

A comparison of self-reported ethnicity and genetic ancestry

Kristjan Eerik Kaseniit, MEng; Christine Lo, PhD; Shivani Nazareth, MS CGC; Kenny K. Wong, MS CGC; Imran S. Haque, PhD

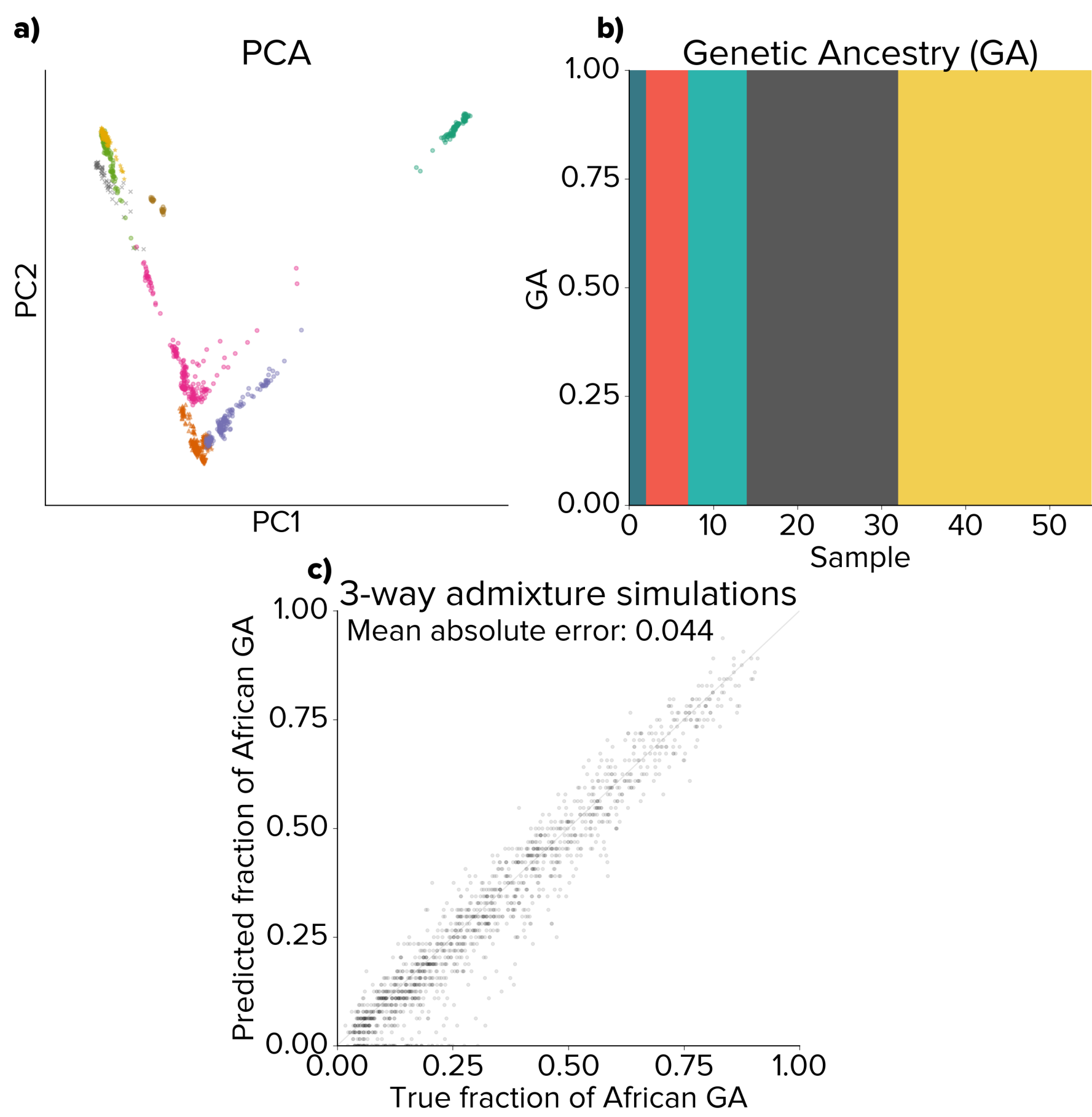
South San Francisco, California

Introduction

Current medical guidelines dictate the use of self-reported ethnicity (**SRE**) in determining disease selection for carrier screening. However, prior studies in personal genetics have noted that self-reported ethnicity is not fully concordant with genetic ancestry (**GA**) and that admixture is pervasive¹. We sought to explore the discordance between self-reported and true genetic ancestry in an expanded carrier screening population in the US.

Methods

Next-generation DNA sequencing was performed for patients taking the Counsyl Family Prep Screen test who self-reported ethnicity choosing from 15 options for ethnic categorization, including one for mixed ethnicity. To compute genetic ancestry, we adapted existing algorithms² for use with off-target next-generation sequencing (NGS) reads that provide ultra-low coverage across the genome (<0.01x).



(a) Principal component analysis (PCA) of genetic data separates populations in a reference panel^{3,4}. (b) Genetic ancestry fractions (stacked barplot) determined from PCA data. (c) Results of determining GA for simulated admixtures of African, Eurasian, Native American reference samples.

Results for N = 119,930

Our study population is among the largest and most diverse reported so far^{1,5}, with 25% of the 119,930 patients self-reporting as Other/Mixed Caucasian, 23% as non-European and 31% as European. We observed a wide range of admixture in the patients choosing to self-report as Hispanic.

In a clinical setting, ethnicity is commonly unreported/unknown or potentially misreported, for example among South/South East/East Asians. Of the patients with unreported or unknown ancestry, nearly one sixth had genetic ancestry that would have prompted a clinician following ethnicity-based screening guidelines to recommend additional tests. Low levels of unreported Ashkenazi Jewish ancestry are observed in self-reported Europeans, as well as in patients who did not report their ethnicity.

“Other / Mixed Caucasian” is chosen mostly by people of European descent

97.6% of these patients have at least 50% European or South Asian genetic ancestry.

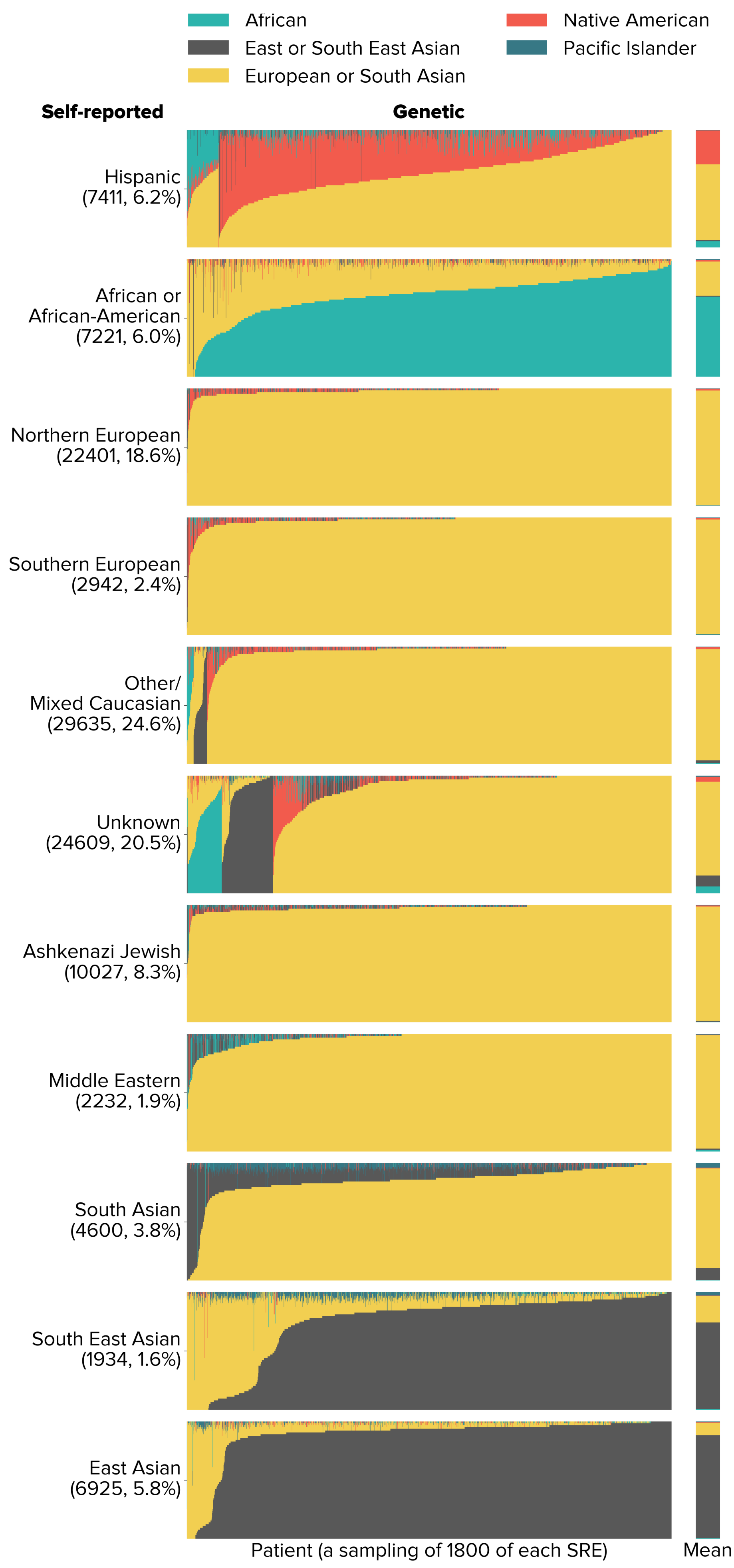
Patients with unreported SRE are representative of the general patient population

At least 25% of GA	% of “Unknown” group
Native American	6%
African	8%
East Asian	10%
European / South Asian	88%

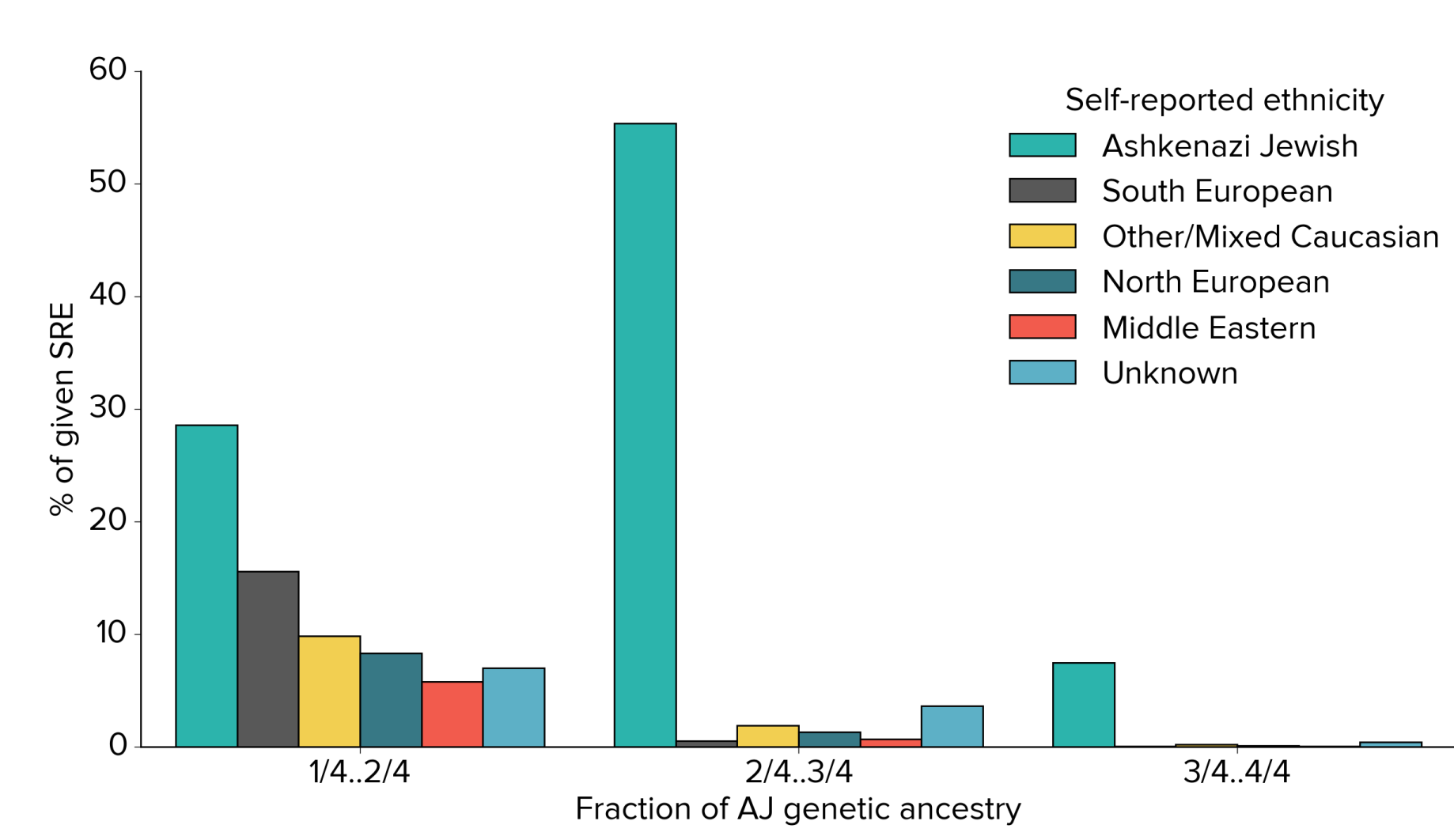
The South Asian, South East Asian, and East Asian self-reporting categories seem to not have hard boundaries

SRE GA ≥25%	South Asian	East Asian
South Asian	98%	4%
South East Asian	22%	86%
East Asian	8%	94%

We observed a remarkable fraction of patients who self-reported in one of the Asian categories, but had nearly 100% of their GA from a different Asian population. The relatively abrupt transitions from South Asian to South East or East Asian genetic ancestry suggests misreporting rather than admixture, while some patients with one parent or grandparent of a different Asian population can also be observed.

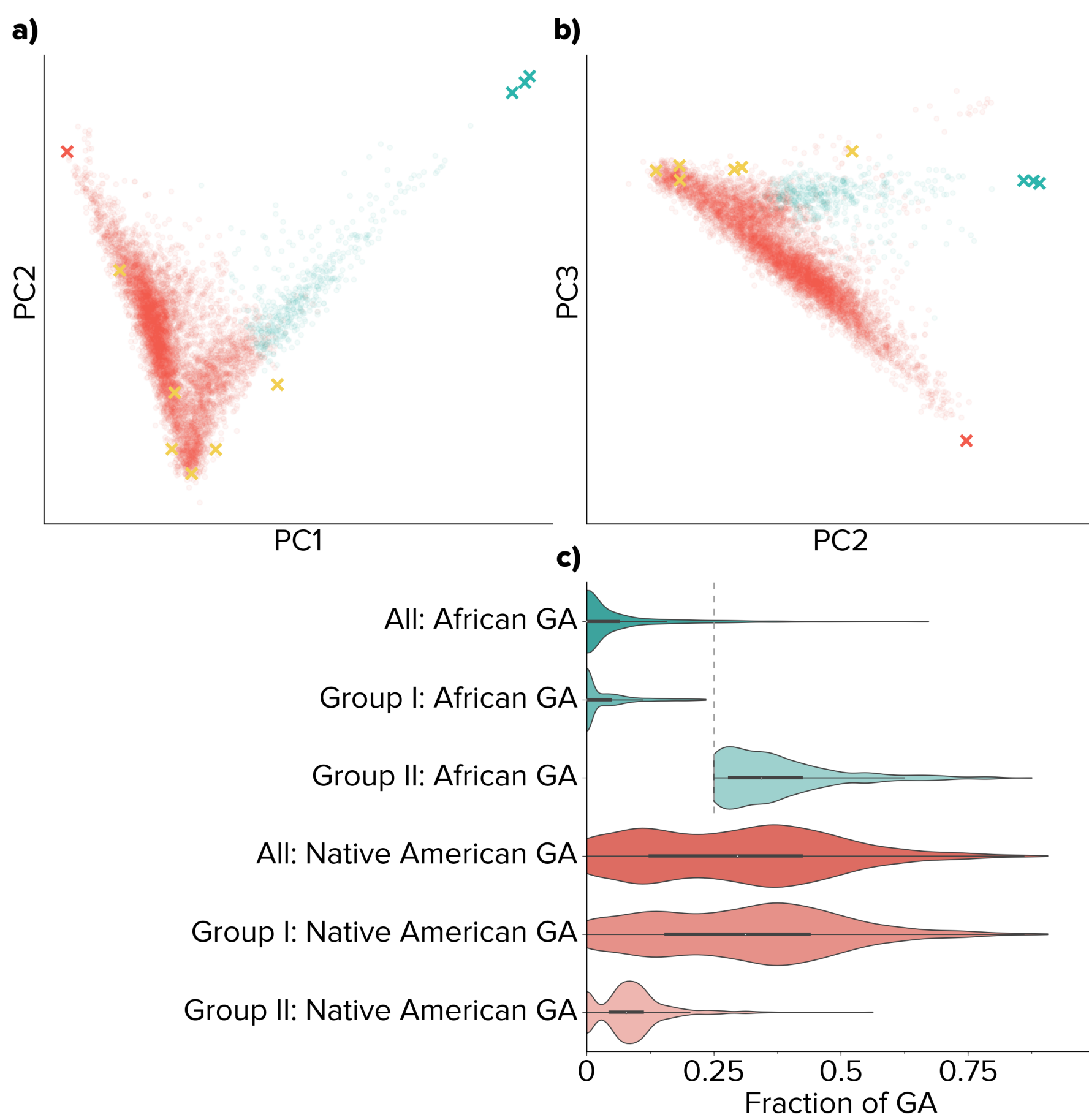


Patients often do not report medically relevant ancestry



Current guidelines from ACOG and ACMG recommend ethnicity-based screening panels, the biggest of which is for the Ashkenazi Jewish (AJ)⁶. We observed that between 6.5% (Middle Easterners) and 16% (South Europeans) of largely European-descent patients who did not self-re port as AJ were predicted to have at least 25% AJ genetic ancestry.

Patients self-reporting as Hispanic have diverse genetic ancestry, even within subgroups



	African GA cutoff	% of Hispanics
Group I	<1/4	94%
Group II	≥1/4	6%

Conclusions

- Genetic ancestry can be determined using ultra-low coverage next-generation sequencing
- A wide range of genetic ancestry is observed
- Genetic ancestry can be more informative than self-reported ethnicity
- Ethnicity can sometimes be mis- or un-reported with implications for screening guidelines and genetic counselling

We observed that self-reported ethnicity is a good predictor of genetic ancestry in only some populations, while in others it can mask the presence or range of true genetic ancestry. These results can inform medical questionnaire design (e.g. specifying examples of Asian populations to reduce misreporting; offering more options for Hispanics, e.g. White/Black Hispanic), and inform carrier screening panel and guideline design.

REFERENCES 1. Bryc, K., et al. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics*, 96(1), 37–53. 2. Wang, C., et al. (2015). Improved Ancestry Estimation for both Genotyping and Sequencing Data using Projection Procrustes Analysis and Genotype Imputation. *The American Journal of Human Genetics*, 926–937. 3. Li, JZ, et al. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 319(5866), 1100–4. 4. Bray SM, et al. (2010). Signatures of founder effects, admixture, and selection in the Ashkenazi Jewish population. *Proc Natl Acad Sci USA* 107(37):16222-7. 5. Banda, Y, et al. (2015). Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics*, 200(4), 1285–95. 6. Gross, S. J., et al. (2008). Carrier screening in individuals of Ashkenazi Jewish descent for the Professional Practice and Guidelines Committee. *Genetics in Medicine*, 10(1), 54–56.