

Hispanic Community Health Study / Study of Latinos (HCHS - SOL)



National Heart, Lung, and Blood Institute
National Institute of Diabetes and Digestive and Kidney Diseases
National Institute of Neurological Disorders and Stroke
National Institute of Deafness and Other Communication Disorders
National Institute of Dental and Craniofacial Research
National Institute for Minority Health and Health Disparities
Office of Dietary Supplements

Challenges to Health Disparities Research

- Barriers
 - Legal Status
 - Past Traumas
 - Discrimination
 - Marginalization
 - Distrust of Government
 - Cultural or Linguistic Differences



Hispanic/Latino Defined

A



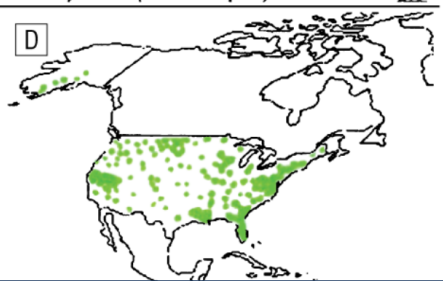
B



C



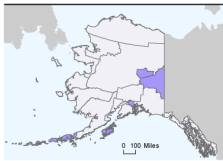
D



A- Latino refers to persons whose origin or ancestries are from countries of Latin America.

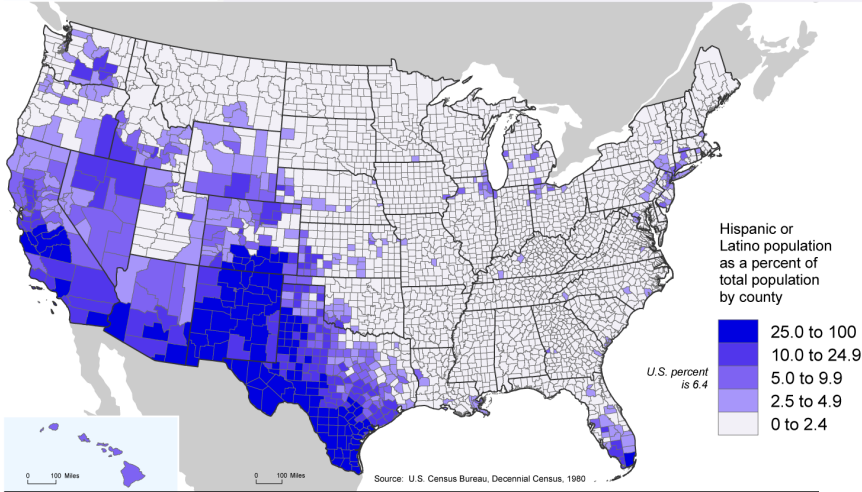
B-The US Office of Management and Budget uses the terms Hispanic and Latino interchangeably to refer to persons who indicated that their origin is Mexican, Puerto Rican, Cuban, Central and South American, or other Spanish culture or Spanish-speaking country or origin, regardless of race.

**C-Hispanic include individuals whose origin or ancestry comes from Hispania, or
D-Spanish-speaking persons of Latin American descent living in the United States.**

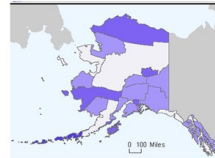


Percent of Population 1980
Hispanic or Latino

6% of US

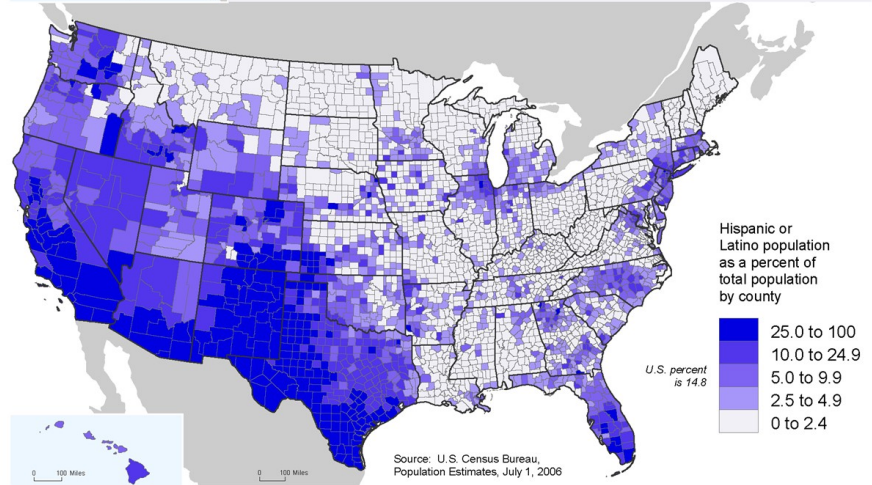


Growth and diversification of US Latino population

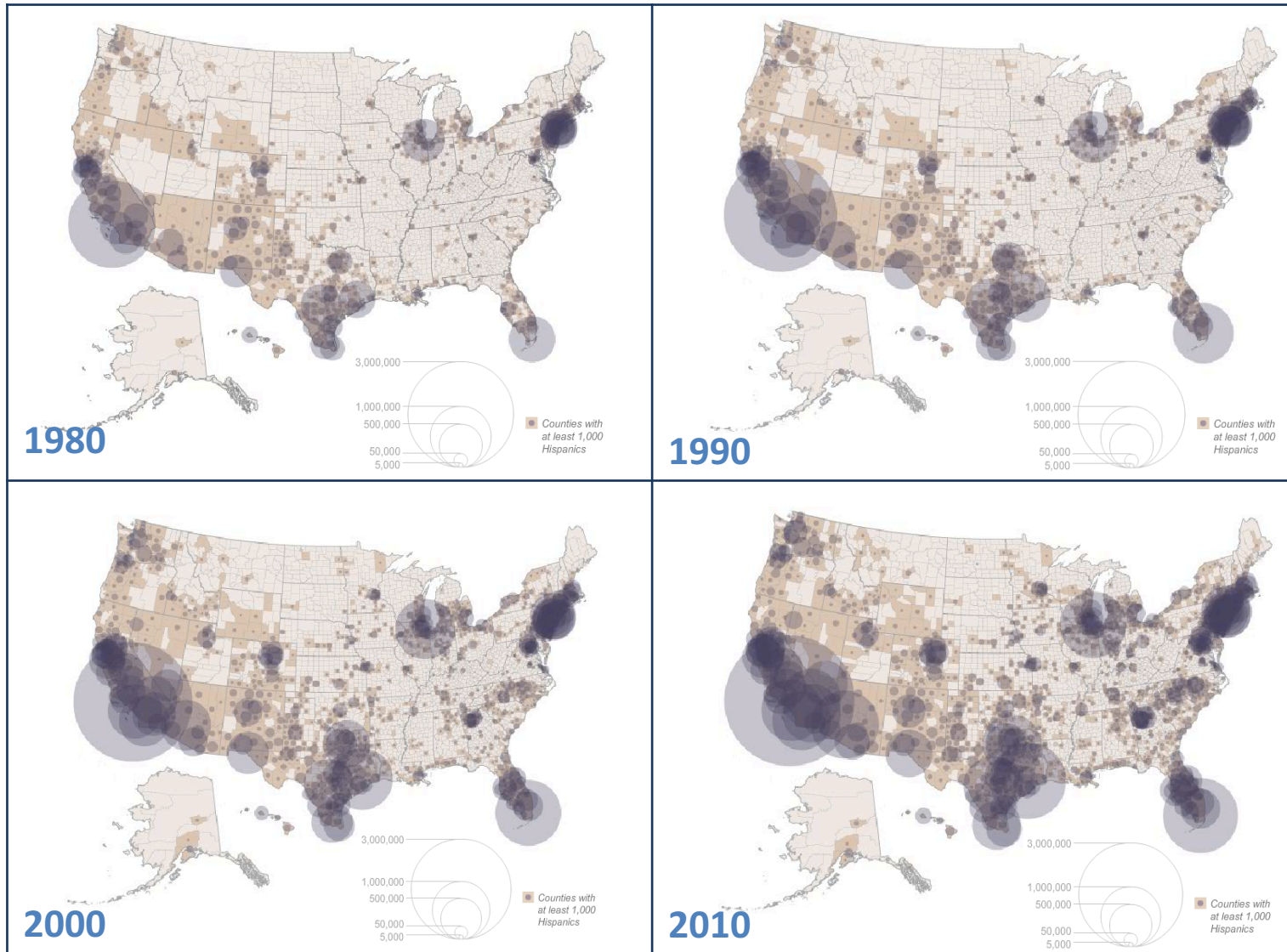


Percent of Population 2006
Hispanic or Latino

15% of US



US Hispanic/Latinos



HCHS - SOL cohort overview

- Multicenter prospective cohort recruited in 2008-2011
- N=16,415 participants aged 18 to 74 years old
 - 1/3 of participants were 18-45 years at baseline (undersampled)
 - 2/3 of participants were > 45 yrs at baseline (oversampled)
 - Area based household sample in four US cities
- TOPMed contact PIs: Robert Kaplan and Kari North
- Genetic SIG chairs: K. North, E. Boerwinkle, T. Sofer
- Genetic Analysis Center: Bruce Weir and Cathy Laurie
- Other site PIs: M. Daviglus (Chicago), N. Schneiderman (Miami), G. Talavera (San Diego), J. Cai (CC)
- Spirometry reading center PIs: Graham Barr, Paul Enright, John Hankinson

Study Timeline



**Recruitment
N=16,415, ages
18-74**

**Baseline clinic
visit 2008-2011**

**Annual follow-up interviews to determine outcomes and
changes in key exposures 2009 -**

**Second clinic visit
2014-2017**

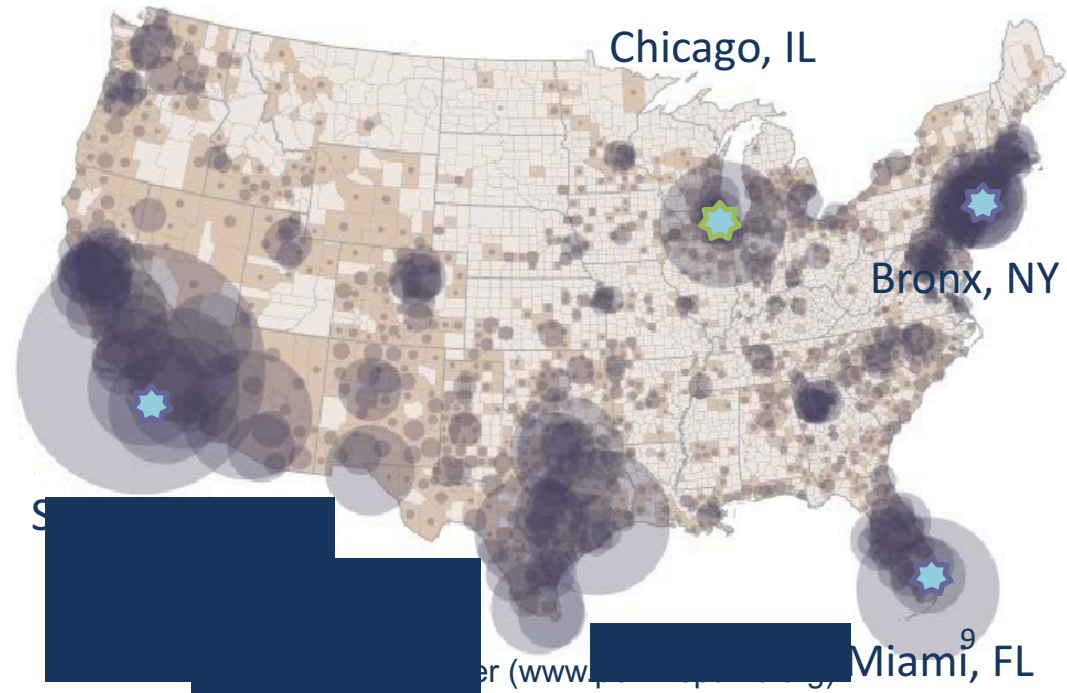
2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019

Disease and biometric phenotypes

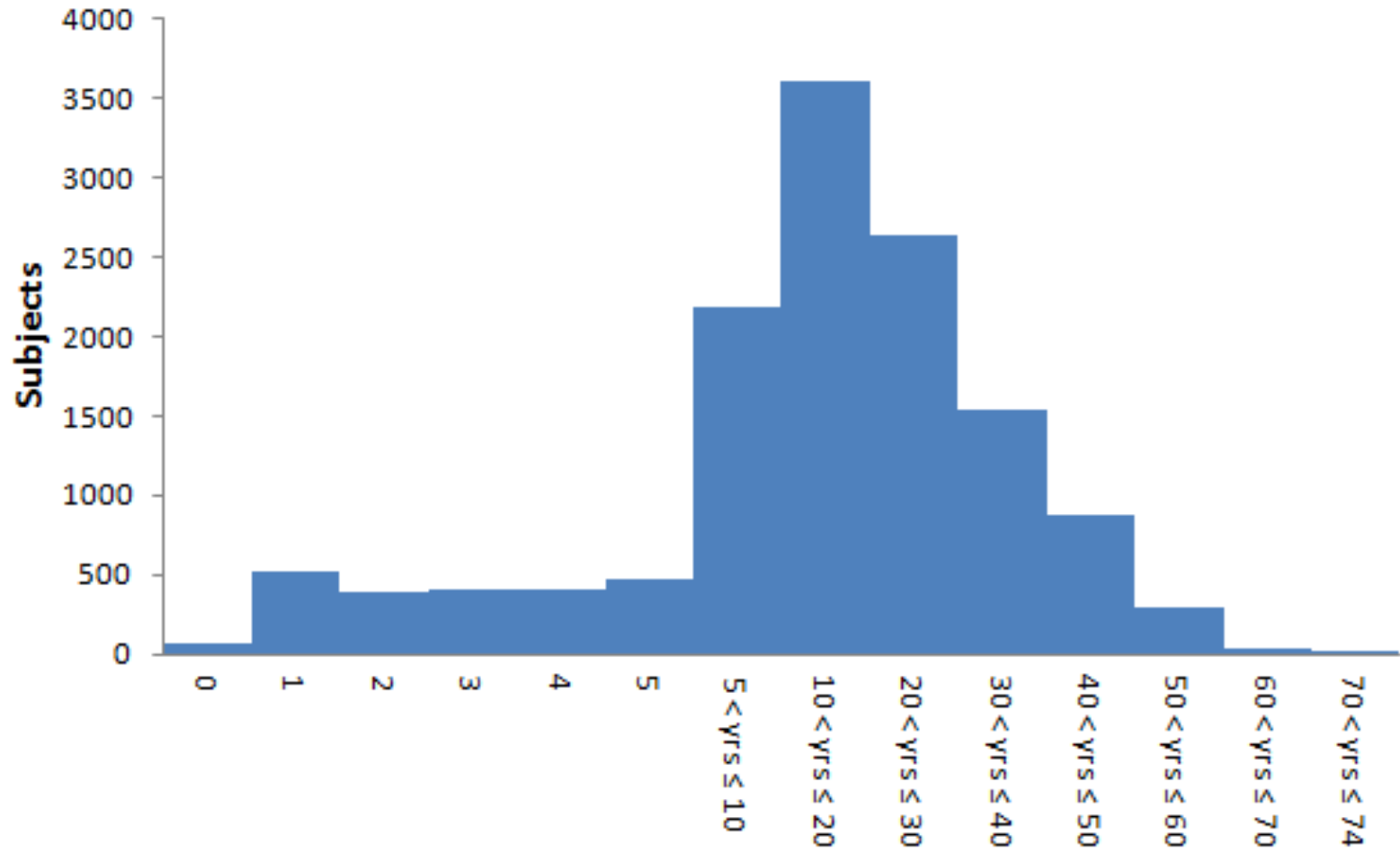
- **Prevalent and incident pulmonary diagnoses + exacerbations**
- **Pulmonary Function Testing (basal and post-bronchodilator)**
- Mortality (total and cause-specific)
- Prevalent and incident CVD (MI, stroke, heart failure)
- Anthropometry/Weight loss/gain
- Lab values: lipids, glucose, OGTT, insulin, inflammation, CBC, Hepatitis A, B C, Total and HDL cholesterol, LFT, renal, etc
- Oral/dental health
- Pregnancy complications
- Hearing
- Blood pressure, ECG, echocardiography
- Sleep quality and disorders
- Cognitive Function
- Substance Abuse
- Diabetes
- Health behaviors (smoking, diet, physical activity, etc)

Study Population

- Hispanic Community Health Study/Study of Latinos (HCHS/SOL)
 - Community-based study of 16,415 men and women, 18-76 years of age at baseline examination (2008-2011)
 - Complex sampling design
 - Self-identified as
 - Central American
 - Cuban
 - Dominican
 - Mexican
 - Puerto Rican
 - South American
 - Other/Multiple



4 out of 5 cohort members were born outside of the 50 states



Length of time living in the 50 states
on date of enrollment

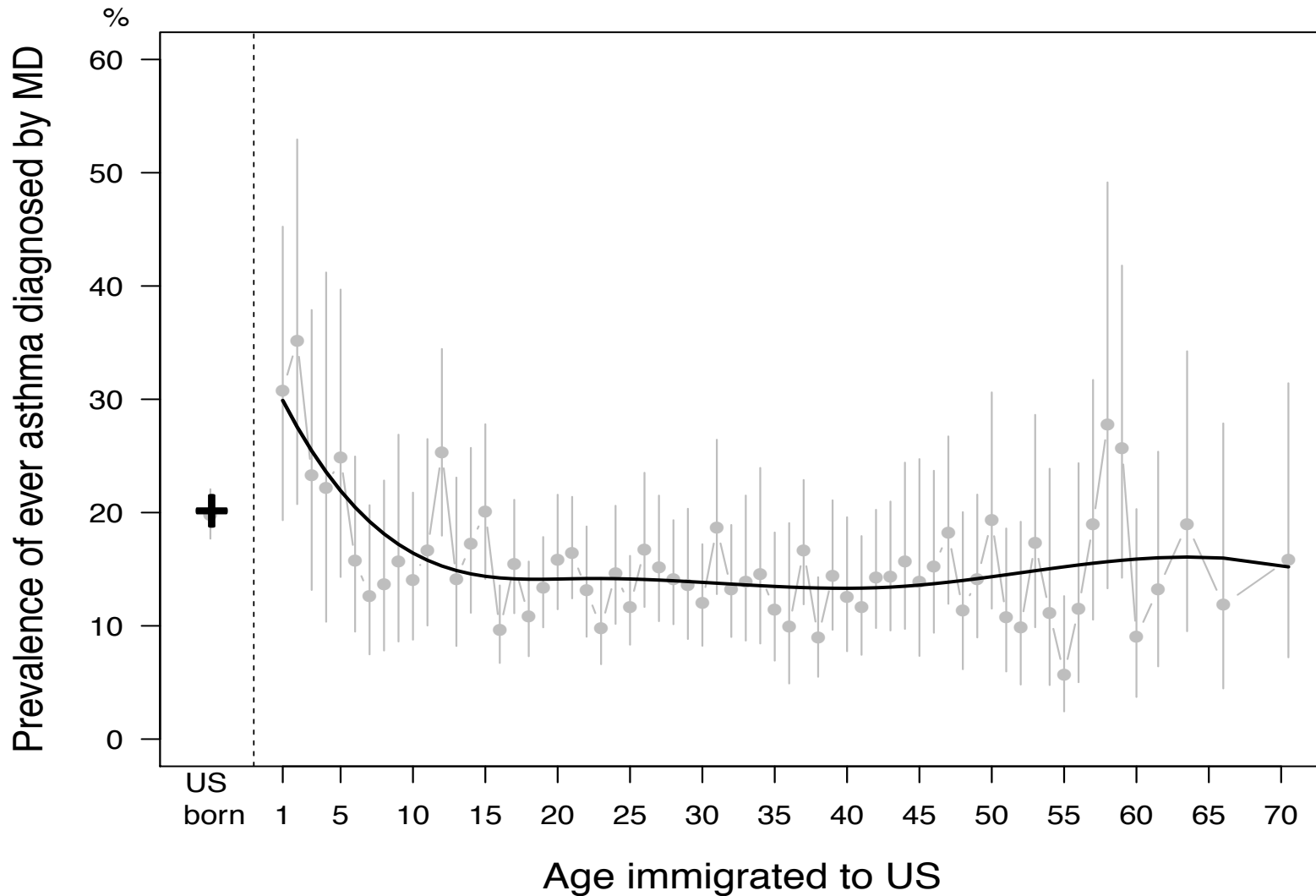
Demographic Characteristics

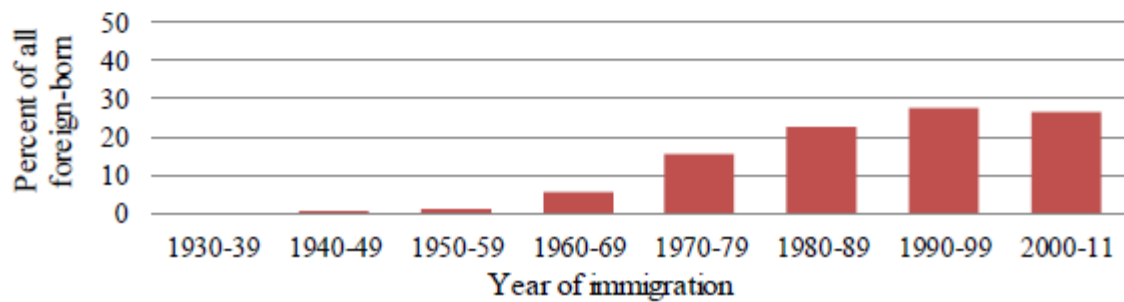
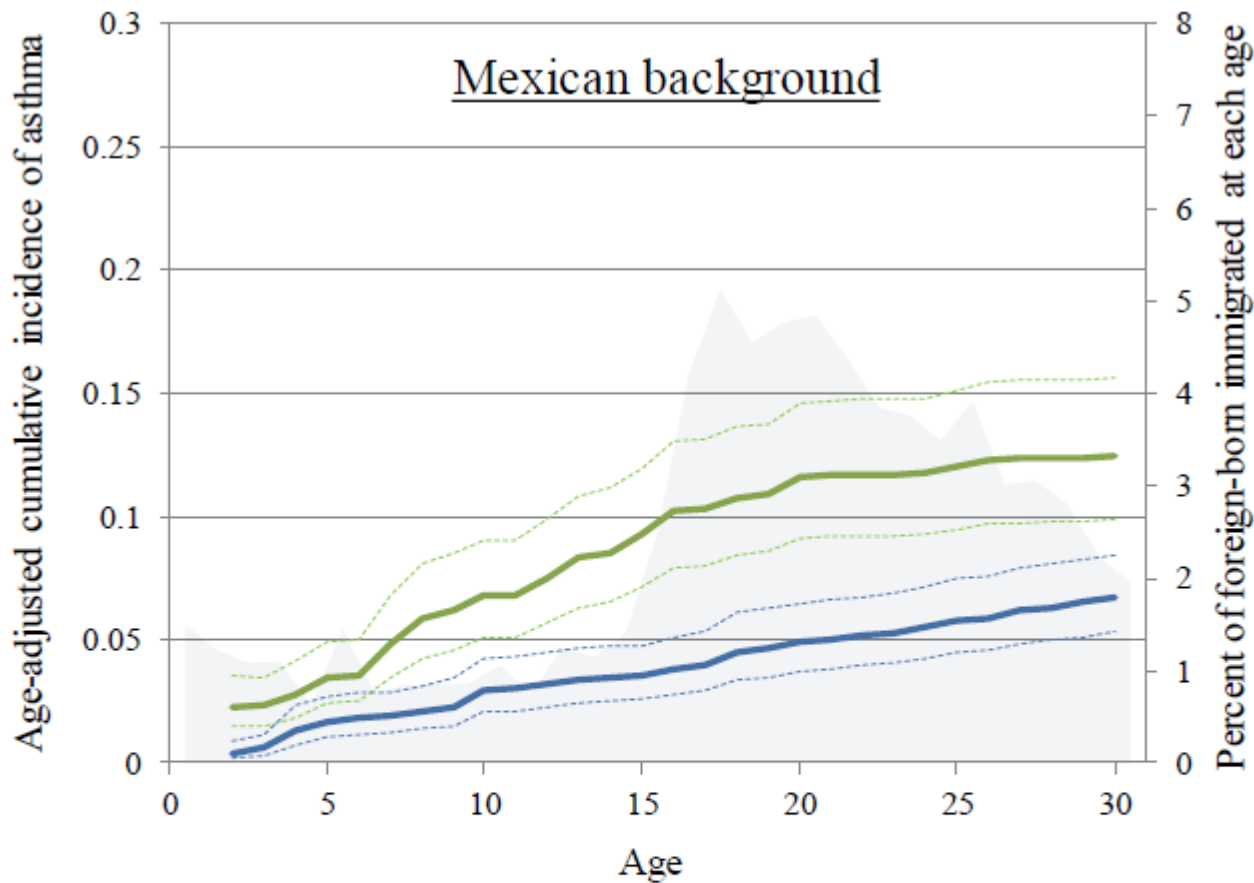
	All	Cuba	Domin. Republic	Mexico	Puerto Rico	Cent. Amer.	So. Amer.
N	16,415	2,201	1,400	6,232	2,590	1,634	1,022
Men, %	40	46	34	37	41	39	40
College, %	15	20	15	12	14	14	22
Income >\$50K, %	11	8	7	14	14	7	11
Prefer Spanish, %	77	91	80	81	42	89	89
US residence <10 years, %	31	55	27	27	8	38	47

Asthma prevalence

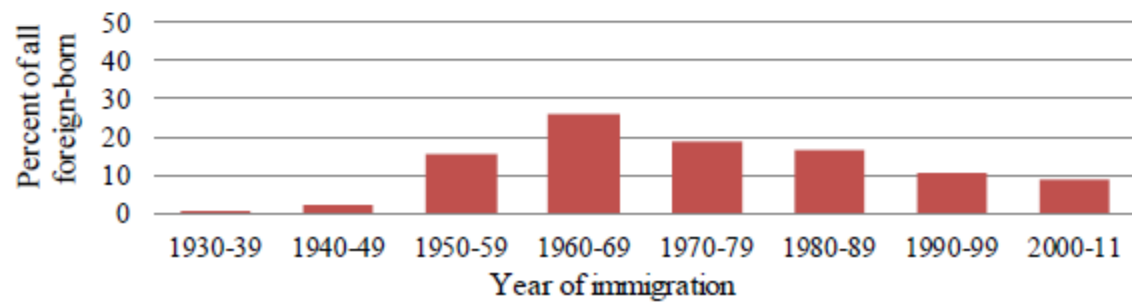
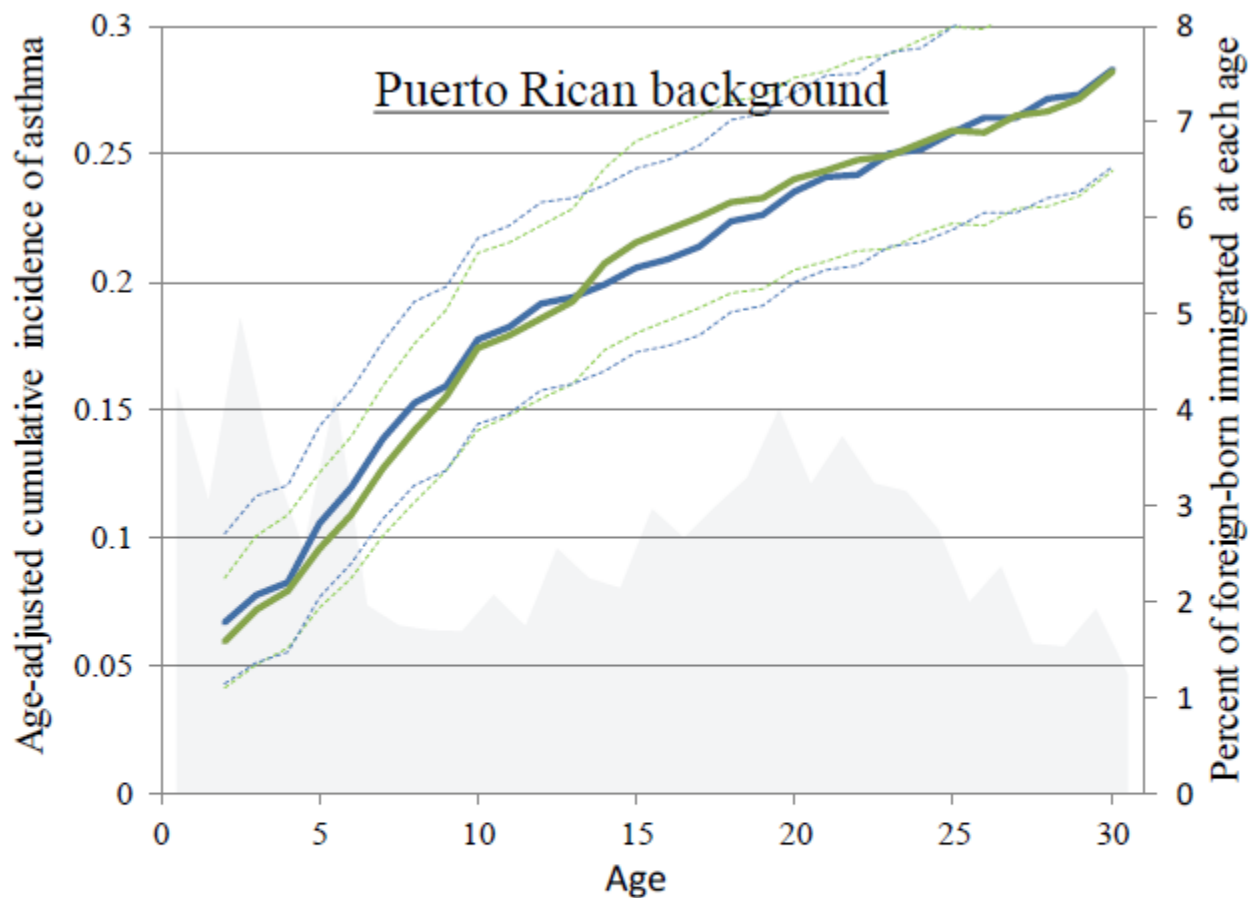
	Mexican	Puerto Rican	Cuban	Central American	Dominican	South American
Physician-diagnosed asthma, ever	7.8 (6.8–8.9)	30.9 (27.8–34)	23.3 (21.1–25.7)	12.8 (10.7–15)	15.7 (13.0–18.7)	10.4 (7.9–13.5)
Physician-diagnosed asthma, current	3.4 (2.7–4.1)	15.3 (13.2–17)	8.6 (7.1–10.4)	4.5 (3.4–6.0)	6.7 (5.3–8.5)	4.3 (2.9–6.2)

Asthma prevalence is highest in Latinos who came to the 50 states during early childhood





- Age-specific immigration trends
- Born outside US mainland
- Born within US mainland
- 95% confidence intervals



- Age-specific immigration trends
- Born outside US mainland
- Born within US mainland
- 95% confidence intervals

SOL Genetic Projects

- Omics in Latinos (OLa) GWAS
- PAGE consortium
 - MetaboChip
 - MEGAChip
- Whole genome sequencing

SOL Genetic Projects

- *Omics in Latinos (OLa) GWAS*
- PAGE consortium
 - MetaboChip
 - MEGACHIP
- Whole genome sequencing

OLa GWAS Project

SNP Microarray Data

12,803 study participants provided consent for genetic studies

Illumina Omni2.5M + ~150k custom content

'SoL_HCHS_Custom_15041502_B'

Array design – Papanicolaou (NHLBI), Rotter and Taylor (LABiomed)

Genotyping performed by Illumina & QC by LABiomed

QA by SOL Genetic Analysis Center (Univ. Washington)

Imputation to 1000G phase 3

SOL Genetic Analysis Working Groups

Ankle Brachial Index	Dental
Anthropometrics	Diabetes
Anxiety/Depression	Electrocardiography
Blood Cell Count	Pulmonary Disease
Blood Pressure	Sleep
Chronic Kidney Disease	mtDNA/Y Chr
Lipids	Reproductive traits
	Smoking



Genetic Analysis Center

Home > GAC Home

Genetic Analysis Center

- GAC Home
- dbGaP Posting
- ▷ Directory
- ▷ Calendar
- ▷ GAC Meetings
- GAC Reports
- GAC Working Groups Wiki
- "HCHS/SOL Confidential" Watermark
- News
- Non-Disclosure Agreement
- ▷ Phenotype Metadata
- Publications and Proposals

Search

Search this site:

gac-user

- My account
- ▷ Create content
- Log out

GAC Home

www.olgastudy.org

The Genetic Analysis Center (GAC) at the University of Washington, Seattle is the working arm of the Genetic Analysis Committee (GAC) of the Hispanic Community Health Study / Study of Latinos (HCHS/SOL). The GAC and the GSC are led by Bruce Weir, Professor of Biostatistics, and Cathy Laurie, a senior principal research scientist at the University of Washington. The GAC's approach was developed during its work as the Coordinating Center for the GENEVA and GARNET consortia.

All documents posted on this web site are **HCHS/SOL-confidential** and a signed **non-disclosure agreement** is required for access. Please contact **Kate Wehr** (katewehr@uw.edu) for more information.

The Genetic Analysis Center will perform genotypic data cleaning on genome-wide SNP microarray data for approximately 13,000 subjects for HCHS/SOL. The cleaned data, quality metrics and quality control report will be used for downstream analyses and posted on the **Database of Genotypes and Phenotypes (dbGaP)**, which is housed at the National Center for Biotechnology Information (NCBI).

The GAC will impute genotypic data using a **1000 Genomes** reference panel for downstream analysis and dbGaP posting.

The GAC provides statistical support to the HCHS/SOL genetic working groups that will publish papers focusing on risk factors and traits of interest for specific conditions.

The NIH Heart, Lung, and Blood Institute (NHLBI) program officials facilitate the achievement of scientific goals and provide institutional oversight and guidance to the project.

HCHS/SOL Genetic Analysis Center Working Groups

Please log in to see [this page with full email lists of members](#).

Working Group

Ankle Brachial Index (ABI) Genetics

Members

Matthew Allison, Co-chair
 Michael Criqui, Co-chair
 Nora Franceschini
 Cecelia Laurie
 Mary McDermott McGrae
 Christina Wassel
 Carmen Isasi, Chair
 Kari North, Co-chair
 Lindsay Fernández-Rhodes
 Nick Gekakis
 Stephanie Gogarten

Scroll down for all working groups and their members
 Directory has contact information

De-convolution of ancestry and relatedness in admixed populations

Matt Conomos and Tim Thornton (University of Washington)

1. Estimate relatedness using KING-robust, which is robust to discrete population structure but not to admixture or departures from HWE within sub-populations.
2. Partition the sample into a mutually unrelated set and the remaining (relatives of the unrelated set and possibly each other)
3. Perform standard principal components analysis (PCA) on the set of unrelated individuals and project onto related individuals
4. Re-estimate relatedness using 'PC-Relate', which provides unbiased kinship coefficients in the presence of population structure, admixture and HWE departures, using individual-specific allele frequencies estimated from sample eigenvectors.
5. Repeat steps 2-5 to get final sets of eigenvectors and kinship coefficients

Genetic Analysis Issues

1. Tiered consent allows variable levels of data sharing. Reconsenting at each in person visit may necessitate withdrawal of subjects

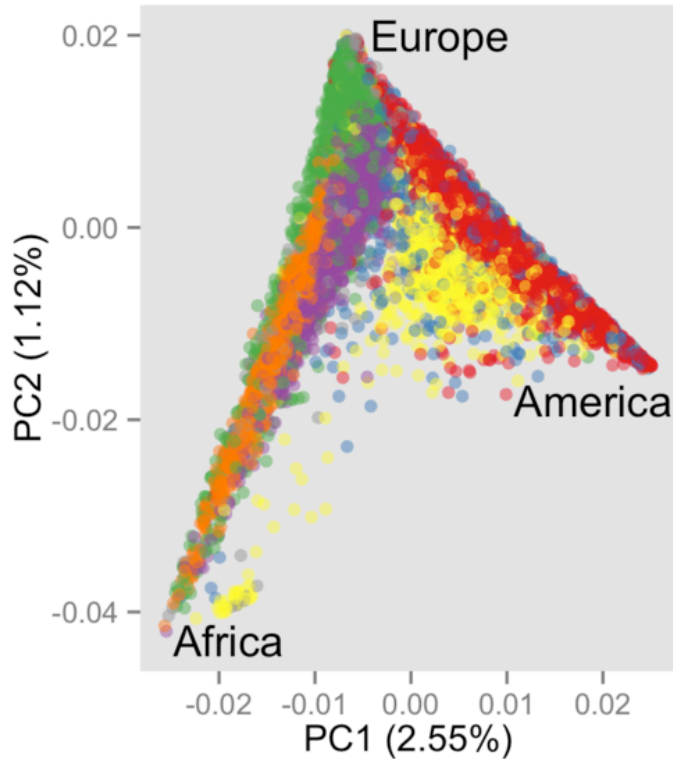
1. Sample survey design, 2-stage probability sampling
 - a. Primary sampling unit = US census block group
 - b. Secondary sampling unit = household
 - c. Household based sampling (average 1.8 enrolled per HH)

2. Relatedness, population structure and admixture
 - a. ~85% of subjects are mutually unrelated
 - b. ~15% are each related to someone in the unrelated set
 - c. SOL participants are very diverse ethnically and genetically
 - i. Caribbean groups: Cuban, Dominican, Puerto Rican
 - ii. Mainland groups: Mexico, Central America, South America
 - d. Novel method used to de-convolute ancestry and relatedness (Conomos & Thornton)

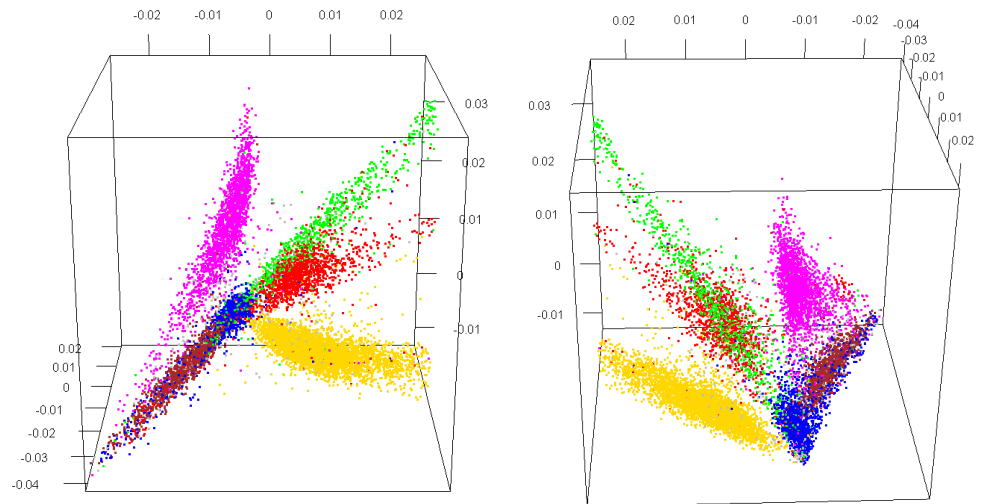
Mixed model and GEE approaches have been developed to handle these issues in association tests (GWAS). Both control Type I error well, but mixed model has more power. (Mixed model – Conomos + Thornton; GEE - Lin + Tao)

Population structure

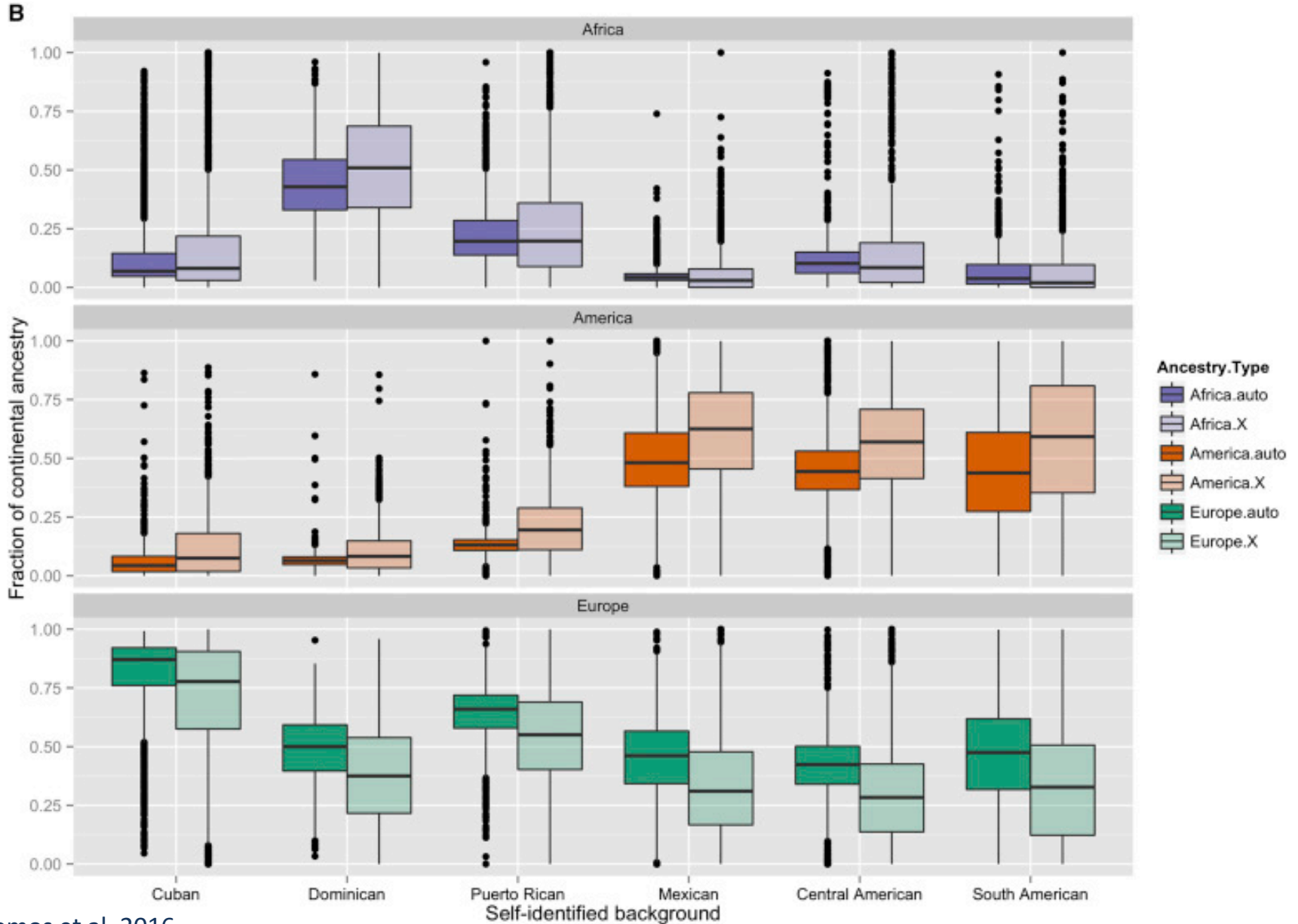
Principal components analysis (PCA)



PCA plots (EV 1-3)

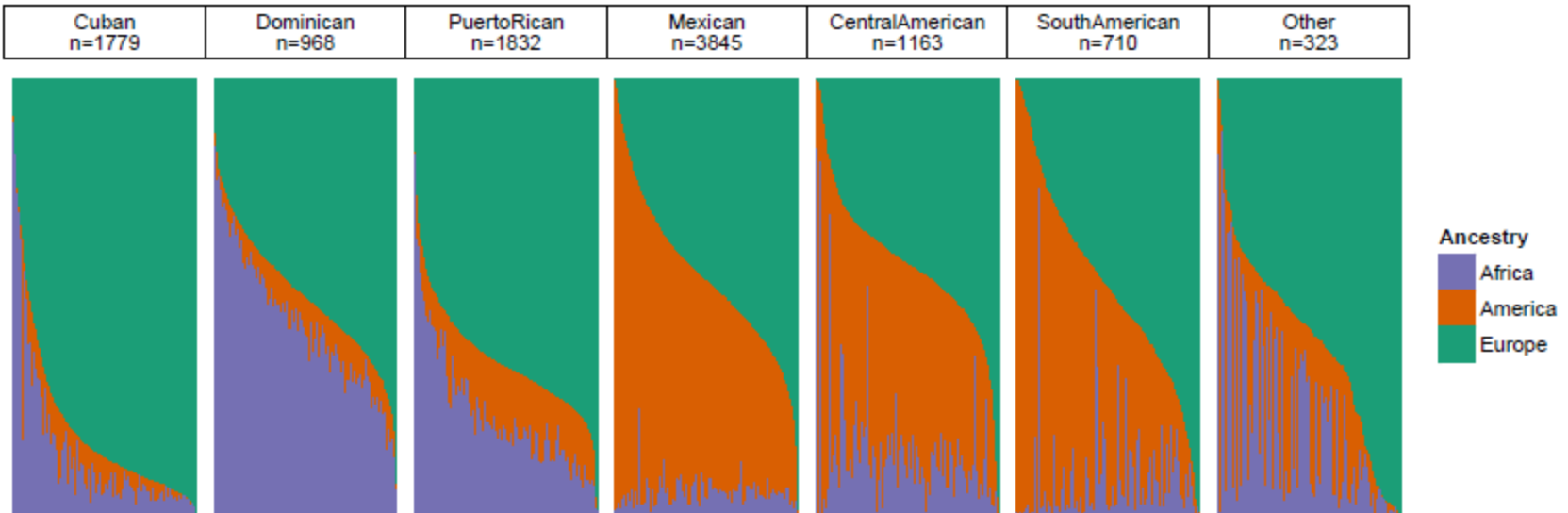


Proportion of ancestral backgrounds

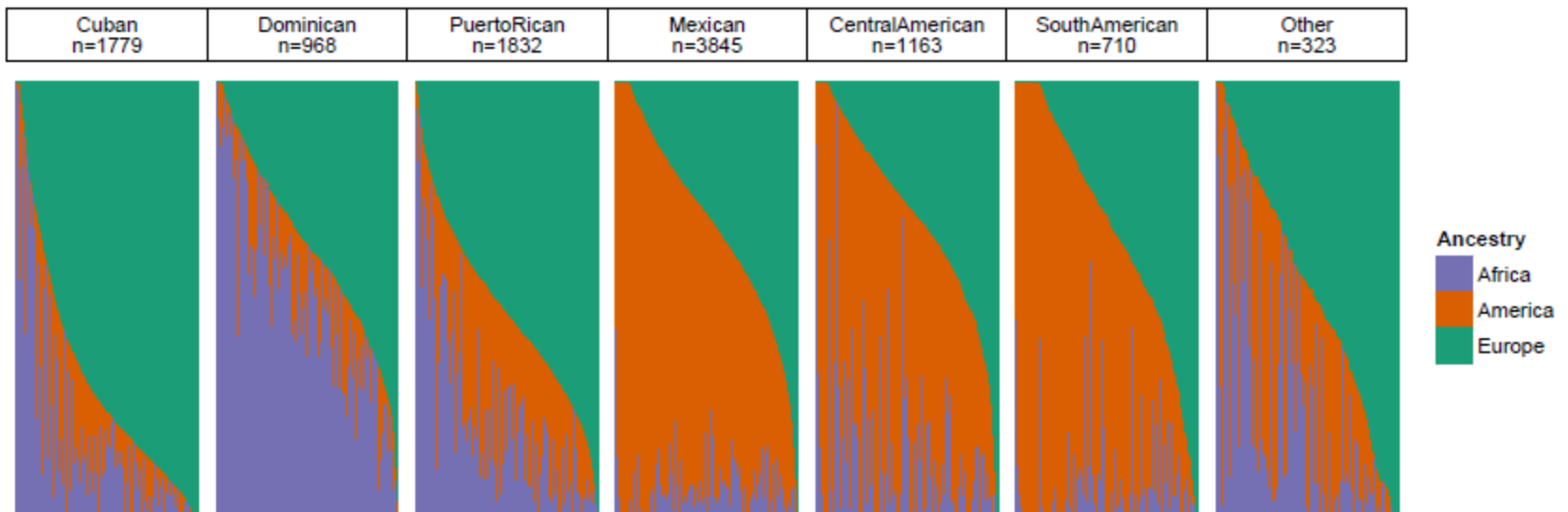


Continental ancestry proportions estimated using ADMIXTURE software

Autosomal

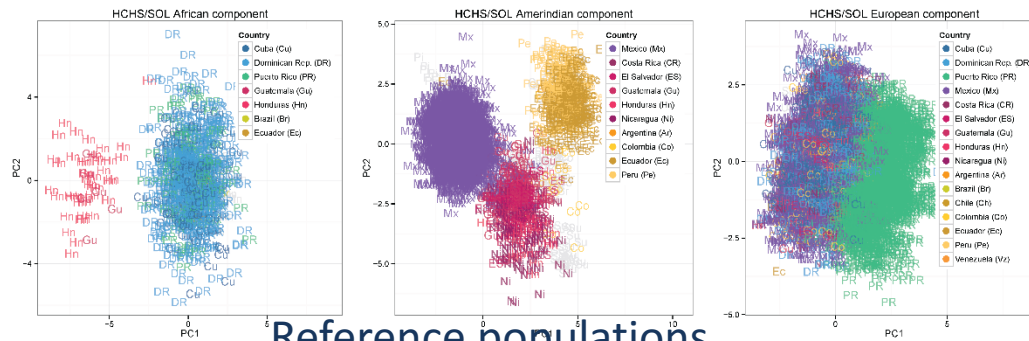


X chromosome

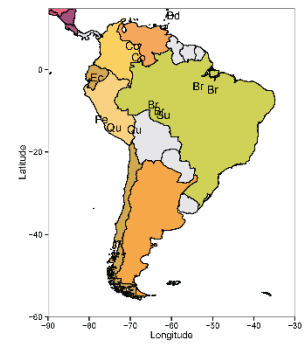
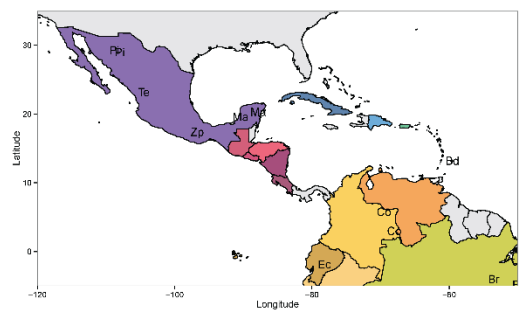
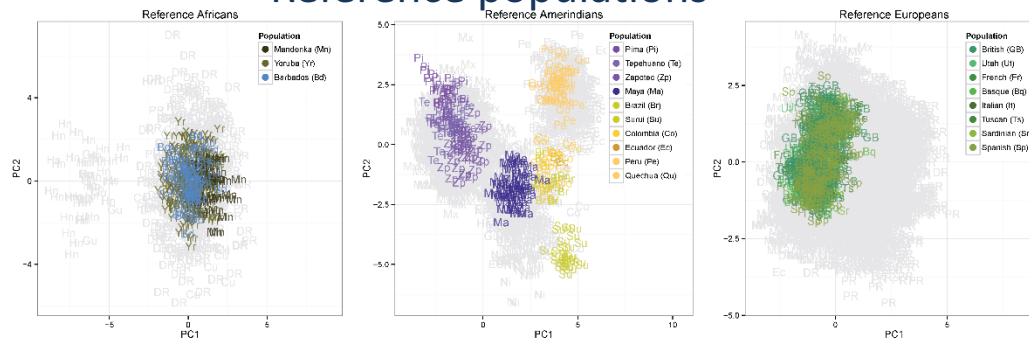


Separate PCA for each of the 3 ancestries

HCHS-SOL groups



Reference populations

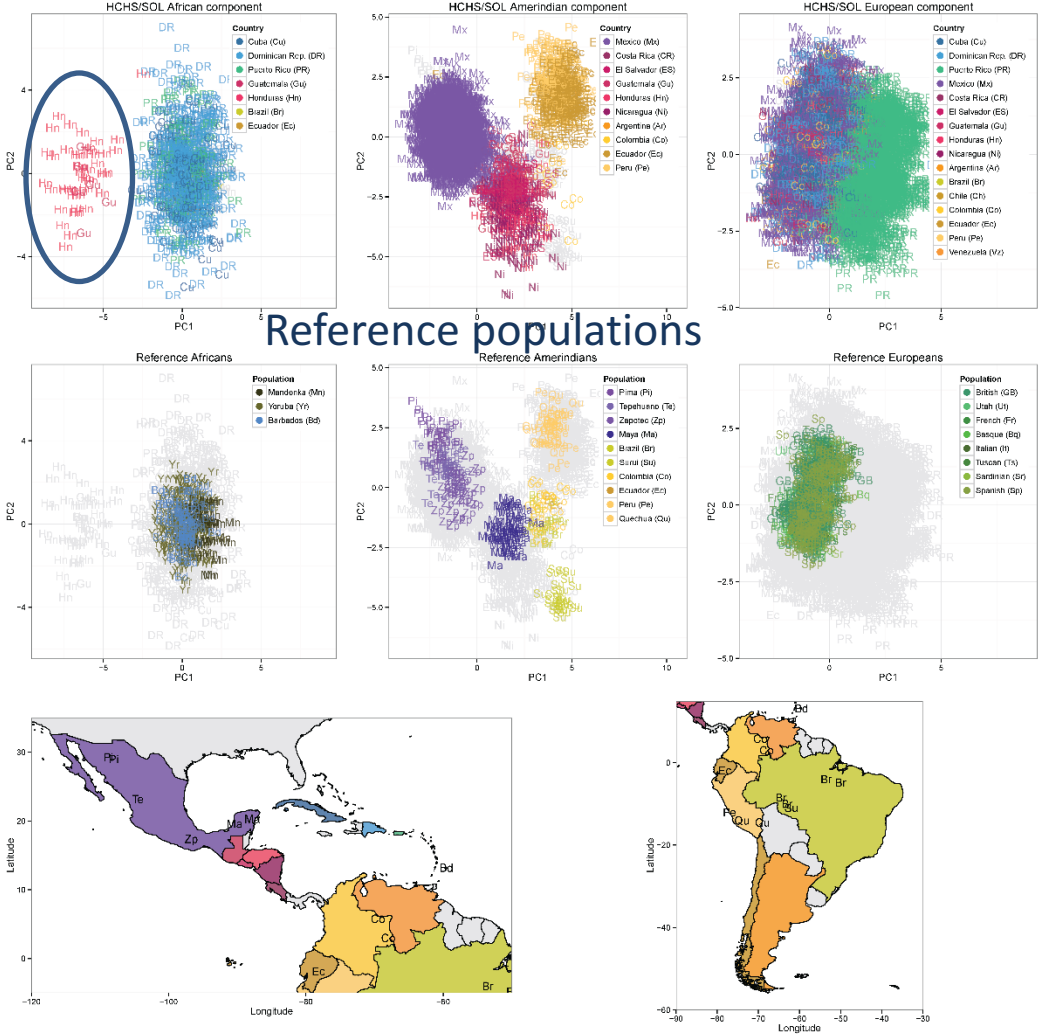


Africa

HCHS-SOL groups

Separate clustering of African component of Hondurans/Guatemalans

- Garífuna population.
 - May have ancestry from specific parts of Africa.
 - Or drift due to small population size.

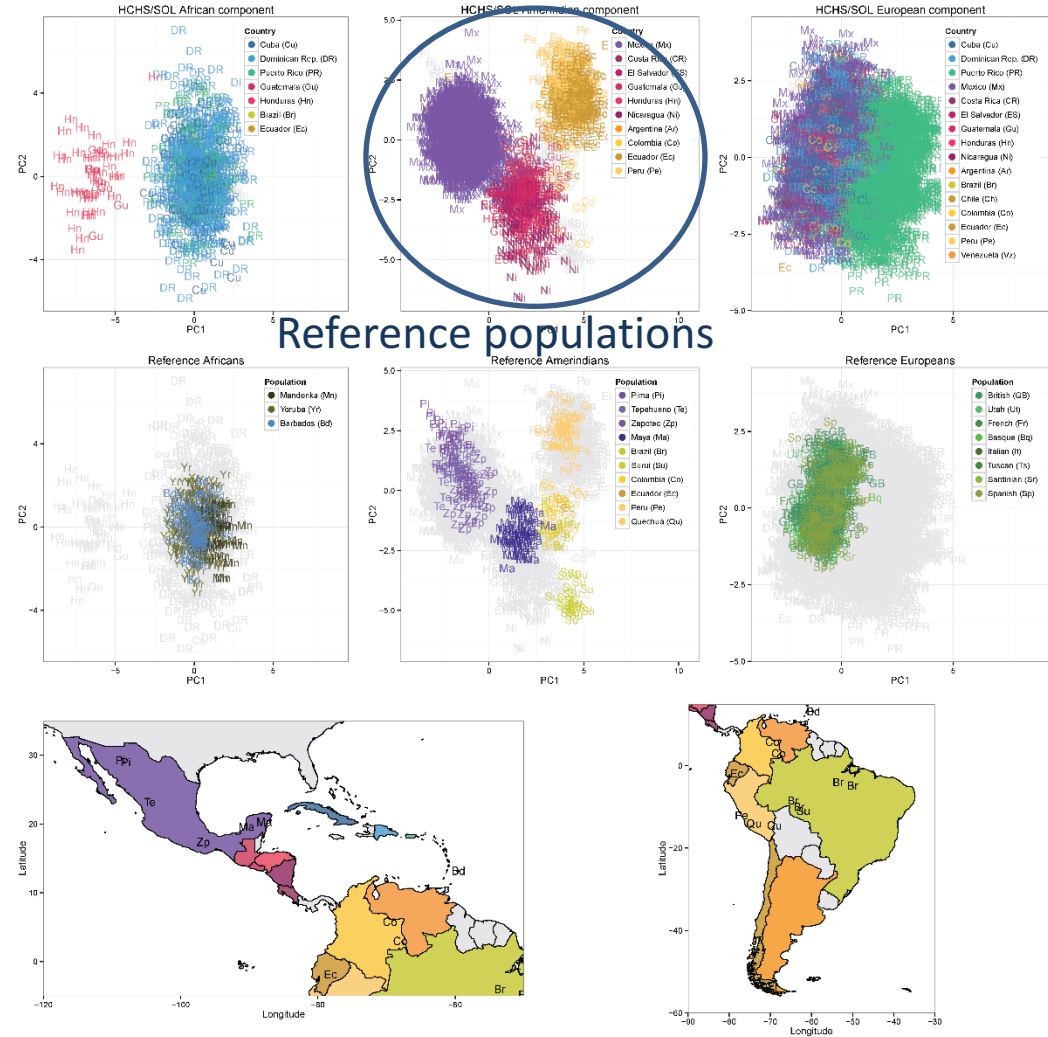


Americas

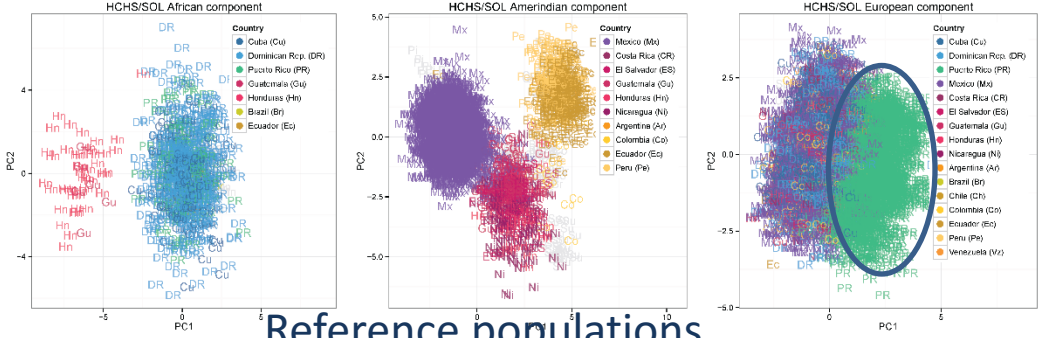
HCHS-SOL groups

Three distinct Amerindian clusters

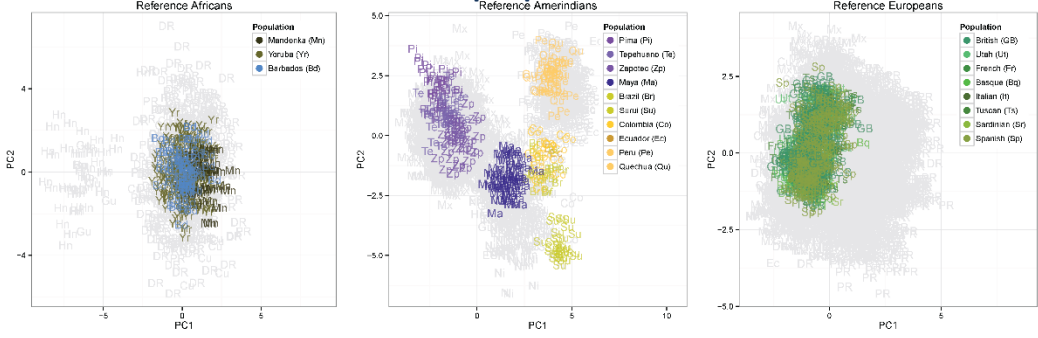
- Mexico, Central America, South America.



HCHS-SOL groups

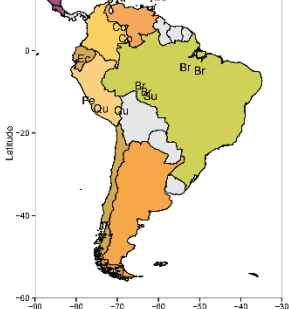
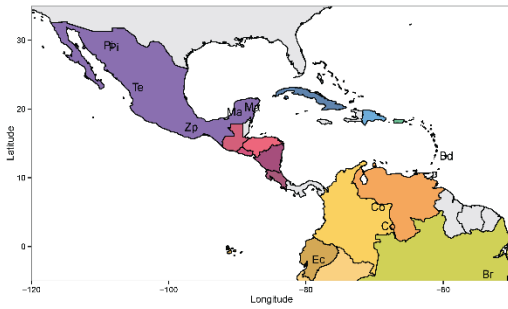


Reference populations

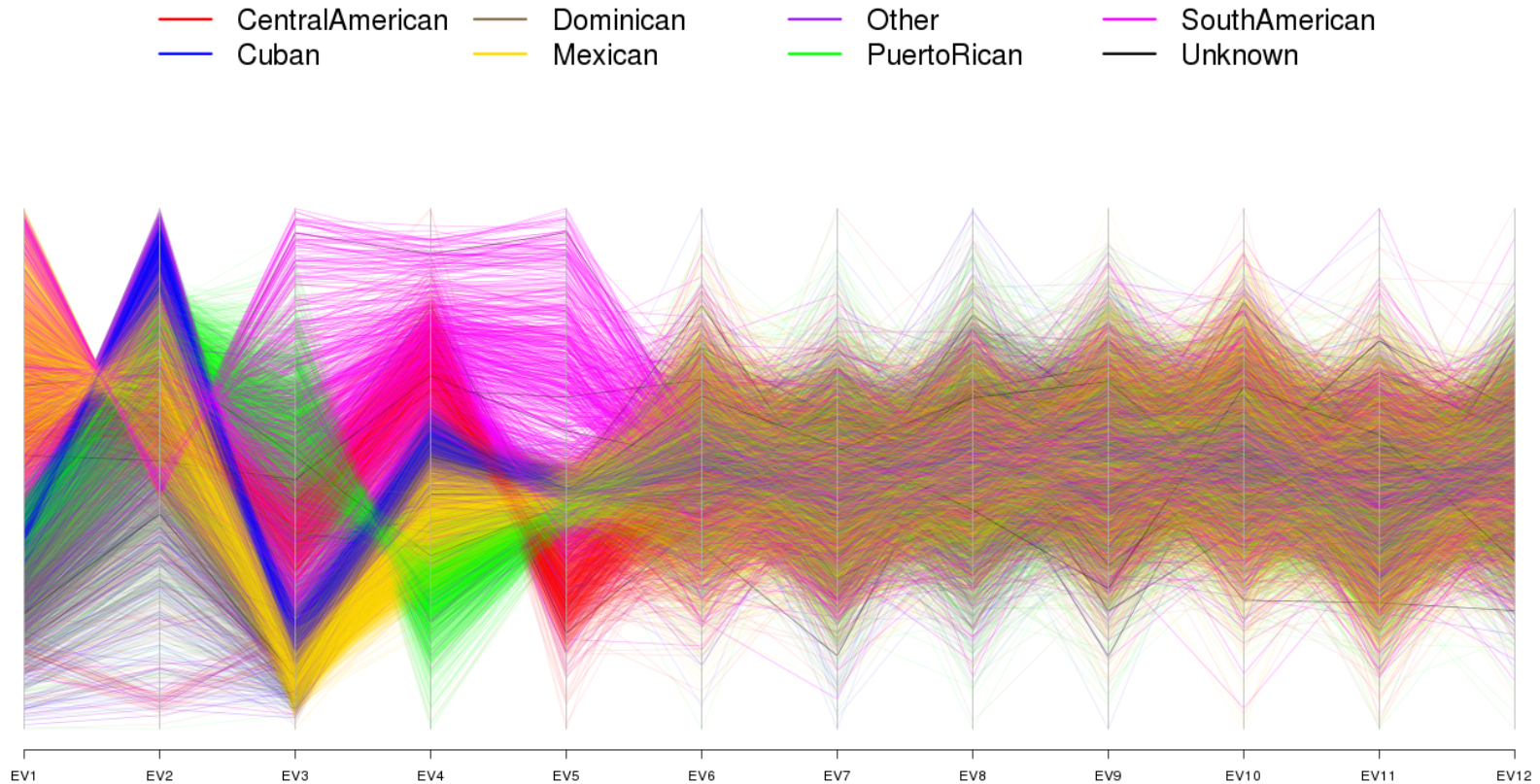


Separate clustering of European component of Puerto Ricans

- May include ancestry from North African or Middle East.
- Or drift due to small population size.



Each of the first 5 Eigenvectors differentiates among self-identified groups




- Use these 5 eigenvectors to adjust for ancestry in association tests and to define genetic analysis groups for stratified analysis
- Compared with self-identified background (personal or family place of origin), the groups are more homogeneous genetically and include individuals with missing or “other” self-identification

SOL Genetic Projects

- Omics in Latinos (OLa) GWAS
- *PAGE consortium*
 - *MetaboChip*
 - *MEGACHIP*
- Whole genome sequencing




Charles Kooperberg
FHCRC
Women's Health Initiative


Chris Haiman
USC
Multiethnic Cohort Study


Tara Matisse
Rutgers University
Coordinating Center


Ruth Loos
Mount Sinai School of Medicine
MSSM Biobank

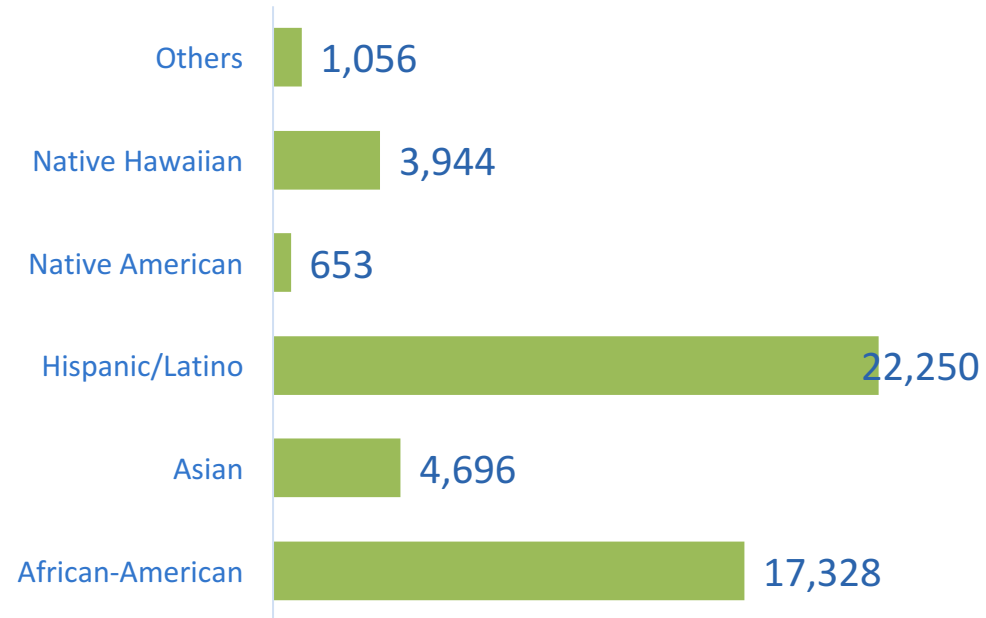

Kari North
UNC-CH
CALiCO Consortium

Population Architecture using Genomics and Epidemiology (PAGE) Study

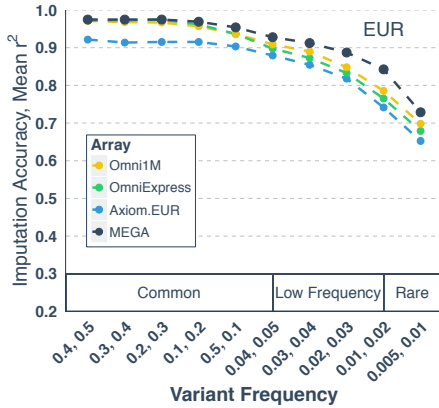
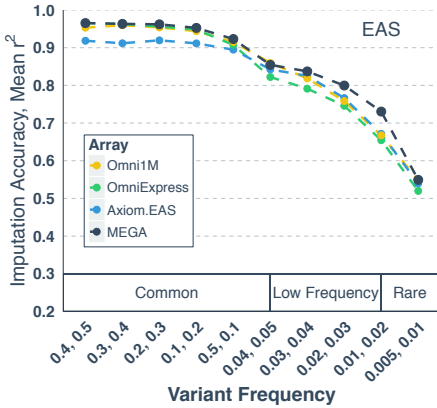
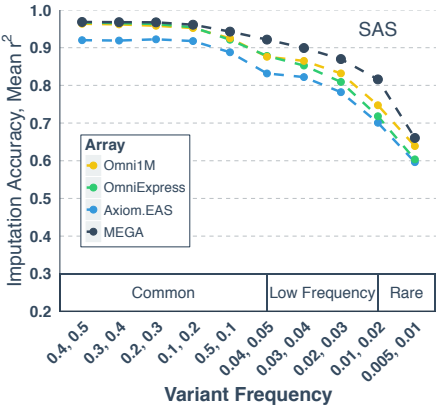
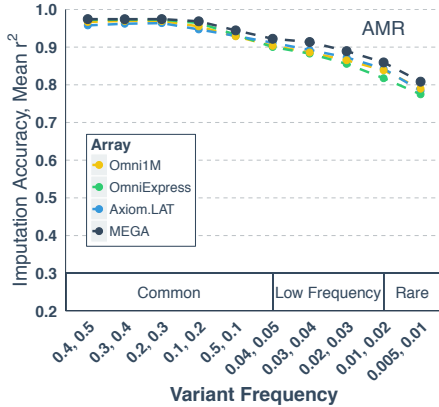
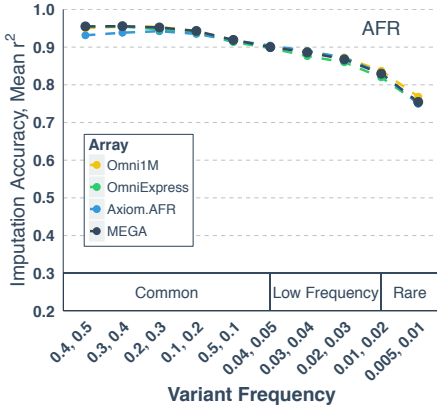
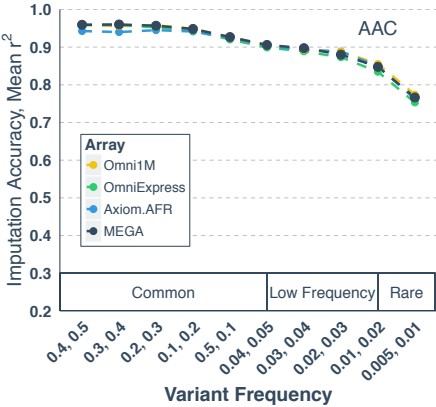
The PAGE Study

Population Architecture using Genomics and Epidemiology II Network

- **Goal:** Investigate ancestrally diverse populations to gain a better understanding of how genetic factors influence susceptibility to disease.
- **Focus on US minority populations.**



Comparison of MEGA Imputation Accuracy to Other Commercial Chips (Supplementary Figure)



SOL Genetic Projects

- **Omics in Latinos (OLa) GWAS**
- PAGE consortium
 - MetaboChip
 - MEGACHIP
- ***Whole genome sequencing***

HCHS/SOL sample selection for whole genome sequencing

N=1917 asthma cases (ever diagnosed by a physician)

N=4503 controls (never diagnosed by a physician)

NHGRI/PAGE ~270 samples selected for high Amerindian ancestry

Primarily Central and South American (only one Puerto Rican)

Washington U. sequencing center

NHGRI/CCDG ~4000 samples selected at random Baylor sequencing center

TOPMed Total approved allocation of 2150 Baylor sequencing center

1. All ever-asthma cases not previously selected by PAGE or CCDG, N=1277. (Bringing total of ever-asthma cases to 1917 across all three sample selections)
2. Six samples previously selected by PAGE, which will serve as cross-sequencing center controls (one per genetic analysis group e.g. Cuban, Mexican, Dominican, etc), N=6
3. Remaining Puerto Ricans not included in PAGE or CCDG selections, N=867.

We thank the participants and staff of the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) for their contributions to this study. The baseline examination of HCHS/SOL was carried out as a collaborative study supported by contracts from the NHLBI to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The following institutes, centers, and offices contributed to the first phase of HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research (NIDCR), National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Neurological Disorders and Stroke, and NIH Office of Dietary Supplements. The Genetic Analysis Center at the University of Washington was supported by NHLBI and NIDCR contracts (HHSN268201300005C AM03 and MOD03). Additional analysis support was provided by 1R01DK101855-01 and 13GRNT16490017. Genotyping efforts were supported by the NIH Department of Health and Human Services (HSN26220/20054C), National Center for Advancing Translational Science Clinical Translational Science Institute (UL1TR000124), and NIDDK Diabetes Research Center (DK063491).