



Immune microenvironment of triple-negative breast cancer in African-American and Caucasian women

Tess O'Meara¹ · Anton Safonov² · David Casadevall^{1,3} · Tao Qing¹ · Andrea Silber¹ · Brigid Killelea¹ · Christos Hatzis¹ · Lajos Pusztai¹

Received: 23 January 2019 / Accepted: 29 January 2019 / Published online: 6 February 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Purpose African-American (AA) patients with triple-negative breast cancer (TNBC) are less likely to achieve pathologic complete response from neoadjuvant chemotherapy and have poorer prognosis than Caucasian patients with TNBC, suggesting potential biological differences by race. Immune infiltration is the most consistent predictive marker for chemotherapy response and improved prognosis in TNBC. In this study, we test the hypothesis that the immune microenvironment differs between AA and Caucasian patients.

Methods RNA-seq expression data were obtained from The Cancer Genome Atlas (TCGA) database for 162 AA and 697 Caucasian breast cancers. Estrogen receptor (ER)-positive, human epidermal growth factor receptor-2 (HER2)-positive, and TNBC subtypes were included in the analyses. Tumor infiltrating lymphocyte (TIL) counts, immunomodulatory scores, and molecular subtypes were obtained from prior publications for a subset of the TNBC cases. Differences in immune cell distributions and immune functions, measured through gene expression and TIL counts, as well as neoantigen, somatic mutation, amplification, and deletion loads, were compared by race and tumor subtype.

Results Immune metagene analysis demonstrated marginal immune attenuation in AA TNBC relative to Caucasian TNBC that did not reach statistical significance. The distributions of immune cell populations, lymphocyte infiltration, molecular subtypes, and genomic aberrations between AA and Caucasian subtypes were also not significantly different. The MHC1 metagene demonstrated increased expression in AA ER-positive cancers relative to Caucasian ER-positive cancers.

Conclusions This study suggests that the immunological differences between AA and Caucasian breast cancers represented by TCGA data are subtle, if they exist at all. We observed no consistent racial differences in immune gene expression or TIL counts in TNBC by race. However, this study cannot rule out small differences in immune cell subtype distribution and activity status that may not be apparent in bulk RNA analysis.

Keywords Triple-negative breast cancer · Race · Immune microenvironment · Immunotherapy · Genetics

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10549-019-05156-5>) contains supplementary material, which is available to authorized users.

✉ Lajos Pusztai
lajos.pusztai@yale.edu

- ¹ Breast Medical Oncology, Yale School of Medicine, Yale Cancer Center, 300 George St, Suite 120, Rm 133, New Haven, CT 06520, USA
- ² University of Pennsylvania School of Medicine, Philadelphia, PA, USA
- ³ Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain

Introduction

Breast cancer survival rates are lower in African-American (AA) women compared to Caucasian women [1]. AA women are about 40% more likely to die from breast cancer than their Caucasian counterparts, despite improvements in mortality in both groups [2]. In addition to socioeconomic factors that affect access to health care and generally higher comorbidity rates among AA patients, biological factors may also contribute to the survival gap. For example, estrogen-receptor (ER), progesterone-receptor (PR), and human epidermal growth factor receptor-2 (HER2)-negative breast cancer (i.e., triple-negative breast cancer, TNBC) that carries a less favorable prognosis is more common among

premenopausal AA women [3, 4]. The increased incidence of TNBC among AA women is attributed to a combination of modifiable risk factors including differential rates of breastfeeding, earlier and higher parity, obesity, as well as non-modifiable factors such as predisposing single nucleotide polymorphisms in the African ancestry [5–7]. We recently showed that response to neoadjuvant chemotherapy, as measured by pathological complete response (pCR) rate, in early stage TNBC is also lower in AA patients [8]. This difference in chemotherapy responsiveness remained significant after adjusting for age, clinical tumor characteristics, length of therapy, as well as social and geographical variables including patients' insurance status and census-derived median income, suggesting that race-associated biological differences may influence treatment response.

The degree of immune infiltration in the tumor microenvironment is the most consistently reported prognostic and chemotherapy treatment response marker in TNBC [9]. Both a high tumor infiltrating lymphocyte (TIL) count and high immune-related gene expression predict higher pCR rates and better prognosis [10–12]. What drives differences in immune infiltration between breast cancers is unknown. A pooled analysis of solid tumors from The Cancer Genome Atlas (TCGA) found that both the total number of somatic mutations and the number of new antigen epitopes from tumor cells (i.e., neoantigen load) correlated positively with immune infiltration across all cancers [13]. However, subsequent analyses focused on breast cancer found that high immune cell gene expression signatures were in fact inversely associated with clonal heterogeneity in TNBC [14, 15], suggesting that an effective immune surveillance continuously eliminates clonal and genomic heterogeneity. Furthermore, a combined analysis of mutational load and clonality indicated that TNBC cancers with high mutational load and low clonality have increased cytolytic activity and better response to chemotherapy [16], suggesting that an adequately high “clonal fraction” of neoantigens may be critical for tumor elimination [17, 18].

Potential differences between the tumor immune microenvironment in AA and Caucasian breast cancer patients have not been clearly characterized. We hypothesized that the poorer chemotherapy sensitivity and worse prognosis of AA TNBC could be due to lower anti-tumor immunity in the tumor microenvironment. Less effective anti-tumor immune surveillance may arise from differences in inherited DNA mutations, the cancer somatic mutational spectra, or environmental factors such as higher levels of stress, which has been shown to dampen immune activities. The goal of this study was to compare TIL counts and immune gene expression in breast cancer between races and between breast cancer subtypes by race in the TCGA cohort. If race-associated immunobiological differences exist in breast cancers, this could potentially suggest novel immuno-oncology

therapeutic strategies to reduce the survival difference between AA and Caucasian patients.

Materials and methods

Gene expression data sources

Gene-level RNA-seq expression data and corresponding clinical data were obtained for 1,105 patients with primary breast cancer from the public portal of the TCGA (<https://portal.gdc.cancer.gov>). RNA-seq by Expectation Maximization (RSEM) data were log₂-transformed and quantile normalized for metagene analysis. After filtering for availability of race information, $n = 697$ Caucasian and $n = 162$ AA patients were selected for final analysis. Breast cancers were categorized as ER-positive/HER2-negative, referred to as ER-positive ($n = 540$), HER2-positive with any ER status ($n = 147$), or ER-negative/HER2-negative, referred to as TNBC ($n = 172$) groups, as described previously [14].

Characterizing the tumor immune microenvironment

Gene expression analysis

The expression of 13 previously reported immune metagenes (CTL, Macrophages, MHC1, MHC2, T-reg, IF1, NK, STAT1, Giam, Tfh, Tinh, Tstim, LCK) that represent various immune cell types and immune functions were calculated as the median log₂-transformed, normalized expression of the member genes as described previously [14, 19–28]. The metagenes and their gene memberships are listed in Supplementary Table 1. The prognostic and chemotherapy response predictive value of each of these metagenes was previously assessed in TCGA and other datasets [19, 28].

We also assessed two recently reported immune-checkpoint blockade (ICB) response predictive markers: the “T cell inflamed” gene signature by Ayers et al. [26] and the Tumor immune dysfunction and exclusion (TIDE) analysis proposed by Jiang et al. [29, 30]. A higher TIDE score indicates greater dysfunction and exclusion of T cells from the immune microenvironment, corresponding to lower likelihood of benefit from ICB.

The relative fractions of 22 immune cell sub-populations were estimated using the CIBERSORT tool [31, 32]. Immune cell populations were aggregated into 9 broader categories to yield estimates of more comprehensive cell populations as described previously [29]. Metagene heatmaps were created in R using the “pheatmap” package version 1.0.10.

Differential gene expression analysis

To identify differentially expressed immune genes between AA and Caucasian cancers, we used the “limma” package in R [33]. Significant differential expression was defined as absolute log₂-fold change > 1.5 with Bonferroni-adjusted *p* value < 0.05. In this analysis, we included 749 of 770 annotated genes from the NanoString IO360 panel [34] that had available RNA-seq data (Supplementary Table 1).

TIL counts, immunomodulatory scores, and TNBC subtypes

Stromal (percent-area of stroma) and total (percent-area of whole-section) TIL counts for TCGA TNBC cases with race annotation (Supplementary Table 2) were obtained from previously published datasets [35–37]. Previously published TNBC immunomodulatory scores and refined-Vanderbilt TNBC subtypes (basal-like 1, basal-like 2, luminal androgen receptor, and mesenchymal) for TCGA TNBC cases were also compared between races [35].

Measuring genomic aberrations and clonality

Gene-level somatic mutation data were obtained from Ciriello et al. [27]. Mutation load was calculated as the number of somatic mutations in a sample, normalized by the total length of sequences with adequate read coverage (> 30X). Predicted neoantigen load data were taken from a previous publication [14]. We assessed mutational heterogeneity using the Mutant-Allele Tumor Heterogeneity (MATH) score, which uses the standardized variance of the variant allele frequency distribution of somatic mutations to approximate clonal heterogeneity [38]. Overall deletion load was defined as the number of genes with GISTIC value of “−2”, and amplification load was defined as the number of genes with GISTIC value of “+2”, indicating definite deletion or amplification of a given segment, respectively. All genomic features were log₂-transformed for visualization and statistical analysis. The availability of gene expression data for each metric is summarized in Supplementary Table 2.

Table 1 Patient characteristics

All patients (<i>n</i> = 859)	Caucasian (<i>n</i> = 697)	AA (<i>n</i> = 162)	<i>p</i> value
Age (years), median	59	54	0.037
Interquartile range (years)	49–67	47–67	
Clinical subtype, <i>n</i> (%)			1.81e−07
ER+	458 (65.7%)	82 (50.6%)	4.21e−04
HER2+	125 (17.9%)	22 (13.6%)	0.204
ER−/HER2− (TNBC)	114 (16.4%)	58 (35.8%)	1.78e−07
Stage, <i>n</i> (%)			0.379
I	131 (18.8%)	30 (18.5%)	
II	382 (54.8%)	93 (57.4%)	
III	163 (23.4%)	30 (18.5%)	
IV	11 (1.6%)	4 (2.5%)	
Not available	10 (1.4%)	5 (3.1%)	
TNBC only (<i>n</i> = 172)	Caucasian (<i>n</i> = 114)	AA (<i>n</i> = 58)	<i>p</i> value
Age (years), median	54.5	51.5	0.839
Interquartile range (years)	48–62	47–66	
Stage			0.788
I	22 (19.3%)	10 (17.2%)	
II	70 (61.4%)	34 (58.6%)	
III	17 (14.9%)	12 (20.7%)	
IV	2 (1.8%)	0 (0%)	
Not available	3 (2.6%)	2 (3.5%)	

Bold values indicate *p* value of < 0.05, from Fisher’s exact test, unadjusted for multiple comparisons, is considered significant

ER Estrogen receptor, HER2 human epidermal growth factor receptor-2, TNBC triple-negative breast cancer

Statistical analysis

All statistical analyses were performed in R version 3.5.0. The Wilcoxon Rank Sum test was used to compare patient characteristics, immune measures and genomic aberrations between the AA and Caucasian cohorts. p values are listed in Supplementary Table 3. Student's t test was used to compare normally distributed TIDE scores and CIBERSORT immune cell populations. Chi-square was employed for the comparison categorical variables. Statistical tests were two-sided, and $p < 0.05$ was considered significant. p values were adjusted for multiple comparisons using the Bonferroni method.

Results

Comparison of immune cell functions and sub-populations by race in all breast cancer subtypes combined

Patient characteristics are summarized in Table 1. The proportion of TNBC was higher in the AA cohort compared to the Caucasian cohort (35.8% vs. 16.4%; $p = 1.78e-07$), while the proportion of ER-positive cancers was higher in the Caucasian group (65.7% vs. 50.6%; $p = 4.21e-04$). The distribution of clinical stage by race or within subtypes by race did not vary significantly.

We compared the expression of 14 immune metagenes between AA ($n = 162$) and Caucasian ($n = 697$) patients (Supplementary Table 3). Overall, we observed higher

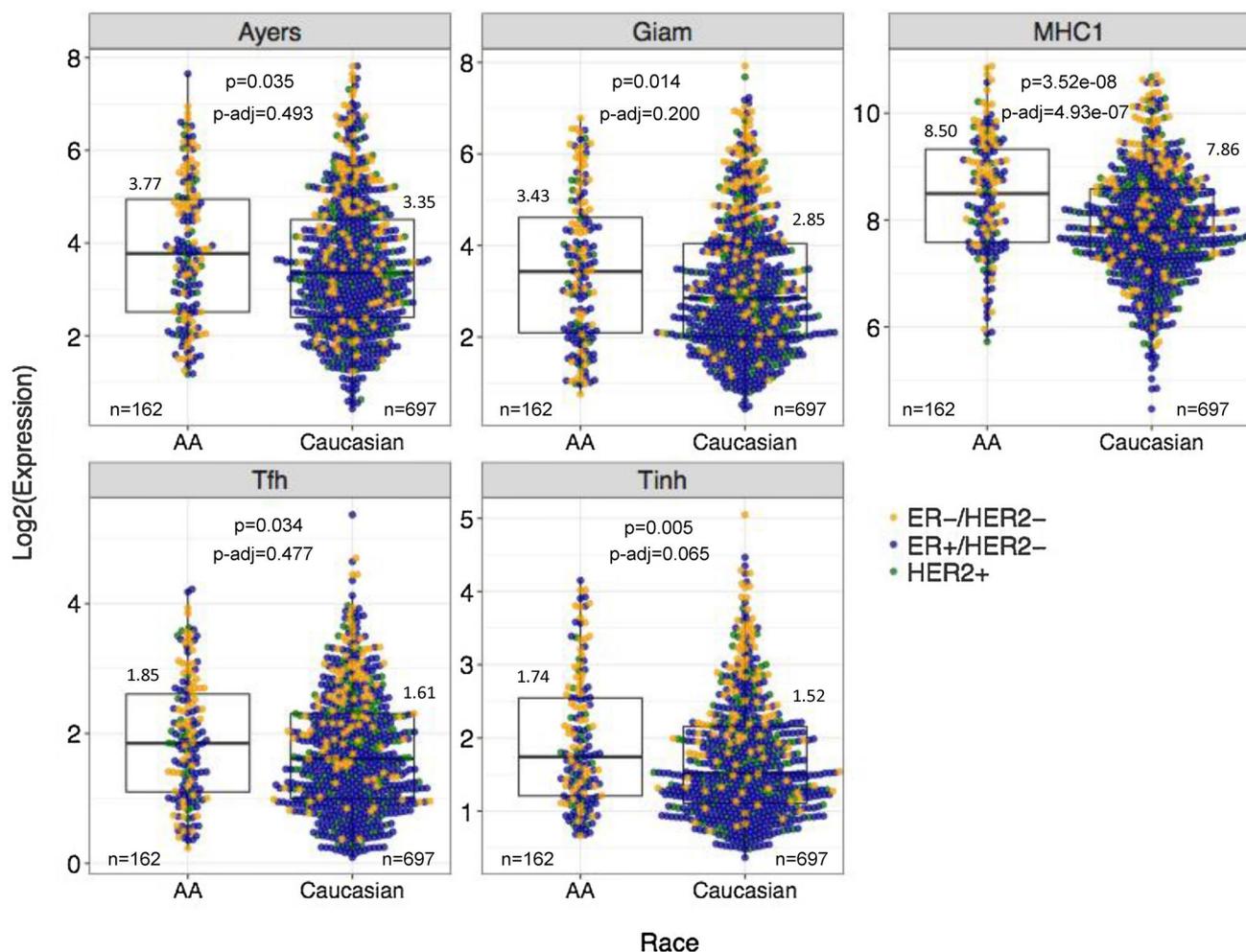


Fig. 1 Expression of representative immune metagenes in breast cancer by patient race across all tumor subtypes. Boxplots summarize the median and IQR expression of five representative immune metagenes by race in the TCGA cohort. All tumor subtypes are included. Log₂-transformed median values, available cases (n), Wilcoxon rank sum

p values, and Bonferroni adjusted p values are displayed for each metagene. Colored points represent the log₂-transformed, normalized expression of ER-positive, HER2-positive, and TNBC cases for each metagene

Fig. 2 Tumor immune dysfunction and exclusion (TIDE) scores by patient race across all tumor subtypes. The mean TIDE, IFN-gamma (“IFNG”), T cell dysfunction (“Dysfunction”), and T cell exclusion (“Exclusion”) scores are displayed for each race cohort, with error bars representing standard error of the mean. All tumor subtypes are included. The means of AA scores are in dark grey; the means of Caucasian scores are in light grey. **Student’s *t* test $p=0.014$

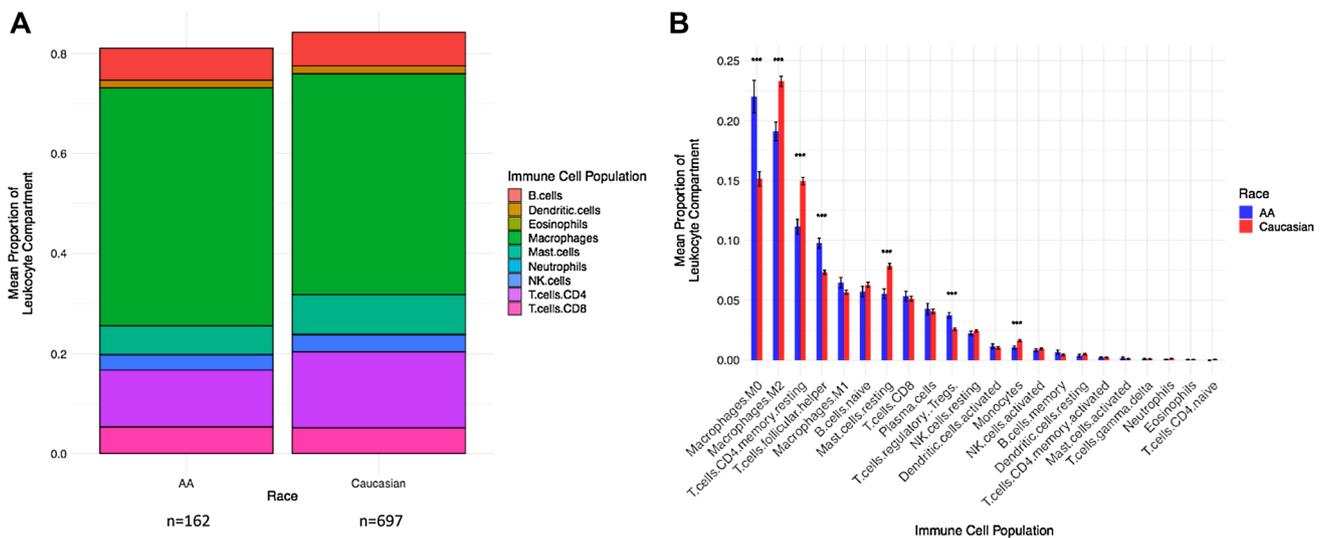
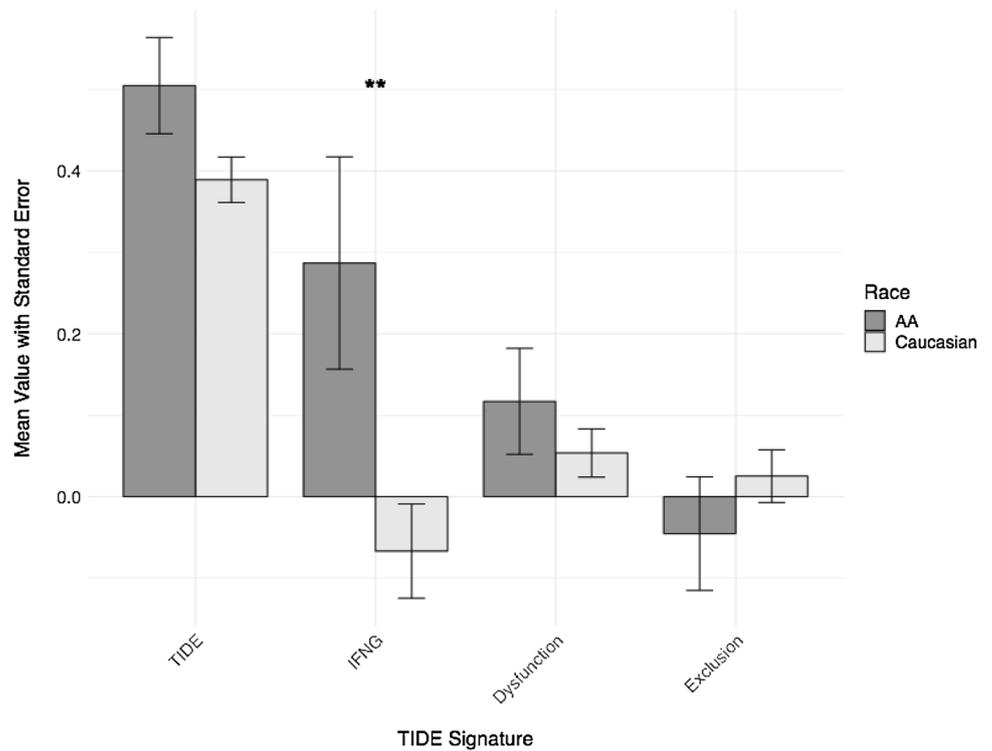


Fig. 3 Relative proportions of immune cell populations by patient race across all tumor subtypes. **a** Bars display the mean proportions of 9 aggregated immune cell populations. Race is on the *x*-axis, and the mean proportion of the leukocyte compartment is on the *y*-axis. All tumor subtypes are included ($n=162$ AA, $n=697$ Caucasian). The height of the bars is <1.0 due to the fact that the mean proportion of each immune cell population was independently calculated, then aggregated. Colors represent individual immune cell populations. **b** Bars display the mean proportions of 22 immune cell populations. Immune cell population is on the *x*-axis, and the mean pro-

portion of the leukocyte component, with standard error of the mean, is on the *y*-axis. All tumor subtypes are included. The mean of AA patients is in blue; the mean of Caucasian patients is in red. All *p* values are Bonferroni adjusted. *** = statistical significance by Student’s *t* test $p < 0.005$: Macrophages M0 p adjusted = $5.87e-06$, Macrophages M2 p adjusted = $3.97e-06$, T cells CD4+ memory resting p adjusted = $5.17e-08$, T cells follicular helper p adjusted = $1.41e-07$, Mast cells resting p adjusted = $1.16e-06$, T cells regulatory p adjusted = $1.28e-05$, Monocytes p adjusted = $5.44e-04$

median expression for all immune metagenes in AA cancers. However, only the MHC1, Ayers/T cell inflamed, Giam, Stat1, Tfh, and Tnh metagenes were nominally statistically significant (Fig. 1), and after Bonferroni adjustment, only the MHC1 metagene remained significant (p adjusted = $4.93E-07$).

Tumor immune dysfunction and exclusion (TIDE) analysis also showed a numerically higher median TIDE score in AA tumors, suggesting decreased responsiveness to ICB, but the difference did not reach statistical significance (Fig. 2). However, one component of the TIDE score, IFN-gamma activation, was significantly higher in AA tumors compared to Caucasian tumors (unadjusted Student t test $p = 0.014$).

CIBERSORT was used to estimate the proportions of immune cell populations in the tumor microenvironment. In both race groups, the aggregated macrophage population (comprised of Monocytes, M0, M1, and M2 populations)

represented the largest proportion of the leukocyte compartment (Fig. 3a). After adjusting for multiple comparisons, the mean proportions of M0 macrophages, T-follicular helper, and T-regulatory cells were significantly higher in AA cancers relative to Caucasian breast cancers (p adjusted = $5.87e-06$, $1.41e-07$, $1.28e-05$, respectively), while the mean proportions of M2 macrophages, resting CD4+ memory cells, resting Mast cells, and Monocytes were higher in Caucasian cancers (p adjusted = $3.97e-06$, $5.17e-08$, $1.16e-06$, $5.44e-04$, respectively) (Fig. 3b). Since AA breast cancers included more TNBC and fewer ER-positive cancers, these differences may be caused by subtype-associated differences in the tumor immune microenvironment. Therefore, next we examined immune variables within molecular subtypes.

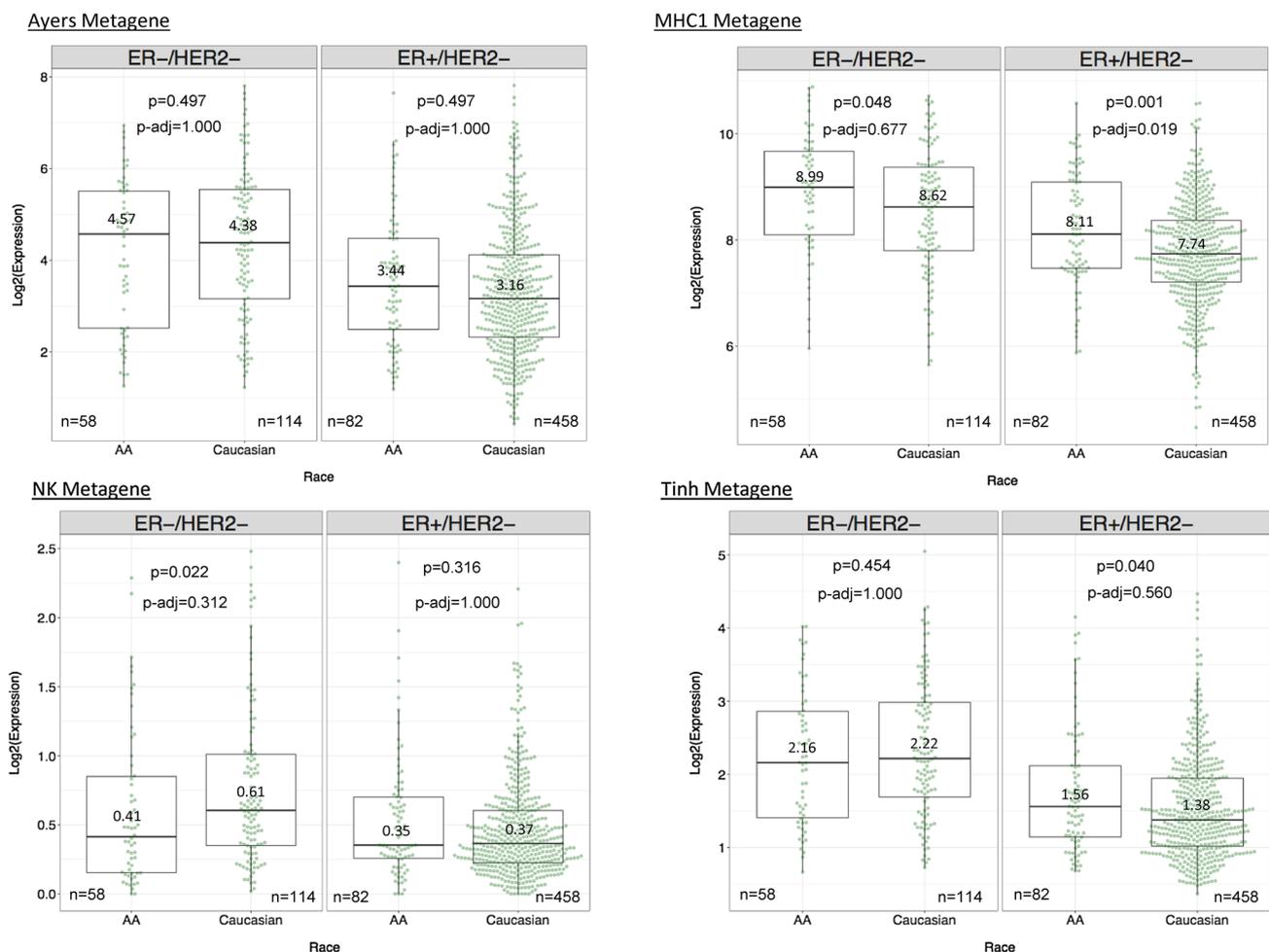


Fig. 4 Differential expression of representative immune metagenes by patient race and individual tumor subtype. Boxplots demonstrate the median and IQR expression of four representative immune metagenes, grouped by tumor subtype and patient race. TNBC cases are in the left panel and ER-positive cases in the right panel. Race

is on the x -axis, and \log_2 -transformed, normalized expression of the metagene is on the y -axis. \log_2 -transformed median values, available cases (n), Wilcoxon rank sum p values, and Bonferroni-adjusted p values are displayed for each metagene

Comparison of immune cell functions and sub-populations by race and breast cancer subtype

We next compared immune metagene expression between AA and Caucasian within the TNBC and ER-positive subtypes. HER2-positive cases were excluded from these analyses due to the heterogeneity of HER2-positive cancers that include both ER-positive and ER-negative subsets that continue to express the typical molecular features of their respective ER status. When TNBC cases were analyzed by race, 10 of 14 metagenes had lower median expression in AA ($n=58$) compared to Caucasian ($n=114$) cancers, whereas the MHC1 metagene showed higher median expression in AA cases. However, after Bonferroni adjustment, these differences were no longer significant (Fig. 4). This suggests that no major immune differences exist in TNBC by race as measured by immune metagene expression.

ER-positive cancers from AA patients ($n=82$) had significantly higher median expression of the MHC1 metagene relative to Caucasian cancers ($n=458$) (p adjusted=0.019). This explains the overall higher MHC1 metagene expression in AA patients when all breast cancer subtypes were combined (see “Results”).

The mean TIDE scores and the Ayers/T cell inflamed metagene scores were similar between AA and Caucasian TNBC and between AA and Caucasian ER-positive cancers, suggesting no difference in ICB sensitivity by race (Fig. 5). Both AA and Caucasian IFN-gamma signatures, measured

by TIDE, were numerically higher in TNBC vs. ER-positive cancers. Likewise, in a heatmap of Ayers gene member expression, TNBC cases, irrespective of race, tended to cluster with highly expressed genes in the signature (Fig. 6). CIBERSORT analysis showed that the aggregated macrophage population represented the largest proportion of the leukocyte compartment in TNBC in both race cohorts (Fig. 7a). The proportion of resting CD4+ memory cells was significantly higher in Caucasian TNBC compared to AA TNBC (p adjusted=0.014), (Fig. 7b).

We also compared immune metagene expression differences between breast cancer subtypes within each race cohort separately. In AA patients, TNBC ($n=58$) demonstrated higher median expression of the MHC1 (p adjusted=0.005) and STAT1 (p adjusted=4.79e−04) metagenes relative to ER-positive cancers ($n=82$). In Caucasian patients, TNBC ($n=114$) demonstrated higher median expression of 11 of 14 metagenes (with the exception of the Macrophage, MHC2, and T-reg metagenes) relative to ER-positive cancers ($n=458$). These findings are consistent with multiple earlier reports that showed higher immune infiltration in TNBC compared to ER-positive cancers.

Differentially expressed immune genes between AA and Caucasian patients

Next, we examined immune gene expression differences at the individual gene level for 749 immune-related genes. After Bonferroni adjustment, 141 genes were significantly up-regulated and 125 genes were significantly

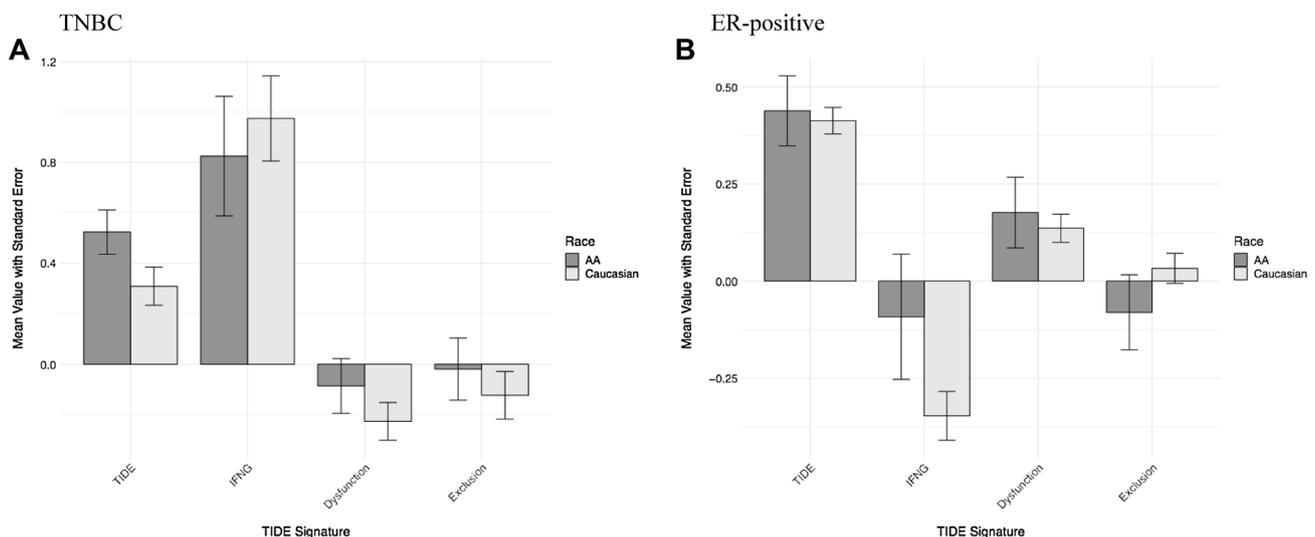


Fig. 5 Tumor immune dysfunction and exclusion (TIDE) scores by patient race and tumor subtype. The mean TIDE, IFN-gamma (“IFNG”), T cell dysfunction (“Dysfunction”), and T cell exclusion (“Exclusion”) scores are displayed for each race cohort with error

bars representing standard error of the mean. The means of AA scores are in dark grey; the means of Caucasian scores are in light grey. **a** Signatures for TNBC cases only, **b** signatures for ER-positive cases only

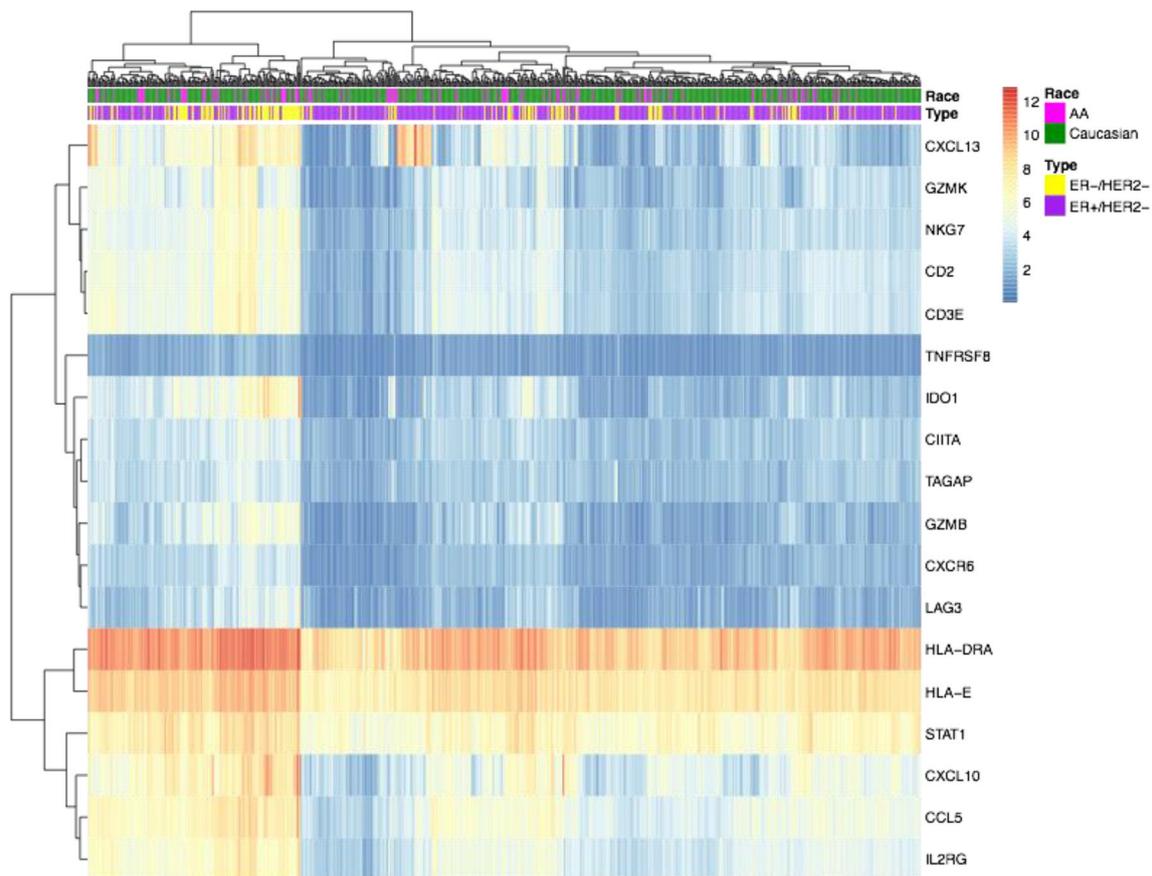


Fig. 6 Expression of Ayers metagene member genes, clustered by rows and columns and annotated by patient race and tumor subtype. Heatmap displaying the clustered expression of the genes in the Ayers metagene ($n=18$), annotated by patient race and tumor subtype. Columns represent patients, with pink (AA) versus green (Caucasian)

reflecting patient race and yellow (TNBC) versus purple (ER-positive) reflecting tumor subtype. Rows represent individual genes, with normalized expression level represented from blue (low expression) to red (high expression). The heatmap is clustered by rows and columns

down-regulated in Caucasian tumors across all subtypes relative to AA tumors (Supplementary Table 4).

When we compared AA and Caucasian TNBC, 13 genes were significantly up-regulated and 12 were significantly down-regulated in Caucasian TNBC relative to AA TNBC. None of these genes had an absolute log-fold-change difference of > 1.5 between races (Supplementary Table 4). The 5 most differentially expressed genes between Caucasian and AA TNBC were *APOE*, *TNFRSF25*, *TYMP* (2.13-, 2.06-, and 1.9-fold higher in AA TNBC, respectively), *MRC1* and *TLR3* (2.0 and 1.85 higher in Caucasian TNBC, respectively). When we compared AA and Caucasian ER-positive cancers, we found 88 genes were up-regulated and 82 genes were down-regulated in Caucasian cancers relative to AA cancers. *ISG15* was significantly up-regulated in AA ER-positive cancers (3.20-fold-change, p adjusted = $2.5e-10$), while *C7* was significantly up-regulated in Caucasian ER-positive cancers (2.83-fold-change, p adjusted = $4.51e-03$).

Comparison of TIL counts, immunomodulatory scores, and Vanderbilt Subtype distribution in TNBC by race

TIL counts were analyzed using two different quantification methods: expressing TILs as percent-area of stroma ($n=108$) and as percent area of the whole-section ($n=142$) according to Loi et al. [36] and Lehmann et al. [35], respectively. In the Loi et al. dataset, there were no statistically significant differences between stromal TILs in AA ($n=20$) and Caucasian ($n=88$) TNBC cases (Wilcoxon $p=0.350$, Fig. 8a). In the Lehmann et al. dataset, we also observed no difference in whole-section mononuclear cell infiltrate by race (AA $n=51$, Caucasian $n=91$, Wilcoxon $p=0.546$, Fig. 8b). The distributions of lymphocyte-predominant ($> 50\%$ TIL), lymphocyte-moderate (10–50% TIL), and lymphocyte-poor ($< 10\%$ TIL) cases were also not significantly different between races in either dataset ($p=0.515$ stromal and $p=0.672$ whole-section).

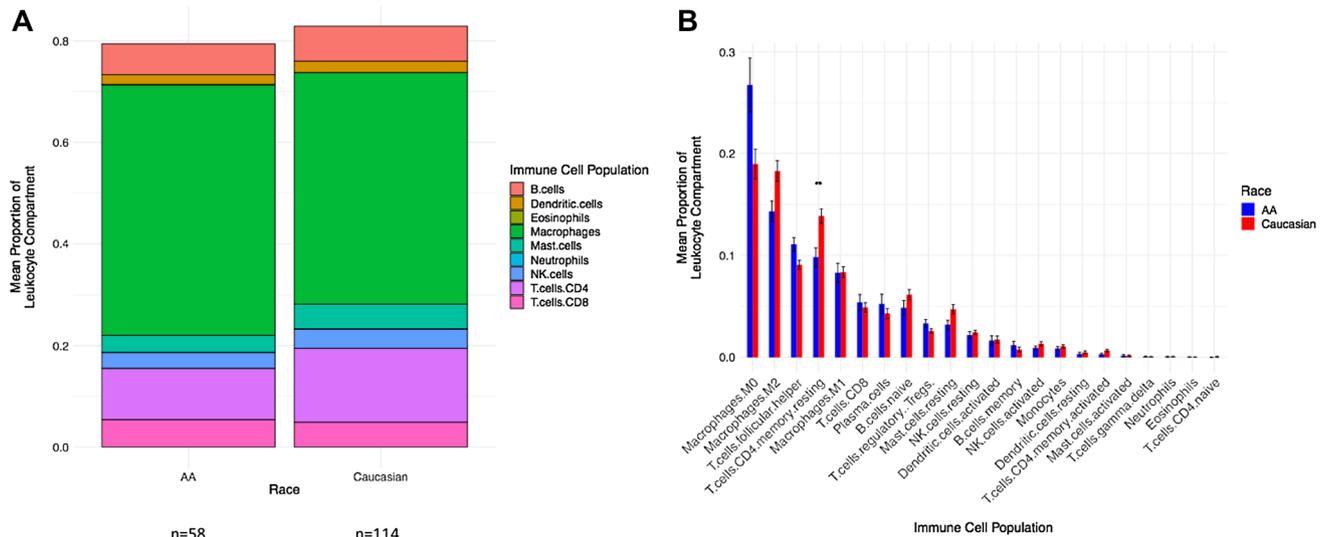


Fig. 7 Relative proportions of immune cell populations by patient race in TNBC cases alone. **a** Bars display the mean proportions of 9 aggregated immune cell populations. Race is on the *x*-axis, and the mean proportion of the leukocyte compartment is on the *y*-axis. Only TNBC cases are included ($n=58$ AA, $n=114$ Caucasian). The height of the bars is <1.0 due to the fact that the mean proportion of each immune cell population was independently calculated, then aggregated. Colors represent individual immune cell populations. **b** Bars

display the mean proportions of 22 immune cell populations. Immune cell population is on the *x*-axis, and the mean proportion of the leukocyte component, with standard error of the mean, is on the *y*-axis. All tumor subtypes are included. The mean of AA patients is in blue; the mean of Caucasian patients is in red. All *p* values are Bonferroni adjusted for FDR. **Statistical significance by Student's *t* test $p < 0.05$: T cells CD4+ memory resting p adjusted = 0.014

The distribution of immunomodulatory scores, classified by Lehmann et al. [35], was also not significantly different between AA ($n=51$) and Caucasian ($n=91$) TNBC patients. Cases with an “intense” immunomodulatory score were few in both the AA and Caucasian cohorts, 14% and 11%, respectively. There was no difference in the distribution of refined-Vanderbilt subtypes between races ($p=0.377$), with the largest proportion of both AA and Caucasian TNBC classified as basal-like 1 (31% for both race groups).

Comparison of genomic aberration metrics and clonal heterogeneity by race

Next, we compared neoantigen, total somatic mutation, deletion, amplification loads, and clonal heterogeneity between AA and Caucasian cancers. Data were available for variable subsets of patients in both race cohorts (Supplementary Table 2). Overall, the neoantigen load was higher in AA compared to Caucasian patients (median = 10 vs. 7, p adjusted = 0.027) when all subtypes were combined. The somatic mutation burden (AA median = 45 vs. Caucasian median = 34, p adjusted = 0.007), amplification load (AA median = 481 vs. Caucasian median = 258, p adjusted = 0.032) and deletion load (AA median = 12 vs. Caucasian median = 4, p adjusted = 0.003) were higher in AA cases (Fig. 9). All these genomic derangement metrics were significantly higher in TNBC compared to ER-positive

cancers (Supplementary Table 3). When TNBC and ER-positive subtypes were compared separately between AA and Caucasian cohorts, these racial differences disappeared, indicating that the apparent racial difference across all subtypes was driven by the higher proportion of TNBC among AA patients. We also compared estimates of clonal heterogeneity captured by the MATH score and found no significant differences between AA and Caucasian patients overall or within individual tumor subtypes.

Discussion

In this paper, we examined immunological differences, measured by immune gene mRNA expression and TIL counts, between breast cancers of AA and Caucasian patients in the TCGA. We sought to detect race-associated differences in the immune microenvironment of TNBC, where immune infiltration has been shown to have prognostic and chemotherapy response predictive functions.

We found no statistically significant difference in TIL counts, TIL distribution, molecular subtype distribution, or immunomodulatory scores between AA and Caucasian TNBC. Unfortunately, these analyses were limited by the small number of TCGA cases with race, TIL, and subtype annotation. There were many more cancers with available mRNA expression data, and analysis of immune metagenes

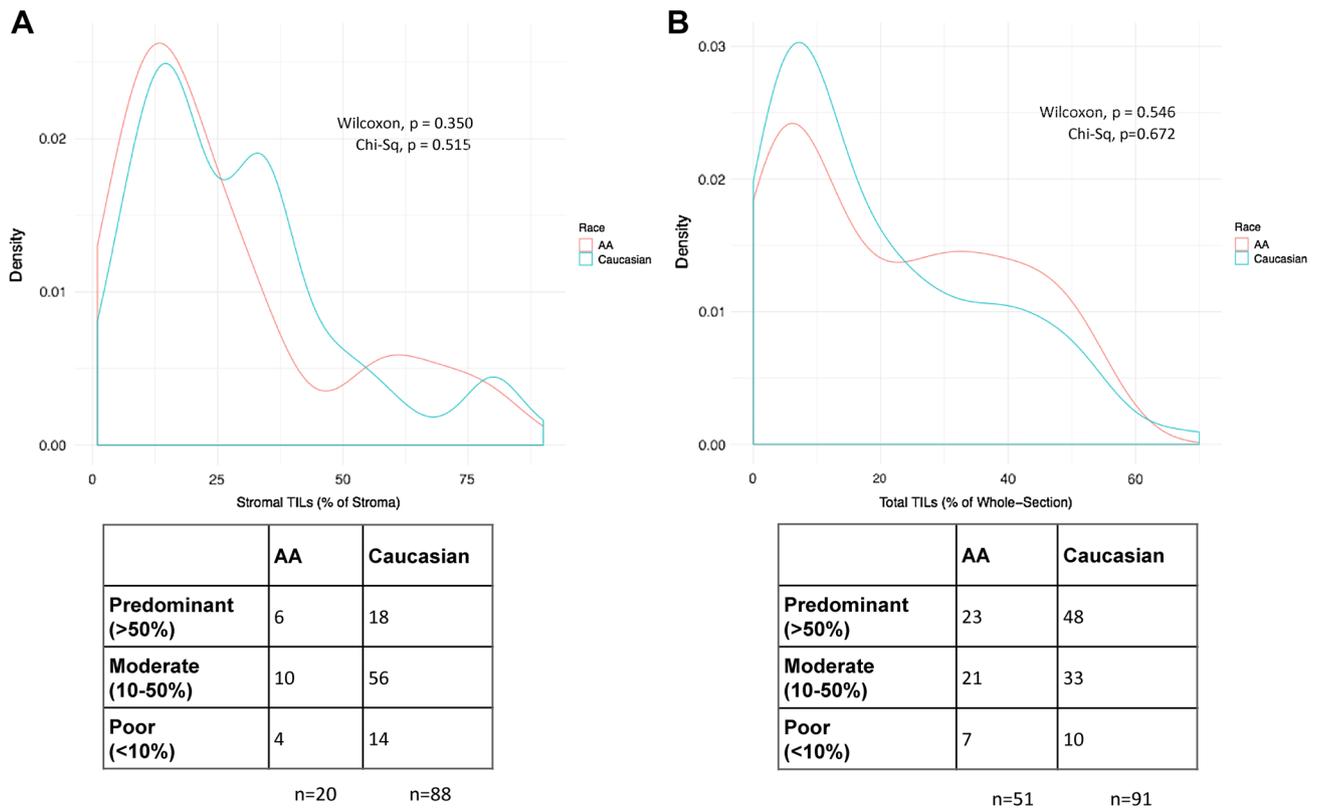


Fig. 8 Stromal and total TIL counts by race in TNBC cases alone. Density plots display the distributions of TIL counts in TNBC cases by race. Red and blue lines represent AA and Caucasian TNBC, respectively. The corresponding table represents the number of lymphocyte-predominant, lymphocyte-moderate, and lymphocyte-poor cases in each cohort, classified as >50% TIL infiltration, 10–50% TIL infiltration, and <10% TIL infiltration, respectively. **a** Stromal TIL counts from Loi et al. are displayed as % of tumor stroma. $n = 20$ AA Caucasian and $n = 88$ Caucasian TNBC were considered. Neither

the median stromal TIL counts (Wilcoxon $p = 0.350$) nor the distribution of TIL burden classification (Chi-square $p = 0.515$) was significantly different between races. **b** Whole-section total TIL counts from Lehmann et al. are displayed as % of whole section. $n = 51$ AA Caucasian and $n = 91$ Caucasian TNBC were considered. Neither the median stromal TIL counts (Wilcoxon $p = 0.546$) nor the distribution of TIL burden classification (Chi-square $p = 0.672$) was significantly different between races

and immunotherapy predictive gene signatures yielded more nuanced results, which suggested small, but potentially important differences. Overall, almost all immune metagenes had higher expression in AA compared to Caucasian breast cancers, and this reached significance for the MHC1 metagene. The higher overall expression of immune genes in AA across all subtypes is likely driven by the higher proportion of TNBC in the AA patient cohort, as TNBC in general contains more inflammatory cells. When patients with TNBC were analyzed separately by race, 10 of the 14 metagenes had lower expression in AA patients, suggesting a trend for lower immune activity in AA TNBC ($n = 58$) compared to Caucasian TNBC ($n = 114$). In contrast, ER-positive cancers showed higher expression of 11 of 14 immune metagenes in AA compared to Caucasian patients, with significantly higher expression of MHC1 metagene (p adjusted = 0.019). These results suggest subtle immunological differences between breast cancers in AA and Caucasian patients, but larger datasets will be required to validate these differences.

Racial differences in MHC1 metagene expression may relate to recent findings that germline HLA-I genotypes influence anti-tumor immunity. Chowell et al. [39] found that maximal heterozygosity at HLA-I loci, as well as the presence of HLA-B44 superfamily alleles, improved patients' overall survival after immune checkpoint therapy. A more diverse set of HLA-I molecules from a heterozygous individual is predicted to present a larger repertoire of tumor-derived neoantigens to cytotoxic T cells, resulting in a more robust immune response. Racial differences in the distributions of HLA-I types have been well described; in one study, AA participants had higher rates of heterozygosity at all HLA-I loci compared to Caucasian participants, but the HLA-B44*02 allele (a member of the B44 superfamily) was more frequent in Caucasian participants [40]. These differences potentially reflect that immune adaptations suited to ancestral environments differentially influence anti-tumor immunity following immunotherapy and are important areas of future investigation.

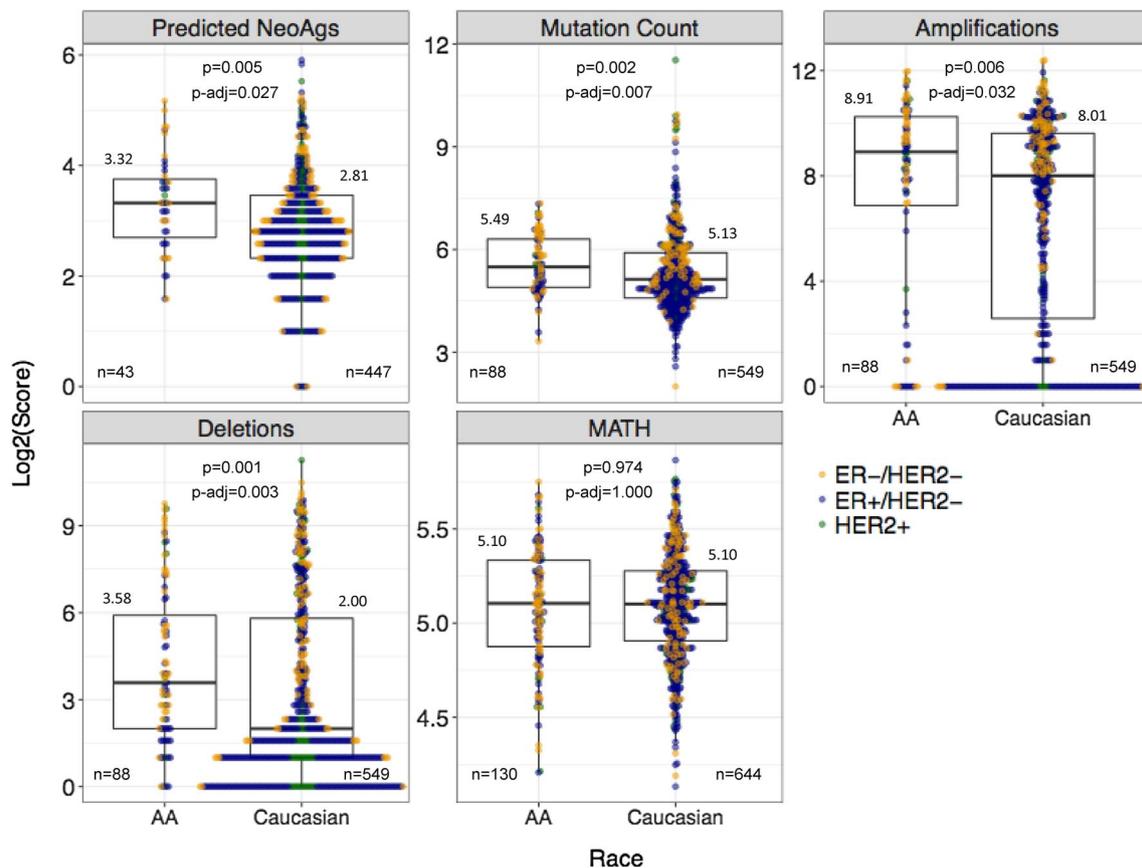


Fig. 9 Comparison of genomic aberrations and clonal heterogeneity by race across all tumor subtypes. Boxplots demonstrate the median and IQR scores of four genomic aberration measures and the MATH clonal heterogeneity score, grouped by race cohort. All tumor subtypes are included. Race is on the x-axis; log₂-transformed score of the measure is on the y-axis. Log₂-transformed median values, avail-

able cases (*n*), Wilcoxon rank sum *p* values, and Bonferroni-adjusted *p* values are displayed for each measure. Colored points represent the log₂-transformed scores of ER-positive, HER2-positive, and TNBC cases for each genomic metric. Abbreviations: Predicted NeoAgs = Predicted Neoantigens, MATH = Mutant-Allele Tumor Heterogeneity

A recent study of AA and Caucasian cases from the TCGA, along with 194 Nigerian breast cancer cases, also found only modest racial differences in mRNA immune gene expression [41]. More significant molecular differences were detected at the DNA level, with increased homologous recombination deficiency and more frequent *TP53* mutations in Nigerian, ER-positive breast cancers compared to ER-positive cancers in Caucasian patients in the USA. Also pointing to racial differences at the somatic DNA level, two previous studies detected more varied and more frequent somatic mutations in AA TNBC than in European-descent TNBC, suggesting more genomic instability in AA cancers [42, 43]. Given these findings, further profiling of germline and somatic aberrations in AA TNBC and the interplay with the tumor immune microenvironment is warranted.

Pitt et al. [41] found enriched IFN-gamma activation in Nigerian and AA ER-positive cancers and increased macrophage infiltration in the Nigerian basal-like TNBC. Our study demonstrated no significant difference in IFN-gamma

activation, reflected by the TIDE scores, between AA and Caucasian TNBC; however, the TIDE-estimated IFN-gamma score was higher in AA tumors across all subtypes. This difference appears to be driven by a low expression of IFN-gamma in Caucasian ER-positive cancers. Nonetheless, neither the overall TIDE score nor the Ayers metagene score was significantly higher in AA tumors of any subtype, suggesting that the higher IFN-gamma signature measured by TIDE did not translate to higher predicted benefit from immune checkpoint therapy. In another study, an interferon gene signature that was observed to be up-regulated in AA prostate cancers was also found to be up-regulated in the AA breast cancer microenvironment [44].

CIBERSORT deconvolution allowed us to infer the relative proportions of immune cell populations in AA and Caucasian TNBC. We detected significantly increased proportions of T-regulatory cells and M0 macrophages in AA tumors across all subtypes. However, these differences did not meet statistical significance in AA TNBC compared to

Caucasian TNBC. These findings are discordant with a previous CIBERSORT analysis that found a higher proportion of T-regulatory cells in AA TNBC/basal subtype cancers [45]. Interestingly, tumoral expression of MHC1 genes has been shown to augment T-regulatory cell function, and we did observe enrichment of MHC1 expression in AA cancers relative to Caucasian cancers [46].

When differential gene expression analyses on the TNBC cases was performed, we found that *APOE*, encoding Apolipoprotein E, was the most differentially expressed gene between AA and Caucasian TNBC, with 2.13-fold higher expression in AA TNBC. A recent study demonstrated that ApoE signaling mediates levels of circulating myeloid-derived suppressor cells (MDSC), an immunosuppressive innate cell population, and in turn, T cell activation and immunotherapy efficacy in cancer models [47]. We observed no significant difference in T cell activation or T cell exclusion (estimated in part by MDSC abundance) between AA and Caucasian TNBC. Nonetheless, the degree of lipoprotein signaling and the effects on T cell activation in the tumor microenvironment may be an important distinguishing factor between AA and Caucasian TNBC biology.

In conclusion, we found no large-scale immunogenic differences between AA and Caucasian TNBC based on gene expression analysis. Overall, we observed a trend toward lower immune gene expressions in AA TNBC, but this did not reach statistical significance for most immune metrics. The power of our analyses is limited by the relatively small number of AA TNBC in the TCGA. Molecular, histologic and outcome data from the large, ongoing neoadjuvant and adjuvant trials (NCT02489448, NCT02101385, NCT03366844) will be required to answer the question of whether subtle, race-associated differences in the tumor immune microenvironment influence clinical outcome.

Funding This research was supported by an NCI R01 Grant (R01CA219647) to L.P.

Compliance with ethical standards

Conflict of interest Tess O'Meara, Anton Safonov, David Casadevall, Tao Qing, Brigid Killelea declares they have no conflict of interest. Andrea Silber has received remuneration from Astra Zeneca. Christos Hatzis is now an employee of Bristol-Myers Squibb Co. Lajos Pusztai has received consulting fees and honoraria from Merck, Astra Zeneca, Novartis, Seattle Genetics, Pfizer, and Almac.

Ethical approval This article does not contain studies with human participants or animals performed by any of the authors.

Informed consent No informed consent was needed for this study. Human subjects were not involved.

References

1. DeSantis C et al. (2014) Breast cancer statistics, 2013. *CA: A Cancer J Clin* 64(1):52–62
2. DeSantis CE et al (2016) Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA Cancer J Clin* 66(1):31–42
3. Stark A et al (2010) African ancestry and higher prevalence of triple-negative breast cancer. *Cancer* 116(21):4926–4932
4. Eric C, Dietze CS, Miranda-Carboni G, Ruth O'Regan, Seewaldt VL (2015) Triple-negative breast cancer in African-American women: disparities versus biology. *Nat Rev Cancer* 15:248–254
5. Shinde SS et al (2010) Higher parity and shorter breastfeeding duration. *Cancer* 116(21):4933–4943
6. Lawrence TS, Rosenberg SA (2010) L, Definition of triple-negative breast cancer and relationship to basal-like molecular subtype. In: VT DeVita Jr (eds) *PPO updates: principles and practice of oncology*. Lippincott Williams & Wilkins, New York, pp 1–6
7. Huo D, Hu H, Rhie SK, Gamazon ER, Cherniack AD, Liu J, Yoshimatsu TF, Pitt JJ, Hoadley KA, Troester M, Ru Y, Lichtenberg T, Sturtz LA, Shelley CS, Benz CC, Mills GB, Laird PW, Shriver CD, Perou CM, Olopade OI (2017) Comparison of breast cancer molecular features and survival by African and European Ancestry in the cancer genome atlas. *JAMA Oncol* 3(12):1654–1662
8. Killelea BK et al (2015) Racial differences in the use and outcome of neoadjuvant chemotherapy for breast cancer: results from the national cancer data base. *J Clin Oncol* 33(36):4267–4276
9. Karn T et al (2015) The influence of host factors on the prognosis of breast cancer: stroma and immune cell components as cancer biomarkers. *Curr Cancer Drug Targets* 15(8):652–664
10. Bianchini G et al (2010) Molecular anatomy of breast cancer stroma and its prognostic value in estrogen receptor–positive and –negative cancers. *J Clin Oncol* 28(28):4316–4323
11. Adams S et al (2014) Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *J Clin Oncol* 32(27):2959–2966
12. Wimberly H et al (2015) PD-L1 expression correlates with tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy in breast cancer. *Cancer Immunol Res* 3(4):326–332
13. Efremova M, Finotello F, Rieder D, Trajanoski Z (2017) Neoantigens generated by individual mutations and their role in cancer immunity and immunotherapy. *Front Immunol* 8:1679
14. Safonov A et al (2017) Immune gene expression is associated with genomic aberrations in breast cancer. *Can Res* 77(12):3317–3324
15. Karn T, Hatzis C, Sanger N, El-Balat A, Rody A, Holtrich U, Becker S, Bianchini G, Pusztai L (2017) Association between genomic metrics and immune infiltration in triple-negative breast cancer. *JAMA Oncol* 3(12):1707–1711
16. Jiang T et al (2016) Predictors of chemosensitivity in triple negative breast cancer: an integrated genomic analysis. *PLoS Med* 13(12):e1002193
17. McGranahan N et al (2016) Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* 351(6280):1463
18. Gejman RS et al (2018) Rejection of immunogenic tumor clones is limited by clonal fraction. *eLife* 7:e41090
19. Rody A et al (2009) T-cell metagene predicts a favorable prognosis in estrogen receptor-negative and HER2-positive breast cancers. *Breast Cancer Res* 11(2):R15

20. Callari M et al (2016) Subtype-specific metagene-based prediction of outcome after neoadjuvant and adjuvant treatment in breast cancer. *Clin Cancer Res* 22(2):337–345
21. Denkert C et al (2010) Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol* 28(1):105–113
22. Esteva FJ et al (2007) CD40 signaling predicts response to pre-operative trastuzumab and concomitant paclitaxel followed by 5-fluorouracil, epirubicin, and cyclophosphamide in HER-2-over-expressing breast cancer. *Breast Cancer Res* 9(6):R87
23. Lennerz V et al (2005) The response of autologous T cells to a human melanoma is dominated by mutated neoantigens. *Proc Natl Acad Sci USA* 102(44):16013–16018
24. Heemskerk B, Kvistborg P, Schumacher TN (2013) The cancer antigenome. *EMBO J* 32(2):194–203
25. Brown SD et al (2014) Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome Res* 24(5):743–750
26. Hsiao TH et al (2013) Identification of genomic functional hotspots with copy number alteration in liver cancer. *EURASIP J Bioinform Syst Biol* 2013(1):14
27. Ciriello G et al (2015) Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* 163(2):506–519
28. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N (2015) Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160:48–61
29. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, Albright A, Cheng JD, Peter Kang S, Shankaran V, Piha-Paul SA, Yearley J, Seiwert TY, Ribas A, McClanahan TK (2017) IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* 127(8):2930–2940
30. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, Li Z, Traugh N, Bu X, Li B, Liu J, Freeman GJ, Brown MA, Wucherpfennig KW, Shirley Liu X (2018) Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. *Nat Med.* <https://doi.org/10.1038/s41591-018-0136-1>
31. Newman AM et al (2015) Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 12:453
32. Thorsson V et al (2018) The immune landscape of cancer. *Immunity* 48(4):812–830.e14
33. Ritchie ME et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43(7):e47–e47
34. Cesano A, Warren S (2018) Bringing the next generation of immuno-oncology biomarkers to the clinic. *Biomedicines* 6(1):14
35. Lehmann BD et al (2016) Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. *PLoS ONE* 11(6):e0157368
36. Loi S et al (2016) RAS/MAPK activation is associated with reduced tumor-infiltrating lymphocytes in triple-negative breast cancer: therapeutic cooperation between MEK and PD-1/PD-L1 immune checkpoint inhibitors. *Clin Cancer Res* 22(6):1499
37. Luen S et al (2016) The genomic landscape of breast cancer and its interaction with host immunity. *The Breast* 29:241–250
38. Edmund A, Moza JWR (2013) MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol* 49(3):211–215
39. Chowell D et al. (2018) Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* 359(6375):582–587
40. Cao K et al (2001) Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Hum Immunol* 62(9):1009–1030
41. Pitt JJ et al (2018) Characterization of Nigerian breast cancer reveals prevalent homologous recombination deficiency and aggressive molecular features. *Nat Commun* 9(1):4181
42. Chang C-S et al. (2018) Genomic analysis of racial differences in triple negative breast cancer. *Genomics.* <https://doi.org/10.1016/j.ygeno.2018.10.010>
43. Keenan T et al (2015) Comparison of the genomic landscape between primary breast cancer in African American versus white women and the association of racial differences with tumor recurrence. *J Clin Oncol* 33(31):3621–3627
44. Martin DN et al (2009) Differences in the tumor microenvironment between African-American and European-American breast cancer patients. *PLoS ONE* 4(2):e4531
45. Elkhany A, Katsuta E, Takabe K (2018) Racial disparity in breast cancer immune microenvironment. *J Clin Oncol* 36(15_suppl):1081–1081
46. Mu J et al (2014) Regulation of MHC class I expression by Foxp3 and its effect on regulatory T cell function. *J Immunol* 192(6):2892
47. Tavazoie MF et al (2018) LXR/ApoE activation restricts innate immune suppression in cancer. *Cell* 172(4):825–840.e18

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.