# Association of Genetic Ancestry and Molecular Signatures with Cancer Survival Disparities: **A Pan-Cancer Analysis**



Kara Keun Lee<sup>1,2,3</sup>, Lavanya Rishishwar<sup>2,3,4</sup>, Dongjo Ban<sup>1,2,3</sup>, Shashwat Deepali Nagar<sup>2,3</sup>, Leonardo Mariño-Ramírez<sup>5</sup>, John F. McDonald<sup>1,2</sup>, and I. King Jordan<sup>1,2,3,4</sup>

## ABSTRACT

While overall cancer mortality has steadily decreased in recent decades, cancer health disparities among racial and ethnic population groups persist. Here we studied the relationship between cancer survival disparities (CSD), genetic ancestry (GA), and tumor molecular signatures across 33 cancers in a cohort of 9,818 patients. GA correlated with race and ethnicity but showed observable differences in effects on CSD, with significant associations identified in four cancer types: breast invasive carcinoma (BRCA), head and neck squamous cell carcinoma (HNSCC), kidney renal clear cell carcinoma (KIRC), and skin cutaneous carcinoma (SKCM). Differential gene expression and methylation between ancestry groups associated cancer-related genes with CSD, of which, seven protein-coding genes [progestin and adipoQ receptor family member 6 (PAQR6), Lck-interacting transmembrane adaptor 1 (LIME1), Sin3A-associated protein 25 (SAP25), MAX dimerization protein 3 (MXD3), coiled-coil glutamate rich protein 2 (CCER2), refilin A (RFLNA), and cathepsin W (CTSW)] significantly interacted with GA and exacerbated observed survival disparities. These findings indicated that regulatory changes mediated by epigenetic mechanisms have a greater contribution to CSD than population-specific mutations. Overall, we uncovered various molecular mechanisms through which GA might impact CSD, revealing potential population-specific therapeutic targets for groups disproportionately burdened by cancer.

Significance: This large-cohort, multicancer study identifies four cancer types with cancer survival disparities and seven

cancer-related genes that interact with genetic ancestry and contribute to disparities.



<sup>1</sup>Integrated Cancer Research Center, School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia. <sup>2</sup>School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia. <sup>3</sup>PanAmerican Bioinformatics Institute, Cali, Colombia. <sup>4</sup>Applied Bioinformatics Laboratory, Atlanta, Georgia. <sup>5</sup>National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, Maryland.

Corresponding Author: I. King Jordan, School of Biological Sciences, Georgia Institute of Technology, North Avenue, Atlanta, GA 30332. Phone: 404-385-2224; Fax: 404-894-0519; E-mail: king.jordan@biology.gatech.edu

Cancer Res 2022:82:1222-33

 ${\small @2022}$  The Authors; Published by the American Association for Cancer Research



Cancer is a complex genetic disease that disproportionately affects population groups in the United States. While cancer health disparities (CHD), or differences in disease burden and outcome between population groups, can be measured by indicators of morbidity, including disease incidence and prevalence, we focus on mortality in this study (1). Mortality is a uniquely important measure that captures the likelihood of an individual both developing and surviving cancer (2). Thus, mortality rates and survival probabilities are key indicators of disease progression that also reflect improvements made in cancer treatment. On the contrary, incidence and prevalence rates may not accurately reflect progress made against cancer, as they are influenced by overdiagnoses, or detection of tumors that do not cause symptoms or death, due to an increase in screening tests (3). Survival and mortality are direct measure of disease burden that are also clinically important for patients suffering from chronic, heavyburden disease such as cancer, increasingly being used in clinical studies as markers of disease prognosis and treatment efficacy (3, 4).

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

doi: 10.1158/0008-5472.CAN-21-2105

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 International (CC BY-NC-ND).

Although overall cancer mortality has steadily decreased in the last decades, it remains a major public health concern with over 600,000 estimated deaths nationally in 2020 (2, 5). The landscape of cancer mortality is further complicated by persistent inequalities across distinct race and ethnicity groups. The innovations and advancements made in cancer surveillance, diagnostics, and therapeutics that have resulted in an overall reduction of cancer incidence and deaths are undermined by the lack of progress made in reducing CHD, which in some cases have worsened over the years (5–7). For example, the survival gap between African American and White women diagnosed with breast cancer widened over the last 20 years, notwithstanding the overall decline in breast cancer mortality rate by 40% (8, 9). Investigations of the underlying factors that cause and exacerbate CHD in certain population groups are critical for promoting cancer health equity.

CHD between race and ethnicity groups have been thoroughly described in previous studies, with an emphasis on the contributions of nongenetic factors, including socioeconomic status, access to healthcare, and environmental exposures (10). However, CHD can remain even when group differences among nongenetic factors are controlled for, pointing to a potential role for genetic and biological factors in CHD. Recent studies have uncovered contributions of genetic differences, including pharmacogenomic variants and differential gene expression to CHD (2, 8, 11-14). The vast expansion of multiomics data for numerous cancer types, through projects such as The Cancer Genome Atlas (TCGA), has enabled deep genomic profiling of CHD across population groups, along with the use of estimated genetic ancestry (GA) instead of self-identified race and ethnicity (SIRE). GA is known to affect cancer survival disparities (CSD) for breast carcinoma, where African ancestry is associated with higher mortality, but the impact of GA on CSD for other cancers is largely unknown (15-17).

It should be stressed that GA and SIRE are distinct concepts, reflecting different aspects of human identity and biology. Race and ethnicity are socially defined, based on shared heritage, culture, and social experiences. More importantly, race and ethnicity are not biological or genetic categories, and considered to be a poor proxy for genetic diversity (10, 18). GA, on the other hand, is a characteristic of the genome based on correlated allele frequency differences among ancestral populations. GA can be inferred objectively and with precision, as a categorical or continuous variable, and it can be defined independently of the social dimensions of race and ethnicity. When GA is inferred as a continuous variable, it can be used to characterize patterns of admixture commonly seen among individuals from modern, cosmopolitan populations.

For this study, we performed a pan-cancer analysis of survival disparities between population groups using the TCGA data, and characterized the differences in GA-associated molecular signatures and their impacts on CSD. We identified the cancer types exhibiting significant disparity in overall survival (OS) outcomes of patients with cancer based on both SIRE and GA. Moreover, differential expression and mutational analyses on different cancer types and GA groups were used to identify molecular signatures affecting cancer survival that display significant interactions with ancestry. These findings suggest that GA and GA-associated features contribute to differential survival outcomes in several cancer types. The results also underscore the potential for population-specific therapeutic targets for groups disproportionately affected by cancer.

## **Materials and Methods**

## TCGA omics data curation and analysis

Affymetrix Human SNP Array 6.0 raw data for TCGA participants across 33 cancer types were downloaded in two formats from the Genomic Data Commons (GDC) Legacy Archive and harmonized with the 1000 Genomes Project (1KGP) reference population variants with the program PLINK (19, 20). GA was characterized at the continental level, via comparison with 1KGP surrogate ancestral source populations from distinct continental population groups, as both categorical and continuous variables. Categorical GA inference yields the most likely continental group assignments for individuals, whereas continuous GA inference yields percent contributions from different continental source populations, which we refer to as GA proportions. Continuous GA ancestry inference is used to capture patterns of admixture for TCGA participants. The continental GA proportions for African (AFR), East Asian (EAS), Native American (NAT), and European (EUR) were estimated with the program ADMIXTURE and were used to categorize participants into continental GA groups, AFR, EAS, Admixed American (AMR), and EUR, using the principal component analysis (PCA) and k-nearest neighbor classifier (k-NN; ref. 21).

Association of Ancestry and Molecular Signatures with CSD

Gene-level count files for RNA-sequencing data and somatic mutation annotation files (MAF) were obtained from the GDC Data Portal. For germline mutations, germline variant data files for 10,389 TCGA participants were downloaded from the GDC Legacy Archive. Differential gene expression and differential mutation analysis between GA groups were performed using the program DESeq2 and *mafCompare* function in the Maftools Bioconductor/ R-package, respectively (22, 23). For differentially expressed genes (DEG), enrichment analysis with the Hallmark Pathway gene sets in the Molecular Signatures Database (MSigDB, version 7.1) was performed. TCGA participant methylation data were taken from the Shiny Methylation Analysis Resource Tool (SMART) App. Additional details can be found in the Supplementary Methods.

## Statistical analysis

## Survival analysis

Clinical data for the all TCGA participants were downloaded from the GDC Data Portal and merged with survival outcome endpoints in the TCGA Pan-Cancer Clinical Data Resource (TCGA-CDR) for survival analysis using R version 3.6.1 (4). Univariable analysis using Kaplan-Meier method and multivariable modeling using Cox proportional hazard models were performed using the OS of TCGA participants. Kaplan-Meier curves were constructed and compared between each GA and SIRE groups using the log-rank test. Then for each cancer type, three types of general Cox proportional hazard models were constructed for: (i) SIRE groups, (ii) categorical GA groups, and (iii) continuous GA proportions. For four cancer types showing significant disparity in relative risk of mortality between GA groups in the univariable and survival analysis, cancer-specific multivariable Cox proportional hazard models were constructed with additional clinically relevant covariates including age at cancer diagnosis, American Joint Committee on Cancer stage, tumor morphology, tumor status at survival event, and gender. Other variables, such as molecular subtype, postmenopause status, and alcohol use, were also included based on data availability and completeness.

#### Differential molecular signature analysis

Identified DEGs were feature selected using the ANOVA and filtered for *F*-statistics >1 and  $P \le 0.05$ . Selected differentially expressed gene sets were additionally tested for classification performance of GA groups using elastic net logistic regression. Accuracy in classification rates were validated using 10-fold cross-validation and the area under the ROCs. Differential mutational rates (somatic and germline) between GA groups were evaluated using the Fisher exact test of independence (P < 0.05).

#### Gene-GA association & interaction analysis

Feature-selected DEGs and identified differentially mutated genes (DMG) were evaluated for significant association with GA using multiple linear regression modeling and logistic regression models, respectively. Genes with expression level or mutation status showing significant association with GA at  $\alpha = 0.05$  were selected for further analysis for gene-by-genetic ancestry (G × GA) interaction on cancer survival outcomes with an interaction term in the final Cox proportional hazard model. For genes displaying significant joint effect with ancestry, methylation level of associated CpG islands (CGI) were tested for differential methylation using the Wilcoxon rank sum test and also added to the G × GA interaction survival model. Additional details including model formulas can be found in the Supplementary Methods.

#### Data availability

The data analyzed in this study were obtained from 1KGP (http://www.internationalgenome.org/data/), TCGA Research Network (https://portal.gdc.cancer.gov/ and https://portal.gdc.cancer. gov/legacy-archive/), TCGA-CDR (https://gdc.cancer.gov/aboutdata/publications/PanCan-Clinical-2018), and the SMART App (http://www.bioinfo-zs.com/smartapp/).

## Results

#### GA inference versus self-identified race and ethnicity

Genome-wide GA, at continuous and categorical scale, was inferred for 9,818 TCGA participants across 33 cancer types. For GA at continuous scale, we used the ADMIXTURE program with reference populations from 1KGP to estimate continental ancestry proportions: AFR, EAS, NAT, EUR, and other (Fig. 1A). 98.95% of TCGA participant ancestry was explained by the four defined continental ancestry groups (Supplementary Fig. S1). TCGA participants show predominantly EUR ancestry (82.5%), followed by AFR (8.19%), EAS (6.76%), and NAT (1.50%). 81.23% (*n* = 7,975) of the TCGA participants had self-identified race information (American Indian or Alaska native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, and White) and 70.08% (n = 6,880) had selfidentified ethnicity (either Hispanic/Latino or not Hispanic/Latino) information available (Supplementary Table S1). A total of 8,010 TCGA participants had race and/or ethnicity information available, with 79.01% identified as White (White race and non-Hispanic or unknown ethnicity), 9.78% as Black (Black or African American race and non-Hispanic or unknown ethnicity), 7.12% as Asian (Asian or Native Hawaiian or other Pacific Islander race and non-Hispanic or unknown ethnicity), and 4.09% as Hispanic (Hispanic/Latino ethnicity). TCGA participants from each of these four SIRE groups show distinct and characteristic ancestry patterns, albeit with some overlap between groups (Fig. 1B).

For GA at categorical scale, PCA followed by k-NN classification with the IKGP was performed on the 9,818 TCGA participants for

categorization into four continental GA groups: EUR, AFR, AMR, and EAS. While high concordance rates were generally observed between SIRE and GA groups (White - EUR: 94.2%, Black/African-American -AFR: 95.79%, Asian - EAS: 90.88%), the concordance rate between Hispanic ethnicity and AMR ancestry was much lower at 68.6% (Supplementary Fig. S2). Differences between SIRE and GA classification can be seen in the PCA plots (Fig. 1C and D), which show visibly distinctive clustering patterns for GA groups, in contrast to the more continuous overlapping patterns seen for the SIRE groups. In addition, lack of concordance between GA and SIRE was seen for a small fraction of individuals with majority EUR or non-EUR ancestry. Of those with over 60% EUR ancestry and SIRE information, 47 (0.74%) participants self-identified as either Asian or Black instead of White. Similarly, of those with over 60% in EAS or AFR, 9 (1.58%) and 22 (2.81%) participants identified as White, respectively. In contrast, GA group assignment based on the PCA and k-NN classifier did not yield any participants with 60% or more of a single ancestry that were classified into a nonmajority GA group.

#### **Cancer survival disparity**

To determine cancer types with CSD among SIRE and GA groups, univariable and multivariable survival analyses were performed, modelling the effect SIRE and GA on OS and OS time. Survival data were available for 33 cancer types for 98.28% of the TCGA participants with inferred GA (n = 9,649). The estimated median follow-up time using the reverse Kaplan-Meier method for all participants was 34.5 months with shortest median follow-up in East Asian/Asian and the longest in European/White (Supplementary Table S2). Univariable analyses using the Kaplan-Meier method and multivariable analyses using Cox proportional hazard modeling revealed significant CSD among different SIRE and GA groups in four cancers: breast invasive carcinoma (BRCA), head and neck squamous cell carcinoma (HNSCC), kidney renal clear cell carcinoma (KIRC), and skin cutaneous carcinoma (SKCM; Fig. 2). Specifically, Kaplan-Meier analysis (Fig. 2A) and pairwise log-rank tests of survival curves (Fig. 2B) showed significantly different survival for the following SIRE groups: Hispanic versus others in KIRC and White or Hispanic versus Asian in SKCM. Significant survival probability difference for the GA groups were: EUR versus EA in HNSCC, AMR versus others in KIRC, and AMR versus both EUR and EAS, and EUR versus EAS in SKCM. Differences in the number and strength of CSD were observed when using GA compared with SIRE for survival analysis. For SKCM, comparison of survival analysis results between SIRE and GA groups was not possible due to lack of Black/African-American participants with survival outcomes. For other cancers including cholangiocarcinoma, glioblastoma multiforme, and rectum adenocarcinoma, SIRE data were unavailable and hence no pairwise tests of survival were performed (Supplementary Table S3).

Subsequently, we modeled the patient cancer survival outcomes in 33 cancers by SIRE and for GA at both categorical and continuous scales, while adjusting for relevant clinical features as covariates using Cox proportional hazard models (Supplementary Table S4). Multivariable modeling results showed significant disparity in survival probabilities among SIRE groups and for GA at categorical and continuous scales for four cancers: BRCA, HNSCC, KIRC, and SKCM (**Fig. 2C**). Final models for these four cancers were built with additional relevant covariates specific to each cancer and checked for concordance rate, power, and proportional hazard assumptions (Supplementary Table S5). For BRCA, there was a significantly greater risk of mortality for the AFR compared with the EUR ancestry group [HR = 1.84; 95% confidence interval (CI), 1.1-3.08; P = 0.021], and a



#### Figure 1.

Genome-wide GA and admixture estimates for TCGA participants. **A**, ADMIXTURE plot showing K = 5 GA and admixture proportions for 1000 Genomes Project continental reference populations–EUR, AF, AMR, and EAS—along with 9,818 TCGA participants. The five ancestry proportions are shown as EUR (orange), AFR (blue), NAT (red), EAS (green), and unknown (gray). **B**, ADMIXTURE ancestry proportions are shown for TCGA participants only, organized by SIRE groups as shown above the plot. **C** and **D**, PCA plots for TCGA participants, color-coded by SIRE (**C**) and GA groups (**D**).

10% increase in the AFR ancestry proportion was associated with an 8% increase in the relative risk of mortality (HR = 1.08; 95% CI, 1.02–1.15; P = 0.013). For HNSC, the AFR ancestry group showed significantly worse survival compared with the EUR group, with two times the relative risk of death (HR = 2.01; 95% CI, 1.23–3.28; P = 0.005), and a 10% increase in AFR ancestry was associated with a 9% increase in the relative risk of mortality (HR = 1.09; 95% CI, 1.03–1.16; P = 0.005). This disparity in survival was also seen in multivariable modeling of SIRE groups for Black/African-American versus White with HNSCC (HR = 2.05; 95% CI, 1.19–3.52; P = 0.009). In addition, a significantly higher risk of mortality was associated with Hispanics

compared with Whites with HNSCC (HR = 2.49; 95% CI, 1.17–5.31; P = 0.018), and a 10% increase in NAT ancestry was associated with a 36% increase in the relative risk of mortality (HR = 1.36; 95% CI, 1.18–1.57; P = <0.001). For KIRC, NAT ancestry was associated with better survival, with a 37% reduction in relative risk for every 10% increase in NAT ancestry (HR = 0.63; 95% CI, 0.4–0.99; P = 0.043). Lastly, for SCKM, a greater risk of mortality was associated with the EAS ancestry group compared with the EUR group (HR = 5.5; 95% CI, 2.51–12.04; P = <0.001), similar to what was seen for SIRE modeling of the Asian versus White groups (HR = 8.37; 95% CI, 3.54–19.77; P = <0.001). In contrast, better survival was associated with the AMR ancestry group



#### Figure 2.

CSD by SIRE and GA. A, Kaplan-Meier plots of the four cancer types that show significant disparities between SIRE and/or GA groups. B, Statistical significance of SIRE and GA CSD shown as  $-\log_{10}$  of P values of pairwise log-rank tests comparing Kaplan-Meier (KM) survival curves, C. Forest plots for exponentiated hazard ratios from the Cox proportional hazard multivariable models for the four cancer types that show significant disparities between SIRE and/or GA groups. Values for categorical GA groups are shown as squares, and values for continuous GA proportions are shown as triangles. Values for SIRE groups are shown as circles. Symbols are color coded according to the SIRE or GA groups: White/EUR, orange; Black/AFR, blue; Hispanic/AMR or NAT, red; Asian/EAS, green.

compared with the EUR group (HR = 0.15; 95% CI, 0.04–0.56; P = 0.005). In summary, Cox proportional hazard modeling using GA allowed for interpretation of survivorship in both categorical and continuous ancestry scales, and only three out of the six significant pairwise disparities detected with GA were also found by modeling SIRE (**Fig. 2C**).

We also evaluated the effect of African subcontinental ancestry on CSD, in light of the high levels of genetic diversity in Africa. To do so, we inferred African subcontinental ancestry proportions for all TCGA participants with  $\geq$ 50% African continental ancestry. African subcontinental ancestry proportions were inferred for three geographically and genetically coherent reference population groups, representing the three main regions of Africa that participated in the transatlantic slave

trade: West Africa (Senegambia and Sierra Leone), West Central Africa (Bight of Benin), and Southwest Africa (Bight of Biafra and the Loango Coast; Supplementary Fig. S3A an S3B). TCGA participants show all three African subcontinental ancestries, with predominantly West Central African ancestry corresponding to the Esan (ESN) and Yoruba (YRI) reference populations from Nigeria (Supplementary Fig. S3C). Multivariable survival modeling with three African subcontinental ancestry proportions as continuous variables did not yield any significant differences in the risk of mortality in AFR participants with BRCA or HNSCC, where significant CSD between AFR and EUR were detected (Supplementary Table S4D). Moreover, models for all other cancer-types in TCGA with AFR participants also generally showed no significant impact of African subcontinental ancestry on cancerspecific relative risk of mortality, with the exception of West Central African ancestry in kidney renal papillary cell carcinoma (KIRP; HR = 0.77; 95% CI, 0.60–0.99; P = 0.041).

#### DEGs by GA

Next, we evaluated the relationship between gene expression and GA for the four cancers that showed significant CSD. Differential gene expression analysis was performed for all significant GA group-cancer pairs: AFR versus EUR in BRCA, AFR versus EUR in HNSC, AMR versus EUR in KIRC, and AMR versus EUR and EAS versus EUR in SKCM. Initial analysis with DESeq2 identified 672 DEGs for BRCA, 443 DEGs for HNSCC, 386 DEGs for KIRC, and 169 and 316 DEGs

Figure 3.

DEGs for cancers showing CSD between GA groups. A, Volcano plots for DEGs in four cancer types and five GA pairs showing significant CSD. The x-axes show log<sub>2</sub> gene expression fold-change values for reference/comparison ancestry groups, and the y-axes show the  $\log_{10} P$  values for DEGs. B, Heatmap for DEGs in patients with BRCA comparing AFR versus EUR ancestry groups; each row is a single participant, and each column is a single gene. Normalized gene expression values are color-coded as shown in the legend, AFR ancestry participants are shown at the top of the heatmap (blue bar), and EUR participants are shown at the bottom (orange bar). Distributions of GA group mean expression values are shown above the heatmap.

respectively for SKCM (**Fig. 3A**). With the exception of BRCA, where 67.41% of all significant DEGs were significantly upregulated in the AFR group over EUR, the majority of DEGs were upregulated in the reference EUR group rather than in the comparison GA group for HNSCC, KIRC, and SKCM (**Fig. 3A**). A heatmap of BRCA gene expression values shows a clear demarcation between AFR and EUR groups, further illustrating differences in gene expression levels between GA groups in BRCA (**Fig. 3B**).

#### Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed with a focus on cancer-related pathways. For this analysis, the EUR ancestry groups



were taken as the reference, and significant DEGs were characterized as either up- or downregulated in the comparison ancestry group relative to the reference EUR group and analyzed for enrichment in the hallmark gene sets from the MSigDB. DEG sets across all four cancers were significantly enriched among 16 different hallmark pathways (**Fig. 4A**). KIRC and BRCA were significantly associated with the greatest number of functional pathways at 12 and 10, respectively (Supplementary Table S6). For BRCA, the AFR upregulated gene set was enriched for early and late estrogen response, apical junction, and KRAS signaling pathways, while the AFR downregulated gene set overlapped with other pathways, such as bile acid metabolism, adipogenesis, and fatty acid metabolism. In KIRC, only the AMR downregulated gene set showed any enrichment, including epithelial-mesenchymal transition (EMT), inflammatory response, angiogenesis, and peroxisome pathways. Similarly, the AFR downregulated gene set in HNSCC only showed enrichment with pathways for genes upregulated by KRAS activation, genes defining early response to estrogen, genes encoding components of apical junction complex, and myogenesis. Lastly, for SKCM, AMR and EAS downregulated genes were enriched for the complement and coagulation pathways, while AMR-only downregulated genes were enriched among the late estrogen response and KRAS signaling pathways.

DEG sets for these four cancers showed significant enrichment among both noncancer-related pathways (i.e., myogenesis) and several





GSEA of DEGs. **A**, Survival and gene set enrichment are shown for AFR (blue arrows), AMR (red arrows), and EAS (green arrows) ancestry groups compared with the reference EUR group. Up arrows indicate higher survival or expression compared with the EUR reference group, and down arrows indicate lower survival or expression. **B**, Illustration of three cancer-related hallmark pathways (inflammatory response, EMT, and angiogenesis) and their associated functions, which are enriched for genes that are underexpressed in the AMR ancestry group for KIRC. cancer-related pathways, such as angiogenesis, EMT, inflammatory response, peroxisome in KIRC and KRAS signaling, estrogen response, fatty and bile acid metabolism in BRCA. In particular, a number of AMR downregulated genes, such as IL6, IL8, matrix metallopeptidase (MMP) 1, are involved in pathways critical to tumorigenesis, including angiogenesis, EMT, and inflammatory response (Supplementary Table S6). These three pathways are intimately connected in a feedback loop: EMT and upregulation of the inflammatory response leads to secretion of inflammatory mediators including cytokines, chemokines, and matrix MMPs that create a protumor microenvironment that enhances angiogenesis, induces EMT, and maintains inflammation through paracrine and autocrine effects (Fig. 4B; ref. 24). For BRCA, there was AFR upregulation of 12 genes involved in response to estrogen and several genes involved in KRAS activation including HSD11B2, which is implicated in hormone metabolism and response and overexpressed in both breast cancer cell lines and breast tumors (Supplementary Table S6) (25). Estrogen has also been linked to breast cancer for its potential role as a mitogen stimulating cell division of breast tissue or as a carcinogen (26).

#### Differential gene expression signatures

DEG sets were further refined by characterizing gene expression signatures as reduced sets of genes that maximally distinguish each of the ancestry groups in the four cancers. First, DEG sets were subject to feature selection using ANOVA and F test. Second, to ensure the GA discriminatory power of the resulting feature-selected DEGs, we evaluated their ability to accurately classify GA groups using elastic net logistic regression. GA classification performances were validated based on 10-fold cross-validation (CV), and the average area under the receiver operating characteristic curve (AUC-ROC) was calculated for each of the four cancers. DEG sets were filtered and reduced to 37 genes for BRCA, 50 genes for HNSCC, nine genes for KIRC, and eight genes for SKCM (AMR vs. EUR). There were no DEGs that passed the assumption checks for EAS versus EUR in SKCM and therefore these were excluded from downstream analyses. The average accuracy rates of feature-selected gene sets were generally high with 87.2% for BRCA, 98.2% for HNSCC, 82.0% for KIRC, and 86.7% for SKCM (Supplementary Fig. S4). The high classification accuracies the GA differential gene expression signatures gave confidence to proceed with the smaller gene set for downstream analyses.

#### $\mathbf{G}\times\mathbf{G}\mathbf{A}$ associations

Before investigating the G  $\times$  GA interaction effects on cancer survival for the four cancer types, the genes making up the GA differential gene expression signatures for the four cancer types were examined for significant associations between gene expression and GA, while accounting for relevant clinical covariates using multiple linear regression (MLR). For BRCA, AFR ancestry showed significant associations (adjusted P < 0.05) with 30 out of 37 selected DEGs, accounting for patient age and molecular subtypes of their breast cancer. Similarly, 12 out of 50, four out of nine, and three out of eight DEGs showed significant associations between gene expression levels and GA for HNSCC, KIRC, and SKCM, respectively (Supplementary Table S7).

#### DMGs by GA

We explored the differences in gene mutational signatures between GA groups showing survival disparities for the four previously identified cancers. For each GA pair, genes displaying significantly different rates of somatic or germline mutations between GA groups were identified by Fisher exact test with a minimum number of patients with cancer with gene mutation in any one group set at five. There were total of 15 significantly DMGs with somatic variants in BRCA (AFR vs. EUR), 20 for HNSCC (AFR vs. EUR), two for KIRC (AMR vs. EUR), and 85 (AMR vs. EUR) and 27 (EAS vs. EUR) for SKCM (Supplementary Fig. S5 and Supplementary Table S8). Germline mutations were far less common than somatic mutations with only *ATM*, *BRCA1*, and *BRCA2* meeting the minimum mutation requirement. None of those three genes showed significantly different germline mutation frequencies between GA groups. We proceeded to gene-GA modeling of DMGs with different rates of somatic mutations using logistic regression. There were zero DMGs significantly associated with GA in BRCA, one in HNSCC, two in KIRC, and 15 (AMR vs. EUR) and four (EAS vs. EUR) in SKCM (Supplementary Table S7).

#### $\mathbf{G}\times\mathbf{G}\mathbf{A}$ interactions and cancer survival

DEGs with significant association between expression levels and GA, and DMGs with significant association between mutation status and GA, were tested for joint effects between gene and GA on survival outcomes for the four previously identified cancers with CSD. To do so, independent variable for DEG or DMG, and a term for  $G \times GA$  interaction, were added to the final survival model and assessed for statistical significance (Supplementary Table S5). While DEGs showed significant  $G \times GA$  interactions for three out of four cancers (BRCA, HNSCC, and SKCM), no DMGs showed significant interactions with GA on survival outcomes (**Fig. 5**). For BRCA, five protein-coding DEGs [Progestin and AdipoQ Receptor Family Member 6 (*PAQR6*),



#### Figure 5.

G × GA interactions associated with CSD. DEGs with significant G × GA interactions that are associated with the relative risk of death in patients with cancer are shown. The *x*-axis shows the log<sub>2</sub> gene expression fold-change values for reference/comparison ancestry groups, and the *y*-axis shows the change in HRs from the Cox proportional hazard models. Genes are grouped by cancer-type/GA combinations: AFR/HNSCC, triangles; AFR/BRCA, circles; AMR/SKCM, square. The set of AFR/HNSCC genes shown in light blue show the reestimated change in HRs after differential methylation levels were added into model.

Lck interacting transmembrane adaptor 1 (LIME1), MAX dimerization protein 3 (*MXD3*), Sin3A Associated Protein 25 (*SAP25*), and coiled-coil glutamate rich protein 2 (CCER2)], showed significant G × GA interactions, all of which were overexpressed in the AFR ancestry group compared with the reference EUR group (**Fig. 5**).

The greatest increase in relative risk of death for patients with BRCA with AFR ancestry was associated with increased in expression of *MXD3* (HR = 2.19; 95% CI, 1.13–2.43; *P* = 0.012), followed by *LIME1* (HR = 2.11; 95% CI, 1.01-4.02; P = 0.023), SAP25 (HR = 1.79; 95% CI,1.19-4.03; P = 0.042), CCER2 (HR = 1.66; 95% CI, 1.02-3.15; P = 0.009), and PAQR6 (HR = 1.56; 95% CI, 1.03–2.38; P = 0.038). For HNSCC, one protein-coding gene cathepsin W (CTSW) and one long noncoding RNA transcript (AC005330.1) showed significant G  $\times$ GA interactions for patients of AFR ancestry (Fig. 5). Interestingly, CTSW, which is underexpressed in the AFR ancestry group compared with the EUR group, was associated with an increased relative risk in mortality (HR = 1.32; 95% CI, 1.004-1.73; P = 0.047) for pstients with HNSCC of AFR ancestry, while AC005330.1, overexpressed in AFR compared with EUR, was associated with decreased relative risk (HR = 0.69; 95% CI, 0.50–0.96; P = 0.025). refilin A (RFLNA), a proteincoding gene overexpressed in the AMR ancestry group compared with the EUR group in SKCM, was also associated with increased relative risk of mortality (HR = 2.69; 95% CI, 1.42–5.08; P = 0.002; Fig. 5). There were nearly no mutations found for these seven protein-coding genes, with less than 1% of each GA group carrying any genetic variants. Here, the G  $\times$  GA interaction results for DEGs highlight varying effects of selected gene expression on cancer survivorship based on GA. Furthermore, the expectation of inverse relationship between survival probability and gene expression levels of genes associated with worse survival (e.g., the GA group with lower survival probability will see a higher expression level for genes associated with worse survival) did not necessarily hold for all DEGs across four cancers. This suggests that directionality of relationship between cancer survival and selected gene expression may also vary by GA and/or cancer type.

#### Differential methylation and $\mathbf{G} \times \mathbf{G}\mathbf{A}$ interactions

Investigation of differential mutation signatures revealed that there are nearly no genetic variants for the seven protein-coding genes showing significant  $G \times GA$  interactions. Therefore, to test for evidences of regulatory differences contributing to the differential expression of these DEGs, we performed methylation analysis of CpG sites specific to these seven genes. Methylation levels of gene-associated CpG sites were downloaded for four cancers, and Spearman rank correlations were calculated between DNA methylation levels and corresponding gene expression levels. There were significant correlations between DNA methylation levels and gene expression levels for seven out of 13 CpG probes in PAQR6, nine out of 18 for LIME1, six out of 10 for SAP25, seven out of 26 for MXD3, one out of one for CCER2, zero out of one for CTSW, and 17 out of 66 for RFLNA (Supplementary Fig. S6). Gene-methylation level associations, adjusted for GA and other clinically relevant covariates, and differential methylation between GA pairs, were evaluated using MLR and the Wilcoxon rank sum test, respectively. There were five CpG sites for MXD3, four for LIME1, and one for PAQR6, associated with both differential expression and differential methylated between GA pairs. The final model used to interrogate the impact of differential methvlation on  $G \times GA$  interactions included the CpG sites cg07598367 for PAQR6, cg12413156 and cg01242400 for LIME1, and cg16616449, cg13278795, and cg08293303 for MXD3, of which, all but one CpG site (cg16616449 for MXD3) showed significantly lower methylation level in AFR compared with EUR with BRCA (Supplementary Fig. S7). The addition of these differentially methylated CpG sites yielded a better model fit with higher concordance rate and lower Akaike information criterion (AIC) for all three DEGs, and the interaction terms for  $G \times GA$  remained significant (Supplementary Table S5). Interestingly, the change in HR between GA groups increased in all three genes, from 1.56 to 2.96 (P = 0.012) for PAQR6, 2.11 to 2.53 (P = 0.022) for LIME1, and 2.19 to 2.45 (P = 0.03) for SAP25, indicating even greater survival disparity between GA groups with the addition of differential methylation data (**Fig. 5**). Methylation analysis results points to the regulatory role of differential methylation on gene expression and provide further evidence for  $G \times GA$  interactions and their contribution to CSD.

## Discussion

In this study, we performed a pan-cancer analysis of 9,818 TCGA participants across 33 cancers types in an effort to discover CSD between SIRE and GA groups along with the molecular genetic features–gene expression, methylation, and mutation–that are associated with such disparities. There are four main implications from this study, each of which shed light on a different aspect of how GA and molecular genetic features interact to affect CSD (1). GA and SIRE are correlated but have different impact on CSD (2). GA is associated with survival disparities in four cancer types: BRCA, HNSC, KIRC, and SKCM (3). Differential gene expression between ancestry groups associates cancer-related hallmark pathways and cancer-related genes with CSD, seven of which contribute to disparities via interactions with GA (4). Gene methylation differences between ancestry groups are associated with differential gene expression and its impact on CSD.

First, we demonstrated that effect of GA and SIRE on survival outcomes of TCGA participants are different. The strength and number of significant CSD found using SIRE varied from GA in the survival analyses, suggesting that different underlying effects may be attributable for the observed differences. Thus, we used GA for all downstream analyses with differential molecular features related to CSD. GA also offers several advantages over SIRE including a greater spectrum of analysis and interpretation by virtue of two scales of measurement, categorical and continuous. Using the continuous scale of continental GA proportions, we estimated the effect of an incremental increase in a particular GA on the HR of patients with cancer. For example, in patients with BRCA with EUR GA, a 10% increase in AFR ancestry was associated with an 8% increase in relative risk of mortality. Meanwhile, a 10% increase in NAT ancestry was associated with a 37% reduction in relative risk in KIRC. This kind of increased resolution for individuals' GA will be important as the number of genetically admixed individuals continues to grow through increased globalization, immigration, and intermarriage (27). Moreover, using GA inferences based on genomic data will help reduce bias and misclassification associated with either self-identification or health-worker identification of race and ethnicity based on subjective perceptions of skin color, cultural background, and other social factors (28). The issue of dissonance between SIRE and GA, especially when persons of non-EURmajority GA identify themselves as White, can also lead to a loss of minority samples underrepresented in genomic studies.

Second, GA was associated with CSD in four cancers: BRCA, HNSC, KIRC, and SKCM. AFR ancestry was associated with significantly worse survival relative to EUR ancestry in both BRCA and HNSC. EAS ancestry also showed increased mortality risk for patients with cancer with SKCM compared with EUR ancestry. Having AMR or NAT ancestry, however, had reversed effects on survival outcomes for different cancers: it negatively affected the survival outcome in HNSCC compared with EUR ancestry but had a positive effect on survival in KIRC compared with AFR and in SKCM compared with EUR. Even after controlling for available clinically relevant factors, the ancestry effect on cancer survival varied in direction and strength across cancers, suggesting that GA-associated survival disparity exists and are cancer-specific.

Third, our results indicate that differential gene expression between ancestry groups associates cancer-related hallmark pathways and cancer-related genes with CSD. Notably, DEGs between AMR and EUR with KIRC were enriched in pathways including inflammatory response, EMT, and angiogenesis, all of which are all well-known hallmark features of cancer and crucial for tumor invasion, growth, and metastasis (29). These genes were underexpressed in AMR compared with EUR with KIRC, which corresponds with the reduced risk of mortality seen in the AMR group. Similarly, patients with BRCA with AFR ancestry showed nearly twice the relative risk of mortality compared with EUR, associated with overexpression of genes related to KRAS activation and estrogen response. The role of both KRAS, an oncogene and a tumor-inducer, and estrogen, in the development of breast cancer has been described in other studies, suggesting that upregulation of these genes may contribute to the higher mortality risk of AFR compared with EUR observed in BRCA (30).

We identified seven protein-coding genes that are associated with CSD via interactions between GA and differential gene expression. There were five such genes uncovered for BRCA-PAQR6, LIME1, MXD3, SAP25, and CCER2-all of which are both upregulated in the AFR ancestry group relative to EUR and previously implicated in tumorigenesis. For example, the PAQR6 gene encodes a plasma membrane progesterone receptor that has been shown to mediate progestin-induced inhibition of apoptosis in breast cancer cells (31). In prostate cancer, elevated expression of PAQR6 was associated with worse patient survival and act as tumor promoter via regulation of the MAPK signaling pathway (5). Both the SAP25 and MXD3 genes encode proteins involved in transcription repression, which are potential targets for cancer therapy. Interference with the interaction between SAP25 and Sin3A/B protein has been shown to inhibit tumor growth in breast cancer cell lines (32). Knockdown of the MXD3 transcription factor protein induced apoptosis in neuroblastoma cell lines, also demonstrating its potential as a therapeutic target (33). The LIME1 gene is another potential target for cancer treatment due to its role in regulation of T-cell functions and genes involved in DNA repair such as MLH1 and BRCA1 (34). Interestingly, increased expression of CCER2 was associated with taxane-induced peripheral neuropathy, a chemotherapy toxicity with an increased risk in AFR ancestry populations due to Taxane, one of the most commonly used chemotherapeutic agents for early and metastatic breast cancer (35-36). Conversely, in HNSCC, expression of the gene CTSW was downregulated in the AFR group compared with EUR. CTSW is a candidate tumorsuppressor gene that is expressed in immune cells such as natural killer and T-cells (37). In a recent study, CTSW showed a positive correlation with breast cancer patient survival and is believed to improve immunity against early cancer cells (37). Finally, the RFLNA gene was over expressed in the AMR ancestry group relative to EUR in SKCM. The RFLNA protein, also called Refilin-A, interacts with filamins and plays a regulatory role in the actin-cytoskeleton network, important in cell adhesion and migration (38). Overexpression of this gene has been associated with several types of cancer and can increase the risk of cancer metastasis (38, 39). While the potential role of these proteincoding genes in tumorigenesis has been suggested, their disparate impact on cancer patient survival linked to specific GA has not been previously shown. These cancer-related genes interacting with specific GA offer potential therapeutic targets for population groups disproportionately burdened by CSD.

Fourth, gene methylation differences, but not mutation differences, are associated with differential expression between ancestry groups and contribute to CSD. DMG analysis revealed that there are no genes with significantly different mutation frequencies between GA groups that show  $G \times GA$  interaction. The mutational rates for the seven protein-coding DEGs were also exceptionally low and indifferent across GA groups, suggesting that regulatory differences are more relevant to CSD. To test this hypothesis, differentially methylated CGIs correlated to each gene were fitted in the survival model with  $G \times GA$ interaction term. GA HR changes increased and the overall model fits were improved, all the while the interaction term remained significant. Epigenetic dysregulation and aberrant methylation are often linked to cancer, including hypomethylation of the GD3 (ganglioside D3 synthase) gene in OS of patients with triple-negative breast cancer (40). Yet these somatically inherited changes are reversible, unlike mutations, and present new opportunities for development of drugs that target epigenetic enzymes to modify these changes (41, 42). Understanding of differential methylation patterns and other epigenetic changes that modify the expression of genes associated with CSD may help to improve health equity in patients with cancer.

There were several limitations to our study. TCGA samples are mostly of EUR ancestry (>80%) and have significantly smaller samples for other ancestry groups. There is far less diversity in TCGA dataset compared with the U.S. population, and it is not a representative sample of the general or cancer populations in the country. This issue of limited minority group samples played a role in using overall survival, most widely available survival endpoint across all cancers, for our analysis. In addition to the significance of mortality as a direct indicator for cancer burden, death is also the least ambiguous endpoint to define compared with other survival endpoints, such as disease-free survival and progression-free survival, reducing risk of misclassification (4). However, some of the patients may not have had sufficient follow-up time to experience death, potentially biasing the results for more aggressive cancer types with higher rates of mortality (4). Another important caveat to note is the potential disparity in the time of diagnosis across different population groups. Since the time-toevent is defined as time between diagnosis and death in our study, differences in frequency, quality, and access to cancer screening tests between GA groups can lead to lead-time bias (due to early detection) or length-time bias (due to non- or slow-growing cancers). For example, White women are more likely to have mammograms at regular basis and have higher in quality 2D or 3D imaging compared to African women with BRCA (43). This can lead to earlier discovery of both symptomatic and asymptomatic tumors in White women and potentially inflate survival time and probabilities in the study of CSD. Since mortality is relatively robust against overdiagnosis, however, we chose OS as endpoint with tumor status at death and age at diagnosis as constant covariates in our models to reduce potential time biases. In addition, we adjusted or stratified for several other clinical characteristics relevant to cancer survival including tumor stage and morphology, to ensure that the models reflect differences in cancer mortality and not differences in diagnosis between GA groups.

Moreover, it must be emphasized that TCGA clinical data do not capture many important environmental, demographic, and behavioral factors that may also contribute to CSD, such as socioeconomic status, access to health care, diet and exercise habits, and stress associated with racial discrimination (44–45). Some studies have shown that equity in or adjustment for these nongenetic factors significantly reduce the levels of CSD (46). Therefore, there may be hidden factors that confound the level of genetic contributions of identified DEGs to CSD that are unaccounted for in our analysis.

Finally, it should be noted that we have mainly characterized GA and performed analyses at the continental level, for both categorical and continuous ancestry variables. Continental level analysis allowed for the most direct comparison with SIRE and was appropriate given the limited minority or non-EUR participants in TCGA. While we performed a separate analysis for African ancestry at the subcontinental level (West Central, West, and Bantu) to investigate its impacts on relative risk of mortality for AFR patients, there were mostly no observed effects and HR was not always estimable due to low sample size. Future studies that perform fine-scale GA analysis on a dataset with more minority samples may be able to detect CSD-ancestry associations potentially missed in this study.

In summary, our pan-cancer analysis of survival outcomes in different population groups and the associated differential molecular features based on GA highlights the molecular genetic contributions to CSD. A number of DEGs were identified in this study that are associated and interact with specific GA and impact cancer survival. Many of these genes have been previously implicated in tumorigenesis and therefore may serve as potential targets for the development of new cancer therapies that can alleviate persistent CSD. Furthermore, our results indicate that disparities in cancer survival are not significantly associated with genetic variants, such as germline or somatic mutations, but instead are influenced by regulatory changes modified by epigenetics including gene methylation. This is in contrast to much of traditional cancer research focusing on the mutational spectrum of oncogenes and tumor-suppressor genes. Instead, our findings point to importance of epigenetics in tumorigenesis. More studies are needed to further characterize the underlying biological differences and mechanisms contributing disparate mortality and morbidities in patients with cancer. Findings of this kind can inform the discovery of new

#### References

- National Academies of Sciences Engineering, and Medicine. Communities in action: pathways to health equity. Washington (DC): National Academies Press; 2017.
- Bugos K, Curtiss CP. Evidence-based survivorship care: current and future challenges in survivorship research. In: Haylock P, Curtiss C, editors. Cancer survivorship: interprofessional, patient-centered approaches to the seasons of survival Pittsburgh (PA): Oncology Nursing Society;2019. p. 3–13.
- Ellis L, Woods LM, Esteve J, Eloranta S, Coleman MP, Rachet B. Cancer incidence, survival and mortality: explaining the concepts. Int J Cancer 2014; 135:1774–82.
- 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.
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70: 7–30.
- DeSantis CE, Siegel RL, Sauer AG, Miller KD, Fedewa SA, Alcaraz KI, et al. Cancer statistics for African Americans, 2016: Progress and opportunities in reducing racial disparities. CA Cancer J Clin 2016;66:290–308.
- Zeng C, Wen W, Morgans AK, Pao W, Shu XO, Zheng W. Disparities by race, age, and sex in the improvement of survival for major cancers: results from the national cancer institute surveillance, epidemiology, and end results (SEER) program in the United States, 1990 to 2010. JAMA Oncol 2015;1: 88–96.
- Aizer AA, Wilhite TJ, Chen MH, Graham PL, Choueiri TK, Hoffman KE, et al. Lack of reduction in racial disparities in cancer-specific mortality over a 20-year period. Cancer 2014;120:1532–9.

druggable targets for cancer treatments and prevention methods that are precise and population-specific, thereby helping to combat health disparities in cancer.

#### **Authors' Disclosures**

D. Ban reports other support from Ovarian Cancer Institute (Atlanta, GA) during the conduct of the study. No disclosures were reported by the other authors.

#### **Authors' Contributions**

K.K. Lee: Conceptualization, data curation, formal analysis, visualization, methodology, writing-original draft, writing-review and editing. L. Rishishwar: Formal analysis, visualization, methodology, writing-review and editing. D. Ban: Data curation, formal analysis, methodology. S.D. Nagar: Data curation, formal analysis, methodology. L. Mariño-Ramírez: Supervision, funding acquisition, project administration, writing-review and editing. J.F. McDonald: Conceptualization, resources, supervision, funding acquisition, project administration, writing-review and editing. I.K. Jordan: Conceptualization, resources, supervision, funding acquisition, writing-review and editing. the supervision, funding tration, writing-review and editing.

#### Acknowledgments

K.K. Lee, L. Rishishwar, and S.D. Nagar were supported by the IHRC-Georgia Tech Applied Bioinformatics Laboratory (Atlanta, GA; grant no. RF383). L. Mariño-Ramírez was supported by the NIH Distinguished Scholars Program (DSP) and the Division of Intramural Research (DIR) of the National Institute on Minority Health and Health Disparities (Bethesda, MD; NIMHD) at NIH (1ZIAMD000016 and 1ZIAMD000018). D. Ban was supported by the Ovarian Cancer Institute. J.F. McDonald was supported by the Ovarian Cancer Institute, Deborah Nash Endowment, and Northside Hospital Research Foundation. The results here are in whole or part based upon data generated by the TCGA Research Network: https:// www.cancer.gov/tcga.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 30, 2021; revised September 20, 2021; accepted January 18, 2022; published first January 21, 2022.

- 9. Simon S. editor, Facts & figures 2019: US cancer death rate has dropped 27% in 25 years. Atlanta (GA): American Cancer Society; 2019.
- Zavala VA, Bracci PM, Carethers JM, Carvajal-Carmona L, Coggins NB, Cruz-Correa MR, et al. Cancer health disparities in racial/ethnic minorities in the United States. Br J Cancer 2021;124:315–32.
- Ahmad A, Azim S, Zubair H, Khan MA, Singh S, Carter JE, et al. Epigenetic basis of cancer health disparities: Looking beyond genetic differences. Biochim Biophys Acta Rev Cancer 2017;1868:16–28.
- Deng J, Chen H, Zhou D, Zhang J, Chen Y, Liu Q, et al. Comparative genomic analysis of esophageal squamous cell carcinoma between Asian and Caucasian patient populations. Nat Commun 2017;8:1533.
- Wang BD, Ceniccola K, Hwang S, Andrawis R, Horvath A, Freedman JA, et al. Alternative splicing promotes tumour aggressiveness and drug resistance in African American prostate cancer. Nat Commun 2017;8:15921.
- Yuan J, Hu Z, Mahal BA, Zhao SD, Kensler KH, Pi J, et al. Integrated analysis of genetic ancestry and genomic alterations across cancers. Cancer Cell 2018;34: 549–60.
- Huo D, Hu H, Rhie SK, Gamazon ER, Cherniack AD, Liu J, et al. Comparison of breast cancer molecular features and survival by african and european ancestry in the cancer genome atlas. JAMA Oncol 2017;3:1654–62.
- Ademuyiwa FO, Tao Y, Luo J, Weilbaecher K, Ma CX. Differences in the mutational landscape of triple-negative breast cancer in African Americans and Caucasians. Breast Cancer Res Treat 2017;161:491–9.
- Zhao F, Copley B, Niu Q, Liu F, Johnson JA, Sutton T, et al. Racial disparities in survival outcomes among breast cancer patients by molecular subtypes. Breast Cancer Res Treat 2021;185:841–9.

- Yudell M, Roberts D, DeSalle R, Tishkoff S. SCIENCE AND SOCIETY. Taking race out of human genetics. Science 2016;351:564–5.
- 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.
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015;4:7.
- Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinf 2011;12:246.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
- Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 2018;28: 1747–56.
- Palena C, Hamilton DH, Fernando RI. Influence of IL-8 on the epithelialmesenchymal transition and the tumor microenvironment. Future Oncol 2012; 8:713–22.
- 25. Yang J, Yu H, Zhang L, Deng H, Wang Q, Li W, et al. Overexpressed genes associated with hormones in terminal ductal lobular units identified by global transcriptome analysis: An insight into the anatomic origin of breast cancer. Oncol Rep 2016;35:1689–95.
- Mufudza C, Sorofa W, Chiyaka ET. Assessing the effects of estrogen on the dynamics of breast cancer. Comput Math Methods Med 2012;2012:473572.
- Perez AD, Hirschman C. the changing racial and ethnic composition of the US population: emerging american identities. Popul Dev Rev 2009;35: 1–51.
- Khan MA, Patel GK, Srivastava SK, Carter JE, Pierce JY, Rocconi RP, et al. Looking at cancer health disparities without the colored lenses. Cancer Health Disparities 2019;3:e1–e9.
- 29. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011; 144:646–74.
- Cui Y, Deming-Halverson SL, Shrubsole MJ, Beeghly-Fadiel A, Fair AM, Sanderson M, et al. Associations of hormone-related factors with breast cancer risk according to hormone receptor status among white and African American women. Clin Breast Cancer 2014;14:417–25.
- Dressing GE, Alyea R, Pang Y, Thomas P. Membrane progesterone receptors (mPRs) mediate progestin induced antimorbidity in breast cancer cells and are expressed in human breast tumors. Horm Cancer 2012;3:101–12.

- 32. Farias EF, Petrie K, Leibovitch B, Murtagh J, Chornet MB, Schenk T, et al. Interference with Sin3 function induces epigenetic reprogramming and differentiation in breast cancer cells. Proc Natl Acad Sci U S A 2010;107:11811–6.
- Duong C, Yoshida S, Chen C, Barisone G, Diaz E, Li Y, et al. Novel targeted therapy for neuroblastoma: silencing the MXD3 gene using siRNA. Pediatr Res 2017;82:527–35.
- Bommhardt U, Schraven B SLBTCR Signaling: Emerging functions of lck in cancer and immunotherapy. Int J Mol Sci 2019;20.
- Schneider BP, Lai D, Shen F, Jiang G, Radovich M, Li L, et al. Charcot-Marie-Tooth gene, SBF2, associated with taxane-induced peripheral neuropathy in African Americans. Oncotarget 2016;7:82244–53.
- Gradishar WJ. Taxanes for the treatment of metastatic breast cancer. Breast Cancer (Auckl) 2012;6:159–71.
- Zhang Y, Manjunath M, Yan J, Baur BA, Zhang S, Roy S, et al. The cancerassociated genetic variant Rs3903072 modulates immune cells in the tumor microenvironment. Front Genet 2019;10:754.
- Savoy RM, Ghosh PM. The dual role of filamin A in cancer: can't live with (too much of) it, can't live without it. Endocr Relat Cancer 2013;20:R341–56.
- Yue J, Huhn S, Shen Z. Complex roles of filamin-A mediated cytoskeleton network in cancer progression. Cell Biosci 2013;3:7.
- Li W, Zheng X, Ren L, Fu W, Liu J, Xv J, et al. Epigenetic hypomethylation and upregulation of GD3s in triple negative breast cancer. Ann Transl Med 2019; 7:723.
- Cheng Y, He C, Wang M, Ma X, Mo F, Yang S, et al. Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. Signal Transduct Target Ther 2019;4:62.
- 42. Fardi M, Solali S, Farshdousti Hagh M. Epigenetic mechanisms as a new approach in cancer treatment: An updated review. Genes Dis 2018;5:304–11.
- Iqbal J, Ginsburg O, Rochon PA, Sun P, Narod SA. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. JAMA 2015;313:165–73.
- Dean LT, Gehlert S, Neuhouser ML, Oh A, Zanetti K, Goodman M, et al. Social factors matter in cancer risk and survivorship. Cancer Causes Control 2018;29: 611–8.
- 45. Massetti GM, Dietz WH, Richardson LC. Excessive weight gain, obesity, and cancer: opportunities for clinical intervention. JAMA 2017;318:1975–6.
- Brawley OW. Prostate cancer and the social construct of race. Cancer 2021;127: 1374–6.