RESEARCH ARTICLE

African Ancestry–Associated Gene Expression Profiles in Triple-Negative Breast Cancer Underlie Altered Tumor Biology and Clinical Outcome in Women of African Descent

Rachel Martini[1,2], Princesca Delpe[3], Timothy R. Chu[4], Kanika Arora[4], Brittany Lord[1,2], Akanksha Verma[3], Deepa Bedi[5], Balasubramanyam Karanam[6], Isra Elhussin[7], Yalei Chen[8], Endale Gebregzabher[9], Joseph K. Oppong[10], Ernest K. Adjei[11], Aisha Jibril Suleiman[12], Baffour Awuah[13], Mahteme Bekele Muleta[14], Engida Abebe[14], Ishmael Kyei[15], Frances S. Aitpillah[10,15], Michael O. Adinku[15], Kwasi Ankomah[16], Ernest Baawuah Osei-Bonsu[13], Dhananjay A. Chitale[17], Jessica M. Bensenhaver[18], David S. Nathanson[18], LaToya Jackson[8], Lindsay F. Petersen[18], Erica Proctor[18], Brian Stonaker[1], Kofi K. Gyan[1], Lee D. Gibbs[19], Zarko Monojlovic[19], Rick A. Kittles[20], Jason White[6], Clayton C. Yates[7], Upender Manne[21,22], Kevin Gardner[23], Nigel Mongan[24,25], Esther Cheng[26], Paula Ginter[26], Syed Hoda[26], Olivier Elemento[3,27], Nicolas Robine[4], Andrea Sboner[26], John D. Carpten[19], Lisa Newman[1], and Melissa B. Davis[1,2,3,4,8]

Illustration by Bianca Dunn

ABSTRACT Women of sub-Saharan African descent have disproportionately higher incidence of triple-negative breast cancer (TNBC) and TNBC-specific mortality across all populations. Population studies show racial differences in TNBC biology, including higher prevalence of basal-like and quadruple-negative subtypes in African Americans (AA). However, previous investigations relied on self-reported race (SRR) of primarily U.S. populations. Due to heterogeneous genetic admixture and biological consequences of social determinants, the true association of African ancestry with TNBC biology is unclear. To address this, we conducted RNA sequencing on an international cohort of AAs, as well as West and East Africans with TNBC. Using comprehensive genetic ancestry estimation in this African-enriched cohort, we found expression of 613 genes associated with African ancestry and 2,000+ associated with regional African ancestry. A subset of African-associated genes also showed differences in normal breast tissue. Pathway enrichment and deconvolution of tumor cellular composition revealed that tumor-associated immunologic profiles are distinct in patients of African descent.

SIGNIFICANCE: Our comprehensive ancestry quantification process revealed that ancestry-associated gene expression profiles in TNBC include population-level distinctions in immunologic landscapes. These differences may explain some differences in race–group clinical outcomes. This study shows the first definitive link between African ancestry and the TNBC immunologic landscape, from an African-enriched international multiethnic cohort.

INTRODUCTION

Breast cancer is the most frequently diagnosed cancer among women globally and the leading cause of cancer-related death among women (1, 2). Despite having lower breast cancer incidence, mortality rates are among the highest across most sub-Saharan African nations, compared with other nations worldwide. Although poorer survival is typically attributed to advanced-stage disease at presentation and limited access to treatment options in lower-middle-income countries (LMIC; ref. 2), triple-negative breast cancer (TNBC) incidence rates across African nations represent approximately 33% of breast cancer diagnoses compared with less than 20% in other nations (3, 4), with the highest incidence of TNBC in West African nations compared with East African nations (3, 5, 6). Globally, overall breast cancer mortality and TNBC burden appear higher across the African diaspora at large, corresponding with a higher prevalence of TNBC disease among women with African ancestry (6), who reside in nations throughout Europe (7, 8), South Africa, and admixed African American (AA) populations in the United States (5, 9, 10). We previously reported a higher risk of TNBC, compared with other types of breast cancer, associated with West African ancestry (5, 6). Therefore, we hypothesized that there may be genetic drivers associated with West African ancestry that predispose and/or lead to aggressive breast cancer, including TNBC.

TNBC continues to have the worst prognosis of breast cancer subtypes and the worst survival outcomes due to a lack of targeted therapy options for these tumors (11, 12). Given TNBC incidence rates across the African diaspora, our efforts in oncologic anthropology have shifted to a molecular focus to uncover and characterize the influence of African ancestry on breast cancer disease etiology and progression (4, 13, 14). Previous comparative breast cancer studies among

[1]Department of Surgery, Weill Cornell Medical College, New York, New York. [2]Department of Genetics, University of Georgia, Athens, Georgia. [3]Englander Institute for Precision Medicine, Weill Cornell Medical College, New York, New York. [4]New York Genome Center, New York, New York. [5]Department of Biomedical Sciences, Tuskegee University, Tuskegee, Alabama. [6]Department of Biology, Tuskegee University, Tuskegee, Alabama. [7]Center for Cancer Research, Tuskegee University, Tuskegee, Alabama. [8]Department of Public Health Sciences, Henry Ford Health System, Detroit, Michigan. [9]Department of Biochemistry, St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia. [10]Department of Surgery, Komfo Anokye Teaching Hospital, Kumasi, Ghana. [11]Department of Pathology, Komfo Anokye Teaching Hospital, Kumasi, Ghana. [12]Department of Pathology, St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia. [13]Directorate of Oncology, Komfo Anokye Teaching Hospital, Kumasi, Ghana. [14]Department of Surgery, St. Paul's Hospital Millennium Medical College, Addis Ababa, Ethiopia. [15]Department of Surgery, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. [16]Directorate of Radiology, Komfo Anokye Teaching Hospital, Kumasi, Ghana. [17]Department of Pathology, Henry Ford Health System, Detroit, Michigan. [18]Department of Surgery, Henry Ford Health System, Detroit, Michigan. [19]Department of Translational Genomics, Keck School of Medicine, University of Southern California, Los Angeles, California. [20]Department of Population Sciences, City of Hope, Duarte, California. [21]Department of Pathology, University of Alabama at Birmingham, Birmingham, Alabama. [22]O'Neal Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, Alabama. [23]Department of Pathology and Cell Biology, Columbia University, New York, New York. [24]Biodiscovery Institute, University of Nottingham, Nottingham, United Kingdom. [25]Department of Pharmacology, Weill Cornell Medical College, New York, New York. [26]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, New York. [27]Institute of Computational Biomedicine, Weill Cornell Medical College, New York, New York.

Corresponding Author: Melissa B. Davis, Department of Surgery, Weill Cornell Medical College, 420 E. 70th Street, New York, NY 10021. Phone: 646-962-2855; Fax: 646-962-0023; E-mail: mbd4001@med.cornell.edu

Cancer Discov 2022;12:1–22

doi: 10.1158/2159-8290.CD-22-0138

patients of diverse race groups have focused on comparing tumors from AA and European American (EA) self-reported race (SRR) groups in the United States. Although there were inherent limitations of cohort size and heterogeneity of race, these approaches were useful in determining that broad biological differences do exist across diverse patient populations (15, 16). Some of these discoveries included race–group distinctions in genomic differences in frequencies of single-nucleotide variants (SNV; refs. 6, 17, 18), somatic tumor mutation signatures (19, 20), structural copy-number variations (CNV; ref. 21), and differences in DNA methylation patterns in both estrogen receptor–positive and estrogen receptor–negative tumors (22). Our work and that of others have uncovered racial differences in gene expression that revealed distinctions in immune response signatures, repeatedly across independent cohorts, implicating differences in the tumor microenvironment (TME; refs. 16, 23) as a possible cause of outcome disparities. The emerging promise of curative immunotherapies in overcoming treatment resistance in TNBC highlights an important opportunity to target the immune microenvironment. Given our previous and current findings that uncover race–group differences in tumor immune responses, these findings have increasing relevance to overcome disparities, particularly in regard to the potential of immunotherapies to improve treatment response (24, 25). However, there are limitations to using SRR in genomic studies, mainly due to complexity in genomic backgrounds of admixed groups.

Recently, our work was the first to use quantified genetic ancestry in admixed AA women to identify African ancestry–specific gene expression differences in TNBC tumors compared with EA women, which we also found had some overlap with SRR-associated gene networks (26). Of the African ancestry–associated genes, 48.1% were distinct from the SRR-associated genes, indicating the functional influence of the genetic ancestry background upon gene expression, apart from SRR alone. Similarly, a recent study by our collaborators characterized gene expression of TNBCs from the Bantu tribe from Kenya and found Bantu population–specific gene expression signatures as compared with TNBCs of AA and EA TNBCs (14). However, the implications of ancestry are still untested, lacking the inclusion of the contemporary and appropriate representative ancestry groups that are specific and relevant to the admixed patient groups.

Therefore, our current study utilized an African-enriched international cohort from the International Center for the Study of Breast Cancer Subtypes (ICSBCS; ref. 27), which will help to resolve a more precise understanding of genetic influences associated with African ancestry in race-associated gene signatures. We have measured the influence of African ancestry on TNBC tumor biology, derived from gene expression differences, that includes West African/Ghanaian and East African/Ethiopian women with TNBC compared with admixed AAs. Our rationale was based firmly on prior studies indicating that shared African ancestry harbors both the unique genetic risk of TNBC tumor etiology and the distinct gene signatures of TNBC among women of the African diaspora (6, 18, 26). We identified both African ancestry–associated gene expression signatures and TME cell-type differences from bulk RNA sequencing (RNA-seq) data. We demonstrate that the inclusion of native Africans with admixed AA patients, who share the same genetic ancestry, can overcome population complexity to help deduce the shared genetic drivers observed in race–group differences and discern these from environmental or other exogenous drivers of gene expression changes. We have identified subpopulation differences in gene expression between East versus West African ancestry lineages, which can be applied throughout population studies of the African diaspora in Europe (7, 8), the United States (5, 9, 10), and abroad (i.e., Afro-Latinx and Afro-Caribbean).

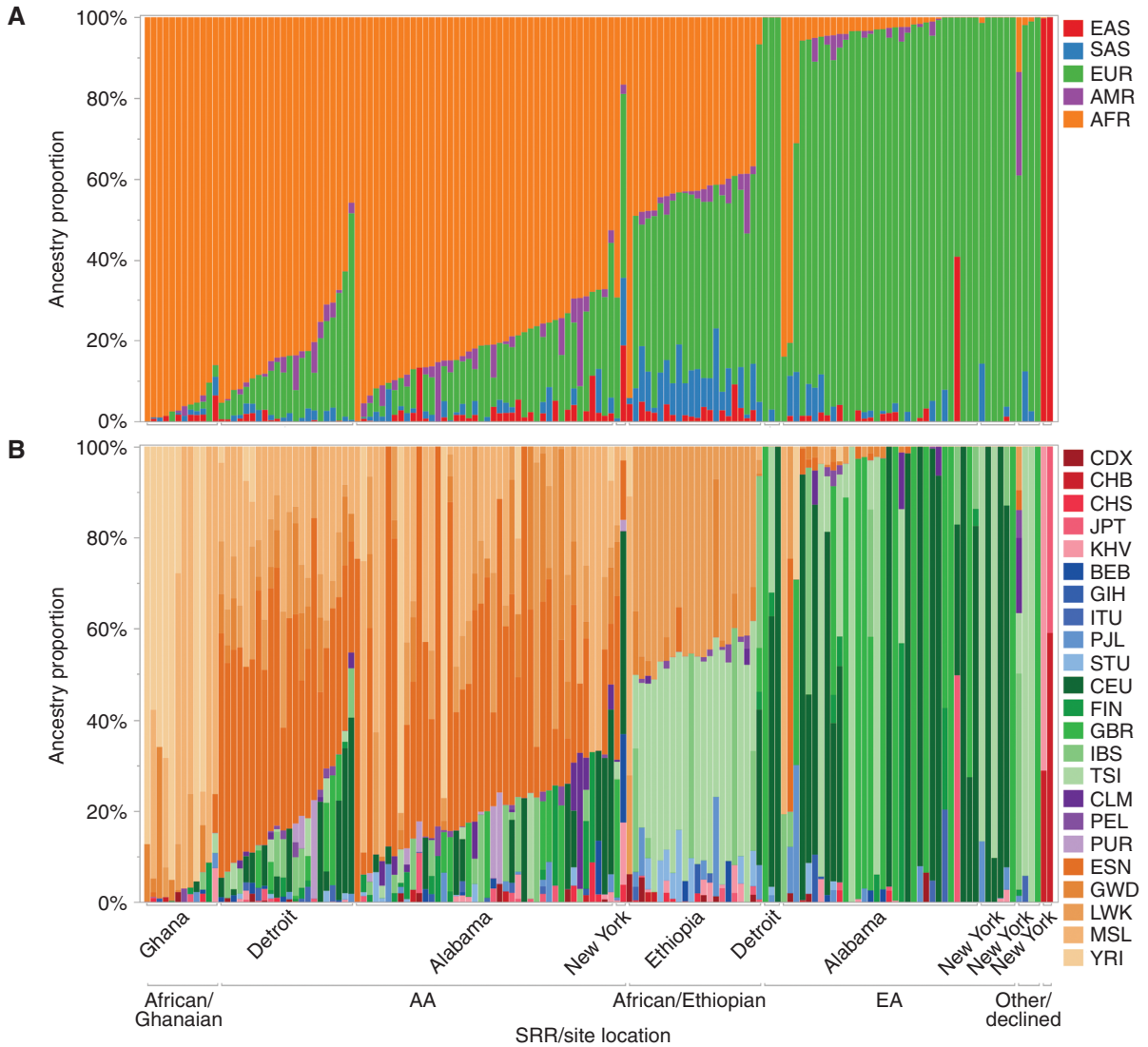
RESULTS

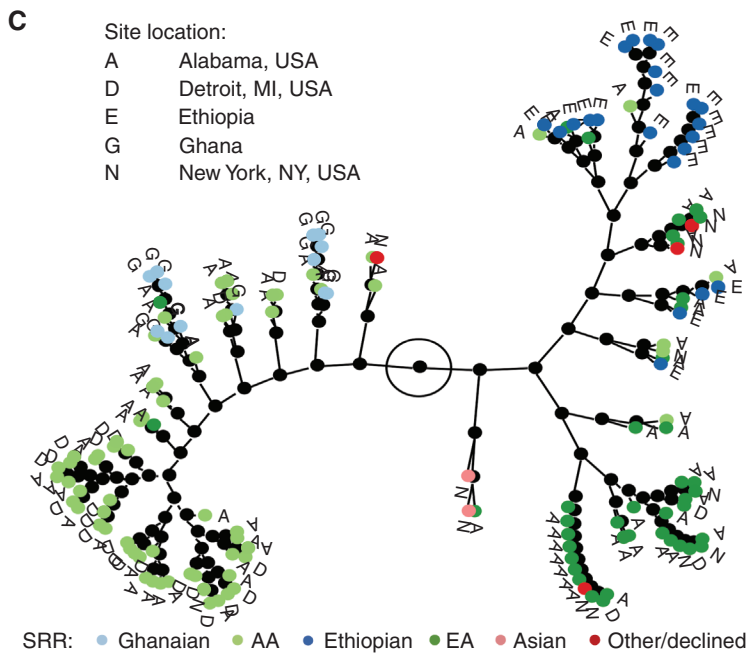
Characterization of Ancestry Profiles Reveals Complex Admixture in African and AA Cohort

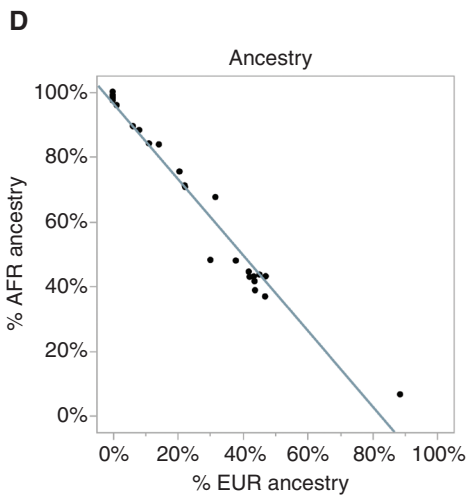
We estimated the global genomic ancestry for each patient in our cohort to determine the varying levels of admixture based on the 1000 Genomes superpopulation and subpopulations (Supplementary Table S1; ref. 28). Our cross-sectional set of 148 TNBC cases includes 66 AAs and 41 European American (EA), enriched with 13 West African Ghanaians, and 22 East African Ethiopians, with four individuals who declined to report SRR (Supplementary Fig. S1). African (AFR) ancestry comparisons indicated significant differences in African ancestry across our cohort (ANOVA $P < 0.001$), with Ghanaian patients having the highest levels of AFR ancestry (median 97.3%) and AAs having an average 15% less AFR ancestry (median 82.6%). Ethiopian patients had a surprisingly lower AFR ancestry (median 43.0%), with nearly equal amounts of European (EUR) ancestry (median 43.5%), which is consistent with previous anthropologic studies (refs. 29–31; Fig. 1A; Supplementary Table S2). Our EA patients generally showed exceptionally low levels of AFR ancestry (median 2.4%); however, three self-identified EA patients had between 30% and 80% AFR ancestry.

For more precise ancestry estimations that reflect regional origins, we estimated genetic ancestry for five African subpopulations, which includes four populations representing

Figure 1. Estimated genetic ancestry distribution in an African-enriched TNBC RNA-seq cohort. Genetic ancestry was estimated from genotypes of the ancestry-informed markers obtained from our RNA-seq alignments, in which we have superpopulation ancestry estimations, relative to the 1000 Genomes superpopulation populations (**A**), and subpopulation ancestry estimations for each individual in our cohort (**B**). In both **A** and **B**, each column represents an individual in the cohort, in which estimated ancestry from a given superpopulation or subpopulation is shown on the *y*-axis, and the *x*-axis is annotated by SRR and location. Superpopulation populations in **A** are East Asian (EAS, red), South Asian (SAS, blue), European (EUR, green), American (AMR, purple), and African (AFR, orange). Subpopulations in **B** are shown in variations of their corresponding superpopulation population color (i.e., AFR populations are in varying shades of orange), and population codes are reported in Supplementary Table S1. Samples are ordered by decreasing AFR ancestry [x-axis left to right: African/Ghanaian (Ghana), AA (Alabama, Detroit, New York), African/Ethiopian (Ethiopia), EA (Alabama, Detroit, New York), other/declined (New York), and Asian (New York)]. **C,** Constellation plot showing phylogeny of samples based on ancestry estimations. SRR of samples are indicated by the colored dots (Ghanaian = light blue, AA = light green, Ethiopian = dark blue, EA = dark green, Asian = light pink, and other/declined = dark pink). Site location of samples is annotated next to the colored dots (A = Alabama, USA; D = Detroit, MI, USA; E = Ethiopia; G = Ghana; and N = New York, NY, USA). **D,** Scatter plot showing inverse correlation of AFR and EUR ancestry in our gene expression cohort.







West African ancestry, including Esan in Nigeria (ESN), Yoruba in Ibadan, Nigeria (YRI), Gambian in Western Divisions in the Gambia (GWD), and Mende in Sierra Leone (MSL). There was only one population representing East Africa in 1000 Genomes—Luhya, in Webuye, Kenya (LWK; Fig. 1B; Supplementary Table S1). As anticipated, AA patients presented with AFR ancestry primarily of West African origin, which included ESN (median 36.1%) and MSL (median 19.7%) ancestry, with less than 10% estimated East African ancestry (LWK median 7.5%). Interestingly, the heterogeneity of African origin within AAs is more extensive and wide-ranging than the origin of African ancestry in Ghanaians or Ethiopians, in whom the amount of specific subpopulation ancestry can range from 0% to 90% for a given individual, indicating the complex diversity of African admixture in AAs. African patients' subpopulation ancestry is highly concordant with their regions of origin, in which Ghanaian ancestry is overwhelmingly enriched with the West African reference groups from YRI (median 66.0%), and MSL (median 24.1%) and Ethiopian patients have almost exclusively East African ancestry, represented as LWK (median 43.0%). Relatedness of patients' estimated genetic ancestry shows separation of Ghanaian and AA patients from Ethiopian and EA patients (Fig. 1C), and AFR and EUR ancestry were significantly inversely correlated among patients (Fig. 1D). Interestingly, the EUR ancestry in our Ethiopian patients was primarily Italian [Toscani in Italia (TSI), median 41.2%]. Ethiopian patients also showed substantial levels of East and South Asian ancestry (EAS median 1.9%, SAS median 9.0%), with more SAS compared with other SRR groups. All of these admixture revelations are consistent with the social histories of each SRR group and reflect the diverse complexity of ancestry across the African diaspora (29–33).

Influence of Ancestry in Gene Expression Profiles of TNBC Tumors Results in Ancestry-Associated Differential Immune Signatures

To investigate AFR ancestry–specific gene expression profiles in our ICSBCS TNBC samples, we isolated our analyses to patients with significant (>35%) AFR ancestry. As previously described (26), we performed gene-by-gene linear regression, using genetic ancestry as a continuous variable, which determined ancestry-associated gene expression. We identified gene signatures associated with AFR ($n = 613$) and EUR ($n = 345$) ancestry ($P < 0.001$), with 293 genes shared between these gene signatures (Fig. 2A; Supplementary Table S3). Given the significant inverse correlation of AFR versus EUR ancestry in our patient cohort (Fig. 1D), we compared the polarity of gene expression levels of the 293 overlapping genes and found that genes upregulated in association with AFR ancestry are conversely downregulated in association with EUR ancestry (Fig. 2B). This may represent genes that have expression drivers that are ancestral informative variants, which are isolated to certain ancestry groups.

Unsupervised hierarchical clustering of the 613 AFR-associated genes separated patients into two distinct clusters correlating with levels of AFR ancestry, which we denote as a low AFR cluster, including primarily Ethiopian TNBC patients, and a high AFR cluster, including primarily AA and Ghanaian patients (Fig. 2C). AFR ancestry is significantly higher among the high AFR cluster (median 86.13%, mean 85.87%) compared with the low AFR cluster (median 43.08%, mean 44.55%; $P < 0.0001$; Supplementary Fig. S2). The high AFR subcluster includes two subnodes: one representing ESN, MSL, and LWK ancestry (6/9 AAs and 1/6 Ghanaians) and a second representing YRI, GWD, and ESN ancestry (3/9 AAs and 4/6 Ghanaians; Fig. 2C, red asterisk; Fig. 2D, red box). The subnodes reflect differences in the origin of AFR ancestry composition observed among West Africans and AAs in our cohort.

We calculated AFR-associated genes (Fig. 2D) for functional pathway enrichment, using the $\log_2$ fold change between the distinct high AFR and low AFR clusters to measure differential expression (Fig. 2E). Top canonical pathways included some previously implicated processes in race–group comparisons, such as RNA posttranscriptional modification through spliceosomal cycle pathway enrichment ($P = 0.0002$, z-score = 3.804; ref. 34), cell-to-cell and extracellular matrix interactions in the integrin signaling pathway ($P = 0.004$, z-score = 0; ref. 35), and chronic inflammation in atherosclerosis signaling ($P = 0.006$, no z-score predicted; ref. 36). Upregulation of WNT family member genes drives enrichment in a colorectal cancer metastasis signaling pathway ($P = 0.004$, z-score = −0.302; ref. 37) and HOTAIR regulatory pathway ($P = 0.006$, z-score = −0.707). One of the top enriched functions identified was immune cell trafficking (*P* value range of subterms 0.0119–0.000502; Fig. 2F and G). Specifically, there was a predicted upregulation of signals relating to immune cell movement and migration, but conversely a predicted inhibition of signals relating to immune cell activation. This finding was of particular interest, given our previous findings related to DARC-regulated immune cell infiltration, which is associated with race groups (38).

Our results thus far establish AFR ancestry–associated differential gene expression in TNBC tumors; however, this differential regulation could be due to a diverse baseline biological context. Therefore, we investigated the expression of our core set of 613 AFR ancestry–associated genes in normal mammary tissue data using the Genotype-Tissue Expression (GTEx) cohort. Out of the 396 GTEx breast/mammary samples with ancestry, 47 were of AFR ancestry (Supplementary Fig. S3), and only 20 of those were females. We found that 17 of the 613 genes (2.7%) were significantly associated with AFR ancestry (Fig. 2H; Supplementary Table S4). Of these 17 genes, seven had a positive expression correlation with AFR ancestry in both normal and tumor, whereas the other 10 showed relatively opposite correlations of expression levels with AFR ancestry. Eight were positively correlated (upregulated) with respect to AFR ancestry in normal tissue, but negatively correlated in tumor, and two were negatively correlated in normal and positively correlated in tumor (Fig. 2H and I). Tumor subtype–agnostic survival analysis using The Cancer Genome Atlas (TCGA) BRCA cohort revealed lower expression of *AVPR2* and *CYBA* and higher expression of *AGMAT* and *SNORA53* show benefit among AA patients, but not EA patients, and this was significant for *CYBA* in AAs ($P = 0.0302$; Fig. 2I). Pathway enrichment analysis suggests that these 17 genes were related to the inflammatory response ($P = 2.5$ E-07) and growth of mammary tumor ($P = 7.3$ E-07) disease and function terms (Fig. 2J). This suggests that already established biological mechanisms, which are altered in the course

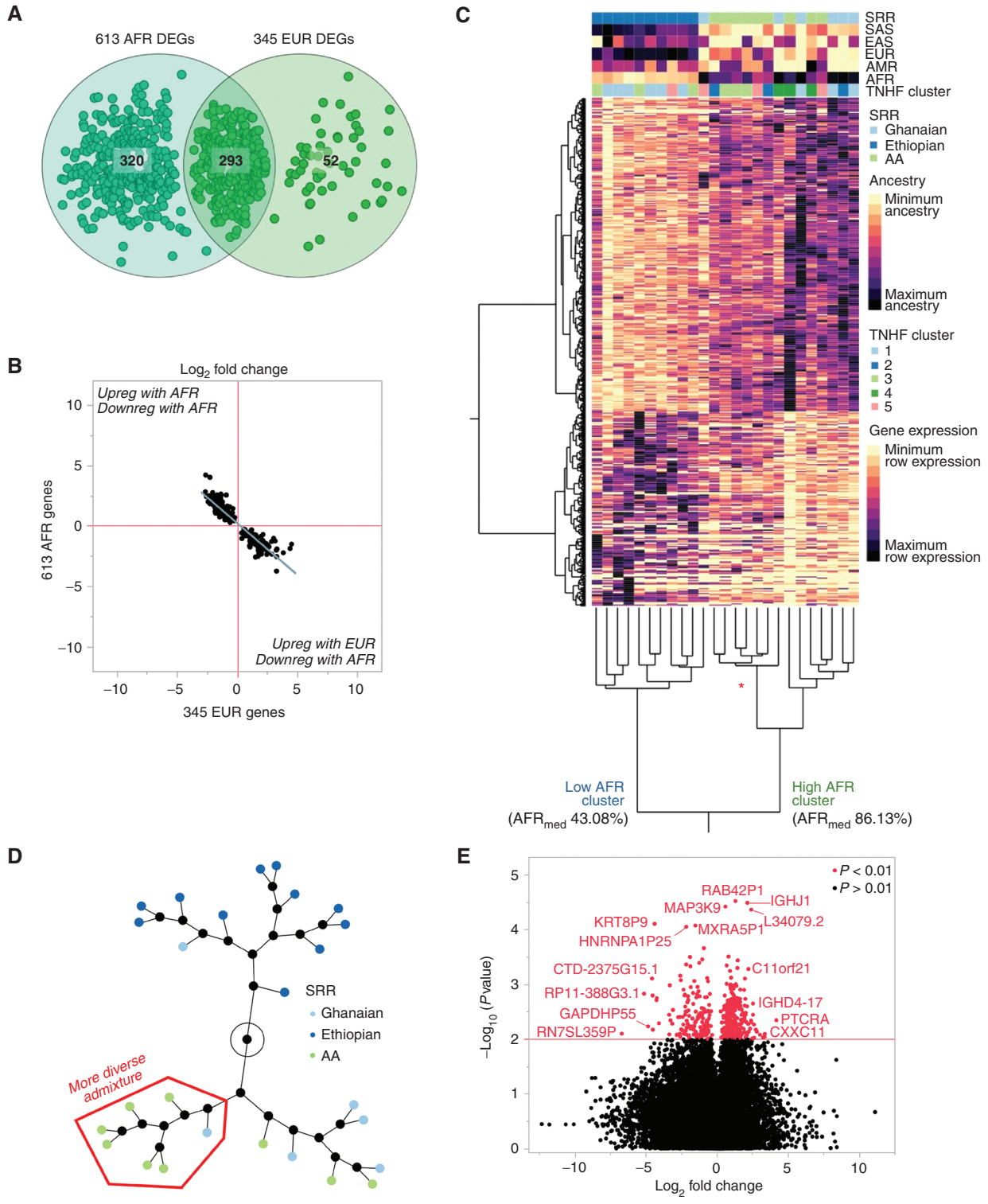


Figure 2. AFR ancestry–associated genes show enrichment in the immune response. **A,** Venn diagram of ancestry-associated genes identified from the AFR and EUR genetic ancestry linear regression model, in which ancestry was used as a continuous variable. DEG, differentially expressed gene. **B,** Scatter plot showing $\log_2$ fold change of 293 overlapping genes from AFR- and EUR-associated gene signatures. The top left quadrant represents those genes upregulated with increasing AFR ancestry (positive $\log_2$ fold change on the *y*-axis) and subsequently downregulated with increasing EUR ancestry (negative $\log_2$ fold change on the *x*-axis). **C,** Unsupervised hierarchical clustering of 613 AFR ancestry–associated genes. Columns represent individuals, where SRR, ancestry estimates, and TNBC subtypes are indicated in the colormap at the top of the heat map. Rows represent genes, where lighter yellow indicated minimum row expression and darker purple shows maximum row expression. AMR, American; TNHF, Triple-Negative Hetero Fluid. **D,** Constellation plot representing the nodal structure of individuals from **C**, where points are colored by SRR (Ghanaian = light blue, Ethiopian = dark blue, AA = light green). Node highlighted by red box indicates increased admixture node, which is highlighted in **C** by the red star. **E,** Volcano plot of AFR-associated genes, showing 613 significant genes in red. (*continued on next page*)

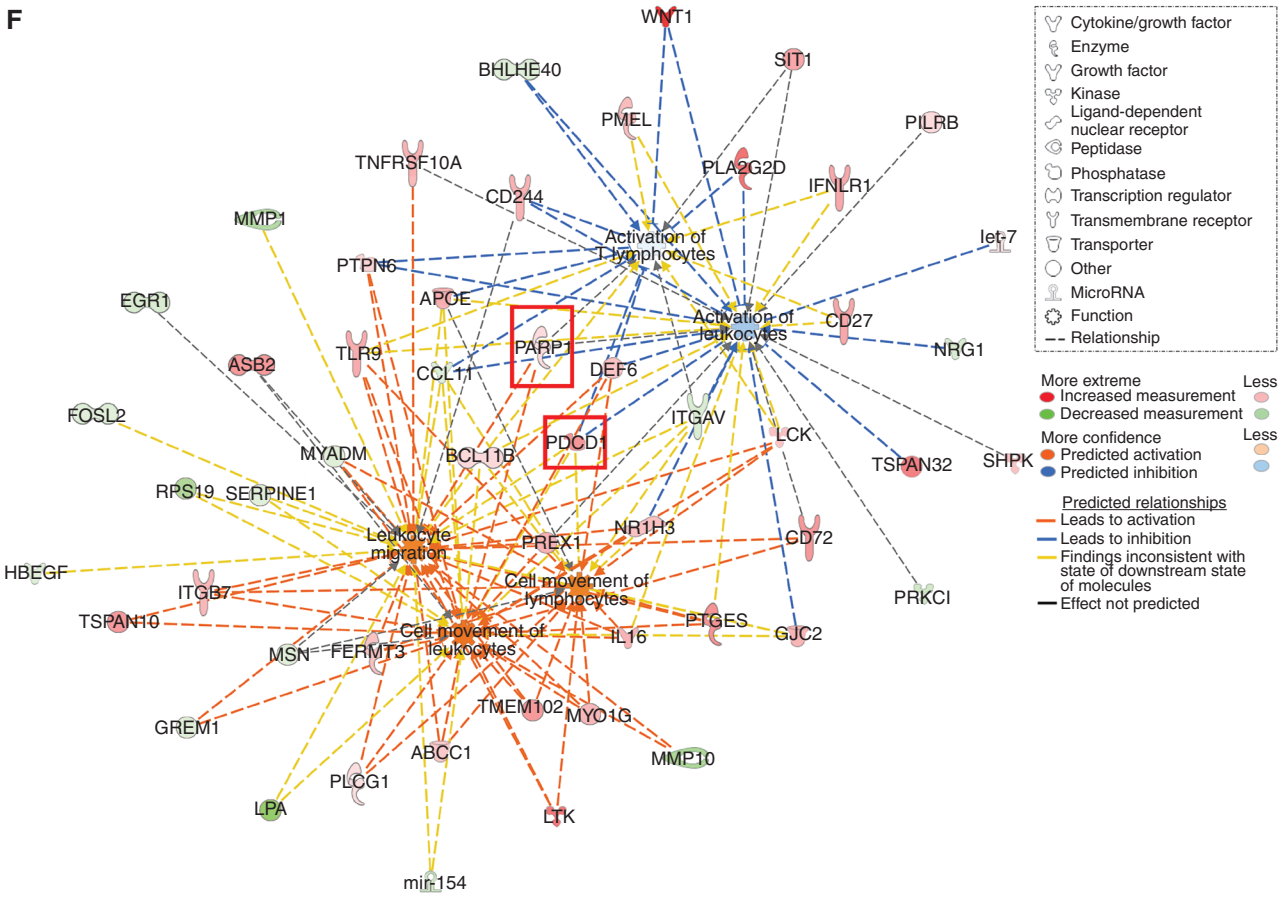


Figure 2. ***(Continued)*** Network (**F**) and treemap (**G**) diagrams of Ingenuity Pathway Analysis (IPA) immune cell trafficking disease and function terms, in which the 613 AFR gene signature was enriched (*P* value range of terms = 0.0119–0.000502). Genes in red or green are upregulated or downregulated among high AFR individuals, respectively. Genes and treemap boxes in orange represent a positive z-score (predicted activation), and those in blue represent a negative z-score (predicted inhibition). **H,** Parallel plot depicting positive or negative correlation (correl) of 17 overlapping AFR ancestry-associated genes in normal (GTEx) or TNBC tumor (ICSBCS) tissue. (*continued on following page*)

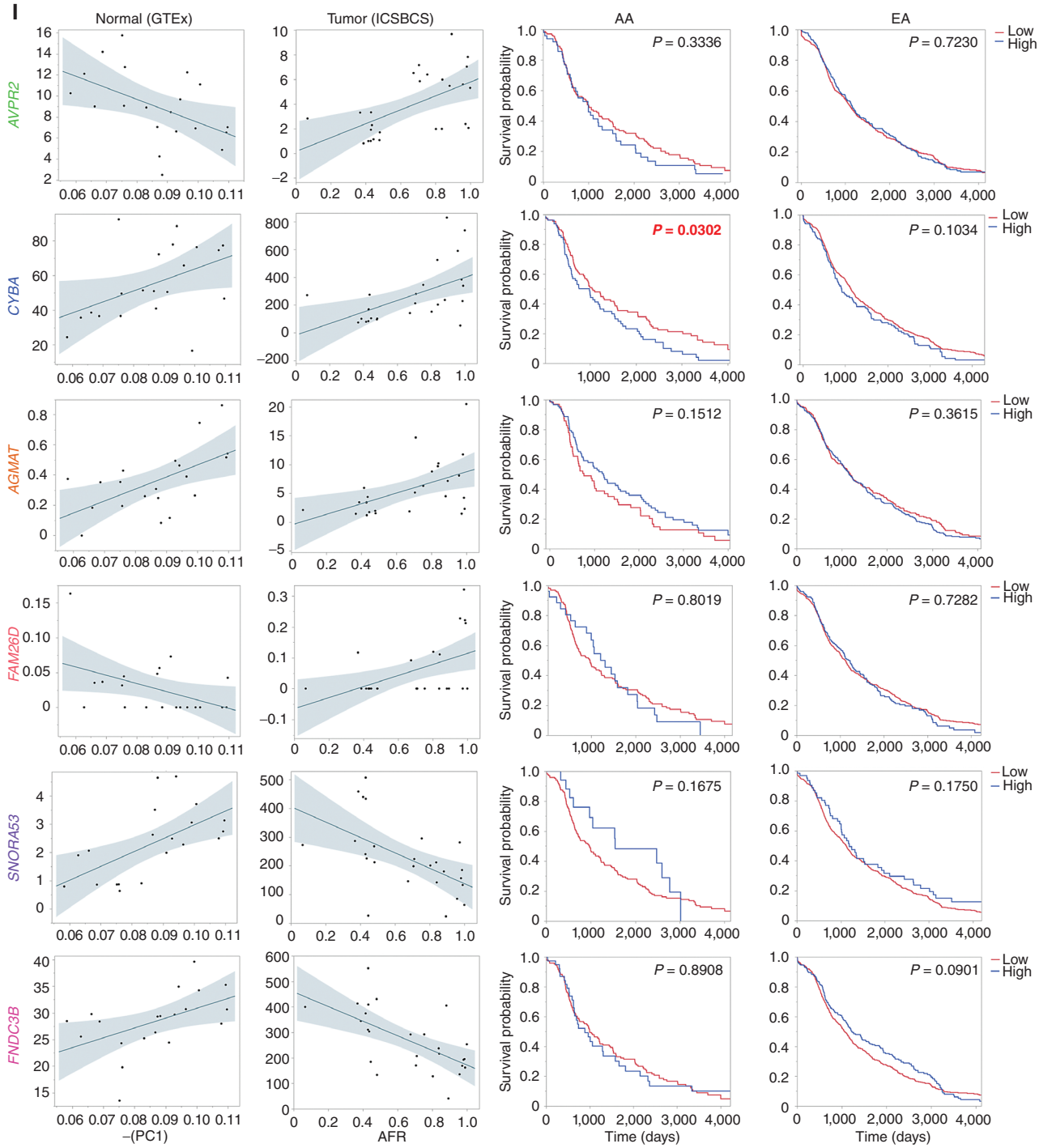


Figure 2. ***(Continued)*** **I,** Scatter plots of representative genes by ancestry in both normal (GTEx) and tumor (ICSBCS), and Kaplan–Meier survival plots in TCGA BRCA data among AA and EA patients with breast cancer. In the GTEx normal tissue cohort, AFR ancestry is increasing with increasing principal component 1 (PC1). The red *P* value highlights $P < 0.05$. (*continued on next page*)

of malignancy from normal to tumor, have divergent or more inflated regulatory drivers in populations of African descent.

Resolution of Subpopulation AFR Ancestry Influence on Gene Expression Signatures

We determined a higher resolution of African subpopulation origins to harness the shared genetic diversity and identify subpopulation-associated gene signatures. Specifically, we utilized the 1000 Genomes African subpopulation ancestry (Fig. 1) estimates for West African (YRI, ESN, GWD, and MSL) and East African (LWK) populations (Supplementary Table S1). By repeating the gene-by-gene statistical model with African subpopulations, we identified a combined 2,567 genes associated with the five African population groups

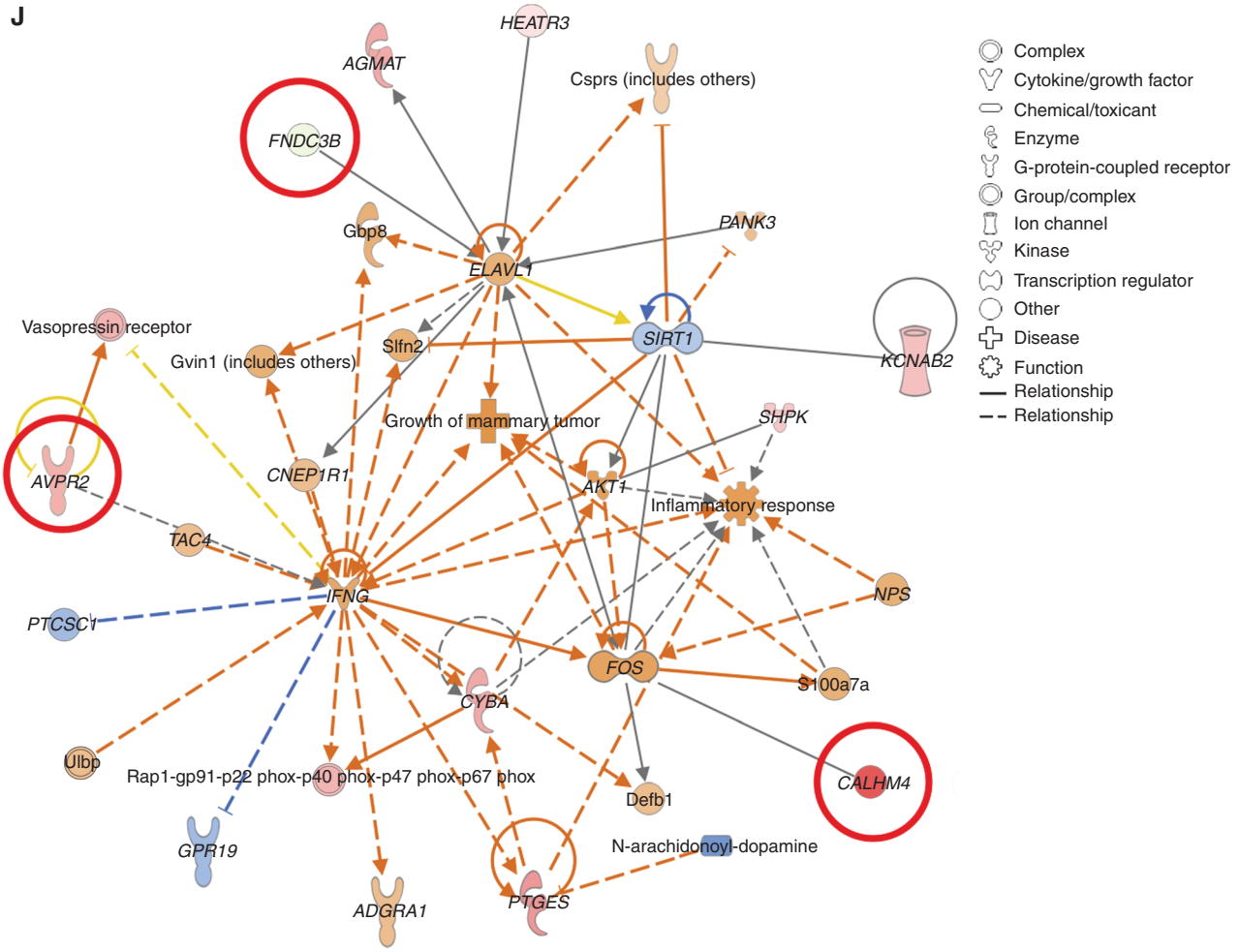


Figure 2. ***(Continued)*** **J,** Top *de novo* network from IPA of 17 overlapping AFR ancestry–associated genes. Color coding of expression is relative to the TNBC tumor (ICSBCS) tissue. Genes circled in red show opposite association with AFR ancestry in normal (GTEx) tissue (i.e., *CALHM4/FAM26D* is upregulated in TNBC tumor tissue, but negatively correlated with AFR ancestry in nondiseased breast tissue).

(Fig. 3A). African subpopulation–specific gene associations included 338 YRI genes, 643 ESN genes, 201 GWD genes, 146 MSL genes, and 1,229 LWK genes (Supplementary Table S5). These gene lists included, but extended beyond, the genes identified in the African superpopulation ancestry analysis (Fig. 3A). Surprisingly, there were no subpopulation-specific genes shared among all five populations, suggesting that there are unique gene expression drivers from each ancestry group. However, a small fraction of each set of individual West African subpopulation genes were shared with the East African LWK population (total $n = 210$). As may be anticipated, we found that the largest overlap of African subpopulation–associated genes was shared between ancestry groups that are geographically adjacent nations (29.0% of YRI and 48.8% of GWD shared between these populations). However, the closest West African groups, YRI and ESN of Nigeria, did not share any associated genes.

Therefore, we considered which SRR/nationality groups carried the specific subpopulation ancestry and therefore in which groups these gene signatures may be found. The East African population with the largest set of associated genes ($n = 1,229$), LWK, predominantly represented our Ethiopian patients with a small portion of AA ancestry represented by LWK (median ~8%). Pathway analysis predicted a decrease in immune response–related function in East African ancestry gene sets, including inhibition of CSF1 and various interleukins, CD28 and lymphopoiesis, and the canonical Tec kinase signaling pathway (Fig. 3B). This inhibitive effect of LWK ancestry on these functions clearly distinguishes the differences in tumor biology between west and East Africans and suggests important differences in immune cell development, response, and activation. MSL ancestry was predominantly found among AAs and Ghanaians and included genes ($n = 146$) that also involved the function of immune response signals that suggest activation of immune-related functions (Fig. 3C). Specifically, there was activation of REL and IL21, both playing important roles in immune response regulation. To validate our findings, we conducted a reanalysis of the AA subset of our previously published cohort and identified African subpopulation–associated gene signatures enriched for pathways that involve immune function (Supplementary Fig. S4A–S4H).

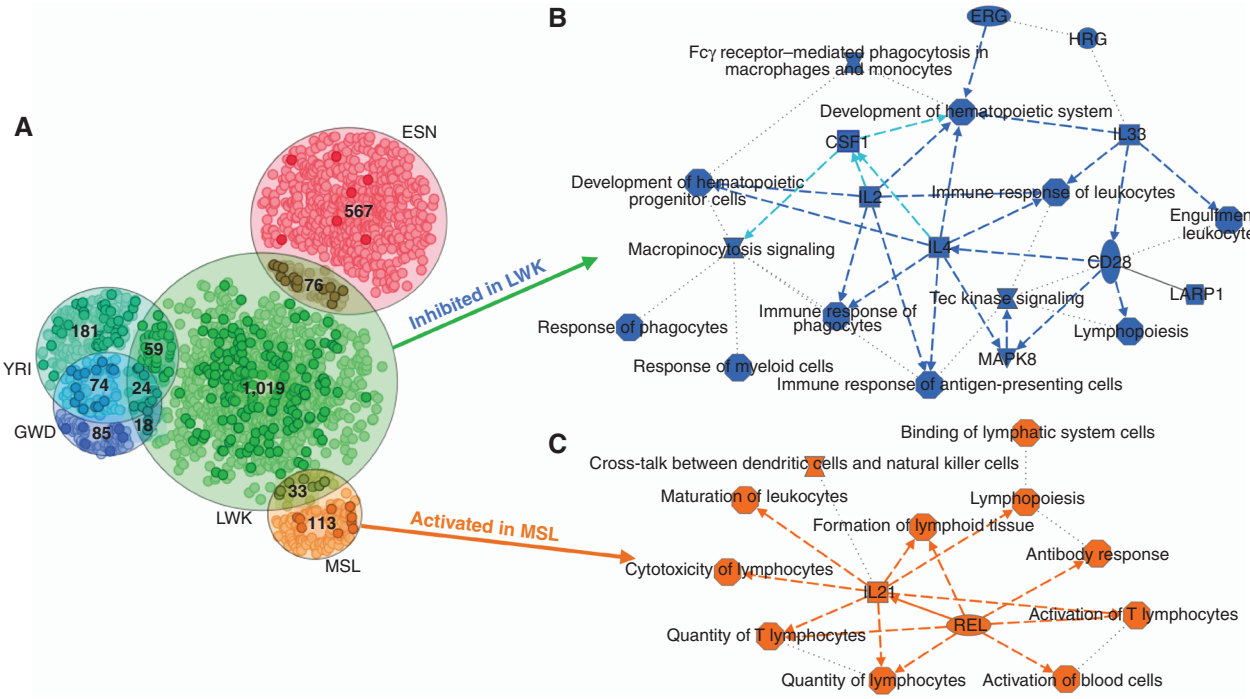


Figure 3. African subpopulation–associated genes are also enriched in the immune response. **A,** Venn diagram of unique and overlapping gene signatures associated with LWK, ESN, MSL, YRI, and GWD ancestry, respectively. Dots that are bolded are genes that overlap with the 613 AFR-associated gene signature. Ingenuity Pathway Analysis of LWK-associated (**B**) and MSL-associated (**C**) genes. **B,** Colors in blue indicate inhibition of regulators, disease/function terms, and canonical pathways among individuals with increasing LWK ancestry. **C,** Colors in orange indicate activation or regulators, disease/function terms, and canonical pathways among individuals with increasing MSL ancestry.

Immune Cell Enrichment Expression Signatures Are Associated with AFR Ancestry

We estimated immune cell populations and overall tumor-associated leukocyte (TAL) abundance with the deconvolution and cell-type enrichment methods CIBERSORTx (39) and xCell (40), respectively. Absolute scores, the sum of all estimated immune populations, were significantly higher in patients with high AFR ancestry compared with patients with low AFR ancestry (Fig. 4A; $P = 0.0076$). The immune cell types accounting for the bulk of the AFR ancestry–associated infiltrating cells included naïve B cells, CD8$^+$ T cells, helper T cells, regulatory T cells (Treg), and activated mast cells (Fig. 4B). Linear association testing showed that the increasing proportions of immune cells directly correlate with increasing AFR ancestry (Fig. 4C), suggesting a direct genetic dose response. The largest proportion of AFR-associated immune cells are naïve B cells, a nonactivated immune cell population. These findings concur with the preceding expression pathway analysis, indicating AFR ancestry–associated gene expression signatures are enriched for stimulated immune cell "migration/movement" (i.e., infiltration) and simultaneous repression of "cell-type activation" (i.e., naïve cells). Conversely, the "activated" cell population, such as "activated mast cell," is more prominent in tumors of patients with low AFR ancestry. To verify this finding with an independent algorithm, we used the xCell cell-type enrichment analysis (36). The results replicated the CIBERSORTx findings, showing AFR-associated immune cell infiltration, and further discerned that the specific T populations associated with AFR ancestry are CD8$^+$ T cells and CD8$^+$ T effector memory cells (Supplementary Fig. S5; $P < 0.05$), building on the observation of CD8$^+$ T cells from CIBERSORTx (Fig. 4C; $P < 0.05$).

To investigate immune-suppressive versus immune-stimulating TME marker associations (41) with African ancestry, we compared the relative expression of several well-known immune-checkpoint genes, including *CD274* (PD-L1 marker), *CTLA4*, and *PDCD1* (PD-1 marker; Fig. 4D). We found that *PDCD1* was significantly associated with AFR ancestry and SRR (ANOVA $P < 0.01$), with both Ghanaian (mean 12.42) and AA (mean 13.72) patients' tumors having 4× higher expression than Ethiopian patients' tumors (mean 3.68). To ensure these immunosuppressive marker gene expression patterns were derived from the immune cell population within the bulk tumor, we tested the correlation of specific immune cell estimates from CIBERSORTx with the *CTLA4*, *CD3D*, and *PDCD1* markers. We found each correlated with the abundance of relevant T-cell subtypes, indicating these cells are the likely source of the RNA expression (Fig. 4E).

To validate the RNA-seq–based immune cell estimations, we used two validation cohorts to quantify immune cell populations from protein-level data. Our first validation set represents clinical-grade IHC marker assays to score the infiltration of immune cells in an independent set of ICSBCS TNBC patients ($n = 40$), distributed across each ethnicity group represented in the RNA-seq cohort. We found similar immune cell infiltration trends across race/ethnic groups, with Ghanaian and AA tumors having higher counts of CD3$^+$

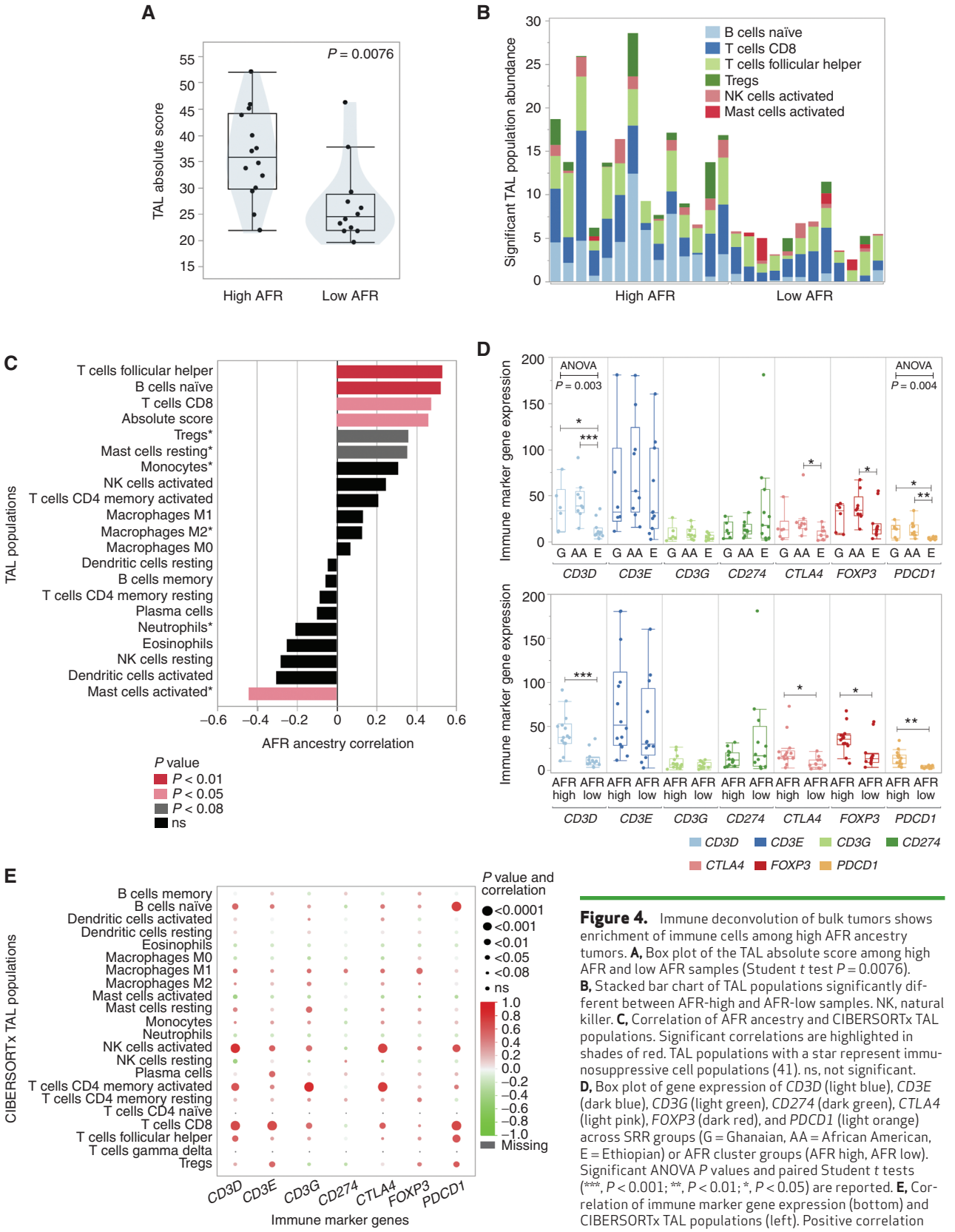


Figure 4. Immune deconvolution of bulk tumors shows enrichment of immune cells among high AFR ancestry tumors. **A,** Box plot of the TAL absolute score among high AFR and low AFR samples (Student *t* test $P = 0.0076$). **B,** Stacked bar chart of TAL populations significantly different between AFR-high and AFR-low samples. NK, natural killer. **C,** Correlation of AFR ancestry and CIBERSORTx TAL populations. Significant correlations are highlighted in shades of red. TAL populations with a star represent immunosuppressive cell populations (41). ns, not significant. **D,** Box plot of gene expression of *CD3D* (light blue), *CD3E* (dark blue), *CD3G* (light green), *CD274* (dark green), *CTLA4* (light pink), *FOXP3* (dark red), and *PDCD1* (light orange) across SRR groups (G = Ghanaian, AA = African American, E = Ethiopian) or AFR cluster groups (AFR high, AFR low). Significant ANOVA *P* values and paired Student *t* tests (***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$) are reported. **E,** Correlation of immune marker gene expression (bottom) and CIBERSORTx TAL populations (left). Positive correlation is shown in red, and negative correlation is shown in green. Size of the dot represents the significance of the correlations. (*continued on following page*)

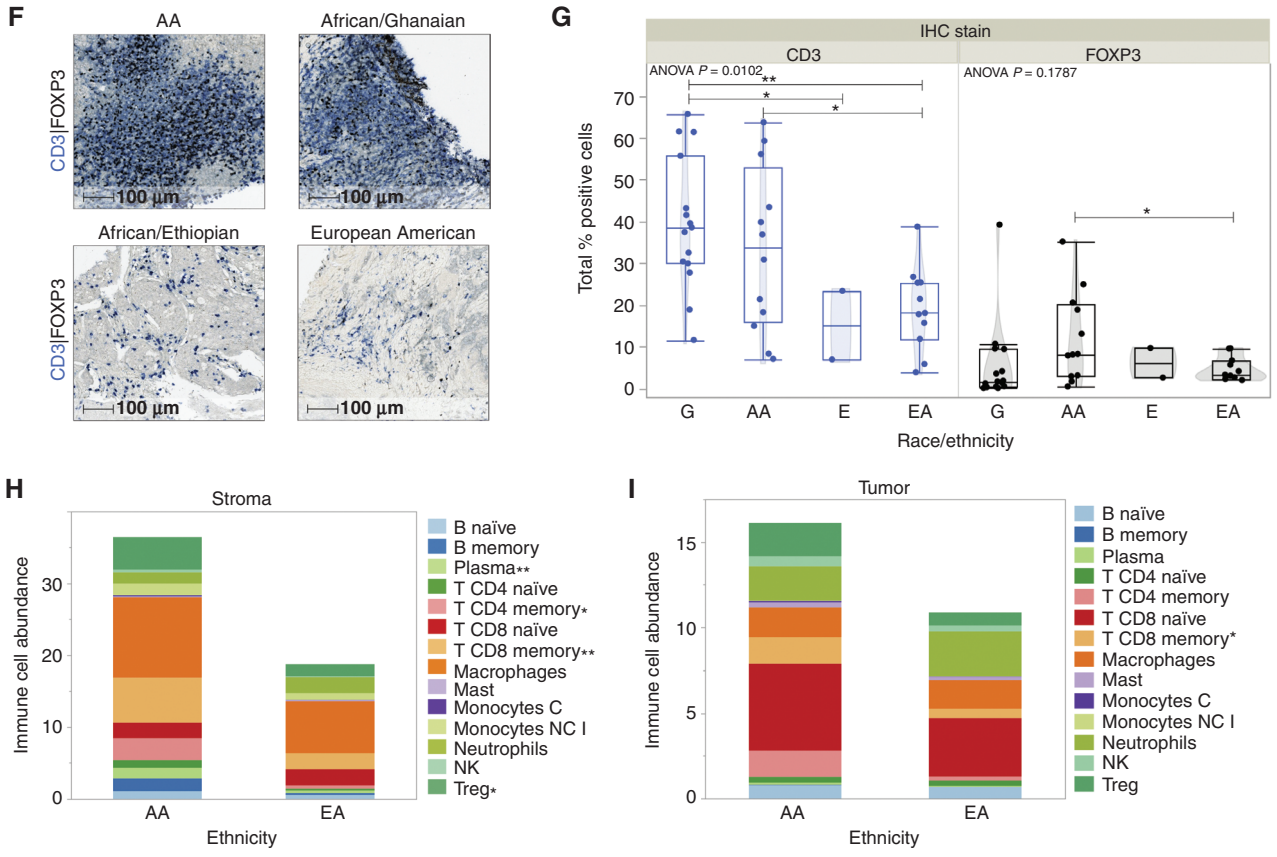


Figure 4. ***(Continued)*** **F,** Representative IHC images of CD3 (blue) and FOXP3 (black) staining in AA (top left), Ghanaian (top right), Ethiopian (bottom left), and EA (bottom right) TNBC cases. **G,** Box plots of percent positive CD3 (blue) and FOXP3 (black) stained cells from IHC images by SRR groups. ANOVA *P* values and paired Student *t* tests (**, $P < 0.01$; *, $P < 0.05$) are reported on the plot. Stacked bar charts of stroma (**H**) and tumor (**I**) segmented regions representing immune cell abundance from GeoMx staining data. Student *t* tests were conducted to determine immune cell populations with differential abundance between SRR groups and are reported on the plots (**, $P < 0.01$; *, $P < 0.05$). Monocytes C, monocytes classical; Monocytes NC I, monocytes nonclassical I.

and FOXP3$^+$ cells compared with Ethiopian and EA tumors (Fig. 4F). CD3$^+$ cells showed significant variation across all race/ethnic groups (ANOVA $P = 0.0102$; Fig. 4G), with significant pair-wise differences between Ghanaians and Ethiopians ($P = 0.0457$) and between AAs and EAs ($P = 0.0379$). Our second validation set represents a pilot cohort of multiplexed imaging data of TNBC patients representing AA ($n = 2$) and EA ($n = 2$) patients using the GeoMx platform. After completed segmentation across ~5 regions of interest (ROI) per patient, stromal and tumor cellular subsets were segregated, and immune cell abundances were determined. In the stromal compartment, we see significantly higher levels of plasma cells ($P = 0.0083$), CD4$^+$ and CD8$^+$ memory T cells ($P = 0.0344$ and $P = 0.0041$, respectively), and Tregs ($P = 0.0232$) among AA patients/ROIs compared with EA (Fig. 4H). In the tumor compartment, CD8$^+$ memory T cells were also significantly higher among AAs ($P = 0.0126$), and we also observed borderline significantly higher levels of Tregs ($P = 0.0601$; Fig. 4I).

FOXP3 expression in Tregs is correlated with a suppressive immune TME, suggesting patients with higher AFR ancestry may have a TME that is more suppressive versus stimulating compared with patients of EUR ancestry. The RNA-based findings in our cohort matched both IHC and multiplexed immunofluorescent protein staining among SRR groups in independent cohorts (Fig. 4D). We found that the most significant differences in immune cell infiltration were found when comparing populations by AFR ancestry rather than across SRR groups, emphasizing the importance of genetic ancestry in immunologic differences. Taken together, this further suggests an immune-suppressive tumor environment being associated specifically with West African ancestry, as opposed to East African or EUR ancestry.

TNBC Subtyping Reveals Ancestry Bias in Composition of Mosaic Heterogeneity

Gene expression signatures are used to categorize TNBC tumors into subtypes that have been shown to predict clinical outcomes (42, 43). The landmark report of these subtypes, the well-known Vanderbilt TNBCtype tool, correlated gene expression signatures of TNBC tumors to their tumor training set, which initially designated tumors into six distinct subtypes, including basal-like 1 (BL1), basal-like 2 (BL2), luminal androgen receptor (LAR), mesenchymal (M), mesenchymal stem–like (MSL), and immunomodulary (IM). An additional "subtype" category of exclusion harbors any tumors with "unsure calls" (UNS), which masked tumors with either multiple subtype correlations or no positive

correlations with any of the established phenotypes from the training set. Further consideration of the histologic context of these tumor subtypes determined that the MSL and IM subtypes are stromal and tumor-associated immune-derived rather than distinct phenotypes, which are currently corrected by the tool to be manually reassigned with their secondary correlation call (43). Correlated subtypes called from the Vanderbilt TNBC tool found tumors in our cohort from Ghanaian and Ethiopian patients were more often BL1, and tumors from AA patients were more often IM subtype (Fig. 5A, top). Interestingly, all IM tumors were from the high AFR ancestry Ghanaian and AA individuals, indicating the strong tumor immune signatures in these ancestry groups. After the suggested reassignment of the IM and MSL subtypes, AAs had a predominance of UNS calls, indicating an unresolved heterogeneity that would not allow designation of a single subtype (Fig. 5A, middle). Therefore, to ascribe a biological phenotype to these tumors, we used a previously described median ranking (26), which excludes the confounding influence of the immune signature genes by only including the gene expression signatures of the BL1, BL2, M, and LAR TNBC subtypes. Our results indicate that BL1 is the predominant subtype for Ghanaians and that M is the predominant subtype for both AAs and Ethiopians (Fig. 5A, bottom).

Given the previously reported heterogeneity of TNBC tumors among AA patients (26), we used our previously established Triple-Negative Hetero Fluid (TNHF) subtyping method (26). In brief, by utilizing both TNBCtype correlations and our median ranks from ~4,000 genes, we are able to determine heterogeneous categories of mixed TNBC subtypes. This assignment of multiple subtypes in a tumor's composition more appropriately informs us of the diverse cell types present in each patient's tumor. Our method also considers the exclusion of certain subtypes to establish unique combinations of subtype composition. Unsupervised hierarchical clustering of these mixed subtype categories resolved into five distinct nodes (Fig. 5B and C). Cluster 1 is the largest node, composed of $BL1^+/M^+/BL2^-/LAR^-$ tumors, which happen to be derived from Ghanaians and Ethiopians. Cluster 2 is $BL2^+/M^-$ and is composed exclusively of AFR-high cases. Cluster 3 tumors are $BL2^+$ and M^+ tumors originating from AA and Ethiopian tumors. Cluster 4 includes $BL1^+/BL2^-$ with only AFR-high cases. Lastly, cluster 5 includes LAR^+ tumors derived from all of the patient groups (Fig. 5C–E). Interestingly, the tumors in clusters 2 and 4 would also be classified as IM (with one UNS) and are exclusively derived from high AFR patients (Fig. 5E). We investigated the immune deconvolution CIBERSORTx signatures of the IM/cluster 2 and 4 tumors and found the TAL populations show enrichment in specific immune cell populations that included B cells, $CD8^+$ T cells, M2 macrophages, and natural killer (NK) cells (Fig. 5F). Overall, we find that the majority of these TNBC tumors in our African-enriched cohort are basal-like combined with mesenchymal, showing a distinction of BL1 among Ghanaians and Ethiopians ($BL1^+/M^+$) and BL2 among Ghanaians and AAs ($BL2^+$ and/or M^+; Fig. 5D). The clusters with high immune cell infiltration (2 and 4) are exclusively patients with substantial West African ancestry (Ghanaian and AA).

Genes Associated with SRR Are Involved in Comorbidity Pathways

In the interest of determining any distinct impact of racial social constructs on the biological phenotypes of tumors, we also investigated the functional pathway enrichment of SRR-associated genes that were not associated with genetic ancestry. We hypothesized that functional pathways of gene signatures associated with SRR would differ from our ancestry-associated gene signatures, noting that the social construct of race is not a comprehensively reliable assessment of factors related to the implicit bias and systemic racism correlated with SRR. Nevertheless, the SRR-associated comparison model identified 1,071 gene signatures as differentially expressed across SRR groups, and these were compared with our 613 AFR-associated genes and 345 EUR-associated genes (Fig. 6A). The overlap of ancestry-associated genes (AFR and/or EUR) with SRR genes included 320 genes, whereas 751 genes were uniquely associated with SRR. These distinctions of race- versus ancestry-associated signatures demonstrate the importance of both individual genetic ancestry and SRR on gene expression signatures in the context of tumor gene expression profiles.

Upon investigation of the 1,071 genes associated with SRR, unsupervised hierarchical clustering grouped patients by ancestry, where Ethiopian cases cluster distinctly from Ghanaian and AA cases (Fig. 6B), which is likely due to the influence of the overlapping ancestry-associated genes included in the analysis. However, unsupervised clustering of the 751 genes unique to the SRR analysis drastically changed the grouping (Fig. 6C), with AA cases diverging from the node containing all African cases. The African cluster also separated into two subnodes, but the divergence was largely driven by an upregulated gene signature found among AA and not seen among our African patients. We hypothesized that this signature represents distinct environmental influences that are unique to AA patients. Strikingly, signature pathway analyses determined that several known canonical pathways were enriched in AAs compared with Ghanaians and Ethiopians ($P < 0.05$), including several related to comorbidities that may reflect unknown comorbidity health status in our patient cohort. Specifically, comorbidity pathways related to cardiac function, adiposity/obesity, diabetes, and insulin signaling were found to be activated among AA patients, but not activated in Ghanaian and Ethiopian patients (Fig. 6D). This supports the premise of social determinants linked to racial constructs, which are known to be associated with these diseases. These comorbidities potentially influence the race-specific biological differences of TNBC microenvironments, which we detected in our cohort (Supplementary Fig. S6A and S6B). Although these comorbidities are known risk factors for negative outcomes, they could be addressed with interventions that target these pathways in tumors.

DISCUSSION

This study is the first comparative RNA-seq study of TNBC that utilized an African-enriched cohort of east and West Africans along with AAs to discern the influence of genetic ancestry on TNBC tumor biology related to racial disparities. Our findings support the emerging concept that the inclusion of multiethnic patient groups in genomic research increases

Figure 5. TNBC subtyping reveals the heterogeneity of tumors. **A,** Pie charts showing the distribution of TNBC subtypes across SRR groups for the TNBCtype initial call [Vanderbilt call (Vandy call), top row], TNBCtype call after removing/reassigning IM and MSL calls (middle row), and the calls using our median ranks method (bottom row). **B,** Heat map of correlations from the Vanderbilt TNBC subtyping tool and our median ranks calling for TNBC subtypes. Color map at the top indicates SRR/ethnicity, high or low AFR cluster sample, Vanderbilt TNBC subtyping call, Vandy call after removal of IM/MSL, and our median ranks call. Samples are clustered into five groups, which are color coded and labeled 1 to 5 on the dendrogram at the bottom. **C,** Line plot depicting positive and negative correlations with the Vanderbilt tool and the median-ranking subtype calls in each of the TNHF clusters. (*continued on next page*)

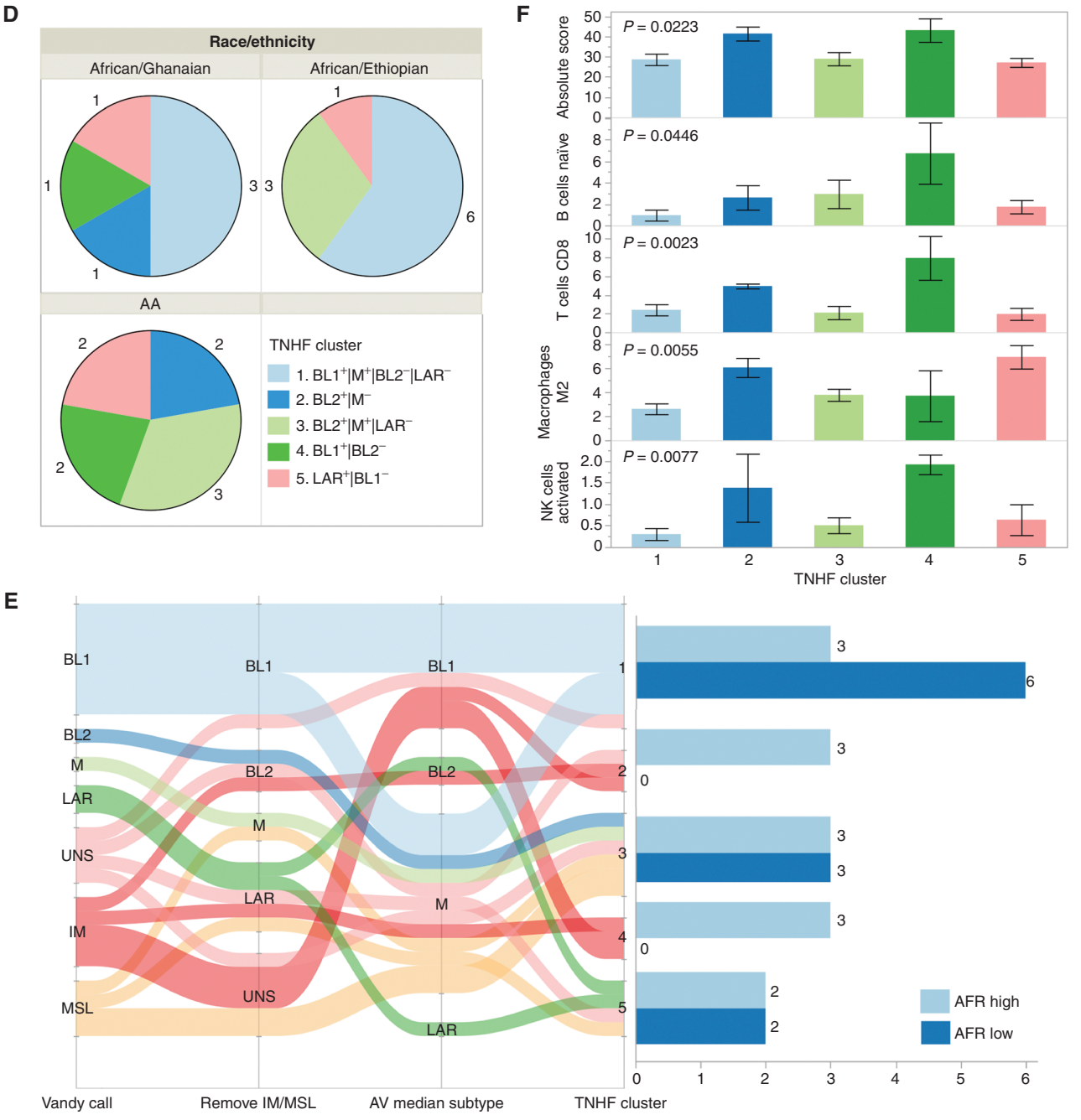


Figure 5. ***(Continued)*** **D,** Pie charts showing the distribution of TNHF clusters across SRR groups. **E,** Sankey plot showing the distribution of calls from initial Vanderbilt TNBCtype results to Vanderbilt call after removal of IM/MSL, to our median ranks method, and to the final TNHF clusters from **B**. Color coding is based on the initial Vandy Call (left). Bar chart to the right shows the number of tumors from AFR high or AFR low in a given cluster. **F,** Stacked bar chart of CIBERSORTx TAL populations in each of the TNHF clusters.

rigor and can have a transformative impact on cancer discoveries and overcoming disparities. Previous approaches have attempted to uncover the causal variables of disparate mortality and TNBC incidence but used SRR as a proxy for genetics or used global ancestry threshold cutoffs to categorize patient groups, which missed the broad range of genetic admixture we have detected here (AFR ancestry ranging from 16.64% to 99.99%) as well as the variability of ancestral origin among AAs due to unique social histories across the African diaspora (33). Even European admixture among AA patients varies regionally, where less European admixture was reported among AA individuals in the Southeast United States compared with the Northeast or Pacific Northwest (32, 33). The unique composition of patient ancestry from the African diaspora in our international ICSBCS cohort gave us a novel power and perspective to show that using linear regression models with estimated genetic ancestry as a continuous variable can unveil a distinct set of African ancestry–associated genes. The high level of

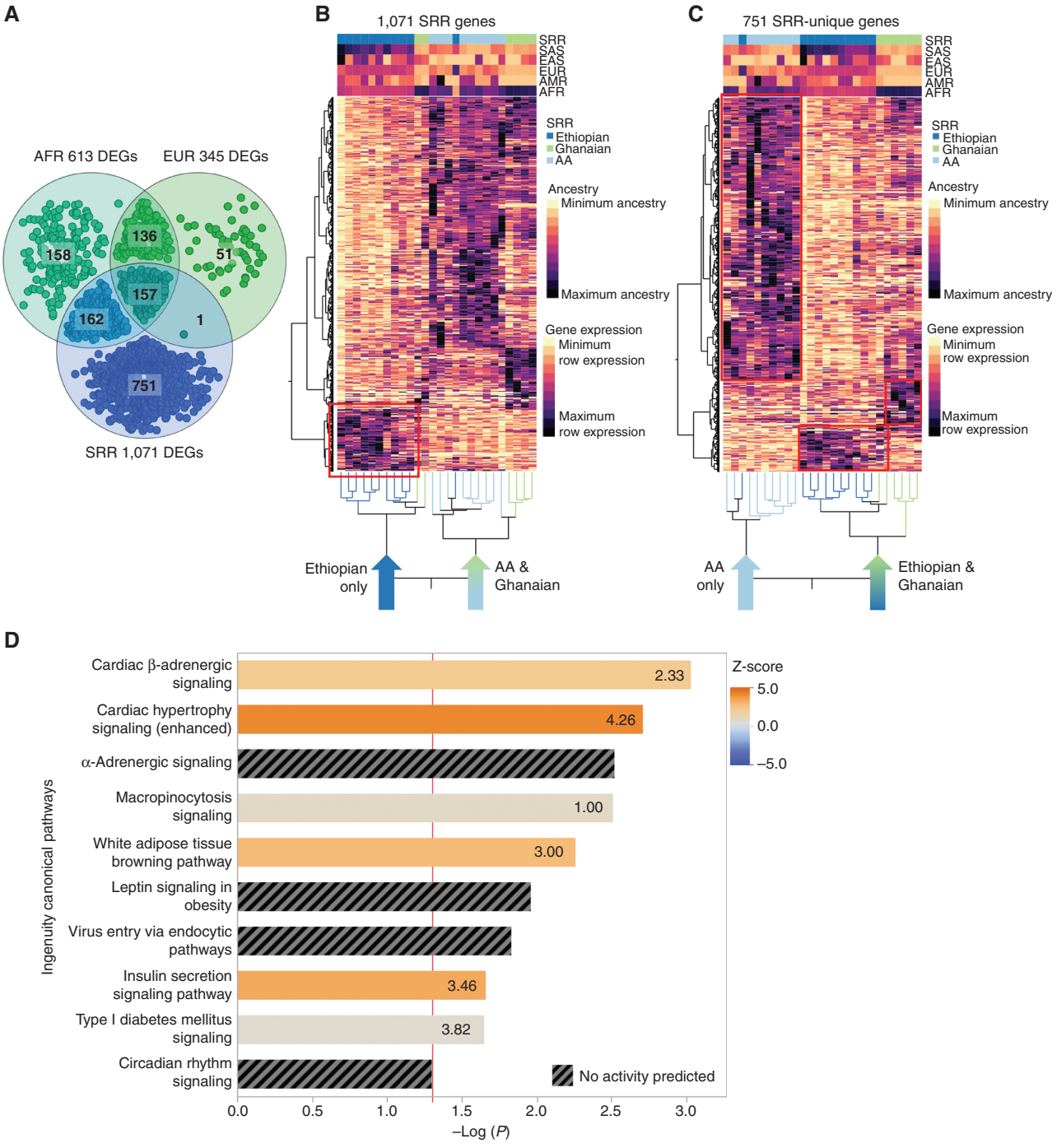


Figure 6. SRR-unique gene signature enriched in comorbid canonical pathways. **A,** Venn diagram depicting the overlap of AFR-, EUR-, and SRR-associated genes. DEG, differentially expressed gene. **B,** Unsupervised hierarchical clustering of the 1,071 SRR-associated genes. AMR, American. **C,** Unsupervised clustering of 751 genes unique to SRR. In both **B** and **C**, columns represent individuals, where SRR and ancestry are shown in the color map at the top, and rows represent genes. Node structure of individuals is shown at the bottom of the heat maps, where clustering was the individual node structure significantly changed between **B** and **C**. **D,** Comparing gene expression values from the node structure in **C**, we determined enrichment of genes in known canonical pathways that would be associated with environmental exposures and/or potential patient comorbidities. Z-scores indicated predicted activation (positive z-score, orange) or inhibition (negative z-score, blue) of the pathway based on the expression of the genes in the pathway in the directionality of AAs. Black striped bars indicated pathways where no z-score/predication was indicated due to insufficient evidence in the Ingenuity Pathway Analysis knowledge base. The red line indicates a *P* value cutoff of 0.05 [–log(0.05) = ~1.3].

European admixture we detected in our Ethiopian patients, nearly equal to AFR ancestry, combined with a significant proportion of SAS admixture is a key example of the power of our oncologic anthropology approach (13). The non-African ancestral origins in Ethiopians were previously reported in genetic anthropology studies as a significant proportion of mitochondrial DNA haplotypes (29) and Y chromosome haplotypes (30) with up to 50% of non-African ancestry (31).

Further, our subcontinental ancestry estimates revealed that although the predominant AFR ancestry of AA and Ghanaian patients is generally West African, the regional origin of African ancestry was distinct between these groups. Specifically, Ghanaians were primarily represented by YRI ancestry with less than 0.01% ESN ancestry, and AAs were primarily represented by ESN, with only two of 66 AAs reporting YRI ancestry over 30% (AA median YRI 0.00%). This is a relevant distinction for genome-wide association studies (GWAS) that attempt to identify AA-specific risk alleles utilizing a YRI reference genome for a genotype imputation template. Our work indicates that the YRI genetic background is less appropriate for our AA patients, and any genetic risk study that utilizes a single inappropriate AFR genome reference could adversely affect the relevance and rigor/reproducibility of the findings. Interestingly, the shared West African origins between Ghanaian and AA patients correspond with MSL ancestry (medians: 24.1% and 19.7%, respectively), which is rarely cited as a genome reference background.

We found that TNBC gene expression signatures associated with the superpopulation AFR ancestry estimates can be distinct from the signatures associated with regional/national AFR ancestry. Interestingly, no overlapping gene signatures were found between YRI and ESN, the predominant African origin of Ghanaians and AAs, respectively. Only YRI- and GWD-associated genes overlap among the West African population–associated genes. The lack of gene overlap suggests that there is a significant population-level divergence of genetic drivers that direct gene expression signatures even within African nations. This was surprising given that the YRI and ESN populations are geographically closer and presumably would have more similar genetic backgrounds that would likely lead to significant gene signature overlaps. However, the only source of ESN ancestry in our cohort is derived from AA patients. These findings all suggest that AFR subpopulations harbor functional/regulatory population-specific genetic drivers of gene expression that are relevant to disease pathology. Of note, AA patients also have unique environmental influences mediating genetic impact on gene signatures (36) compared with African patients, and a lack of gene expression signature overlap may represent the influence of mediating environmental factors.

The 600+ AFR-associated gene signatures effectively separated the node of Ethiopian patients from the Ghanaian and AA patients, revealing a broad distinction of tumor biological traits. The predominant functions of these genes are mechanisms of immune cell trafficking and activation of migration signals, and this was validated in independent cohorts with classic IHC methods, establishing a higher infiltration of tumor-associated leukocytes in the TME of patients with West African ancestry. Our findings are in agreement with previous studies that indicated higher inflammation and immune cell enrichment in race groups of African descent (16, 23). In addition, human evolution studies on general immunologic responses support our TNBC findings, where distinct immune expression signatures have been reported in response to pathogens, including differential immune cell activity between European and African populations (44, 45). We hypothesize that our previous observations, related to TME inflammation in the context of the African-specific Duffy-null blood group status (6, 18), may further account for these ancestry-specific clinical consequences. Interestingly, the heightened immune response was specifically associated with the MSL subgroup ancestry (Fig. 4C–E and Fig. 5B, respectively), implicating a shared MSL ancestral origin of our cohort as the likely source of a genetic factor that modifies immunologic responses. Further studies are needed to untangle the actual alleles that may be MSL-specific and functionally involved in immune responses.

Although these analyses primarily focused on the proof of principle of ancestry conveying an influence on tumor biology, cancer is well known to be influenced by the intersection of genetic and environmental variables (36, 46). Therefore, we also demonstrated the importance of a bimodal approach to cancer disparities research (47) that engages both social and biological variables. To investigate any social determinant effect, we modeled SRR associations with gene expression and were able to detect biological differences that were not associated with genetic ancestry. Although SRR alone does not capture certain social determinants, SRR-associated gene signatures involved canonical pathways that are known to be influenced by individual and area deprivation, suggesting SRR captures, at least in part, the influence of social structure on tumor biology. Our findings emphasize once again a need to assess social constructs and quantify the related racial discrimination practices, such as issues with marginalization and community redlining (48), to establish the impact of these practices on the underlying biology that determines treatment outcomes. It is equally important to note that beyond just characterizing the causal factors of disparities, finding novel biological traits presents an opportunity for therapeutic targets or interventions that incorporate these signatures in clinical treatment decision-making to improve outcomes.

Uncovering AFR-associated gene expression in normal breast tissue that overlaps with gene signatures in tumors indicates that cancer-related gene networks may have baseline differences across the diverse population even before tumors develop (20, 47). We have shown systemic differences in inflammation among patients with cancer and healthy controls that is associated with African-specific alleles (i.e., Duffy-null; ref. 38), which are related to both normal and tumor enrichment of inflammatory response gene expression. It is imperative to understand the extent of population differences in the regulation of gene network baselines that may ultimately sustain tumorigenesis pathways. Several of the baseline AFR-ancestry gene expression changes we detected in normal tissue were switched to opposing directions in tumor tissue. This finding provides preliminary evidence that the ancestry-specific differences in tumor expression could be initiated in response to malignancy as opposed to normal biological variation. For example, the *AVPR2* gene is negatively correlated with AFR ancestry in normal tissue but switches to be positively correlated with AFR ancestry in tumors, showing a population-specific pattern of drastic upregulation in TNBC tumors. The function of *AVPR2* was defined as maintaining homeostatic levels of water and electrolytes in renal cells but has been detected in multiple cancer types, including breast cancer (49), with both pro- and antimalignant actions

depending on the tumor type (49–52). The apparent clinical associations with survival trends appear to only emerge in AAs (Fig. 2I) compared with EA patients and further suggest population-private functionality that could be implicated in disease prognosis or novel therapeutic opportunities. Similarly, the *FNDC3B* gene, previously known as the factor for adipocyte differentiation, is an RNA binding protein that has been shown to be predominantly expressed in white adipose tissue and increases expression to play a role in early stages of adipocyte differentiation. It has a normal to tumor tissue expression transition from positive to negative correlation with AFR ancestry, respectively. This effective downregulation could represent an important loss of metabolic pathway regulation, as variants of *FNDC3B* have been GWAS hits in hemoglobin A1C measurements, body mass index, and waist–hip ratio. Interestingly, there is a higher expression of *FNDC3B* in TNBC (TCGA) compared with non-TNBC and persistently lower expression in AA compared with EA patients, with a subtype-agnostic association with survival that differs among race groups. Specifically, we observe better survival trending with high *FNDC3B* expression in EA patients, but low expression is associated with better survival in AA patients. Intriguingly, this suggests that SRR-associated gene network changes could be derived from distinct normal breast biology. However, one key limitation to these observations is the relatively low number of diverse patients in publicly available datasets with normal tissue expression.

As we continue to uncover the genetic regulation and environmental cues that influence TNBC differences across patient populations, our findings provide benchmarks for gene candidates of effective interventions to reduce mortality disparities and even cancer prevention. Ultimately, the unique tumor traits that hinge upon ancestry-associated gene expression signatures described here represent an important opportunity to fully characterize functional differences in tumor biology and opens a path to novel theragnostic options for these highly aggressive tumors.

METHODS

Ancestry Patient Cohort

ICSBCS Patient Cohort. The ICSBCS biorepository represents the efforts of an international consortium of breast cancer clinicians and researchers with the goal to characterize breast cancer disease in diverse populations worldwide. We have prospectively recruited patients with breast cancer since 2006, from whom formalin-fixed, paraffin-embedded (FFPE) tumor tissue has been collected. Across all institutions, written informed consent was obtained from the patients, and the work has been conducted in accordance with recognized ethical guidelines. Institutional Review Board (IRB) approval for the utilization of biorepository samples was obtained from participating sites in the United States (Weill Cornell Medical College, New York, NY; Henry Ford Health System, Detroit, MI; and University of Michigan, Ann Arbor, MI) and our international African partnering institutions (Komfo Anokye Teaching Hospital, Kumasi, Ghana, and the Millennium Medical College St. Paul's Hospital, Addis Ababa, Ethiopia). In the present study, TNBC tumor tissue was obtained from a total of 45 patients, including nine AAs, three EAs, 12 Ghanaians, and 21 Ethiopians (Supplementary Fig. S1). Confirmation of TNBC diagnosis by IHC was completed for Ghanaian and Ethiopian cases at our ICSBCS U.S. site locations in Michigan (University of Michigan, Henry Ford Health System) and New York (Weill Cornell Medical College). Samples collected in this cohort were used in both the ancestry (n = 61) and gene expression analyses (n = 26; Supplementary Fig. S1; Supplementary Table S6).

University of Alabama at Birmingham Patient Cohort. The University of Alabama at Birmingham (UAB) TNBC cohort has been previously described (26) and consists of a convenience cohort of retrospective FFPE TNBC tissue collected between 2000 and 2012 at the UAB. Samples were collected and used under the UAB IRB. In the present study, samples were analyzed from 74 patients, including 42 AA and 32 EA patients. Samples in the UAB patient cohort were used in ancestry comparisons and as a validation cohort for gene expression analyses findings (Supplementary Fig. S1).

Englander Institute for Precision Medicine Patient Cohort. All samples were collected and used under the Weill Cornell Medical College IRB. In the present study, we have estimated ancestry from TNBC tissue of 13 patients, including one AA, six EA, two Asian, and four patients who responded "other" or declined to provide race/ethnicity information (Supplementary Fig. S1).

RNA Extraction from Archival FFPE Tissue

RNA was extracted from archival FFPE tissue using a modified Qiagen RNeasy FFPE Kit protocol. Briefly, prior to deparaffinization of the FFPE tissue, the samples were incubated with 1× acidic antigen retrieval solution at 90°C for 5 minutes. Following incubation, samples were cooled to room temperature and any excess paraffin was removed from the tube. We then proceeded through the standard kit protocol. RNA yield was quantified using the Qubit RNA Broad Range kit and Qubit 4.0 fluorometer.

RNA Library Preparation and Sequencing

The quality of each RNA is assessed using RNA High-Sensitivity Screen on TapeStation (Agilent Technologies). For RNA sequencing, 100 ng of total RNA molecules were used to construct libraries using Illumina TruSeq RNA Exome Library Prep Kit following the manufacturer's protocols. The final libraries were then quantified using Agilent D1000 Screen Tape as well as sequenced on Illumina MiSeq V2 Micro Kit to assess insert sizes and integrity before sequencing on a high-throughput sequencer. Each library was normalized to 4 nmol/L and pooled and sequenced on an Illumina NextSeq500 High Output Kit (Illumina). All sequencing reads were converted to industry-standard FASTQ files using BCL2FASTQ (version 1.8.4).

RNA-seq Data Processing and Quality Control of Samples

Raw RNA-seq reads were assessed with Fast QC (version 0.11.8; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and Trimmomatic (ref. 53; version 0.36) was utilized for read trimming and adapter removal. Reads were aligned using HISAT2 (ref. 54; version 2.0.4) with the GrCh37 reference genome. Picard tools (version 2.18.3; https://broadinstitute.github.io/picard/) were used to pull alignment metrics for the samples, in which a number of sequenced reads were found to have high levels of read duplication. Duplicate reads were removed using Picard, and only samples that had 10M reads after deduplication were utilized in subsequent gene expression analyses.

Ancestry Estimation Using Variants Called from RNA-seq Alignments

Ancestry proportion is determined by the Admixture v1.3.0 (55) software, which uses a maximum likelihood–based method to estimate the proportion of reference population ancestries in a sample. We genotyped the reference markers generated from 1,964 unrelated 1000 Genomes project (28) samples directly on the RNA-seq samples using GATK pileup. Individuals from populations MXL (Mexican ancestry from Los Angeles in United States), ACB (African Caribbean

in Barbados), and ASW (African ancestry in the Southwest United States) were excluded from the reference due to being putatively admixed. The reference was further filtered by using only SNP markers with a minimum minor allele frequency of 0.01 overall and 0.05 in at least one 1000 Genomes superpopulation. Variants were additionally linkage disequilibrium–pruned using PLINK v1.9 (56) with a window size of 500 kb, a step size of 250 kb, and an r^2 threshold of 0.2, resulting in 122,377 markers remaining. The analysis results in a proportional breakdown of each sample into five superpopulations [AFR, American (AMR), EAS, EUR, and SAS] and 23 subpopulations (Supplementary Table S1).

Gene Expression Quantification and Differential Gene Analysis

Stringtie (ref. 54; version 1.3.3) was used to quantify gene expression from our deduplicated aligned reads. Quantified genetic ancestry and SRR groups were used to identify ancestry- or SRR-associated genes in our cohort, using linear regression analysis comparing gene expression either with the continuous ancestry variable or ANOVA analysis to determine associations with the categorical SRR variable. Genes with $P < 0.01$ were included in further analyses. Unsupervised hierarchical clustering of our gene lists was completed using JMP Pro 16 (SAS Institute Inc.).

Network Analyses of Differentially Expressed Genes

Ingenuity Pathway Analysis software (Qiagen, version 01-16) was used to determine the involvement of our gene lists in various canonical pathways, to determine upstream regulators, and to draw *de novo* networks involving our gene lists. For each analysis and gene list, the log fold change was calculated based on the resulting node structure of the samples when the gene lists underwent unsupervised hierarchical clustering, as our ancestry-associated versus SRR gene lists resulted in different clustering patterns of our samples.

Gene Expression Analysis in Nondiseased Breast/Mammary Tissue

To explore our AFR-associated gene signatures in normal mammary tissue, we obtained ancestry estimations and gene expression values [transcripts per million (TPM)] from the GTEx cohort (https://gtexportal.org/home/datasets). African ancestry was determined by principal coordinates as described by Gay and colleagues (57), where PC1 was associated primarily with AFR ancestry. Using PC1 <–0.04 as a threshold for AFR ancestry, we identified 47 individuals of AFR ancestry. Twenty of the 47 AFR ancestry individuals identified as female, and expressed the female-specific long noncoding RNA X-inactive–specific transcript and no genes from chromosome Y, and were used in subsequent analyses.

Survival Analysis of AFR Ancestry–Associated Genes in TNBC Tumors and Nondiseased Breast/Mammary Tissues

The TCGA BRCA cohort was used to determine if there was any prognostic benefit associated with our AFR ancestry–associated genes found in both TNBC and normal breast tissue (GTEx). Survival data (58) were obtained from https://gdc.cancer.gov/about-data/publications/pancanatlas. We used an upper quartile cutoff to separate patients into high- and low-expression categories, in which differences in survival outcomes were visualized by fitting Kaplan–Meier curves. *P* values from log-rank tests were reported where significant differences were found.

Tumor-Associated Immune Cell Abundance in Tumors Using RNA-seq Deconvolution and Enrichment Methods

To determine estimated abundance of tumor-associated immune cell populations, we used the online CIBERSORTx (39) platform (https://cibersortx.stanford.edu/) with our gene expression values as input. The LM22 signature matrix file was used as a reference, and the estimation was completed with quantile normalization disabled (as recommended for RNA-seq data) with 500 permutations. Only CIBERSORTx output that was determined to be significant ($P < 0.05$) was included in our analyses.

We have additionally used xCell for deconvolution of immune and other cell populations from our bulk RNA-seq data (40). Normalized TPM expression was used as input for the xCell algorithm.

IHC of CD3 and FOXP3

FFPE tumor blocks were obtained from the ICSBCS biorepository. Slide preparations were conducted through the Henry Ford Health System Histology Core using standard operating protocols. From the FFPE blocks, 4-μm sections were obtained. Multiplex staining was done using FOXP3 at a 1:100 dilution (BioLegend, cat. no. 320101) with CD3 predilute (Agilent, IR503) as the antibody diluent. Clinical attributes of the IHC cohort are reported in Supplementary Table S7.

Tumor-Infiltrating Leukocyte Analysis from IHC

Tumor-infiltrating leukocyte markers from multiplex IHC staining were analyzed using HALO software (V3, Indica Labs). Stained slides were electronically scanned using the Leica Aperio scanner and transferred into the HALO program. Positively stained tumor cells were annotated from hematoxylin and eosin staining and matched to a serial section with FOXP3 and CD3 multiplex staining. A custom algorithm optimized to detect color differences between the two markers was used to determine the number of positively stained cells for each marker. Positive tumor cells for each marker were divided by the total number of tumor cells and converted to a percent for subsequent data analysis.

Protein-Level Immune Cell Deconvolution of TNBC Cases

The NanoString GeoMx Digital Spatial Profiler was used to analyze immune profiles of four TNBC cases representing two AA and two EA patients. Patients were consented at Tuskegee University under IRB approval. FFPE tissue was sectioned and stained with fluorescently labeled antibodies specific for the epithelial cell marker PanCK, pan-leukocyte marker CD45, and macrophage marker CD68. Data from four to five ROIs were captured per patient, in which segmentation was performed on each ROI to distinguish the stromal and tumor compartments. Immune cell quantifications for each segmented stromal and tumor ROI were determined with the R/Bioconductor package spatialdecon (59).

TNBC Subtyping

To determine TNBC subtypes of our samples, we input gene expression values into the Vanderbilt TNBCtype online tool (https://cbc.app.vumc.org/tnbc/; ref. 42). The TNBC subtypes IM and MSL have been determined to primarily represent infiltrating immune cells and tumor-associated stroma, respectively, and therefore these calls are reassigned to their second most correlated call and significant call (43). UNS are where multiple correlations are significantly associated with a tumor gene expression profile, and in our cohort, these were able to be resolved after disregarding IM and MSL calls.

As a supplementary validation method to the gene expression correlation-based Vanderbilt TNBC classification tool, a summarized ranks measure was computed using the original TNBC subtypes signatures for all samples using normalized RNA-seq expression data. TNBC subtype signatures were obtained from Lehmann and colleagues (42). Across all samples, all genes expressed were ranked from low to high expression using the rank function in R statistical software, with a minimum rank method used to resolve duplicate expression ties. For each sample, ranks for each gene in the given subtype signature were extracted and a representative median of ranks for the gene signature was calculated to estimate the overall

regulation of the signature with respect to the total expression. The TNBC subtype signature with max median signature rank per sample was the assigned TNBC subtype for the sample.

Data Availability Statement

Sequencing data from our gene expression cohort can be accessed in the Gene Expression Omnibus (GSE211167).

Authors' Disclosures

C.C. Yates reports grants from the NCI (U54 CA118623), the NIH/National Institute on Minority Health and Health Disparities (NIMHD; U54-MD007585-26), and the Department of Defense (PC170315P1 and W81XWH-18-1-0589) during the conduct of the study; personal fees from Riptide Biosciences, QED Therapeutics, and Amgen and other support from Riptide Biosciences outside the submitted work; and a patent for peptides having immunomodulatory properties issued and licensed to Riptide Biosciences and Aurinia Pharmaceuticals and a patent for three peptides having anti-inflammatory properties issued to Riptide Biosciences. U. Manne reports grants from the NIH/NCI (5U54CA118948) during the conduct of the study. O. Elemento reports other support from Owkin, Freenome, and OneThree Bio, personal fees and other support from Volastra Therapeutics and Pionyr Immunotherapeutics, and personal fees from Champions Oncology during the conduct of the study. J.D. Carpten reports grants from Susan G. Komen for the Cure and the NCI during the conduct of the study and is a member of the Board of Directors for the American Association for Cancer Research. M.B. Davis reports grants from the NCI and Weill Cornell Medical College during the conduct of the study; grants from Genentech outside the submitted work; and a patent for inflammation and immune markers in cancer pending. No disclosures were reported by the other authors.

Authors' Contributions

R. Martini: Conceptualization, data curation, formal analysis, investigation, methodology, writing-original draft, writing–review and editing. **P. Delpe:** Formal analysis, investigation, methodology. **T.R. Chu:** Data curation, formal analysis, writing-original draft. **K. Arora:** Formal analysis, investigation, writing–review and editing. **B. Lord:** Formal analysis, validation, investigation, visualization, writing-original draft. **A. Verma:** Formal analysis, investigation, methodology, writing-original draft. **D. Bedi:** Data curation, methodology. **B. Karanam:** Data curation, methodology. **I. Elhussin:** Data curation, methodology. **Y. Chen:** Data curation, validation, investigation. **E. Gebregzabher:** Resources, investigation. **J.K. Oppong:** Resources, investigation, project administration. **E.K. Adjei:** Resources, investigation, methodology, project administration. **A. Jibril Suleiman:** Resources, investigation. **B. Awuah:** Resources, investigation, project administration. **M.B. Muleta:** Resources, investigation, project administration. **E. Abebe:** Resources, data curation, project administration. **I. Kyei:** Resources, data curation, investigation. **F.S. Aitpillah:** Resources, data curation. **M.O. Adinku:** Resources, data curation. **K. Ankomah:** Resources, data curation, investigation. **E.B. Osei-Bonsu:** Resources, investigation. **D.A. Chitale:** Resources, data curation, investigation, methodology. **J.M. Bensenhaver:** Resources, data curation, supervision, investigation, project administration. **D.S. Nathanson:** Resources, funding acquisition, investigation, project administration, writing–review and editing. **L. Jackson:** Resources, methodology. **L.F. Petersen:** Resources, investigation, writing–review and editing. **E. Proctor:** Resources, investigation. **B. Stonaker:** Resources, data curation, project administration. **K.K. Gyan:** Resources, project administration. **L.D. Gibbs:** Formal analysis, investigation, methodology. **Z. Manojlovic:** Formal analysis, supervision, investigation, methodology, writing-original draft. **R.A. Kittles:** Methodology, writing–review and editing. **J. White:** Resources, validation, investigation, methodology. **C.C. Yates:** Conceptualization, funding acquisition, validation, writing–review and editing. **U. Manne:** Resources, funding acquisition, validation, writing–review and editing. **K. Gardner:** Conceptualization, writing–review and editing. **N. Mongan:** Writing-review and editing. **E. Cheng:** Resources, data curation, investigation, methodology. **P. Ginter:** Resources, data curation, investigation, methodology. **S. Hoda:** Resources, data curation, investigation, methodology. **O. Elemento:** Resources, supervision, methodology, writing–review and editing. **N. Robine:** Data curation, software, formal analysis, supervision, investigation, writing–review and editing. **A. Sboner:** Data curation, formal analysis, supervision, investigation. **J.D. Carpten:** Resources, formal analysis, supervision, investigation, methodology, writing–review and editing. **L. Newman:** Conceptualization, resources, funding acquisition, project administration, writing–review and editing. **M.B. Davis:** Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing-original draft, project administration, writing–review and editing.

Acknowledgments

This work was supported by funding from Susan G. Komen (awarded to L. Newman) and R01 CA259396-01 (awarded to M.B. Davis). This work was also supported by U54-MD007585-26 (NIH/NIMHD), U54 CA118623 (NIH/NCI), and Department of Defense grants (PC170315P1 and W81XWH-18-1-0589) awarded to C.C. Yates. These studies were partly supported by 5U54CA118948 (NIH/NCI) and by institutional funds (Department of Pathology and School of Medicine of the UAB) awarded to U. Manne. We acknowledge the help provided by the UAB Tissue Biorepository Shared Facility grant of the UAB O'Neal Comprehensive Cancer Center, P30CA013148. This work was also supported by U54CA233465 for L.D. Gibbs and J.D. Carpten and P30CA014089 for L.D. Gibbs, Z. Manojlovic, and J.D. Carpten. This research was further supported by the University of Southern California Institute of Translational Genomics Keck Genomics Platform. We thank all the members of the International Center for the Study of Breast Cancer Subtypes, from the United States and Africa, for their dedication to our mission. We also extend our most sincere gratitude to all the patients and their families for their contribution and trust in this work. The GTEx Project was supported by the Common Fund of the Office of the Director of the NIH and by the NCI, the National Human Genome Research Institute (NHGRI), the National Heart, Lung, and Blood Institute (NHLBI), the National Institute on Drug Abuse (NIDA), the National Institute of Mental Health (NIMH), and the National Institute of Neurological Disorders and Stroke (NINDS). The data used for the analyses described in this article were obtained from the GTEx Portal in May 2022.

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

Note

Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

Received February 3, 2022; revised June 17, 2022; accepted August 23, 2022; published first September 19, 2022.

REFERENCES

1. Global Burden of Disease Cancer C, Fitzmaurice C, Akinyemiju TF, Al Lami FH, Alam T, Alizadeh-Navaei R, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2016: a systematic analysis for the Global Burden of Disease Study. JAMA Oncol 2018;4:1553–68.

2. Torre LA, Islami F, Siegel RL, Ward EM, Jemal A. Global cancer in women: burden and trends. Cancer Epidemiol Biomarkers Prev 2017;26:444–57.
3. Martini R, Newman L, Davis M. Breast cancer disparities in outcomes; unmasking biological determinants associated with racial and genetic diversity. Clin Exp Metastasis 2022;39:7–14.
4. Newman LA, Kaljee LM. Health disparities and triple-negative breast cancer in African American women: a review. JAMA Surg 2017;152:485–93.
5. Jiagge E, Jibril AS, Chitale D, Bensenhaver JM, Awuah B, Hoenerhoff M, et al. Comparative analysis of breast cancer phenotypes in African American, White American, and West Versus East African patients: correlation between African ancestry and triple-negative breast cancer. Ann Surg Oncol 2016;23:3843–9.
6. Newman LA, Jenkins B, Chen Y, Oppong JK, Adjei E, Jibril AS, et al. Hereditary susceptibility for triple-negative breast cancer associated with Western Sub-Saharan African ancestry: results from an international surgical breast cancer collaborative. Ann Surg 2019;270:484–92.
7. Copson E, Maishman T, Gerty S, Eccles B, Stanton L, Cutress RI, et al. Ethnicity and outcome of young breast cancer patients in the United Kingdom: the POSH study. Br J Cancer 2014;110:230–41.
8. Rapiti E, Pinaud K, Chappuis PO, Viassolo V, Ayme A, Neyroud-Caspar I, et al. Opportunities for improving triple-negative breast cancer outcomes: results of a population-based study. Cancer Med 2017;6:526–36.
9. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006;295:2492–502.
10. Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y, McCue P, et al. Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. Cancer 2007;110:876–84.
11. DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, et al. Breast cancer statistics, 2019. CA Cancer J Clin 2019;69:438–51.
12. Costa RLB, Gradishar WJ. Triple-negative breast cancer: current practice and future directions. J Oncol Pract 2017;13:301–3.
13. Davis MB, Newman LA. Oncologic anthropology: an interdisciplinary approach to understanding the association between genetically defined African ancestry and susceptibility for triple-negative breast cancer. Curr Breast Cancer Rep 2021;13:247–58.
14. Saleh M, Chandrashekar DS, Shahin S, Agarwal S, Kim HG, Behring M, et al. Comparative analysis of triple-negative breast cancer transcriptomics of Kenyan, African American and Caucasian Women. Transl Oncol 2021;14:101086.
15. Lindner R, Sullivan C, Offor O, Lezon-Geyda K, Halligan K, Fischbach N, et al. Molecular phenotypes in triple-negative breast cancer from African American patients suggest targets for therapy. PLoS One 2013;8:e71915.
16. Martin DN, Boersma BJ, Yi M, Reimers M, Howe TM, Yfantis HG, et al. Differences in the tumor microenvironment between African-American and European-American breast cancer patients. PLoS One 2009;4:e4531.
17. Huo D, Feng Y, Haddad S, Zheng Y, Yao S, Han YJ, et al. Genome-wide association studies in women of African ancestry identified 3q26.21 as a novel susceptibility locus for oestrogen receptor-negative breast cancer. Hum Mol Genet 2016;25:4835–46.
18. Martini R, Chen Y, Jenkins BD, Elhussin IA, Cheng E, Hoda SA, et al. Investigation of triple-negative breast cancer risk alleles in an international African-enriched cohort. Sci Rep 2021;11:9247.
19. Keenan T, Moy B, Mroz EA, Ross K, Niemierko A, Rocco JW, et al. 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 2015;33:3621–7.
20. Carrot-Zhang J, Chambwe N, Damrauer JS, Knijnenburg TA, Robertson AG, Yau C, et al. Comprehensive analysis of genetic ancestry and its molecular correlates in cancer. Cancer Cell 2020;37:639–54.
21. Chen Y, Sadasivan S, Ruicong S, Datta I, Taneja K, Chitale D, et al. Breast and prostate cancers harbor common somatic copy number alterations that consistently differ by race and are associated with survival. BMC Genomics 2020;13:116.
22. Espinal AC, Buas MF, Wang D, Cheng DT, Sucheston-Campbell L, Hu Q, et al. FOXA1 hypermethylation: link between parity and ER-negative breast cancer in African American women? Breast Cancer Res Treat 2017;166:559–68.
23. Davis M, Tripathi S, Hughley R, He Q, Bae S, Karanam B, et al. AR negative triple negative or "quadruple negative" breast cancers in African American women have an enriched basal and immune signature. PLoS One 2018;13:e0196909.
24. Perez-Garcia J, Soberino J, Racca F, Gion M, Stradella A, Cortes J. Atezolizumab in the treatment of metastatic triple-negative breast cancer. Expert Opin Biol Ther 2020;20:981–9.
25. Schmid P, Cortes J, Pusztai L, McArthur H, Kummel S, Bergh J, et al. Pembrolizumab for early triple-negative breast cancer. N Engl J Med 2020;382:810–21.
26. Davis M, Martini R, Newman L, Elemento O, White J, Verma A, et al. Identification of distinct heterogenic subtypes and molecular signatures associated with African ancestry in triple-negative breast cancer using quantified genetic ancestry models in admixed race populations. Cancers 2020;12:1220.
27. Jiagge E, Oppong JK, Bensenhaver J, Aitpillah F, Gyan K, Kyei I, et al. Breast cancer and African ancestry: lessons learned at the 10-year anniversary of the Ghana-Michigan Research Partnership and International Breast Registry. J Glob Oncol 2016;2:302–10.
28. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature 2015;526:68–74.
29. Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, et al. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. Am J Hum Genet 2004;75:752–70.
30. Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. Am J Hum Genet 2002;70:265–8.
31. Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Gallego Romero I, et al. Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am J Hum Genet 2012;91:83–96.
32. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. Am J Hum Genet 2015;96:37–53.
33. Kittles RA, Santos ER, Oji-Njideka NS, Bonilla C. Race, skin color and genetic ancestry. Calif J Health Promot 2007;5:9–23.
34. Al Abo M, Hyslop T, Qin X, Owzar K, George DJ, Patierno SR, et al. Differential alternative RNA splicing and transcription events between tumors from African American and White patients in The Cancer Genome Atlas. Genomics 2021;113:1234–46.
35. Slattery ML, John E, Torres-Mejia G, Stern M, Lundgreen A, Hines L, et al. Matrix metalloproteinase genes are associated with breast cancer risk and survival: the Breast Cancer Health Disparities Study. PLoS One 2013;8:e63165.
36. Dietze EC, Sistrunk C, Miranda-Carboni G, O'Regan R, Seewaldt VL. Triple-negative breast cancer in African-American women: disparities versus biology. Nat Rev Cancer 2015;15:248–54.
37. Ashktorab H, Daremipouran M, Devaney J, Varma S, Rahi H, Lee E, et al. Identification of novel mutations by exome sequencing in African American colorectal cancer patients. Cancer 2015;121:34–42.
38. Jenkins BD, Martini RN, Hire R, Brown A, Bennett B, Brown I, et al. Atypical chemokine receptor 1 (DARC/ACKR1) in breast tumors is associated with survival, circulating chemokines, tumor-infiltrating immune cells, and African ancestry. Cancer Epidemiol Biomarkers Prev 2019;28:690–700.
39. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol 2019;37:773–82.

40. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol 2017;18:220.
41. Salemme V, Centonze G, Cavallo F, Defilippi P, Conti L. The crosstalk between tumor cells and the immune microenvironment in breast cancer: implications for immunotherapy. Front Oncol 2021;11:610303.
42. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest 2011;121:2750–67.
43. Lehmann BD, Jovanovic B, Chen X, Estrada MV, Johnson KN, Shyr Y, et al. Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. PLoS One 2016;11:e0157368.
44. Yeyeodu ST, Kidd LR, Kimbro KS. Protective innate immune variants in racial/ethnic disparities of breast and prostate cancer. Cancer Immunol Res 2019;7:1384–9.
45. Nedelec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. Cell 2016;167:657–69.
46. Warner ET, Tamimi RM, Hughes ME, Ottesen RA, Wong YN, Edge SB, et al. Racial and ethnic differences in breast cancer survival: mediating effect of tumor characteristics and sociodemographic and treatment factors. J Clin Oncol 2015;33:2254–61.
47. Lord BD, Martini RN, Davis MB. Understanding how genetic ancestry may influence cancer development. Trends Cancer 2022;8:276–9.
48. Collin LJ, Gaglioti AH, Beyer KM, Zhou Y, Moore MA, Nash R, et al. Neighborhood-level redlining and lending bias are associated with breast cancer mortality in a large and diverse metropolitan area. Cancer Epidemiol Biomarkers Prev 2021;30:53–60.
49. Keegan BP, Akerman BL, Pequeux C, North WG. Provasopressin expression by breast cancer cells: implications for growth and novel treatment strategies. Breast Cancer Res Treat 2006;95:265–77.
50. Heidman LM, Peinetti N, Copello VA, Burnstein KL. Exploiting dependence of castration-resistant prostate cancer on the arginine vasopressin signaling axis by repurposing vaptans. Mol Cancer Res 2022;20:1295–304.
51. North WG. Gene regulation of vasopressin and vasopressin receptors in cancer. Exp Physiol 2000;85:27–40.
52. Sinha S, Dwivedi N, Tao S, Jamadar A, Kakade VR, Neil MO, et al. Targeting the vasopressin type-2 receptor for renal cell carcinoma therapy. Oncogene 2020;39:1231–45.
53. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20.
54. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 2016;11:1650–67.
55. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009;19:1655–64.
56. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
57. Gay NR, Gloudemans M, Antonio ML, Abell NS, Balliu B, Park Y, et al. Impact of admixture and ancestry on eQTL analysis and GWAS colocalization in GTEx. Genome Biol 2020;21:233.
58. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell 2018;173:400–16.
59. Danaher P, Kim Y, Nelson B, Griswold M, Yang Z, Piazza E, et al. Advances in mixed cell deconvolution enable quantification of cell types in spatial transcriptomic data. Nat Commun 2022;13:385.