

# Population Descriptors in Genomics Research: What classifications are used?

**Shawneequa L. Callier, JD, MA**

Associate Professor	Special Volunteer
Clinical Research and Leadership	Center for Research on Genomics
School of Medicine and Health Sciences	& Global Health
George Washington University	National Human Genome Research Institute
	National Institutes of Health

---

International HundredK+ Cohorts Consortium

Virtual Conference

November 4, 2021

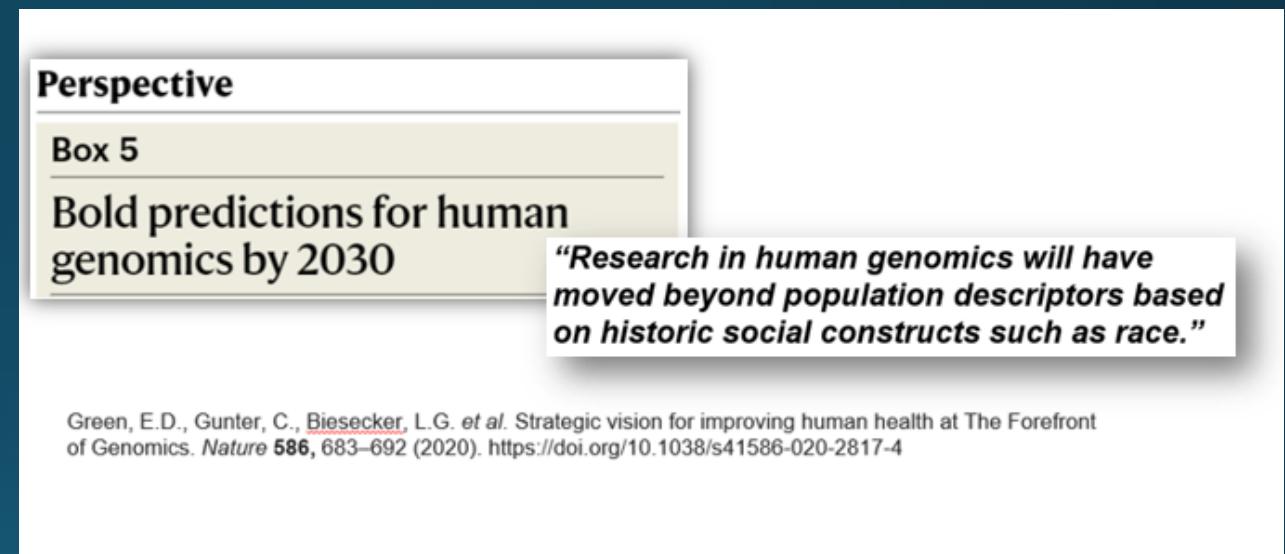
# Agenda

1. The goal to harmonize cohorts in global genomics research
2. An IHCC cohort study to understand how researchers categorize populations
3. The challenges of using population descriptors in research
4. Opportunities for IHCC

# Genomics Research Landscape

## Goals

1. Global diversity and inclusion
2. Harmonization or standardization of terms
3. Determination of what descriptors we should use



# IHCC Survey

- Objectives
  - Understand how IHCC cohorts record population descriptors
  - Assess IHCC cohort opportunities and challenges related to collecting, defining, and reporting data
  - Use findings to inform discussion about harmonization and standardization of global data
- Status
  - IRB-approved study
  - Data-collection phase

# IHCC Survey & Research Project

## Harmonization of race, ethnicity, and ancestry measures across the globe

Collaboration with International HundredK+ Cohorts Consortium (IHCC)



	Race	Ethnicity	Ancestry	Tribe	Other
Brazil	Yes	Yes	Yes	No	No
France	No	No	Yes	No	Yes
Malaysia	Yes	Yes	Yes	No	No
Mexico	No	No	No	No	No
Qatar	No	No	No	No	No
United States	Yes	Yes	Yes	No	No
US Cohort 1	Yes	Yes	Yes	No	No
US Cohort 2	Yes	Yes	No	No	No

# Race / Ethnicity

## Race:

- 1 White
- 2 Black/African-American
- 3 Asian
- 4 American Indian/Alaska Native
- 5 Native Hawaiian/Pacific Islander
- 6 Other

## Ethnicity:

Spanish/Hispanic/Latino  
yes/no



## What is your race?

- 1 White
- 2 Black, African-American, or Negro
- 3 American Indian or Alaska Native
- 4 Asian Indian
- 5 Chinese
- 6 Filipino
- 7 Japanese
- 8 Korean
- 9 Vietnamese
- 10 Other A Black
- 11 Native f Mixed
- 12 Guamar Asian
- 13 Samoan Indigenous
- 14 Other Pacific Islander

## Are you Spanish/Hispanic/Latino?

- 0 No, not Spanish/Hispanic/Latino
- 1 Yes, Puerto Rican
- 2 Yes, Mexican, Mexican American, or Chicano
- 3 Yes, Cuban
- 4 Yes, other Spanish/Hispanic/Latina

National Statistics Department (no detailed information provided)



# Ancestry

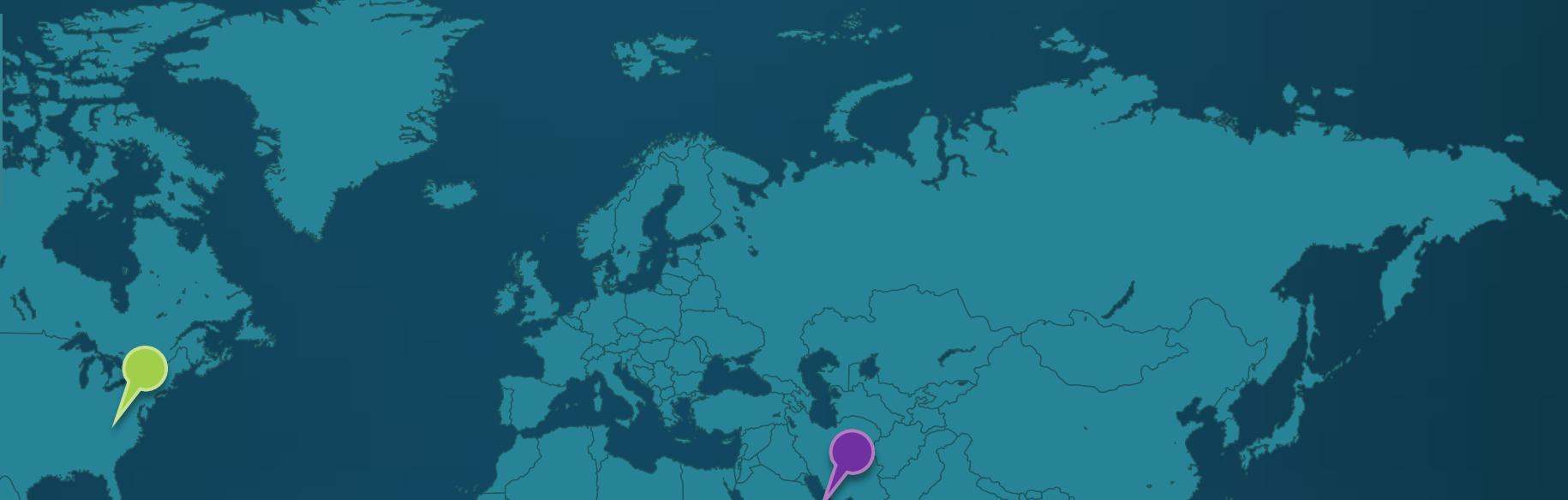
European  
African  
Amerindian

Participants and  
their parents'  
country of birth

Note:

For Malaysia cohort and US cohort 2, no detailed information was provided in terms of Ancestry categories.

# Genetic Assessment



	Genetic Assessment	Ancestry Assignment
Brazil	Yes	Yes
France	No	No
Malaysia	No	No
Mexico	No	No
Qatar	Yes	Yes
US Cohort 1	Yes	Yes
US Cohort 2	No	No

# The challenges of using population descriptors in research



ONTOLOGY SEARCH

## Human Ancestry Ontology

Human ancestry ontology for the NHGRI GWAS Catalog

- ancestry category
  - Aboriginal Australian
  - African
    - Sub-Saharan African
  - African American or Afro-Caribbean
    - African American
    - Afro-Caribbean
    - Anguillan
    - Antiguan or Barbudan
    - Barbadian
    - Haitian
    - Jamaican

- ancestry status
  - admixed ancestry
  - genetically isolated ancestry
- continent
  - Africa
  - Asia
  - Europe
  - Latin America and the Caribbean
  - Northern America
  - Oceania
- country
- region

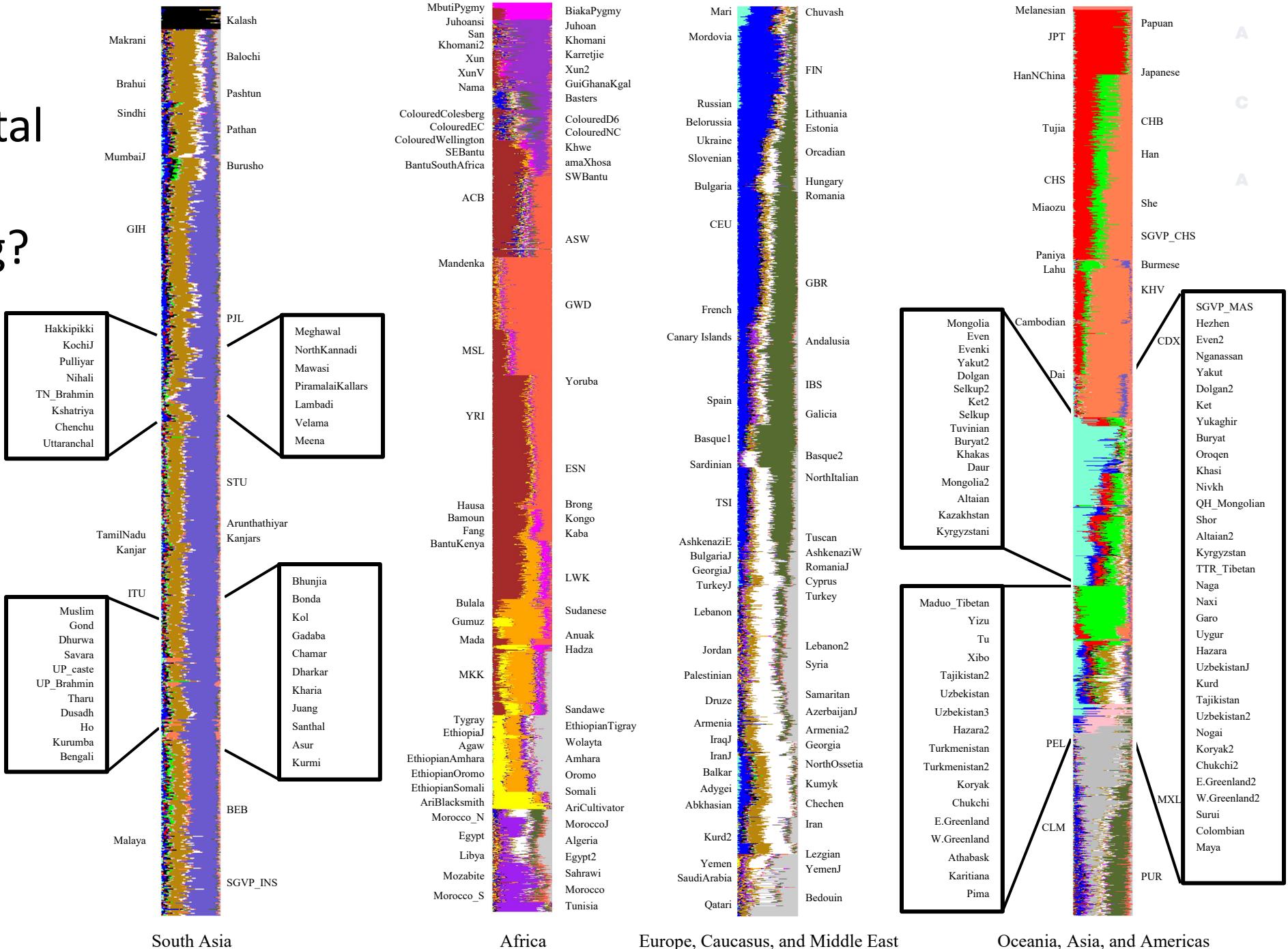
- Asian
  - Central Asian
  - East Asian
    - Filipino
    - Nepalese
  - South Asian
  - South East Asian

- Hispanic or Latin American
  - Brazilian
  - Costa Rican
    - Costa Rican founder
  - Cuban
  - Dominican
  - Mexican
  - Puerto Rican
- Native American
  - Pima Indian
  - Plains American Indian
- Oceanian
  - Kosraen
  - Micronesian
  - Native Hawaiian
  - Pacific Islander
  - Papua New Guinean
  - Solomon Islander

- European
  - Amish
  - British
- Canadian
  - French Canadian founder
- Dutch
  - Dutch founder
    - Erasmus Rucphen
    - European American
- Finnish
  - Hutterite
  - Icelandic
- Irish
  - genetically homogenous Irish
- Italian
  - Carlantino
  - Cilento
  - Friuli Venezia Giulia
  - Italian founder
    - Sardinian
    - Talana
    - Val Borbera
  - Korculan
  - Old Order Amish
  - Orcadian
  - Romanian founder
  - Sorbian
  - South Tyrolean
  - Spanish
  - Split founder
  - Vis founder

# What are continental labels capturing?

Baker, J.L., Rotimi, C.N. & Shriner, D. Human ancestry correlates with language and reveals that race is not an objective genomic classifier. *Sci Rep* 7, 1572 (2017).



# What are continental labels capturing?

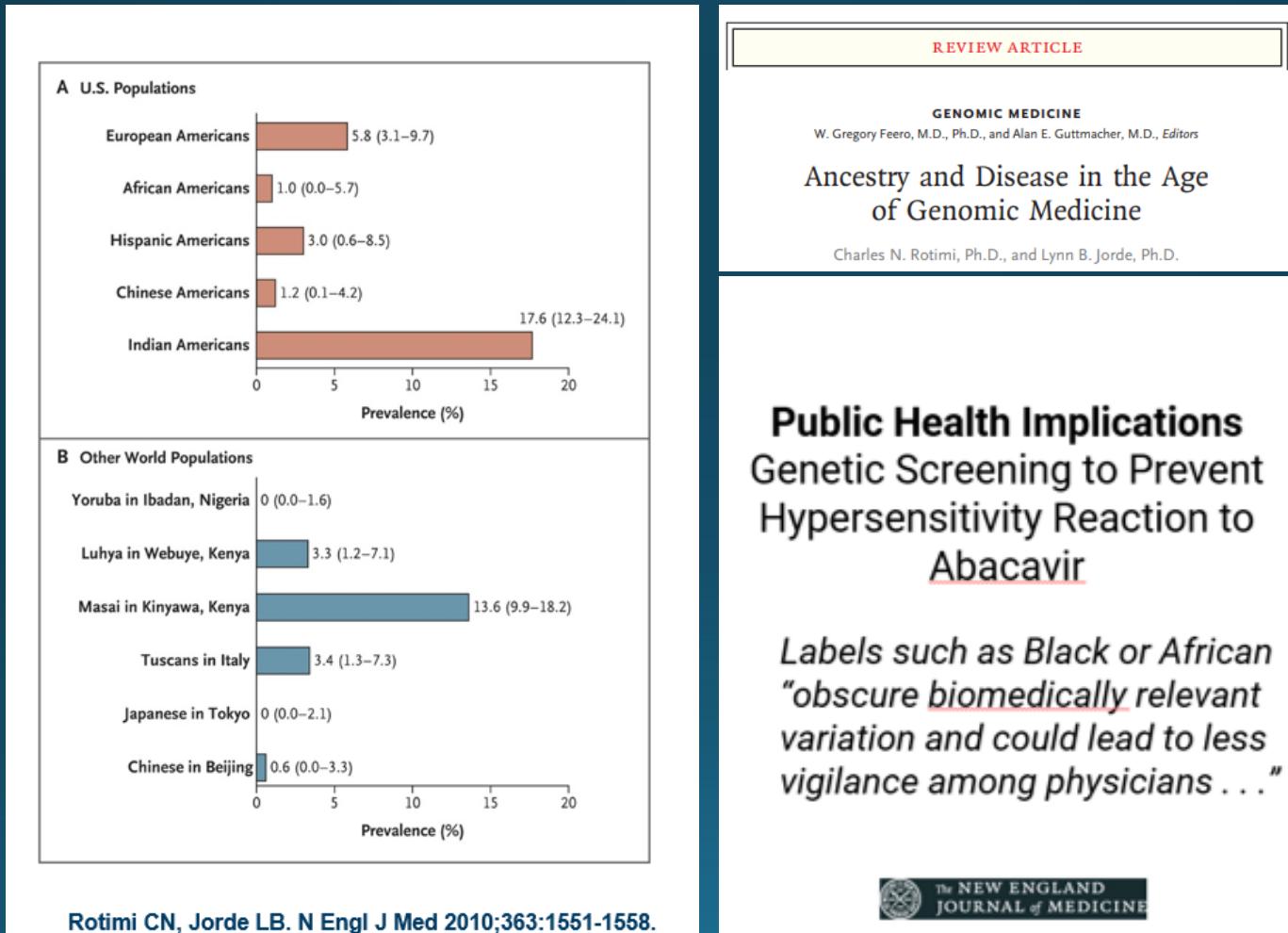


Figure 1. Variation in the HLA-B\*5701 Locus in 11 HapMap Samples.

# What are racial and ethnic labels capturing?



## Race in a Genetic World

MAY-JUNE 2008

HARVARD MAGAZINE

"I AM AN African American," says Duana Fullwiley, "but in parts of Africa, I am white." To do fieldwork as a medical anthropologist in Senegal, she says, "I take a plane to France, a seven- to eight-hour ride. My race changes as I cross the Atlantic. There, I say, '*Je suis noire*', and they say, 'Oh, okay—*métisse*—you are mixed.' Then I fly another six to seven hours to Senegal, and I am white. In the space of a day, I can change from African American, to *métisse*, to *tubaab* [Wolof for "white/European"]. This is not a joke, or something to laugh at, or to take lightly. It is the kind of social recognition that even two-year-olds who can barely speak understand. '*Tubaab*,' they say when they greet me."

Dr. Duana Fullwiley  
Anthropologist  
Stanford University

Nationalgeographic.com/magazine/2018/04/race-twins-black-white-biggs

### ARTICLE

#### Chad Genetic Diversity Reveals an African History Marked by Multiple Holocene Eurasian Migrations

Marc Haber,<sup>1,\*</sup> Massimo Mezzavilla,<sup>1,2</sup> Anders Bergström,<sup>1</sup> Javier Prado-Martinez,<sup>1</sup> Pille Hallast,<sup>1,3</sup> Riyadah Saif-Ali,<sup>4</sup> Molham Al-Habori,<sup>4</sup> George Dedoussis,<sup>5</sup> Eleftheria Zeggini,<sup>1</sup> Jason Blue-Smith,<sup>6,10</sup> R. Spencer Wells,<sup>7</sup> Yali Xue,<sup>1</sup> Pierre A. Zalloua,<sup>8,9</sup> and Chris Tyler-Smith<sup>1,\*</sup>

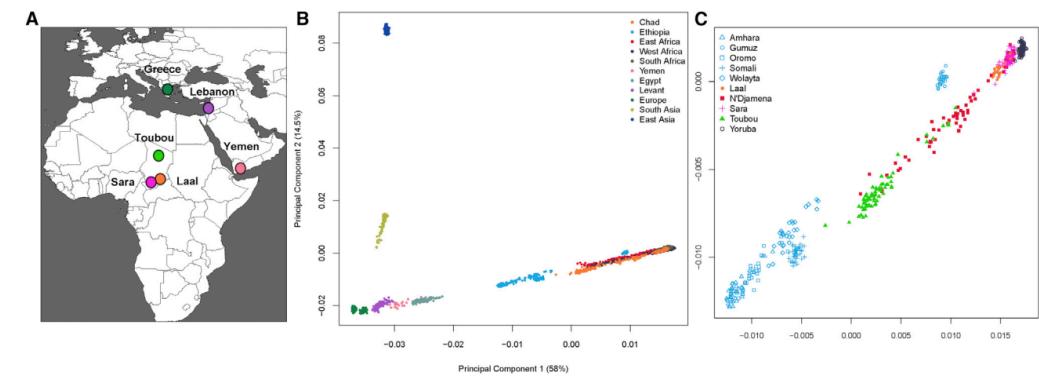


Figure 1. Population Locations and Genetic Structure

(A) The map shows the location of newly genotyped or sequenced populations.  
(B) PCA of worldwide populations shows that Near Easterners and East Africans are intermediate to Eurasians and sub-Saharan Africans on PC1. Chad populations are close to sub-Saharan Africans and have some samples drawn toward Ethiopians.  
(C) Magnification of the African PCA shows different affinities of the Chad populations to other Africans: the Toubou cluster close to Ethiopians, whereas the Sara and Laal speakers are close to the Yoruba. The mixed samples from N'Djamena, the capital, are intermediate to the Toubou, Sara, and Laal speakers.

The American Journal of Human Genetics 99, 1316–1324, December 1, 2016

# Opportunities for IHCC

- Understand practices regarding population descriptors within a global consortium
- Assess ambiguities and inconsistencies in how cohorts collect, record, and report population data
- Contribute to the discourse on global harmonization in ethical and scientifically appropriate ways