

# Article TCGA Expression Analyses of 10 Carcinoma Types Reveal Clinically Significant Racial Differences

Brian Lei <sup>1,2</sup>, Xinyin Jiang <sup>3,4</sup> and Anjana Saxena <sup>2,4,\*</sup>



- Biology Department, Brooklyn College, New York, NY 11210, USA
- <sup>3</sup> Department of Health and Nutrition Sciences, Brooklyn College, New York, NY 11210, USA
- <sup>4</sup> Biology and Biochemistry Programs, CUNY Graduate Center, New York, NY 10016, USA

Correspondence: asaxena@brooklyn.cuny.edu

**Simple Summary:** Racial disparities in cancer incidence and outcome rates are prevalent in the US, with a variety of contributing factors, such as socioeconomic status, differences in lifestyle, environmental exposures, and biological and genetic determinants. The goal of this research was to broadly analyze public data to identify critical differences in molecular signatures and pathway regulation between races. Additionally, to support the clinical translatability of our work, we evaluated the association of differences in gene expression with patients' survival outcomes. Our findings help inform the use of novel biomarkers in clinical settings and the future development of race-adjusted precision therapies.

**Abstract:** Epidemiological studies reveal disparities in cancer incidence and outcome rates between racial groups in the United States. In our study, we investigated molecular differences between racial groups in 10 carcinoma types. We used publicly available data from The Cancer Genome Atlas to identify patterns of differential gene expression in tumor samples obtained from 4112 White, Black/African American, and Asian patients. We identified race-dependent expression of numerous genes whose mRNA transcript levels were significantly correlated with patients' survival. Only a small subset of these genes was differentially expressed in multiple carcinomas, including genes involved in cell cycle progression such as *CCNB1*, *CCNE1*, *CCNE2*, and *FOXM1*. In contrast, most other genes, such as transcriptional factor *ETS1* and apoptotic gene *BAK1*, were differentially expressed and clinically significant only in specific cancer types. Our analyses also revealed race-dependent, cancer-specific regulation of biological pathways. Importantly, homology-directed repair and ERBB4-mediated nuclear signaling were both upregulated in Black samples compared to White samples in four carcinoma types. This large-scale pan-cancer study refines our understanding of the cancer health disparity and can help inform the use of novel biomarkers in clinical settings and the future development of precision therapies.

**Keywords:** The Cancer Genome Atlas (TCGA); carcinoma; racial disparity; differential gene expression; race-based survival; gene-based survival; cancer pathway; DNA repair

## 1. Introduction

Race is a very broad categorizer that groups populations according to common ancestral and phenotypic characteristics. Racial groups are social constructs that exhibit immense biological variation within them. Even so, previous work has identified genetic, epigenetic, and expression-level differences between racial groups that are implicated in a wide range of biological mechanisms [1]. Cancer epidemiologists have long recognized race-based differences in cancer incidence and outcomes, especially in the United States [2]. In addition to genetic variations linked to ancestry, external factors, such as culture, access to healthcare, socioeconomic status, and environmental exposures (e.g., pollution), have



Citation: Lei, B.; Jiang, X.; Saxena, A. TCGA Expression Analyses of 10 Carcinoma Types Reveal Clinically Significant Racial Differences. *Cancers* 2023, *15*, 2695. https://doi.org/ 10.3390/cancers15102695

Academic Editors: Shailesh Singh, Rajesh Singh and Sikander Ailawadhi

Received: 23 February 2023 Revised: 2 May 2023 Accepted: 4 May 2023 Published: 10 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). substantial impact on health disparity. Furthermore, studies suggest that self-reported race aligns closely with genetic ancestry and is a useful proxy for external determinants [2–5].

The debate is ongoing regarding the relative contribution of intrinsic and extrinsic factors in the cancer disparity. Some studies suggest that differences in poor outcomes cannot fully be explained by disparate access to care [6–8], while others suggest that no differences persist in mortality rates in equal-access healthcare environments or after adjustment for non-biological factors [9–14]. Such results are complicated by further stratifications along the lines of sex and age. For example, Lin et al. found no significant survival differences in renal cell carcinoma between White people and Black people after stratifying by age and sex [13]. On the other hand, Andaya et al. found that in colon cancer, survival differences between Black people and White people in an equal-access care system were evident only in patients younger than 50 years old [6].

American cancer registries have shown that five-year cancer survival rates are lower in Black people at each stage of diagnosis across a wide spectrum of cancer types [15]. The current published literature provides further insights into the cancer disparity in individual cancer types. For example, prostate cancer occurs more often and has greater mortality rates in African American men compared to Caucasian Americans, which has been increasingly linked to genetic and molecular alterations in addition to socioeconomic factors [16]. Likewise, incidence rates of aggressive endometrial cancers are significantly higher in non-Hispanic Black women compared to White women, and 5-year relative survival for Black women is significantly less than White people [17]. Islami et al. identified a major racial difference in liver cancer death rates that are linked to differences in risk factor prevalence, ranging from 5.5 per 100,000 in non-Hispanic White people to 11.9 per 100,000 in American Indians/Alaska Natives [18]. A racial examination of the California Cancer Registry revealed that advanced-stage and high-grade bladder cancer are especially prevalent in Black patients, along with significantly poorer 5-year disease-specific survival [19]. Tannenbaum et al. identified improved survival in Asian people compared to White people in Florida for non-small cell lung cancer [20].

Regardless of the relative importance of biological and non-biological factors, racial differences in incidence and survival are linked to many differentially expressed genes (DEGs). For example, Li et al. identified the XKR9 gene, implicated in the exposure of phosphatidylserine during apoptosis, as being differentially expressed between Asian Americans, Caucasian Americans, and African Americans in many cancer types and significantly associated with overall survival. The authors suggest that XKR9 could act as a potential race-dependent target for immunotherapy [21]. However, results from smallerscale studies suggest that DEG patterns vary significantly across cancer types: Grunda et al. identified that the prognostically significant genes AR, BCL2, CCND1, CDKN1A, CDKN1B, CDKN2A, ERBB2, ESR1, GATA3, IGFBP2, IL6ST, KRT19, MUC1, PGR, and SERPINE1 are differentially expressed in non-Hispanic White and African American breast cancer patients [22]. Vazquez et al. used The Cancer Genome Atlas (TCGA) data to discover that differential expression of chemokine receptors, when assessed with race and molecular subtype, could explain racial differences in the tumor microenvironment and response to immunotherapy [23]. Powell et al. highlighted race-dependent expression of genes, such as AKT1, ALOX12, IL8, CXCR4, FASN, and TIMP3, in prostate cancer, suggesting opportunities for targeted therapies [24]. Other prostate cancer research points to aberrant activation of signaling pathways, such as androgen receptor (AR), epidermal growth factor receptor (EGFR), and inflammation, across races as potential causes of racial difference in incidence and mortality [25]. Table 1 depicts a selection of DEGs between samples collected from White and Black patients that encompass a variety of cancer types and are implicated in many critical biological pathways. Additionally, other biological criteria, such as genetic polymorphism, mutational variation, and epigenetic variation, have also been linked to race-based differences in cancer incidence and outcome rates. For example, Asif et al. discovered CpG hypomethylation in Black patients with endometrial cancer compared to White patients indicated worse oncogenic transformation [26].

Cancer Type	Higher Expression in	DEGs	References
	Black	IGF1, IL2RA, IL6	[1 01]
All	White	XKR9, CST1, MTRNR2L1	[1,21]
	Black	CDKN2A, CRYBB2, SRC, SCN1A	
Breast	White	AR, BCL2, CASP8, CCND1, CDKN1A, CDKN1B, ERBB2, ESR1, GATA3, IGFBP2, IL6ST, KRT19, MUC1, PGR, SERPINE1	[1,22,27]
Calamatal	Black	EGFR, TOP2A, CRYBB2, PSPH, IL33	[28.20]
Colorectal	White	IL27, ADAL, ARSE, SMOC1, HESX1, TRNP1	[20,29]
Endometrial	Black	MCM2, PLK1, MCM7, LAMA5, LAMC1	[30]
Encometrial	White	SMC1B, CDC7, CCNE2, BRCA2, ERBB2, FGFR3	[50]
Lung	White	CD274	[28]
Prostate	Black	AKT1, SPINK1, AR, ARA55, GNB3, POLR2L, TBP, EGFR, CRYBB2, IL6, IL8, CXCR4, FASN	[1,24,25]
	White	ALOX12, TIMP3	

**Table 1.** Differentially expressed genes (DEGs) in breast, colorectal, endometrial, lung, and prostate cancers.

In this study, we aimed to analyze pre-existing datasets from The Cancer Genome Atlas (TCGA) with a focus on molecular differences between three races: White, Black, and Asian patients. We described the definitions of these race categories in detail in the Methodology section below. While poorer cancer prognoses and incidence rates have been associated with specific racial groups [15,31], the precise molecular mechanisms contributing to these differences remain the subject of ongoing research. Large transcriptomic and proteomic datasets available from a variety of studies and patient groups are often under-analyzed with respect to the racial component. In this study, we assessed these large datasets through the lens of race to gain insights about the racial dependency of DEGs, biological pathways, and patient survival. These revelations will be pivotal not only in providing cancer-specific sets of diagnostic and prognostic biomarkers but also in designing and developing race-adjusted precision therapies in the future.

# 2. Methodology

## 2.1. Data Information

We used TCGA datasets generated as part of the PanCancer Atlas tumor molecular analysis project, cataloged in Table 2. These datasets are already normalized to facilitate comparative analyses. We accessed and analyzed these data using the open-access, opensource cancer database cBioPortal [32–34]. The selected studies encompassed ten different cancer types, including the PanCancer Atlas datasets of colorectal adenocarcinoma (CORE, n = 338), uterine corpus endometrial carcinoma (UCEC, n = 445), invasive breast carcinoma (BRCA-P, n = 903), kidney renal clear cell carcinoma (KIRC, n = 383), kidney renal papillary cell carcinoma (KIRP, n = 253), liver hepatocellular carcinoma (LIHC, n = 329), lung adenocarcinoma (LUAD, n = 434), stomach adenocarcinoma (STAD, n = 344), and thyroid carcinoma (THCA, n = 370) [35]. In addition to these nine PanCancer Atlas datasets, we also included and analyzed two TCGA datasets profiled in 2015 Cell studies, namely, Invasive Breast Carcinoma (BRCA-C, n = 684) [36] and Prostate Adenocarcinoma (PRAD, n = 313) [37]. These two datasets also included gene methylation data. We chose both BRCA-C and BRCA-P datasets despite their significant overlap, as BRCA-P has nearly twice the number of Black patient samples compared to BRCA-C, evaluating reproducibility in our study. PRAD was the only study that did not include survival information associated with its samples. We assessed race-based survival with the samples that included all the

information on race, mRNA expression, copy number alterations, protein expression and methylation; we filtered these data in the cBioPortal web tool.

**Table 2.** Breakdown of selected cancer datasets and respective sample counts. Gene expression analyses were performed either between White vs. Black or White vs. Asian samples, for a given cancer type.

Dataset	Compary Trung	Sample Count		
Abbreviation		White	Black	Asian
BRCA-C	Breast Invasive Carcinoma (Cell 2015)	594	90	
BRCA-P	Breast Invasive Carcinoma (PanCancer Atlas)	726	177	
CORE	Colorectal Adenocarcinoma	277	61	
KIRC	Kidney Renal Clear Cell Carcinoma	330	53	
KIRP	Kidney Renal Papillary Cell Carcinoma	193	60	
LUAD	Lung Adenocarcinoma	382	52	
PRAD	Prostate Adenocarcinoma	270	43	
UCEC	Uterine Corpus Endometrial Carcinoma	344	101	
LIHC	Liver Hepatocellular Carcinoma	174		155
STAD	Stomach Adenocarcinoma	258		86
THCA	Thyroid Carcinoma	320		50

In the TCGA database, each sample is assigned to only one of five racial categories, although in this study, we considered only three: "White", "Black or African American", and "Asian". In this study, we use the term "Black" samples to refer to samples categorized as "Black or African American". We excluded samples assigned to "American Indian or Alaska Native" and "Native Hawaiian or other Pacific Islander" due to low sample sizes, and samples for which race data were absent or unknown. The racial categories reported by TCGA are based on the definitions used by the US Census Bureau in classifying written responses to the race question [38,39]. Other than these self-identified categories, further information about the ancestry or racial admixture data are not available in TCGA. When discussing the literature, we have retained the racial designation as stated in corresponding publications.

#### 2.2. Race-Based Kaplan–Meier Survival Analysis

We performed race-based survival analysis for each cancer type in the cBioPortal web tool using race as the comparison factor. For BRCA-C, BRCA-P, CORE, KIRC, KIRP, LUAD, and UCEC, we compared samples labeled as "White" against "Black". For LIHC, STAD, and THCA, samples labeled "White" were compared against samples labeled "Asian". In addition to these baseline racial analyses, we performed separate analyses after stratifying the racial groups by sex (e.g., White males vs. Black males, White females vs. Black females). Due to unequal levels of racial representation in the different cancer datasets, the racial groups that we selected for analyses varied. Survival was evaluated using four different survival metrics: disease-free survival, progression-free survival, disease-specific survival, and overall survival. We used log rank tests performed by cBioPortal to assess statistical significance; a *p*-value  $\leq 0.05$  was considered statistically significant.

#### 2.3. Gene-Based Kaplan–Meier Survival Analysis

To determine the most clinically relevant genes for each cancer type, we used the Kaplan–Meier (KM) Plotter web tool [40]. The KM plotter database integrates expression data and clinical outcomes from the Gene Expression Omnibus (GEO), European Genomephenome Archive (EGA), and TCGA [41]. We evaluated the prognostic value of the mRNA expression levels of a compiled set of 155 target oncogenes and tumor suppressor genes (see Supplemental S1) for all cancer types listed in Table 2, except for PRAD, for which these data were not available. For each gene, samples were divided into two groups of low and high mRNA expression levels, based on the median expression, and the overall patients' survival rates were compared. *p*-values and hazard ratios with 95% confidence intervals were calculated. To reduce the false discovery rate (FDR), we used  $p \le 0.01$  as a threshold of statistical significance.

## 2.4. mRNA, Protein, and Methylation Analysis

In the cBioPortal web tool, we performed group comparison analyses with race as the comparison factor. For BRCA-C, BRCA-P, CORE, KIRC, KIRP, LUAD, PRAD, and UCEC, we compared "White" with "Black" samples, as reported. For LIHC, STAD, and THCA, "White" was compared against "Asian" samples. The racial groups selected for analysis varied due to unequal levels of racial representation in the different cancer datasets. For each study, results from mRNA and protein expression analyses (and DNA methylation analyses results for BRCA-C and PRAD) were downloaded as tab-separated values (TSV) files. We wrote and used a pandas-based Python (version 3.7.7) program to filter the data for the 155 target genes. Statistically significant differences in mRNA and protein expression between the two racial cohorts, as determined by the Benjamini–Hochberg procedure performed by the cBioPortal web tool, were identified.

#### 2.5. Reactome Pathway Analysis

To perform integrative pathway analysis, we downloaded the PanCancer Atlas Reversephase protein arrays (RPPA) and clinical data from the original publication [42,43]. These data were not filtered for the 155 target genes for the purposes of thorough pathway analysis. We created a pandas-based Python program (https://github.com/brian-lei/cancerracial-disparity, accessed on 18 January 2021) to format and export the data to Reactome, an open-source, open-access peer-reviewed pathway database [44,45]. We grouped samples by cancer type and by race, deleted duplicate samples, and removed data for a few selected proteins, namely, ADAR1, alpha-catenin, TTF-1, caspase-3, caspase-9, PARP1, and JAB1, with missing corresponding values. Post-translationally modified versions of proteins were also excluded from the analysis. We manually annotated samples according to their corresponding race in the Reactome web tool and performed a Pathway Analysis with Down-weighting of Overlapping Genes (PADOG) microarray analysis with race as the comparison factor.

#### 2.6. Statistical Analysis as Presented in Results

Genes that received *p*-values less than or equal to 0.01 in gene-based KM survival analyses were considered "clinically significant". The survival relationship, as evaluated by KM analysis, is denoted in subsequent tables with a "+" for positive and a "-" for negative correlations between specific mRNA gene expression and overall survival. The associated log rank *p*-value is denoted in parentheses.

Genes that received q-values less than or equal to 0.05 in cBioPortal Benjamini-Hochberg comparison analysis were considered "differentially expressed" for mRNA expression, protein expression, or both. The race denoted in tables under "mRNA" is the racial group with *higher* mRNA expression of the gene, followed by the associated q-value in parentheses. The race denoted in tables under "Protein" is the race with *higher* protein expression and the associated q-value. "NS" represents non-significant values. Empty cells represent missing data from the original TCGA database.

Each table in the results section presents gene expression information for a particular cancer type. In all cases, we only included genes that were both clinically significant and differentially expressed between the denoted races. We excluded genes for which neither mRNA nor protein expression was significantly different.

In the Reactome Pathway Analysis subsection, differential regulation of pathways was considered significant with an adjusted *p*-value  $\leq 0.05$  and denoted in parentheses in each table header.

#### 6 of 19

# 3. Results

The analyses are organized alphabetically by cancer type. Each section reports the results of cBioPortal differential expression analysis. All sections, except for PRAD, include cBioPortal race-based survival and KM gene-based survival analysis results. Gene methylation analyses were performed for only two studies for which data were available, BRCA-C and PRAD. Gene expression analysis was performed on only one dataset for each section/cancer type, except for Breast Invasive Carcinoma, which includes results from both BRCA-C and BRCA-P with significant overlap in samples. However, it is important to emphasize that even when BRCA-P included almost two-fold "Black" samples and an additional 132 "White" samples than BRCA-C, the overall analyses yielded similar results. Complete results and statistical significance on all 155 target genes are included in Supplemental S2–S4. The original Reactome report with all queried pathways is included in Supplemental S5.

## 3.1. Breast Invasive Carcinoma (BRCA)

## 3.1.1. Race-Based Survival

In BRCA-C and BRCA-P, no significant differences were identified between White patients and Black patients for any of the survival metrics.

## 3.1.2. Gene-Based Survival

High BCL3, FHIT, and IGF1R transcript levels were correlated with increased overall patient survival, while high CCNE1 and TGFBR1 transcript levels were correlated with decreased overall patient survival.

## 3.1.3. Differential Expression

BRCA-C and BRCA-P data revealed a relatively high number of differentially expressed genes compared to the other studies. Most notably, in BRCA-C, CCNE1 mRNA and protein levels were significantly increased in Black samples compared to White samples. However, in BRCA-P, only the mRNA levels were significant, while protein levels remained non-significant (Table 3). Similarly, genes, such as APC, AXL, CDK4, CXCR1, ERCC4, IGF1, IGF1R, KRAS, MYB, PDGFD, RASSF8, and TGFBR1, had similar patterns of differential mRNA expression in both studies but lacked corresponding protein data. There were 87 other genes that exhibited differential expression across races; however, these were clinically non-significant, suggesting overall patient survival was not impacted. For example, BAK1, CCNB1, NOTCH1, RAD51, and STMN1 mRNA, and protein levels were enriched in Black samples in both studies, while KDR, MAPK9, and RAD50 mRNA, and protein were enriched in White samples in both studies.

**Table 3.** BRCA-C and BRCA-P, breast invasive carcinoma. Of all 155 target genes, 5 genes were identified by KM analysis as clinically significant ( $p \le 0.01$ ). Of these, the four genes shown here were differentially expressed ( $q \le 0.05$ ) between White vs. Black samples, in at least one of the two datasets.

Gene	Survival Relationship	BRCA-C mRNA	BRCA-C Protein	BRCA-P mRNA	BRCA-P Protein
	( <i>p</i> -value in parentheses)		(q-value in p	arentheses)	
BCL3	+(0.0063)	NS (0.127)		Black (0.002278)	
IGF1R	+(0.0026)	White $(1.32 \times 10^{-4})$		White $(3.27 \times 10^{-6})$	
CCNE1	-(0.0093)	Black ( $1.082 \times 10^{-5}$ )	Black (0.0087)	Black ( $2.22 \times 10^{-10}$ )	NS (0.577)
TGFBR1	-(0.0086)	White (0.0348)		White (3.081 $ imes 10^{-5}$ )	

## 3.1.4. Methylation

In BRCA-C, CCNE1 methylation was significantly increased in White samples compared to Black samples. Thirty-eight other target genes exhibited differential methylation across racial groups, including tumor-suppressor genes, BRCA2 (higher methylation in White samples), and PTEN (higher methylation in Black samples).

## 3.2. Colorectal Adenocarcinoma (CORE)

# 3.2.1. Race-Based Survival

In CORE, Black patients suffered from significantly worse progression-free survival (p = 0.0476) than White patients, while differences in disease-free (p = 0.443), disease-specific (p = 0.0513), and overall survival (p = 0.961) were not significant. While survival of Black vs. and White females did not differ significantly in any metric, Black males had significantly worse progression-free (p = 0.004526) and disease-specific survival (p = 0.0362) than White males.

## 3.2.2. Gene-Based Survival

Out of all 155 target genes, only higher BIRC2 and FGF16 transcript levels were significantly associated with increased and decreased overall survival, respectively, in rectal adenocarcinoma. Expression of the remaining 153 genes did not impact overall survival and hence, were clinically non-significant.

# 3.2.3. Differential Expression

While BIRC2 mRNA expression was not significantly different between White samples and Black samples, BIRC2 protein expression was significantly increased in Black samples (Table 4). Unfortunately, neither mRNA nor protein data were available for FGF16 in the original TCGA data, despite being clinically significant.

**Table 4.** CORE, colorectal adenocarcinoma. Of all 155 target genes, 2 (BIRC2 and FGF16) were identified by KM analysis as clinically significant ( $p \le 0.01$ ). Of these, only BIRC2 was differentially expressed ( $q \le 0.05$ ) between White vs. Black samples.

Gene	Survival Relationship	mRNA	Protein
	( <i>p</i> -value in parentheses)	(q-value in	parentheses)
BIRC2	+(0.0017)	NS (0.605)	Black (0.004757)

From the other 153 clinically non-significant genes, two genes, DIABLO and FHIT, differed in their expressions between Black and White samples. DIABLO mRNA expression was higher in White samples (protein levels were not significantly different), while FHIT expression was higher in Black samples (protein data not available). A total of 24 genes exhibited differential protein expression: BAK1, BIRC2, FASN, FOXM1, IRS1, MSH2, and STMN1 protein levels were significantly increased in Black samples, while CCNE2, CDH2, CDKN1A, CHEK1, DIRAS3, ERCC1, ETS1, GAPDH, HSPA1A, MAPK9, MYC, NF2, PRDX1, TGM2, TIGAR, VHL, and XRCC1 protein levels were significantly increased in White samples. The remaining clinically non-significant genes either had no data in TCGA or transcript levels did not differ between the examined races.

# 3.3. Kidney Renal Clear Cell Carcinoma (KIRC)

## 3.3.1. Race-Based Survival

No significant differences were identified between White patients and Black patients, even after stratifying by sex, for any of the survival metrics.

# 3.3.2. Gene-Based Survival

High APC, BCL2, CCND1, CDKN1B, CTNNB1, ERBB2, ERBB3, ERCC4, ERG, ETS1, FGF1, FGF12, FLT1, GAB2, IGF1R, IL6R, KDR, KRAS, MAPK1, MAPK9, MSH2, MYCN,

NRAS, PDGFC, PDGFD, PTCH1, PTEN, RAD50, RASSF8, RB1, SCFD2, SMAD4, TAL2, TGFA, TGFBR2, and TGFBR3 transcript levels were correlated with increased overall patient survival.

High AXL, BAK1, BCL3, CCNB1, CCNE1, CCNE2, CRP, ERCC1, ETV4, FASN, FGF5, FGF8, FGF17, FGF21, FGF23, FOXM1, GAPDH, GNAS, GRP, IL6, NKX2-5, NKX3-1, NTRK1, SHBG, and SRC transcript levels were correlated with decreased overall patient survival.

#### 3.3.3. Differential Expression

AXL, FGF5, KRAS, NKX2-5, RB1, and TAL2 transcript levels were significantly higher in White samples compared to Black samples, although protein data were absent from the original TCGA data. NRAS transcript levels were also increased in White samples, but protein expression was not significantly different across races. ETV4 and FGF17 transcript levels were higher in Black samples, although protein data were absent. ERBB2 transcript levels were higher in Black samples, but protein did not differ significantly across races. ETS1, MAPK1, MAPK9, and RAD50 mRNA, and protein levels were significantly increased in White samples. BCL2, CCNE2, ERCC1, and GAPDH protein levels were also increased in White samples, although transcript levels were not significantly different across races. Interestingly, CCNB1, MSH2, FOXM1, and RAD51 protein levels were increased in Black samples, while transcript levels were increased in White samples (Table 5).

**Table 5.** KIRC, kidney renal clear cell carcinoma. Of all 155 target genes, 61 genes were clinically significant in KM analysis ( $p \le 0.01$ ); 21 genes were differentially expressed ( $q \le 0.05$ ) between White vs. Black samples, as depicted below.

Gene	Survival Relationship	mRNA	Protein
	( <i>p</i> -value in parentheses)	(q-value in p	parentheses)
BCL2	$+(5.9 \times 10^{-6})$	NS (0.244)	White (0.008112)
ERBB2	$+(1.7 \times 10^{-5})$	Black (0.007076)	NS (0.426)
ETS1	+(0.0002)	White (0.0322)	White (0.0285)
KRAS	+(0.0005)	White (0.0111)	
MAPK1	$+(6.4 \times 10^{-6})$	White $(5.349 \times 10^{-6})$	White (0.008112)
MAPK9	+(0.0003)	White (0.0002776)	White (0.0307)
MSH2	+(0.0011)	White (0.0353)	Black (0.0169)
NRAS	+(0.0073)	White $(4.941 \times 10^{-5})$	NS (0.878)
RAD50	$+(1.4 \times 10^{-6})$	White (0.004998)	White $(3.4  imes 10^{-7})$
RB1	+(0.0001)	White (0.005863)	
TAL2	$+(1.1 \times 10^{-8})$	White (0.0123)	
AXL	-(0.0083)	White (0.0113)	
CCNB1	-(0.0074)	White (0.0286)	Black (0.007517)
CCNE2	-(0.0006)	NS (0.1)	White (0.0259)
ERCC1	-(0.001)	NS (0.785)	White (0.000253)
ETV4	-(0.0062)	Black (0.0263)	
FGF5	$-(6.8  imes 10^{-5})$	White (0.004835)	
FGF17	-(0.0028)	Black (0.0148)	
FOXM1	$-(1.8 imes 10^{-5})$	White (0.006838)	Black ( $6.343 \times 10^{-5}$ )
GAPDH	-(0.0039)	NS (0.174)	White $(5.163 \times 10^{-6})$
NKX2-5	$-(1.5  imes 10^{-7})$	White (0.0372)	

Other 37 genes exhibited differential expression across races but were clinically nonsignificant without any impact on overall survival. Interestingly, BAP1 transcript levels were increased in Black samples, while protein levels were higher in White samples. Several other genes had differential protein expression but not differential mRNA expression, namely, BECN1, BIRC2, CDKN1A, DIABLO, DIRAS3, IRS1, NF2, PRDX1, RAF1, STMN1, TGM2, TP53, VHL, and XRCC1.

## 3.4. Kidney Renal Papillary Cell Carcinoma (KIRP)

## 3.4.1. Race-Based Survival

No significant differences were identified between White patients and Black patients, even after stratifying by sex, for any of the survival metrics.

## 3.4.2. Gene-Based Survival

High AXL and SMAD4 transcript levels were correlated with increased overall patient survival, while high BRCA2, CCNE1, CCNE2, FASN, FGF5, FGF7, FGF8, FGF11, FGF18, FOXM1, GRP, IGF2, and PDGFRB transcript levels were correlated with decreased overall patient survival.

## 3.4.3. Differential Expression

For all 155 target genes, including the genes identified by KM analysis as clinically significant for kidney renal papillary cell carcinoma, mRNA expression levels were not significantly different across races. Differences in protein levels were either not significant (e.g., BRCA2) or indeterminable due to a lack of associated protein data (e.g., AXL).

# 3.5. Liver Hepatocellular Carcinoma (LIHC)

## 3.5.1. Race-Based Survival

No significant differences were identified between White patients and Asian patients, even after stratifying by sex, for any of the survival metrics.

#### 3.5.2. Gene-Based Survival

High BAK1, CCNB1, CCNE2, CDK4, CDKN2A, CHEK1, DIABLO, ETV1, ETV4, FOXM1, GAPDH, HRAS, MEN1, MSH2, NKX2-5, NRAS, PRDX1, RAD51, RASSF7, STMN1, TIGAR, and VHL transcript levels were all correlated with decreased overall patient survival. Interestingly, none of the target genes were correlated with improved survival.

#### 3.5.3. Differential Expression

In LIHC, protein data for the genes identified by KM analysis as clinically significant were either absent or displayed non-significant differences across races, except for BAK1, whose protein levels were significantly increased in White samples compared to Asian samples. BAK1 transcript levels, however, were not significantly different across races. CCNB1, CCNE2, CHEK1, FOXM1, HRAS, MSH2, RAD51, and STMN1 transcript levels were significantly higher in Asian samples, while TIGAR transcripts were increased in White samples (Table 6).

**Table 6.** LIHC, liver hepatocellular carcinoma. Of all 155 target genes, 22 were identified by KM analysis as clinically significant ( $p \le 0.01$ ). Of these, the following 10 were differentially expressed ( $q \le 0.05$ ) between White vs. Asian samples.

Gene	Survival Relationship	mRNA	Protein
	( <i>p</i> -value in parentheses)	(q-value in p	arentheses)
BAK1	-(0.008)	NS (0.366)	White (0.0446)
CCNB1	-(0.0008)	Asian (0.004704)	NS (0.176)
CCNE2	-(0.0011)	Asian (0.002332)	NS (0.22)
CHEK1	-(0.0002)	Asian (0.0256)	NS (0.22)
FOXM1	-(0.0015)	Asian (0.0238)	NS (0.374)
HRAS	-(0.0012)	Asian (0.0375)	
MSH2	-(0.0004)	Asian (0.00474)	NS (0.152)
RAD51	-(0.007)	Asian (0.003367)	NS (0.302)
STMN1	-(0.0004)	Asian (0.01)	NS (0.731)
TIGAR	-(0.0092)	White (0.002248)	NS (0.857)

Other 31 genes exhibited differential expression across races but were clinically nonsignificant without any impact on overall survival; all these genes differed in mRNA expression only. Specifically, AKT1, ALK, BCL2, CDKN1A, CRP, EGFR, ERCC4, ETS1, FGF7, FLT1, GDNF, GRP, HBEGF, IGF1, IL6, IL6R, KIT, MDM2, NKX3-1, PDGFC, PDGFD, PTCH1, PTCH2, RASSF6, RET, TGFBR1, TGFBR2, and TGFBRAP1 mRNA levels were enriched in White samples, while CCNE1, GNAS, and MAPK9 mRNA levels were enriched in Asian samples. Protein expression data for these genes were either not significant or unavailable from the original TCGA database.

## 3.6. Lung Adenocarcinoma (LUAD)

# 3.6.1. Race-Based Survival

No significant differences were identified between White patients and Black patients, even after stratifying by sex, for any of the survival metrics.

#### 3.6.2. Gene-Based Survival

High BCL2, FGF18, GAB2, KIT, NTRK1, PTCH1, RASSF2, and TMPRSS2 transcript levels were correlated with increased overall patients' survival, while high CCNB1, CCNE2, FGF5, FGF19, FOXM1, GAPDH, IRS1, TGFA, and VEGFC transcript levels were correlated with decreased overall survival.

#### 3.6.3. Differential Expression

The mRNA transcript levels were not significantly different between White samples and Black samples for any of the genes identified by KM analysis as clinically significant for lung adenocarcinoma. Notably, IRS1 protein levels were significantly increased in Black samples (Table 7).

**Table 7.** LUAD, lung adenocarcinoma. Of all 155 target genes, 17 were identified by KM analysis as clinically significant ( $p \le 0.01$ ). Of these, only IRS1 was differentially expressed ( $q \le 0.05$ ) between White vs. Black samples.

Gene	Survival Relationship	mRNA	Protein
	( <i>p</i> -value in parentheses)	(q-value in	parentheses)
IRS1	-(0.0099)	NS (0.669)	Black (0.0486)

Seven clinically non-significant genes exhibited differential expression across races. ERCC1 and TIGAR protein expression was significantly increased in White samples, while NOTCH1 and RAD51 protein levels were increased in Black samples. mRNA levels for these four genes were not different between White samples and Black samples. MAPK1, MAPK9, and TGFBR1 transcript levels were enriched in White samples compared to Black samples; protein data were not significant for MAPK1/MAPK9 and absent for TGFBR1.

#### 3.7. Prostate Adenocarcinoma (PRAD)

## 3.7.1. Differential Expression

The mRNA transcript levels of CTNNB1, FGF19, NF2, PTCH1, and RB1 were elevated in White samples, while transcript levels of FGFR4 and RASSF7 were elevated in Black samples. Protein data for these genes were either not significant or missing. SMAD4 protein levels were increased in White samples, although mRNA levels were not significantly different between Black samples and White samples.

#### 3.7.2. Methylation

No significant methylation differences were identified in PRAD for any of the 155 examined genes.

#### 3.8. Stomach Adenocarcinoma (STAD)

# 3.8.1. Race-Based Survival

In STAD, Asian patients suffered from significantly worse disease-free survival (p = 0.0145) than White patients, while differences in progression-free (p = 0.729), disease-specific (p = 0.901), and overall survival (p = 0.304) were not significant. White females suffered from significantly worse overall survival compared to Asian females (p = 0.0453), while Asian males suffered from significantly worse disease-free survival compared to White males (p = 0.009962).

## 3.8.2. Gene-Based Survival

High MSH2 transcript levels were correlated with increased overall patients' survival, while high EGF, GRP, PDGFD, PDGFRB, TGFBR1, and VEGFC transcript levels were correlated with decreased overall survival.

## 3.8.3. Differential Expression

The mRNA transcript levels were not significantly different between White samples and Asian samples for any of the genes identified by KM analysis as clinically significant for stomach adenocarcinoma. The GRP transcript data were absent from the database. Protein data were absent for all the genes that were identified as clinically significant, except for MSH2 protein, for which the data was not significantly different between White samples and Asian samples.

Two genes, PRDX1 and VHL, exhibited differential expression across races but were not clinically significant and did not impact overall survival. While PRDX1 transcript levels were not significantly different, protein levels were increased in White samples. VHL mRNA expression was increased in Asian samples, while protein levels were not different.

## 3.9. Thyroid Carcinoma (THCA)

# 3.9.1. Race-Based Survival

No significant differences were identified between White patients and Asian patients, even after stratifying by sex, for any of the survival metrics.

#### 3.9.2. Gene-Based Survival

High EGF and FGF5 transcript levels were correlated with decreased overall patients' survival.

## 3.9.3. Differential Expression

Of the 155 target genes, none exhibited differential protein expression between White samples and Asian samples. EGF and FGF5 mRNA expressions did not differ significantly between White samples and Asian samples, and protein data for both genes were absent from THCA. The sole significant result was that TIGAR, a clinically non-significant gene, had enriched transcript levels in White samples.

# 3.10. Uterine Corpus Endometrial Carcinoma (UCEC)

## 3.10.1. Race-Based Survival

In UCEC, Black patients suffered from significantly worse disease-specific survival (p = 0.0392), while differences in progression-free (p = 0.263), disease-free (p = 0.272), and overall survival (p = 0.331) were not significant.

#### 3.10.2. Gene-Based Survival

High FGF1, IGF1, and MDM2 transcript levels were correlated with increased overall patients' survival, while high ALK, BAK1, CCNE1, CDKN2A, FGF11, FGF12, MSH2, and RET transcript levels were correlated with decreased overall survival.

## 3.10.3. Differential Expression

MDM2 transcript levels were significantly higher in White samples, while FGF12 and CDKN2A transcripts were higher in Black samples; protein data were absent for these genes in the original TCGA data. CCNE1 mRNA and protein levels were both increased in Black samples compared to White samples. Additionally, while MSH2 mRNA levels were not significantly different across races, protein levels were elevated in Black samples (Table 8).

**Table 8.** UCEC, uterine corpus endometrial carcinoma. Of all 155 target genes, 11 were identified by KM analysis as clinically significant ( $p \le 0.01$ ). Of these, the following five were differentially expressed ( $q \le 0.05$ ) between White vs. Black samples.

Gene	Survival Relationship	mRNA	Protein
	( <i>p</i> -value in parentheses)	(q-value in p	parentheses)
MDM2	+(0.0013)	White (0.016)	
CCNE1	-(0.0001)	Black (0.0298)	Black (0.0146)
CDKN2A	$-(9.7  imes 10^{-5})$	Black (0.0104)	
FGF12	-(0.0042)	Black (0.00292)	
MSH2	-(0.0081)	NS (0.106)	Black (0.0277)

Other 30 genes exhibited differential expression across races but were clinically nonsignificant and did not affect overall survival. Specifically, ABL1, ETV1, FGF2, FGF5, GL11, PDGFRA, PDGFRB, RASSF6, and TP53 transcript levels were elevated in White samples, while CASP12, CDH2, MYCL, and VEGFD transcript levels were elevated in Black samples. Protein data for these genes were either not significant or not available.

Additionally, BECN1, BIRC2, ERBB3, FASN, FOXM1, IRS1, NRAS, RAD51, and STMN1 protein levels were significantly increased in Black samples while CDKN1A, CTNNB1, ERCC1, ETS1, GAPDH, RAD50, and TGM2 protein levels were increased in White samples. However, mRNA levels were not significantly different across races for any of these genes. Interestingly, while NOTCH1 mRNA expression was significantly higher in White samples, protein levels were enriched in Black samples.

## 3.11. Reactome Pathway Analysis

We queried 987 pathways, of which 189 pathways were differentially regulated between races, at least in 1 cancer type. We observed that for 74 out of these 189 pathways, for each pathway, data for at least 3 proteins were available (see Supplemental S5). Importantly, homology-directed repair (HDR) was differentially regulated between Black and White samples in BRCA, CORE, KIRC, LUAD, and UCEC due to the differential protein expression of ten genes, namely, ATM, BRCA2, CHEK1, ERCC1, MRE11, PCNA, RAD50, RAD51, TP53BP1, and XRCC1 (Table 9). In CORE, HDR was overall downregulated in Black samples compared to White samples. In contrast, in four other cancer types, HDR was upregulated in Black samples indicating the existence of other cancer-specific differences. **Table 9.** Differential expression of homology-directed repair (HDR) genes in CORE, UCEC, BRCA, KIRC, and LUAD. A yellow or blue highlighted header indicates overall HDR pathway upregulation in White or Black samples, respectively; the associated *p*-value is denoted in parentheses with the study abbreviation. Two yellow or blue dots indicate significantly higher protein expression in the corresponding race. A single-colored dot represents a non-significantly higher expression.

Protein	CORE (0.027)	UCEC (0.015)	BRCA (0.034)	KIRC (0.011)	LUAD (0.01)
ATM	••	•	••	•	••
BRCA2	•	••	•	••	•
CHEK1	••	•	••	•	•
ERCC1	••	••	••	••	•
MRE11	•	••	••	••	•
PCNA	•	•	••	•	•
RAD50	•	••	••	••	•
RAD51	••	••	••	••	••
TP53BP1	•	•	•	•	•
XRCC1	••	•	•	••	•

In CORE, nucleotide excision repair was significantly downregulated in Black samples compared to White samples (adjusted p = 0.019) due to decreased XRCC1 and ERCC1 protein expression.

In LIHC, DNA mismatch repair pathways were upregulated in Asian samples compared to White samples due to elevated MSH6 protein levels (adjusted p = 0.024). Additionally, the G2/M DNA damage checkpoint pathway was upregulated in Asian samples in LIHC (adjusted p = 0.018). In Asian samples, ATM levels were higher while CDK1, TP53, YWHAB, and YWHAE levels were lower compared to White samples.

In BRCA, CORE, KIRC, and LUAD, ERBB4-mediated nuclear signaling was upregulated in Black samples compared to White samples due to the differential protein expression of six genes, namely, ESR1, PGR, SRC, STAT5A, STMN1, and YAP1 (Table 10).

**Table 10.** Differential expression in CORE, BRCA, KIRC, and LUAD of ERBB4-mediated nuclear signaling proteins. A blue highlighted header indicates overall signaling pathway upregulation in Black samples; the associated *p*-value is denoted in parentheses with the study abbreviation. Two yellow or blue dots indicate significantly higher protein expression in the corresponding race. A single-colored dot represents a non-significantly higher expression.

Protein	CORE (0.005)	BRCA (0.009)	KIRC (0.013)	LUAD (0.018)
ESR1	••	•	••	•
PGR	••	••	••	•
SRC	•	••	•	•
STAT5A	•	••	•	•
STMN1	••	••	••	••
YAP1	•	••	•	•

In LUAD, TP53-mediated regulation of apoptotic gene transcription was increased in Black samples compared to White samples (adjusted p = 0.002): Black samples exhibited higher protein levels of ATM and BID, while TP53 and BAX expression was not significantly different between Black samples and White samples. In THCA, TP53 regulation of apoptotic gene transcription was increased in Asian samples compared to White samples (adjusted p = 0.017): Asian samples exhibited higher levels of BID protein. ATM, TP53, and BAX expression was also higher in Asian samples, although these differences were not significant.

In STAD, FOXO-mediated transcription of cell cycle genes was downregulated in Asian samples compared to White samples (adjusted p = 0.014).

# 4. Discussion

In this study, we dissected large transcriptomic, proteomic, and survival data available from The Cancer Genome Atlas (TCGA) to identify molecular differences between White, Black, and Asian samples that correlate with cancer health disparities. It is important to re-emphasize that TCGA data do not include detailed ancestry information, and hence, we have used the available racial categories instead. Although these racial categories are inherently subjective, others have emphasized the usefulness of race as it mostly aligns with genetic ancestry and can be a proxy to explain total variance that cannot be fully explained otherwise [3–5]. Thus, the findings in this study must not be confused with any intrinsic biological differences between populations. It is also worth mentioning that this study does not suggest any causal relationships; the underlying "cause-effect" relationships of much of the reported differential gene/protein expressions and survival disparities are unknown and must be addressed in further experimental and clinical setups. Nevertheless, these findings help illustrate the complex nature of the tumor molecular differences that exist between racial groups, particularly between White patients and Black patients. Importantly, there were insufficient data for extensive comparisons between White patients and Asian patients, highlighting the need for more proportional racial representation. However, our methods provide a valuable pipeline to analyze the race dependency of DEGs, proteins, and pathways and their survival association. We investigated the expression of 155 cancerrelevant genes and their negative or positive associations with patient survival to pinpoint race-dependent molecular factors in multiple cancer types.

We observed that most of the targeted genes were both differentially expressed and clinically significant in no more than one cancer type, except for a few that were significant in at least two cancer types; this is suggestive of cancer-specific variations in gene expression. For example, the cell cycle progression gene *CCNE1* was both differentially expressed and implicated in patient survival in breast and endometrial cancer. Additionally, the DNA repair gene MSH2 exhibited differential expression and clinical significance in kidney clear cell, liver, and endometrial cancer. *CCNB1*, *CCNE2*, and *FOXM1* were three other genes that were differentially expressed in more than one carcinoma with survival implications, specifically in kidney clear cell and liver carcinoma. The clinical implications of multiple members of the cyclin family of proteins and the cell cycle regulator *FOXM1* across multiple cancer types suggest that differences in cell cycle regulation are associated with survival disparities. Previous cancer-specific studies have also proposed this connection in breast [46] and endometrial carcinoma [30].

It is important to note that although we focused on genes that had a significant correlation with overall survival, we also identified many genes that had no impact on overall clinical survival yet were differentially expressed between the examined races. It may be useful to perform further research on this subset of genes and their relevant pathways to uncover any unapparent relationships, if any, between expression and patients' survival.

Additionally, the source datasets varied in their abundance of differentially expressed biomarkers. While several pathways were identified in multiple cancer types as being regulated in a race-dependent manner with direct correlation to altered molecular patterns, in stomach adenocarcinoma, none of the clinically significant genes were differentially expressed between White and Asian samples.

A striking group of pathways that were significantly associated with race was DNA repair mechanisms. Most notably, racial differences between Black and White samples in the expression of genes pertaining to homology-directed repair, including *RAD50*, *RAD51*, *MRE11*, *ERCC1*, and *BRCA2*, were observed in five cancer types. Moreover, genes involved in nucleotide excision repair and DNA mismatch repair were differentially expressed in colorectal and liver cancer, respectively. Nuclear signaling mediated by ERBB4 was upregulated in Black samples in four cancer types, an important finding considering that overexpression of the kinase is associated with cancer development [47]. Table 11 lists the major findings of this study, indicating differentially regulated major pathways.

Pathway	Carcinoma Type (s)	<b>Race with Higher Expression</b>
DNA mismatch repair	Liver hepatocellular	A (vs. W)
ERBB4-mediated nuclear signaling	Breast, colorectal, kidney renal clear cell, lung	B (vs. W)
FOXO-mediated transcription of cell cycle genes	Stomach	W (vs. A)
G2/M DNA damage checkpoint	Liver hepatocellular	A (vs. W)
	Breast, endometrial, kidney renal clear cell, lung	B (vs. W)
Homology-directed repair	Colorectal	W (vs. B)
Nucleotide excision repair	Colorectal	W (vs. B)
TP53-mediated regulation of apoptotic	Lung	B (vs. W)
gene transcription	Thyroid	A (vs. W)

**Table 11.** Differentially expressed pathways and their respective carcinoma types. A: Asian, B: Black, W: White.

Protein data for many genes such as *CDKN2A* were absent from the original TCGA datasets; even so, our study corroborates the reported enrichment of *CDKN2A* transcript levels in Black patients with endometrial cancer [30]. We can also attempt to reconcile the mRNA differential expression analysis results with the existing literature on protein imbalances across racial groups. In UCEC, *MDM2* oncogene expression was higher in White samples, a result consistent with the reported elevation of MDM2 protein expression in Caucasian Americans when compared to African Americans in prostate cancer [48]. However, elevated transcript levels of *FGF12* in Black samples in UCEC conflict with the reported enrichment of fibroblast growth factor 12 in White patients [30]. We found many such contradictory results where transcript level differences did not accurately predict protein level differences (for example, MSH2, CCNB1, and FOXM1 in KIRC). Such inconsistencies between mRNA and protein differential expression may be attributable to post-transcriptional regulations of gene expression with racial impact; this is a promising topic for future research.

The autocrine-paracrine growth factor, insulin-like growth factor 1 (IGF-1), has been reported to be a promoter of cancer that inhibits the sex hormone-binding globulin (SHBG) and exists in elevated levels in African American individuals [1]. Interestingly, although they were not identified as clinically significant in breast cancer, cBioPortal analyses revealed significantly decreased levels of *IGF1* mRNA and elevated levels of *SHBG* mRNA in Black samples in both BRCA-P and BRCA-C studies.

This study provides a methodological pipeline to identify clinically valuable molecular targets with preexisting datasets. However, we recognize certain limitations. The use of broadly defined racial groups and cancer types in our analyses neglects their great variability and diversity (e.g., racially admixed people, different carcinoma subtypes, etc.). Genetic ancestry and tumor typing information would certainly be of interest to consider in future research. There is also a consistent lack of representation of Black and Asian patients in the data, an issue that must be adequately addressed by the broader research community. The source data only included protein information for a few hundred genes; hence our pathway analyses are not conclusive. We were unable to perform a comprehensive analysis of the prostate adenocarcinoma dataset as it lacked survival data. Finally, we were unable to perform multiple analyses of similar datasets, except BRCA-C and BRCA-P, as other datasets in TCGA and cBioPortal are predominantly represented by White patients and contain limited or no data for Black and Asian patients for adequate statistical comparison.

Nevertheless, in this study, we highlight correlative trends between patients' survival and differentially expressed genes, suggesting that disparities in cancer outcomes overlap with key molecular differences. Future research should work with larger tumor datasets with a greater representation of Asian and Black patients to yield more statistically accurate analyses. Additionally, further work is needed to clarify the root causes of observed molecular differences to find direct relationships between differential expression and disparate outcomes. Ultimately, observed differences in cancer incidence and survival outcome are the result of a combination of many factors, including genetic polymorphisms, distinct molecular signatures, the synergistic effects of many traits, lifestyle, and environmental influences. Associations between self-reported race and disease outcome as presented in this study are thus valuable in predicting health impact of such structural inequalities along with genetic variations. With more proportional racial representation in tumor datasets and increased collection of proteomics data, we will be able to solidify our knowledge of molecular racial differences and work towards the development of personalized cancer therapeutics. Genome projects, such as the "1000 Genomes Project" [49] and the NIHinitiated "All of Us" [50], will provide more opportunities to identify unique molecular signatures in the future.

## 5. Conclusions

Racial patterns of differential gene expression and pathways' regulation varied between carcinomas. The expression of certain cell cycle genes (*CCNB1*, *CCNE1*, *CCNE2*, and *FOXM1*) was race-dependent in several cancer types and correlated with significant survival differences. We observed racial variation in the regulation of several DNA repair mechanisms and oncogenic pathways. Importantly, homology-directed repair pathway was upregulated in Black samples compared to White samples in breast, endometrial, kidney renal clear cell, and lung carcinomas. ERBB4-mediated nuclear signaling, which is associated with cancer development, was upregulated in Black samples compared to White samples in breast, colorectal, kidney renal clear cell, and lung carcinomas. Mapping race-dependent differentially expressed genes that influence overall patients' survival and illustrating their impact on proteomics and functional molecular pathways may provide valuable insights into targeting disparities in cancer.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers15102695/s1, Supplemental S1: List of 155 analyzed genes; Supplemental S2: Complete differential mRNA expression results; Supplemental S3: Complete differential protein expression results; Supplemental S4: Complete differential methylation results; Supplemental S5: Original Reactome report.

**Author Contributions:** This paper was conceptualized, written, and reviewed by A.S. and B.L.; B.L. designed the methodology, wrote the Python (version 3.7.7) program, performed all data analyses, and prepared the original draft; A.S. provided overall supervision and guidance during data collection, analysis, and writing; X.J. reviewed and edited. All authors have read and agreed to the published version of the manuscript.

**Funding:** This project is funded by the National Cancer Institute's Center to Reduce Cancer Health Disparities Geographic Management Program (GMaP) Region 4 and GMaP Region 1 to B.L. This study was in part supported by the PSC-CUNY Enhanced Award (ID 63814-00 51) to A.S. This research was made possible in part with funding from The Tow Foundation as a Tow Faculty Research and Creativity Grant (90266-53-01) to A.S.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study as in this study, we used publicly available open-access data from "The Cancer Genome Atlas" (TCGA) Program.

**Informed Consent Statement:** All samples in TCGA have been collected and utilized following strict human subjects protection guidelines, informed consent, and IRB review of protocols. Click here to view the TCGA ethics and policies.

**Data Availability Statement:** The results shown in this study are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga (accessed on 18 July 2021). Data used in this article is publicly available and has been referenced. The computer code can be found at https://github.com/brian-lei/cancer-racial-disparity (accessed on 18 January 2021).

Acknowledgments: We thank Gianni Nigita (Department of Cancer Biology and Genetics and Comprehensive Cancer Center, The Ohio State University, Columbus, OH 43210, USA) for his valuable comments and insights on this study during the editing of this manuscript.

**Conflicts of Interest:** The authors do not have any conflict of interest.

# References

- 1. Özdemir, B.C.; Dotto, G.-P. Racial Differences in Cancer Susceptibility and Survival: More Than the Color of the Skin? *Trends Cancer* 2017, *3*, 181–197. [CrossRef] [PubMed]
- Zavala, V.A.; Bracci, P.M.; Carethers, J.M.; Carvajal-Carmona, L.; Coggins, N.B.; Cruz-Correa, M.R.; Davis, M.; de Smith, A.J.; Dutil, J.; Figueiredo, J.C.; et al. Cancer Health Disparities in Racial/Ethnic Minorities in the United States. *Br. J. Cancer* 2021, 124, 315–332. [CrossRef]
- 3. Banda, Y.; Kvale, M.N.; Hoffmann, T.J.; Hesselson, S.E.; Ranatunga, D.; Tang, H.; Sabatti, C.; Croen, L.A.; Dispensa, B.P.; Henderson, M.; et al. Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **2015**, *200*, 1285–1295. [CrossRef] [PubMed]
- Borrell, L.N.; Elhawary, J.R.; Fuentes-Afflick, E.; Witonsky, J.; Bhakta, N.; Wu, A.H.B.; Bibbins-Domingo, K.; Rodríguez-Santana, J.R.; Lenoir, M.A.; Gavin, J.R.; et al. Race and Genetic Ancestry in Medicine—A Time for Reckoning with Racism. *N. Engl. J. Med.* 2021, 384, 474–480. [CrossRef] [PubMed]
- 5. Wickland, D.P.; Sherman, M.E.; Radisky, D.C.; Mansfield, A.S.; Asmann, Y.W. Lower Exome Sequencing Coverage of Ancestrally African Patients in The Cancer Genome Atlas. *JNCI J. Natl. Cancer Inst.* **2022**, *114*, 1192–1199. [CrossRef]
- 6. Andaya, A.A.; Enewold, L.; Zahm, S.H.; Shriver, C.D.; Stojadinovic, A.; McGlynn, K.A.; Zhu, K. Race and Colon Cancer Survival in an Equal-Access Health Care System. *Cancer Epidemiol. Biomark. Prev.* **2013**, *22*, 1030–1036. [CrossRef]
- Hanson, H.A.; Martin, C.; O'Neil, B.; Leiser, C.L.; Mayer, E.N.; Smith, K.R.; Lowrance, W.T. The Relative Importance of Race Compared to Health Care and Social Factors in Predicting Prostate Cancer Mortality: A Random Forest Approach. J. Urol. 2019, 202, 1209–1216. [CrossRef]
- 8. Nipp, R.; Tramontano, A.C.; Kong, C.Y.; Pandharipande, P.; Dowling, E.C.; Schrag, D.; Hur, C. Disparities in Cancer Outcomes across Age, Sex, and Race/Ethnicity among Patients with Pancreatic Cancer. *Cancer Med.* **2018**, *7*, 525–535. [CrossRef] [PubMed]
- Dess, R.T.; Hartman, H.E.; Mahal, B.A.; Soni, P.D.; Jackson, W.C.; Cooperberg, M.R.; Amling, C.L.; Aronson, W.J.; Kane, C.J.; Terris, M.K.; et al. Association of Black Race With Prostate Cancer–Specific and Other-Cause Mortality. *JAMA Oncol.* 2019, 5, 975–983. [CrossRef]
- Spratt, D.E.; Chen, Y.-W.; Mahal, B.A.; Osborne, J.R.; Zhao, S.G.; Morgan, T.M.; Palapattu, G.; Feng, F.Y.; Nguyen, P.L. Individual Patient Data Analysis of Randomized Clinical Trials: Impact of Black Race on Castration-Resistant Prostate Cancer Outcomes. *Eur. Urol. Focus* 2016, 2, 532–539. [CrossRef]
- Lee, S.; Reha, J.L.; Tzeng, C.-W.D.; Massarweh, N.N.; Chang, G.J.; Hetz, S.P.; Fleming, J.B.; Lee, J.E.; Katz, M.H. Race Does Not Impact Pancreatic Cancer Treatment and Survival in an Equal Access Federal Health Care System. *Ann. Surg. Oncol.* 2013, 20, 4073–4079. [CrossRef] [PubMed]
- 12. Hassan, M.O.; Arthurs, Z.; Sohn, V.Y.; Steele, S.R. Race Does Not Impact Colorectal Cancer Treatment or Outcomes with Equal Access. *Am. J. Surg.* 2009, *197*, 485–490. [CrossRef] [PubMed]
- 13. Lin, J.; Zahm, S.H.; Shriver, C.D.; Purdue, M.; McGlynn, K.A.; Zhu, K. Survival among Black and White Patients with Renal Cell Carcinoma in an Equal-Access Health Care System. *Cancer Causes Control* **2015**, *26*, 1019–1026. [CrossRef] [PubMed]
- 14. Tramontano, A.C.; Nipp, R.; Mercaldo, N.D.; Kong, C.Y.; Schrag, D.; Hur, C. Survival Disparities by Race and Ethnicity in Early Esophageal Cancer. *Dig. Dis. Sci.* 2018, *63*, 2880–2888. [CrossRef]
- DeSantis, C.E.; Siegel, R.L.; Sauer, A.G.; Miller, K.D.; Fedewa, S.A.; Alcaraz, K.I.; Jemal, A. Cancer Statistics for African Americans, 2016: Progress and Opportunities in Reducing Racial Disparities: Cancer Statistics for African Americans. *CA Cancer J. Clin.* 2016, 66, 290–308. [CrossRef] [PubMed]
- 16. Karakas, C.; Wang, C.; Deng, F.; Huang, H.; Wang, D.; Lee, P. Molecular Mechanisms Involving Prostate Cancer Racial Disparity. *Am. J. Clin. Exp. Urol.* **2017**, *5*, 34–48.
- 17. Cote, M.L.; Ruterbusch, J.J.; Olson, S.H.; Lu, K.; Ali-Fehmi, R. The Growing Burden of Endometrial Cancer: A Major Racial Disparity Affecting Black Women. *Cancer Epidemiol. Biomark. Prev.* **2015**, 24, 1407–1415. [CrossRef] [PubMed]
- 18. Islami, F.; Miller, K.D.; Siegel, R.L.; Fedewa, S.A.; Ward, E.M.; Jemal, A. Disparities in Liver Cancer Occurrence in the United States by Race/Ethnicity and State: Liver Cancer in the United States. *CA A Cancer J. Clin.* **2017**, *67*, 273–289. [CrossRef]
- Sung, J.M.; Martin, J.W.; Jefferson, F.A.; Sidhom, D.A.; Piranviseh, K.; Huang, M.; Nguyen, N.; Chang, J.; Ziogas, A.; Anton-Culver, H.; et al. Racial and Socioeconomic Disparities in Bladder Cancer Survival: Analysis of the California Cancer Registry. *Clin. Genitourin. Cancer* 2019, 17, e995–e1002. [CrossRef]
- 20. Tannenbaum, S.L.; Koru-Sengul, T.; Zhao, W.; Miao, F.; Byrne, M.M. Survival Disparities in Non–Small Cell Lung Cancer by Race, Ethnicity, and Socioeconomic Status. *Cancer J.* **2014**, *20*, 237–245. [CrossRef]
- Li, Y.; Pang, X.; Cui, Z.; Zhou, Y.; Mao, F.; Lin, Y.; Zhang, X.; Shen, S.; Zhu, P.; Zhao, T.; et al. Genetic Factors Associated with Cancer Racial Disparity—An Integrative Study across Twenty-one Cancer Types. *Mol. Oncol.* 2020, 14, 2775–2786. [CrossRef] [PubMed]

- Grunda, J.M.; Steg, A.D.; He, Q.; Steciuk, M.R.; Byan-Parker, S.; Johnson, M.R.; Grizzle, W.E. Differential Expression of Breast Cancer-Associated Genes between Stage- and Age-Matched Tumor Specimens from African- and Caucasian-American Women Diagnosed with Breast Cancer. *BMC Res. Notes* 2012, *5*, 248. [CrossRef] [PubMed]
- Vazquez, E.D.; Fang, X.; Levesque, L.A.; Huynh, M.; Venegas, C.; Lu, N.; Salazar, N. Chemokine Receptors Differentially Expressed by Race Category and Molecular Subtype in the Breast Cancer TCGA Cohort. *Sci. Rep.* 2022, *12*, 10825. [CrossRef] [PubMed]
- Powell, I.J.; Dyson, G.; Land, S.; Ruterbusch, J.; Bock, C.H.; Lenk, S.; Herawi, M.; Everson, R.; Giroux, C.N.; Schwartz, A.G.; et al. Genes Associated with Prostate Cancer Are Differentially Expressed in African American and European American Men. *Cancer Epidemiol. Biomark. Prev.* 2013, 22, 891–897. [CrossRef] [PubMed]
- Bhardwaj, A. Racial Disparities in Prostate Cancer a Molecular Perspective. *Front. Biosci.* 2017, 22, 772–782. [CrossRef] [PubMed]
   Asif, H.; Foley, G.; Simon, M.; Roque, D.; Kim, J.J. Analysis of Endometrial Carcinoma TCGA Reveals Differences in DNA
- Methylation in Tumors from Black and White Women. *Gynecol. Oncol.* 2023, 170, 1–10. [CrossRef]
  27. Huo, D.; Hu, H.; Rhie, S.K.; Gamazon, E.R.; Cherniack, A.D.; Liu, J.; Yoshimatsu, T.F.; Pitt, J.J.; Hoadley, K.A.; Troester, M.; 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. [CrossRef]
- Heath, E.I.; Lynce, F.; Xiu, J.; Ellerbrock, A.; Reddy, S.K.; Obeid, E.; Liu, S.V.; Bollig-Fischer, A.; Separovic, D.; Vanderwalde, A. Racial Disparities in the Molecular Landscape of Cancer. *Anticancer Res.* 2018, *38*, 2235–2240. [CrossRef]
- 29. Jovov, B.; Araujo-Perez, F.; Sigel, C.S.; Stratford, J.K.; McCoy, A.N.; Yeh, J.J.; Keku, T. Differential Gene Expression between African American and European American Colorectal Cancer Patients. *PLoS ONE* **2012**, *7*, e30168. [CrossRef]
- 30. Dubil, E.A.; Tian, C.; Wang, G.; Tarney, C.M.; Bateman, N.W.; Levine, D.A.; Conrads, T.P.; Hamilton, C.A.; Maxwell, G.L.; Darcy, K.M. Racial Disparities in Molecular Subtypes of Endometrial Cancer. *Gynecol. Oncol.* **2018**, *149*, 106–116. [CrossRef]
- 31. Ashktorab, H.; Kupfer, S.S.; Brim, H.; Carethers, J.M. Racial Disparity in Gastrointestinal Cancer Risk. *Gastroenterology* **2017**, 153, 910–923. [CrossRef] [PubMed]
- Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The CBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* 2012, 2, 401–404. [CrossRef] [PubMed]
- Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the CBioPortal. *Sci. Signal.* 2013, *6*, pl1. [CrossRef] [PubMed]
- 34. CBioPortal for Cancer Genomics. Available online: https://www.cbioportal.org/ (accessed on 18 July 2021).
- Hoadley, K.A.; Yau, C.; Hinoue, T.; Wolf, D.M.; Lazar, A.J.; Drill, E.; Shen, R.; Taylor, A.M.; Cherniack, A.D.; Thorsson, V.; et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* 2018, 173, 291–304.e6. [CrossRef] [PubMed]
- Ciriello, G.; Gatza, M.L.; Beck, A.H.; Wilkerson, M.D.; Rhie, S.K.; Pastore, A.; Zhang, H.; McLellan, M.; Yau, C.; Kandoth, C.; et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* 2015, 163, 506–519. [CrossRef]
- Abeshouse, A.; Ahn, J.; Akbani, R.; Ally, A.; Amin, S.; Andry, C.D.; Annala, M.; Aprikian, A.; Armenia, J.; Arora, A.; et al. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 2015, *163*, 1011–1025. [CrossRef] [PubMed]
- TCGA Forms and Documents. Available online: https://www.nationwidechildrens.org/research/areas-of-research/ biopathology-center/nci-ccg-project-team/the-cancer-genome-atlas/tcga-forms-and-documents (accessed on 20 March 2023).
- 39. About the Topic of Race. Available online: https://www.census.gov/topics/population/race/about.html (accessed on 20 March 2023).
- 40. Balázs, G. KM-Plot. Available online: http://www.kmplot.com/analysis (accessed on 18 July 2021).
- Nagy, Á.; Lánczky, A.; Menyhárt, O.; Győrffy, B. Author Correction: Validation of MiRNA Prognostic Power in Hepatocellular Carcinoma Using Expression Data of Independent Datasets. Sci. Rep. 2018, 8, 11515. [CrossRef]
- The Cancer Genome Atlas Research Network; Weinstein, J.N.; Collisson, E.A.; Mills, G.B.; Shaw, K.R.M.; Ozenberger, B.A.; Ellrott, K.; Shmulevich, I.; Sander, C.; Stuart, J.M. The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* 2013, 45, 1113–1120. [CrossRef]
- 43. PanCanAtlas Publications | NCI Genomic Data Commons. Available online: https://gdc.cancer.gov/about-data/publications/ pancanatlas (accessed on 18 July 2021).
- 44. Fabregat, A.; Sidiropoulos, K.; Viteri, G.; Marin-Garcia, P.; Ping, P.; Stein, L.; D'Eustachio, P.; Hermjakob, H. Reactome Diagram Viewer: Data Structures and Strategies to Boost Performance. *Bioinformatics* **2018**, *34*, 1208–1214. [CrossRef]
- 45. Reactome Pathway Database. Available online: https://reactome.org/ (accessed on 18 July 2021).
- Porter, P.L.; Lund, M.J.; Lin, M.G.; Yuan, X.; Liff, J.M.; Flagg, E.W.; Coates, R.J.; Eley, J.W. Racial Differences in the Expression of Cell Cycle-Regulatory Proteins in Breast Carcinoma: Study of Young African American and White Women in Atlanta, Georgia. *Cancer* 2004, 100, 2533–2542. [CrossRef]
- El-Gamal, M.I.; Mewafi, N.H.; Abdelmotteleb, N.E.; Emara, M.A.; Tarazi, H.; Sbenati, R.M.; Madkour, M.M.; Zaraei, S.-O.; Shahin, A.I.; Anbar, H.S. A Review of HER4 (ErbB4) Kinase, Its Impact on Cancer, and Its Inhibitors. *Molecules* 2021, 26, 7376. [CrossRef] [PubMed]

- 48. Wang, G.; Firoz, E.F.; Rose, A.; Blochin, E.; Christos, P.; Pollens, D.; Mazumdar, M.; Gerald, W.; Oddoux, C.; Lee, P.; et al. MDM2 Expression and Regulation in Prostate Cancer Racial Disparity. *Int. J. Clin. Exp. Pathol.* **2009**, *2*, 353–360. [PubMed]
- 1000 Genomes Project Consortium; Auton, A.; Brooks, L.D.; Durbin, R.M.; Garrison, E.P.; Kang, H.M.; Korbel, J.O.; Marchini, J.L.; McCarthy, S.; McVean, G.A.; et al. A Global Reference for Human Genetic Variation. *Nature* 2015, 526, 68–74. [CrossRef] [PubMed]
- 50. The All of Us Research Program Investigators The "All of Us" Research Program. N. Engl. J. Med. 2019, 381, 668–676. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.