The American Journal of Human Genetics

Supplemental Data

Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos

Matthew P. Conomos, Cecelia A. Laurie, Adrienne M. Stilp, Stephanie M. Gogarten, Caitlin P. McHugh, Sarah C. Nelson, Tamar Sofer, Lindsay Fernández-Rhodes, Anne E. Justice, Mariaelisa Graff, Kristin L. Young, Amanda A. Seyerle, Christy L. Avery, Kent D. Taylor, Jerome I. Rotter, Gregory A. Talavera, Martha L. Daviglus, Sylvia Wassertheil-Smoller, Neil Schneiderman, Gerardo Heiss, Robert C. Kaplan, Nora Franceschini, Alex P. Reiner, John R. Shaffer, R. Graham Barr, Kathleen F. Kerr, Sharon R. Browning, Brian L. Browning, Bruce S. Weir, M. Larissa Avilés-Santa, George J. Papanicolaou, Thomas Lumley, Adam A. Szpiro, Kari E. North, Ken Rice, Timothy A. Thornton, and Cathy C. Laurie

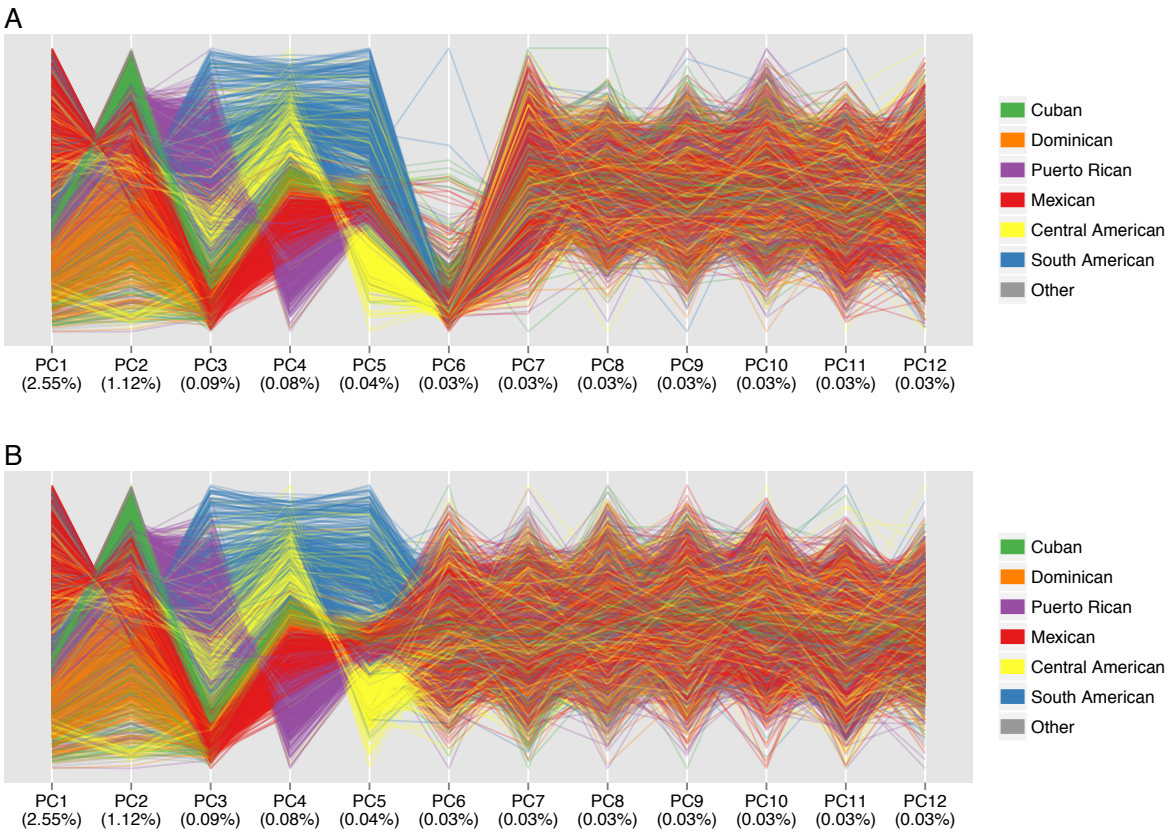

Figure S1: Principal components with and without 19 individuals having a high proportion of East Asian ancestry

(A) Parallel coordinates of the first 12 PCs using all HCHS/SOL individuals. PC6 separates a small number of individuals with predominantly East Asian ancestry.

(B) Parallel coordinates for the first 12 PCs for all individuals except for the 19 individuals with high East Asian ancestry. Percent variance accounted for is given for each PC in the horizontal axis labels. Color-coding is by self-identified background. "Other" includes subjects whose background is multiple, other or had a missing value.

The 12 parallel vertical lines of equal length correspond to the first 12 PCs. Each individual is represented by a set of line segments connecting their PC values.

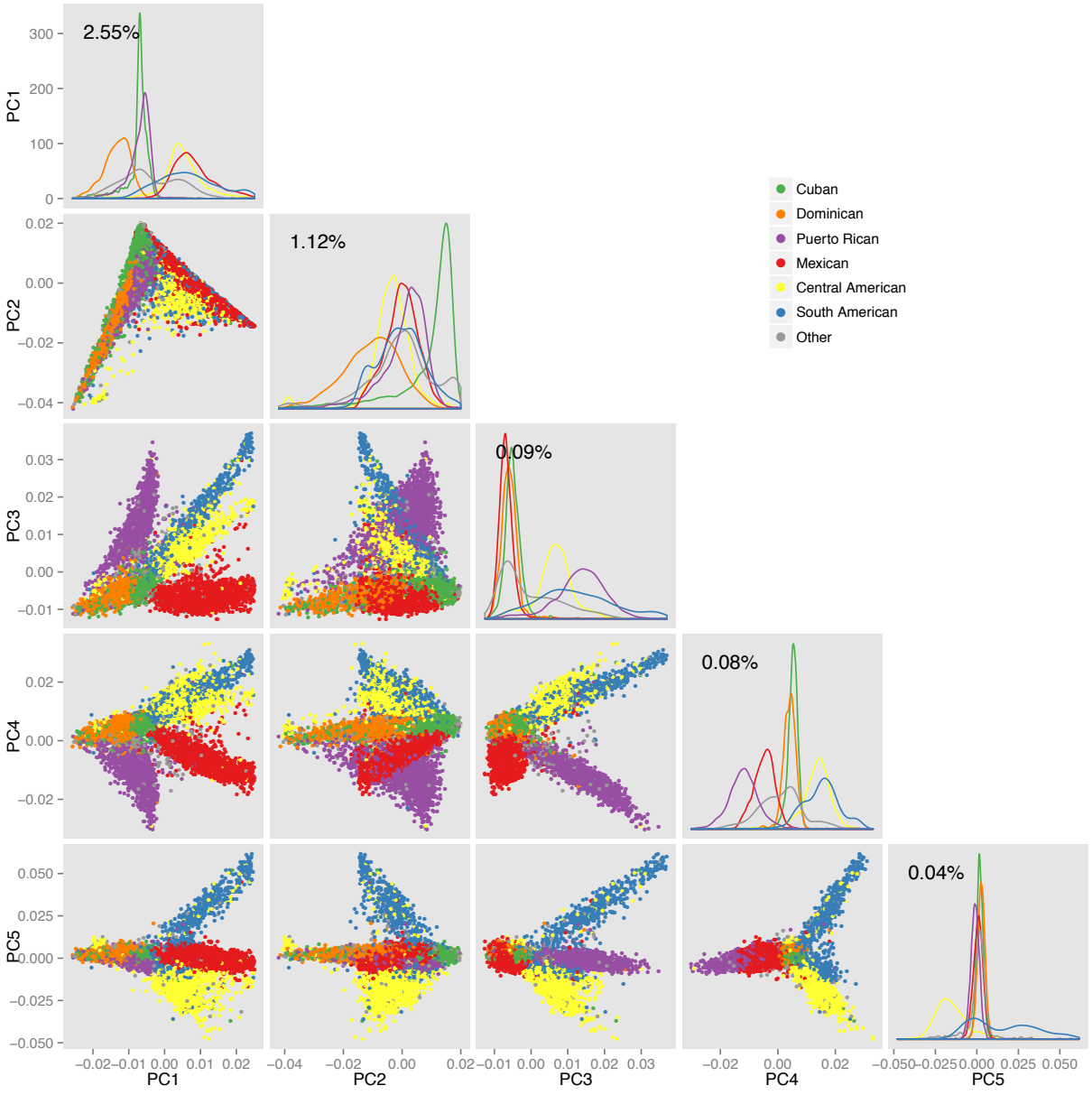

Figure S2: Pairwise PC plots for PCA of HCHS/SOL participants
All participants except the outliers with high East Asian ancestry were included in the PCA. Color-coding is by self-identified background. "Other" includes subjects whose background is multiple, other or had a missing value. The diagonal has density plots for the PC on the horizontal axis and the percent of variance accounted for by that PC.

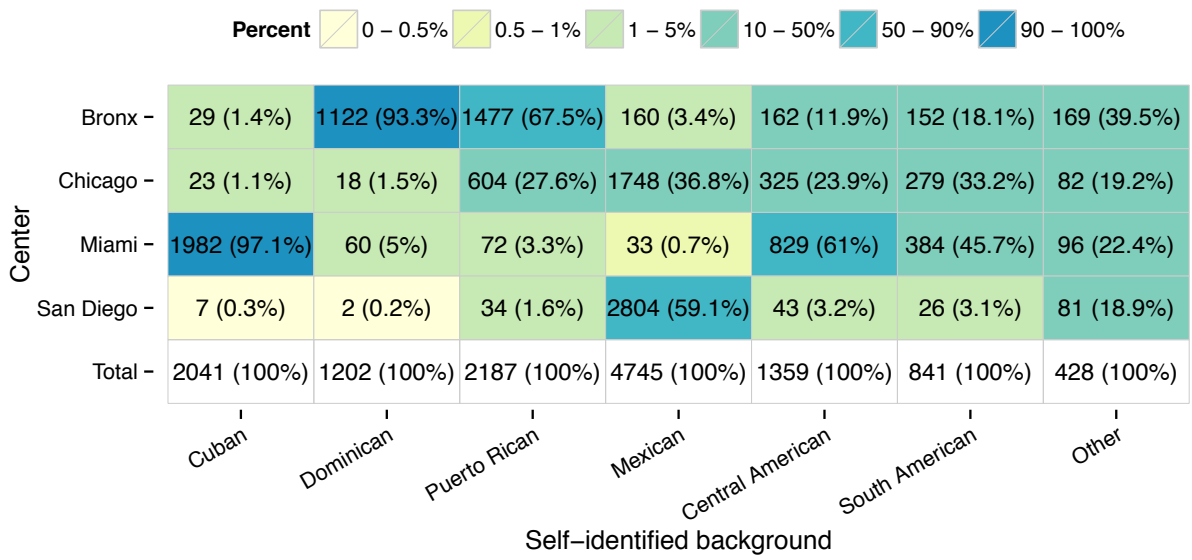

Figure S3: Cross-classification of recruitment center and self-identified background
The number in each box shows the number of genotyped participants from each recruitment center and self-identified background, while the color indicates the percentage of participants from each center for a given self-identified background. “Other” includes subjects whose background is multiple, other or had a missing value.

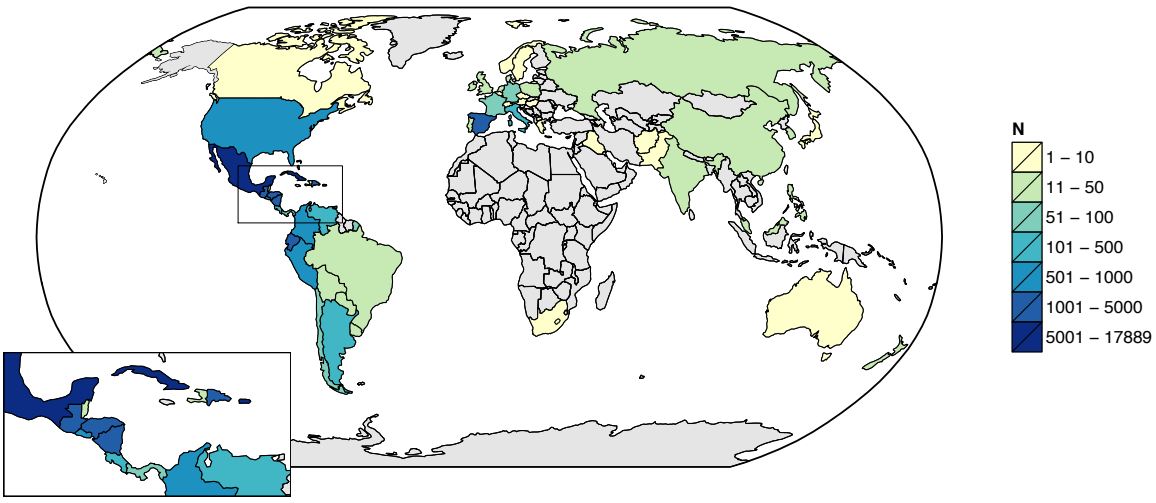

Figure S4: Grandparental origins for HCHS/SOL participants
World map with countries color-coded by the number of grandparents reported to originate from each country, with an inset that shows a larger version of Central America. Specific country origins for 49,626 grandparents were reported by 12,698 genotyped participants. The majority of grandparents come from the Americas. Grandparents from the U.S. are all placed in the lower 48 states (Alaska and Hawaii are colored grey even though it is possible that some U.S. grandparents could have come from those states). Grandparents from Puerto Rico are assigned to Puerto Rico, rather than the U.S.

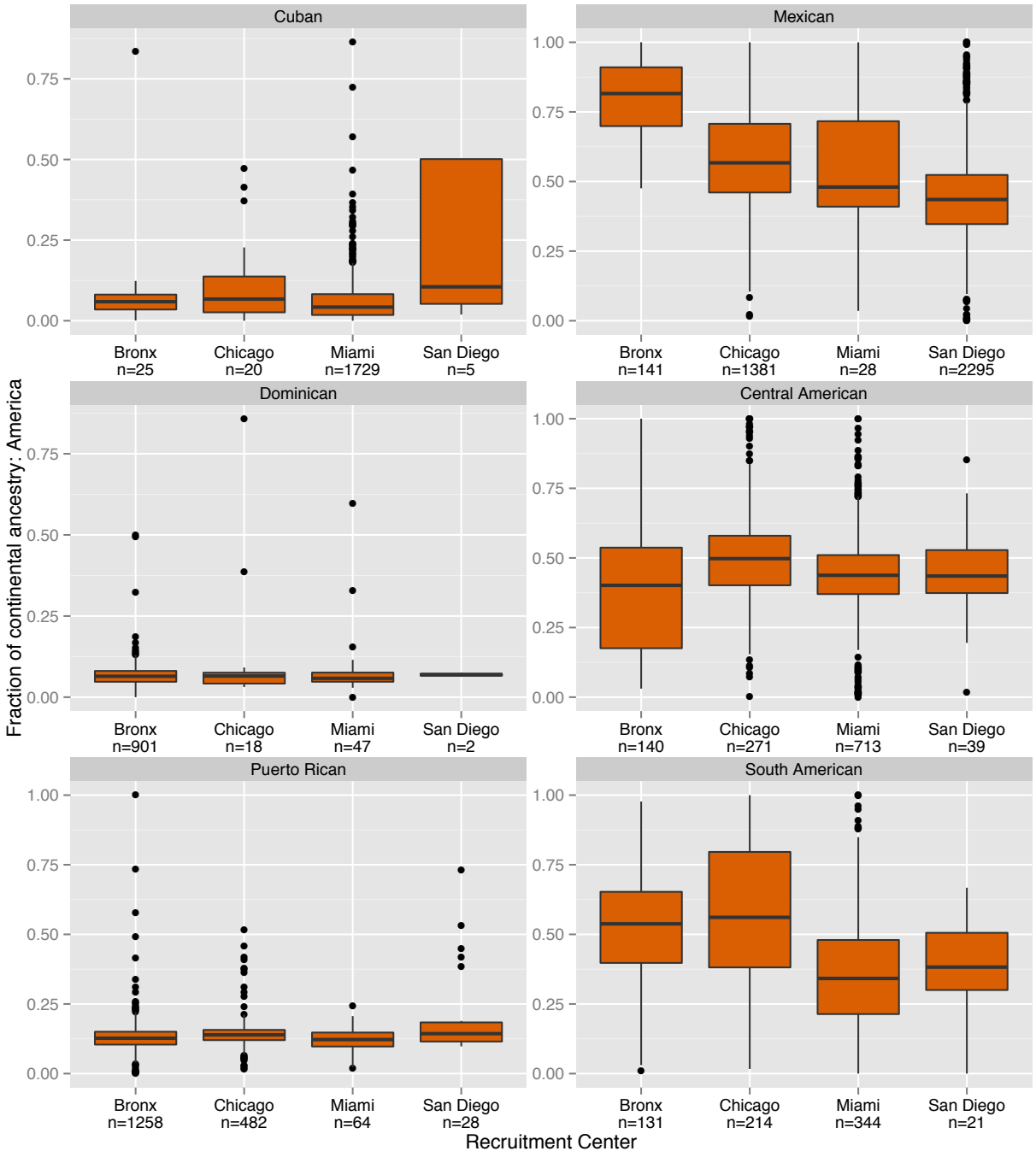

Figure S5: Distributions of the fraction of Amerindian ancestry by recruitment center and self-identified background

Boxplots showing the distributions of Amerindian ancestry proportions estimated for unrelated subjects only. Sample size for each center within a background group is indicated in the axis labels.

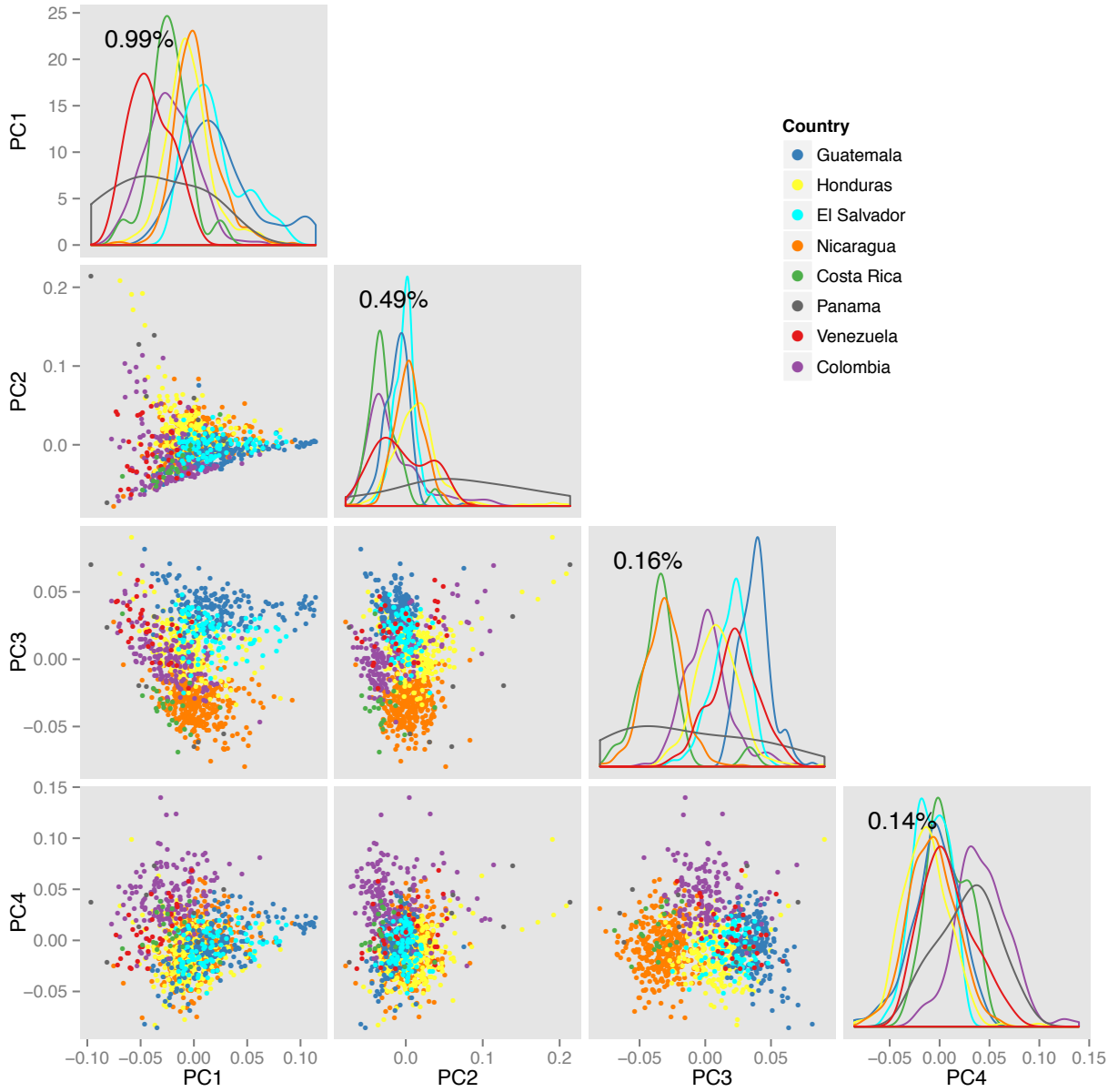

Figure S6: PCA of HCHS/SOL participants with all four grandparents from the same country for Central America, Colombia and Venezuela

Pairs plot showing PCs calculated using 1,094 unrelated subjects with all four grandparents from the countries given in the legend. The group of 37 outliers with high proportions of African ancestry (see Methods) were excluded. The diagonal has density plots for the PC on the horizontal axis and the percent of variance accounted for by that PC. The SNP set used was identical to the overall PCA excluding the outliers with high East Asian ancestry.

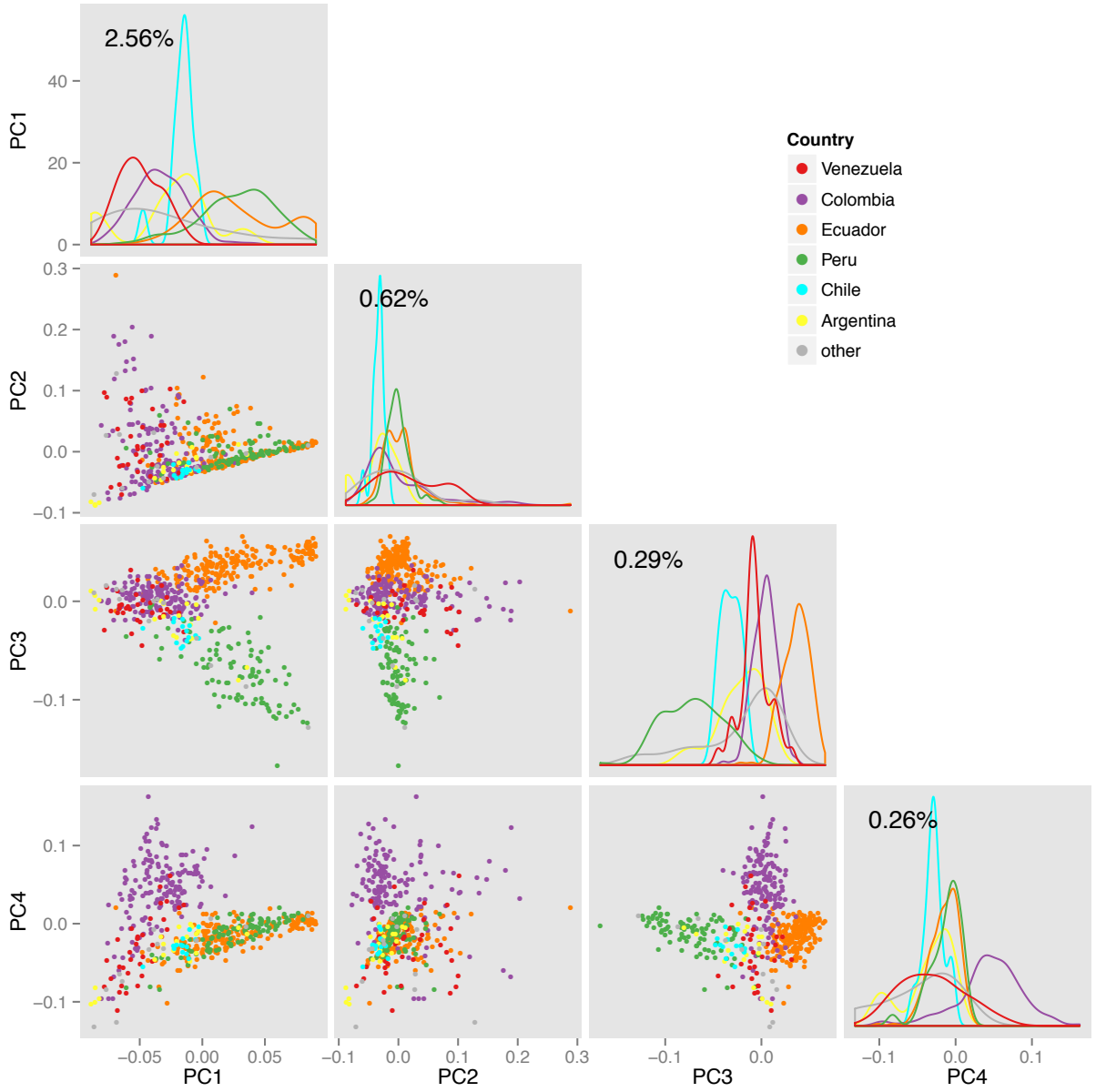

Figure S7: PCA of HCHS/SOL participants with all four grandparents from the same South American country

Pairs plots showing PCs calculated using 552 unrelated subjects with all four grandparents from the same South American country. Two outliers were excluded. The diagonals show density plots of each principal component. The SNP set used was identical to the overall PCA excluding the outliers with high East Asian ancestry.

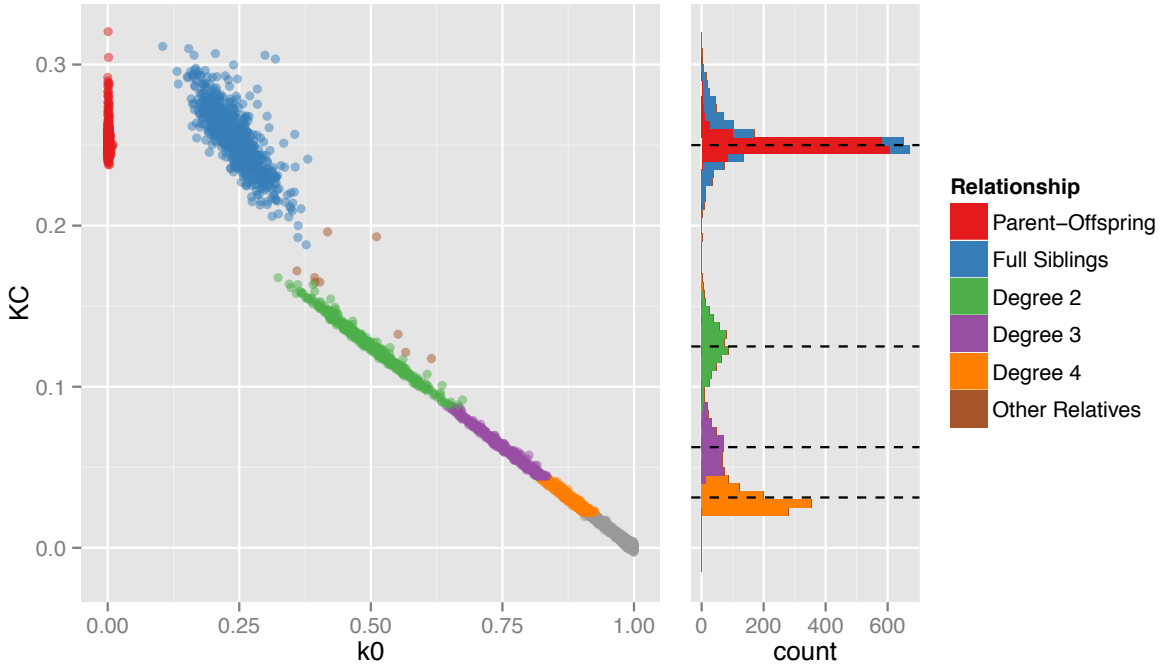

Figure S8: Relatedness estimates for pairs of HCHS/SOL participants
Relatedness was estimated with PC-Relate, excluding outliers with a high proportion of East Asian ancestry. Each point represents a pair of individuals. Because of the large number of unrelated pairs, only a random sample was included in the plot. The horizontal axis has estimates of k_0, which is the probably that, among the two alleles at a locus, zero are identical by descent. The vertical axis has estimates of the kinship coefficient, KC. A marginal histogram of KC color-coded by relationship (excluding unrelated subjects) is shown in the right panel. The black dashed lines represent the expected kinship coefficient for each degree of relatedness. A pair of individuals was assigned a relationship of degree d if $2^{-(d+3/2)} < KC \leq 2^{-(d+1/2)}$. Parent-Offspring pairs were distinguished from full siblings using a threshold of $k_0 = 2^{-11/2}$.

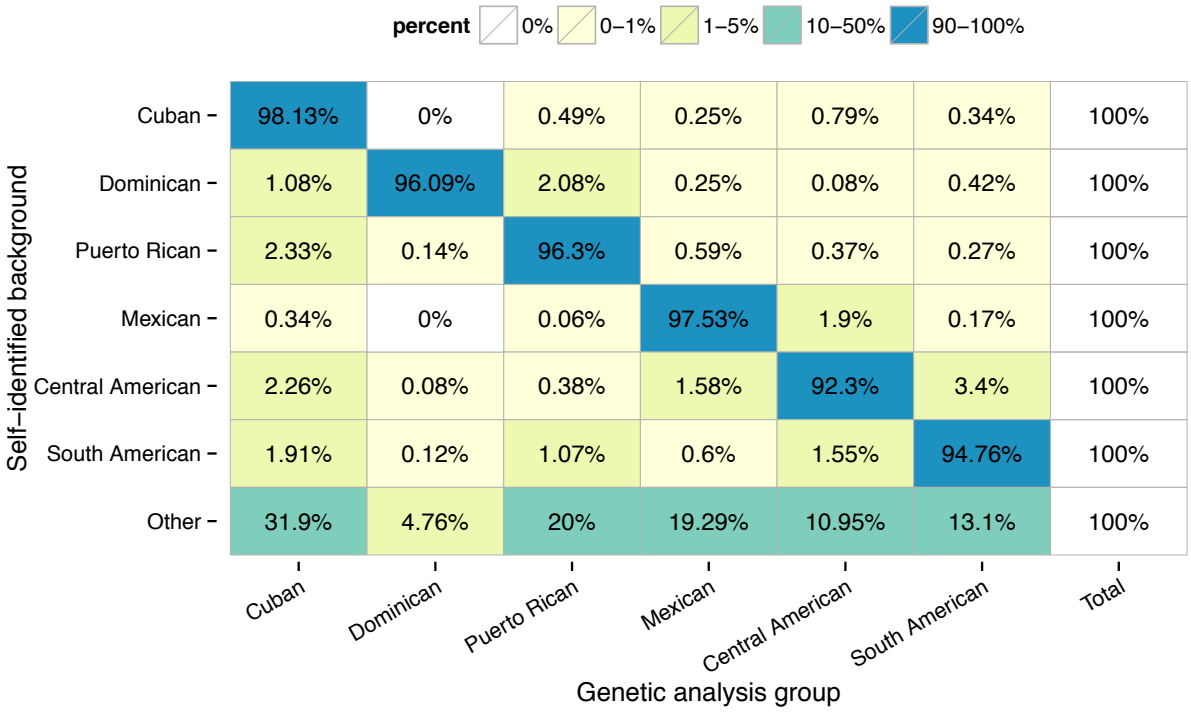

Figure S9: Cross-classification of genetic analysis group and self-identified background group Each cell has the percentage of individuals within a background group that are in the specified genetic analysis group (i.e. percentage of the row total). Color coding is based on this percentage. Individuals in the genetic analysis group having the same value appear on the diagonal. “Other” includes subjects whose background is multiple, other or had a missing value.

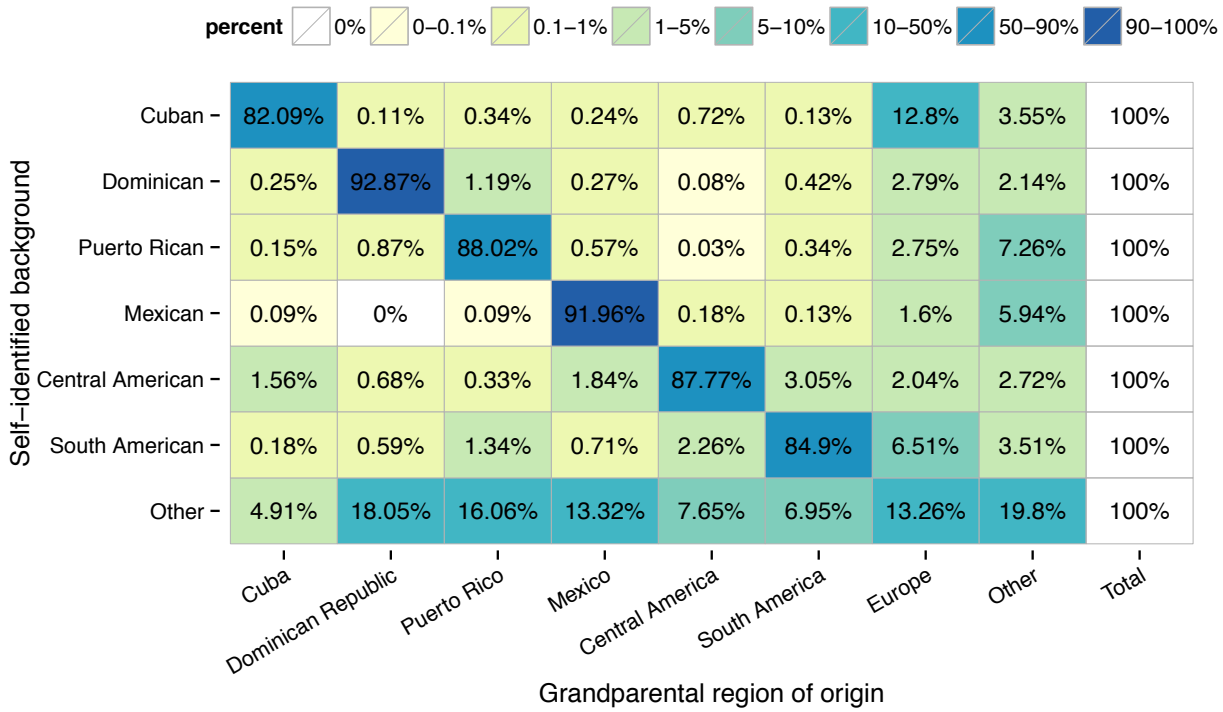

Figure S10: Cross-classification of grandparental region of origin and self-identified background group

Each cell (i,j) has the percentage of all grandparents for participants in the background group for the i^{th} row that originate from the region given in the j^{th} column. Color-coding is based on this percentage. "Other" background includes subjects whose background is multiple, other or had a missing value. "Other" grandparental region includes countries in regions other than the ones listed.

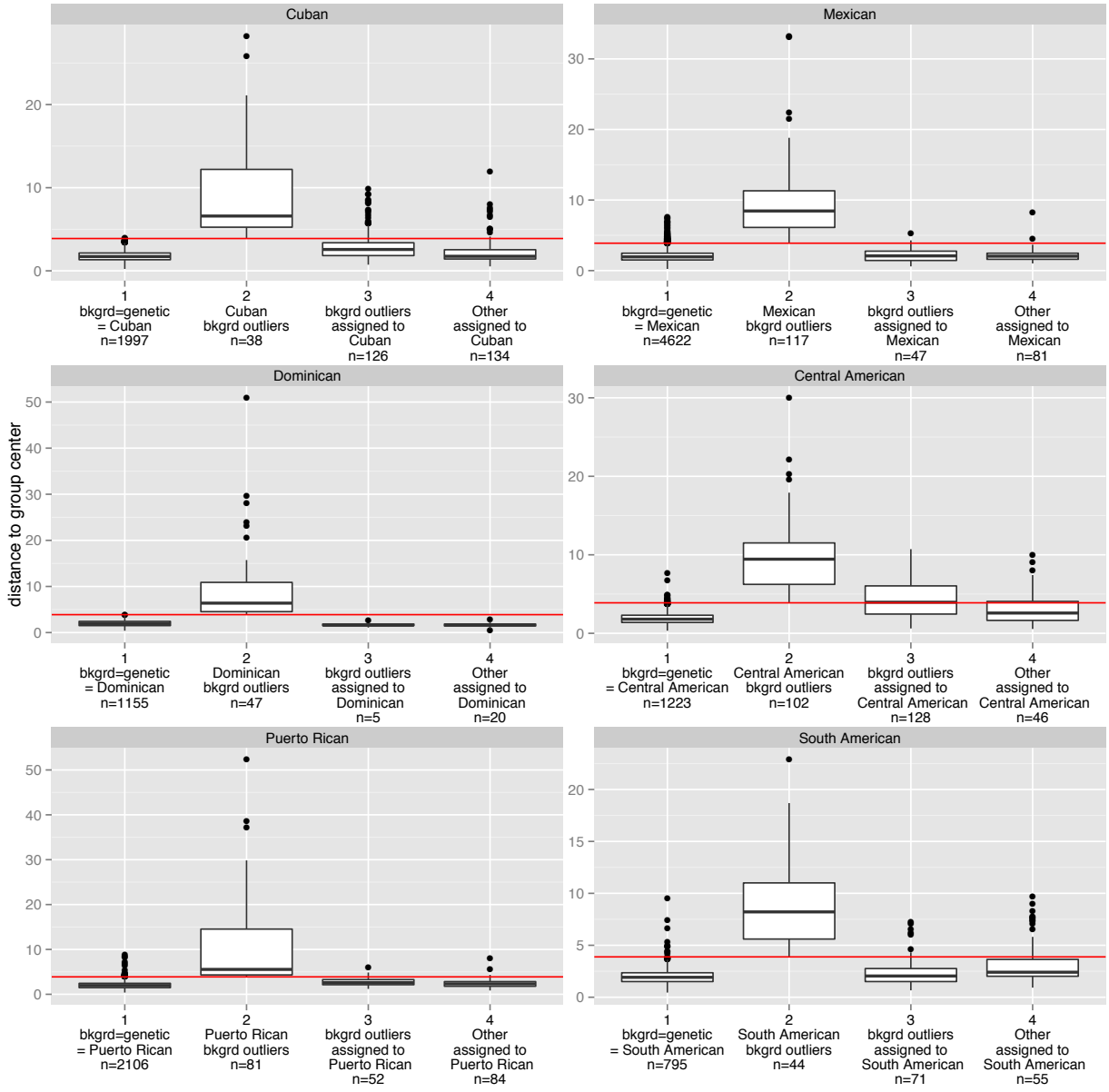

Figure S11: Genetic homogeneity of genetic analysis groups evaluated with Mahalanobis distances Distributions of the Mahalanobis distances between individual points in the five-dimensional PC space and the center of a given hyper-ellipsoid. For the Mexican hyper-ellipsoid, the four boxplots include individuals who belong to (1) both the Mexican self-identified background and Mexican genetic analysis groups, (2) the Mexican self-identified background group and another (not Mexican) genetic analysis group, (3) one of the other (not Mexican) specific self-identified background groups and the Mexican genetic analysis group, and (4) the "Other" (i.e. multiple, other or missing values) self-identified background group and the Mexican genetic analysis group. The red line indicates the distance from the Mexican hyper-ellipsoid boundary to its center, which was one of the criteria for defining genetic analysis group. These descriptions apply to each plot, substituting for "Mexican" the appropriate hyper-ellipsoid label at the top of the plot.

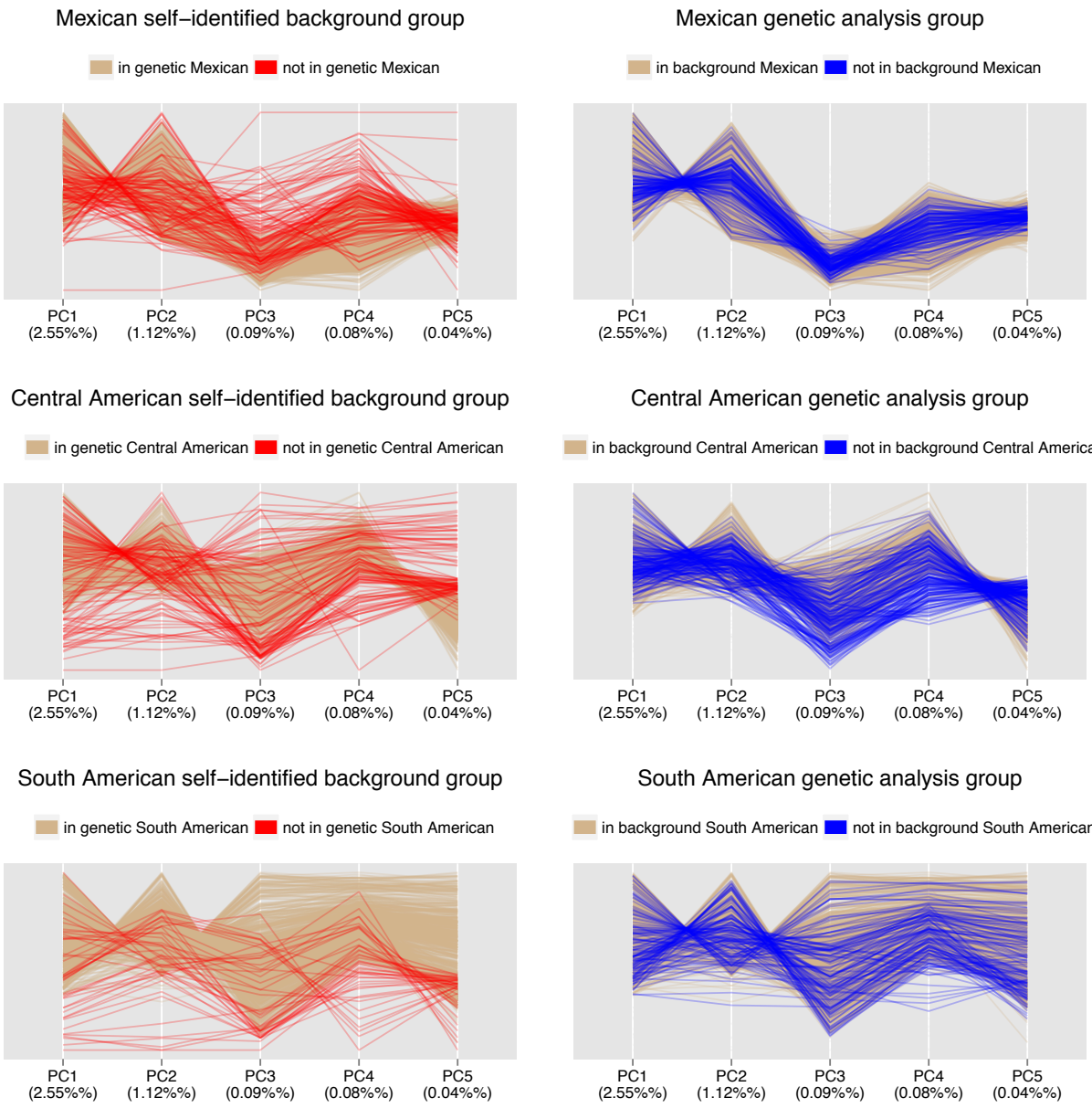

Figure S12: Genetic homogeneity of genetic analysis groups evaluated with PCs: Mainland groups The two Mexican group plots show parallel coordinates for individuals of either Mexican self-identified background (left) or Mexican genetic analysis group (right), from the PCA of all individuals except the outliers with high East Asian ancestry. The vertical scale is the same for both plots. The left plot shows only individuals in the Mexican self-identified background group, distinguishing those that are also in the Mexican genetic analysis group from those that are not. The right plot shows only individuals that are in the Mexican genetic analysis group, distinguishing those that are also in the Mexican self-identified background group from those that are not. The left plot shows that individuals with self-identified Mexican background that are not in the Mexican genetic analysis group (red) consist of outliers for one or more PCs. The right plot shows that self-identified non-Mexican background individuals who are in the Mexican genetic analysis group are not outliers. The same description applies to the other two groups shown here, substituting for “Mexican” the group name in the plot title.

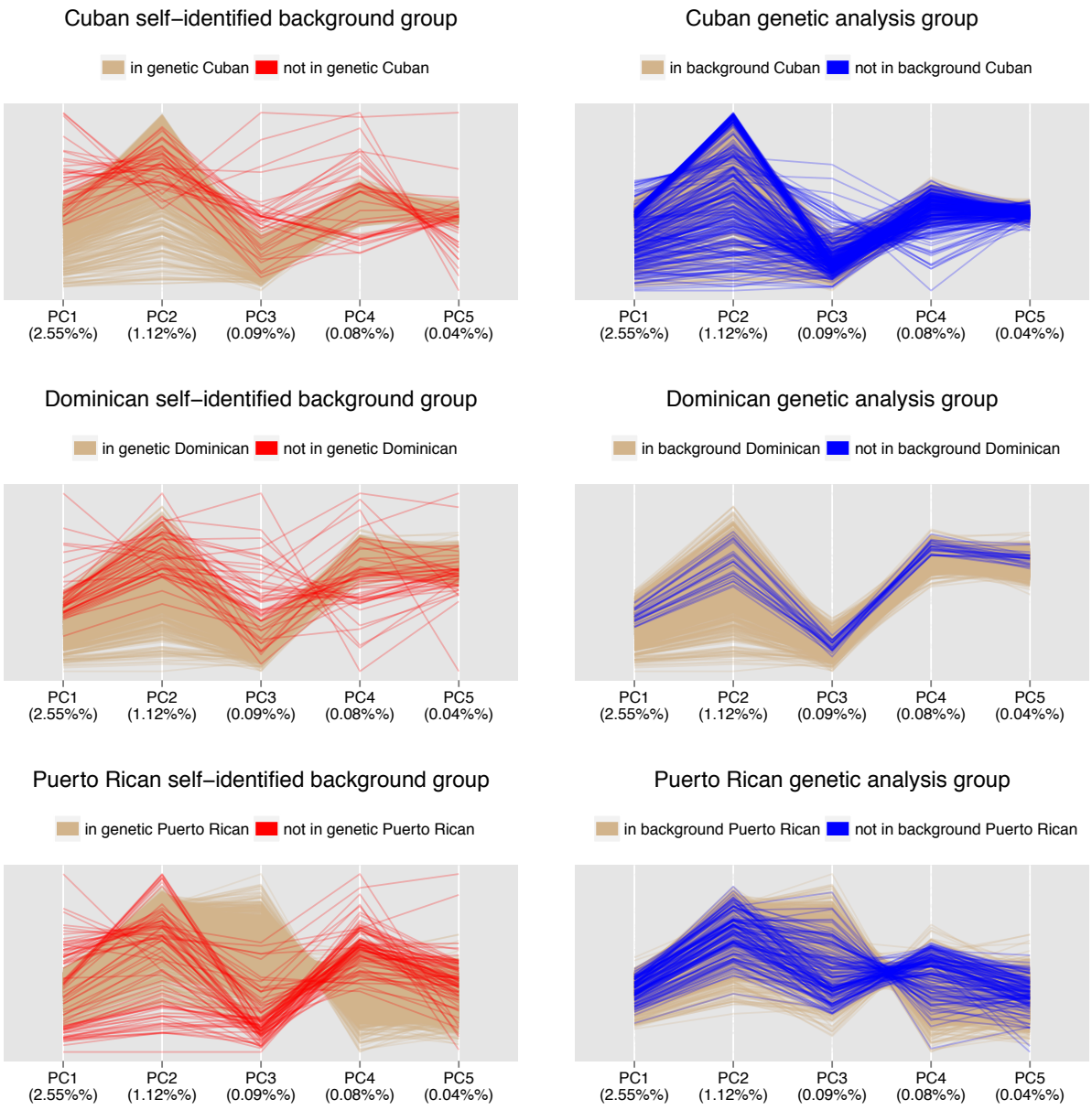

Figure S13: Genetic homogeneity of genetic analysis groups evaluated with PCs: Caribbean groups The two Cuban group plots show parallel coordinates for individuals of either Cuban self-identified background (left) or Cuban genetic analysis group (right), from the PCA of all individuals except the outliers with high East Asian ancestry. The vertical scale is the same for both plots. The left plot shows only individuals in the Cuban self-identified background group, distinguishing those that are also in the Cuban genetic analysis group from those that are not. The right plot shows only individuals that are in the Cuban genetic analysis group, distinguishing those that are also in the Cuban self-identified background group from those that are not. The left plot shows that individuals with self-identified Cuban background that are not in the Cuban genetic analysis group (red) consist of outliers for one or more PCs. The right plot shows that self-identified non-Cuban background individuals who are in the Cuban genetic analysis group are not outliers. The same description applies to the other two groups shown here, substituting for “Cuban” the group name in the plot title.

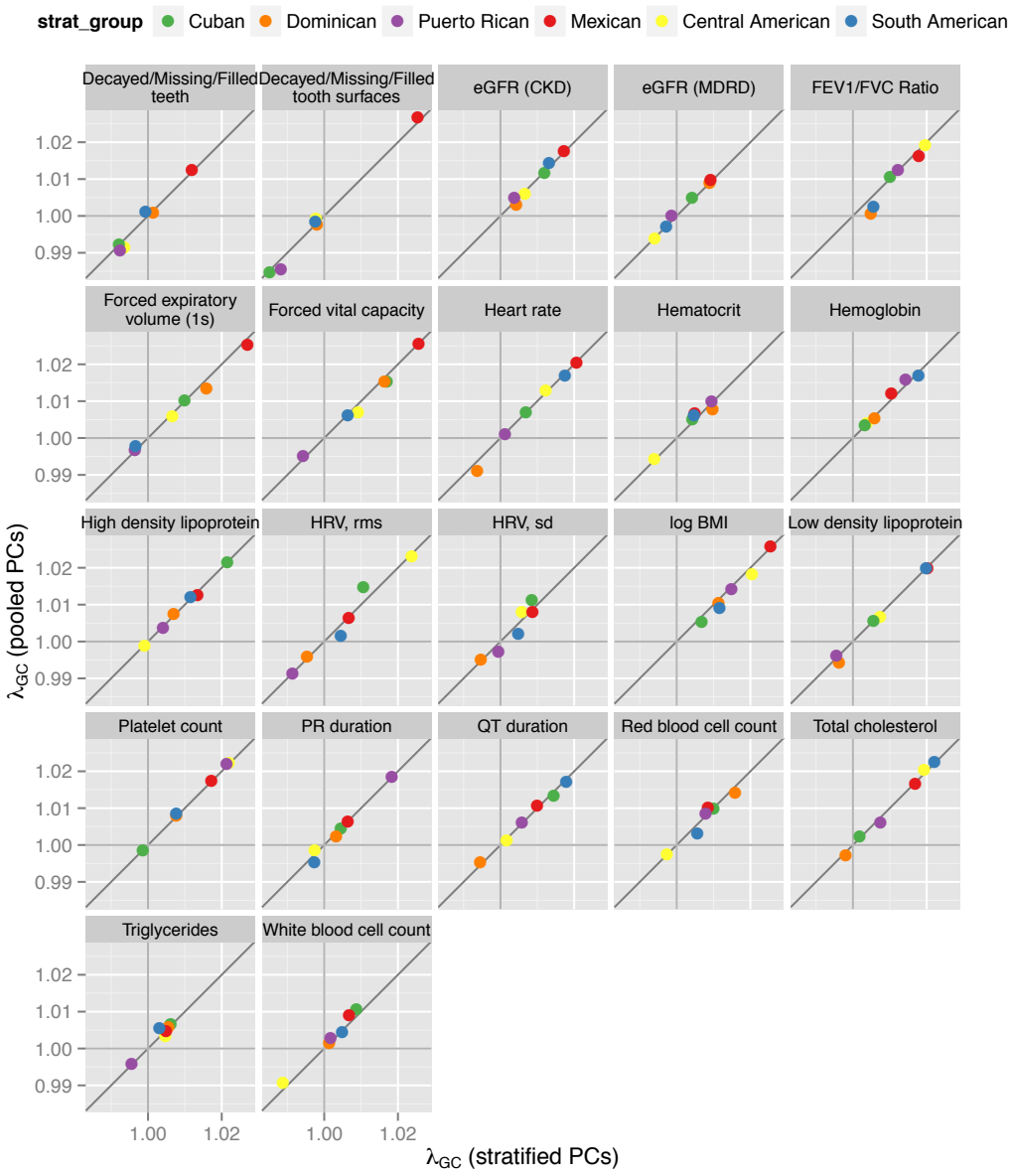

Figure S14: Genomic inflation factor (λ_{GC}) within each genetic analysis group for pooled PCs versus group-specific PCs

For each of 22 biomedical traits, λ_{GC} was calculated within each genetic analysis group for analyses with pooled principal components versus group-specific principal components. Five PCs were used in each case. SNPs were filtered on effective N $>$ 120 within each group. The median number of SNPs passing this filter for each group was: Cuban (8,206,000), Dominican (7,608,000), Puerto Rican (8,607,000), Mexican (8,660,000), Central American (6,736,000), and South American (5,817,000).

			λ_{GC}			
			LMM			LM
Trait group	Trait	N	no PCs	PC 1-5	PC 1-20	PC 1-5
Anthropometrics	log BMI	12,705	1.081	1.050	1.049	1.152
Blood cell count	Hematocrit	12,502	1.706	1.022	1.022	1.072
Blood cell count	Hemoglobin	12,502	1.901	1.030	1.030	1.073
Blood cell count	Platelet count	12,491	1.140	1.045	1.044	1.103
Blood cell count	Red blood cell count	12,502	1.152	1.058	1.059	1.124
Blood cell count	White blood cell count	11,809	2.168	1.057	1.057	1.121
Chronic kidney disease	eGFR (CKD)	12,739	5.672	1.047	1.045	1.111
Chronic kidney disease	eGFR (MDRD)	12,739	5.840	0.999	0.998	1.073
Dental	Decayed/Missing/Filled teeth	11,803	1.797	0.997	0.997	1.044
Dental	Decayed/Missing/Filled tooth surfaces	11,803	1.557	1.019	1.019	1.065
Electrocardiography	Heart rate	10,216	1.889	1.044	1.046	1.076
Electrocardiography	HRV, rms	10,798	1.253	1.025	1.025	1.051
Electrocardiography	HRV, sd	10,798	1.339	1.018	1.020	1.034
Electrocardiography	PR duration	11,686	1.645	1.041	1.041	1.099
Electrocardiography	QT duration	11,932	1.272	1.021	1.021	1.086
Lipids	High density lipoprotein	12,730	1.938	1.032	1.032	1.099
Lipids	Low density lipoprotein	12,467	1.090	1.045	1.046	1.115
Lipids	Total cholesterol	12,731	1.098	1.031	1.031	1.106
Lipids	Triglycerides	12,730	4.407	1.001	1.001	1.053
Pulmonary disease	FEV1/FVC Ratio	11,832	1.402	1.073	1.073	1.126
Pulmonary disease	Forced expiratory volume (1s)	11,833	6.972	1.043	1.043	1.101
Pulmonary disease	Forced vital capacity	11,832	9.287	1.036	1.036	1.097

Table S1: Genomic inflation factor (λ_{GC}) from GWAS for models that differ in accounting for relatedness and ancestry

λ_{GC} values for GWAS of 22 different biomedical traits run with 1) no PCs, 2) the first five or 3) the first twenty PCs to control for confounding by ancestry, and 4) without random effects to control for correlation structure. The model for each analysis included sex, center, age, sampling weight, and other trait-specific covariates. Models in the "LMM" category also included random effects for household, block group, and polygenic effects due to relatedness, and the "LM" model used a simple linear model that ignored correlations among subjects due to relatedness, household and block group. Genetic analysis group was not included. Sample size (N) varied by trait-specific exclusion criteria. All λ_{GC} values were calculated using autosomal SNPs filtered by an effective minor allele count, effN > 120, as described in Methods. The number of SNPs used to calculate λ_{GC} varied by analysis due to the varying sample size, but a median of 1,898,000 genotyped SNPs and 12,030,000 imputed SNPs were used in the calculations.

		AIC Difference	
Trait group	Trait	PC5 - PC5g	PC20 - PC20g
Anthropometrics	log BMI	18.4	16.8
Blood cell count	Hematocrit	-6.6	-6.9
Blood cell count	Hemoglobin	-6.6	-7.4
Blood cell count	Platelet count	-4.7	-4.3
Blood cell count	Red blood cell count	-6.6	-7.0
Blood cell count	White blood cell count	-0.5	-0.9
Chronic kidney disease	eGFR (CKD)	-1.9	-0.9
Chronic kidney disease	eGFR (MDRD)	-1.1	-0.2
Dental	Decayed/Missing/Filled teeth	19.3	18.9
Dental	Decayed/Missing/Filled tooth surfaces	21.8	21.6
Electrocardiography	Heart rate	1.8	1.1
Electrocardiography	HRV, rms	-0.4	-0.8
Electrocardiography	HRV, sd	-2.7	-2.9
Electrocardiography	PR duration	-0.8	-0.3
Electrocardiography	QT duration	3.3	2.2
Lipids	High density lipoprotein	10.1	8.7
Lipids	Low density lipoprotein	-5.4	-5.1
Lipids	Total cholesterol	-4.7	-4.2
Lipids	Triglycerides	3.7	2.8
Pulmonary disease	FEV1/FVC Ratio	0.1	-0.4
Pulmonary disease	Forced expiratory volume (1s)	10.9	11.9
Pulmonary disease	Forced vital capacity	15.8	15.6

Table S2: Evaluation of the contribution of genetic analysis group to regression models using AIC Each of 22 traits was analyzed using an LMM in which the model included fixed effects for sex, age, recruitment center, sampling weight, and random effects for census block group, household, and polygenic effects due to relatedness. Models also included trait-specific covariates in some cases. In addition, each model included either the first five or the first twenty PCs as fixed effects to adjust for ancestry, either with or without genetic analysis group. The column "PC5-PC5g" is PC5 (AIC for the model with PC1-5 but no genetic analysis group) minus PC5g (AIC for the model with PC1-5 and genetic analysis group); similarly for the column labeled as "PC20 - PC20g", except using PCs 1-20. A positive difference indicates that the model with genetic analysis group is a better fit than the model without genetic analysis group.

		AIC difference
Trait group	Trait	PC5b - PC5g
Anthropometrics	log BMI	3.0
Blood cell count	Hematocrit	-3.0
Blood cell count	Hemoglobin	-2.1
Blood cell count	Platelet count	-4.2
Blood cell count	Red blood cell count	-1.3
Blood cell count	White blood cell count	5.2
Chronic kidney disease	eGFR (CKD)	0.5
Chronic kidney disease	eGFR (MDRD)	3.2
Dental	Decayed/Missing/Filled teeth	-10.2
Dental	Decayed/Missing/Filled tooth surfaces	-14.4
Electrocardiography	Heart rate	4.3
Electrocardiography	HRV, rms	2.5
Electrocardiography	HRV, sd	1.6
Electrocardiography	PR duration	0.8
Electrocardiography	QT duration	-6.8
Lipids	High density lipoprotein	0.4
Lipids	Low density lipoprotein	0.0
Lipids	Total cholesterol	0.8
Lipids	Triglycerides	1.1
Pulmonary disease	FEV1/FVC Ratio	0.3
Pulmonary disease	Forced expiratory volume (1s)	1.8
Pulmonary disease	Forced vital capacity	-3.0

Table S3: Evaluation of the contribution of genetic analysis group versus self-identified background group to regression models using AIC

Each of 22 traits was analyzed using an LMM in which the model included fixed effects for sex, recruitment center, age, sampling weight, and the first five PCs, and random effects for census block group, household, and polygenic effects due to relatedness. The model labeled as "PC5g" also had genetic analysis group as a fixed effect, while "PC5b" had self-identified background as a fixed effect. Models also included trait-specific covariates in some cases. Only participants with one of the six specific self-identified background groups were included in both analyses. A positive AIC difference indicates that the model with genetic analysis group is a better fit than the model with background group.

		CV	
Trait group	Trait	gengrp	bkgrd
Lipids	High density lipoprotein	0.06	0.07
Dental	Decayed/Missing/Filled teeth	0.06	0.06
Dental	Decayed/Missing/Filled tooth surfaces	0.07	0.06
Lipids	Triglycerides	0.07	0.06
Electrocardiography	QT duration	0.07	0.06
Electrocardiography	HRV, sd	0.08	0.08
Chronic kidney disease	eGFR (MDRD)	0.08	0.09
Lipids	Low density lipoprotein	0.08	0.09
Electrocardiography	HRV, rms	0.09	0.08
Electrocardiography	PR duration	0.09	0.09
Lipids	Total cholesterol	0.10	0.10
Pulmonary disease	Forced vital capacity	0.12	0.11
Chronic kidney disease	eGFR (CKD)	0.12	0.12
Blood cell count	Hematocrit	0.13	0.12
Blood cell count	Hemoglobin	0.14	0.15
Anthropometrics	log BMI	0.15	0.14
Blood cell count	White blood cell count	0.15	0.15
Blood cell count	Platelet count	0.15	0.16
Blood cell count	Red blood cell count	0.16	0.14
Electrocardiography	Heart rate	0.16	0.16
Pulmonary disease	Forced expiratory volume (1s)	0.17	0.17
Pulmonary disease	FEV1/FVC Ratio	0.27	0.27

Table S4: Comparison of residual variance heterogeneity for genetic analysis versus self-identified background groups

Coefficient of variation for residual variance by group is given for each of 22 biomedical traits. These statistics derive from LMMs assuming homoscedasticity and using either genetic analysis group ("gengrp") or self-identified background group ("bkgrd") as a fixed effect covariate. The sample set was the same for both models, and it included only individuals who had one of the six specific values for both genetic analysis group and self-identified background. Both models included fixed effects for sex, age, center, sampling weight, and PCs 1-5, and random effects for block group, household, and polygenic effects due to relatedness. In some cases, trait-specific fixed effect covariates were also included in both models.

		λ_{GC}	
Trait group	Trait	homoscedastic	heteroscedastic
Dental	Decayed/Missing/Filled teeth	0.996	1.006
Chronic kidney disease	eGFR (MDRD)	0.999	1.013
Lipids	Triglycerides	1.001	1.009
Electrocardiography	HRV, sd	1.019	1.010
Dental	Decayed/Missing/Filled tooth surfaces	1.019	1.013
Electrocardiography	QT duration	1.021	1.016
Blood cell count	Hematocrit	1.022	1.013
Electrocardiography	HRV, rms	1.025	1.016
Blood cell count	Hemoglobin	1.029	1.020
Lipids	Total cholesterol	1.031	1.014
Lipids	High density lipoprotein	1.032	1.025
Pulmonary disease	Forced vital capacity	1.035	1.031
Electrocardiography	PR duration	1.041	1.024
Pulmonary disease	Forced expiratory volume (1s)	1.042	1.026
Electrocardiography	Heart rate	1.044	1.020
Blood cell count	Platelet count	1.045	1.026
Lipids	Low density lipoprotein	1.046	1.027
Chronic kidney disease	eGFR (CKD)	1.047	1.029
Anthropometrics	log BMI	1.049	1.041
Blood cell count	White blood cell count	1.056	1.027
Blood cell count	Red blood cell count	1.058	1.030
Pulmonary disease	FEV1/FVC Ratio	1.072	1.027

Table S5: Genomic inflation (λ_{GC}) for models with homoscedasticity and heteroscedasticity
Each of 22 traits was analyzed using an LMM in which the model included fixed effects for sex, recruitment center, age, sampling weight, genetic analysis group, and the first five PCs, and random effects for census block group, household, and polygenic effects due to relatedness. The model labeled "homoscedastic" fit one residual variance for all subjects, while the model labeled "heteroscedastic" fit a different residual variance for each genetic analysis group. Models also included trait-specific covariates in some cases. The data in this table are plotted in Figure 8B.

		λ_{GC}		
Trait group	Trait	gengrp	background	no group
Lipids	Triglycerides	1.009	1.012	1.001
Blood cell count	Hematocrit	1.013	1.018	1.022
Lipids	Total cholesterol	1.014	1.014	1.031
Electrocardiography	QT duration	1.016	1.016	1.022
Blood cell count	Hemoglobin	1.020	1.023	1.030
Lipids	High density lipoprotein	1.025	1.025	1.032
Blood cell count	Platelet count	1.026	1.026	1.045
Blood cell count	White blood cell count	1.027	1.028	1.057
Lipids	Low density lipoprotein	1.027	1.026	1.045
Pulmonary disease	FEV1/FVC Ratio	1.027	1.026	1.073
Blood cell count	Red blood cell count	1.030	1.035	1.058
Anthropometrics	log BMI	1.041	1.042	1.050

Table S6: Comparison of λ_{GC} for models using different group definitions
The table shows λ_{GC} for models with three different group definitions. All GWAS models were adjusted for sex, center, age, sampling weight, and PCs 1-5, and trait-specific covariates as fixed effects, and random effects for block group, household, and genetic relatedness were included. In addition, both models with group variables were run using heterogeneous residual variance. The "no group" model uses the full set of 12,784 subjects, minus any trait-specific exclusions. The "gengrp" model includes genetic analysis group as a fixed effect and excludes 37 subjects that have missing genetic analysis group. The "background" model includes self-reported background as a fixed effect, and excludes 425 subjects with missing self-reported background.

Trait group	Trait	N hits	slope [95% CI] background	slope [95% CI] no group	Refs
Blood cell count	White blood cell count	14	0.925 [0.909-0.941]	1.059 [1.044-1.073]	[1–4]
Lipids	High density lipoprotein	72	0.931 [0.922-0.940]	0.989 [0.987-0.991]	[5–7]
Anthropometrics	log BMI	104	0.936 [0.913-0.958]	0.988 [0.975-1.001]	[8–12]
Lipids	Triglycerides	40	0.951 [0.945-0.957]	1.020 [1.017-1.023]	[6,7]
Pulmonary disease	FEV1/FVC Ratio	21	0.960 [0.925-0.995]	0.946 [0.893-1.000]	[13–15]
Blood cell count	Hemoglobin	19	0.965 [0.944-0.986]	1.023 [1.015-1.031]	[2,16–18]
Blood cell count	Hematocrit	10	0.966 [0.904-1.028]	1.047 [1.017-1.077]	[2,17,19]
Blood cell count	Platelet count	59	0.969 [0.948-0.990]	0.970 [0.960-0.979]	[2,19–22]
Electrocardiography	QT duration	35	0.981 [0.963-0.999]	0.976 [0.970-0.982]	[23]
Lipids	Total cholesterol	74	0.989 [0.981-0.997]	0.976 [0.972-0.981]	[6,7]
Blood cell count	Red blood cell count	18	0.995 [0.963-1.027]	1.009 [0.992-1.026]	[2,17,18,24]
Lipids	Low density lipoprotein	58	0.999 [0.991-1.008]	0.976 [0.972-0.980]	[6,7]
	mean		0.964 [0.950-0.978]	0.998 [0.979-1.017]	

Table S7: Comparison of Wald test statistics for the effects of SNPs with previously published trait associations, using models with different ethnic group definitions.
For 12 biomedical traits, association tests were performed in HCHS/SOL to assess power to detect known hits from the literature for models utilizing genetic analysis group, self-identified background group, or no group variable. See Figure 10 for plots of these data. All models included adjustment for sex, center, age, sampling weight, PCs 1-5, and trait-specific covariates. Random effects included block group, household, and genetic relatedness. The models also included genetic analysis group ("gengrp", up to 12,747 subjects); self-reported background ("background", up to 12,359 subjects); or no group variable ("no group", up to 12,784 subjects) as a covariate. For models using a group variable ("gengrp" and "background"), heterogeneous residual variance was fit by that group. The slopes and 95% confidence intervals are from a linear regression (through the origin) of test statistics for a given trait from either the background or no-group model, regressed on the corresponding test statistics from the genetic analysis group model. See Figure 10A for an example using log BMI. All test statistics are from the Wald test for the effect of the published SNP hit ($\chi^2_{(1)}$) in HCHS/SOL data and were adjusted for λ_{GC} before fitting the linear model. The number of previously identified variants affecting each trait ("N hits") is shown. The "mean" row gives the mean of the slope estimates for each model type and its 95% confidence interval. References for papers used to determine known hits for each trait are shown in the "Refs" column.

References

[1] Crosslin, D. R., McDavid, A., Weston, N., Nelson, S. C., Zheng, X., Hart, E., de Andrade, M., Kullo, I. J., McCarty, C. A., Doheny, K. F., et al. (2012). Genetic variants associated with the white blood cell count in 13,923 subjects in the eMERGE Network. Hum Genet *131*, 639–652.

[2] Kamatani, Y., Matsuda, K., Okada, Y., Kubo, M., Hosono, N., Daigo, Y., Nakamura, Y., and Kamatani, N. (2010). Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet *42*, 210–215.

[3] Nalls, M. A., Couper, D. J., Tanaka, T., van Rooij, F. J. A., Chen, M.-H., Smith, A. V., Toniolo, D., Zakai, N. A., Yang, Q., Greinacher, A., et al. (2011). Multiple loci are associated with white blood cell phenotypes. PLoS Genet *7*, e1002113.

[4] Reiner, A. P., Lettre, G., Nalls, M. A., Ganesh, S. K., Mathias, R., Austin, M. A., Dean, E., Arepalli, S., Britton, A., Chen, Z., et al. (2011). Genome-wide association study of white blood cell count in 16,388 African Americans: the continental origins and genetic epidemiology network (COGENT). PLoS Genet *7*, e1002108.

[5] Elbers, C. C., Guo, Y., Tragante, V., Van Iperen, E. P., Lanktree, M. B., Castillo, B. A., Chen, F., Yanek, L. R., Wojczynski, M. K., Li, Y. R., et al. (2012). Gene-centric meta-analysis of lipid traits in African, East Asian and Hispanic populations. Plos One *7*.

[6] Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature *466*, 707–713.

[7] Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M. L., Mora, S., et al. (2013). Discovery and refinement of loci associated with lipid levels. Nat Genet *45*, 1274–1283.

[8] Berndt, S. I., Gustafsson, S., Magi, R., Ganna, A., Wheeler, E., Feitosa, M. F., Justice, A. E., Monda, K. L., Croteau-Chonka, D. C., Day, F. R., et al. (2013). Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat Genet *45*, 501–512.

[9] Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J., et al. (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature *518*, 197–206.

[10] Monda, K. L., Chen, G. K., Taylor, K. C., Palmer, C., Edwards, T. L., Lange, L. A., Ng, M. C. Y., Adeyemo, A. A., Allison, M. A., Bielak, L. F., et al. (2013). A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. Nat Genet *45*, 690–696.

[11] Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., Lango Allen, H., Lindgren, C. M., Luan, J., Magi, R., et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet *42*, 937–948.

[12] Wen, W., Cho, Y.-S., Zheng, W., Dorajoo, R., Kato, N., Qi, L., Chen, C.-H., Delahanty, R. J., Okada, Y., Tabara, Y., et al. (2012). Meta-analysis identifies common variants associated with body mass index in east Asians. Nat Genet *44*, 307–311.

[13] Hancock, D. B., Eijgelsheim, M., Wilk, J. B., Gharib, S. A., Loehr, L. R., Marciante, K. D., Franceschini, N., van Durme, Y. M. T. A., Chen, T.-H., Barr, R. G., et al. (2010). Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. Nat Genet *42*, 45–52.

[14] Repapi, E., Sayers, I., Wain, L. V., Burton, P. R., Johnson, T., Obeidat, M., Zhao, J. H., Ramasamy, A., Zhai, G., Vitart, V., et al. (2010). Genome-wide association study identifies five loci associated with lung function. Nat Genet *42*, 36–44.

[15] Soler Artigas, M., Loth, D. W., Wain, L. V., Gharib, S. A., Obeidat, M., Tang, W., Zhai, G., Zhao, J. H., Smith, A. V., Huffman, J. E., et al. (2011). Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet *43*, 1082–1090.

[16] Chambers, J. C., Zhang, W., Li, Y., Sehmi, J., Wass, M. N., Zabaneh, D., Hoggart, C., Bayele, H., McCarthy, M. I., Peltonen, L., et al. (2009). Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. Nat Genet *41*, 1170–1172.

[17] Ganesh, S. K., Zakai, N. A., van Rooij, F. J. A., Soranzo, N., Smith, A. V., Nalls, M. A., Chen, M.-H., Kottgen, A., Glazer, N. L., Dehghan, A., et al. (2009). Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nat Genet *41*, 1191–1198.

[18] van der Harst, P., Zhang, W., Mateo Leach, I., Rendon, A., Verweij, N., Sehmi, J., Paul, D. S., Elling, U., Allayee, H., Li, X., et al. (2012). Seventy-five genetic loci influencing the human red blood cell. Nature *492*, 369–375.

[19] Li, J., Glessner, J. T., Zhang, H., Hou, C., Wei, Z., Bradfield, J. P., Mentch, F. D., Guo, Y., Kim, C., Xia, Q., et al. (2013). GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children. Hum Mol Genet *22*, 1457–1464.

[20] Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic, J., Elling, U., Goodall, A. H., Labrune, Y., et al. (2011). New gene functions in megakaryopoiesis and platelet formation. Nature *480*, 201–208.

[21] Qayyum, R., Snively, B. M., Ziv, E., Nalls, M. A., Liu, Y., Tang, W., Yanek, L. R., Lange, L., Evans, M. K., Ganesh, S., et al. (2012). A meta-analysis and genome-wide association study of platelet count and mean platelet volume in African Americans. PLoS Genet *8*, e1002491.

[22] Shameer, K., Denny, J. C., Ding, K., Jouni, H., Crosslin, D. R., de Andrade, M., Chute, C. G., Peissig, P., Pacheco, J. A., Li, R., et al. (2014). A genome- and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. Hum Genet *133*, 95–109.

[23] Arking, D. E., Pulit, S. L., Crotti, L., van der Harst, P., Munroe, P. B., Koopmann, T. T., Sotoodehnia, N., Rossin, E. J., Morley, M., Wang, X., et al. (2014). Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. Nat Genet *46*, 826–836.

[24] Soranzo, N., Spector, T. D., Mangino, M., Kuhnel, B., Rendon, A., Teumer, A., Willenborg, C., Wright, B., Chen, L., Li, M., et al. (2009). A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. Nat Genet *41*, 1182–1190.