Ethnic-Specific Gene Expression Patterns Impacting Prostate Cancer in African-American Men

Junseok Lee[1] **& Roberto Aguilar**[2]

Received May 6, 2025
Accepted September 25, 2025
Electronic access November 15, 2025

Prostate cancer (PCa) exhibits pronounced racial disparities in incidence, aggressiveness, and mortality, with African American (AA) men experiencing earlier onset and poorer outcomes compared to European American (EA) men. To investigate ancestry-associated molecular differences, we analyzed RNA-sequencing data (GSE104131) from high-grade (Gleason score ≥ 7 (4+3)) PCa tumors and matched adjacent normal prostate tissues of AA and EA patients using DESeq2 and Gene Set Enrichment Analysis (GSEA). Differential expression analysis revealed no individual genes significantly altered between AA and EA tumors; however, pathway-level analysis identified 33 enriched gene sets distinguishing the two groups. AA tumors displayed upregulation of keratinization and cornified envelope formation pathways, and downregulation of IgG immunoglobulin complex and endoplasmic reticulum protein-folding processes. Within-group analyses showed 574 differentially expressed genes (DEGs) in AA tumors and 330 in EA tumors ($(|\log_2 \mathrm{FC}| > 1,\ \mathrm{FDR} < 0.05)$). In AA tumors, notable upregulated genes included AMACR and NPY, with enrichment of xenobiotic detoxification and proliferation pathways, alongside strong downregulation of epithelial-to-mesenchymal transition (EMT) and stromal activation programs. EA tumors were characterized by upregulation of androgen-responsive epithelial programs and protein synthesis machinery, and downregulation of striated muscle development and inflammatory signaling. These findings suggest ancestry-specific differences in epithelial differentiation, immune signaling, and drug resistance mechanisms, with keratin pathway enrichment in AA tumors potentially linked to aggressive phenotypes and poorer androgen deprivation therapy response. This study underscores the utility of pathway-level approaches for detecting subtle but coordinated molecular differences and highlights the need for ancestry-informed biomarker discovery and therapeutic strategy development in PCa.
Keywords: prostate cancer, African-American men, health disparity, RNA-seq, differentially expressed genes, deregulated pathways, DESeq2, GSEA, biomarkers

Introduction

Prostate cancer (PCa) is the most common cancer in men in the United States, other than skin cancer. Mortality-wise, PCa is the second-leading cause of cancer death in American men, behind only lung cancer. About 1 in 44 men will die of prostate cancer. The American Cancer Society's estimates for PCa in the United States for 2025 are about 313,780 new cases of PCa and about 35,770 deaths from PCa[1].

Prostate cancer has the largest racial disparities of any cancer in the United States. PCa in African-American (AA) men exhibits the largest racial disparities in cancer-related mortality, with more than double the risk of death compared to other ethnic groups[2]. Black men have a risk of diagnosis that is 1.67 times (180 vs 108 per 100,000) and a risk of cancer death that is 2.06 times (36 vs 17 per 100,000) that of White men[3].

While much of these health disparities have been ascribed to socio-economic differences and access to care[4], previous studies show that AA men are at higher risk of being diagnosed with PCa, present more frequently with high-grade disease, are at increased risk of biochemical progression after radical prostatectomy (RP), and have higher expression of adverse molecular features, even in an equal access setting[5]. This strongly suggests that biological differences contribute to PCa racial disparities.

Recent studies further reveal that AA men with PCa exhibit unique biological profiles, including altered androgen receptor signaling, genomic instability, metabolic dysregulation, and inflammatory signaling, all of which may contribute to more aggressive disease progression[6]. Despite decades of research, the specific biological mechanisms underlying these disparities remain incompletely understood. There is a need for studies that integrate molecular characterization in racially diverse patient cohorts to identify biological differences that may underlie disparities in PCa[7]. By defining these differences at the pathway level, such research can effectively uncover biological systems perturbed in ethnic-specific tumor cells, guiding future studies toward the most promising targets for prevention, precision medicine, and therapy development[8].

Building on this need, this study aims to understand the biological underpinnings of PCa racial disparities by identifying

[1] *Western Reserve Academy, Hudson, OH*
[2] *Cancer Immunology, Western Reserve Academy, Hudson, OH*

and characterizing deregulated biological pathways in African American men, who experience a higher incidence and more aggressive forms of the disease. Given prior evidence of racial disparities in prostate cancer outcomes and immune-related pathways, we hypothesize that AA prostate tumors would exhibit distinct immune signaling profiles compared to EA tumors [9–13].

Methods

1. RNA-seq Data Acquisition and Preprocessing

RNA-seq raw count data for matched high-grade (Gleason score ≥ 7 (4+3)) prostate tumor and adjacent normal tissue samples from African-American (AA) and European-American (EA) patients were obtained from the Gene Expression Omnibus (GEO; accession GSE104131) via the National Center for Biotechnology Information (NCBI) RNA-seq counts resource [14].

The original dataset comprised 32 unique biological samples: 8 tumor samples from AA patients, 8 tumor samples from EA patients, 8 adjacent normal tissue samples from AA patients, and 8 adjacent normal tissue samples from EA patients. Each sample was sequenced in two technical replicates.

The NCBI RNA-seq processing pipeline aligns sequencing runs labeled as Homo sapiens and transcriptomic to GRCh38.p13 human genome assembly (NCBI Assembly GCA_000001405.15) using HISAT2. Runs passing 50% alignment rate are then processed with Subread featureCounts to produce gene-level raw counts, with counts from multiple SRA runs for the same GEO sample summed to yield a single count value per gene per sample. Data derived from single cell samples are skipped, and values in raw count matrices are rounded for compatibility with common differential expression analysis software like DESeq2, edgeR or limma voom [15].

In the processed raw count matrix downloaded from GEO, 4 samples from the original dataset were absent due to NCBI pipeline exclusions: 2 adjacent normal tissue samples from EA patients, 1 tumor sample from an EA patient, and 1 adjacent normal tissue sample from an AA patient. The final dataset analyzed, therefore, consisted of 28 biological GEO samples, all generated under uniform experimental conditions within a single study.

2. Quality Control and Normalization

All analyses were conducted in R version 4.4.3 using the DESeq2 package (version 1.46.0) for differential expression analysis [16]. Genes with zero total counts across all samples were removed prior to analysis. Raw counts were normalized using DESeq2's median-of-ratios method to adjust for sequencing depth and RNA composition differences. PCA was performed on variance-stabilized count data to assess sample clustering by condition and verify the absence of batch effects. All samples originated from the same dataset (GSE104131) and were processed using a uniform protocol, minimizing the likelihood of batch effects.

This quality control ensured that observed differences in gene expression between AA and EA tumors reflected biological variation rather than technical variation, which was critical for robustly testing our hypothesis.

3. Differential Expression Analysis and GSEA Analysis

Differential expression was assessed using the Wald test, and p-values were adjusted for multiple testing with the Benjamini-Hochberg false discovery rate (FDR) method. Genes with $|\log_2$ fold change (FC) $| > 1$ and an adjusted p-value (FDR) < 0.05 were considered statistically significant. Gene annotations were obtained using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) Bioinformatics Resources 2021 update [17].

Gene Set Enrichment Analysis (GSEA) was performed using the clusterProfiler package in R, with all genes ranked by $\log_2$ fold changes from the differential expression analysis [18]. Gene sets were obtained from the Human Molecular Signatures Database (MSigDB) via the msigdbr package in R for Homo sapiens [19]. Enrichments were considered significant at an adjusted p-value (FDR) < 0.05 and an $|$ Normalized Enrichment Score (NES) $| > 1$. Representative enriched pathways were visualized using the enrichplot package in R, specifically with the gseaplot2 function [20].

Results

We compared tumor and matched normal tissues within African-American (AA) and European-American (EA) groups, and between AA and EA tumors. Results are presented for each contrast below.

1. AA Tumor and EA Tumor Samples

Direct comparison of prostate cancer (PCa) samples from AA and EA patients did not yield any differentially expressed genes (DEGs) meeting the significance threshold ($|\log_2$ fold change (FC) $| > 1$ and adjusted p-value (FDR) < 0.05).

However, Gene Set Enrichment Analysis (GSEA) using the full ranked gene list revealed 33 significantly enriched pathways (FDR < 0.05, $|$ Normalized Enrichment Score (NES) $| > 1$). Positively enriched pathways in AA tumors included GOBP keratinization (NES = 1.51, FDR = 0.0074), REACTOME formation of the cornified envelope (NES = 1.48, FDR = 0.0032), REACTOME keratinization (NES = 1.46, FDR = 0.000015), and pathways related to olfactory signaling and transduction (Table 1). Negatively enriched pathways included DURANTE adult olfactory neuroepithelium plasma cells (NES = -3.04, FDR

= 0.000015), GOCC IgG immunoglobulin complex (NES = -2.73, FDR = 0.000044), and multiple ER protein-folding and transport processes such as the calnexin/calreticulin cycle (NES = -2.62, FDR = 0.0089) (Table 2).

These results suggest that while no single gene reached statistical significance, coordinated expression changes in pathways related to keratinization, immune complex formation, and ER proteostasis may underlie ancestry-associated differences in prostate cancer biology.

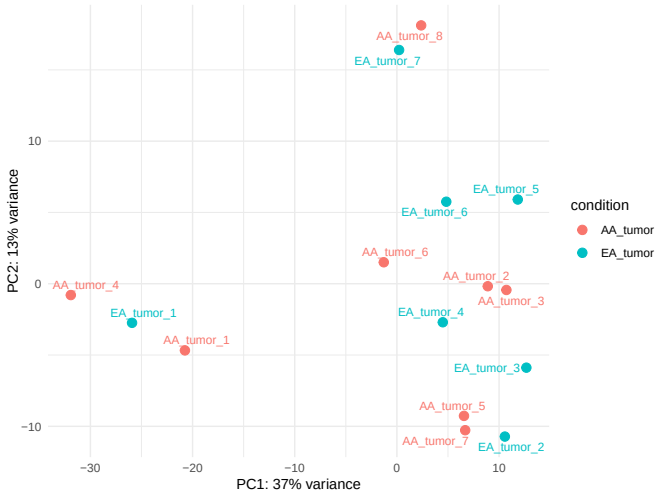


Fig. 1 PCA of AA Tumor versus EA Tumor. PCA plot of African American (AA, red) and European American (EA, blue) prostate tumor samples showing partial separation by ancestry. The PCA plot shows the absence of batch effect due to technical issues.

Table 1 Top upregulated pathways (FDR < 0.05 and NES > 1). The top 5 pathways with the highest NES were selected.

Pathways	NES	FDR
GOBP Keratinization	1.51	0.007
REACTOME Formation of the Cornified Envelope	1.48	0.0032
REACTOME Keratinization	1.46	1.48×10^{-5}
REACTOME Olfactory Signaling Pathway	1.31	0.0021
KEGG Olfactory Transduction	1.3	0.0032

2. AA Tumor and AA Adjacent Normal Samples

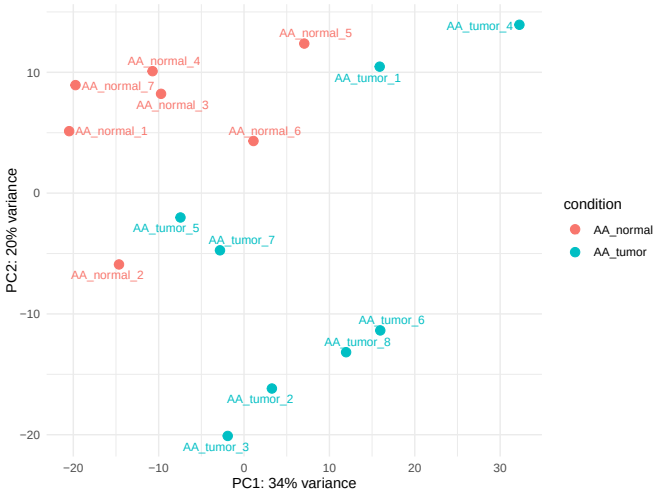
Differential expression analysis between AA tumor and adjacent normal samples identified 926 genes with an FDR < 0.05. Of these, 574 DEGs (330 upregulated, 244 downregulated genes) also had an $\mid \log_2 \text{FC} \mid > 1$ and were considered DEGs. The

Table 2 Top downregulated pathways (FDR < 0.05 and NES < -1). The top 5 pathways with the lowest NES were selected.

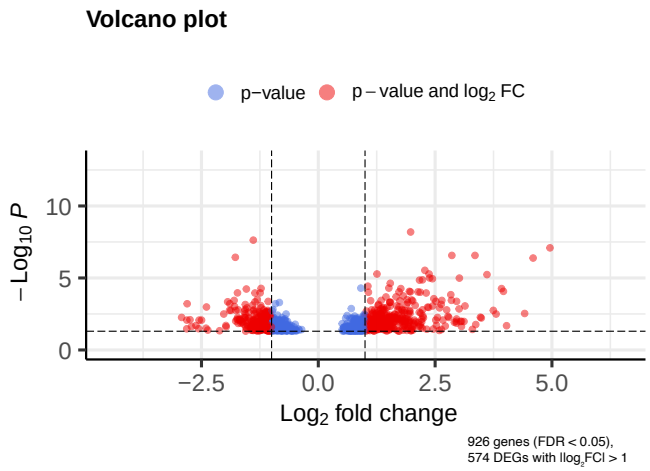
Pathways	NES	FDR
DURANTE Adult Olfactory Neuroepithelium Plasma Cells	-3	1.48×10^{-5}
GOCC IgG Immunoglobulin Complex	-2.7	4.41×10^{-5}
GOBP Endoplasmic Reticulum to Cytosol Transport	-2.7	0.0042
MODULE 155	-2.7	0.0074
REACTOME Calnexin/Calreticulin Cycle	-2.6	0.0089

volcano plot displays all FDR-significant genes, with those exceeding both statistical and fold-change thresholds highlighted in red (Figure 2B).

We ranked genes by $\log_2$ fold change among those meeting the significance threshold (FDR < 0.05) and identified 10 upregulated and 10 downregulated genes in AA tumor versus AA adjacent normal samples. Among the upregulated genes, AMACR had the highest fold change ($\log_2$ FC = 4.96), followed by C1QTNF3-AMACR ($\log_2$ FC = 4.60), ANKRD34B ($\log_2$ FC = 4.42), NPY ($\log_2$ FC = 4.03), MUC6 ($\log_2$ FC = 3.96), CRISP3 ($\log_2$ FC = 3.92), LOC105377527 ($\log_2$ FC = 3.81), OR51E1 ($\log_2$ FC = 3.75), TRC-GCA6-1 ($\log_2$ FC = 3.61), and DLX1 ($\log_2$ FC = 3.49) (Table 3). The top downregulated genes included IGHG1 ($\log_2$ FC = -2.93), C1QL1 ($\log_2$ FC = -2.82), IGHGP ($\log_2$ FC = -2.81), WIF1 ($\log_2$ FC = -2.81), IGKV3-20 ($\log_2$ FC = -2.75), IGKV1-5 ($\log_2$ FC = -2.71), CCDC54-AS1 ($\log_2$ FC = -2.62), IGKC ($\log_2$ FC = -2.56), OLFM4 ($\log_2$ FC = -2.55), and NELL2 ($\log_2$ FC = -2.50) (Table 4).



(A) PCA of AA tumor versus adjacent normal prostate tissue



(B) Volcano plot of DEGs in PCa and adjacent normal prostate tissue from AA patients

Fig. 2 Plots comparing PCa and adjacent normal prostate tissue from AA patients. (A) PCA plot of AA tumor (blue) and AA adjacent normal (red) samples showing separation. The PCA plot shows the absence of batch effect due to technical issues. (B) Volcano plot shows DEGs between PCa and adjacent noncancerous prostate tissue in AA patients. The horizontal axis is the $\log_2$ FC between PCa and adjacent noncancerous prostate tissue. The negative $\log_{10}$ of the adjusted p-value (FDR) obtained from DESeq2 is plotted on the vertical axis. Each gene is represented by one point on the graph. The red spots represent significant DEGs ($\log_2$ FC > 1: upregulated; $\log_2$ FC < -1: downregulated). The blue spots represent genes that had FDR < 0.05 but didn't match the $|\log_2 \text{FC}| > 1$ threshold.

GSEA identified a total of 920 deregulated pathways with FDR < 0.05 and $|\log_2 \text{FC}| > 1$, including 25 upregulated and 895 downregulated pathways in PCa and adjacent normal prostate tissue from AA patients.

Among positively enriched pathways, notable examples included the LIU prostate cancer up signature (NES = 1.62, FDR = 3.72×10^{-4}) reflects activation of established tumor-associated transcriptional programs, while GNF2 MKI67 (NES = 1.60, FDR = 0.0492) underscores enhanced proliferative activity. Additionally, upregulation of xenobiotic detoxification regulators (NES = 1.76, FDR = 0.0100) suggests potential for treatment resistance, and negative regulation of glycogen metabolic processes (NES = 1.71, FDR = 0.0315) indicates shifts in metabolic demands associated with tumorigenesis (Table 5).

The most strongly downregulated pathways included the TOMLINS prostate cancer dn signature (NES = -4.16, $FDR = 1.30 \times 10^{-18}$), a gene set defined by reduced expression in prostate tumors relative to benign tissue (Table 6).

Table 3 Top 10 upregulated genes in AA tumor versus AA adjacent normal prostate tissue. The ten most upregulated genes in African-American (AA) prostate tumor tissue relative to matched adjacent normal tissue, ranked by $\log_2$ FC among those with FDR < 0.05.

Genes	Gene Annotation	$\log_2$ FC	FDR
AMACR	alpha-methylacyl-CoA racemase	4.96	8.11×10^{-8}
C1QTNF3-AMACR	C1QTNF3-AMACR readthrough (NMD candidate)	4.6	4.14×10^{-7}
ANKRD34B	ankyrin repeat domain 34B	4.42	0.00293
NPY	neuropeptide Y	4.03	0.0207
MUC6	mucin 6, oligomeric mucus/gel-forming (gene/pseudogene)	3.96	8.51×10^{-5}
CRISP3	cysteine rich secretory protein 3	3.92	5.46×10^{-5}
LOC105377527	uncharacterized LOC105377527	3.81	0.0032
OR51E1	olfactory receptor family 51 subfamily E member 1	3.75	0.00178
TRC-GCA6-1	tRNA-Cys (anticodon GCA) 6-1	3.61	5.84×10^{-6}
DLX1	distal-less homeobox 1	3.49	0.00602

Table 4 Top 10 downregulated genes in AA tumor versus AA adjacent normal prostate tissue. The ten most downregulated genes in AA prostate tumor tissue relative to matched adjacent normal tissue, ranked by $\log_2$ FC among those with FDR < 0.05.

Genes	Gene Annotation	$\log_2$ FC	FDR
IGHG1	immunoglobulin heavy constant gamma 1 (G1m marker)	-2.93	0.00551
C1QL1	complement C1q like 1	-2.82	0.0339
IGHGP	immunoglobulin heavy constant gamma P (non-functional)	-2.81	0.00847
WIF1	WNT inhibitory factor 1	-2.81	6.19×10^{-4}
IGKV3-20	immunoglobulin kappa variable 3-20	-2.75	0.00796
IGKV1-5	immunoglobulin kappa variable 1-5	-2.71	0.0236
CCDC54-AS1	CCDC54 antisense RNA 1	-2.62	0.02
IGKC	immunoglobulin kappa constant	-2.56	0.0332
OLFM4	olfactomedin 4	-2.55	0.00844
NELL2	neural EGFL like 2	-2.5	0.00847

3. EA Tumor and EA Adjacent Normal Samples

Differential expression analysis between the EA tumor and adjacent normal samples identified 540 genes with an FDR < 0.05. Of these, 330 DEGs (212 upregulated, 118 downregulated genes) also had an $|\log_2 \text{FC}| > 1$ and were considered

Table 5 Top 10 upregulated pathways in AA tumor versus AA adjacent normal prostate tissue. The ten most upregulated pathways in AA prostate tumor tissue relative to matched adjacent normal tissue, ranked by NES among those with FDR < 0.05.

Pathways	NES	FDR
chr15q11	1.78	3.51×10^{-9}
chr21q11	1.77	0.0149
GOBP regulation of xenobiotic detoxification by transmembrane export across the plasma membrane	1.76	0.01
GOBP negative regulation of glycogen metabolic process	1.71	0.0315
chr11q11	1.69	0.0453
LIU prostate cancer up	1.62	3.72×10^{-4}
GNF2 MKI67	1.6	0.0492
ZHAN multiple myeloma PR up	1.55	0.033
HOWARD NK cell inact monov influenza A Indonesia 05 2005 H5N1 age 18 49yo 3dy up	1.55	0.0189
GNF2 HMMR	1.52	0.0407

Table 6 Top 10 downregulated pathways in AA tumor versus AA adjacent normal prostate tissue. The ten most downregulated pathways in AA prostate tumor tissue relative to matched adjacent normal tissue, ranked by NES among those with FDR < 0.05.

Pathways	NES	FDR
TOMLINS prostate cancer dn	-4.16	1.30×10^{-18}
GAVISH 3CA malignant metaprogram 12 EMT 1	-3.97	7.53×10^{-17}
GAVISH 3CA metaprogram fibroblasts pericyte like	-3.82	1.49×10^{-14}
GAVISH 3CA metaprogram epithelial EMT like 1	-3.77	6.52×10^{-14}
GAVISH 3CA metaprogram endothelial endo 6	-3.58	1.37×10^{-11}
GAVISH 3CA metaprogram epithelial interferon MHC	-3.51	5.77×10^{-11}
PAPASPYRIDONOS unstable atherosclerotic plaque dn	-3.47	1.05×10^{-10}
GAVISH 3CA malignant metaprogram 18 interferon MHC II 2	-3.45	1.94×10^{-10}
GAVISH 3CA metaprogram endothelial HEV 1	-3.4	6.44×10^{-10}
WINTER hypoxia dn	-3.39	5.77×10^{-11}

DEGs. The volcano plot displays all FDR-significant genes, with those exceeding both statistical and fold-change thresholds highlighted in red (Figure 3B). Among these, we ranked genes by $\log_2$ fold change and selected the ten most upregulated and ten most downregulated genes in EA tumor versus EA adjacent normal tissue.

The top upregulated genes included CRISP3 ($\log_2$ FC = 5.90), a cysteine-rich secretory protein, followed by OR51E2 ($\log_2$ FC = 4.39), OR51S1 ($\log_2$ FC = 3.63), PCAT14 ($\log_2$ FC = 3.57), LOC105370113 ($\log_2$ FC = 3.33), SIM2 ($\log_2$ FC = 3.22), TDRD1 ($\log_2$ FC = 3.17), LOC105374286 ($\log_2$ FC = 2.95), TFF3 ($\log_2$ FC = 2.94), and LOC107985492 ($\log_2$ FC = 2.90) (Table 7).

The most downregulated gene was TGM4 ($\log_2$ FC = -4.02), followed by PENK ($\log_2$ FC = -3.18), SERPINB11 ($\log_2$ FC = -2.84), LTF ($\log_2$ FC = -2.61), TCAP ($\log_2$ FC = -2.60), FAM167A ($\log_2$ FC = -2.37), LRP2 ($\log_2$ FC = -2.17), IQSEC3 ($\log_2$ FC = -1.82), APOBEC3C ($\log_2$ FC = -1.80), and PGM5-AS1 ($\log_2$ FC = -1.79) (Table 8).

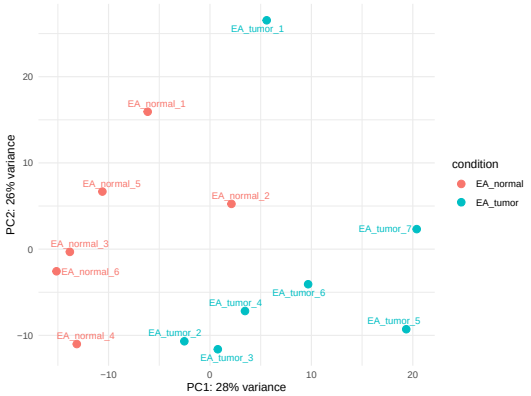
Table 7 Top 10 upregulated genes in EA tumor versus EA adjacent normal prostate tissue. The ten most upregulated genes in EA prostate tumor tissue relative to matched adjacent normal tissue, ranked by $\log_2$ FC among those with FDR < 0.05.

Genes	Gene Annotation	$\log_2$ FC	FDR
CRISP3	cysteine rich secretory protein 3	5.9	0.00506
OR51E2	olfactory receptor family 51 subfamily E member 2	4.39	6.15×10^{-7}
OR51S1	olfactory receptor family 51 subfamily S member 1	3.63	0.0245
PCAT14	prostate cancer associated transcript 14	3.57	6.90×10^{-6}
LOC105370113	uncharacterized LOC105370113	3.33	0.042
SIM2	SIM bHLH transcription factor 2	3.22	1.25×10^{-5}
TDRD1	tudor domain containing 1	3.17	0.00842
LOC105374286	uncharacterized LOC105374286	2.95	0.027
TFF3	trefoil factor 3	2.94	0.0287
LOC107985492	uncharacterized LOC107985492	2.9	0.0155

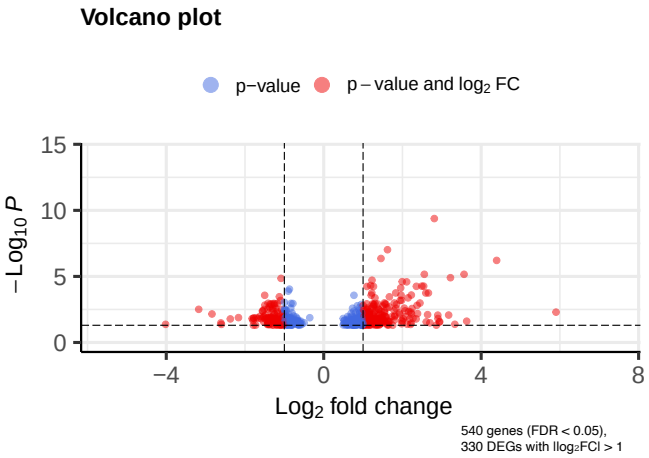
GSEA identified a total of 1880 deregulated pathways with FDR < 0.05 and $|\log_2 \text{FC}| > 1$, including 1786 upregulated and 94 downregulated pathways in PCa and adjacent normal prostate tissue from EA patients.

Among positively enriched pathways, the LIU prostate cancer up signature (NES = 3.74, FDR = 5.49×10^{-37}) reflected strong activation of previously reported tumor-associated transcriptional programs, while the GAVISH 3CA malignant metaprogram 9 unfolded protein response (NES = 3.10, FDR = 3.37×10^{-12}) indicated heightened endoplasmic reticulum stress adaptation. Upregulation of cytoplasmic ribosomal protein pathways (NES = 2.97, FDR = 1.18×10^{-13}) and translation initiation (NES = 2.94, FDR = 2.49×10^{-12}) suggested an overall increase in protein synthesis demands (Table 9).

The most strongly downregulated pathways included the LIU prostate cancer dn signature (NES = -2.16, FDR = 7.33×10^{-33}), which represents transcriptional programs typically suppressed in prostate tumors, as well as multiple striated muscle contraction and muscle cell development pathways (NES $\approx$ -1.90),



(A) PCA of EA tumor versus adjacent normal prostate tissue



(B) Volcano plot of DEGs in PCa and adjacent normal prostate tissue from EA patients

Fig. 3 Plots comparing PCa and adjacent normal prostate tissue from EA patients. (A) PCA plot of EA tumor (blue) and EA adjacent normal (red) samples showing separation. It shows the absence of batch effect due to technical issues. (B) Volcano plot shows DEGs between PCa and adjacent noncancerous prostate tissue in EA patients, serving as a parallel comparison to the AA patient-specific analysis. The horizontal axis is the $\log_2$ FC between PCa and adjacent noncancerous prostate tissue. The negative $\log_{10}$ of the adjusted p-value (FDR) obtained from DESeq2 is plotted on the vertical axis. Each gene is represented by one point on the graph. The red spots represent significant DEGs ($\log_2$ FC > 1: upregulated; $\log_2$ FC < -1: downregulated). The blue spots represent genes that had FDR < 0.05 but didn't match the $|\log_2 \text{FC}| > 1$ threshold.

consistent with loss of muscle-related gene expression. Notably, downregulation of the HIV Tat to TLR2/4 NF-B signaling pathway (NES = -1.88, FDR = 0.0118) points to reduced inflammatory signaling potential in EA tumors (Table 10).

Table 8 Top 10 downregulated genes in EA tumor versus EA adjacent normal prostate tissue. The ten most downregulated genes in EA prostate tumor tissue relative to matched adjacent normal tissue, ranked by $\log_2$ FC among those with FDR < 0.05.

Genes	Gene Annotation	$\log_2$ FC	FDR
TGM4	transglutaminase 4	-4.02	0.0424
PENK	proenkephalin	-3.18	0.0031
SERPINB11	serpin family B member 11	-2.84	0.00698
LTF	lactotransferrin	-2.61	0.0331
TCAP	titin-cap	-2.6	0.0428
FAM167A	family with sequence similarity 167 member A	-2.37	0.0165
LRP2	LDL receptor related protein 2	-2.17	0.0131
IQSEC3	IQ motif and Sec7 domain ArfGEF 3	-1.82	0.0156
APOBEC3C	apolipoprotein B mRNA editing enzyme catalytic subunit 3C	-1.8	0.0218
PGM5-AS1	PGM5 antisense RNA 1	-1.79	0.0493

Table 9 Top 10 upregulated pathways in EA tumor versus EA adjacent normal prostate tissue. The ten most upregulated pathways in EA prostate tumor tissue relative to matched adjacent normal tissue, ranked by NES among those with FDR < 0.05.

Pathways	NES	FDR
LIU prostate cancer up	3.74	5.49×10^{-37}
GAVISH 3CA malignant metaprogram 9 unfolded protein response	3.1	3.37×10^{-12}
TOMLINS prostate cancer up	3.09	3.55×10^{-13}
MORF UBE2I	3.08	8.38×10^{-28}
MORF NPM1	2.98	1.64×10^{-21}
WP cytoplasmic ribosomal proteins	2.97	1.18×10^{-13}
GNF2 EIF3S6	2.96	5.10×10^{-17}
KEGG medicus reference translation initiation	2.94	2.49×10^{-12}
GAVISH 3CA metaprogram epithelial androgen prostate	2.92	3.24×10^{-10}
REACTOME SRP dependent cotranslational protein targeting to membrane	2.87	2.01×10^{-14}

Discussion

African American (AA) men are disproportionately affected by prostate cancer, experiencing earlier onset, more aggressive disease progression, and higher mortality rates compared to European American (EA) men[21,22]. The racial disparities observed in prostate cancer among AA men are multifactorial, involving not only cultural and socioeconomic factors but also a significant lack of racial diversity in genomic and precision medicine research[23,24]. In this study, we compared gene expression profiles

Table 10 Top 10 downregulated pathways in EA tumor versus EA adjacent normal prostate tissue. The ten most downregulated pathways in EA prostate tumor tissue relative to matched adjacent normal tissue, ranked by NES among those with FDR < 0.05.

Pathways	NES	FDR
LIU prostate cancer dn	-2.16	7.33×10^{-33}
WP striated muscle contraction pathway	-1.98	0.000365
GOBP striated muscle cell development	-1.9	0.000123
REACTOME striated muscle contraction	-1.9	0.00157
GNF2 MYL2	-1.88	0.00233
KEGG medicus pathogen HIV Tat to TLR2 4 NFkB signaling pathway	-1.88	0.0118
MODULE 387	-1.88	0.00164
MODULE 329	-1.88	0.00146
RICKMAN head and neck cancer F	-1.87	0.000871
HP myocardial fibrosis	-1.87	0.0175

between AA and EA prostate tumors, as well as between tumors and matched adjacent normal tissues within each ancestry group. The goal was to identify molecular features associated with ancestry and tumorigenesis in prostate cancer (PCa).

Direct differential expression analysis did not reveal any individual genes surpassing the defined thresholds for statistical and biological significance ($|\log_2 \text{FC}| > 1$, FDR < 0.05). However, gene set enrichment analysis (GSEA) identified coordinated transcriptional changes across multiple biological pathways. AA tumors showed positive enrichment for keratinization[25] and cornified envelope formation, which are processes related to epithelial cell differentiation, as well as for olfactory signaling pathways. Negative enrichment was observed for IgG immunoglobulin complex-related processes and ER protein-folding pathways, including the calnexin/calreticulin cycle. These findings suggest that while individual gene differences may be subtle, ancestry-associated differences in PCa can be captured at the pathway level.

Within AA patients, differential expression analysis identified 574 DEGs (330 upregulated, 244 downregulated) meeting both statistical and fold-change thresholds. The most upregulated genes included AMACR[26] and C1QTNF3-AMACR, both previously implicated in PCa, as well as genes involved in metabolic processes (e.g., GSTA1) and neural/olfactory signaling (e.g., NPY, OR51E1)[27,28]. The most downregulated genes included immunoglobulin components (IGHG1, IGHGP, IGKV3-20, IGKV1-5, IGKC) and WIF1, a known Wnt pathway inhibitor[29].

Pathway analysis revealed enrichment of prostate cancer-associated transcriptional signatures (LIU prostate cancer up) and proliferation markers (MKI67)[30], along with pathways regulating xenobiotic detoxification and glycogen metabolism. These findings align with established roles of metabolic reprogramming and treatment resistance in tumor biology. Downregulated pathways included the TOMLINS prostate cancer dn signature and multiple epithelial-to-mesenchymal transition (EMT)-associated gene programs (GAVISH 3CA malignant metaprogram 12 EMT 1, GAVISH 3CA epithelial EMT like 1) as well as fibroblast/pericyte-like stromal programs and hypoxia-related gene sets.

Within EA patients, 330 DEGs (212 upregulated, 118 downregulated) met both statistical and fold-change thresholds. Upregulated genes included CRISP3, OR51E2, and PCAT14, all previously associated with PCa biology[31–33]. Downregulated genes featured TGM4, PENK, and SERPINB11, among others. Pathway enrichment indicated strong upregulation of multiple prostate cancer-specific signatures (LIU prostate cancer up, TOMLINS prostate cancer up), unfolded protein response, androgen-prostate epithelial programs, and protein synthesis machinery. Downregulated pathways were predominantly related to striated muscle development and contraction, myocardial fibrosis, and inflammatory signaling.

Implications

This study identifies distinct transcriptomic differences between African American and European American prostate tumors, highlighting ancestry-associated variation in pathways linked to keratinization, xenobiotic detoxification, immune complex formation, and stromal-epithelial interactions. The enrichment of keratinization-related pathways in African American tumors aligns with prior reports linking keratin family gene expression to aggressive disease subtypes and poorer androgen deprivation therapy response. Upregulation of xenobiotic detoxification pathways suggests an enhanced drug efflux capacity, potentially contributing to treatment resistance. Conversely, the downregulation of immunoglobulin complex pathways and TLR-NFB signaling in European American tumors indicates suppression of specific inflammatory and immune processes, which may influence tumor-immune interactions. Collectively, these findings underscore the potential biological underpinnings of racial disparities in prostate cancer outcomes and point toward ancestry-informed therapeutic strategies and biomarker development.

Limitations

This study has several limitations that should be considered when interpreting the results. First, the RNA-seq data were obtained from a publicly available dataset (GSE104131) with

a relatively small sample size (n = 28), which may limit the statistical power to detect subtle differences and increase the possibility of false negatives in differential expression analysis. Second, the analysis was restricted to a single cohort and lacked validation in an independent patient population, which is necessary to confirm the reproducibility and generalizability of the findings. Third, although all samples were processed under uniform experimental conditions, unmeasured biological variables such as treatment history, comorbidities, or tumor microenvironment heterogeneity could have influenced gene expression patterns. Fourth, the study design was observational and cross-sectional, precluding causal inferences regarding the functional role of the identified pathways in disease progression or racial disparities. Finally, while pathway-level analyses provided insights into coordinated biological changes, these approaches rely on existing gene set annotations, which may not capture novel or context-specific biology, particularly in underrepresented populations. Future work incorporating larger, racially diverse cohorts, multi-omics integration, and functional validation will be essential to further elucidate the biological mechanisms underlying prostate cancer disparities.

Conclusion

These findings reinforce the value of pathway-level analysis in detecting ancestry-associated molecular differences in PCa. They indicate both shared and distinct biological processes underlying tumorigenesis in AA and EA patients, which have the potential to inform future investigations into ancestry-specific disease mechanisms and the development of tailored therapeutic strategies and biomarkers.

Acknowledgments

Special thanks to Dr. Aguilar, my Cancer-Immunology teacher.

References

1 A. C. Society, *Key statistics for prostate cancer*, 2024, https://www.cancer.org/cancer/types/prostate-cancer/about/key-statistics.html.

2 A. Elbialy, A. Sood, S.-J. Wang, P. Wang, A. Fadiel, A. V. Parwani, S. Huang, G. Shvets, N. Putluri, J. Li and X. Liu, *Cell Insight*, 2025, **4**, 100238.

3 R. Gulati, Y. A. Nyame, J. M. Lange, J. E. Shoag, A. Tsodikov and R. Etzioni, *J. Natl. Cancer Inst. Monogr.*, 2023, **2023**, 212–218.

4 A. J. Vickers, B. Mahal and O. O. Ogunwobi, *J. Clin. Oncol.*, 2023, **41**, 2151–2154.

5 J. Farrell, G. Petrovics, D. G. McLeod and S. Srivastava, *Int. J. Mol. Sci.*, 2013, **14**, 15510–15531.

6 M. Kim, P. Tamukong, G. C. Galvan, Q. Yang, A. D. Hoedt, M. R. Freeman, S. You and S. Freedland, *Genome Med.*, 2024, **16**, 92.

7 B. A. Mahal, T. Gerke, S. Awasthi, H. R. Soule, J. W. Simons, A. Miyahira, S. Halabi, D. George, E. A. Platz, L. Mucci and K. Yamoah, *Eur. Urol. Oncol.*, 2022, **5**, 18–29.

8 P. Creixell, J. Reimand, S. Haider, G. Wu, T. Shibata, M. Vazquez, V. Mustonen, A. Gonzalez-Perez, J. Pearson, C. Sander, B. J. Raphael, D. S. Marks, B. F. F. Ouellette, A. Valencia, G. D. Bader, P. C. Boutros, J. M. Stuart, R. Linding, N. Lopez-Bigas and L. D. Stein, *Nat. Methods*, 2015, **12**, 615–621.

9 J. S. Myers, A. K. von Lersner, C. J. Robbins and Q.-X. A. Sang, *PLoS One*, 2015, **10**, e0145322.

10 E. A. Pudova, G. S. Krasnov, A. A. Kobelyatskaya, M. V. Savvateeva, M. S. Fedorova, V. S. Pavlov, K. M. Nyushko, A. D. Kaprin, B. Y. Alekseev, D. Y. Trofimov, G. T. Sukhikh, A. V. Snezhkina and A. V. Kudryavtseva, *Front. Genet.*, 2021, **11**, 613162.

11 L. Albarracn-Navas, M. Almonte-Becerril, E. Guerrero, J. Rivadeneira, M. Telechea-Fernndez, E. Guzmn, F. Caldern, M. J. Hernndez-Leal, T. Otzen, C. Manterola, G. Duque and . L. Riffo-Campos, *Biomedicines*, 2024, **12**, 2509.

12 W. Rayford, A. T. Beksac, J. Alger, M. Alshalalfa, M. Ahmed, I. Khan, U. G. Falagario, Y. Liu, E. Davicioni, D. E. Spratt, E. M. Schaeffer, F. Y. Feng, B. Mahal, P. L. Nguyen, R. B. Den, M. D. Greenberger, R. Bradley, J. M. Watson, M. Beamer, L. Stamatakis, D. J. Carmen, S. Awasthi, J. Hwang, R. Weil, H. Merisaari, N. Mohamed, L. A. Deane, D. Chakravarty, K. K. Yadav, K. Yamoah, S. S. Nair and A. K. Tewari, *Commun. Biol.*, 2021, **4**, 670.

13 S. Ramakrishnan, E. Cortes-Gomez, S. R. Athans, K. M. Attwood, S. R. Rosario, S. J. Kim, D. E. Mager, E. G. Isenhart, Q. Hu, J. Wang and A. Woloszynska, *Genome Med.*, 2024, **16**, 52.

14 E. L. Teslow, B. Bao, G. Dyson, C. Legendre, C. Mitrea, W. Sakr, J. D. Carpten, I. Powell and A. Bollig-Fischer, *Mol. Oncol.*, 2018, **12**, 1138–1152.

15 National Center for Biotechnology Information, *RNA-seq counts*, 2024, https://www.ncbi.nlm.nih.gov/geo/info/rnaseqcounts.html.

16 M. I. Love, W. Huber and S. Anders, *Genome Biology*, 2014, **15**, 550.

17 B. T. Sherman, M. Hao, J. Qiu, X. Jiao, M. W. Baseler, H. C. Lane, T. Imamichi and W. Chang, *Nucleic Acids Research*, 2022, **50**, W216–W221.

18 S. Xu, E. Hu, Y. Cai, Z. Xie, X. Luo, L. Zhan, W. Tang, Q. Wang, B. Liu, R. Wang, W. Xie, T. Wu, L. Xie and G. Yu, *Nature Protocols*, 2024, **19**, 3292–3320.

19 I. Dolgalev, *msigdbr: MSigDB gene sets for multiple organisms in a tidy data format*, https://CRAN.R-project.org/package=msigdbr, 2025, R package version 24.1.0.

20 G. Yu, *enrichplot: Visualization of functional enrichment result*, https://bioconductor.org/packages/enrichplot, 2025, R package version 1.28.4.

21 J. Farrell, D. Young, Y. Chen, J. Cullen, I. L. Rosner, J. Kagan, S. Srivastava, D. G. McLeod, I. A. Sesterhenn, S. Srivastava and G. Petrovics, *Molecular and Clinical Oncology*, 2014, **2**, 982–986.

22 C. M. Zeigler-Johnson, E. Spangler, M. Jalloh, S. M. Gueye, H. Rennert and T. R. Rebbeck, *Canadian Journal of Urology*, 2008, **15**, 3872–3882.

23 J. W. J. Lillard, K. A. Moses, B. A. Mahal and D. J. George, *Cancer*, 2022, **128**, 3787–3795.

24 W. Liu, L. Zheng, R. Na, L. Wei, J. Sun, J. Gallagher, J. Wei, W. K. Resurreccion, S. Ernst, K. S. Sfanos, W. B. Isaacs and J. Xu, *Molecular Cancer Research*, 2020, **18**, 1815–1824.

25 J. O. Mori, J. White, I. Elhussin, B. M. Duduyemi, B. Karanam, C. Yates and H. Wang, *Frontiers in Oncology*, 2022, **12**, 928357.

26 N. Jiang, S. Zhu, J. Chen, Y. Niu and L. Zhou, *PLOS ONE*, 2013, **8**, e74386.

27 A. Pronin and V. Slepak, *Journal of Biological Chemistry*, 2021, **296**, 100475.

28 T. Henry, K. Wang, J. Pluznick, O. Zhuravel, S. Zozulya and V. Slepak, *Journal of Pharmacology and Experimental Therapeutics*, 2024, **389**, 283.

29 Z. Xia, D. Du, Z. Zhang, Z. Liu, Z. Hu, X. Li, X. Guo and Z. He, *Translational Andrology and Urology*, 2024, **13**, 2601–2616.

30 Z. Song, Q. Zhou, J.-L. Zhang and J. Ouyang, *World Journal of Clinical Cases*, 2024, **12**, 32–41.

31 M. Volpert, L. Furic, J. Hu, A. E. OConnor, R. J. Rebello, S. Keerthikumar, J. Evans, D. J. Merriner, J. Pedersen, G. P. Risbridger, P. McIntyre and M. K. OBryan, *Endocrine-Related Cancer*, 2020, **27**, 415–430.

32 M. T. Thomsen, M. Busk, D. Zhang, C. L. Chiu, H. Zhao, F. J. Garcia-Marques, A. Bermudez, S. Pitteri, M. Borre, J. D. Brooks and J. R. Nyengaard, *BMC Cancer*, 2025, **25**, 535.

33 Y. Yan, J. Liu, Z. Xu, M. Ye and J. Li, *Disease Markers*, 2021, **2021**, 9494619.