ACCEPTED MANUSCRIPT

Online Repository Tables

Genetic Ancestry Influences Asthma Susceptibility and Lung Function Among Latinos

Maria Pino-Yanes, PhD[1,2], Neeta Thakur, MD, MPH[1], Christopher R. Gignoux, PhD[3], Joshua M. Galanter, MD, MAS[1,3], Lindsey A. Roth, MA[1], Celeste Eng, BS[1], Katherine K. Nishimura, MPH[1], Sam S. Oh, PhD[1], Hita Vora, MS[4], Scott Huntsman, MS[1], Elizabeth A. Nguyen, BS[1], Donglei Hu, PhD[1], Katherine A. Drake, PhD[3], David V. Conti, PhD[4], Andres Moreno-Estrada, MD, PhD[5], Karla Sandoval, PhD[5], Cheryl A. Winkler, PhD[6], Luisa N. Borrell, DDS, PhD[7], Fred Lurmann, MS[8], Talat S. Islam, MD[4], Adam Davis, MA, MPH[9], Harold J. Farber, MD, MSPH[10], Kelley Meade, MD[9], Pedro C. Avila, MD[11], Denise Serebrisky, MD[12], Kirsten Bibbins-Domingo, MD, PhD[13], Michael A. Lenoir, MD[14], Jean G. Ford, MD[15], Emerita Brigino-Buenaventura, MD[16], William Rodriguez-Cintron, MD[17], Shannon M. Thyne, MD[18], Saunak Sen, PhD[19], Jose R. Rodriguez-Santana, MD[20], Carlos D. Bustamante, PhD[5], L. Keoki Williams, MD, PhD[21,22], Frank D. Gilliland, MD, PhD[4], W. James Gauderman, PhD[4], Rajesh Kumar, MD, MSPH[23], Dara G. Torgerson, PhD[1], and Esteban G. Burchard, MD, MPH[1,3]

[1] Department of Medicine, University of California, San Francisco, CA, USA
[2] CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain.
[3] Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA
[4] Department of Preventative Medicine, University of Southern California, Los Angeles, CA, USA
[5] Department of Genetics, Stanford University, Palo Alto, CA, USA
[6] Basic Research Laboratory, SAIC-Frederick, Inc., Center for Cancer Research, National Cancer Institute, Frederick, MD, USA
[7] Department of Health Sciences, Graduate Program in Public Health, City University of New York, Bronx, NY, USA
[8] Sonoma Technology, Inc, Petaluma, CA, USA
[9] Children's Hospital and Research Center Oakland, Oakland, CA, USA
[10] Department of Pediatrics, Section of Pulmonology, Baylor College of Medicine and Texas Children's Hospital, Houston, TX, USA
[11] Department of Medicine, Northwestern University, Chicago, IL, USA
[12] Pediatric Pulmonary Division, Jacobi Medical Center, Bronx, NY, USA
[13] Division of General Internal Medicine, University of California, San Francisco, USA
[14] Bay Area Pediatrics, Oakland, CA, USA
[15] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[16] Department of Allergy and Immunology, Kaiser Permanente-Vallejo Medical Center, Vallejo, CA
[17] Veterans Caribbean Health Care System, San Juan, Puerto Rico
[18] Department of Pediatrics, University of California, San Francisco, San Francisco General Hospital, CA, USA
[19] Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA
[20] Centro de Neumología Pediátrica, San Juan, Puerto Rico
[21] Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, Michigan, USA
[22] Department of Internal Medicine, Henry Ford Health System, Detroit, Michigan
[23] Children's Memorial Hospital and the Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

Corresponding Author:
Maria Pino-Yanes, Department of Medicine, University of California, San Francisco, CA, USA. 1550 4th Street, Rock Hall Room 582. Phone: 415-514-9931. Fax: 415-514-4365. Email: mdelpino@ull.edu.es

ACCEPTED MANUSCRIPT

Table E1. Baseline characteristics of GALA II participants included in the current study.

	Puerto Ricans			Mexicans			Other Latinos		
Characteristic	Cases (n=910)	Controls (n=890)	*p*-value	Cases (n=618)	Controls (n=667)	*p*-value	Cases (n=365)	Controls (n=324)	*p*-value
Age (years)	12 [10:14]	13 [11:16]	**$3.2x10^{-13}$**	13 [10:15]	14 [11:17]	**$2.5x10^{-6}$**	13 [10:15]	14 [11:17]	**$1.1x10^{-8}$**
Sex (male)	502 (55%)	410 (46%)	**$1.4x10^{-4}$**	344 (56%)	283 (42%)	**$2.8x10^{-6}$**	200 (55%)	129 (40%)	**$1.2x10^{-4}$**
Family history of allergies or asthma	629 (71%)	301 (36%)	**$<2.2x10^{-16}$**	226 (39%)	115 (19%)	**$1.1x10^{-14}$**	193 (59%)	103 (35%)	**$6.1x10^{-9}$**
Preterm birth	182 (23%)	117 (15%)	**$2.6x10^{-5}$**	57 (11%)	49 (8%)	0.190	50 (16%)	41 (14%)	0.629
BMI percentile	67.4 [44.0:96.0]	63.1 [39.0:92.2]	**0.014**	79.1 [71.0:98.0]	75.4 [59.0:96.0]	**$1.0x10^{-4}$**	79.9 [56.8:98.0]	71.9 [52.0:96.0]	0.075
In utero smoking exposure	66 (7%)	48 (5%)	0.138	25 (4%)	13 (2%)	**0.041**	14 (4%)	12 (4%)	0.916
Early life secondhand smoke	240 (27%)	196 (25%)	0.195	148 (28%)	110 (20%)	**0.001**	79 (24%)	65 (22%)	0.619
Current secondhand smoke	197 (22%)	177 (20%)	0.426	120 (20%)	146 (23%)	0.205	70 (20%)	58 (18%)	0.686
Annual household income									
<$12,000	302 (38%)	248 (41%)	0.599	94 (17%)	117 (21%)	**$9.5x10^{-5}$**	71 (24%)	65 (25%)	0.860
$12,000-25,000	225 (33%)	215 (30%)		173 (31%)	205 (36%)		89 (30%)	83 (32%)	
$25,000-50,000	142 (20%)	134 (19%)		163 (30%)	100 (18%)		71 (24%)	55 (21%)	
>$50,000	73 (9%)	60 (10%)		124 (22%)	140 (25%)		64 (22%)	59 (22%)	
Mother's education level (high school degree)	717 (81%)	714 (83%)	0.313	277 (52%)	253 (47%)	0.152	202 (61%)	177 (63%)	0.836
Insurance									
No insurance	16 (2%)	28 (3%)	0.098	35 (6%)	95 (14%)	**$1.6x10^{-11}$**	19 (5%)	27 (8%)	0.250
Government-subsidized	578 (64%)	536 (61%)		467 (76%)	510 (78%)		280 (78%)	239 (74%)	
Private insurance	313 (34%)	320 (36%)		110 (18%)	51 (8%)		63 (17%)	57 (18%)	
Acculturation by language preference[a]									
Low	725 (80%)	742 (87%)	**$3.7x10^{-4}$**	137 (22%)	203 (30%)	**$5.3x10^{-8}$**	87 (24%)	91 (29%)	**$1.2x10^{-4}$**
Medium	53 (6%)	40 (5%)		348 (57%)	397 (60%)		168 (46%)	168 (55%)	
High	123 (14%)	69 (8%)		125 (21%)	64 (10%)		107 (30%)	50 (16%)	
Discrimination[b]	128 (16%)	100 (13%)	0.194	167 (28%)	168 (27%)	0.687	100 (29%)	87 (29%)	0.900
Daycare attendance in the first 3 years of life	258 (29%)	236 (27%)	0.604	110 (18%)	97 (15%)	0.130	120 (33%)	91 (28%)	0.170
Number of siblings	2 [1:2]	2 [1:2]	0.691	2 [1:3]	2 [1:3]	**0.005**	2 [1:2]	2 [1:2]	0.555
Birth order (firstborn, yes/no)	254 (30%)	256 (32%)	0.438	119 (21%)	103 (17%)	0.073	91 (29%)	80 (28%)	0.839
Dogs and/or cats in the home in the first year of life	425 (47%)	444 (51%)	0.165	152 (25%)	201 (30%)	**0.026**	104 (29%)	92 (29%)	0.976
Nitrogen dioxide in the first year of life (NO_2, ppb)	14.7 [9.0:13.3]	12.7 [9.0:11.4]	0.092	23.0 [18.1:27.3]	24.1 [19.1:28.1]	**0.005**	25.3 [19.2:31.7]	26.4 [20.7:32.0]	0.056

[a] Three-level scale defined by the language preference of the mother and the child, assigning the low category if both child and mother speak Spanish, medium if only one of them speak Spanish and the other speaks English, and high if both child and mother speak English.
[b] Dichotomized variable (yes/no), based on reporting having experienced discrimination due to race, ethnicity, color or language in one of the following settings: at school, receiving medical care, getting service in a store or restaurant, or in public settings.

For continuous variables, the mean and interquartile range are displayed and Mann-Whitney U-test *p*-values are presented for the comparison of cases versus controls.
For categorical variables, the number and proportion of subjects in each category are shown and χ^2-test *p*-values are presented for the comparison of cases versus controls.
p-values≤0.05 are in bold.

ACCEPTED MANUSCRIPT

Table E2. Baseline characteristics participants from CHS and GALA I included in the current study.

	Hispanics CHS		Mexicans GALA I	
	Cases (606)	Controls (792)	Cases (170)	Controls (151)
Age (years)	7 [6:8]	8 [6:10]	13 [10:14]	27 [22:32]
Sex (male)	267 (44%)	382 (48%)	108 (64%)	47 (36%)
Family history of allergies or asthma	284 (47%)	221 (28%)	170 (100%)	NA
Preterm birth	71 (12%)	68 (9%)	6 (4%)	NA
BMI percentile	67 [49:93]	62 [38:89]	84 [80:98]	NA
Pre-FEV_1 (L)	2.1 [1.9:2.3]	2.1 [1.8:2.3]	2.6 [2.0:3.2]	NA
Pre-FVC (L)	2.5 [2.2:2.7]	2.3 [2.0:2.6]	3.2 [2.4:3.9]	NA
Pre-FEF_{25-75} (L/s)	2.5 [2.0:2.8]	2.5 [2.1:2.9]	2.6 [1.7:3.3]	NA
Post-FEV_1 (L)	NA	NA	2.8 [2.1:3.4]	NA
Post-FVC (L)	NA	NA	3.2 [2.4:3.9]	NA
Post-FEF_{25-75} (L/s)	NA	NA	2.6 [1.7:3.3]	NA
In utero smoking exposure	62 (11%)	39 (5%)	6 (4%)	NA
Early life secondhand smoke	57 (9%)	54 (7%)	100 (62%)	NA
Current secondhand smoke	53 (9%)	51 (6%)	103 (61%)	NA
Annual household income				
<$7,500	32 (6%)	61 (10%)	NA	NA
$7,500-14,999	64 (13%)	105 (16%)	NA	NA
$15,000-29,999	117 (23%)	155 (24%)	NA	NA
$30,000-49,999	113 (22%)	136 (21%)	NA	NA
$50,000-99,999	139 (27%)	133 (21%)	NA	NA
≥100,000	45 (9%)	49 (8%)	NA	NA
Mother's education level (high school degree)	410 (68%)	405 (51%)	NA	NA
Age at asthma onset	5 [2:9]	NA	6 [2:8]	NA
Daycare attendance in the first 3 years of life	NA	NA	NA	NA
Number of siblings	NA	NA	NA	NA
Birth order (firstborn, yes/no)	179 (30%)	206 (26%)	NA	NA
Dogs and/or cats in the home in the first year of life	242 (40%)	329 (42%)	NA	NA
Exposure to nitrogen dioxide at study entry (NO_2, ppb)	21.6 [13.0-28.1]	20.4 [12.6-26.1]	NA	NA

For continuous variables, the mean and interquartile range are displayed. For categorical variables, the number and proportion of subjects in each category are shown.

Table E3. Median and interquartile range of proportions of each ancestry for each individual study.

Study	Group	African	European	Native American
GALA II	Puerto Ricans	21.1 [14.0:25.5]	67.6 [62.4:74.7]	11.3 [9.4:12.7]
	Other Latinos	17.0 [6.1:26.9]	49.0 [39.6:59.6]	34.6[9.9:51.4]
	Mexicans	4.0 [2.3:5.0]	37.3 [26.7:48.8]	58.7 [47.0:70.7]
GALA I	Mexicans	5.3 [3.7:6.7]	49.2 [41.4:77.4]	45.5 [36.1:52.4]
CHS	Hispanics	4.8 [1.8:5.8]	55.8 [41.4:72.3]	39.4 [23.7:52.3]

ACCEPTED MANUSCRIPT

Table E4. Results of association testing between genetic ancestry and the odds of asthma in GALA II in the subset of individuals with complete data for all significant covariates (n=2,983).

		Model 1[a]		Model 2[b]	
Ancestry	Group	OR	*p*-value	OR	*p*-value
Native American	Mexicans	0.65 (0.54-0.79)	**$1x10^{-5}$**	0.75 (0.62-0.91)	**0.004**
	Other Latinos	0.72 (0.57-0.91)	**0.007**	0.75 (0.58-0.96)	**0.023**
	Puerto Ricans	1.27 (0.70-2.28)	0.434	1.34 (0.74-2.42)	0.335
African	Mexicans	2.77 (0.89-8.57)	0.078	2.12 (0.69-6.58)	0.192
	Other Latinos	1.63 (1.15-2.33)	**0.007**	1.59 (1.12-2.28)	**0.011**
	Puerto Ricans	1.32 (1.08-1.61)	**0.006**	1.34 (1.10-1.64)	**0.004**
European	Mexicans	1.55 (1.27-1.89)	**$1x10^{-5}$**	1.34 (1.09-1.65)	**0.005**
	Other Latinos	1.12 (0.85-1.48)	0.406	1.06 (0.80-1.41)	0.695
	Puerto Ricans	0.72 (0.59-0.89)	**0.002**	0.70 (0.57-0.86)	**0.001**

[a] Odds of asthma in the model adjusted for age, sex and recruitment site as covariates.
[b] Odds of asthma in the model adjusted for age, sex, recruitment site, SES, and early-life exposures variables as covariates.
All the estimates of odds ratios and 95% confidence interval are referred to each increase of 20% in each ancestry.
p-values≤0.05 are in bold.

ACCEPTED MANUSCRIPT

Table E5. Summary of association testing of genetic ancestry with lung function pre-bronchodilator and post-bronchodilator measurements of FEF_{25-75} in GALA II asthma cases.

		Pre-bronchodilator		Post-bronchodilator	
Ancestry	Group	Effect ± SE[a] (ml/s)	*p*-value[a]	Effect ± SE[a] (ml/s)	*p*-value[a]
Native American	Mexicans	71±40	0.078	40±41	0.326
	Other Latinos	11±49	0.816	-8±49	0.865
	Puerto Ricans	16±118	0.891	1±117	0.993
	Meta-analysis	45±30	0.135	19±31	0.530
African	Mexicans	-121±179	0.501	-198±182	0.276
	Other Latinos	-46±64	0.474	-66±64	0.307
	Puerto Ricans	-87±46	0.060	-69±46	0.132
	Meta-analysis	-75±37	**0.041**	-73±37	**0.045**
European	Mexicans	-72±42	0.091	-33±43	0.440
	Other Latinos	18±52	0.736	53±52	0.309
	Puerto Ricans	89±47	0.062	72±47	0.124
	Meta-analysis	4±27	0.872	25±27	0.352

[a]Adjusted for age, sex, height, height^2, use of controller medication, immunoglobulin E levels, and recruitment region as covariates.
All the estimates are referred to each increase of 20% in ancestry.
All meta-analyses were performed using a fixed effect model as heterogeneity was not detected across samples (*p*-value>0.05). The only exception was European ancestry and $preFEF_{25-25}$ where significant heterogeneity was found (*p*-value=0.040) and a random effect model was applied.
p-values≤0.05 are in bold.

ACCEPTED MANUSCRIPT

Table E6. Results of association testing between genetic ancestry pre-bronchodilator measures of FEV_1 and FVC in children without asthma from GALA II (n=404).

Measurement	Ancestry	Beta ± SE (ml)	*p*-value
Pre-FEV_1	Native American	40±27	0.145
	African	-188±58	**0.001**
	European	0±29	0.994
Pre-FVC	Native American	35±32	0.272
	African	-228±69	**0.001**
	European	15±34	0.664

Adjusted by age, sex, height, height^2, and recruitment region as covariates. Effect estimates are expressed in ml and referred to each increase of 20% in each ancestry compared with the other two ancestries.
p-values≤0.05 are in bold.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table E7. Power calculations for each the sub-analysis of asthma susceptibility.

Study	Group	Power African (%)	Power European (%)	Power Native American (%)
GALA II	Puerto Ricans	79.0	46.4	9.0
	Other Latinos	69.2	47.4	**99.7**
	Mexicans	2.0	**81.0**	**99.6**
GALA I	Mexicans	1.0	8.6	22.9
CHS	Hispanics	23.5	**99.1**	**99.9**
Meta-analysis	All samples	**99.9**	**99.9**	**99.9**

Values of statistical power higher than 80% are presented in bold.

ACCEPTED MANUSCRIPT

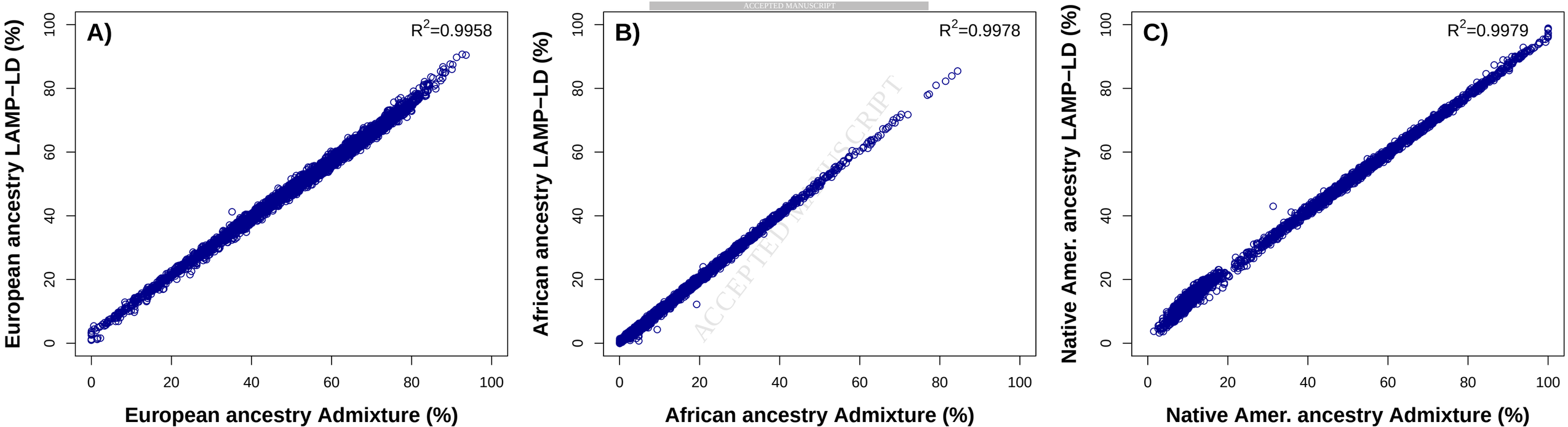
Table E8. Power calculations for each the sub-analysis of lung function among cases (Mexicans [n=602], other Latinos [n=357], and Puerto Ricans [n=903]).

Measurement	Ancestry	Group	Power pre-bronchodilator (%)	Power post-bronchodilator (%)
FEV_1	Native American	Mexicans	16.13	14.20
		Other Latinos	21.11	20.11
		Puerto Ricans	0.76	0.65
		Meta-analysis	**99.88**	**99.75**
	African	Mexicans	1.35	1.22
		Other Latinos	**83.32**	**81.76**
		Puerto Ricans	**98.14**	**95.90**
		Meta-analysis	**99.99**	**99.99**
	European	Mexicans	0.59	0.55
		Other Latinos	0.42	0.41
		Puerto Ricans	0.57	0.49
		Meta-analysis	5.70	5.02
FVC	Native American	Mexicans	8.28	7.90
		Other Latinos	12.78	12.66
		Puerto Ricans	0.49	0.46
		Meta-analysis	**97.62**	**97.20**
	African	Mexicans	0.82	0.80
		Other Latinos	65.13	64.76
		Puerto Ricans	**86.96**	**83.26**
		Meta-analysis	**99.99**	**99.98**
	European	Mexicans	0.40	0.39
		Other Latinos	0.32	0.32
		Puerto Ricans	0.38	0.36
		Meta-analysis	3.09	2.96
FEF_{25-75}	Native American	Mexicans	7.68	6.11
		Other Latinos	8.22	7.20
		Puerto Ricans	0.42	0.37
		Meta-analysis	**95.26**	**99.98**
	African	Mexicans	0.78	0.67
		Other Latinos	48.14	43.32
		Puerto Ricans	78.56	70.68
		Meta-analysis	**97.20**	**91.93**
	European	Mexicans	0.39	0.34
		Other Latinos	0.26	0.25
		Puerto Ricans	0.33	0.30
		Meta-analysis	2.55	2.14

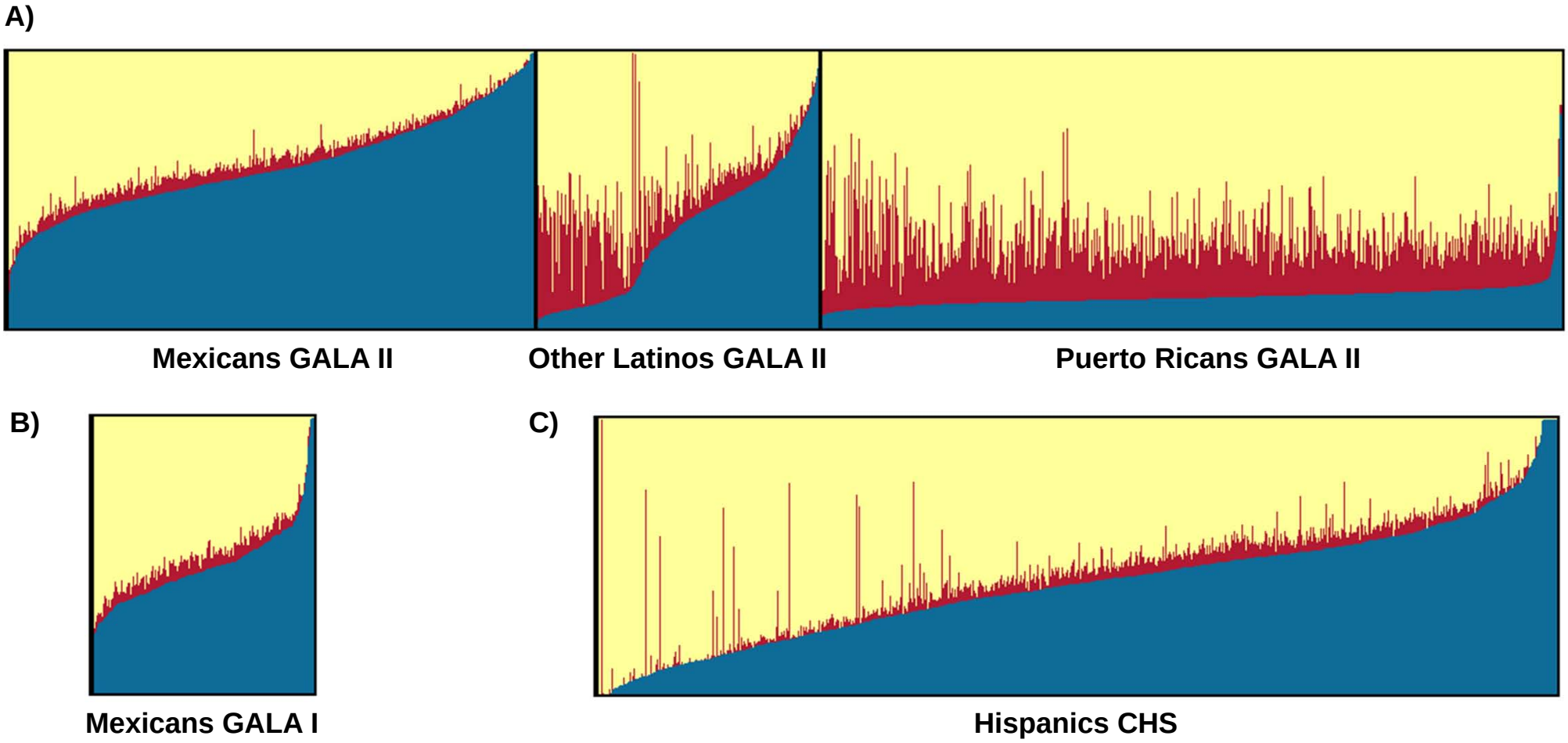
Values of statistical power higher than 80% are presented in bold.

ACCEPTED MANUSCRIPT

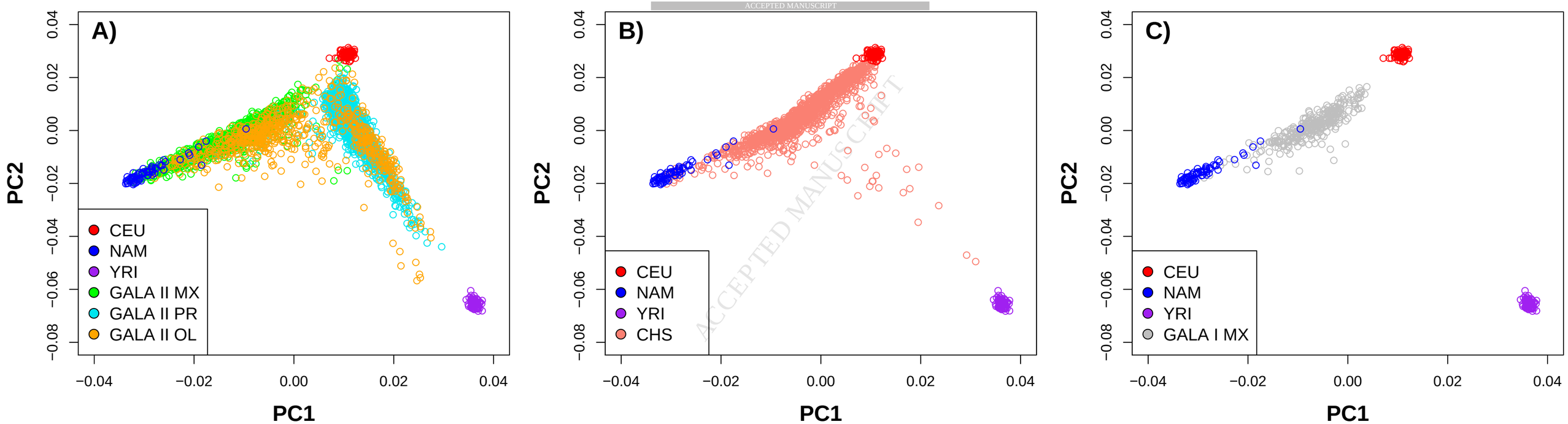
ACCEPTED MANUSCRIPT



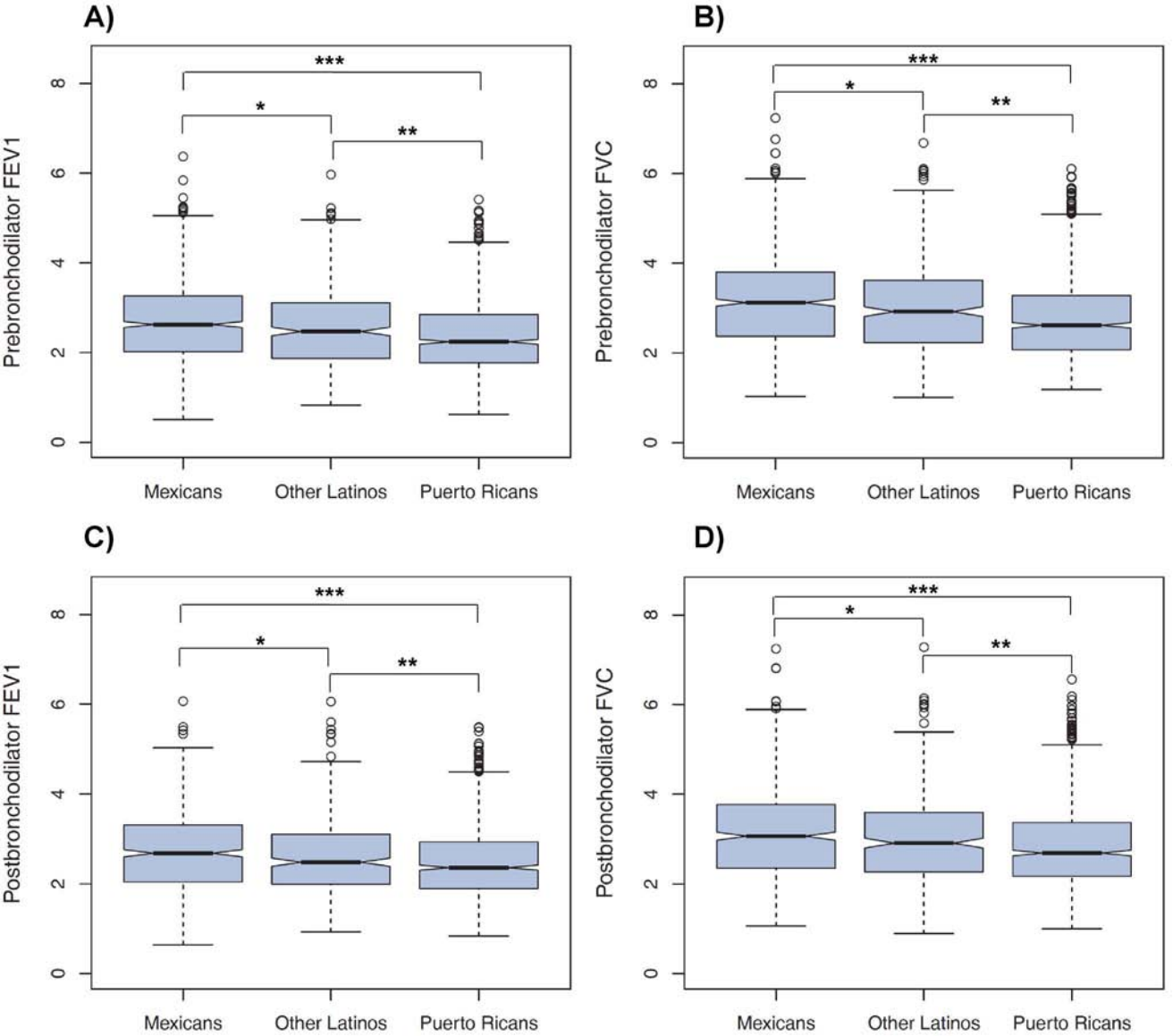
ACCEPTED MANUSCRIPT

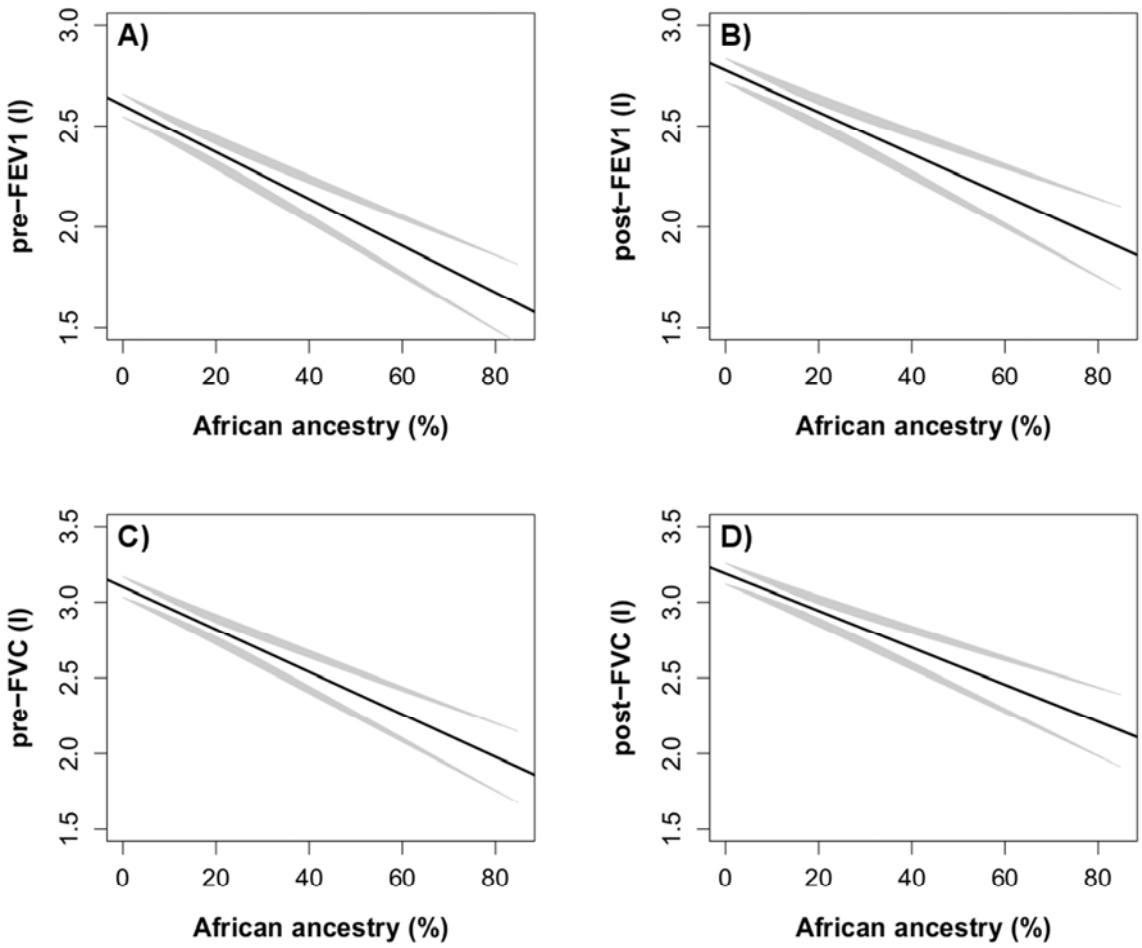

A)
Mexicans GALA II
Other Latinos GALA II
Puerto Ricans GALA II
B)
Mexicans GALA I
C)
Hispanics CHS

ACCEPTED MANUSCRIPT

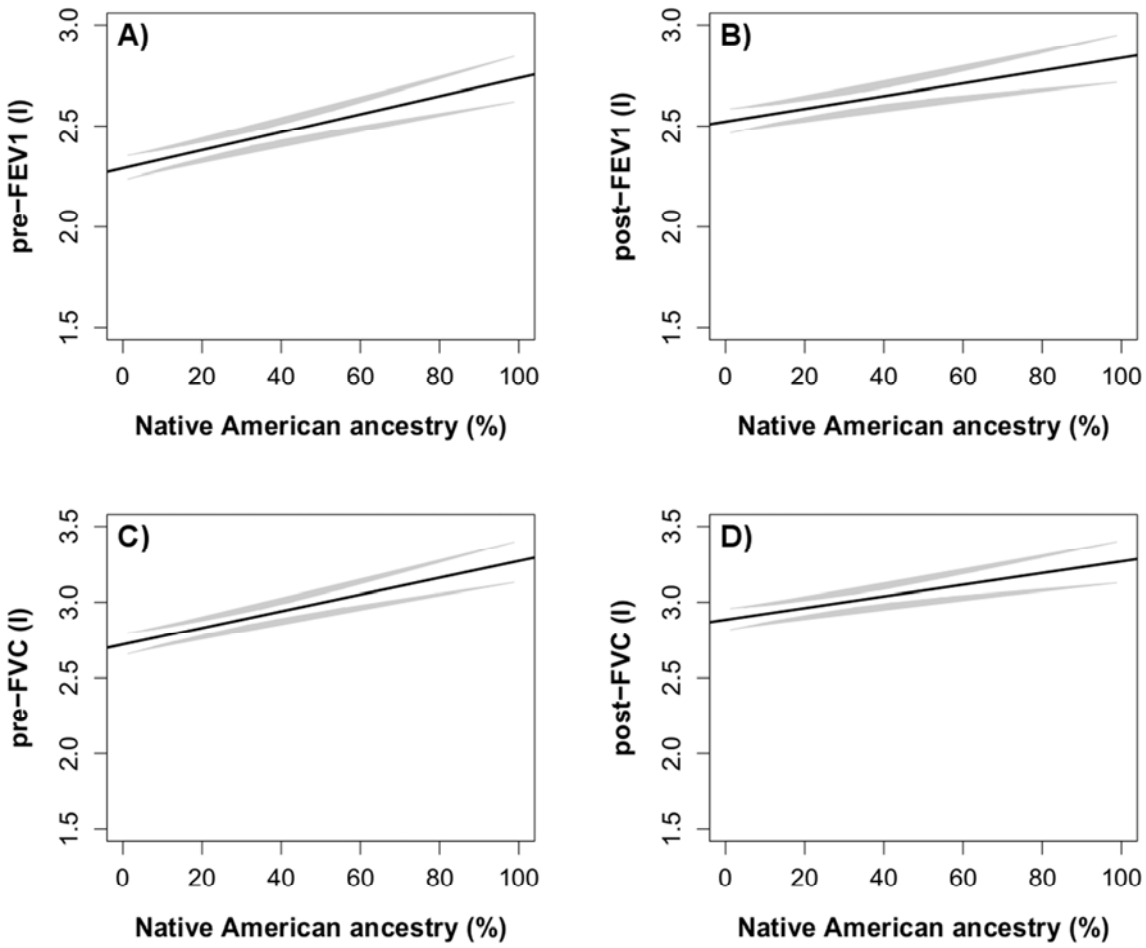


ACCEPTED MANUSCRIPT

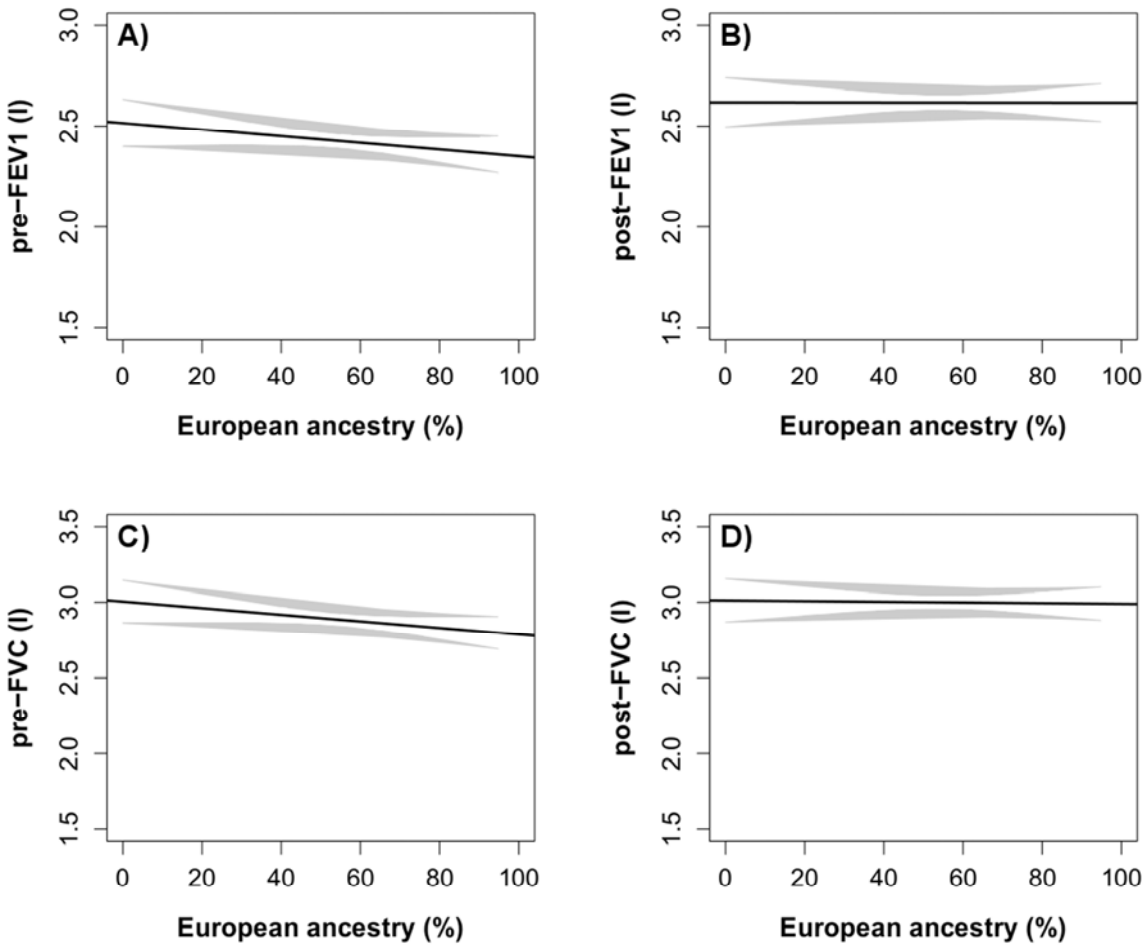

A)
Prebronchodilator FEV1
0
2
4
6
8
***
*
**
Mexicans
Other Latinos
Puerto Ricans
B)
Prebronchodilator FVC
0
2
4
6
8
***
*
**
Mexicans
Other Latinos
Puerto Ricans
C)
Postbronchodilator FEV1
0
2
4
6
8
***
*
**
Mexicans
Other Latinos
Puerto Ricans
D)
Postbronchodilator FVC
0
2
4
6
8
***
*
**
Mexicans
Other Latinos
Puerto Ricans


A)
pre−FEV1 (l)
1.5
2.0
2.5
3.0
0
20
40
60
80
African ancestry (%)
B)
post−FEV1 (l)
1.5
2.0
2.5
3.0
0
20
40
60
80
African ancestry (%)
C)
pre−FVC (l)
1.5
2.0
2.5
3.0
3.5
0
20
40
60
80
African ancestry (%)
D)
post−FVC (l)
1.5
2.0
2.5
3.0
3.5
0
20
40
60
80
African ancestry (%)

ACCEPTED


A)
pre−FEV1 (l)
1.5
2.0
2.5
3.0
0
20
40
60
80
100
Native American ancestry (%)
B)
post−FEV1 (l)
1.5
2.0
2.5
3.0
0
20
40
60
80
100
Native American ancestry (%)
C)
pre−FVC (l)
1.5
2.0
2.5
3.0
3.5
0
20
40
60
80
100
Native American ancestry (%)
D)
post−FVC (l)
1.5
2.0
2.5
3.0
3.5
0
20
40
60
80
100
Native American ancestry (%)

ACCEPTED


A)
pre−FEV1 (l)
1.5
2.0
2.5
3.0
0
20
40
60
80
100
European ancestry (%)
B)
post−FEV1 (l)
1.5
2.0
2.5
3.0
0
20
40
60
80
100
European ancestry (%)
C)
pre−FVC (l)
1.5
2.0
2.5
3.0
3.5
0
20
40
60
80
100
European ancestry (%)
D)
post−FVC (l)
1.5
2.0
2.5
3.0
3.5
0
20
40
60
80
100
European ancestry (%)

ACCEPTED

ACCEPTED MANUSCRIPT

Online Repository Text

Genetic Ancestry Influences Asthma Susceptibility and Lung Function Among Latinos

Maria Pino-Yanes, PhD[1,2], Neeta Thakur, MD, MPH[1], Christopher R. Gignoux, PhD[3], Joshua M. Galanter, MD, MAS[1,3], Lindsey A. Roth, MA[1], Celeste Eng, BS[1], Katherine K. Nishimura, MPH[1], Sam S. Oh, PhD[1], Hita Vora, MS[4], Scott Huntsman, MS[1], Elizabeth A. Nguyen, BS[1], Donglei Hu, PhD[1], Katherine A. Drake, PhD[3], David V. Conti, PhD[4], Andres Moreno-Estrada, MD, PhD[5], Karla Sandoval, PhD[5], Cheryl A. Winkler, PhD[6], Luisa N. Borrell, DDS, PhD[7], Fred Lurmann, MS[8], Talat S. Islam, MD[4], Adam Davis, MA, MPH[9], Harold J. Farber, MD, MSPH[10], Kelley Meade, MD[9], Pedro C. Avila, MD[11], Denise Serebrisky, MD[12], Kirsten Bibbins-Domingo, MD, PhD[13], Michael A. Lenoir, MD[14], Jean G. Ford, MD[15], Emerita Brigino-Buenaventura, MD[16], William Rodriguez-Cintron, MD[17], Shannon M. Thyne, MD[18], Saunak Sen, PhD[19], Jose R. Rodriguez-Santana, MD[20], Carlos D. Bustamante, PhD[5], L. Keoki Williams, MD, PhD[21,22], Frank D. Gilliland, MD, PhD[4], W. James Gauderman, PhD[4], Rajesh Kumar, MD, MSPH[23], Dara G. Torgerson, PhD[1], and Esteban G. Burchard, MD, MPH[1,3]

[1] Department of Medicine, University of California, San Francisco, CA, USA
[2] CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain.
[3] Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA
[4] Department of Preventative Medicine, University of Southern California, Los Angeles, CA, USA
[5] Department of Genetics, Stanford University, Palo Alto, CA, USA
[6] Basic Research Laboratory, SAIC-Frederick, Inc., Center for Cancer Research, National Cancer Institute, Frederick, MD, USA
[7] Department of Health Sciences, Graduate Program in Public Health, City University of New York, Bronx, NY, USA
[8] Sonoma Technology, Inc, Petaluma, CA, USA
[9] Children's Hospital and Research Center Oakland, Oakland, CA, USA
[10] Department of Pediatrics, Section of Pulmonology, Baylor College of Medicine and Texas Children's Hospital, Houston, TX, USA
[11] Department of Medicine, Northwestern University, Chicago, IL, USA
[12] Pediatric Pulmonary Division, Jacobi Medical Center, Bronx, NY, USA
[13] Division of General Internal Medicine, University of California, San Francisco, USA
[14] Bay Area Pediatrics, Oakland, CA, USA
[15] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[16] Department of Allergy and Immunology, Kaiser Permanente-Vallejo Medical Center, Vallejo, CA
[17] Veterans Caribbean Health Care System, San Juan, Puerto Rico
[18] Department of Pediatrics, University of California, San Francisco, San Francisco General Hospital, CA, USA
[19] Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA
[20] Centro de Neumología Pediátrica, San Juan, Puerto Rico
[21] Center for Health Policy and Health Services Research, Henry Ford Health System, Detroit, Michigan, USA
[22] Department of Internal Medicine, Henry Ford Health System, Detroit, Michigan
[23] Children's Memorial Hospital and the Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

Corresponding Author:
Maria Pino-Yanes, Department of Medicine, University of California, San Francisco, CA, USA. 1550 4th Street, Rock Hall Room 582. Phone: 415-514-9931. Fax: 415-514-4365. Email: mdelpino@ull.edu.es

Text E1. Supplementary Methods

Subjects from the Genes-environments & Admixture in Latino Americans (GALA II) study

The genes-environments & Admixture in Latino Americans (GALA II) study is an ongoing multicenter case-control study of asthma in Latinos, organized from the coordinating center based at the University of California, San Francisco.[E1-E3] Cases and healthy controls of Latino ancestry were recruited from 5 urban study centers throughout the U.S. (Chicago, Illinois; Bronx, New York; Houston, Texas; San Francisco Bay Area, California; and San Juan, Puerto Rico). Subjects were eligible if they were 8-21 years of age, self-identified all four grandparents as Latino, and had <10 pack-years of smoking history. Subjects were excluded if they reported any of the following: (1) 10 or more pack-years of smoking; (2) any smoking within 1 year of recruitment date; (3) pregnancy in the third trimester; or (4) history of one of the following conditions: sickle cell disease, cystic fibrosis, sarcoidosis, cerebral palsy, or history of heart or chest surgery. Asthma was defined by physician diagnosis and report of symptoms of coughing, wheezing, or shortness of breath in the 2 years preceding enrollment. Controls had no reported history of asthma, eczema, hives, hay fever, allergic rhinitis, not reported use of medication for allergies and not reported symptoms of wheezing or shortness of breath during their lifetime. All local institutional review boards approved the study and all subjects/parents provided written assent/consent, respectively. Detailed clinical measurements were recorded for each individual and whole blood for DNA, RNA, and plasma extraction was collected. In addition, trained interviewers, proficient in both English and Spanish, administered questionnaires to gather baseline demographic data, as well as information on general health, asthma status, acculturation, social, and environmental exposures. A total of 4,048 subjects (1998 cases, 2050 controls) had been recruited at the time of genotyping.

Subjects were classified into three Latino ethnicity categories according to the self-reported ethnicity of their four grandparents: Puerto Rican, Mexican and other Latino. The category designated as "other Latino" included individuals with mixed Latino ancestry or an ethnicity different from Puerto Rican or Mexican (Cuban, Dominican, El Salvadorian, Guatemalan, Nicaraguan, Honduran, Colombian, Brazilian, or Argentinian).

Pulmonary function testing was conducted with a (KoKo® PFT Spirometer, nSpire Health Inc., Louisville, CO) according to American Thoracic Society recommendations.[E4] Subjects with asthma were instructed to withhold their bronchodilator medications for at least 8 hours testing. Standard measurements of airway obstruction included forced expiratory volume in one second (FEV_1), forced vital capacity (FVC), and forced expiratory flow between 25-75% of vital capacity (FEF_{25-75}). After completing baseline spirometry, subjects were given 4 puffs of albuterol administered through a metered-dose inhaler with a spacer, and spirometry was repeated after 15 minutes to obtain post-bronchodilator measurements.

Genotyping and quality control measures for GALA II samples

GALA II subjects were genotyped with the Axiom LAT1 array (World Array 4, Affymetrix, Santa Clara, CA),[E5] which includes 817,810 single nucleotide polymorphisms (SNPs) and was specifically designed to capture genetic variation in Latino populations.[E6]

Genotyping was blind to the case-control status and to monitor genotyping quality, one sample of known genotype (Affymetrix's sample #103) was included on each plate. Genotype calls were obtained simultaneously for all cases and controls using the Axiom GT1 algorithm on the Affymetrix Genotyping Console™ Software (GTC) following manufacturer's

recommendations. Samples with fewer than 82% of the dish quality control (DQC) were excluded. Then, another round of genotype calling was performed and samples were excluded if the genotyping rate was below 97%. We then excluded samples with discrepancy between genetic sex and reported gender, and with cryptic relatedness (PI_HAT>0.3). SNPs that did not pass quality control procedures were removed from the analyses using the following criteria: failing Axiom quality controls, <95% call rate, or deviation from Hardy-Weinberg equilibrium ($p<10^{-6}$) within each defined population (Puerto Rican, Mexican, and other Latino). Therefore, genotypes for 723,888 high-quality autosomal SNPs were available for the analyses. About 7% of subjects were filtered out, resulting in a total of 3,774 samples with high quality genotype data of self-reported Puerto Ricans (n=1,800), Mexicans (n=1,285) and other Latinos (n=689) for this analysis.[E7] All the quality control measurements where performed using the Affymetrix Genotyping Console™ Software, PLINK ,[E8] and R.[E9]

Genotyping and quality control procedures performed in GALA I and CHS have been described elsewhere. [E10,E11] Briefly, GALA I was genotyped with Affymetrix 6.0 GeneChip Array (Affymetrix, Santa Clara, CA) and markers were filtered based on call rates <95%, Hardy-Weinberg equilibrium deviation in controls ($p<10^{-6}$), and unambiguous mapping to the human reference genome, resulting in a set of 729,685 SNPs. Subjects from GALA I were filtered based on 95% call rates and consistency between genetic and reported sex. [E11] Study samples from the CHS study were genotyped using HumanHap550, HumanHap550-Duo or Human610-Quad BeadChip microarrays (Illumina, San Diego, CA). Individuals were excluded from analysis with call rates <90% and SNPs were filtered if they had a call rate < 95% or departed from Hardy-Weinberg equilibrium ($p<10^{-6}$) in controls or had poor concordance with genotypes a previous

the candidate gene study. The number of SNPs that passed quality control in CHS was 562,843.[E11]

Assessment of genetic ancestry

Estimates of African, European and Native American ancestries were obtained using an unsupervised analysis in ADMIXTURE [E12] assuming 3 ancestral populations: African, European and Native American. We used reference haplotypes from individuals from HapMap phase II (http://hapmap.ncbi.nlm.nih.gov) for the European and African components: Utah residents with Northern and Western European ancestry from the CEPH collection (CEU) and Yoruba in Ibadan, Nigeria (YRI). Genotypes for CEU and YRI were obtained from the Axiom Genotype Data Set (http://www.affymetrix.com/support/technical/sample_data/axiom_db/axiomdb_data.affx) and SNPs were filtered by missing rate (SNPs with <95% call rate were removed). Since there are no Native American reference samples for Native American ancestry in the HapMap database, we genotyped 95 Native American individuals (NAM) on the Axiom LAT1 array.[E2] After quality control measurements, the final Native American reference set included 71 Native American unrelated individuals, as 24 samples did not pass our quality control measurements (20 of them showed high cryptic relatedness [PI_HAT>0.3] and 4 of them were filtered because of their low call genotyping rates [<97%]). These 71 individuals included: 14 Zapotec, 2 Mixe, and 11 Mixtec from the southern State of Oaxaca (which were a subset of the samples described in Price et al. 2007) [E13] and 44 Nahua individuals from Central Mexico.[E2]

Ancestry estimates were obtained from common set of makers across GALA II, GALA I, CHS, and the reference populations (CEU, YRI, and NAM). For that, we first removed C/G and

A/T SNPs to avoid potential artifacts from mislabeling the reference allele across population and then we merged the SNPs passing quality control measurements in each study and the reference populations, resulting in 34,616 markers. After that, linkage disequilibrium pruning was performed to ensure that markers used in ancestry estimation had low pairwise linkage disequilibrium ($R^2 \leq 0.15$), obtaining a final set of 20,669 markers.

Admixture's maximum likelihood estimates where then obtained using an unsupervised analysis in ADMIXTURE to avoid introducing bias in the estimations.[E12] This software is being shown to provide ancestry estimates as accurate as other softwares that use Bayesian estimates[E14] or principal component analysis[E15,E16] In addition, we confirmed that ancestry estimates were accurate using LAMP-LD,[E17] a method to infer local ancestry leveraging the structure of linkage disequilibrium. Mean values of genetic ancestry obtained with LAMP-LD were highly correlated with ancestry estimates provided by ADMIXTURE ($R^2 > 0.995$, Figure E1).We finally performed principal component analysis to compare the genetic ancestry components among the three studies: GALA II, GALA I, and CHS (Figure E2) using EIGENSOFT. [E15,E16]

Statistical analysis of association between genetic ancestry and asthma susceptibility in GALA II

We tested for an association between African, European and Native American ancestry and asthma susceptibility in GALA II using logistic regression and stratifying by the three Latino ethnicity categories (Mexican, other Latino and Puerto Rican). These analyses included recruitment region (Puerto Rico, New York, Chicago, Texas, and San Francisco Bay Area), age and sex as covariates. Other potential confounders were also considered if they were associated with asthma in univariate models for each ethnic group considered by separate ($p \leq 0.05$, Table

E1) to be included in in secondary multivariate models, but were only kept in the final model if they showed an association with asthma at $p \leq 0.05$. Although family history of allergies or asthma was significantly associated with asthma status, we did not include this variable because it is highly correlated with genetic ancestry due to their transmission from parents to children. The other variables evaluated included: body mass index (BMI) percentiles, preterm birth (born before 37 weeks of pregnancy), number of siblings, daycare attendance in the first 3 years of life, birth order (firstborn, yes/no), having dogs or cats in the home in the first year of life, exposure to second hand smoke exposure, exposure to air pollutants, and variables related to socioeconomic status (SES), health care access, discrimination and acculturation.

BMI was calculated for each participant using weight and height measures. Given that BMI varies across age and sex among children, [E18,E19] standardized sex- and age-specific growth charts were used to calculate BMI percentiles (http://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm). Smoke exposure was evaluated based on 3 different variables: maternal smoking during pregnancy, exposure to secondhand smoke in the first 2 years of life from household smokers, and exposure to current adult household smokers.[E20] Exposure to air pollution was evaluated by assigning geographic coordinates for all self-reported residential histories since birth using the TomTom/Tele Atlas EZ-Locate software. Regional ambient air pollution data were acquired from the U.S. Environmental Protection Agency's Air Quality System. Exposures to O_3, NO_2, SO_2, PM_{10}, and $PM_{2.5}$ were calculated for the first year of life. Pollution exposures were estimated by calculating the inverse distance-squared weighted average from the four closest air pollution monitoring stations within 50 km of a residence.[E3]

Variables relating to SES and access to health care included annual household income, mother's highest education level, and child's health insurance status. For annual household income, we defined four categories based on the quartiles of the income distribution (<$12,000, $12,000-25,000, $25,000-50,000, >$50,000). Mother's highest education level was dichotomized based on completion of high school, general education development, or equivalent. Health insurance status was assigned to three categories including no insurance, government-subsidized insurance, or private insurance. Discrimination was evaluated as a dichotomized variable (yes/no), based on reporting having experienced discrimination due to race, ethnicity, color or language in one of the following settings: at school, receiving medical care, getting service in a store or restaurant, or in public settings. Acculturation was considered as a three-level scale defined by the language preference of the mother and the child, assigning the low category if both child and mother speak Spanish, medium if only one of them speak Spanish and the other speaks English and high if both child and mother speak English.

The final adjusted model for Mexicans included the following covariates: recruitment site, age, sex, exposure to secondhand smoke in the first 2 years of life from household smokers, health insurance status, and annual household income. For other Latinos the adjusted model included: recruitment site, age, sex, self-reported ethnicity subgroup, and acculturation. For Puerto Ricans the adjusted models included as covariates: recruitment site, age, sex, and preterm birth.

To account for the multiple comparisons performed we adopted a Bonferroni correction. For asthma we analyzed 3 ancestries and 3 Latino subgroups from GALA II and two from the replication samples, making a total of 15 tests (Bonferroni correction: 0.05/9, therefore corrected $\alpha=3.3\text{x}10^{-3}$). For lung function we tested 54 different tests: 3 ancestries in 3 Latino subgroups for

6 measurements of lung function, consequently the corrected threshold of significance was $\alpha = 9.3 \times 10^{-4}$ (0.05/54).

Post-hoc power calculations were performed with the software Quanto (http://biostats.usc.edu/Quanto.html) based on the Bonferroni corrected thresholds and the effect sizes observed in our meta-analyses for asthma and lung function.

ACCEPTED MANUSCRIPT

References Online Repository Material

E1. Borrell LN, Nguyen EA, Roth LA, Oh SS, Tcheurekdjian H, Sen S, et al. Childhood obesity and asthma control in the GALA II and SAGE II studies. Am J Respir Crit Care Med 2013; 187:697-702.

E2. Kumar R, Nguyen EA, Roth LA, Oh SS, Gignoux CR, Huntsman S, et al. Factors associated with degree of atopy in Latino children in a nationwide pediatric sample: the Genes-environments and Admixture in Latino Asthmatics (GALA II) study. J Allergy Clin Immunol 2013; 132:896-905 e1.

E3. Nishimura KK, Galanter JM, Roth LA, Oh SS, Thakur N, Nguyen EA, et al. Early-Life Air Pollution and Asthma Risk in Minority Children. The GALA II and SAGE II Studies. Am J Respir Crit Care Med 2013; 188:309-18.

E4. American Thoracic Society. Standardization of Spirometry, 1994 Update. American Thoracic Society. Am J Respir Crit Care Med 1995; 152:1107-36.

E5. Drake KA, Torgerson DG, Gignoux CR, Galanter JM, Roth LA, Huntsman S, et al. A genome-wide association study of bronchodilator response in Latinos implicates rare variants. J Allergy Clin Immunol 2014; 133:370-8.

E6. Hoffmann TJ, Zhan Y, Kvale MN, Hesselson SE, Gollub J, Iribarren C, et al. Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. Genomics 2011; 98:422-30.

E7. Galanter JM, Gignoux CR, Torgerson DG, Roth LA, Eng C, Oh SH, et al. Genome-wide association study and admixture mapping identify different asthma-associated loci in Latinos: The Genes-environments & Admixture in Latino Americans study. J Allergy Clin Immunol 2014, in press.

E8. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559-75.

E9. R Development Core Team. R: A Language and Environment for Statistical Computing. Viena, Austria; 2008.

E10. Torgerson DG, Gignoux CR, Galanter JM, Drake KA, Roth LA, Eng C, et al. Case-control admixture mapping in Latino populations enriches for known asthma-associated genes. J Allergy Clin Immunol 2012; 130:76-82 e12.

E11. Torgerson DG, Ampleford EJ, Chiu GY, Gauderman WJ, Gignoux CR, Graves PE, et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. Nat Genet 2011; 43:887-92.

E12. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009; 19:1655-64.

E13. Price AL, Patterson N, Yu F, Cox DR, Waliszewska A, McDonald GJ, et al. A genomewide admixture map for Latino populations. Am J Hum Genet 2007; 80:1024-36.

E14. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000; 155:945-59.

E15. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006; 2:e190.

E16. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006; 38:904-9.

E17. Baran Y, Pasaniuc B, Sankararaman S, Torgerson DG, Gignoux C, Eng C, et al. Fast and accurate inference of local ancestry in Latino populations. Bioinformatics 2012; 28:1359-67.

E18. Hammer LD, Kraemer HC, Wilson DM, Ritter PL, Dornbusch SM. Standardized percentile curves of body-mass index for children and adolescents. Am J Dis Child 1991; 145:259-63.

E19. Pietrobelli A, Faith MS, Allison DB, Gallagher D, Chiumello G, Heymsfield SB. Body mass index as a measure of adiposity among children and adolescents: a validation study. J Pediatr 1998; 132:204-10.

E20. Oh SS, Tcheurekdjian H, Roth LA, Nguyen EA, Sen S, Galanter JM, et al. Effect of secondhand smoke on asthma control among black and Latino children. J Allergy Clin Immunol 2012; 129:1478-83 e7.

ACCEPTED MANUSCRIPT

Supplementary figure legends

Figure E1. Correlation between genetic ancestry estimates obtained by using ADMIXTURE versus LAMP-LD. A) European ancestry, B) African ancestry, C) Native American ancestry.

Figure E2. Admixture proportions obtained with ADMIXTURE for the individuals included from each study: A) GALA II, B) GALA I, C) CHS. Each vertical bar represents one participant and colors display the proportions of Native American (blue), African (red), and European (yellow) ancestry.

Figure E3. Plot displaying the results from the principal component analysis showing the reference populations for Europeans (CEU), Africans (YRI) and Native Americans (NAM) and the participants of the studies A) GALA II, B) CHS, C) Mexicans from GALA I. MX: Mexicans; OL: other Latinos; PR: Puerto Ricans.

Figure E4. Boxplots showing the comparison of different lung function measures among Mexicans, other Latinos and Puerto Ricans children with asthma from GALA II. A) Pre-bronchodilator FEV_1, B) Pre-bronchodilator FVC, C) Post-bronchodilator FEV_1, D) Post-bronchodilator FVC. *, $p<0.05$; **, $p<0.01$; ***, $p<1x10^{-7}$.

Figure E5. Relationship between African ancestry and A) Pre-bronchodilator FEV_1, B) Post-bronchodilator FEV_1, C) Pre-bronchodilator FVC, and D) Post-bronchodilator FVC, for all

Latino children with asthma from GALA II. The black line indicates the predicted values of lung function based on the proportions of African ancestry, with 95% confidence intervals represented by the grey lines.

Figure E6. Relationship between Native American ancestry and A) Pre-bronchodilator FEV_1, B) Post-bronchodilator FEV_1, C) Pre-bronchodilator FVC, and D) Post-bronchodilator FVC, for all Latino children with asthma from GALA II. The black line indicates the predicted values of lung function based on the proportions of African ancestry, with 95% confidence intervals represented by the grey lines.

Figure E7. Relationship between European ancestry and A) Pre-bronchodilator FEV_1, B) Post-bronchodilator FEV_1, C) Pre-bronchodilator FVC, and D) Post-bronchodilator FVC, for all Latino children with asthma from GALA II. The black line indicates the predicted values of lung function based on the proportions of African ancestry, with 95% confidence intervals represented by the grey lines.

ACCEPTED MANUSCRIPT