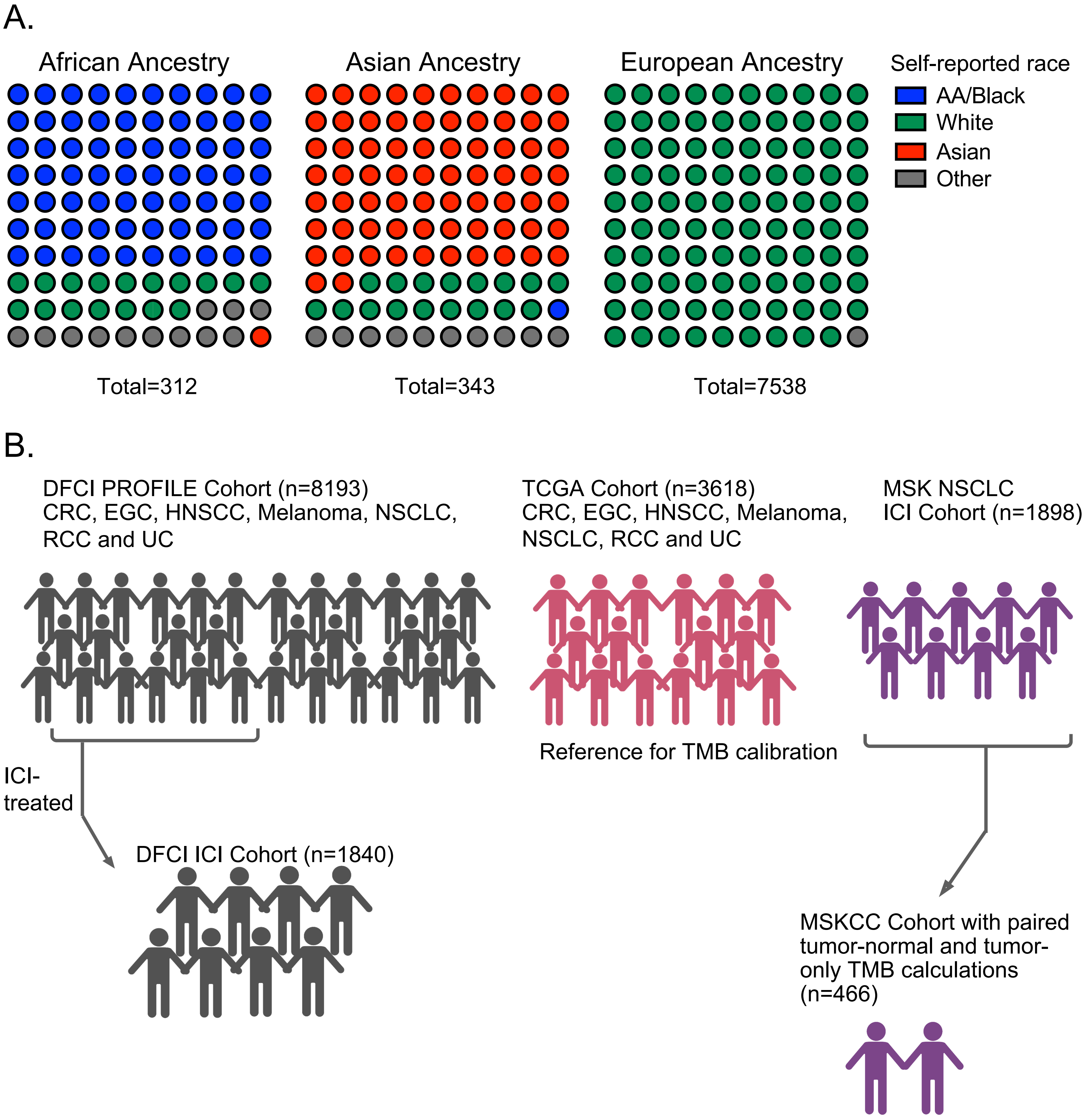


Figure S1: Related to Figure 1 and STAR Methods. Study Cohort. A. Concordance between self-reported race and genetic ancestry among the entire DFCI cohort (n=8193 patients). B. Schematic diagram of cohort study design.

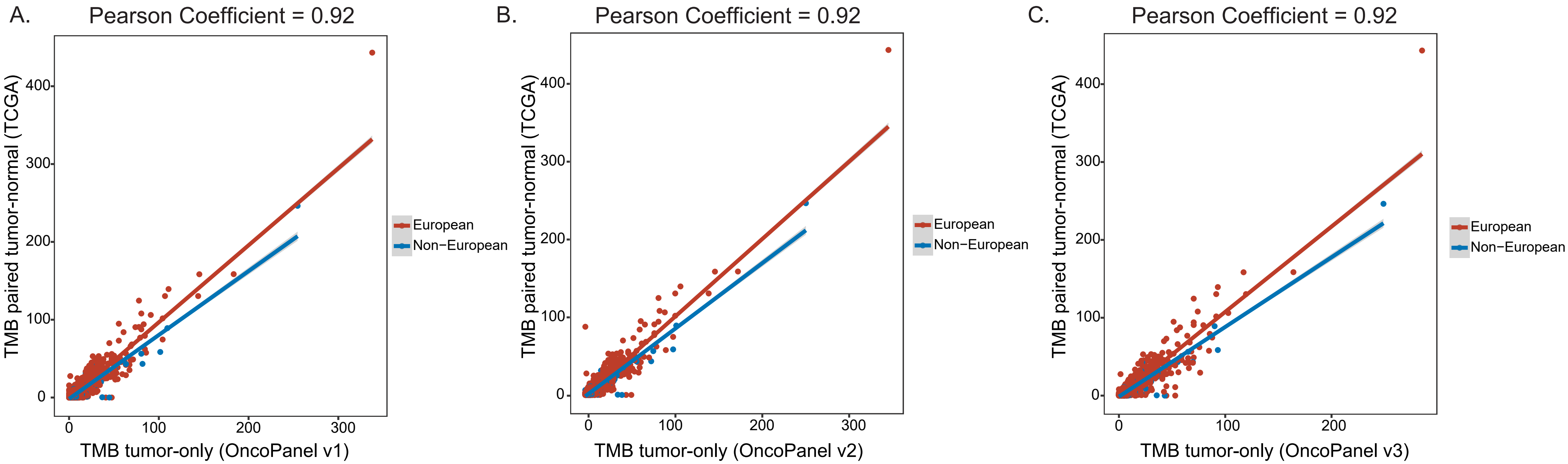


Figure S2: Related to Figure 1 and 2. Ancestral correlations between WES paired tumor-normal TMB in the TCGA cohort and tumor-only TMB using restricted Oncopanel genes. TCGA cohorts included colorectal cancer, esophagogastric cancer, head and neck squamous cell carcinoma, melanoma, non-small cell lung cancer, urothelial carcinoma, and renal cell carcinoma. Paired tumor-normal TMB was calculated using WES data from the TCGA cohorts. Tumor-only TMB was calculated after restricting to A) Oncopanel V1 (n=275), B) V2 (n=300), and C) V3 (n=447) genes.

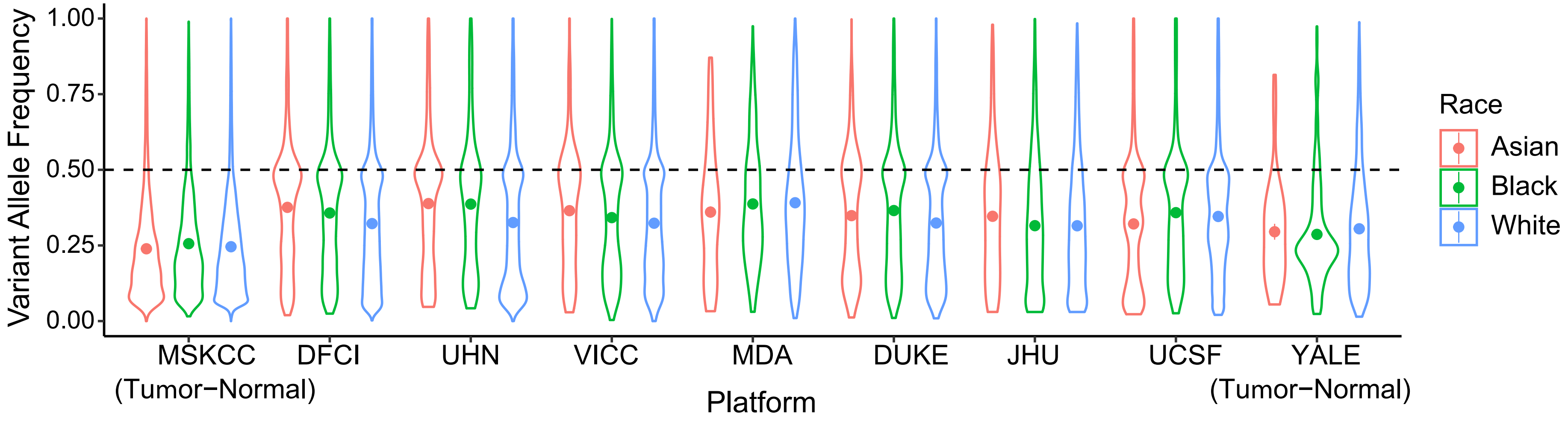


Figure S3: Related to Figure 1 and 2. Distribution of variant allele frequencies (VAF) by self-reported race in Oncopanel and other GENIE sequencing panels. MSK-IMPACT and YALE are paired tumor-normal sequencing panels and other panels including Oncopanel (DFCI) are tumor-only sequencing panels. The dots represent the medians. Horizontal dashed line represents sequencing variants with variant allele fraction of 50%.

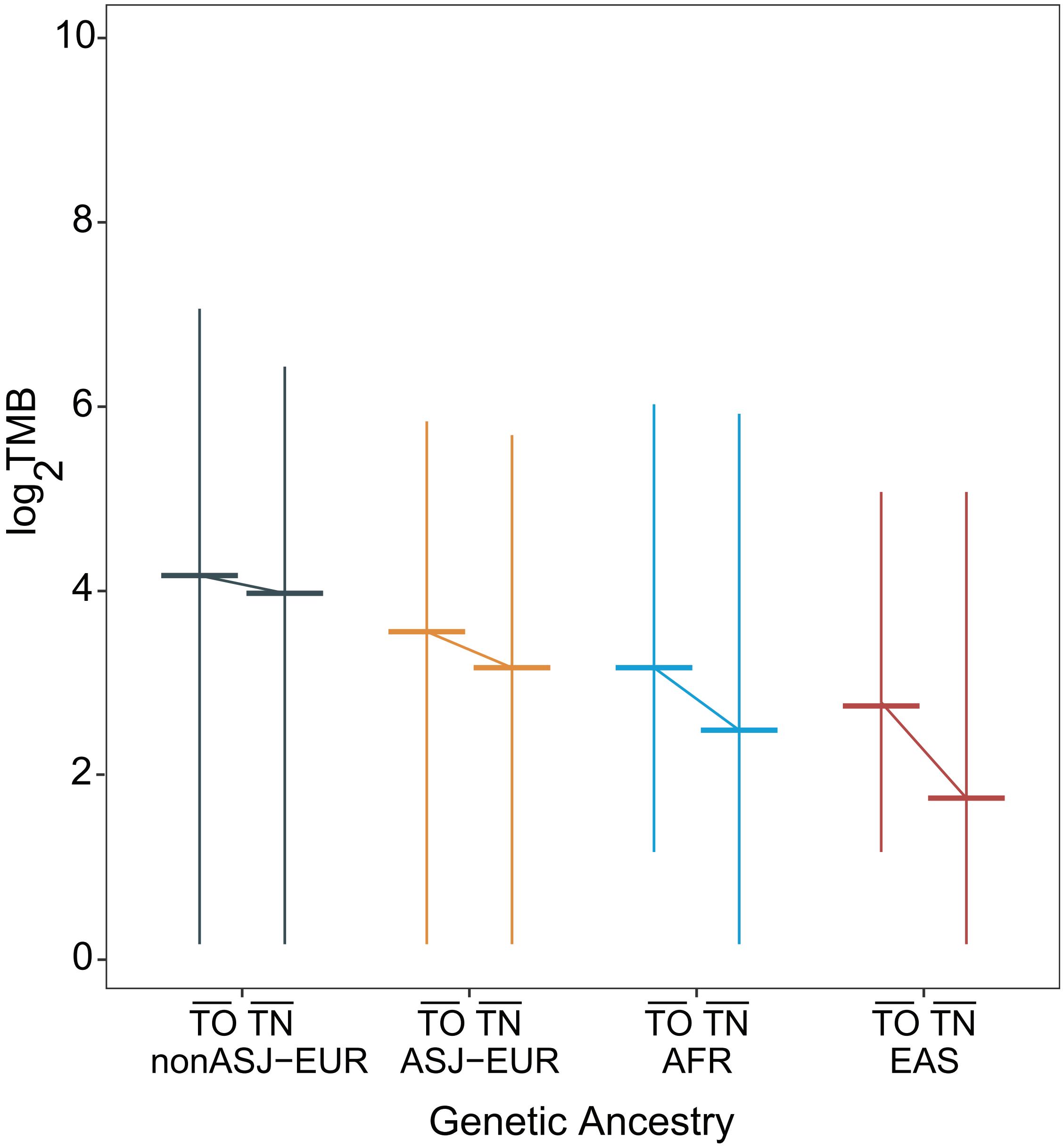
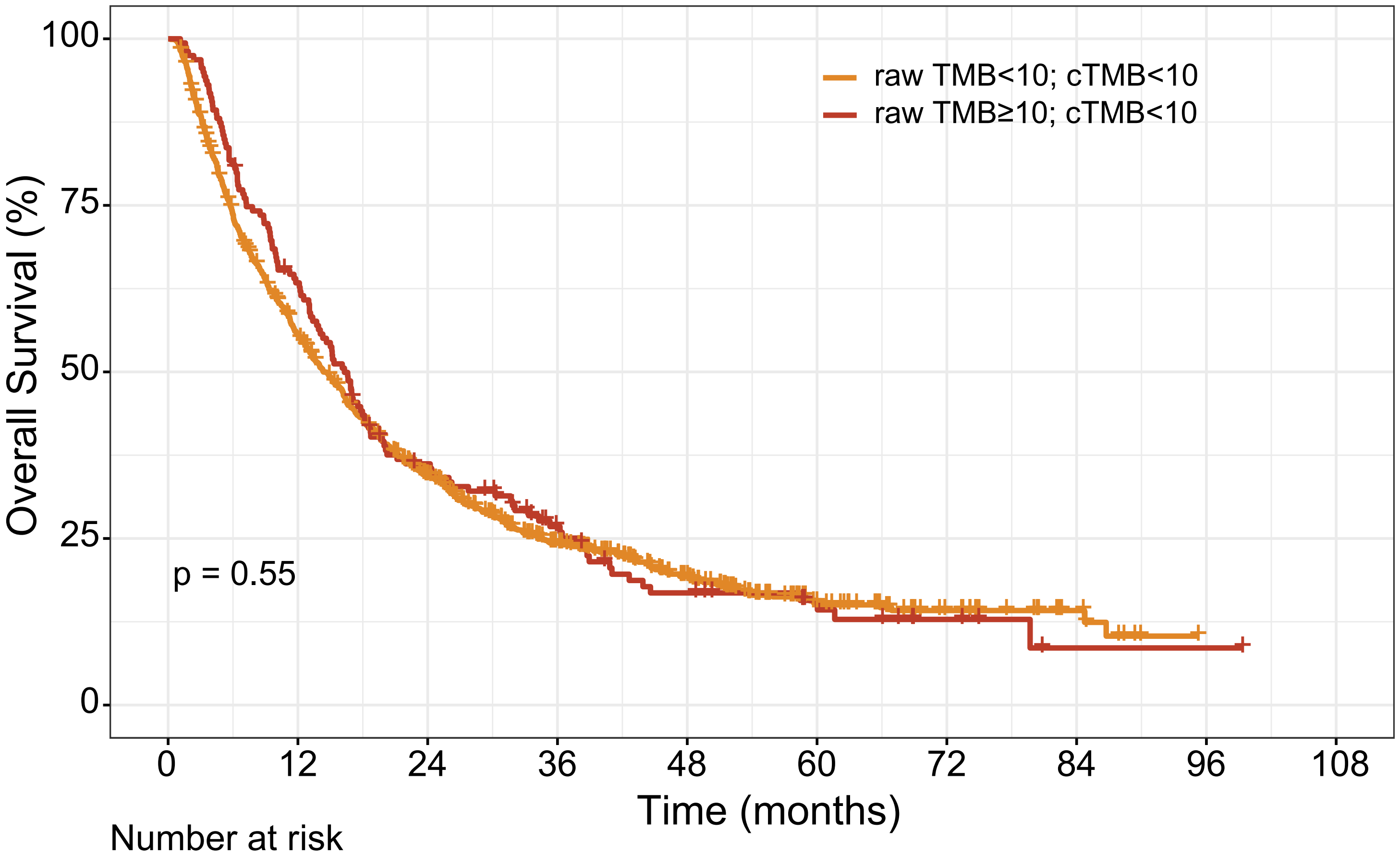


Figure S4: Related to Figure 1. Tumor-only versus tumor matched TMB estimates across ancestries in the MSKCC cohort (n=466 patients). Log2 TMB values across the ancestral populations. The median and 95% CI confidence intervals are shown. NonASJ-EUR: non-Ashkenazi Europeans; ASJ-EUR: Ashkenazi Europeans; AFR: African; EAS: East Asian; TO: Tumor-only; TN: Tumor/Normal.



Number at risk

🟧	1056	566	325	187	109	51	24	9	0	0
🟥	159	99	53	31	18	11	5	1	1	0

Figure S5: Related to Figure 2. Association between false TMB-High (raw TMB≥10; calibrated TMB<10) and true TMB-Low tumors (raw TMB<10; calibrated TMB<10) and overall survival among ICI-treated patients.

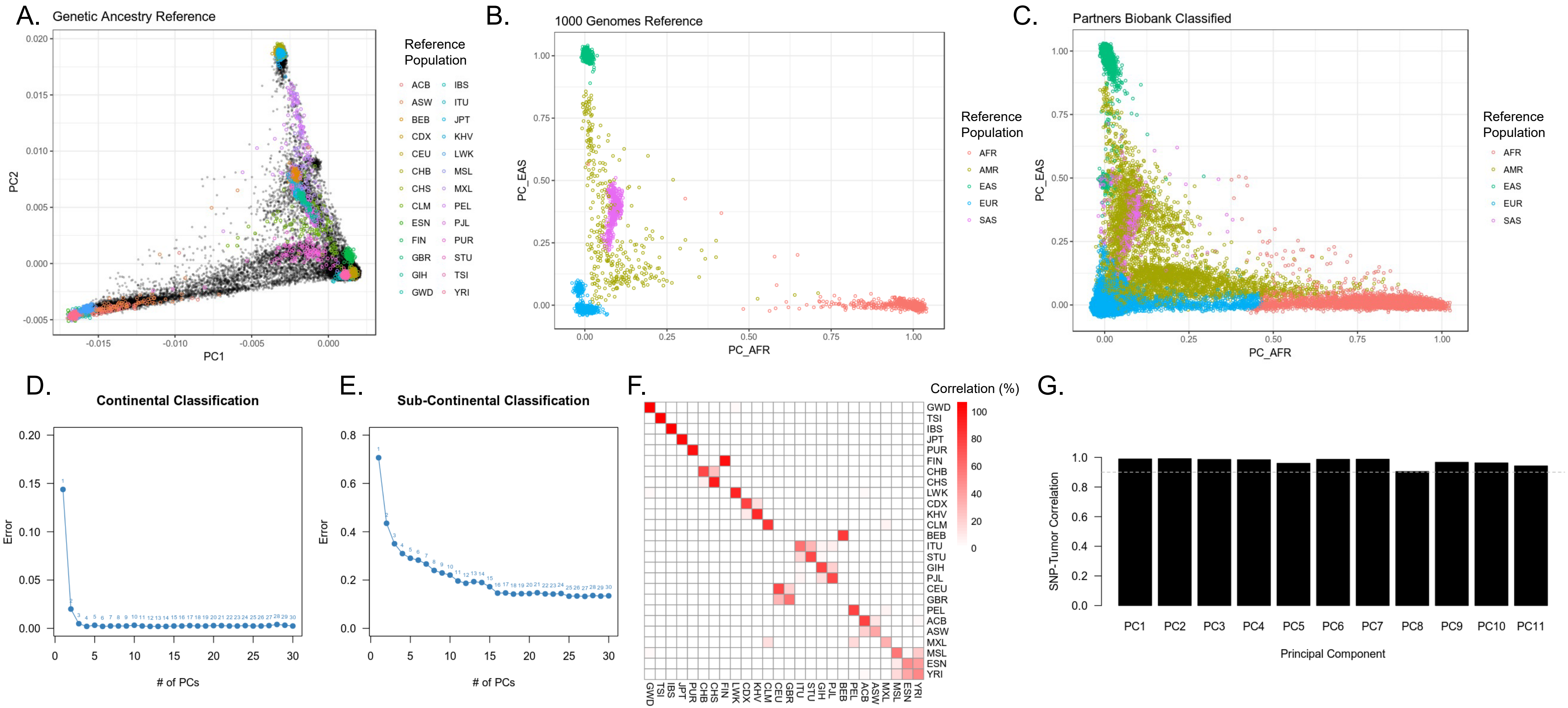


Figure S6: Related to STAR Methods. Inferring Genetic Ancestry from sequencing panels. A. Genetic ancestry inferred in the reference population. Leading two Principal Components (PCs) are shown for each individual (point) along the x-axis (PC1) and y-axis (PC2) color-coded by the 1000 Genomes reference subpopulation. B. 1000 Genomes individuals color coded by their continental population labels. C. Partners Biobank population color coded by their classified continental ancestry trained in the 1000 Genomes reference. For B&C, continental ancestry components are rotated and rescaled to reflect proportion of ancestry for each individual (point) and African (AFR, x-axis) or East Asian (EAS, y-axis) ancestry. D. Principal component (PC) informativeness for population labels in the entire DFCI cohort (n=32,336) with classification error (out-of-bag) for continental population labels in the 1000 Genomes reference data. E. Principal component (PC) informativeness for population labels in the entire DFCI cohort (n=32,336) with classification error (out-of-bag) for fine-scale population (sub-continental) labels in the 1000 Genomes reference data from a Random Forest classifier. For D&E, each point represents the number of PCs included in the classifier for training. F. Heatmap of (out-of-bag) misclassification for the Random Forrest classifier constructed from PC1-PC11 with the 1000 Genomes subcontinental population labels as the response. G. Correlation between the PC inferred from germline SNP array (ground truth) and that projected from tumor imputed data for each PC1-PC11, Related to STAR Methods. ACB: African Caribbean, ASW: African Ancestry in Southwest US; BEB: Bengali in Bangladesh; CDX: Chinese Dai in Xishuangbanna, China; CEU: Utah residents (CEPH) with Northern and Western European ancestry; CHB: Han Chinese in Beijing, China; CHS: Han Chinese South; CLM: Colombian in Medellin, Colombia; ESN: Esan in Nigeria; FIN: Finnish in Finland; GBR: British in England and Scotland; GIH: Gujarati Indians in Houston, TX; GWD: Gambian in Western Division, The Gambia – Mandinka; IBS: Iberian populations in Spain; ITU: Indian Telugu in the UK; JPT: Japanese in Tokyo, Japan; KHV: Kinh in Ho Chi Minh City, Vietnam; LWK: Luhya in Webuye, Kenya; MSL: Mende in Sierra Leone; MXL: Mexican Ancestry in Los Angeles, California; PEL: Peruvian in Lima, Peru; PJL: Punjabi in Lahore, Pakistan; PUR: Puerto Rican in Puerto Rico; STU: Sri Lankan Tamil in the UK; TSI: Toscani in Italia; YRI: Yoruba in Ibadan, Nigeria. AFR: African; AMR: Admixed American; EAS: East Asian; EUR: European; SAS: South Asian

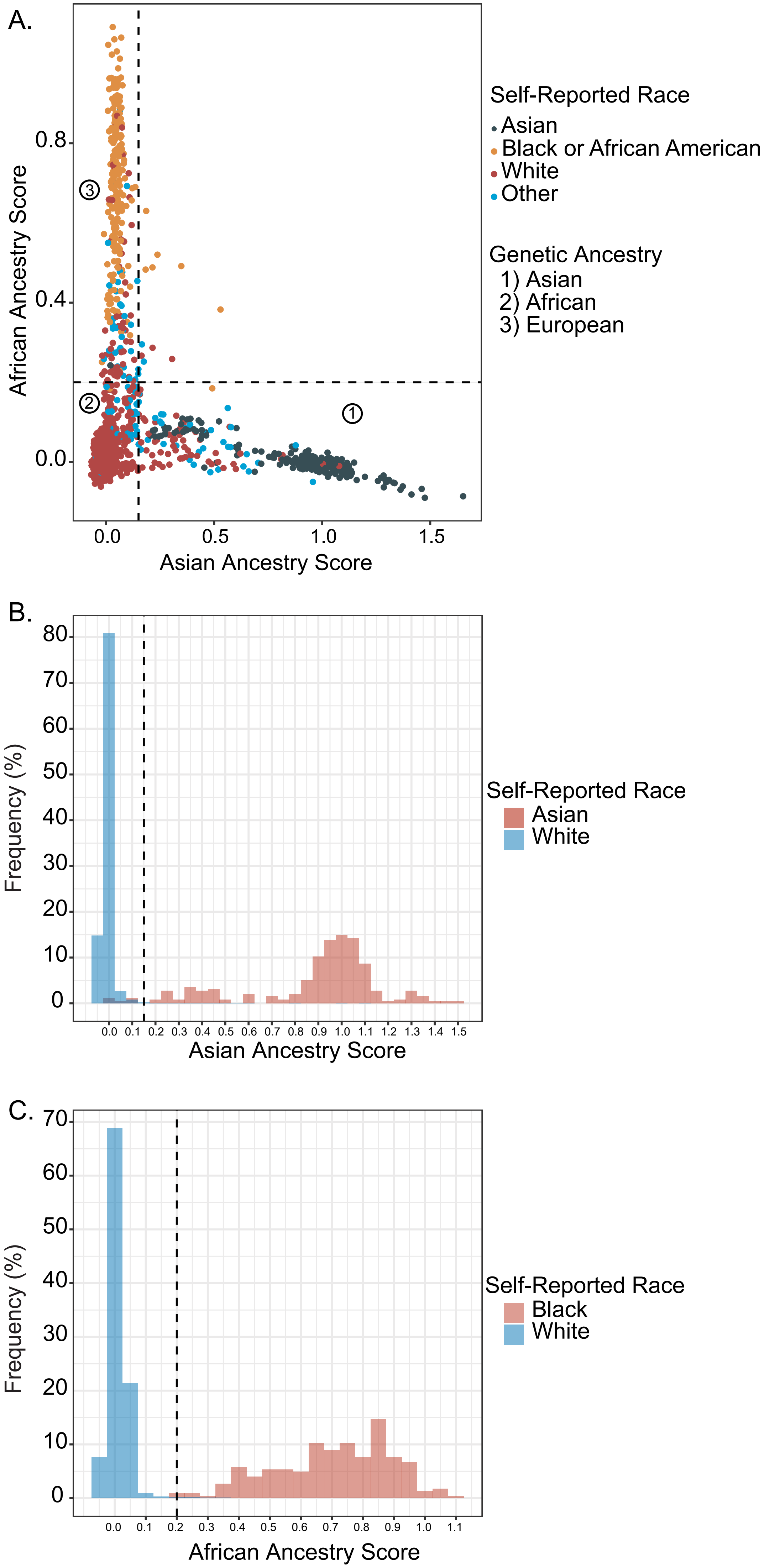

A.
African Ancestry Score
0.8
0.4
0.0
③
②
①
Asian Ancestry Score
0.0
0.5
1.0
1.5
Self-Reported Race
● Asian
● Black or African American
● White
● Other
Genetic Ancestry
1) Asian
2) African
3) European
B.
Frequency (%)
80
70
60
50
40
30
20
10
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5
Asian Ancestry Score
Self-Reported Race
Asian
White
C.
Frequency (%)
70
60
50
40
30
20
10
0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1
African Ancestry Score
Self-Reported Race
Black
White

Figure S7: Related to STAR Methods. Comparison between imputed genetic ancestry and self-reported race. A. Distribution of ICI-treated patients at DFCI (n=1840 patients) by genetic ancestry and self-reported race. B. Self-reported race as a function of Asian ancestry index with a vertical line at 0.15 representing the cut-off above which patients are considered of Asian ancestry. C. Self-reported race as a function of African ancestry index with a vertical line at 0.2 representing the cut-off above which patients are considered of African ancestry.