	2.557	0.004
MC4	0.003	1.669	0.082
MC5	0.004	2.541	0.005
MC6	0.002	1.301	0.223
MC7	0.001	0.665	0.758
MC8	0.001	0.956	0.480
MC9	0.001	0.836	0.594
MC10	0.002	1.069	0.382
MC11	0.006	4.155	0.000
MC12	0.002	1.182	0.297
MC13	0.001	0.541	0.862
MC14	0.002	1.189	0.293
MC15	0.003	1.710	0.072
MC16	0.004	2.912	0.001
MC17	0.002	1.406	0.171
MC18	0.001	0.748	0.679
MC19	0.003	1.836	0.049
MC20	0.004	2.548	0.005

Table S19: Female analysis cohort SFOH PC prediction

SFOHPC	R^2	F-stat	F p-value
SFOHPC1	0.023	130.500	0.000
SFOHPC2	0.050	287.404	0.000
SFOHPC3	0.019	106.193	0.000
SFOHPC4	0.026	148.888	0.000
SFOHPC5	0.068	403.814	0.000
SFOHPC6	0.018	100.656	0.000
SFOHPC7	0.015	82.691	0.000
SFOHPC8	0.023	130.571	0.000
SFOHPC9	0.005	26.119	0.000
SFOHPC10	0.031	176.632	0.000
SFOHPC11	0.003	17.575	0.000
SFOHPC12	0.016	89.621	0.000
SFOHPC13	0.011	58.483	0.000
SFOHPC14	0.007	40.848	0.000
SFOHPC15	0.007	40.672	0.000
SFOHPC16	0.005	26.622	0.000
SFOHPC17	0.006	31.220	0.000
SFOHPC18	0.003	14.879	0.000
SFOHPC19	0.014	79.060	0.000
SFOHPC20	0.002	10.923	0.000

Table S20: Male analysis cohort SFOH PC prediction

SFOHPC	R^2	F-stat	F p-value
SFOHPC1	0.029	93.444	0.000
SFOHPC2	0.029	93.820	0.000
SFOHPC3	0.011	35.157	0.000
SFOHPC4	0.029	93.979	0.000
SFOHPC5	0.050	164.359	0.000
SFOHPC6	0.016	50.351	0.000
SFOHPC7	0.016	49.133	0.000
SFOHPC8	0.019	59.961	0.000
SFOHPC9	0.003	9.312	0.000
SFOHPC10	0.024	76.904	0.000
SFOHPC11	0.004	12.901	0.000
SFOHPC12	0.011	33.572	0.000
SFOHPC13	0.008	26.071	0.000
SFOHPC14	0.003	9.457	0.000
SFOHPC15	0.005	14.536	0.000
SFOHPC16	0.002	5.317	0.000
SFOHPC17	0.004	11.328	0.000
SFOHPC18	0.002	6.340	0.000
SFOHPC19	0.013	40.577	0.000
SFOHPC20	0.002	6.202	0.000

Table S21: Female analysis cohort Genetic PC prediction by SFOH PCs

GENPC	R^2	F-stat	F p-value
GENPC1	0.195	667.330	0.000
GENPC2	0.030	84.979	0.000
GENPC3	0.036	103.870	0.000
GENPC4	0.037	106.096	0.000
GENPC5	0.010	26.425	0.000
GENPC6	0.015	41.636	0.000
GENPC7	0.004	11.087	0.000
GENPC8	0.006	17.743	0.000
GENPC9	0.006	16.389	0.000
GENPC10	0.003	8.948	0.000

Table S22: Male analysis cohort SFOH PC prediction

GENPC	R^2	F-stat	F p-value
GENPC1	0.163	301.686	0.000
GENPC2	0.022	34.238	0.000
GENPC3	0.033	53.451	0.000
GENPC4	0.021	32.894	0.000
GENPC5	0.012	18.112	0.000
GENPC6	0.015	23.474	0.000
GENPC7	0.006	9.524	0.000
GENPC8	0.005	8.019	0.000
GENPC9	0.007	10.692	0.000
GENPC10	0.002	3.775	0.000

Table S23: Female analysis cohort Genetic PC prediction by MCs

GENPC	R^2	F-stat	F p-value
GENPC1	0.216	756.081	0.000
GENPC2	0.046	131.553	0.000
GENPC3	0.045	128.738	0.000
GENPC4	0.039	110.738	0.000
GENPC5	0.011	30.506	0.000
GENPC6	0.018	50.962	0.000
GENPC7	0.004	10.541	0.000
GENPC8	0.009	25.112	0.000
GENPC9	0.004	12.192	0.000
GENPC10	0.004	9.662	0.000

Table S24: Male analysis cohort Geentic PC prediction by MCs

GENPC	R^2	F-stat	F p-value
GENPC1	0.188	358.846	0.000
GENPC2	0.031	50.241	0.000
GENPC3	0.042	67.416	0.000
GENPC4	0.020	31.514	0.000
GENPC5	0.013	19.653	0.000
GENPC6	0.016	25.431	0.000
GENPC7	0.005	8.026	0.000
GENPC8	0.007	11.587	0.000
GENPC9	0.005	7.527	0.000
GENPC10	0.002	3.730	0.000

Table S25: Female analysis cohort SFOH MC prediction

MC	R^2	F-stat	F p-value
MC1	0.040	226.615	0.000
MC2	0.040	229.695	0.000
MC3	0.087	521.314	0.000
MC4	0.046	263.632	0.000
MC5	0.031	172.917	0.000
MC6	0.009	51.786	0.000
MC7	0.009	52.024	0.000
MC8	0.008	43.778	0.000
MC9	0.005	27.700	0.000
MC10	0.003	18.362	0.000
MC11	0.002	9.559	0.000
MC12	0.001	6.562	0.000
MC13	0.002	13.166	0.000
MC14	0.004	24.265	0.000
MC15	0.003	17.397	0.000
MC16	0.036	204.757	0.000
MC17	0.013	73.345	0.000
MC18	0.020	114.420	0.000
MC19	0.031	177.678	0.000
MC20	0.023	127.086	0.000

Table S26: Male analysis cohort SFOH MC prediction

MC	R^2	F-stat	F p-value
MC1	0.052	168.987	0.000
MC2	0.031	100.368	0.000
MC3	0.057	187.569	0.000
MC4	0.041	132.756	0.000
MC5	0.021	66.095	0.000
MC6	0.004	13.878	0.000
MC7	0.008	26.127	0.000
MC8	0.002	7.693	0.000
MC9	0.002	6.648	0.000
MC10	0.003	9.438	0.000
MC11	0.002	6.397	0.000
MC12	0.004	13.320	0.000
MC13	0.002	5.724	0.000
MC14	0.002	7.727	0.000
MC15	0.003	9.151	0.000
MC16	0.037	119.717	0.000
MC17	0.013	39.689	0.000
MC18	0.010	31.378	0.000
MC19	0.025	80.316	0.000
MC20	0.018	56.164	0.000

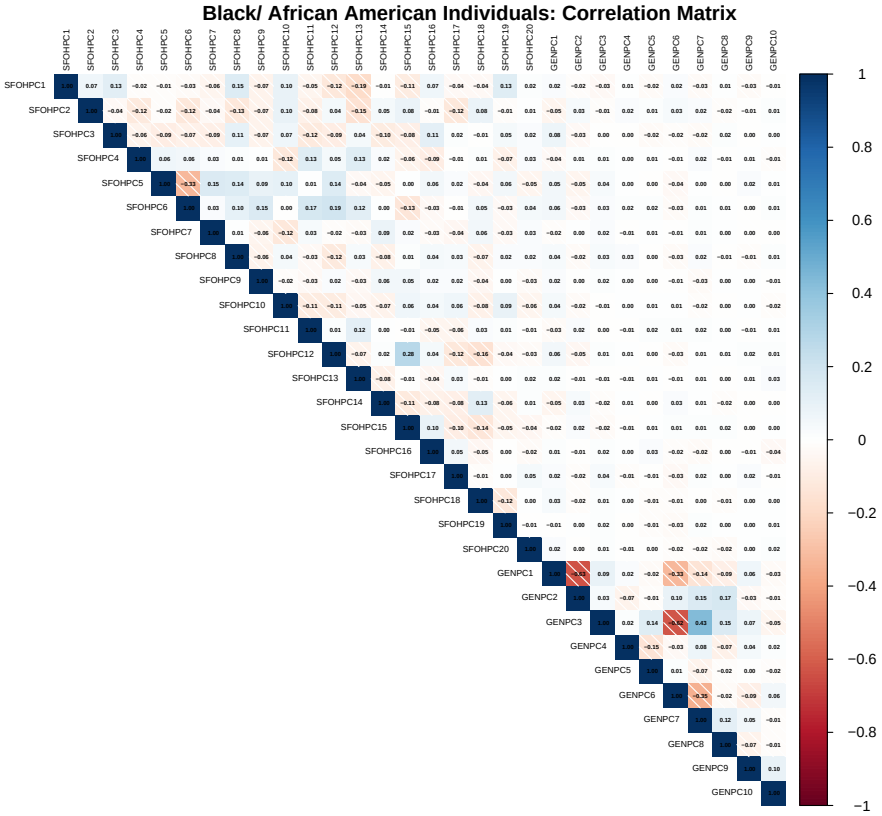

Fig. S5: PC Pearson correlation matrix for Black/African American individuals

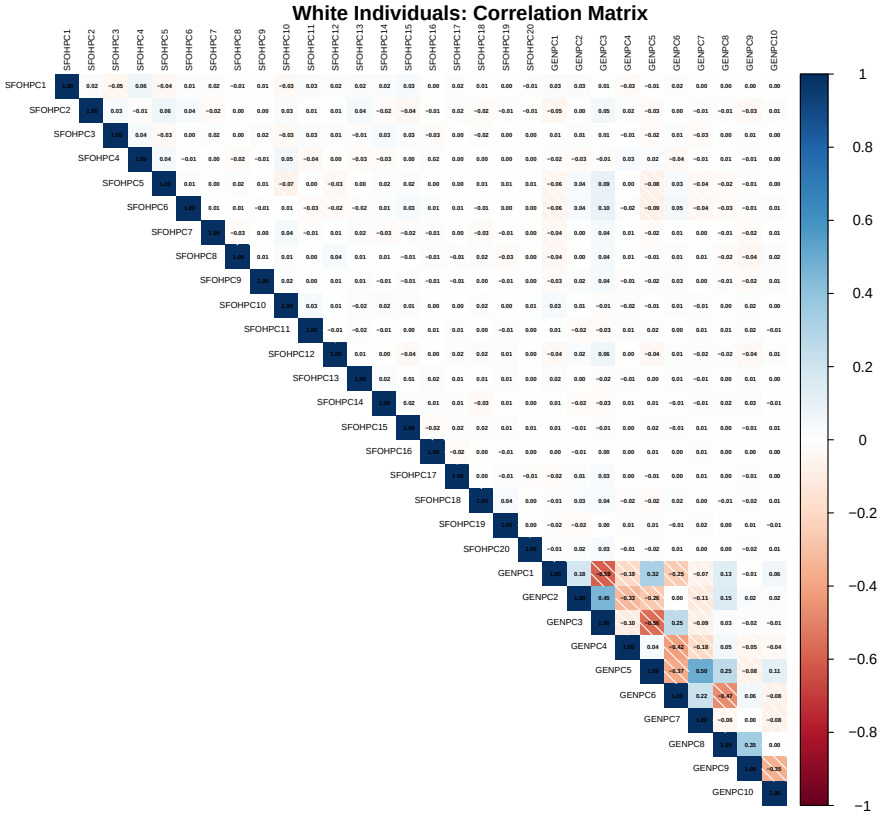

Fig. S6: PC Pearson correlation matrix for White individuals

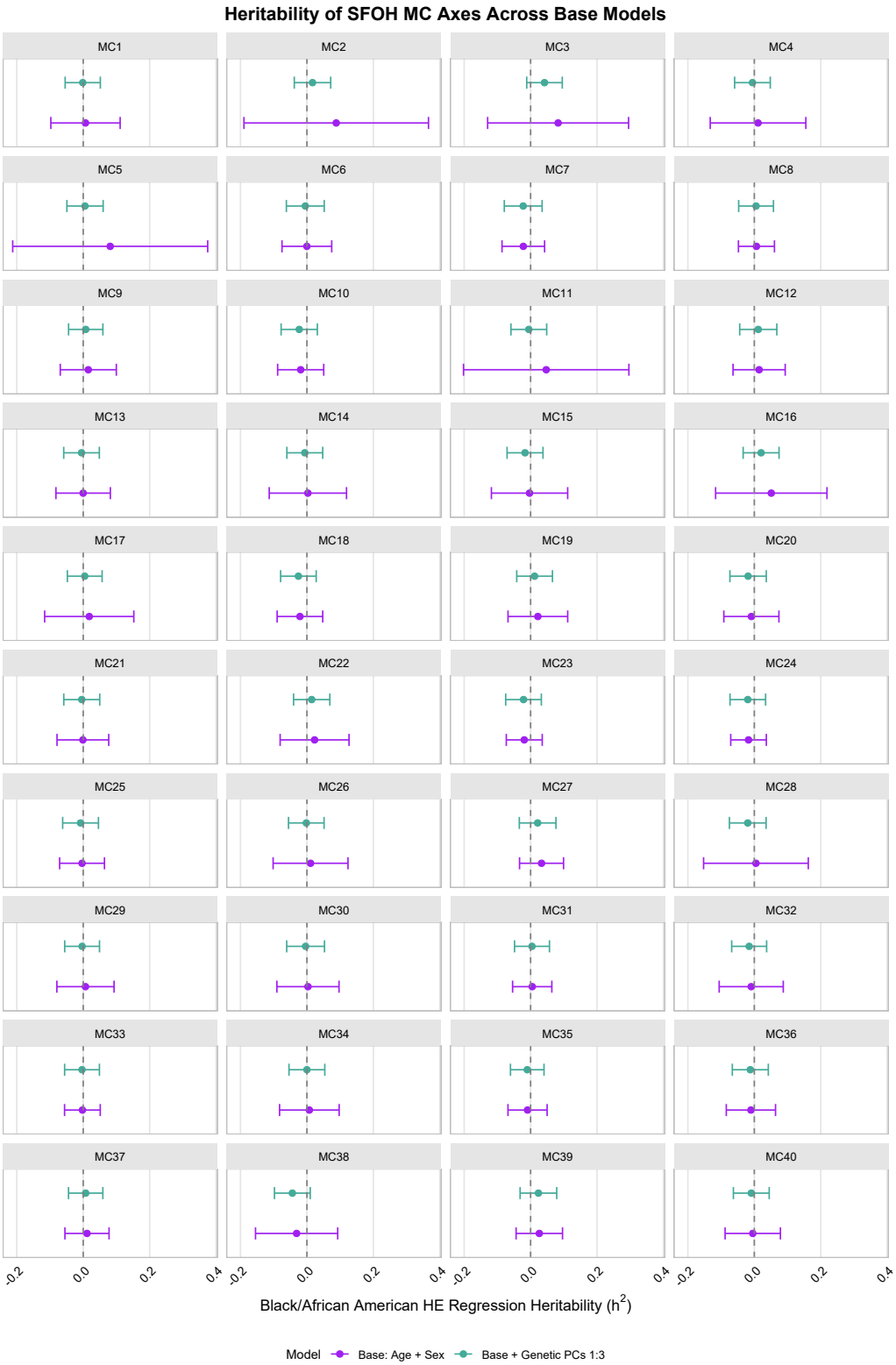

Fig. S7: Black/AA HE heritability estimates of SFOH MC Axes, none are significantly different from zero.

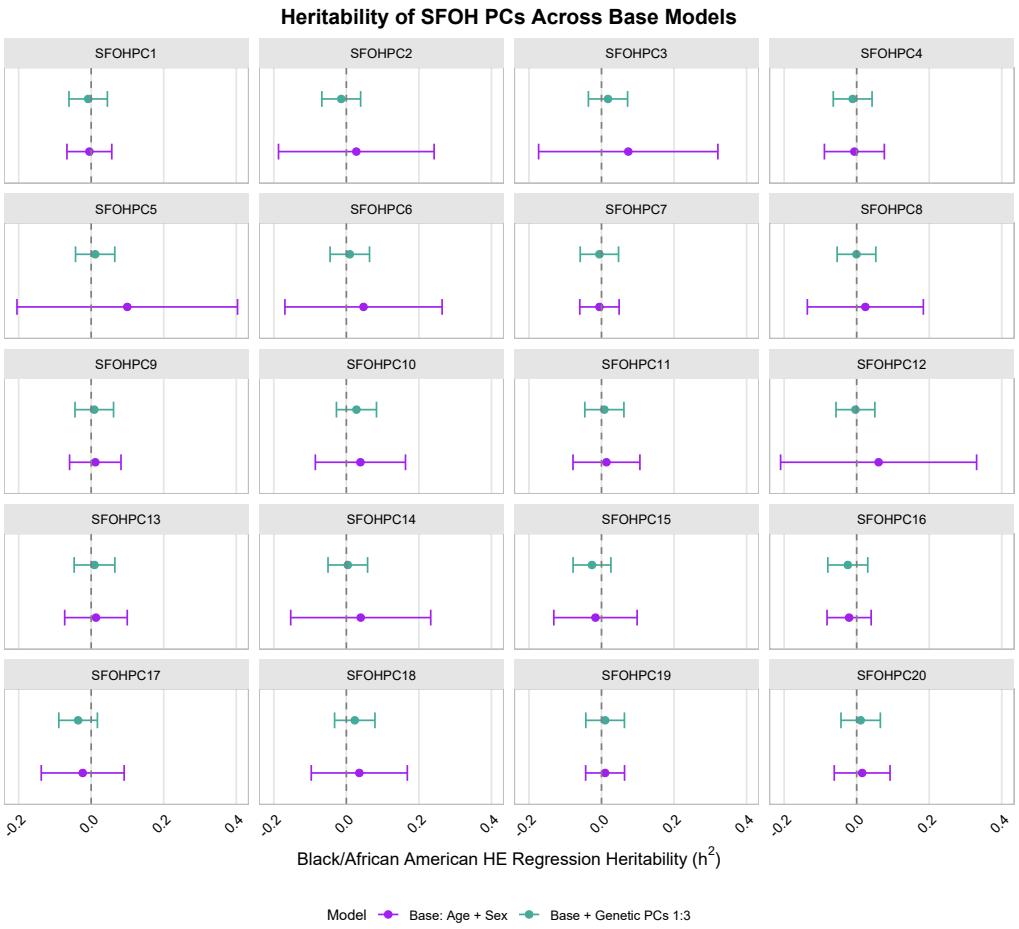

Fig. S8: Black/AA HE heritability estimates of SFOH PCs, none are significantly different from zero.

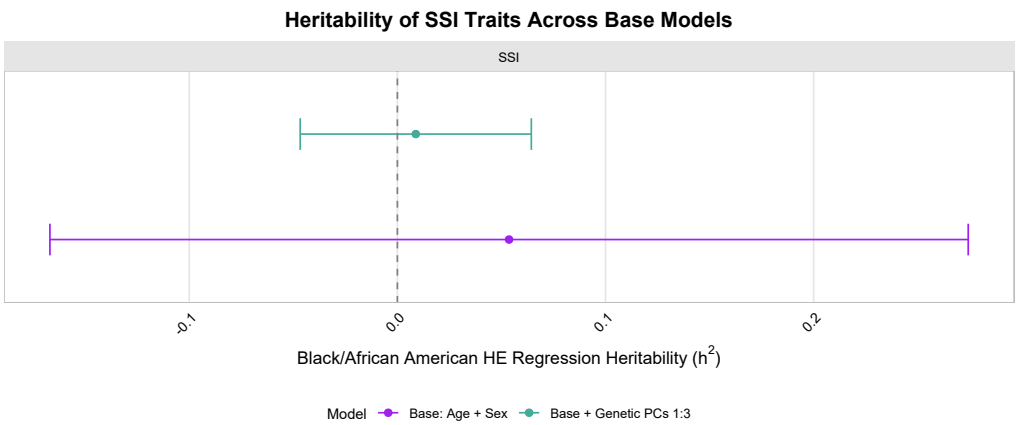

Fig. S9: Black/AA HE heritability estimates of SFOH derived SSI, none are significantly different from zero.

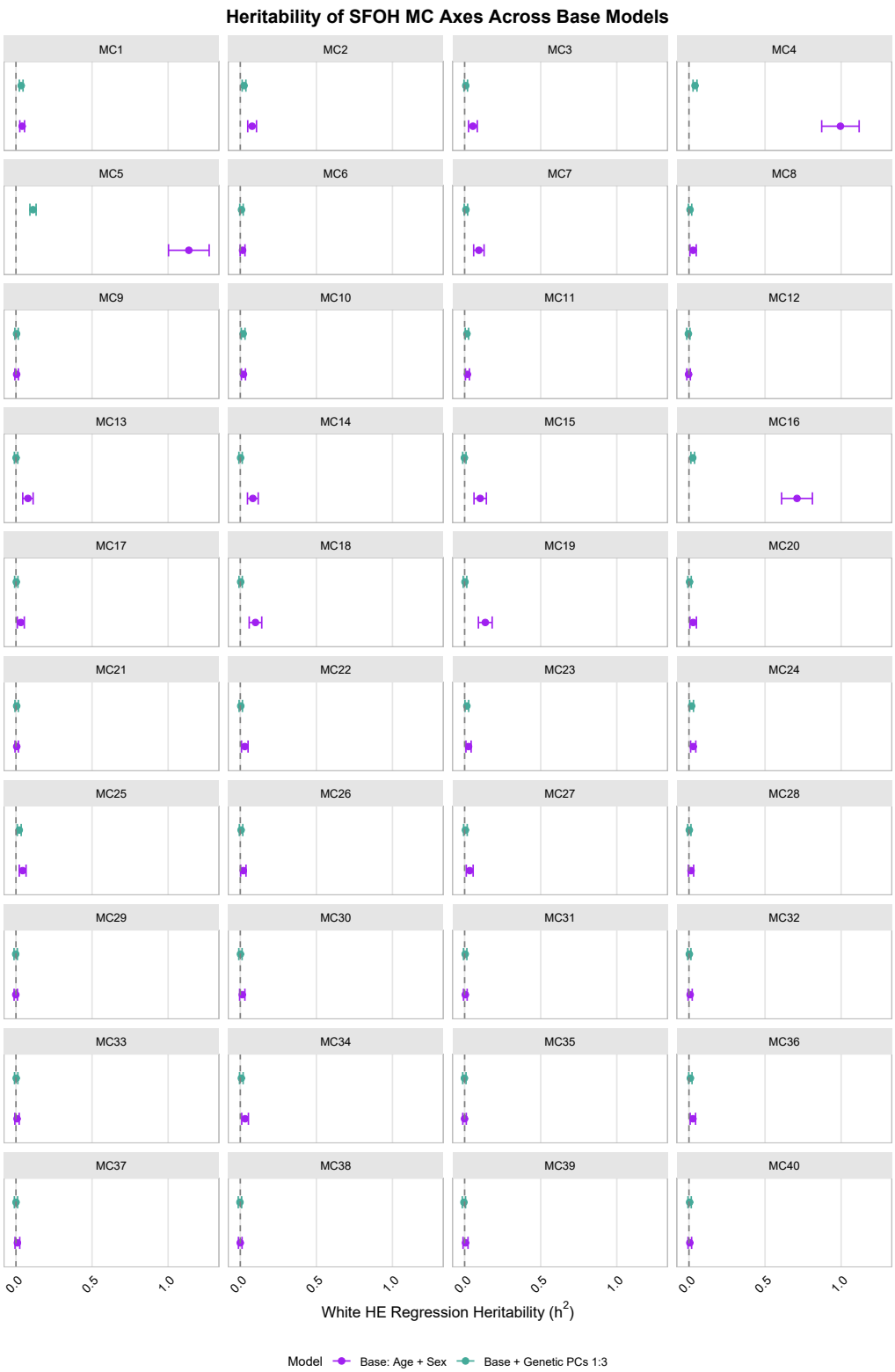

Fig. S10: White HE heritability estimates of SFOH MC Axes. MC Axes 1, 2, 4, 5, 10, 11, 16, and 23:25 remain significantly different from zero when including genetic PCs

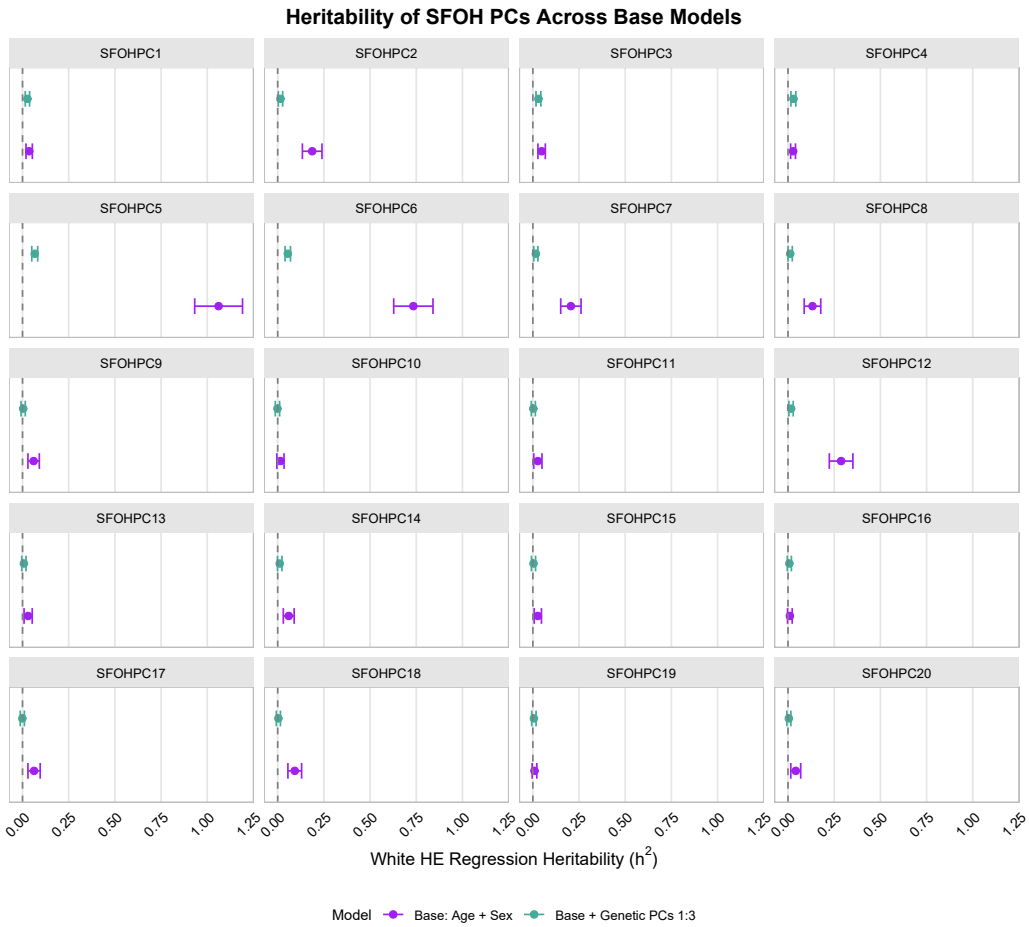

Fig. S11: White HE heritability estimates of SFOH PCs. SFOHPCs 1:8 and 12 significantly different from zero when including genetic PCs

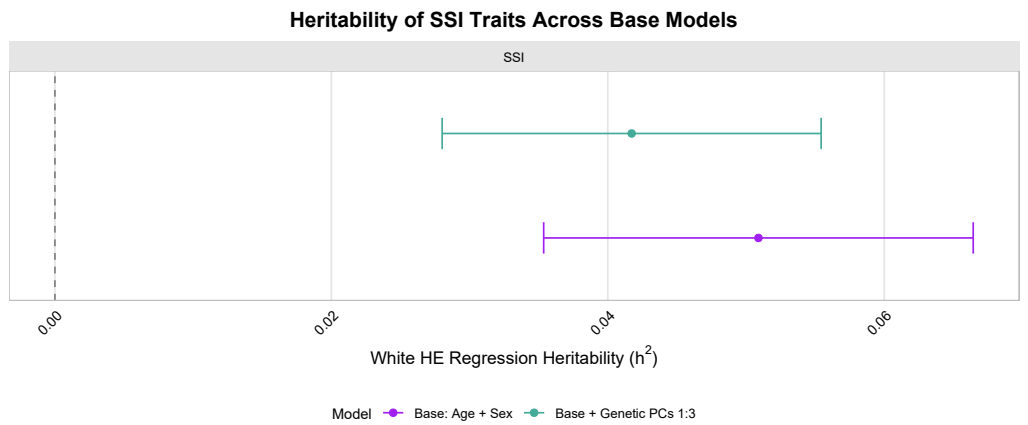

Fig. S12: White HE heritability estimates of SFOH derived SSI, remains significantly different from zero when including genetic PCs.

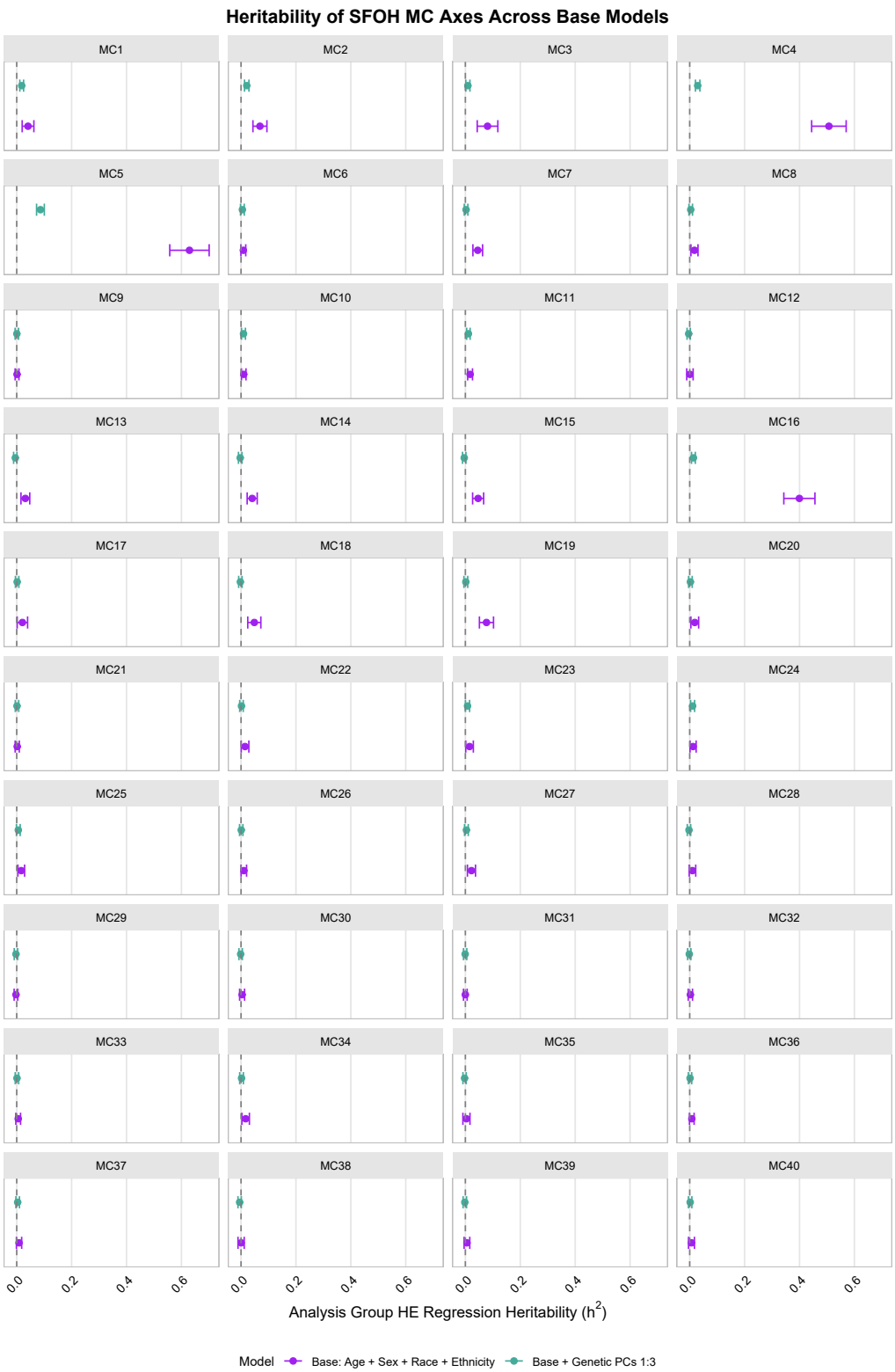

Fig. S13: Analysis Group HE heritability estimates of SFOH MC Axes. MC Axes 1:5, 10, 11, 16, 23, and 24 remain significantly different from zero when including genetic PCs

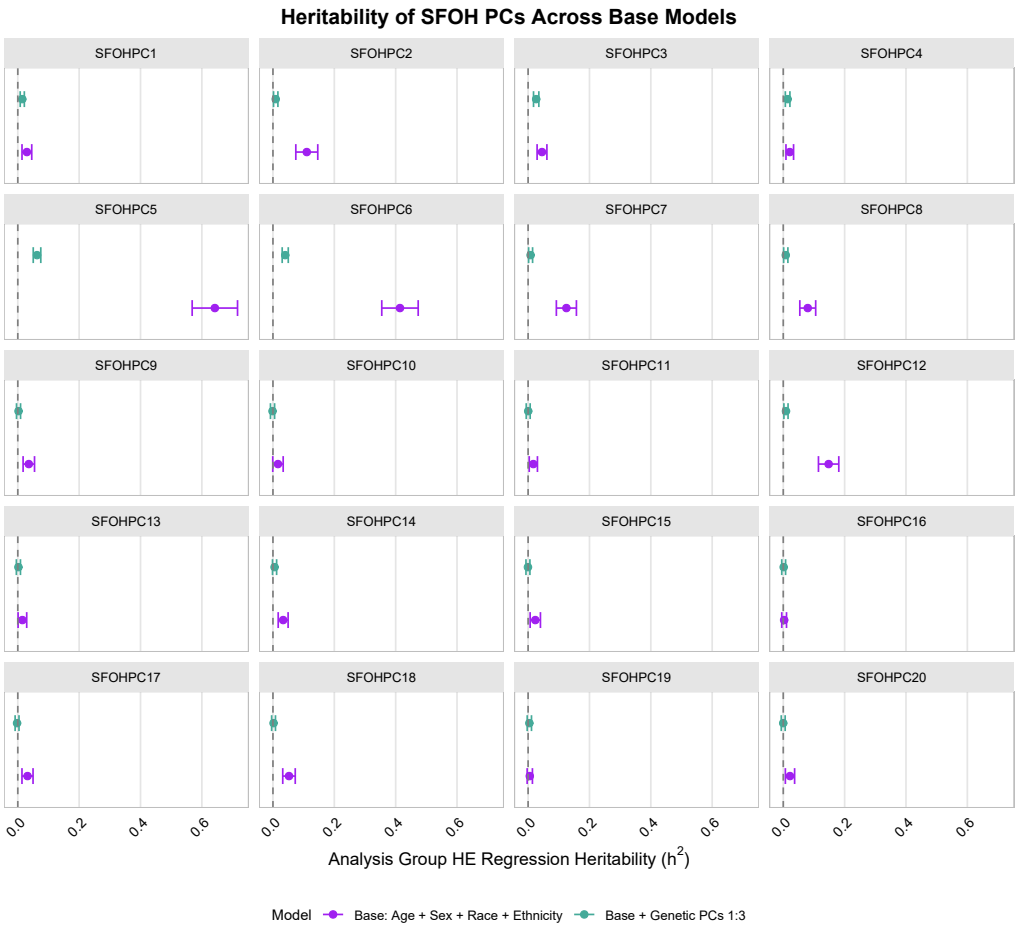

Fig. S14: Analysis Group HE heritability estimates of SFOH PCs. SFOHPCs 1:8 and 12 significantly different from zero when including genetic PCs

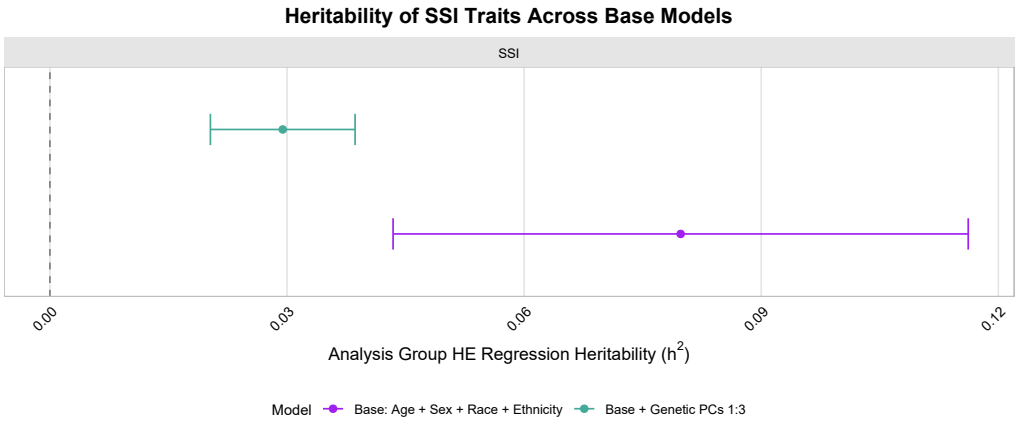

Fig. S15: Analysis Group HE heritability estimates of SFOH derived SSI, remains significantly different from zero when including genetic PCs.

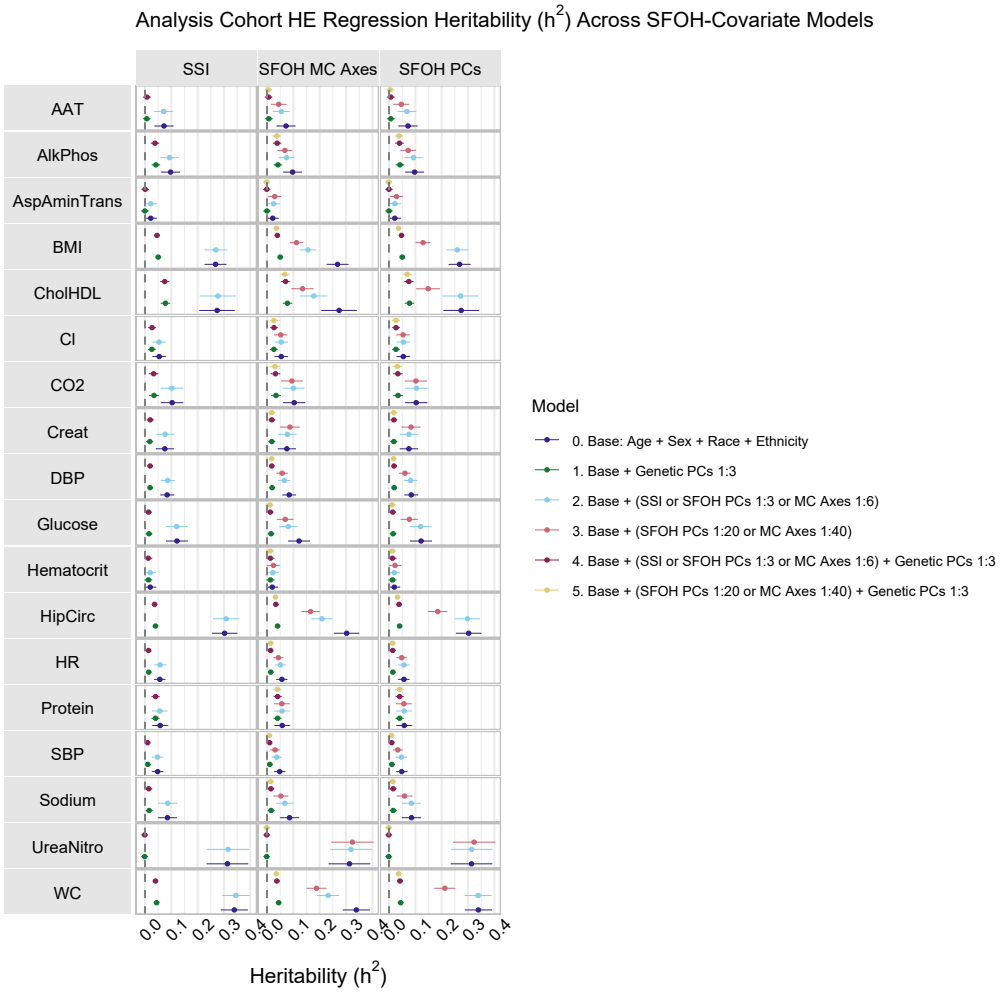

Fig. S16: Analysis Group HE heritability estimates of 18 traits.

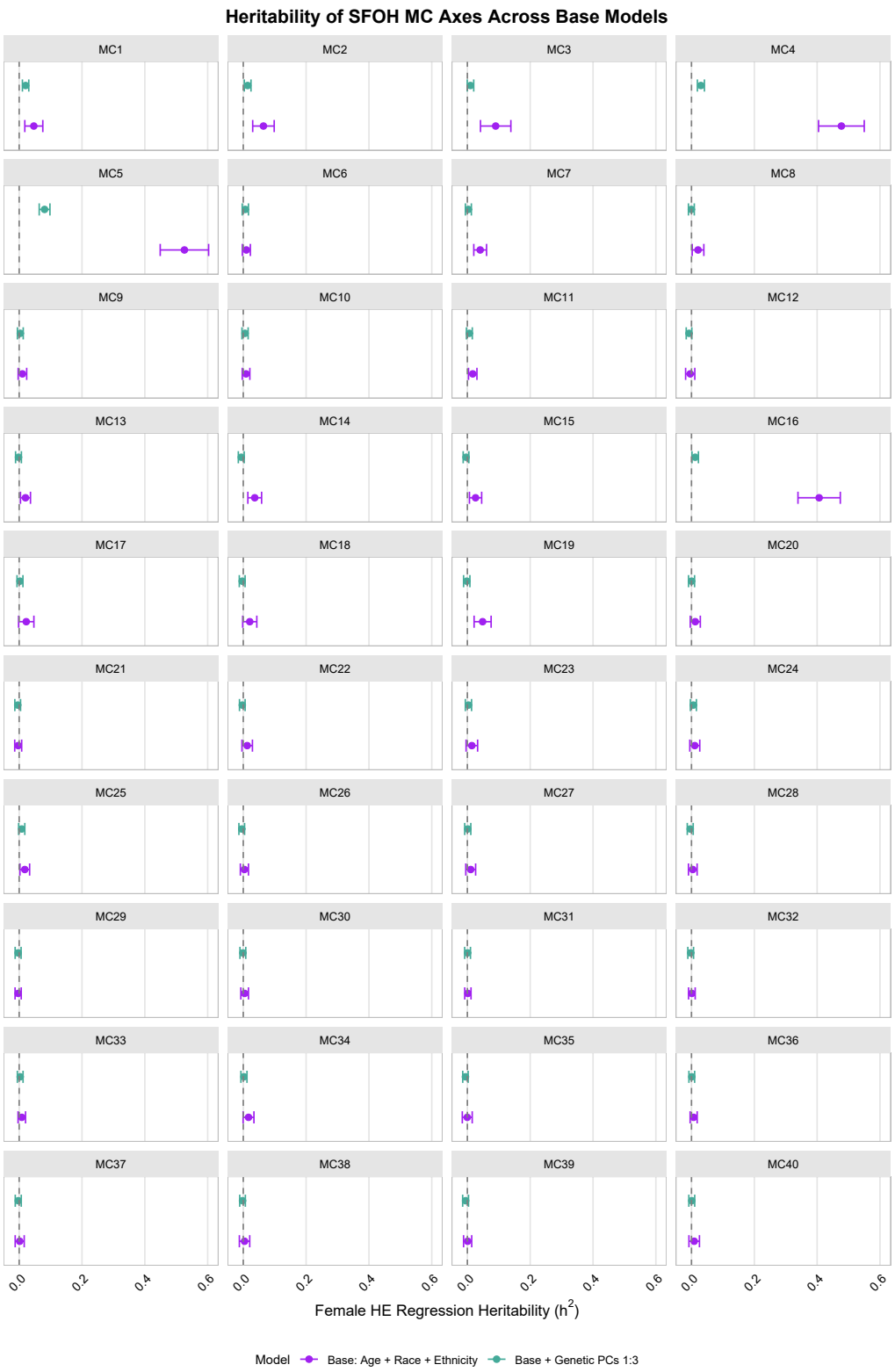

Fig. S17: Female HE heritability estimates of SFOH MC Axes. MC Axes 1, 2, 4, 5, and 16 remain significantly different from zero when including genetic PCs

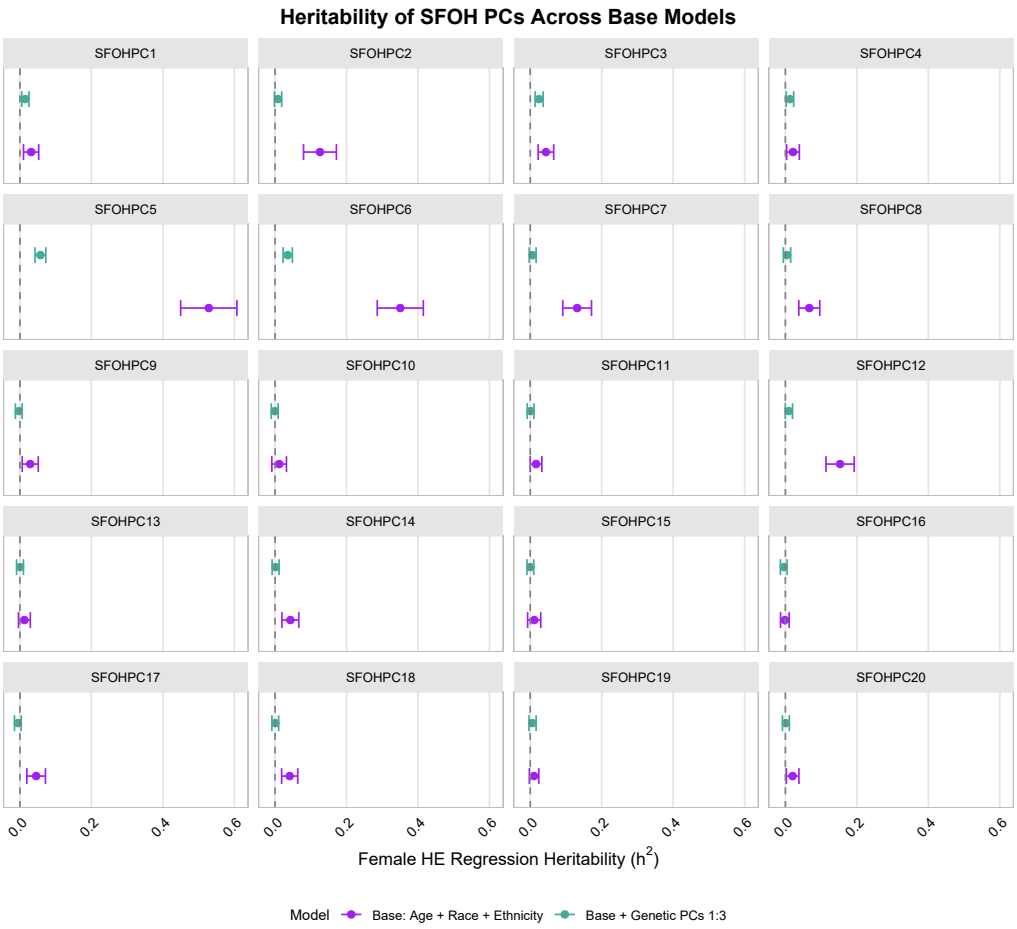

Fig. S18: Female HE heritability estimates of SFOH PCs. SFOHPCs 1, 3, 4, 5, and 6 significantly different from zero when including genetic PCs

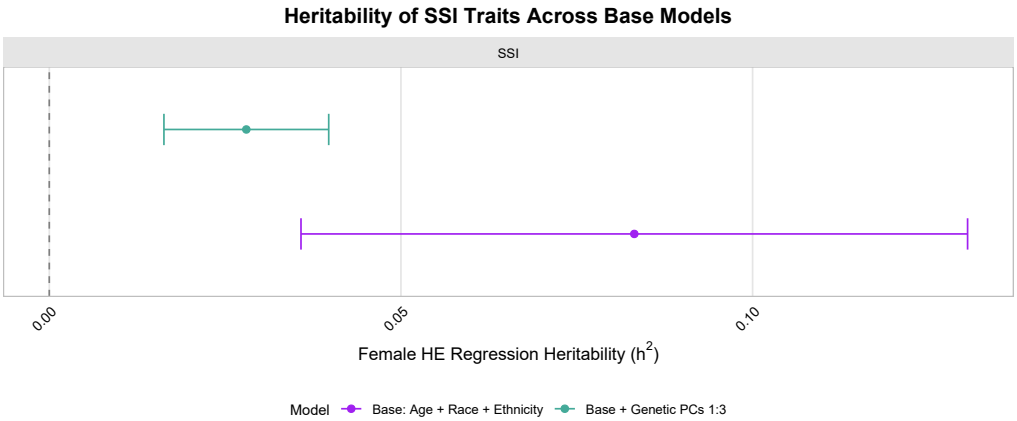

Fig. S19: Female HE heritability estimates of SFOH derived SSI, remains significantly different from zero when including genetic PCs.

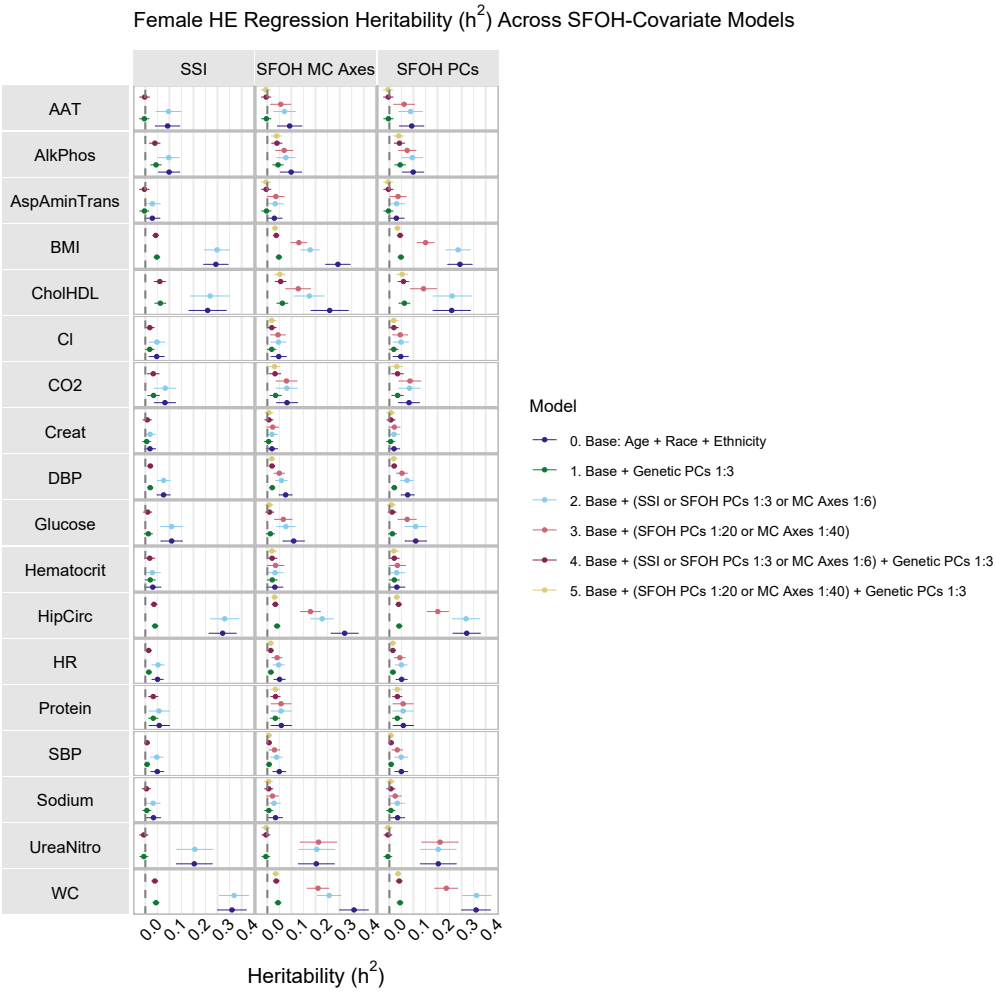

Fig. S20: Female HE heritability estimates of 18 traits.

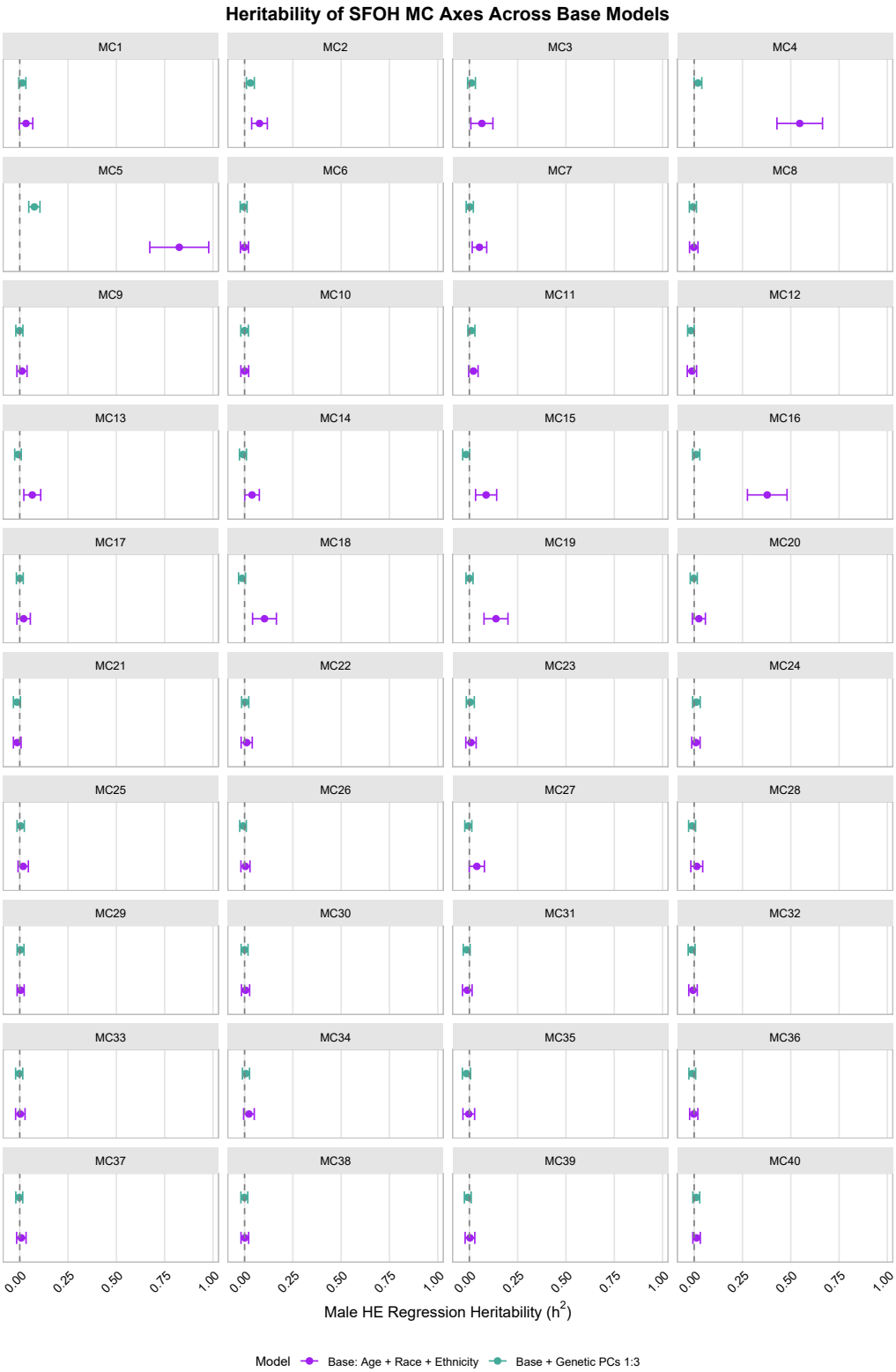

Fig. S21: Male HE heritability estimates of SFOH MC Axes. MC Axes 2 and 5 remain significantly different from zero when including genetic PCs

Fig. S22: Male HE heritability estimates of SFOH PCs. SFOHPCs 3, 5, and 6 significantly different from zero when including genetic PCs

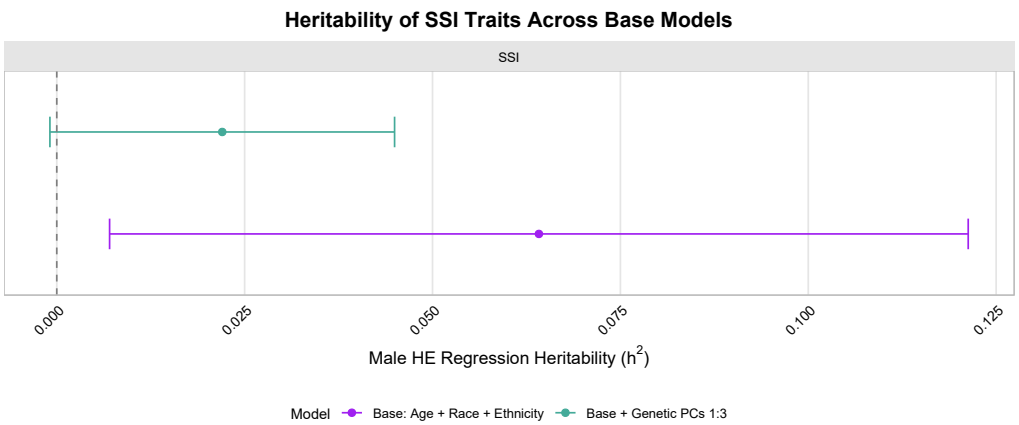

Fig. S23: Male HE heritability estimates of SFOH derived SSI, is not significantly different from zero when including genetic PCs.

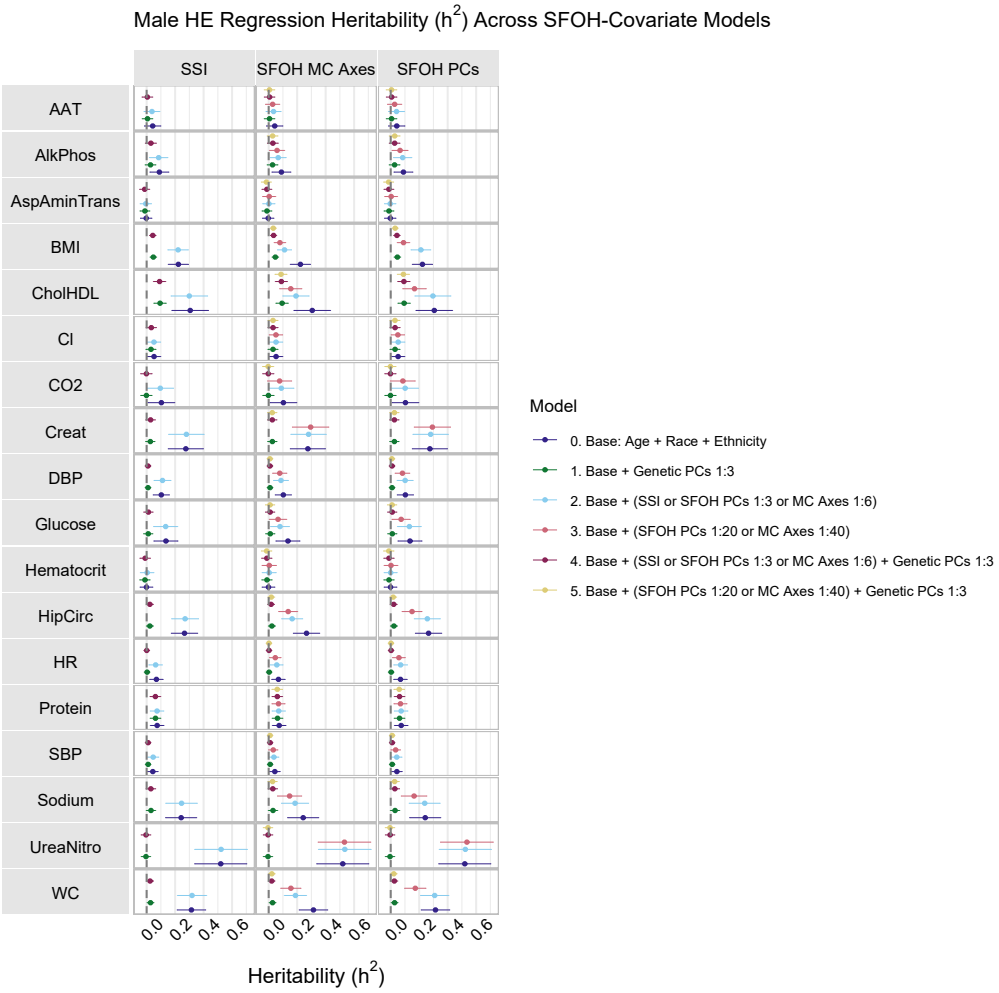

Fig. S24: Male HE heritability estimates of 18 traits.

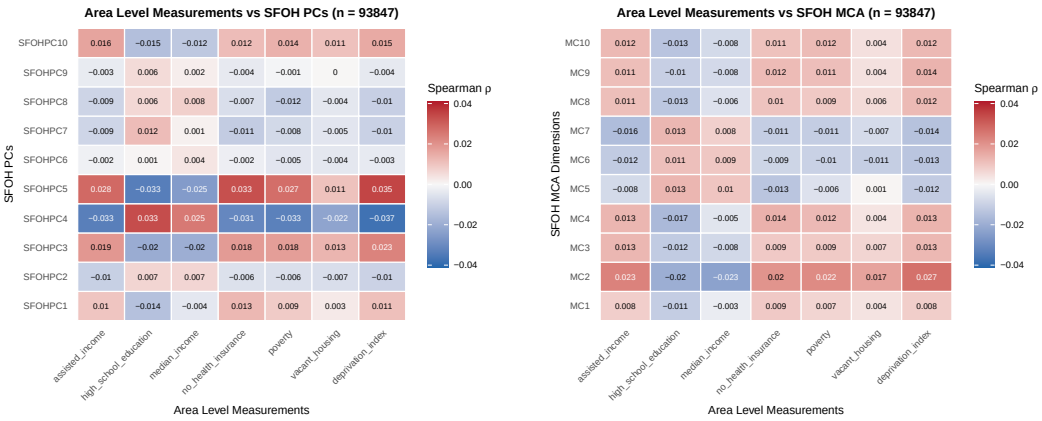

Fig. S25: Relationship between SFOH summaries and area level measurements.

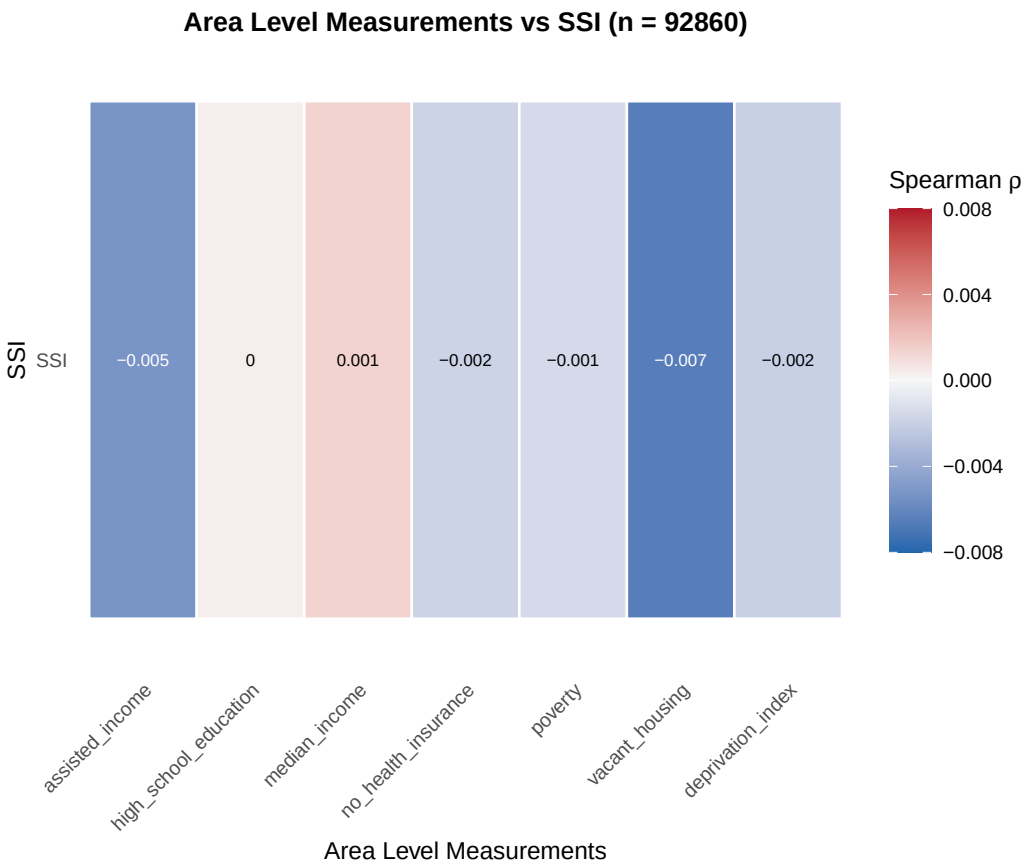

Fig. S26: Relationship between SSI and area level measurements.

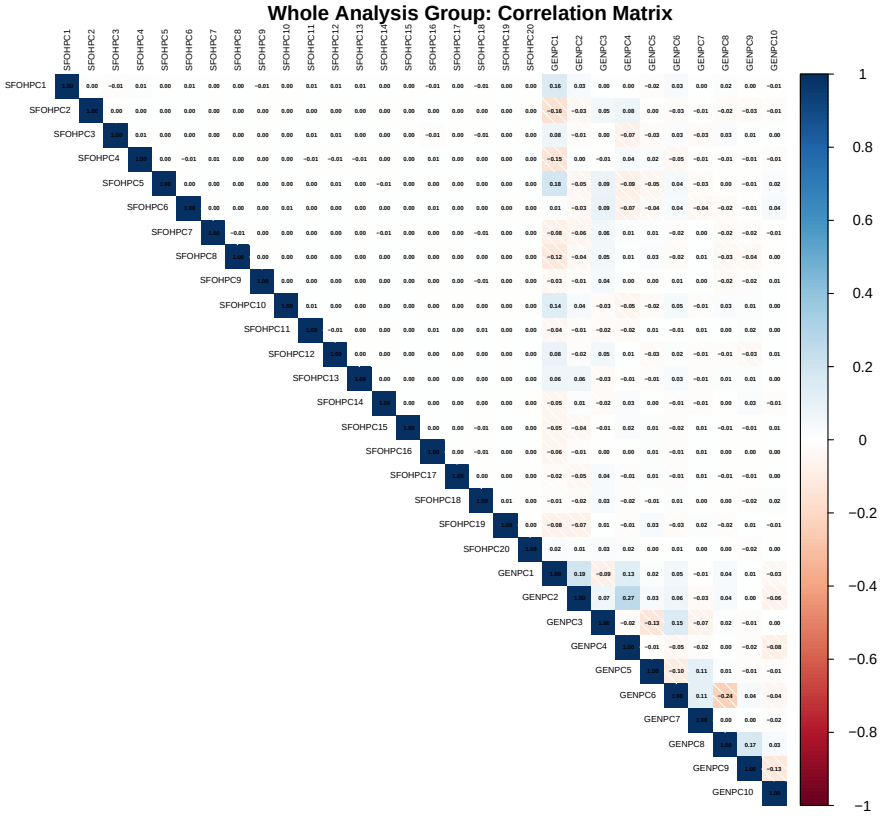

Fig. S27: PC Pearson correlation matrix for full cohort.

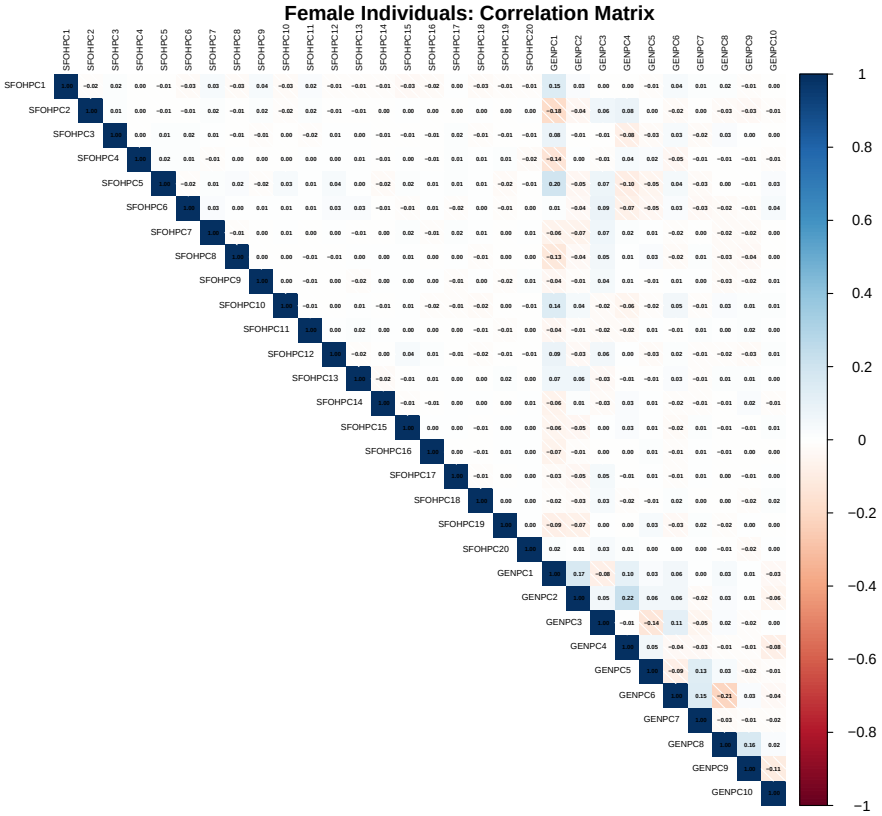

Fig. S28: PC Pearson correlation matrix for Female Analysis Cohort.

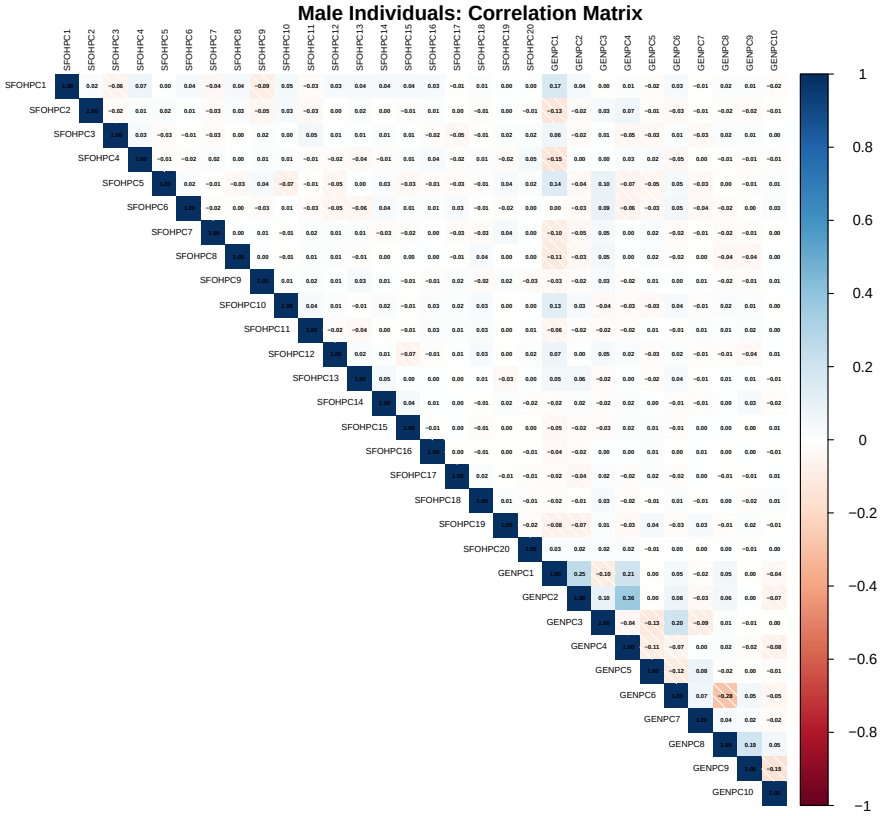

Fig. S29: PC Pearson correlation matrix for Male Analysis Cohort.

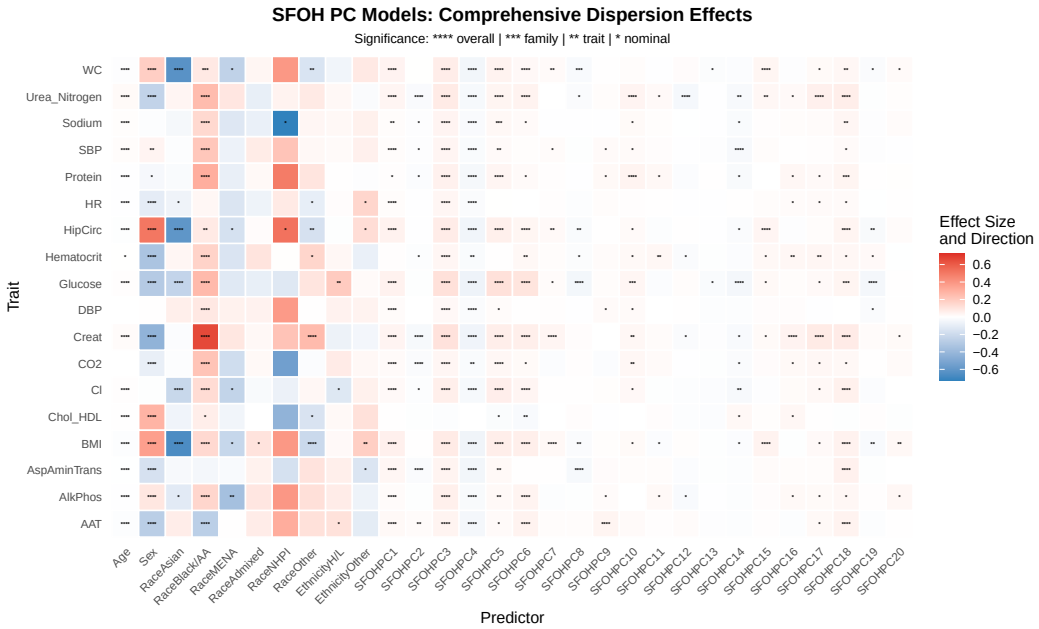

Fig. S30: Heatmap of significant predictors in trait dispersion models with SFOH PCs.

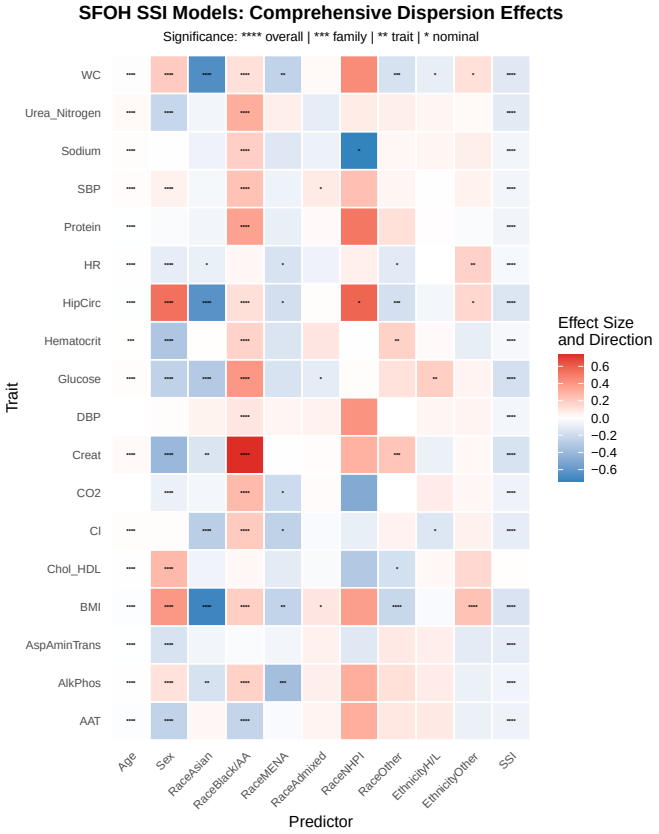

Fig. S31: Heatmap of significant predictors in trait dispersion models with SSI.