

Standardized biogeographic grouping system for annotating populations in pharmacogenetic research

Rachel Huddart Ph.D.

Scientific Curator, PharmGKB

Background

- Wide variation in the frequency of pharmacogenetic alleles between different global populations.
- Grouping pharmacogenomic studies by population facilitates comparison of results across different studies and feeds into CPIC guidelines.
- Current population grouping methods are subjective, vague or are applied inconsistently.



Previous categories

US Office of Management and Budget (OMB) categories (used by PharmGKB):

- White
- Black or African American
- American Indian and Alaska Native
- Asian
- Native Hawaiian or Pacific Islander
- Hispanic/Latino (additional ethnicity category)

Human Genome Diversity Project (HGDP-CEPH) population labels (used by CPIC)

- African
- American
- Caucasian
- Central/South Asian
- East Asian
- Middle Eastern
- Oceanian
- African American (added by CPIC)

TPMT frequency table

| PMID | Major ethnicity | Population |
|----------|-----------------|--------------------------------------|
| 21400026 | Middle Eastern | Turkish |
| 24499706 | Middle Eastern | Palestinian |
| 20408054 | Middle Eastern | Iranian |
| 19048244 | Middle Eastern | Israeli |
| 19048244 | Middle Eastern | Israeli |
| 23398787 | Middle Eastern | Jordanian |
| 20521035 | Middle Eastern | Jordanian |
| 25940902 | Middle Eastern | Middle Eastern |
| 21938428 | Middle Eastern | Iranian |
| 21348397 | Middle Eastern | Arabs |
| 21348397 | Middle Eastern | Druze |
| 21348397 | Middle Eastern | Kurdish |
| 17617792 | Middle Eastern | Turkish |
| 23065291 | Middle Eastern | Tunisian |
| 25573490 | Middle Eastern | Moroccan |
| 22225964 | Middle Eastern | Tunisian |
| 26811598 | Middle Eastern | Egyptian |
| 12814450 | Middle Eastern | Egyptian |
| 25819542 | Middle Eastern | Libyan |
| 25819542 | Middle Eastern | Libyan |
| 25819542 | Middle Eastern | Libyan |
| 25819542 | Middle Eastern | Libyan |
| 17577869 | Middle Eastern | Lebanese, Palestinian, Syrian, Iraqi |

CYP2C19 frequency table

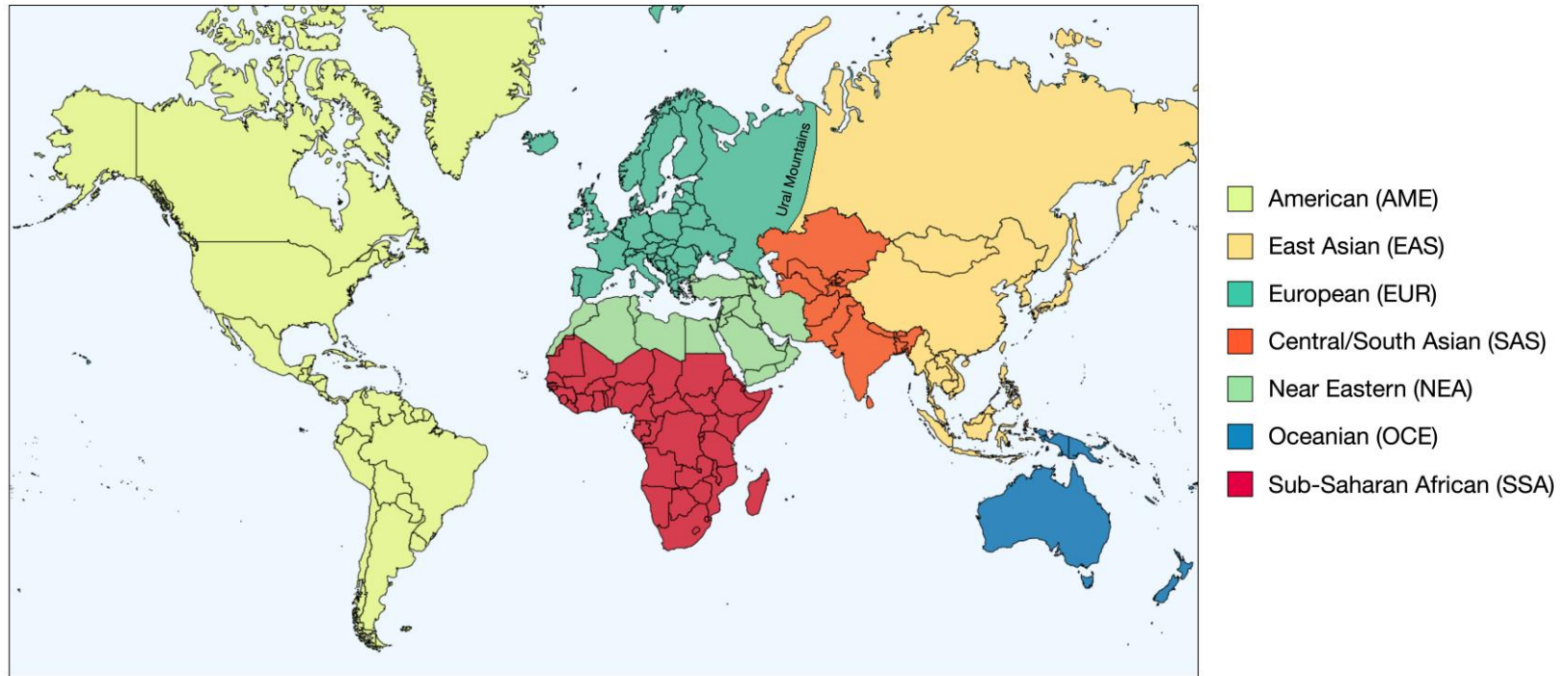
| PMID | Major ethnicity | Population |
|----------|-----------------|---------------------|
| 7781265 | Africa | Zimbabwean |
| 9014201 | Africa | Ethiopian |
| 9797796 | Africa | Tanzanian |
| 10510152 | Africa | Tanzanian |
| 11372584 | Africa | South African |
| 11372584 | Africa | Tanzanian |
| 12047484 | Africa | Egyptian |
| 12142727 | Africa | Ethiopian |
| 14616425 | Africa | Beninese |
| 16413245 | Africa | Ethiopian |
| 18423013 | Africa | Tunisian |
| 19954515 | Africa | Ghanaian |
| 19002442 | Africa | Ugandan |
| 20173083 | Africa | African |
| 20831548 | Africa | Nigerian |
| 20712527 | Africa | South African |
| 20712527 | Africa | South African |
| 22118051 | Africa | Black South African |
| 22352331 | Africa | Egyptian |
| 23356658 | Africa | South African |
| 26021325 | Africa | Ethiopian |
| 26021325 | Africa | Gambia |
| 26021325 | Africa | Kenya |
| 26021325 | Africa | Nigeria |
| 26021325 | Africa | Nigeria |
| 26021325 | Africa | Sierra Leone |
| 26244421 | Africa | South African |
| 26757134 | Africa | Egyptian |

New biogeographical groups

- Based on analysis of data from HGDP and 1000 Genomes
- Geographical clustering pattern – greatest predictor of human genetic variation
- It is important to note that classifying individuals and communities into a few distinct groups with defined boundaries conflicts with our understanding of human variation, history, and social/cultural identities.
- **As a result, we respectfully present these groups as a tool to represent broad differences in frequencies of pharmacogenetic variation rather than as a classification of human diversity.**

New biogeographical groups

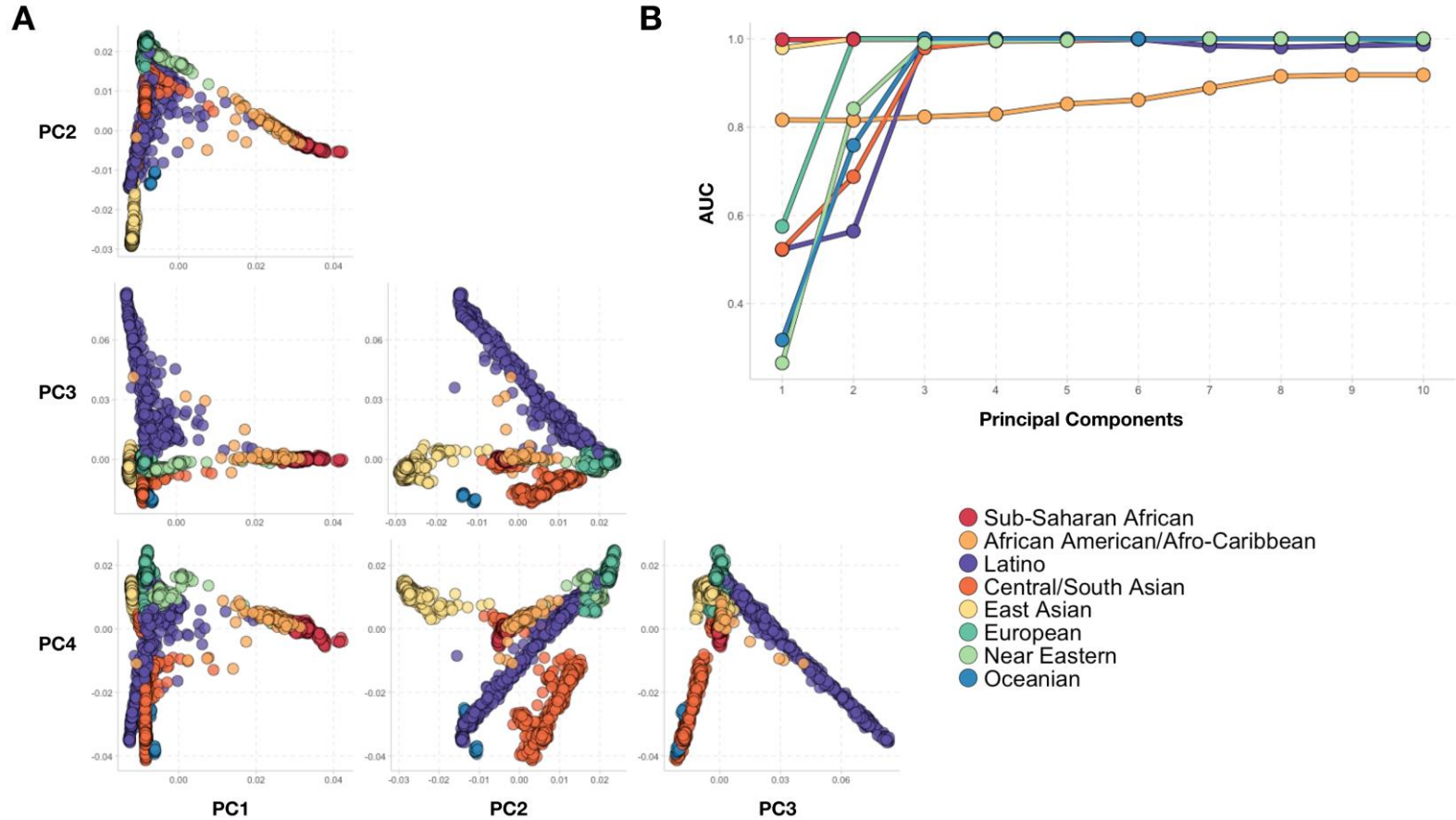
- Seven geographical groups:



- Two admixed groups:
 - African American/Afro-Caribbean (AAC)
 - Latino (LAT)

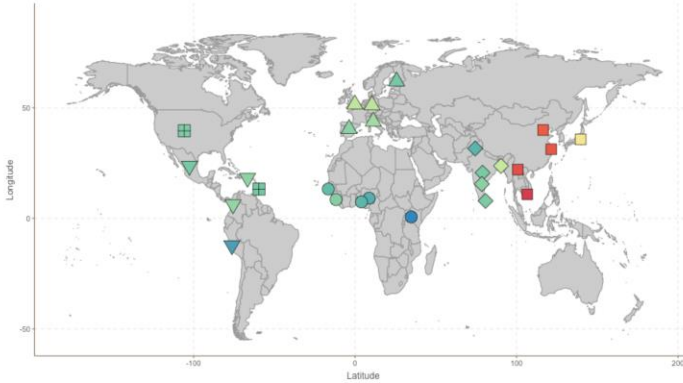


New biogeographical groups

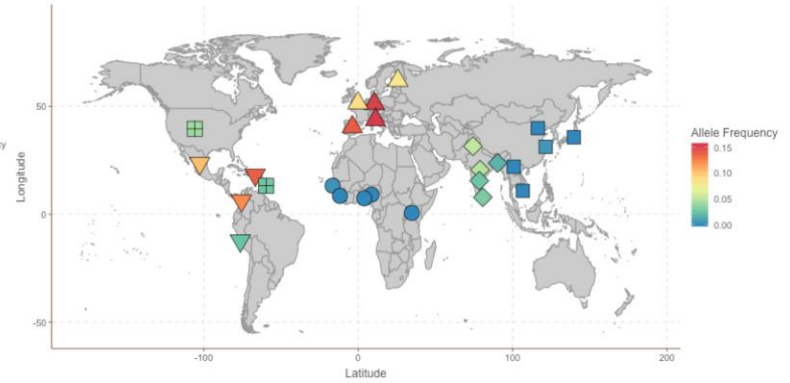


New biogeographical groups

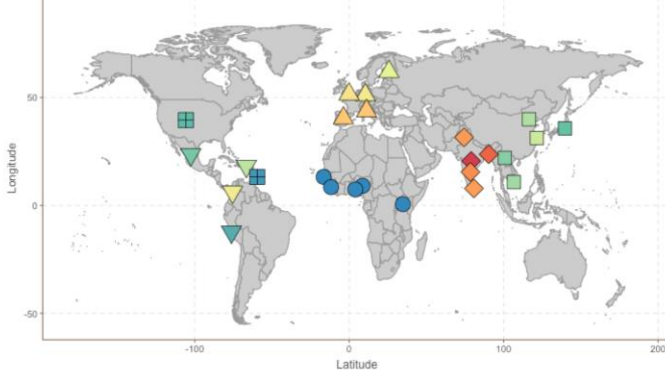
A. *CYP2D6*10* (rs1065852)



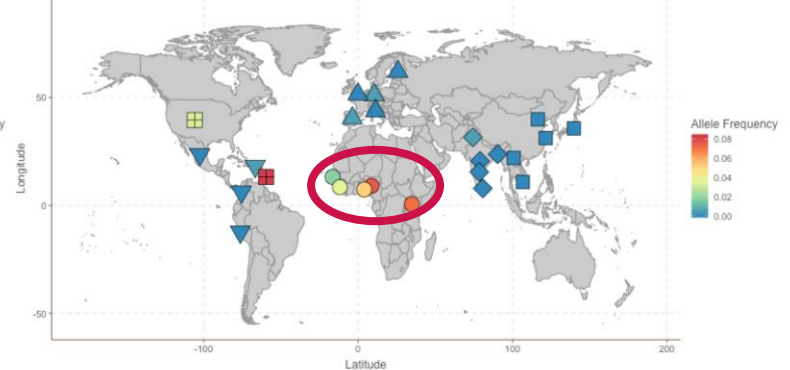
B. *CYP2C9*2* (rs1799853)



C. *CYP2C9*3* (rs1057910)



D. *CYP2C9*8* (rs7900194)



- | | | |
|---------------------|---------------------|---------------------------------|
| East Asian | European | Latino |
| Sub-Saharan African | Central/South Asian | African American/Afro-Caribbean |

Conclusion

- New grouping system represents a more consistent, evidence-based method of illustrating global allele frequencies.
- Now in use at PharmGKB. Recommended as the standard grouping mechanism for population pharmacogenomic studies.
- Need to record detailed self-reported race and ethnicity of study participants.
- **These groups are intended for use in pharmacogenomic research only and not for guiding implementation of pharmacogenomics in the clinic.**



CPIC allele frequency tables

| | A | B | C | D | E | F | G | H | I | J |
|----|--|---------------------------------|--|---|--|------------------------------------|---|----------------------------------|----------------------------------|---|
| 1 | Frequencies^a of CYP2C9 alleles in major race/ethnic groups^b | | | | | | | | | |
| 2 | CYP2C9 allele^c | African Allele Frequency | African American Allele Frequency | Caucasian (European + North American) Allele Frequency | Middle Eastern Allele Frequency | East Asian Allele Frequency | South/Central Asian Allele Frequency | Americas Allele Frequency | Oceanian Allele Frequency | |
| 3 | *1 ^d | 71.30 | 86.88 | 78.54 | 77.24 | 90.15 | 82.80 | 89.96 | 92.21 | |
| 4 | *2 | 3.84 | 2.34 | 12.70 | 13.02 | 0.41 | 6.63 | 4.52 | 0.77 | |
| 5 | *3 | 2.51 | 1.38 | 8.29 | 8.92 | 3.82 | 10.37 | 3.41 | 7.02 | |
| 6 | *4 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| 7 | *5 | 1.12 | 1.37 | 0.07 | 0.08 | 0.00 | 0.00 | 0.45 | 0.00 | |
| 8 | *6 | 1.07 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | |
| 9 | *7 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| 10 | *8 | 5.68 | 5.82 | 0.18 | 0.33 | 0.53 | 0.10 | 1.40 | 0.00 | |
| 11 | *9 | 10.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| 12 | *10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | |
| 13 | *11 | 2.63 | 1.37 | 0.23 | 0.00 | 0.07 | 0.10 | 0.21 | 0.00 | |
| 14 | *12 | 0.00 | 0.00 | 0.00 | 0.10 | 0.13 | 0.00 | 0.00 | 0.00 | |

| | A | B | C | D | E | F | G | H | I | J |
|----|---|---|----------------------------------|---|------------------------------------|----------------------------------|--------------------------------|--------------------------------------|----------------------------------|--|
| 1 | Frequencies^a of CYP2C9 alleles in major population groups^b | | | | | | | | | |
| 2 | CYP2C9 allele^c | African American/Afro-Caribbean Allele Frequency | American Allele Frequency | Central/South Asian Allele Frequency | East Asian Allele Frequency | European Allele Frequency | Latino Allele Frequency | Near Eastern Allele Frequency | Oceanian Allele Frequency | Sub-Saharan Africa Allele Frequency |
| 3 | *1 ^d | 86.88 | 92.85 | 82.48 | 89.67 | 78.57 | 84.49 | 76.44 | 96.45 | 75.47 |
| 4 | *2 | 2.34 | 1.87 | 6.79 | 0.41 | 12.70 | 8.90 | 12.50 | 2.35 | 1.67 |
| 5 | *3 | 1.38 | 2.64 | 10.54 | 4.31 | 8.26 | 4.75 | 8.47 | 1.20 | 0.90 |
| 6 | *4 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.50 | 0.00 | 0.00 |
| 7 | *5 | 1.37 | 0.08 | 0.00 | 0.00 | 0.07 | 0.82 | 0.23 | 0.00 | 1.12 |
| 8 | *6 | 0.85 | 0.00 | 0.00 | 0.00 | 0.00 | 0.11 | 0.00 | 0.00 | 1.07 |
| 9 | *7 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 10 | *8 | 5.82 | 2.44 | 0.10 | 0.53 | 0.18 | 0.66 | 0.45 | 0.00 | 6.29 |
| 11 | *9 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 10.85 |
| 12 | *10 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 13 | *11 | 1.37 | 0.12 | 0.10 | 0.07 | 0.23 | 0.28 | 0.00 | 0.00 | 2.63 |
| 14 | *12 | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 | 0.00 | 0.10 | 0.00 | 0.00 |

Acknowledgements

- PharmGKB
 - Teri Klein
 - Russ Altman
 - Michelle Whirl-Carrillo
 - Katrin Sangkuhl
 - Li Gong
 - Julia Barbarino
 - Caroline Thorn
 - Ryan Whaley
 - Mark Woon
 - Jill Robinson
 - Bonnie Kwong
- Stanford University
 - Carlos Bustamante
 - Genevieve Wojcik
 - Alice Popejoy
- University of Colorado
 - Chris Gignoux
- University of Washington
 - Alison Fohner

Summary <https://www.pharmgkb.org/page/biogeographicalGroups>

Pre-print <https://www.biorxiv.org/content/early/2018/10/11/384016>