Translational Oncology 14 (2021) 101086

Contents lists available at ScienceDirect

Translational Oncology

journal homepage: www.elsevier.com/locate/tranon

Original Research

Comparative analysis of triple-negative breast cancer transcriptomics of Kenyan, African American and Caucasian Women

Mansoor Saleh[a,b,c,1], Darshan Shimoga Chandrashekar[d,1], Sayed Shahin[e], Sumit Agarwal[d], Hyung-Gyoon Kim[d], Michael Behring[d], Asim Jamal Shaikh[e], Zahir Moloo[e], Isam-Eldin A Eltoum[d], Clayton Yates[d,f], Sooryanarayana Varambally[c,d,2], Upender Manne[c,d,2,*]

[a] Department of Medicine, University of Alabama at Birmingham, Birmingham 35233, AL, United States
[b] Department of Hematology-Oncology, the Aga Khan University, Nairobi, Kenya
[c] O'Neal Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham 35233, AL, United States
[d] Department of Pathology, University of Alabama at Birmingham, Birmingham, AL, United States
[e] Department of Pathology, the Aga Khan University, Nairobi, Kenya
[f] Department of Biology & Center for Cancer Research, Tuskegee University, Tuskegee, AL, United States

A B S T R A C T

Purpose: : Triple-negative breast cancer (TNBC) patients of various ethnic groups often have discrete clinical presentations and outcomes. Women of African descent have a disproportionately higher chance of developing TNBCs. The aim of the current study was to establish the transcriptome of TNBCs from Kenyan (KE) women of Bantu origin and compare it to those TNBCs of African-Americans (AA) and Caucasians (CA) for identifying KE TNBC-specific molecular determinants of cancer progression and potential biomarkers of clinical outcomes.

Patients and Methods: : Pathology-confirmed TNBC tissues from Kenyan women of Bantu origin ($n = 15$) and age and stage range matched AA ($n = 19$) and CA ($n = 23$) TNBCs of patients from Alabama were included in this study. RNA was isolated from paraffin-embedded tissues, and expression was analyzed by RNA sequencing.

Results: : At clinical presentation, young KE TNBC patients have tumors of higher stages. Differential expression analysis identified 160 up-regulated and 178 down-regulated genes in KE TNBCs compared to AA and CA TNBCs. Validation analyses of the TCGA breast cancer data identified 45 KE TNBC-specific genes that are involved in the apoptosis (ACTC1, ERCC6 and CD14), cell proliferation (UHRF2, KDM4C, UHMK1, KCNH5, KRT18, CSF1R and S100A13), and Wnt signaling (BCL9L) pathways.

Conclusions: : In this study, we identified biomarkers that are specific for KE TNBC patients of Bantu origin. Further study with a larger sample size of matched tumors could confirm our findings. If biologically confirmed, these molecular determinants could have clinical and biological implications and serve as targets for development of personalized therapeutics for KE TNBC patients.

Introduction

Triple-negative breast cancer (TNBC) is an aggressive phenotype of breast cancer, characterized by absence of estrogen receptor (ER), progesterone receptor (PR), and Her2 protein [1]. TNBCs are more frequently found in women of African descent [2–6]. African women with TNBCs present at younger ages and with more advanced disease than Caucasian (CA) women [7], and they suffer a disparity in outcomes despite being matched for stage. In the absence of funded public health policies, most African women present with advanced breast cancer [8], and, for African women, there appears to be no difference in stage at diagnosis for TNBC versus non-TNBC histology [9].

Although this disparity has often been associated with differences in socioeconomic status, recent studies have demonstrated that biological factors are involved in both the risk and outcome of breast cancers for African American (AA) patients [10–12]. Previous studies of the genomic fingerprint of Nigerian TNBCs described molecular features associated with an aggressive phenotype that are more prevalent than in TNBCs of AA and CA women [13,14]. There is distinct genetic diversity among Africans from various geographic regions [15–17]. Variation in the prevalence of ER negative breast cancers by study location [18,19] supports the importance of ethnic subgroup and locale upon TNBC epidemiology in sub-Saharan Africa. A recent analysis based on the National Program of Cancer Registries and US Cancer Statistics showed

* Corresponding author at: Translational Anatomic Pathology, Wallace Tumor Institute, , University of Alabama at Birmingham, Room # 420A, Birmingham, AL 35233, USA.
E-mail address: upendermanne@uabmc.edu (U. Manne).
[1] Equal Contribution
[2] Shared Senior Authorship

https://doi.org/10.1016/j.tranon.2021.101086
Received 12 December 2020; Received in revised form 4 March 2021; Accepted 23 March 2021

that the prevalence of TNBCs among black women in the United States varied by birthplace [5]. Thus, the finding for Nigerian patients may not be representative of women of East Africa. Furthermore, given the ancestral heterogeneity of the population in Kenya [17], findings for African women of any one ethnicity could well differ from those for women with a different ethnicity.

Recently, the decision-making process for cancer therapies has improved as knowledge of the molecular biology of breast cancer has evolved. In this regard, human breast carcinogenesis and metastasis, including that of TNBCs, has been explored by use of new techniques such as gene expression profiling, genome and transcriptome sequencing, proteomics, and microRNA analysis that led to identification of new molecular sub-type specific biomarkers and therapies. Our recently reported findings suggest that the altered oncologic pathways in TNBCs of AA and CA patients are driven by shared genetic ancestry [20]. However, the molecular profiles of cancers of different populations remain poorly characterized, as few studies have reported molecular characterization of breast cancers for non-CAs, in particular none for Kenyan (KE) TNBCs. Thus, we embarked on a study to determine the molecular fingerprints of TNBCs of women from Kenya and compared them with those of AA and CA women with the aims of determining differences and identifying molecular characteristics of clinical relevance.

Methods

Patients and tissue collection

Using an IRB-approved protocol (REC Ref 2012/05 (v7) of the Aga Khan University Hospital in Nairobi (AKUHN), we consented women with histologically confirmed TNBCs to contribute residual tumor tissue for genetic sequencing at the University of Alabama at Birmingham (UAB). The protocol received expedited review by the IRB at UAB (IRB-100,910,007). All KE patients included in this study were of non-white, African descent, and Bantu ancestry. Where available, clinical information, including age, gender, date of diagnosis, and pathologic stage (AJCC) [21] of the tumor as determined following surgery, was recorded from the medical charts. Patients did not undergo metastatic staging unless they had normal laboratory parameters, including liver function studies, and obvious symptoms of systemic disease. All patients were thus deemed by their treating surgeon to have localized, non-metastatic disease. All Kenyan patients underwent definitive therapeutic mastectomy and axillary lymph node dissection. Information about post-surgical adjuvant therapy and time to and site of recurrence was not available, since most patients returned to their county medical system for post-operative care. However, survival data at 3 yrs was available through direct contact with the patients via phone. The FFPE tissue blocks were cleared for shipping to UAB by the Kenyan Ministry of Health (MOH/ADM/1/1/81 Vol.1). All 40 samples were catalogued and identified only by an assigned number with the identification code securely maintained at the Pathology Department of AKUHN. Clinical information for each of the 40 cases was provided with the accompanying blocks.

ER, PR, and HER2/neu immunostaining

Where available, all FFPE tissues sections underwent repeat ER/PER/HER2 testing at the Department of Pathology of UAB by the following protocol. In brief, tumor blocks were selected, and 5-μM sections were placed on plus slides. Heat-induced epitope retrieval was performed for 20 min at 95 °C on a Dako PT Link Pre-treatment System using a high pH target retrieval solution, pH 9, diluted 1 in 50 as per the manufacturer's instructions. The sections were stained for ER (FLEX RTU Monoclonal Rabbit Anti-Human ER α, Clone EP1) PR, (FLEX RTU Monoclonal Rabbit Anti-Human PgR 636 Antibody clone), and HER2/neu (Polyclonal Rabbit Anti-Human c-erbB-2 Oncoprotein diluted 1:200 with the EnVision™ FLEX antibody diluent) on a Dako Autostainer Link 48 platform using EnVision™ FLEX Kit detection kits. Diaminobenzidine was used as the chromogen. Interpretation of ER and PR staining was accomplished according to the Allred scoring system. Composite scores ≥3 for the intensity and proportion of cells staining were considered to be positive. HER2/neu scoring was according to the ASCO/CAP guidelines. HER2/neu was scored positive if 30% or more of the tumor cells showed complete and strong membrane positivity. A score of 0 or 1+ was considered HER2/neu negative. Cases that were scored as HER2/neu 2+ (equivocal) were subjected to FISH analysis at our reference laboratory to assess gene amplification. A FISH test result was considered positive if the ratio of the LSI HER2/neu/CEP17 signal was >2.2. Normal breast tissue was used as an in-built positive control for ER/PR. Additionally, a section of uterine cervix was used as an on-slide positive control. All stromal, columnar epithelial, and squamous epithelial cells showed a moderate to strong nuclear staining. For HER2/neu, an on-slide control section from a composite tissue block of HER2/neu scores of 1+, 2+, and 3+ was used.

Of the 40 samples from AKUHN, FFPE core biopsy blocks were available for 34 cases, of which 31 were reconfirmed at UAB as TNBCs with ER < 1%, PR <1%, and Her2 < 2+. There was insufficient material for RNA extraction from 13 of the samples, leaving 18 samples for genomic studies. Of these, demographic data (date of diagnosis, age, gender, pathologic stage, clinical stage, and status at 3-year follow up) were available for only 15 of the 18 cases. These fifteen cases were thus selected for further data analysis.

For this study, we used a sub-set of 19 AA and 23 CA women with TNBCs, diagnosed at the University of Alabama at Birmingham (UAB), from a larger, recently described study [20]. These AA and CA cohorts were matched as close as possible for age and stage ranges of KE patients. Kenyan patients underwent diagnostic staging studies only if they demonstrated clinical symptoms or a laboratory abnormality suggestive of metastatic disease. Since the larger study was focused on identifying molecular signatures that differ between TNBCs of AA women and CA women, we performed RNA sequencing (RNAseq) for the TNBC cohort. As described in our recent study [20], a retrospective convenient formalin-fixed, paraffin-embedded (FFPE) archival tissue cohort from the Division of Anatomic Pathology of UAB consisting of 104 AA and CA women diagnosed with TNBC between 2000 and 2012 was selected (Supplemental Figure 1A). For this cohort, personal medical history and clinical records were limited. Following quality control screening, a final set of 75 cases remained (42 AAs and 33 CAs). Of these, samples were separated by treatment status, treatment-naïve ($n = 60$) or recurrent tumors from post-treatment ($n = 15$). Of the treatment-naïve cases, there was a near equal distribution of race categories (31 AAs and 29 CAs). Of the recurrent tumor cases, the representation of AAs was more than twice that of CAs (11 AAs and 4 CAs). All tumors and corresponding normal regions were macro-dissected by pathologists prior to RNA extraction. Stage and grade distribution were similar between racial groups (Table 1).

RNA isolation

Total RNA was extracted from macrodissected FFPE samples using TRIzol reagent kit (Invitrogen) was performed as described earlier [22]. The concentration of the RNA was estimated by a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, California) and an Agilent 2100 Bioanalyzer (Applied Biosystems, Carlsbad, CA) assessed its integrity. RNA from each sample was amplified for RNAseq. The methodology for removing the ribosomal RNA (rRNA), fragmentation, priming for the first-strand and second-strand synthesis, the standard library preparation, A-tailing, and custom adapter ligation were accomplished by following the methods described in our recent breast cancer study [20].

Table 1
Descriptive statistics of clinical variables by race.

Race	AA ($N = 23$)	CA ($N = 19$)	KE ($N = 15$)	p
Age (±STD)	52.4 ± 9.7	52.2 ± 9.1	48.7 ± 10.3	0.472
AJCC Stage				0.027
- II	15 (65.2%)	16 (84.2%)	6 (40.0%)	
- III	8 (34.8%)	3 (15.8%)	9 (60.0%)	
Overall Survival at 3 years				0.595
- Alive	17 (73.9%)	14 (73.7%)	13 (86.7%)	
- Dead	6 (26.1%)	5 (26.3%)	2 (13.3%)	

RNA-seq data analysis

Raw sequencing reads were first trimmed and cleaned using Trim Galore (v0.4.1) [http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/]. Afterwards, reads were aligned to the human reference genome (hg38) using TopHat v2.1 [23]. Samtools (Version: 1.3.1) [24] were used to sort mapped reads, and HTSeq-count [25] was employed to enumerate reads associated with each human gene.

DESeq2 was used to perform differential expression analysis as per standard rules [https://bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.html]. Genes with adjusted P-values <0.05 and absolute fold changes ≥1.5 were considered as differentially expressed. Gene Ontology (GO) and KEGG pathway enrichment analysis of differentially expressed genes (DEGs) was performed using DAVID (Database for Annotation, Visualization and Integrated Discovery) version 6.8 [26]. Comparison of DEGs was accomplished with the online tool Venny [https://bioinfogp.cnb.csic.es/tools/venny/index.html]. Heatmaps were generated using the heatmap.2 function of the gplots R package [https://cran.r-project.org/web/packages/gplots/index.html].

The Cancer Genome Atlas (TCGA) validation

The UALCAN [27] (http://ualcan.path.uab.edu) integrative cancer data analysis portal was used to examine RNA expression of candidate genes in normal breast samples and TNBC samples from CA and AA patients. UALCAN uses Level 3 RNA-seq data and patients' clinical data of TCGA breast invasive carcinoma (BRCA) samples obtained via TCGA-assembler [28]. In total, 116 patient samples were categorized as TNBCs based on the immunohistochemical status of ER, PR, and HER2. Of those, 32 were related to AA and 69 to CA patients. Transcripts per million (TPM) values for each gene in each sample were obtained by multiplying the scaled estimate value by 1000,000.

Statistical analysis

Associations between clinical variables were determined by use of $\chi 2$ tests for categorical variables and t-tests for continuous variables. Fisher's exact tests were used when appropriate for small sample sizes. Unpaired t-tests were performed to assess the significance of differences between TPM values of normal and TNBC samples. Genes with adjusted P-values <0.05 and absolute fold changes of ≥1.5 were considered as differentially expressed using DESeq2.

Results

Overall clinical characteristics of tnbc patient cohort

Table 1 provides clinical and demographic information for the 57 patients included in this study. The largest group of participants was CAs ($n = 23$, 41%), followed by AAs ($n = 19$, 33%) and KEs ($n = 15$, 26%). All patients were female with a median age of 50 (range 33–70). KE women were younger at diagnosis, although there was no significant difference in age across the three race groups. All patients had localized disease, however, KE patients had higher-stage tumors (stage III vs stage II) compared to the other two groups (p-value 0.027). Clinical status (alive, disease free) at three years of follow-up (obtained through phone interviews) did not differ significantly across cohorts. Evidence of disease at 3 years of follow up was different among the races; however, this analysis was limited by missing data for KE patients since post-operative follow-up and diagnosis of any recurrence was accomplished at the local county clinics that often lacked adequate diagnostic facilities and adequate documentation.

Screening of differentially expressed genes on the basis of transcriptome analysis

Global transcriptome sequencing of KE TNBC ($n = 15$), CA TNBC ($n = 19$), and AA TNBC ($n = 23$) samples was performed to identify genes that showed KE TNBC-specific expression. Differential expression analysis identified 683 protein-coding genes as up-regulated and 768 protein-coding genes as down-regulated in KE TNBC samples compared to AA TNBC samples (Fig. 1A; **Supplementary** Table 1). Similarly, 353 protein coding genes were up-regulated, and 372 protein coding genes were down-regulated in KE TNBC samples compared to CA TNBC samples (Fig. 1B**; Supplementary** Table 2). Comparisons of these DEGs showed 160 genes commonly up-regulated and 178 genes commonly down-regulated in KE TNBC samples (Fig. 1C).

Distinct DEGs in KE TNBCs compared to CA and AA patients

The top 50 commonly up/down-regulated genes of the KE TNBCs, as compared to CAs and AAs, are shown in Fig. 2 and **Supplementary Table 3**. Some of the highly upregulated DEGs were DSC1, KRT6C, SPINK5, ACTC1, FCRL1, DEFB1, BCL2L10, SERPINB7, FLG2, INSM2, UHRF2, UHMK1, KDM4C, GFAP, LSM14A, PLA2G4E, ERCC6, and RBM24; downregulated genes included KCNH5, PSG11, AGT, CD14, KRT18, CSF1R, S100A13 and BCL9L.

Pathway analysis predicted a collagen-related enriched pathway for KE TNBCs

Functional analysis of common DEGs showed platelet degranulation, extracellular matrix disassembly, and collagen catabolism as the top enriched biological processes (Fig. 3A). Genes associated with lysosomes, protein processing in the endoplasmic reticulum, and spliceosome KEGG pathways were also enriched in common DEGs. The raw sequencing data were deposited in the NCBI Gene expression omnibus (GEO) [GSE142258, GSE142731].

Assessment of potential biomarkers using TCGA database for KE TNBCs

To identify potential biomarkers for KE TNBC samples, the RNA expression profile of 338 common DEGs (160 up-regulated and 178 down-regulated) was examined. Genes that showed no change in expression across normal breast, CA TNBC, and AA TNBC samples from TCGA dataset were of interest. This analysis revealed 45 KE TNBC-specific genes (**Fig. 4**), which included 25 up-regulated and 20 down-regulated genes.

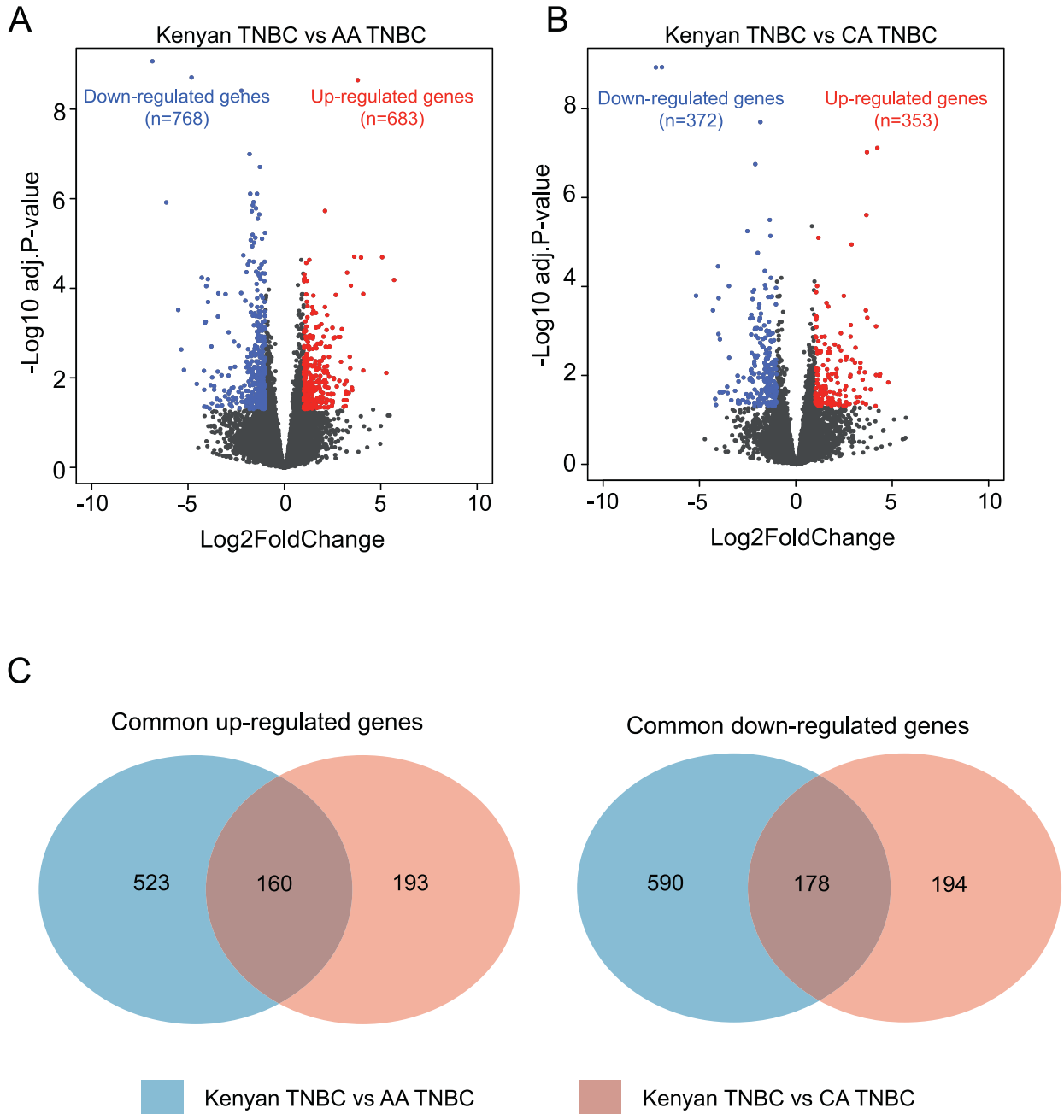

Fig. 1. Fig. 1: RNA-seq data analysis of Kenyan, African American and Caucasian TNBC samples. (A) Volcano plot showing 683 up-regulated protein coding genes and 768 down-regulated protein coding genes in Kenyan TNBC samples compared to African American TNBC samples, (B) Volcano plot showing 353 up-regulated protein coding genes and 372 down-regulated protein coding genes in Kenyan TNBC samples compared to Caucasian TNBC samples. The negative log 10 of adjusted p-value and the log 2 of the fold change is plotted on the Y and X-axis respectively. (C) Venn diagrams show commonly up-/down-regulated genes in Kenyan TNBC samples compared to both African American and Caucasian TNBC samples. Differential expression analysis between African American TNBC samples and Caucasian TNBC samples yielded 312 up-regulated and 177 down-regulated genes (not shown).

Among these KE TNBC-specific genes, we observed a) apoptosis-associated genes such as ACTC1, ERCC6, and CD14; b) cell proliferation-associated genes such as UHRF2, KDM4C, UHMK1, KCNH5, KRT18, CSF1R and S100A13; and c) WNT signaling pathway related BCL9L.

Discussion

The diseases of the developing world are not replicas of those in the West [29]. In addition to the numerous environment and socio-economic factors, genomic characteristics, linked to ancestral heritage, may be involved. The increased incidence of TNBCs and the early age of onset for young African women show a distinctive pattern [2–6]. The current study showed that TNBC patients from Kenya were younger with most presenting with higher stages as compared to their counterparts in Alabama. Prior studies show that younger breast cancer patients of minority ethnic backgrounds present with aggressive features and have poor survival [30,31].

Although genomic studies for some African countries have been conducted [7,8], it is notable that Africans are not a monolithic group. There are variations in genetic structure and exposures across African populations, and they vary by birthplace and ancestral heritage. In the present study, we sequenced the transcriptomes of KE TNBC patients seen at the

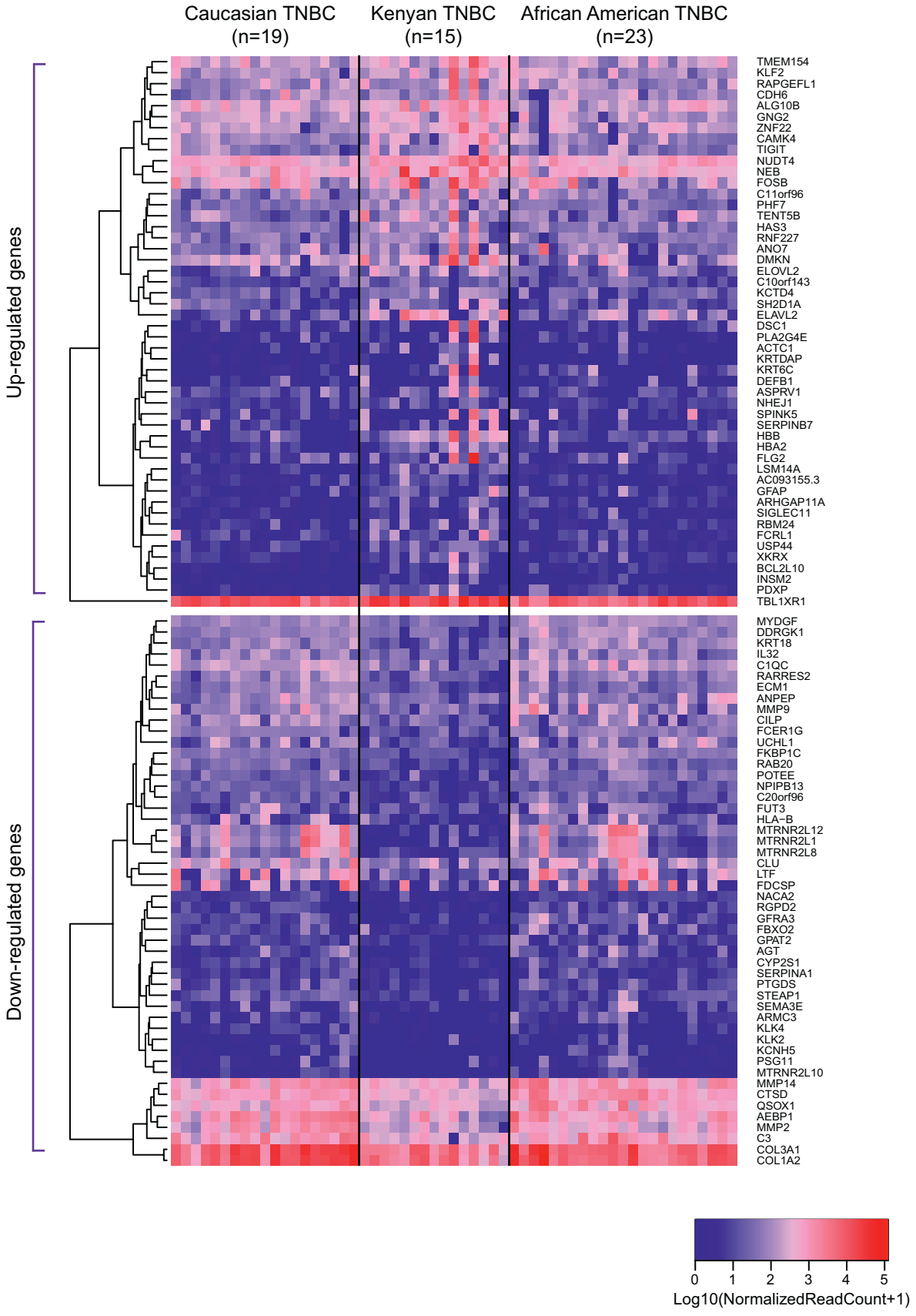

Fig. 2. Top 100 genes differentially expressed in KE TNBC samples. Heatmap showing list of top 50 up-regulated and 50 down-regulated genes (ordered by adj. P-value) in KE TNBC samples compared to AA and CA TNBC samples. “gplots” R package was used for generation of the heatmap.

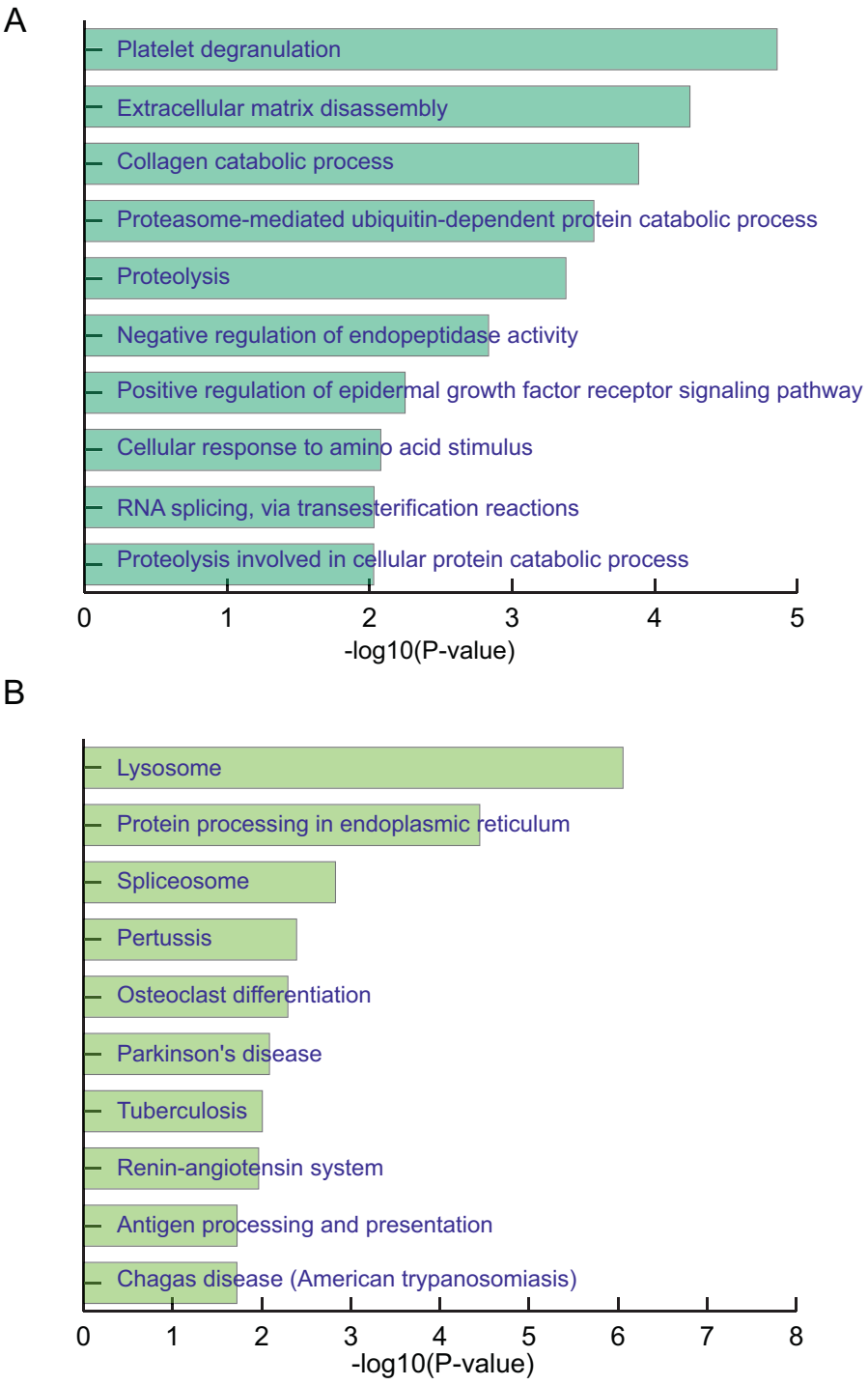

Fig. 3. KEGG pathway and gene ontology enrichment analysis of common, differentially expressed genes of KE TNBCs. (A) and (B) Top 10 biological processes and KEGG pathways enriched in common up-/down-regulated genes. The X-axis represents the negative log 10 of p-values.

Aga Khan University Hospital in Nairobi. Given the demography of the population in Nairobi and the location of the clinic, patients included in this study were predominantly of Bantu ancestry. Transcriptomes of KE TNBC patients were compared with those of CA and AA patients. *In silico* validation using TCGA BRCA dataset aided in narrowing down KE TNBC-specific biomarkers. Our patient population was comprised of 15 women from Kenya with confirmed TNBCs and 19 AA and 23 CA women with TNBCs from Alabama derived from larger cohorts used in our recently published study [20]. We reported previously that genomic profiles of TNBC patients differ for AA and CA ethnic groups; transcriptome analysis identified racial/ethnic-specific genes that were differentially expressed in TNBCs, showing a higher incidence of basal-like tumors and altered TP53, NFB1, and AKT pathways in AA TNBCs compared to CA TNBCs [20]. The higher number of stage III patients in this small KE cohort is a limitation of our current study and may have influenced our findings. However, the differences were identified not only between KE and CA patients, who often present with lower stage disease, but also with the AA cohort who presented with later stage disease compared to CA patients.

To the best of our knowledge, the current investigation is the first comparative transcriptomic profiling of KE patients predominantly of Bantu origin. Here, we compared the results from KE TNBC patients with those of AA and CA TNBCs to determine whether racial differences in genes translate into different clinical biomarkers. Given the small

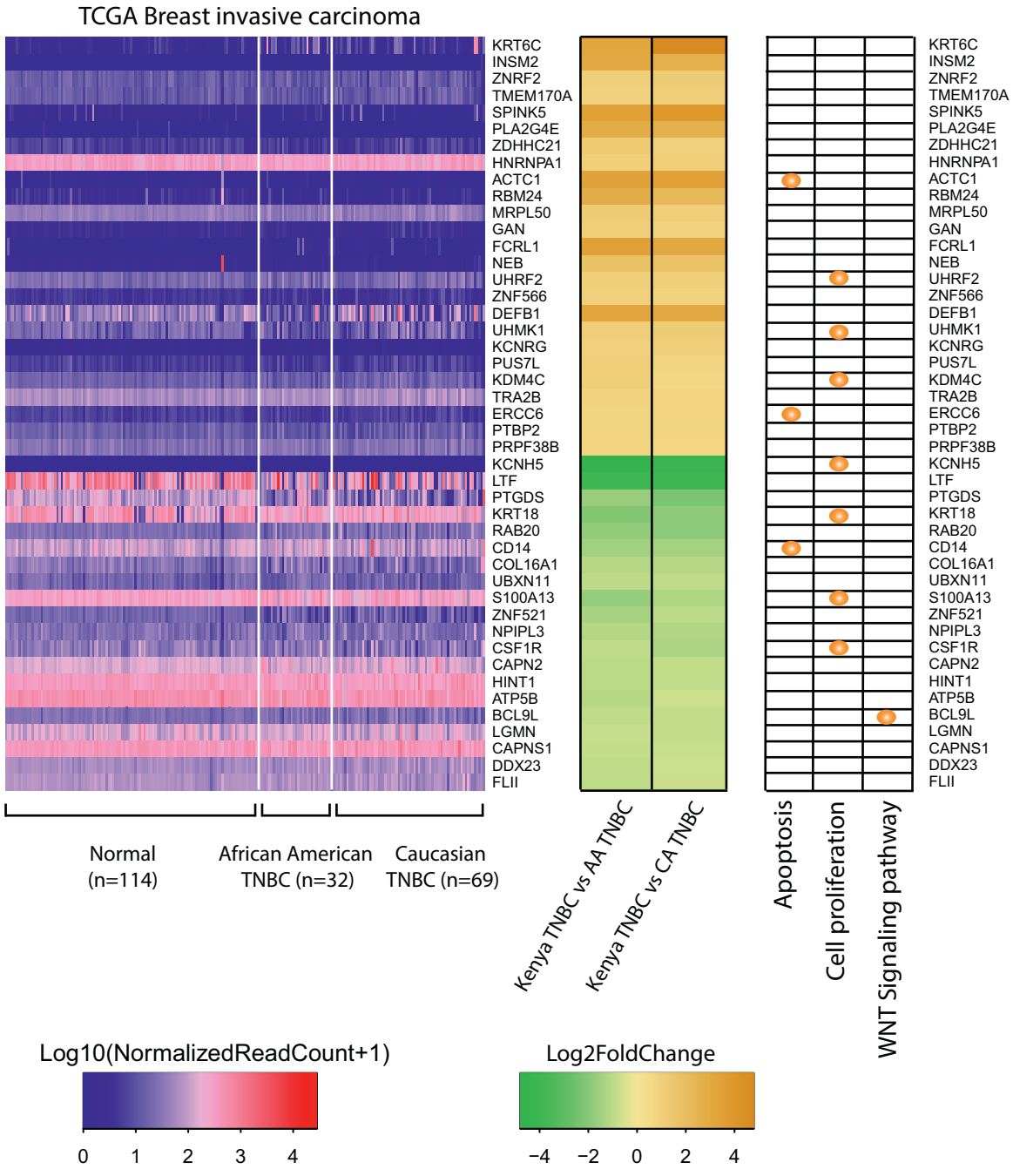

Fig. 4. Potential biomarkers specific for KE TNBCs. Heatmap showing a list of KE TNBC-specific 45 genes (25 up-regulated and 20 down-regulated), which, in TGCA BRCA dataset, show no difference in expression in normal breast, AA TNBC, and CA TNBC samples.

sample size of KE TNBCs, we were unable to perform a more detailed, matched study; a limitation that can be overcome by focusing on the identified difference with a larger, accurately matched population. Despite this limitation, the comparative analysis revealed intriguing findings. KE TNBCs had more genes upregulated and downregulated than those of AA or CA patients. Of the 523 and 193 genes upregulated when comparing KE TNBCs with AA TNBCs, respectively, 160 were shared between the two study populations. Similarly, of the 590 and 194 genes downregulated in KE and CA TNBCs, 178 were shared between the two populations. Women in low- and middle-income countries often present clinically with late stage disease. This is related to socio-economic inequities and lack of available medical resources [7,8] as opposed to de novo aggressive biology [9]. Thus, although the KE cohort included women with higher stages as compared to the AA and CA cohorts, this difference does not explain the differential genetic expression seen in our study. Our review of the transcriptomes revealed a subset of upregulated genes that are singularly overexpressed in KE TNBCs and could serve as biomarkers. These include factors involved in apoptosis, cell proliferation, and the WNT signaling pathway. In particular, dysregulated genes associated with the Wnt/β-catenin pathway, observed for KE TNBCs as compared to AA and CA TNBCs, suggest that this pathway contributes to the aggressive phenotypes of TNBCs of KE women.

Upregulated genes specific for KE TNBCs included (a) Actin Alpha Cardiac Muscle 1 (ACTC1), an actin coding gene that is a prognostic and invasion biomarker for glioblastoma [32] and a diagnostic biomarker for prostate cancer [33]. (b) ERCC Excision Repair 6 (ERCC6), a DNA-

binding protein whose elevated expression is associated with poor overall survival of colorectal cancer patients and resistance to 5-fluorouracil therapy [34]. In the Chinese population, ERCC6 polymorphisms are associated with risk for oral cancer and gastric cancer [35,36]. (c) CD14, a surface protein, whose over-expression in bladder cancer leads to cell proliferation and tumor growth [37]. In gastric cancer, CD14 promotes the epithelial-mesenchymal transition and invasion through TNF-alpha [38]. (d) B-Cell CLL/Lymphoma 9-Like (BCL9L), which is involved in the WNT signaling pathway and in development of breast cancers [39], pancreatic cancers [40], and colon cancers [41]. (e) Ubiquitin-like with PHD and Ring Finger Domains 2 (UHRF2), an E3 ubiquitin ligase involved in regulation of the cell cycle. UHRF2 is also involved in tumorigenesis of intrahepatic cholangeocarcinomas [42], hepatocellular carcinomas [43], colon cancers [44], and breast cancers [45]. However, in non-small cell lung carcinomas [46] and esophageal squamous cell carcinomas [47], UHRF2 functions as a tumor suppressor. (f) Lysine Demethylase 4C (KDM4C), a trimethylation-specific demethylase, which maintains chromosomal stability and promotes cell proliferation in TNBCs [48]. (g) U2AF Homology Motif Kinase 1 (UHMK1), a serine/threonine kinase that promotes the cell cycle. UHMK1 is used as an autoantibody biomarker for early detection of ovarian cancers [49]. (h) Potassium Voltage-Gated Channel Subfamily H Member 5 (KCNH5), which is hyper-methylated in adenocarcinomas of non-small cell lung cancers compared to squamous cell carcinomas [50]. (i) Keratin 18 (KRT18), which promotes growth of invasive breast cancers [51], colorectal cancers [52], and luminal-like prostate cancers [53]. (j) Colony Stimulating Factor 1 Receptor (CSF1R), a receptor for cytokine Colony Stimulating Factor 1 that regulates macrophage production, differentiation, and function. Blocking of CSF1R is a common therapeutic strategy for pancreatic cancers [54], prostate cancers [55], multiple myeloma [56], and acute myeloid leukemia [57]. (k) S100 Calcium Binding Protein A13 (S100A13), which is involved in export of secretory proteins. (l) S100A13, an angiogenic marker for human melanoma [58] and human astrocytic gliomas [59].

If confirmed with a larger cohort of KE TNBCs and validated in a confirmatory data set, these transcriptomic markers may serve as diagnostic biomarkers for KE TNBCs. Further studies are required to determine if these findings also hold for Kenyan women of other ancestral heritages. Several of these upregulated factors, identified in our study, are functional proteins necessary for cell growth, proliferation, migration, and metastasis. Thus, future studies should focus on targeting these molecules to develop new therapeutics, specifically for KE TNBCS.

In conclusion, this is the first transcriptome sequencing study investigating the gene expression profiles of KE TNBCs. We also performed integrative analysis with other datasets that shed light on the specific and common alterations in these TNBCs. This RNA sequencing analysis highlights the role of molecular alterations in TNBCs and the potential benefit of targeting pathways in this disease for the KE population as compared to AAs and CAs. The differentially expressed genes may have pathogenic roles for TNBC patients of KE origin. The small sample size, however, hinders us from making broad conclusions; future studies are needed to validate our findings by performing transcriptomic analyses, as well as the mutational profiling, of a large cohort of KE TNBCs. In addition, while we adjusted for stage in selection, there is a chance that unrecognized/misclassified metastatic disease may have occurred in KE patients. Nevertheless, the findings from this comparative transcriptome analysis reveal intriguing features that remain to be studied, confirmed, and actionably applied for diagnostic and therapeutic applications for TNBCs in low-resource countries. Moreover, the findings from this comparative transcriptome analysis have implications for understanding ethnic differences in relation to TNBCs and, after biological validation, can be applied for the diagnosis of aggressive phenotypes and as targets in the development of personalized therapeutics.

However, it is estimated that 90% of all biomarker research is done in resource-rich countries, while 90% of the global burden of disease is in resource-poor nations [60]. This imbalance in applied research adds both an urgency to results from primary studies and difficulty to validation efforts in resource poor settings. Specific challenges to biomarker research in resource-poor settings include: lack of standardized procedures of sample collection and management, biological diversity, disease heterogeneity, technical limitations, lack of resources/funding [61]. This investigation represents a strategy of addressing the above challenges through symmetric research partnership [62,63], and the authors urge continued pursuit of biomarker discovery/validation through this approach.

Author contributions

Conception and design: Mansoor Saleh, Sayed Shahin, Sooryanarayana Varambally, and Upender Manne

Financial support: Mansoor Saleh and Upender Manne

Administrative support: Hyung-Gyoon Kim

Provision of study materials or patients: Mansoor Saleh, Sayed Shahin, Isam-Eldin A. Eltoum, and Upender Manne

Collection and assembly of data: Mansoor Saleh, Shahin Sayed, Darshan Shimoga Chandrashekar, Sumit Agarwal, Hyung-Gyoon Kim, Asim Jamal Shaikh, and Zahir Moloo

Data analysis, generating figures, tables and interpretation: Darshan Shimoga Chandrashekar, Sumit Agarwal, and Michael Behring

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

Funding

These studies were supported by grants U54CA118948 (UM) and U54 CA118623 (CY) and by institutional funds (Department of Pathology and School of Medicine of the University of Alabama at Birmingham) awarded to MS, SV, and UM. We acknowledge the grant (P30CA013148) funded for the Tissue Biorepository of the UAB O'Neal CCC.

Context

Key Objective

Breast cancer often presents with distinct clinical outcomes among various ethnic populations. There is a need to identify prognostic and predictive biomarkers specific for triple-negative breast cancers (TNBCs) in understudied African patients such as the Kenyan (KE) population. This study specifically focused on KE patients of Bantu ancestry. Established biomarkers could help in understanding the biological mechanisms and pathways underlying the clinical phenotype of the disease.

Knowledge generated

This is the first study that identifies, by using a transcriptomic approach and integrated molecular data across three racial groups, potential biomarkers that are differentially expressed in KE TNBCs. Additionally, the TNBCs of young KE patients manifest aggressive characteristics and are more advanced as compared to those of AAs and CAs.

Relevance

Findings of the present study reflect the characteristics of TNBCs of KE patients and could contribute to understanding of the clinical nature of the disease. The identified biomarkers could be integrated with tumor biology and genomics data to describe the TNBC phenotype for KE women and, in the future, serve as predictive, prognostic, and therapeutic targets.

Supplementary Tables, Related to Methods

Declarations of Competing Interest

All authors state that there are no potential conflicts of interest.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2021.101086.

References

[1] C.K. Anders, et al., The evolution of triple-negative breast cancer: from biology to novel therapeutics, Am Soc Clin Oncol Educ Book 35 (2016) 34–42.
[2] L.A. Carey, et al., Race, breast cancer subtypes, and survival in the Carolina breast cancer study, JAMA 295 (21) (2006) 2492–2502.
[3] K.R. Bauer, et al., Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so–called triple-negative phenotype: a population-based study from the California cancer Registry, Cancer 109 (9) (2007) 1721–1728.
[4] N. Howlader, et al., US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status, J. Natl. Cancer Inst. 106 (5) (2014).
[5] H. Sung, et al., Breast cancer subtypes among Eastern-African-born black women and other black women in the United States, Cancer 125 (19) (2019) 3401–3411.
[6] L.A. Newman, L.M. Kaljee, Health disparities and triple-negative breast cancer in African American women: a review, JAMA Surg 152 (5) (2017) 485–493.
[7] F.O. Ademuyiwa, O.I. Olopade, Racial differences in genetic factors associated with breast cancer, Cancer Metastasis Rev. 22 (1) (2003) 47–53.
[8] S. Siddharth, D. Sharma, Racial disparity and triple-negative breast cancer in African-American women: a multifaceted affair between obesity, biology, and socioeconomic determinants, Cancers (Basel) 10 (12) (2018) 514.
[9] E. Hadgu, et al., Breast cancer in Ethiopia: evidence for geographic difference in the distribution of molecular subtypes in Africa, BMC Womens Health 18 (1) (2018) 40.
[10] J.D. Fackenthal, O.I. Olopade, Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations, Nat. Rev. Cancer 7 (12) (2007) 937–948.
[11] A.D. Deshpande, et al., Racial disparities in breast cancer survival: an analysis by age and stage, J. Surg. Res. 153 (1) (2009) 105–113.
[12] J. Mehrotra, et al., Estrogen receptor/progesterone receptor-negative breast cancers of young African-American women have a higher frequency of methylation of multiple genes than those of Caucasian women, Clin. Cancer Res. 10 (6) (2004) 2052–2057.
[13] J.J. Pitt, et al., Author correction: characterization of Nigerian breast cancer reveals prevalent homologous recombination deficiency and aggressive molecular features, Nat. Commun. 10 (1) (2019) 288.
[14] N. Wright, et al., Panoptic overview of triple-negative breast cancer in Nigeria: current challenges and promising global initiatives, J. Glob. Oncol. 4 (2018) 1–20.
[15] K. Stefflova, et al., Dissecting the within-Africa ancestry of populations of African descent in the Americas, PLoS One 6 (1) (2011) e14495.
[16] A. Stark, et al., African ancestry and higher prevalence of triple-negative breast cancer: findings from an international study, Cancer 116 (21) (2010) 4926–4932.
[17] S. Sayed, et al., Ethnicity and breast cancer characteristics in Kenya, Breast Cancer Res. Treat. 167 (2) (2018) 425–437.
[18] D. Huo, et al., Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer, J. Clin. Oncol. 27 (27) (2009) 4515–4521.
[19] E.J. Kantelhardt, et al., The prevalence of estrogen receptor-negative breast cancer in Ethiopia, BMC Cancer 14 (2014) 895.
[20] M. Davis, 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 12 (5) (2020).
[21] M.B. Amin, et al., The Eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging, CA Cancer J. Clin. 67 (2) (2017) 93–99.
[22] L.C. Bovell, et al., The prognostic value of microRNAs varies with patient race/ethnicity and stage of colorectal cancer, Clin. Cancer Res. 19 (14) (2013) 3955–3965.
[23] D. Kim, et al., TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, Genome Biol. 14 (4) (2013) R36.
[24] H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data, Bioinformatics 27 (21) (2011) 2987–2993.
[25] S. Anders, P.T. Pyl, W. Huber, HTSeq–a Python framework to work with high-throughput sequencing data, Bioinformatics 31 (2) (2015) 166–169.
[26] W. Huang da, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (1) (2009) 44–57.
[27] D.S. Chandrashekar, et al., UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses, Neoplasia 19 (8) (2017) 649–658.
[28] Y. Zhu, P. Qiu, Y. Ji, TCGA-assembler: open-source software for retrieving and processing TCGA data, Nat. Methods 11 (6) (2014) 599–600.
[29] D. Stuckler, Population causes and consequences of leading chronic diseases: a comparative analysis of prevailing explanations, Milbank Q. 86 (2) (2008) 273–326.
[30] V. Fabiano, et al., Breast cancer in young women presents with more aggressive pathologic characteristics: retrospective analysis from an Argentine national database, JCO Glob. Oncol. 6 (2020) 639–646.
[31] E. Copson, et al., Ethnicity and outcome of young breast cancer patients in the United Kingdom: the POSH study, Br. J. Cancer 110 (1) (2014) 230–241.
[32] S. Ohtaki, et al., ACTC1 as an invasion and prognosis marker in glioma, J. Neurosurg. 126 (2) (2017) 467–475.
[33] H.C. Huang, et al., Discovering disease-specific biomarker genes for cancer diagnosis and prognosis, Technol. Cancer Res. Treat. 9 (3) (2010) 219–230.
[34] Z. Zhao, G. Zhang, W. Li, Elevated expression of ERCC6 Confers resistance to 5-fluorouracil and is associated with poor patient survival in colorectal cancer, DNA Cell Biol. 36 (9) (2017) 781–786.
[35] C.F. Chiu, et al., A novel single nucleotide polymorphism in ERCC6 gene is associated with oral cancer susceptibility in Taiwanese patients, Oral Oncol. 44 (6) (2008) 582–586.
[36] J.W. Liu, et al., The DNA repair gene ERCC6 rs1917799 polymorphism is associated with gastric cancer risk in Chinese, Asian Pac. J. Cancer Prev. 14 (10) (2013) 6103–6108.
[37] M.T. Cheah, et al., CD14-expressing cancer cells establish the inflammatory and proliferative tumor microenvironment in bladder cancer, Proc. Natl. Acad. Sci. U. S. A. 112 (15) (2015) 4725–4730.
[38] K. Li, et al., CD14 regulates gastric cancer cell epithelialmesenchymal transition and invasion in vitro, Oncol. Rep. 30 (6) (2013) 2725–2732.
[39] N. Zatula, et al., The BCL9-2 proto-oncogene governs estrogen receptor alpha expression in breast tumorigenesis, Oncotarget 5 (16) (2014) 6770–6787.
[40] G. Sannino, et al., Role of BCL9L in transforming growth factor-beta (TGF–beta)-induced epithelial-to-mesenchymal-transition (EMT) and metastasis of pancreatic cancer, Oncotarget 7 (45) (2016) 73725–73738.
[41] F.H. Brembeck, et al., BCL9-2 promotes early stages of intestinal tumor progression, Gastroenterology 141 (4) (2011) 1359–1370 1370 e1-3.
[42] R. Peng, et al., Overexpression of UHRF2 in intrahepatic cholangiocarcinoma and its clinical significance, Onco Targets Ther 10 (2017) 5863–5872.
[43] T. Zhang, et al., UHRF2 decreases H3K9ac expression by interacting with it through the PHD and SRA/YDG domain in HepG2 hepatocellular carcinoma cells, Int. J. Mol. Med. 39 (1) (2017) 126–134.
[44] S. Lu, et al., Ubiquitin-like with PHD and ring finger domains 2 is a predictor of survival and a potential therapeutic target in colon cancer, Oncol. Rep. 31 (4) (2014) 1802–1810.
[45] J. Wu, et al., Identification and functional analysis of 9p24 amplified genes in human breast cancer, Oncogene 31 (3) (2012) 333–341.
[46] C. Jin, et al., Loss of UHRF2 is associated with non-small cell lung carcinoma progression, J. Cancer 9 (17) (2018) 2994–3005.
[47] T. Iguchi, et al., Identification of UHRF2 as a negative regulator of epithelial-mesenchymal transition and its clinical significance in esophageal squamous cell carcinoma, Oncology 95 (3) (2018) 179–187.
[48] J. Garcia, F. Lizcano, KDM4C activity modulates cell proliferation and chromosome segregation in triple-negative breast cancer, Breast Cancer 10 (2016) 169–175.
[49] B.A. Katchman, et al., Autoantibody biomarkers for the detection of serous ovarian cancer, Gynecol. Oncol. 146 (1) (2017) 129–136.
[50] S.E. Hawes, et al., DNA hypermethylation of tumors from non-small cell lung cancer (NSCLC) patients is associated with gender and histologic type, Lung Cancer 69 (2) (2010) 172–179.
[51] S.C. Doebar, et al., Gene expression differences between ductal carcinoma in situ with and without progression to invasive breast cancer, Am. J. Pathol. 187 (7) (2017) 1648–1655.
[52] J. Zhang, S. Hu, Y. Li, KRT18 is correlated with the malignant status and acts as an oncogene in colorectal cancer, Biosci. Rep. 39 (8) (2019).
[53] S.G. Zhao, et al., Associations of luminal and basal subtyping of prostate cancer with prognosis and response to androgen deprivation therapy, JAMA Oncol. 3 (12) (2017) 1663–1672.
[54] Y. Zhu, et al., CSF1/CSF1R blockade reprograms tumor-infiltrating macrophages and improves response to T-cell checkpoint immunotherapy in pancreatic cancer models, Cancer Res. 74 (18) (2014) 5057–5069.
[55] J. Xu, et al., CSF1R signaling blockade stanches tumor-infiltrating myeloid cells and improves the efficacy of radiotherapy in prostate cancer, Cancer Res. 73 (9) (2013) 2782–2794.
[56] Q. Wang, et al., Therapeutic effects of CSF1R-blocking antibodies in multiple myeloma, Leukemia 32 (1) (2018) 176–183.
[57] D.K.t. Edwards, et al., CSF1R inhibitors exhibit antitumor activity in acute myeloid leukemia by blocking paracrine signals from support cells, Blood 133 (6) (2019) 588–599.
[58] D. Massi, et al., S100A13 is a new angiogenic marker in human melanoma, Mod. Pathol. 23 (6) (2010) 804–813.
[59] M. Landriscina, et al., S100A13, a new marker of angiogenesis in human astrocytic gliomas, J. Neurooncol. 80 (3) (2006) 251–259.
[60] S.K. Sgaier, et al., Biobanks in Developing Countries: needs and Feasibility, Science 318 (5853) (2007) 1074–1075.
[61] S. Gupta, et al., Challenges and prospects for biomarker research: a current perspective from the developing world, Biochimica et Biophysica Acta (BBA) - Proteins Proteomics 1844 (5) (2014) 899–908.
[62] W. Krull, Editorial: helping to create symmetric partnerships: a new approach to supporting research in Sub-Saharan Africa, Trop. Med. Int. Health 10 (2) (2005) 118–120.
[63] D.S. Lawrence, L.A. Hirsch, Decolonising global health: transnational research partnerships under the spotlight, Int. Health 12 (6) (2020) 518–523.