**Global reference panel version 2**

The reference panel included HGDP (PMID: 32193295) and 1000 Genomes Project Phase 3 (1KGP) (PMID: 26432245) populations. The combined HGDP+1KGP genotype callset was obtained from GnomAD version 3.1.2 ([https://gnomAD.broadinstitute.org/downloads#v3-hgdp-1kg](https://gnomad.broadinstitute.org/downloads#v3-hgdp-1kg)). Putative ancestral population origins (i.e. superpopulations and regional populations) for all reference subjects were obtained from the International Genome Sample Resource (IGSR) (PMID: 31584097) (igsr\_samples.tsv). Admixed subjects were removed from usage as references, specifically we removed from the 1KGP the African Caribbean (ACB), African Ancestry SW (ASW), Columbian (CLM), Mexican Ancestry (MXL), Peruvian (PEL), Puerto Rican (PUR) populations and from the HGDP the Hazara, Mozabite, and Uygur populations, as well as one individual who was annotated to both IBS and MSL populations. Within superpopulations, closely related samples (2nd degree relative or closer) were removed using KING using the --unrelated --degree 2 options. The resulting reference panel included 2,767 individuals from 70 populations (table X). The reference panel genotype data was then filtered to bi-allelic, strand unambiguous SNPs that were also found in the test genotype data using BCFtools (PMID: 19505943). SNPs were removed if strand corrected allele codes did not match between reference and test datasets. Principal components were calculated in the reference data using SMARTPCA (PMID: 21448230). The reference data included 7 superpopulations: Africa (AFR), Americas (AMR), Central-South Asia (CSA), East Asia (EAS), Europe (EUR), Oceania (OCE), and South-West Asia (SWA). The SWA and EUR superpopulation labels were collapsed together in order to form K=6 clusters. A K=6 cluster reference panel was created using SNPweights (PMID: 23539302), using the PCs and superpopulation membership data as inputs.