Comparison of the genomic landscape between primary breast cancer in African American versus Caucasian women and the association of racial differences with tumor recurrence

Keenan, et al

Supplemental Material

The Cancer Genome Atlas Data

Data were downloaded from The Cancer Genome Atlas Data Portal site at https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm. The selected options were: Disease – BRCA, Data Type – All, Center/Platform – All, Batch Number – All, Preservation – All, and Access Tier – All. All other options were left unselected, and the Apply button at the bottom of the page was then clicked.

Clinical and whole exome sequencing (IlluminaGA_DNASeq_curated) data were downloaded on April 28, 2014. On the second web page, after the Apply button was clicked as described above, the Clinical – Biotab and Somatic Mutations – WUSM Curated Mutation Calling columns were selected. The Build Archive button on the left hand side of the page was clicked, and the data were downloaded. Follow-up clinical data files were merged with the original clinical data file.

Clinical and RNA sequencing (RNASeqV2) data were downloaded on December 29, 2014. On the second web page, after the Apply button was clicked as described above, the Clinical – Biotab and RNASeqV2 – UNC IlluminaHiSeq_RNASeqV2 columns were selected. The Build Archive button on the left hand side of the page was clicked, and the data were downloaded.

These instructions give the most recently updated data, which change over time. The data files used in the analyses presented in this manuscript are available online as supplementary data.

Methods for Clinical Variables

Clinical subtype was categorized as human epidermal growth factor receptor 2 positive (HER2+) by fluorescence in situ hybridization ≥ 2.2 and/or immunohistochemistry 3+, hormone receptor positive (HR+) if estrogen and/or progesterone receptor positive by immunohistochemistry without HER2 positivity, or TNBC if known not to be HER2+ or HR+. Pathologic stage was designated by the American Joint Committee on Cancer staging system using tumor, node, and metastasis classifications.[27] Tumor recurrence was identified by the tumor status and follow-up times reported by TCGA in follow-up clinical data files, which were merged with the original clinical data file.[26] Median follow-up time was 24.9 months (interquartile range 15.7-45.4). For time to progression, death was treated as a competing risk event. Patients alive and without progression were censored at the last follow-up time. All participants were enrolled in protocols approved by corresponding institutional review boards.

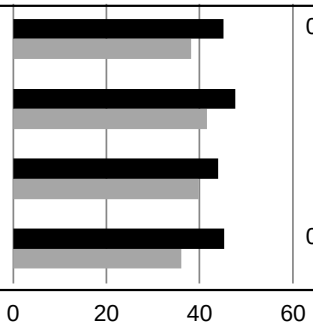
eTable 1. Adjusted racial association of *TP53* and *PIK3CA* mutations by clinical subtype

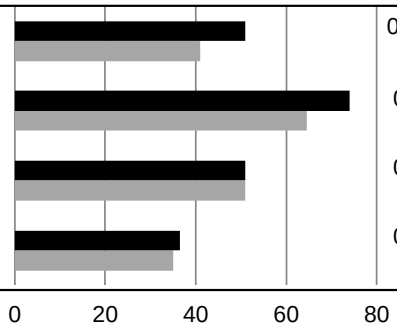
	***TP53* Mutations**		***PIK3CA* Mutations**	
Subtype	**Adj OR (95% CI)**	**p**	**Adj OR (95% CI)**	**p**
Overall	1.90 (1.24, 2.92)	0.003	0.50 (0.30, 0.83)	0.008
TNBC	1.23 (0.49, 3.09)	0.66	0.69 (0.17, 2.78)	0.60
HER2	0.89 (0.30, 2.63)	0.83	0.81 (0.24, 2.79)	0.74
HR	1.31 (0.57, 3.03)	0.53	0.61 (0.30, 1.24)	0.17

Adj, adjusted for age and stage; OR; odds ratio comparing African Americans to Caucasians.

eTable 2. Mutation number per tumor and intratumor heterogeneity by race and subtype

			Mutant Allele Tumor Heterogeneity		**Mutation Number per Tumor**	
Subtype	**Race**	**n**	**Median (IQR)**	**p**	**Median (IQR)**	**p**
Overall	AA	105	45.1 (35, 52)	0.0002	51.0 (34, 91)	0.004
	CA	663	38.2 (29, 47)		41.0 (26, 73)	
TNBC	AA	37	47.7 (37, 57)	0.02	74.0 (43, 113)	0.45
	CA	84	41.6 (32, 51)		64.5 (45, 100)	
HER2	AA	16	44.0 (39, 51)	0.15	51.0 (39, 73)	0.82
	CA	99	39.8 (32, 49)		51.0 (30, 93)	
HR	AA	42	45.2 (30, 51)	0.0001	36.5 (30, 64)	0.02
	CA	395	36.1 (28, 46)		35.0 (24, 64)	





AA, African American; CA, Caucasian American; IQR, interquartile range.
p values calculated by Kruskal-Wallis rank test.

eTable 3. Association of race with intratumor genetic heterogeneity

		Mutant Allele Tumor Heterogeneity
Subtype	**n**	**Beta coefficient[a] (95% CI)**
Overall	768	5.07 (2.41, 7.74)*
Multivariable Overall[b]	672	4.29 (1.47, 7.12)**
TNBC	121	5.92 (0.84, 11.00)
HER2	115	4.09 (-1.54, 9.72)
HR	437	3.00 (-1.22, 7.21)

[a]Beta coefficient comparing African Americans to Caucasians adjusted for age and stage.
[b]Multivariable linear regression model consisting of race, age, stage, TNBC, HER2, and HR.
*p = 0.02, **p = 0.003.

eTable 4. Age and stage adjusted hazard ratios for progression with all available data

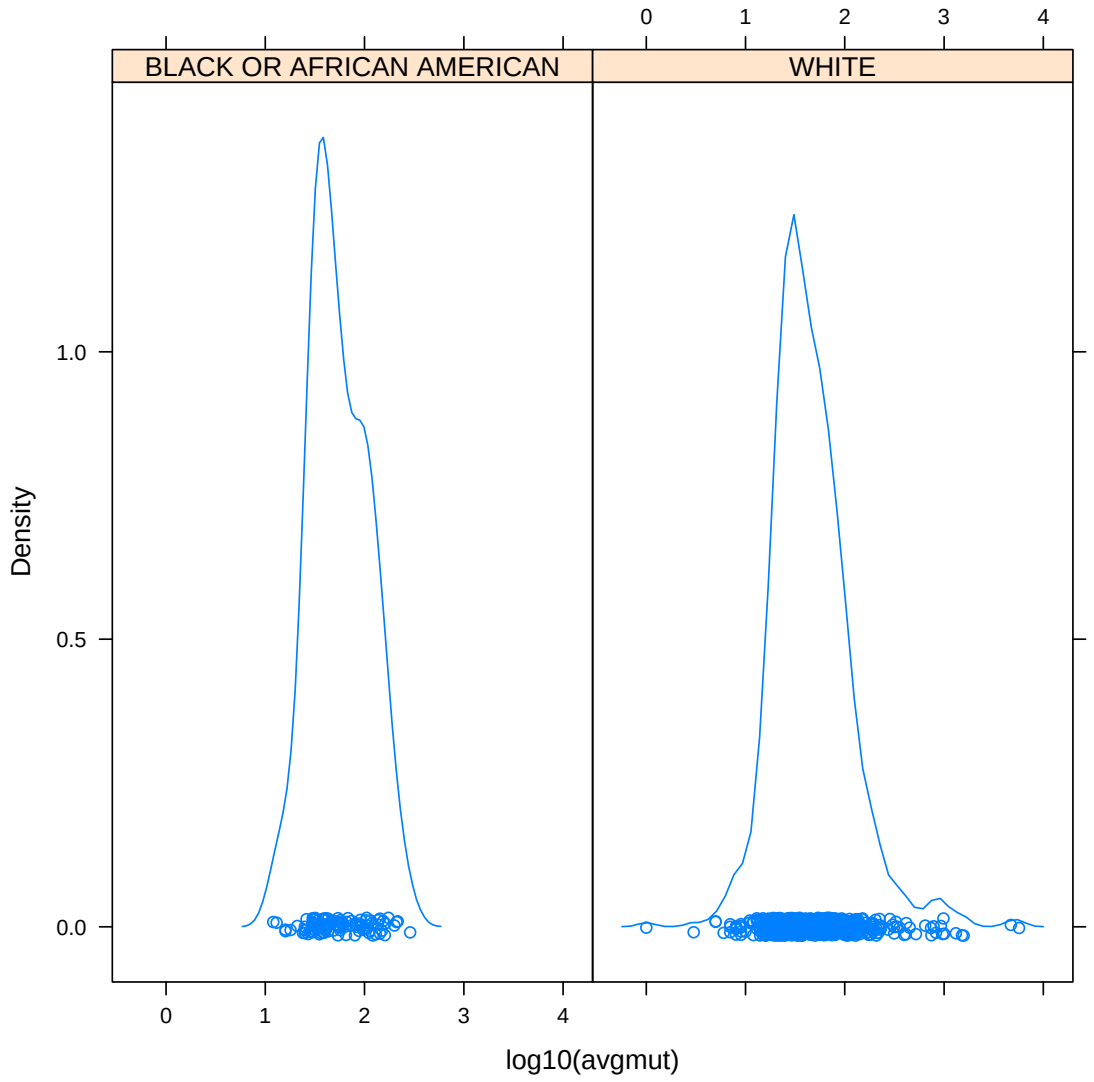
Characteristic	Adj. HR	95% CI	p	N	n
AA vs. CA race	2.04	(1.04, 3.98)	0.04	752	40
AA vs. CA race also adjusted for[a]					
MATH	2.26	(1.17, 4.39)	0.02	667	40
TP53 mutation	1.93	(0.95, 3.90)	0.07	752	40
Basal PAM50 subtype	1.49	(0.73, 3.00)	0.27	750	40
TNBC	1.31	(0.62, 2.78)	0.48	705	34
PAM50 subtype	1.37	(0.69, 2.74)	0.37	750	40
MATH	1.01	(0.99, 1.03)	0.47	667	40
TP53 mutation	1.45	(0.77, 2.76)	0.25	752	40
Basal PAM50 subtype	3.65	(1.87, 7.12)	<0.001	750	40
TNBC	4.11	(1.92, 8.79)	<0.001	705	34
PAM50 subtype					
Luminal A	*Ref*	*Ref*	*Ref*	750	40
Luminal B	2.19	(0.82, 5.80)	0.12	750	40
HER2	1.50	(0.41, 5.46)	0.54	750	40
Basal-like[b]	4.97	(2.00, 12.40)	0.001	750	40
Normal-like	0.41	(0.05, 3.08)	0.39	750	40

Adj, age and stage adjusted; AA, African American; CA, Caucasian American; MATH, mutant allele tumor heterogeneity; N, number of participants; n, number of tumor recurrences; Ref, reference.

[a]Hazard ratio for tumor recurrence comparing African Americans to Caucasians adjusted for age, stage, and separately MATH, *TP53* mutation, basal vs. non-basal PAM50 subtype, TNBC, PAM50 subtype (with indicator variables), or a multivariable model including all of these variables.

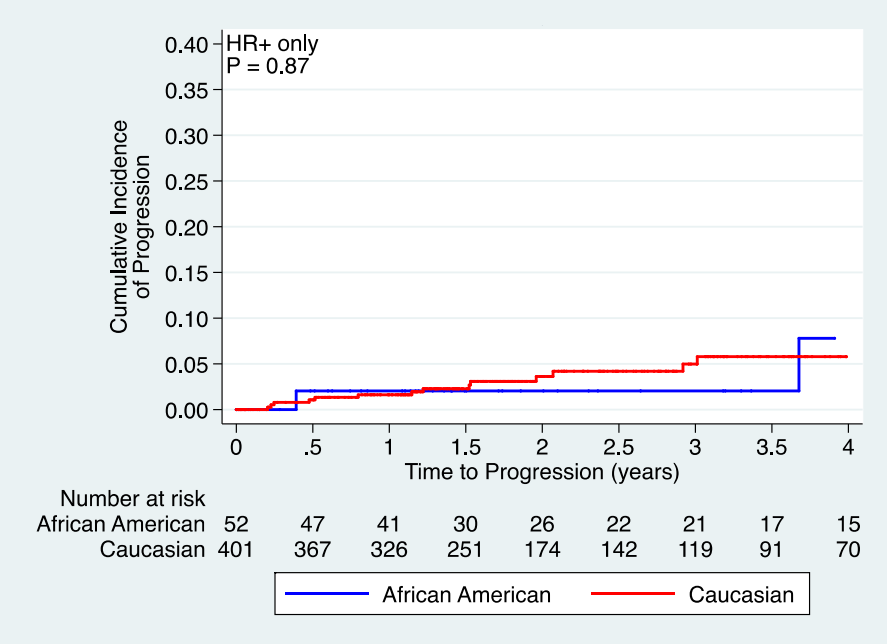
[b]Hazard ratio comparing PAM50 basal subtype to PAM50 luminal A subtype.

eFigure 1. Density plots of mutation number per tumor by race



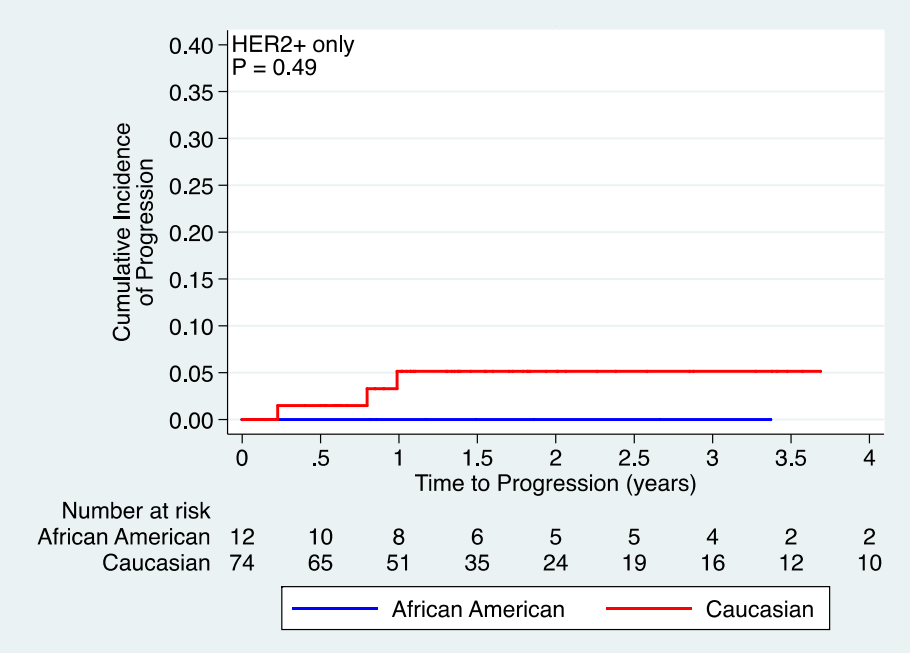
Density plots on a log scale show different racial patterns for mutation number per tumor, such as a sub-population of 25 Caucasian tumors with greater than 300 mutations.

eFigure 2. Cumulative incidence of progression for HR+ tumors by race



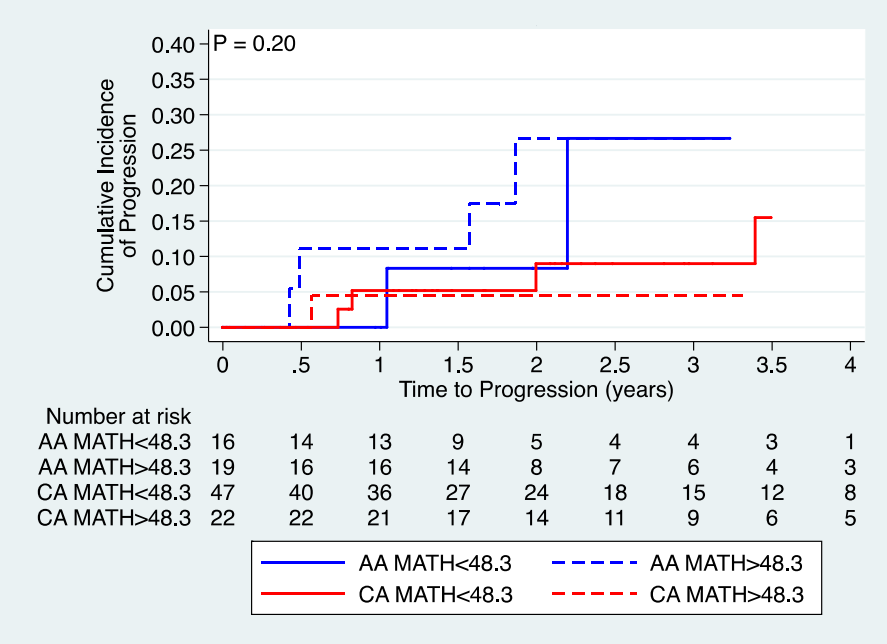
Cumulative incidence of tumor recurrence among HR+ tumors stratified by race.
HR+, hormone receptor positive.

eFigure 3. Cumulative incidence of progression for HER2+ tumors by race



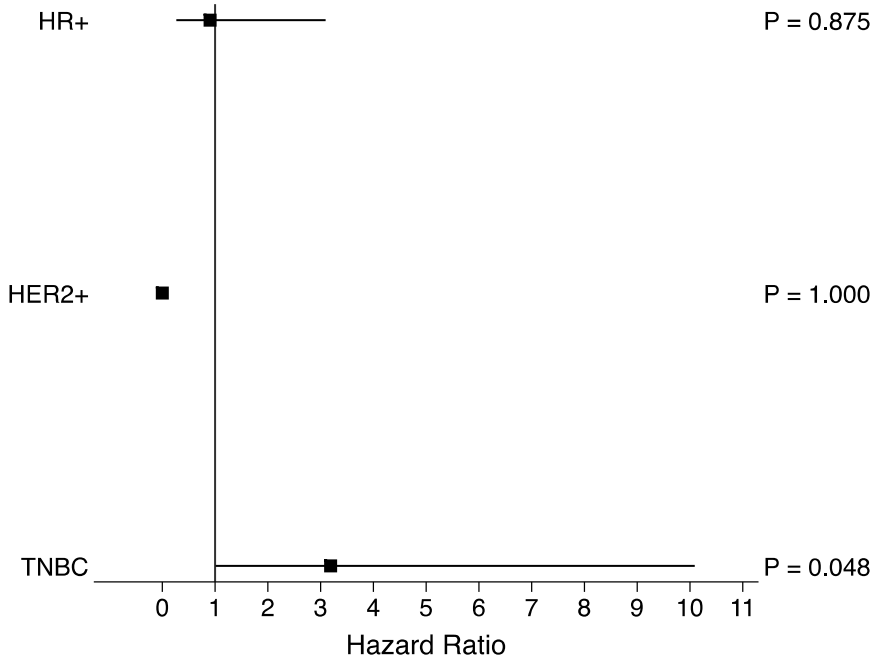
Cumulative incidence of tumor recurrence among HER+ tumors stratified by race.
HER2+, human epidermal growth factor receptor 2 positive.

eFigure 4. Cumulative incidence of progression by race and MATH within TNBC



Cumulative incidence of tumor recurrence among TNBC stratified by race and MATH < or > 75th percentile of 48.3. AA, African American; CA, Caucasian American; MATH, mutant allele tumor heterogeneity; TNBC, triple negative breast cancer.

eFigure 5. Forest plot for progression by race within each clinical subtype



Forest plot of unadjusted hazard ratios for tumor recurrence comparing African Americans to Caucasians for HR+ tumors, HER2+ tumors, and TNBC. The hazard ratio could not be estimated for HER2+ tumors as there were no tumor recurrences among African Americans with HER2+ tumors.