Comparing Therapeutically-Relevant Copy Number Profiles in African Americans and European Americans with Renal Cell Carcinoma

Thesis Highlights

- Kidney cancer is one of the top ten most diagnosed forms of cancer in the US and affects over 70,000 people per year with a five-year survival rate of 76.5%.
- Renal cell carcinoma (RCC) is the most common, making up 85% of all kidney cancer cases: Clear cell RCC (ccRCC) comprises 75%, papillary RCC (pRCC) is 10-15% of RCC, and chromophobe RCC (chRCC) makes up nearly 5%.
- African Americans (AAs) have higher incidence rates of RCC in comparison to European Americans (EAs). When comparing AA and EA stage at diagnosis, AAs have lower 5-year survival for advanced cancers (regional: 62% vs 74%) and (distant: 10.4% vs 13.5%).
- Genomic studies suggest that AA ccRCC patients are less likely to respond to *VHL* targeted therapy due to different mutation types and frequencies. Other genomic factors may be involved, like Somatic Copy Number Variations (SCNVs). SCNVs are genomic amplifications or deletions acquired over time and commonly found in the human genome. Many RCC patients have characteristic SCNV profiles in important tumor suppressor genes and oncogenes.
- Self-reported AAs have population-specific SCNVs and may differ by SCNV status.
- High West African ancestry (WAA) is associated with poor cancer survival in AAs and may differ by SCNV status.
- This honors thesis profiled SCNVs by self-reported race and genetic ancestry to help explain RCC survival disparities. By race, AAs and EAs had population-specific deletions and amplifications that positively correlated with gene expression and patient survival. By ancestry, patients with high and low WAA had distinct gene deletions and amplifications previously associated with cancer-related pathways. WAA predicted *NOVA1* candidate gene expression. Low expression and high WAA was associated with better patient survival. Both self-reported race and genetic ancestry should be considered when categorizing patients in the clinic.

Honors Thesis by Erica Beatson (‘21)
Academic year (2020-2021)
Advisor: Dr. Khadijah A. Mitchell
Committee: Dr. James Dearworth, Dr. Eric Ho, and Dr. Tamara Stawicki

TABLE OF CONTENTS

I. Biographical Sketch

I am Erica Beatson, a senior at Lafayette College working towards a Bachelor of Science degree in Biology. Upon graduation, I will have completed four semesters of research with Dr. Khadijah A. Mitchell. I requested to join her lab during the spring 2019 semester, while taking her course, Precision Medicine, and then began my independent research project the following fall semester.

My interest in cancer was derived from my father's diagnosis with papillary renal cell carcinoma in 2010 and reinforced while I completed the CURE laboratory project, studying lung cancer health disparities, associated with Precision Medicine. Dr. Mitchell was awarded a Department of Defense grant to fund a kidney cancer health disparities project, and offered me a position that combined my interests in racial inequalities and kidney cancer. After a year conducting independent research, I continued my work as a summer EXCEL research fellow, where I then decided to pursue a senior honors thesis. My preliminary findings as an independent researcher served as a basis for the knowledge required to conduct and write my thesis.

After graduation, I will work as a postbaccalaureate CRTA research fellow at the National Institute of Health, under the Principal Investigator William Douglas Figg Sr, PhD. Working in a molecular pharmacology lab, I will continue to study cancer genetics, with a focus on prostate cancer, and I will have the opportunity to further develop my laboratory skills.

II. Abstract

Background: Kidney cancer is one of the top ten most diagnosed forms of cancer in the US and affects over 70,000 people per year with a five-year survival rate of 76.5%. Renal cell carcinoma (RCC) is the most common, making up 85% of all kidney cancer cases. Clear cell RCC (ccRCC) comprises 75%, papillary RCC (pRCC) is 10-15% of RCC, and chromophobe RCC (chRCC) makes up nearly 5%. African Americans (AAs) have higher incidence rates of RCC in comparison to European Americans (EAs). When comparing AA and EA stage at diagnosis, AAs have lower 5-year survival for advanced cancers (regional: 62% vs 74%) and (distant: 10.4% vs 13.5%). Recent genomic studies suggest AA patients are less likely to respond to targeted inhibitors than EA patients based on population-specific mutations. Other genomic determinants may also be involved, such as somatic copy number variations (SCNVs) and genetic ancestry. SCNVs are the most common structural variations in the human genome and can lead to alteration of kidney cancer-related oncogene and tumor suppressor gene expression. Recent work has associated high West African ancestry (WAA) with poor cancer survival in AA patients. Abnormal SCNV patterns have been identified in RCC patients. SCNV frequencies also differ between healthy AAs and EAs. To our knowledge, no study to date has profiled SCNVs by self-reported race and genetic ancestry to help explain RCC survival disparities.

Hypothesis (Chapter 1): Self-reported AAs and EAs have different amplification and deletion profiles, resulting in novel population-specific SCNVs that relate to patient survival.

Hypothesis (Chapter 2): Patients with high WAA have distinct SCNV changes that correlate with more aggressive RCC biology and patient survival.

Methods: A diverse human reference genome was created using Affymetrix SNP 6.0 array data and demographic information for AA (n = 99) and EA (n = 182) 1000 Genomes Project study participants (International Genome Sample Resource Database). Affymetrix SNP 6.0 array data were downloaded for RCC patients (ccRCC: n = 513; pRCC: n = 257; chRCC: n = 62) and merged with clinical and demographic information (National Cancer Institute Genomics Data Commons Legacy Archive and cBioPortal). Differential SCNV and gene expression analyses were performed by self-reported race and WAA at the genome-wide, chromosome, chromosome region, and gene levels. Disease specific survival curves were generated based on median gene expression. Simple linear regression was used to show the relationship between candidate gene expression and WAA.

Results: By race, SCNV profiles had minimal variation at the genome-wide level. A total of 6,099 out of 53,786 chromosomal SCNV segments varied significantly by race (*P* + FDR <0.05). AA ccRCC patients had 71 unique SCNVs, including 15 deleted and 56 amplified genes, and EA ccRCC patients had 258 unique SCNVs, including 42 deleted and 216 amplified genes. Five AA- and EA- unique SCNVs had strong clinical relevance. Four AA- and EA- unique SCNVs have positive correlations with gene expression. AA ccRCC patients with low candidate gene expression have higher 5-year disease specific survival than EAs, and this difference is statistically significant for *LRBA*. WAA better classified the tumor suppressor gene, *PTEN*, than self-reported race (P <0.05 vs not significant). By ancestry, there were subtle genome-wide differences across the ccRCC genome. There were 52,220 out of 111,766 chromosomal SCNV segments that varied significantly by WAA in ccRCC patients (*P* + FDR <0.05). This corresponded with three chromosomes (7, 12, and 19), 8 chromosome region/cytobands, and 12 genes. Three gene SCNVs were associated with high WAA (*ACTR3C*, *IMMP2L*, and *SLCO1B3*). WAA did not predict candidate gene expression. By ancestry, there were also subtle genome-wide differences across the pRCC genome. There were 476 out of 59,868 chromosomal SCNV segments that varied significantly by WAA in pRCC patients (*P* <0.05). This corresponded with 16 chromosomes, 103 chromosome region/cytobands, and 793 genes. After a differential expression analysis, the top three candidate genes were selected (*NOVA1*, *RPL23AP7*, and *NOMO3)*. These three gene SCNVs were associated with high WAA (*ACTR3C*, *RPL23AP7*, and *NOMO3*). WAA predicted *NOVA1* gene expression. *NOVA1* gene expression was not associated with patient survival.

Conclusion: By race, AAs and EAs had population-specific deletions and amplifications that correlated with gene expression and patient survival. By ancestry, patients with high and low WAA had distinct gene deletions and amplifications previously associated with cancer-related pathways. *NOVA1*, a key candidate gene, was not associated with patient survival. Both self-reported race and genetic ancestry should be considered when categorizing patients in the clinic.

Discussion: AA patients with RCC may be less likely to respond to targeted therapies than their EA counterparts because of differences in RCC tumor biology and greater genetic admixture. Future studies should find novel diagnostic, prognostic, and treatment biomarkers to create new targeted therapies in AAs.

III. Introduction

A. Kidney cancer incidence, mortality, and survival statistics

Kidney cancer is amongst the top ten most common cancers, with rates increasing since 1975[1]. In 2021, there will be an estimated 63,060 newly diagnosed kidney cancer cases and 13,780 deaths in the US[1]. The projected increase in incidence is partially attributed to the growth of the aging population and advances in detection rate[2]. Similarly, a decrease in projected deaths is likely accredited to the discovery of novel therapies[2]. The five year survival rate for kidney cancer cases combined is about 76.5 percent for all races, sexes, ages, and stages[3].

B. Histologic types and subtypes

Renal Cell Carcinoma (RCC) is the most prevalent histology, or appearance under a microscope, accounting for 85 percent of kidney cancer cases[1]. RCC is further broken down into three histologic subtypes: clear cell, papillary, and chromophobe. The most frequent is clear cell RCC (ccRCC), consisting of 60 to 80 percent of RCC cases[1]. Broadly speaking, it is distinguished by mutations in the hypoxia signaling pathway, heightened angiogenesis and intra-tumor heterogeneity[4]. Specifically, ccRCC is characterized by an increased ribose metabolism pathway and mRNA expression associated with poor survival[5]. Generally, ccRCC is followed by changes in glucose and fatty acid metabolism and in the tricarboxylic acid cycle[6]. Therefore, ccRCC is understood to be a metabolic disease[6]. ccRCC is additionally distinguished by mutations to the PI3K/AKT pathway and the SWI/SNF chromatin remodelling complex[7]. ccRCC is divided into two subtypes: ccA and ccB, the latter is associated with a more aggressive phenotype and poorer survival outcomes[8].

The second most common is papillary RCC (pRCC), which makes up 15 to 20 percent of RCC cases and is divided into two subtypes[9]. Type I is multifocal and has papillae and tubular

structures surrounded by small cells with small, oval nuclei and basophilic cytoplasm and type II is heterogeneous and has large cells with large, spherical nuclei[10]. Furthermore, most type I tumors are localized and are diagnosed in stage I, while type II tumors are more advanced and found in stages II or III[10]. The most common pRCC-associated mutation is in the oncogene *MET,* which activates intracytoplasmic tyrosine kinase domains and therefore the hepatocyte growth factor (HGF)/MET pathway[11]. Finally, chromophobe RCC (chRCC) accounts for approximately 5 percent of all RCC cases and commonly has a genomic rearrangement within the *TERT* promoter region resulting in higher expression[9].

C. Kidney cancer treatment

RCC is treated according to Tumor-Node-Metastasis (TNM) stage and histology. TNM is a widely used cancer staging system[12]. T refers to the size of the primary tumor, N refers to lymph node presence of cancer, and M defines whether the cancer has metastasized[12]. Due to molecular differences, response to treatment varies among patients[1]. Stages I, II, and III are treated the same; the first line of defense is surgery, followed by surveillance, and targeted treatment with Tyrosine Kinase inhibitors (TKI)[1]. Patients with Stage IV cancer are broken down into surgically resectable and unresectable cancers. Resectable cancers are treated similarly to stage I-III cancers, with the addition of immunotherapies such as Interleukin-2 (IL-2)[1]. Unresectable clear cell tumors are treated by surveillance, TKI, or IL-2[1].

The framework of RCC treatment has evolved over the past 15 year, with a recent increase in combination therapies utilizing immunotherapies[13]. The primary form of treatment remains radical or partial nephrectomy, because its 5-year survival rate ranges up to 93 percent[14]. Secondary treatment is important for RCC patients that are not candidates for surgery, such as the 30 percent of patients diagnosed with advanced and metastatic disease, and the 10-20 percent

treated with early-stage disease who endure recurrence[14]. Survival outcomes drop to 67 percent and 12 percent for patients with regional metastases and distant metastases, respectively[14]. Therefore, the use of targeted therapies has the strong potential to improve patient outcomes[14].

D. Kidney cancer disparities and targeted genomic treatments based on DNA mutations

1. *Incidence and survival by race*

African American (AA) males and females have higher rates of kidney and renal pelvis cancers in comparison to their European American (EA) counterparts[15]. Incidence rates per 100,000 AAs is 18.7; whereas, rates of cancer for EAs is only 16.8[2]. AAs not only have higher rates of ccRCC[1], but also a higher frequency of the more aggressive subtype (ccB)[8]. When diagnosed with localized disease, AAs and EAs survive at 91.9 and 93 percent[3]. Survival rates for AAs diagnosed with regional cancer is 62 percent, compared to 74 percent for EAs[3]. Patients with distant cancer have racial differences in survival, with AAs at 10.4 percent and EAs at 13.5 percent[3]. According to the most recent NCI SEER data, AA kidney cancer patients with localized and regional disease have significantly lower survival than EAs between 2000 and 2012.

2. *VHL mutations*

Genomic alteration profiles vary between healthy AAs and EAs[7], suggesting that biological predispositions may be a driver of the cancer disparity. *VHL* is a tumor suppressor gene linked to kidney cancer; 52 to 82 percent of ccRCC cases have *VHL* inactivations[8] and nearly 60 percent of individuals with the mutation develop cancer[8]. *VHL* is significantly less mutated in AAs, suggesting that there are other forms of genetic variation that initiate tumor development[8]. Additionally, AA patients who lack *VHL* mutations will be less likely to respond to FDA approved *VHL* inhibiting therapies[8].

E. Reducing kidney cancer disparities and potential targeted genomic treatments based on DNA copy number variations

1. *Copy number variations (CNV)*

Copy number variations (CNVs) are the most common structural variations in the human genome and result in an alteration of the number of copies of a gene[16]. CNVs may be a result of a duplication or deletion of a gene which can modify gene expression. Duplication of oncogenic genes can initiate uncontrollable cell cycling and amplify cell growth. In fact, a prior enrichment analysis of ccRCC indicated that amplified genes were involved in cancer-related signaling transduction pathways[17]. Deletion of tumor suppressor genes can prevent DNA repair or apoptosis and continue abnormal cellular growth.

Somatic CNVs (SCNVs) arise *de novo*, whereas germline CNVs (GCNVs) are inherited. Targeting SCNVs, which may arise during tumor progression, could contribute to the development of new targeted therapies. Identification of novel population-specific biomarkers could be utilized to treat AAs, improving their disease-specific survival outcomes.

F. References

1. American Cancer Society. "Cancer Facts & Figures 2021." *American Cancer Society*, 2021, 7.
2. Miller, K. D. *et al.* Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin*69, 363–385 (2019).
3. National Cancer Institute. (2018). Kidney Cancer - Surveillance, Epidemiology, and End Results Program - Cancer stat facts. Retrieved October, 2020
4. Wolf, M. M., Kimryn Rathmell, W. & Beckermann, K. E. Modeling clear cell renal cell carcinoma and therapeutic implications. *Oncogene*39, 3413–3426 (2020).
5. Teh, B. T., Farber, L. J. & Furge, K. Molecular Characterization of Renal Cell Carcinoma. in *Renal Cell Carcinoma*(eds. Figlin, R. A., Rathmell, W. K. & Rini, B. I.) 91–111 (Springer US, 2012). doi:10.1007/978-1-4614-2400-0_5.
6. Wettersten, H. I., Aboud, O. A., Lara, P. N. & Weiss, R. H. Metabolic reprogramming in clear cell renal cell carcinoma. *Nat Rev Nephrol*13, 410–419 (2017).

7. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*499, 43–49 (2013).
8. Krishnan, B., Rose, T. L., Kardos, J., Milowsky, M. I. & Kim, W. Y. Intrinsic Genomic Differences Between African American and White Patients With Clear Cell Renal Cell Carcinoma. *JAMA Oncol*2, 664 (2016).
9. Ricketts, C. J. *et al.*The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep*23, 313-326.e5 (2018).
10. The Cancer Genome Atlas Research Network. Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. New England Journal of Medicine. November 4, 2015.
11. Singer, E. A., Bratslavsky, G., Linehan, W. M. & Srinivasan, R. Targeted therapies for non-clear renal cell carcinoma. *Target Oncol*5, 119–129 (2010).
12. American Joint Committee on Cancer. "Breast Cancer Staging System" *American Joint Committee on Cancer*, 2009, 7.
13. Lalani, A.-K. A. *et al.*Systemic Treatment of Metastatic Clear Cell Renal Cell Carcinoma in 2018: Current Paradigms, Use of Immunotherapy, and Future Directions. *Eur Urol*75, 100–110 (2019).
14. Tannir, N. M., Pal, S. K. & Atkins, M. B. Second-Line Treatment Landscape for Renal Cell Carcinoma: A Comprehensive Review. *Oncologist*23, 540–555 (2018).
15. Actual and Projected Cancer Incidence Rates, United States, 1975 to 2020 | CDC. (2019).
16. Aouiche, C., Shang, X. & Chen, B. Copy number variation related disease genes. *Quant Biol*6, 99–112 (2018).
17. Zhou, W. *et al.*Comprehensive Analysis of Copy Number Variations in Kidney Cancer by Single-Cell Exome Sequencing. *Front Genet*10, 1379 (2019).

IV. Chapter 1: Identifying Novel SCNV Biomarkers by Race in Clear Cell Renal Cell Carcinoma Patients

A. Introduction

1. GCNV profile differences between healthy AAs and EAs

CNVs are more common than single nucleotide polymorphisms (SNPs), implicating the widespread nature of CNVs in the human genome. Approximately 5,600 human CNV loci have been identified[1], with various associations to disease[1]. Understanding population-specific CNVs could help to understand possible driving forces in cancer biology differences.

GCNV profile differences have been studied between healthy, or noncancerous, AAs and EAs. One study found that there were GCNV frequency differences on each of the 23 chromosomes, with varying significance[1]. Specifically, chromosomes 15q and 17q have significant differences in GCNV frequency[1]. Because GCNV profiles differ between AAs and EAs[1], it is possible SCNVs may also be population-specific. However, no known study to date has profiled SCNVs to help explain the racial differences in ccRCC outcomes.

2. Peak chromosomal regions affected by SCNVs in ccRCC patients

SCNV patterns help to provide clear distinction between RCC subtypes[2]. ccRCC tumors are characterized by amplifications and deletions in specific chromosomal regions. A total of 91 percent of ccRCC patients have a deletion of chromosome 3p, 45 percent have 14q deletions and 67 percent have an amplification of 5q[3]. Loss of chromosome 3p results in somatic copy number changes to four genes, most commonly associated with ccRCC: *VHL*, *PBRM1*, *BAP1*, and *SETD2* and deletion of 14q results in loss of *HIF1A*, which has been linked with aggressive disease[3]. Little is known about the amplified region on chromosome 5q[3].

In comparison to all other cancer types, ccRCC has more SCNVs impacting entire chromosome arms and fewer regional SCNVs[3]. Loss or gain of entire chromosome arms can result in changes that promote tumor progression[4]. To our knowledge, no study has profiled population-specific SCNVs to explain the racial differences in ccRCC survival.

3. Hypothesis

I predict that self-reported AAs and EAs have different amplification and deletion profiles, resulting in novel population-specific SCNVs that relate to patient survival.

B. Methods

1. Creating a diverse human reference genome

Downloaded Affymetrix SNP 6.0 array CEL files and demographic information for AA and EA 1000 Genomes (1KG) Project participants from the International Genome Sample Resource database (https://www.internationalgenome.org/category/affy/). Two populations were included: Americans of African Ancestry in the Southwestern USA (ASW, *n* = 99) and Utah Residents with Northern and Western European Ancestry (CEU, *n* = 182). A CN.model file was created using Partek Genomics Suite 7.0 (PGS) to serve as a reference. A statistical analysis was performed for potential confounding due to sex.

2. Accessing and annotating primary tumor data from AA and EA ccRCC patients

Utilized the National Cancer Institute Genomics Data Commons Legacy Archive to download Affymetrix Human SNP 6.0 array data for AA (*n* =56) and EA (*n* = 466) ccRCC patients in The Cancer Genome Atlas (TCGA) study (https://portal.gdc.cancer.gov/legacy-archive/search/f). Downloaded clinical and demographic data for the same populations from cBioPortal (https://www.cbioportal.org/). Statistical analyses were performed for potential confounding variables. Genetic ancestry data for the 513 patients were downloaded from The

Cancer Genetic Ancestry Atlas (TCGAA) and consequently sorted based on percentage of West African ancestry (http://fcgportal.org/TCGAA/).

3. Differential SCNV analysis by self-reported race

Performed differential genome-wide, chromosomal, and gene SCNV analysis by self-reported race in kidney tumor tissues. Imported the .CEL files into PGS and used the Batch Effect Removal Tool to normalize the Affymetrix array values. The tool removed variability due to the experimental scan date and also possible confounding clinical and demographic variables. The PGS Copy Number Workflow was employed to complete a genome-wide segmentation analysis, identify chromosomal regions of deletion or amplification, and discover candidate gene alterations via one-way ANOVAs. Regions with less than or equal to one copy ($P + FDR < 0.05$) were considered to be deleted, while regions with greater than or equal to three copies ($P + FDR < 0.05$) were impacted by amplifications. Candidate genes were prioritized based on their function in GeneCards (https://www.genecards.org) and PubMed (https://pubmed.ncbi.nlm.nih.gov/) and association with human disease in MalaCards (https://www.malacards.org).

4. Differential gene expression analysis by self-reported race

Downloaded normalized mRNA-sequencing expression data (RSEM) from the TCGA Splicing Variant Database (TSVdb) (http://www.tsvdb.com) for candidate genes. Unpaired parametric t-tests with Welch's corrections were individually conducted using GraphPad Prism 8 for each gene.

5. Disease-specific survival based on median gene expression

Utilized cBioPortal (https://www.cbioportal.org) to determine 5-year disease specific survival outcomes for AA and EA ccRCC patients based on high or low mRNA expression of

candidate SCNV genes. Patients were divided into high or low groups based on median expression. Kaplan-Meier survival curves were plotted, and log rank P values were calculated.

6. Incorporating genetic ancestry data for potential gene targets

The ancestry of AAs is primarily from Niger (~71 percent), and more generally from West Africa (73.2 percent)[5,6]. Patients were divided into two categories: individuals with greater than 70 percent West African ancestry (WAA) and those with less than 70 percent WAA. Differences in gene expression according to WAA status were tested via unpaired parametric t-tests with Welch's corrections.

C. Results

1. There were no significant differences by sex in the diverse human reference genome study participants

The proportion of females to males in both populations was nearly 50:50 (Table 1). There were no significant differences in participant sex between the two populations, as determined by a Fisher's exact test (Table 1).

2. There were significant differences by sex, stage, grade, and vital status in the ccRCC patient cohort

Four out of five clinical and demographic variables had significant differences by self-reported (Table 2). The mean age at diagnosis was similar between AAs and EAs, with EAs having a wider range. EAs have a significantly higher percentage of males and nearly twice as many patients diagnosed with Stage IV cancer (Table 2). There are four times as many EA individuals with Grade IV cancer compared to AAs (Table 2). Finally, EAs are 1.5x more likely to have died (Table 2).

3. AA ccRCC patients had 71 unique SCNVs, including 15 deleted and 56 amplified genes, and EA ccRCC patients had 258 unique SCNVs, including 42 deleted and 216 amplified genes

Across the genomes of the AA and EA ccRCC patients, there were 53,786 genomic segments detected, 6,099 of which were significant by race (P + FDR < 0.05) (Figure 1A). Subtle variations were observed across the genome (Figure 1B). AA patients had 38,806 deleted regions, comprising all chromosomes except 9, 16, 18, 20, 21, 22, and X. These regions corresponded to 42 cytobands and 45 genes. Fifteen genes are uniquely deleted in AAs (Table 3). AA patients had 741 amplified regions, which spanned every chromosome except 20, X, and Y. The regions consisted of 121 cytobands and 141 genes. Fifty-six of which were uniquely amplified in AAs (Table 4). EA patients had 38,802 deletions across all chromosomes except 8, 9, 14, 18, 20, 21, 22, and X. These regions correlated to 40 cytobands and 42 genes. Twelve of which were uniquely deleted in EAs (Table 5). EA patients also had 868 amplified regions across all chromosomes except 20, 21, X, and Y, which corresponded to 100 cytobands and 216 genes. A total of 130 genes were uniquely amplified in EAs (Table 6). In summary, we observed population-specific chromosome deletion and amplification patterns for AAs (Figure 2A-B) and EAs (Figure 2C-D).

4. Five AA- and EA- unique SCNVs had strong clinical relevance

The 71 AA-unique and 258 EA-unique SCNVs were ranked based on clinical promise, such as their gene function and association with human disease. Two interesting candidate genes had significant SCNVs in AAs (unique deletion: *RAB3C*, unique amplification: *LRBA*). *RAB3C* is a member of the Ras oncogene family and *LRBA* is an immune system response gene involved in bacterial infection (Table 7). Three interesting candidate genes had significant SCNVs in EAs

(unique deletions: *GSTM1*, *GSTM2*, unique amplification: *PTEN*). Both deleted genes are involved in the detoxification of compounds and *PTEN* is a tumor suppressor gene that negatively regulates the AKT/PKB pathway (Table 7). *LRBA*, *GSTM1*, and *PTEN* have been previously associated with kidney cancer in general, RCC specifically, and other kidney-related diseases (Table 7). *RAB3C* and *GSTM2* are associated with various cancers (Table 7).

5. Four AA- and EA- unique SCNVs have positive correlations with gene expression

AAs had more *RAB3C* deletions and significantly lower gene expression than EAs (Figure 5). AAs had greater *LRBA* amplifications, however, they had significantly less gene expression (Figure 5). Of the AA-unique SCNVs, only *RAB3C* had a positive correlation between the SCNV and gene expression statuses. EAs had deletions of *GSTM1* and *GSTM2* and significantly lower expression when compared to AAs (Figure 5). EAs had more *PTEN* amplifications and higher median expression, although it was not significant (Figure 5). Of the EA-unique SCNVs, all three had a positive correlation between the SCNV and gene expression statuses.

6. AA ccRCC patients with low candidate gene expression have higher 5-year disease specific survival than EAs, and this difference is statistically significant for *LRBA*

In AAs, *RAB3C* was deleted, had lower expression, and better survival than EAs (Figure 6A). *LRBA* was amplified in AAs, with decreased mRNA expression, and improved survival when compared to EAs (Figure 6B). SCNVs may be driving the *RAB3C* finding, but not the *LRBA* observation. *GSTM1* and *GSTM2* are deleted in EAs, had lower expression, and worse survival relative to AAs (Figure 6C-D). *PTEN* amplifications were significant in EAs, along with higher gene expression, and poor survival (Figure 6E). SCNVs may influence the *GSTM1*, *GSTM2*, and *PTEN* findings. Overall, AAs have better survival with low expression of all five

candidate genes. The *LRBA* observation may be driven by other biological factors, like genetic ancestry.

7. WAA better classified the tumor suppressor gene, *PTEN,* than self-reported race

Self-reported race is a social construct. Whereas genetic ancestry is a biological construct. Exploring the impact of genetic ancestry on candidate gene expression in ccRCC patients may serve as a better classifier than self-reported race. Since most AAs have ~70 percent WAA, patients were categorized into groups of greater than or less than 70 percent WAA. Fifteen self-reported AAs had WAA below 70 percent (Table 8). No self-reported EAs had WAA above 70 percent (Table 8).

RAB3C deletions were positively correlated with lower median gene expression by self-reported race and WAA (Figures 3 and 5). It was significantly lower in AAs. *LRBA* amplifications were negatively correlated with lower median gene expression by self-reported race and no noticeable trend by WAA (Figures 3 and 5). The mRNA expression was significantly lower in AAs. *GSTM1* deletions corresponded with significant lower expression by self-reported race and WAA (Figures 3 and 5). *GSTM2* deletions corresponded with lower expression when using both self-reported race and WAA (Figures 3 and 5). This was only significant by self-reported race. *PTEN* amplifications are associated with high gene expression by both self-reported race and WAA (Figures 3 and 5).This was only significant by WAA.

D. Discussion

1. Role of genes significantly deleted and amplified in self-reported AA ccRCC patients

Overall, the data supported my hypothesis that there are different SCNVs for AAs and EAs with ccRCC. *RAB3C* was significantly deleted, had lower expression, and better survival among the AAs in the TCGA cohort. *RAB3C*, a member of the Ras oncogene family, encodes a

small GTPase which may be involved in vesicle trafficking[7]. Somatic alterations to Ras are commonly found in several types of cancers, with over 40,000 publications on its function and role in cancer[8]. The superfamily consists of at least 150 members and has both upstream and downstream effects[8]. Pathways and outcomes related to Ras include cell cycle progression, cellular growth and migration and apoptosis[8]. Therefore, deletion of *RAB3C* could disrupt regulatory functions of members of the Ras family, with tumor promoting effects. AAs are significantly associated with SCNVs affecting *RAB3C,* suggesting that *RAB3C* could be a population-specific regulator of ccRCC development.

AAs had significant *LRBA* amplifications and lower mRNA expression. Low expressing AAs have significantly better 5-year survival rates than EA low expresssors. MalaCards says *LRBA* has a relationship with Chronic Kidney Disease (CKD)[9]. This relationship is significant because CKD is a common risk factor for kidney cancer development[10]. In fact, kidney cancer risk increases by 10 to 80 percent after dialysis, a treatment for kidney failure[10]. *LRBA*, therefore, could be a tumor-promoting gene in AAs with high expression. *RAB3C* and *LRBA* seem to be potential therapeutic targets because of the survival trends observed.

Role of genes significantly deleted and amplified in self-reported EA ccRCC patients

EAs had significant *GSTM1* and *GSTM2* deletions, lower expression, and worse survival. Both *GSTM1* and *GSTM2* belong to the mu class of glutathione S-Transferases[11,12]. The entire mu class has a range of functions, including the detoxification of electrophilic compounds (e.g. carcinogens, drugs, and environmental toxins). Deletions of *GSTM1* and *GSTM2* could result in increased susceptibility to carcinogens and toxins or change individual drug response[11,12]. Interestingly, lack of function mutations have been associated with many cancers[11,12]. Specifically, *GSTM1* and *GSTM2* deletions have been associated with increased risk of RCC

development[13]. Patients with one deleted copy of *GSTM1* were found to have decreased survival rates[13]. These genes could serve as a screening tool to determine whether or not EA patients are at a higher risk of developing ccRCC.

EAs had significant *PTEN* amplifications, higher expression, and worse survival. *PTEN* has cancer related functions, acting as a tumor suppressor gene by down-regulating the AKT/PKB signaling pathway[14]. While *PTEN* did not have significant expression differences by self- reported race (a social construct), it did have significant variation when grouping patients by WAA (a biological construct). This shows the significance of using genetic ancestry for future studies within the clinic. Decreased *PTEN* expression is linked with poor prognosis and survival[15]. Even though *PTEN* is amplified in many ccRCC patients with low WAA that have higher mRNA-expression, some low expressors may benefit from the increased activity of the gene through AKT/PKB signaling pathway drugs.

Possible targeted therapies for ccRCC patients

RAB3C could serve as a therapeutic target, due to its membership to the Ras family. Because over 40,000 publications exist on the role of Ras in cancer, it serves as a relatively well understood source of cancer progression and existing therapies could be applied to AA ccRCC patients[8]. No current FDA approved drugs exist to target *RAB3C,* but inhibitors of Ras (Cetuximab, or Panitumumab) could be utilized as further research is conducted to strengthen the role of *RAB3C* in ccRCC[7]. *LRBA* seems to be a potential therapeutic target because low expressing AAs have significantly better 5-year survival rates than EAs. As the gene is significantly amplified in AAs, some high expressing ccRCC patients could benefit from a novel targeted LRBA inhibitor. Currently, there are no FDA approved targeted therapies for *LRBA*[9].

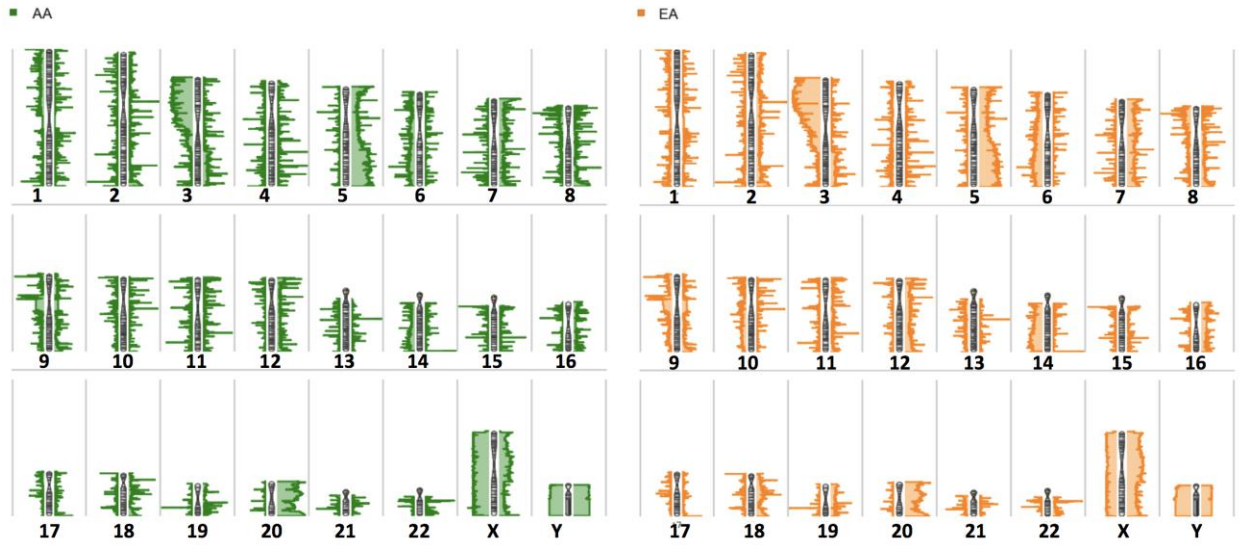
GSTM1 and *GSTM2* deletions in ccRCC patients has been associated with poor survival outcomes. EAs are specifically affected and have greater deletions. A novel therapeutic method could be to supplement GSTM1 protein in an attempt to increase survival rates. Finally, the AKT/PKB signaling pathway activates the mTOR pathway[16]. The PTEN tumor suppressor protein can indirectly down regulate AKT[14], leaving mTOR inactivated. EA patients with low *PTEN* expression and active AKT would benefit from Everolimus treatment, an FDA-approved drug that inhibits the mTOR pathway[14].

E. Tables and Figures

Table 1. Demographic Characteristic of 1000 Genomes Project Study Participants Serving as the Reference Genome			
	AA (ASW, *n* = 99)	**EA (CEU, *n* = 182)**	***P***
Sex (%) ^			*0.71*
Female	53 (54)	93 (51)	
Male	46 (46)	89 (49)	
^ Fisher's exact test			

Table 2. Clinical and Demographic Characteristics of Clear Cell Renal Cell Carcinoma Patients in the TCGA Cohort			
	AA (*n* = 55)	**EA (*n* = 458)**	***P***
Age in Years #			0.62
Mean (SD)	59.9 (10.5)	60.6 (12.2)	
Range	37-81	26-90	
Sex (%) ^			*0.016*
Female	28 (51)	153 (33)	
Male	27 (49)	305 (67)	
Stage (%) +			*0.016*
I	38 (69)	217 (48)	
II	5 (9)	49 (11)	
III	6 (11)	112 (24)	
IV	5 (9)	78 (17)	
Unknown	1 (2)	2 (0)	
Grade (%) +			*0.003*
1	4 (7)	10 (2)	
2	25 (46)	188 (41)	
3	21 (38)	181 (40)	
4	2 (4)	74 (16)	
Unknown	3 (5)	5 (1)	
Vital Status (%) ^ &			*0.033*
Dead	11 (20)	159 (35)	
Alive	44 (80)	296 (65)	
Unknown	0	3 (0)	
^ Fisher's exact test, + Chi square test, # t-test & unknown patients removed from significance testing			

A



B

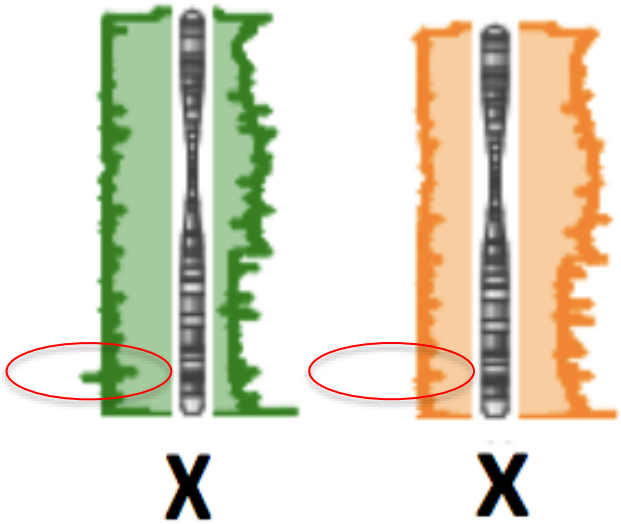


Figure 1. Genome-wide comparison of tumor tissues in AA and EA clear cell renal cell carcinoma patients. Karyograms represent the frequency of SCNVs genome-wide (A) and in select genomic regions (B). Peaks on the left of each chromosome are deletions and on the right are amplifications.

Table 3. List of Significantly Deleted Genes Unique to African American Clear Cell Renal Cell Carcinoma Patients			
Gene Symbol	**Cytoband**	**Average Copy Number**	***P***
ZNF438	10p11.23	1.126	0.00843461
BMS1P5	10q11.22	1.15361	0.00735365
RNA5SP314	10q11.22	1.15385	0.00709993
DMBT1	10q26.13	1.18915	0.00956398
RPL36AP40	11p14.3	1.20472	9.10E-04
RP11-726G1.1	12p13.31	1.14501	9.26E-03
MTCO2P3	13q21.1	1.00931	1.56E-08
HEATR4	14q24.3	1.13329	0.00240574
ACOT1	14q24.3	1.13329	0.00301232
NT5CP2	14q24.3	1.13726	0.00289366
RP11-109E12.1	2q21.1	1.26752	0.0022696
KCNIP4	4p15.2	1.25666	4.90E-06
RPS23P5	5q11.2	1.16163	8.42E-11
RAB3C	5q11.2	1.16163	8.42E-11
SGCD	5q33.2	1.06305	0.00555641

Table 4. List of Significantly Amplified Genes Unique to African American Clear Cell Renal Cell Carcinoma Patients							
Gene Symbol	**Cytoband**	**Average Copy Number**	***P***	**Gene Symbol**	**Cytoband**	**Average Copy Number**	***P***
KIF1B	1p36.22	2.91477	0.00408173	***FNTB***	14q23.3	2.85432	0.00959146
TMEM51	1p36.21	2.78676	0.00838743	***GPX2***	14q23.3	2.77259	0.0080904
GBAP1	1q22	3.13099	2.07E-05	***DPF3***	14q24.2	2.76431	0.00276233
MTX1P1	1q22	3.04791	0.00951446	***RIN3***	14q32.12	2.59579	0.0102744
GBA	1q22	3.04791	0.00951446	***RP11-678G14.2***	19p12	2.48649	0.00293732
SLC9A3P1	10q11.23	3.0243	0.00416477	***RP11-678G14.3***	19p12	2.48347	0.00833659
ASAH2	10q11.23	2.88165	0.000382	***CTC-512J12.6***	19q13.31	2.49028	4.45E-18
SH3PXD2A	10q24.33	3.04602	0.00817994	***ZNF285***	19q13.31	2.49028	4.45E-18
SORCS1	10q25.1	2.8328	0.00832666	***CTC-512J12.4***	19q13.31	2.49028	4.45E-18
VTI1A	10q25.2	2.95285	0.000946818	***ANTXR1***	2p13.3	2.83734	0.000533369
SPRN	10q26.3	2.94699	4.76E-06	***KIF5C***	2q23.1	2.94036	0.00671968
CYP2E1	10q26.3	2.96082	7.43E-06	***PI4KA***	22q11.21	2.50569	0.00952595
SYCE1	10q26.3	2.94142	4.29E-05	***AP000345.1***	22q11.23	2.50091	0.00185458
OR51A2	11p15.4	3.42472	0.00316565	***TTC38***	22q13.31	2.74281	0.0093591
OR5F1	11q12.1	2.906	2.89E-09	***PRSS50***	3p21.31	2.78134	0.00136473
RP11-150C16.1	12q14.1	3.01895	0.0113566	***PRSS45***	3p21.31	2.78134	0.00136473
RPS6P22	12q14.1	3.01895	0.0113566	***RP11-427P5.2***	3p21.31	2.79946	0.000557536
WDR66	12q24.31	2.67687	0.00142616	***DAG1***	3p21.31	2.60363	0.0061434
MTCO2P3	13q21.1	2.87895	1.56E-08	***LRBA***	4q31.3	2.72572	5.27E-10
CDC16	13q34	2.53817	0.0101161	***TRERF1***	6p21.1	3.02237	3.31E-14
MIR4502	13q34	2.53817	0.0092995	***RP11-135M8_A.1***	6q13	2.86612	0.00831128
AE000658.25	14q11.2	2.75241	0.000723884	***RP3-428L16.1***	6q26	2.73618	0.0122058
RP11-58E21.5	14q21.3	2.72587	0.0033793	***RPS6KA2***	6q27	2.73656	0.00209357
RP11-58E21.7	14q21.3	2.72587	0.0033793	***Z98049.1***	6q27	2.73656	0.00209357

RP11-58E21.1	14q21.3	2.72587	0.0033793	***RP11-514O12.4***	6q27	2.79386	0.00209357
ESR2	14q23.2	2.93184	0.0128171	***RPS6KA2-AS1***	6q27	2.79386	0.00209357
CHURC1	14q23.3	2.85432	0.00959146	***PDGFRL***	8p22	2.75108	0.000158473
CHURC1-FNTB	14q23.3	2.85432	0.00959146	***CHCHD4P2***	9q31.2	2.71456	0.00277584

Table 5. List of Significantly Deleted Genes Unique to European American Clear Cell Renal Cell Carcinoma Patients			
Gene Symbol	**Cytoband**	**Average Copy Number**	***P***
C1orf63	1p36.11	0.806918	0.00937989
RP3-465N24.5	1p36.11	0.806918	0.00937989
GSTM2	1p13.3	0.920208	0.000885782
GSTM1	1p13.3	0.920208	0.000885782
AC000032.2	1p13.3	0.923653	0.000761057
NME7	1q24.2	0.90592	0.0126005
AC073218.1	2p22.3	1.00833	0.00898557
CTD-2231H16.1	5p15.33	1.40116	1.34E-11
CTD-2593A12.3	5q35.3	0.965996	0.0115048
CTD-2593A12.2	5q35.3	0.967738	0.00973364
AC000370.2	7q31.33	0.862825	0.0109036
MGAM	7q34	0.977222	0.0121328

Table 6. List of Significantly Amplified Genes Unique to European American Clear Cell Renal Cell Carcinoma Patients							
Gene Symbol	**Cytoband**	**Average Copy Number**	***P***	**Gene Symbol**	**Cytoband**	**Average Copy Number**	***P***
ZRANB2-AS2	1p31.1	3.13653	0.000812063	***RP11-488C13.6***	14q24.3	3.33063	0.00288084
ZBTB7B	1q21.3	3.0049	0.0129286	***RP11-488C13.7***	14q24.3	3.33063	0.00288084
DCST2	1q21.3	3.0049	0.0129286	***RP11-488C13.5***	14q24.3	3.33063	0.00288084
RP11-222A11.1	10q21.3	3.0959	0.0112812	***ANGEL1***	14q24.3	3.53505	0.00159956
PTEN	10q23.31	3.02381	0.00504405	***C14orf166B***	14q24.3	3.46017	0.00240898
STK24	13q32.2	3.10729	0.0113893	***RP11-488C13.1***	14q24.3	3.46017	0.00265441
SMOC1	14q24.2	3.03567	0.0116836	***RN7SKP17***	14q24.3	3.46017	0.00265441
RP3-414A15.11	14q24.3	3.33231	0.0124301	***RP11-7F17.7***	14q24.3	3.46017	0.00292248
ACOT2	14q24.3	3.91875	0.00362493	***RN7SL356P***	14q24.3	3.46017	0.00292248
NT5CP1	14q24.3	3.91875	0.00362493	***RP11-7F17.1***	14q24.3	3.46017	0.00292248
ACOT4	14q24.3	3.49486	0.00578728	***IRF2BPL***	14q24.3	3.46017	0.00292248
ACOT6	14q24.3	3.49486	0.00578728	***RP11-7F17.5***	14q24.3	3.46017	0.00292248
RP3-414A15.10	14q24.3	3.49486	0.00578728	***RP11-7F17.4***	14q24.3	3.46017	0.00292248
NDUFB8P1	14q24.3	3.49486	0.00578728	***RP11-7F17.3***	14q24.3	3.46017	0.00292248
DNAL1	14q24.3	3.49486	0.00578728	***CIPC***	14q24.3	3.46017	0.00292248
RNU6-240P	14q24.3	3.49486	0.00578728	***RP11-463C8.4***	14q24.3	3.46017	0.00292248
RP4-693M11.3	14q24.3	3.49486	0.00578728	***TMEM63C***	14q24.3	3.46017	0.00292248
PNMA1	14q24.3	3.49486	0.00578728	***ZDHHC22***	14q24.3	3.46017	0.00292248
ELMSAN1	14q24.3	3.49486	0.00578728	***AC007375.1***	14q24.3	3.46017	0.00292248
YLPM1	14q24.3	3.24438	0.00929411	***RP11-463C8.5***	14q24.3	3.46017	0.00292248
RP11-316E14.6	14q24.3	3.24438	0.00929411	***SNORA32***	14q24.3	3.46017	0.00292248
DLST	14q24.3	3.25906	0.00938947	***NGB***	14q24.3	3.46017	0.00292248
RPS6KL1	14q24.3	3.25906	0.00938947	***MIR1260A***	14q24.3	3.45007	0.00288084
PGF	14q24.3	3.25906	0.00938947	***POMT2***	14q24.3	3.46017	0.00292248
EIF2B2	14q24.3	3.25906	0.00938947	***GSTZ1***	14q24.3	3.46017	0.00292248

RP11-950C14.3	14q24.3	3.25906	0.00938947	***TMED8***	14q24.3	3.46017	0.00292248
MLH3	14q24.3	3.25906	0.00938947	***RN7SL137P***	14q24.3	3.46017	0.00292248
RNU6-689P	14q24.3	3.25906	0.00938947	***RP11-493G17.4***	14q24.3	3.46017	0.00292248
ACYP1	14q24.3	3.25906	0.00938947	***SAMD15***	14q24.3	3.46017	0.00292248
ZC2HC1C	14q24.3	3.25906	0.00938947	***NOXRED1***	14q24.3	3.46017	0.00292248
NEK9	14q24.3	3.25906	0.00938947	***FKSG61***	14q24.3	3.46017	0.00265441
HIF1AP1	14q24.3	3.25906	0.00938947	***VIPAS39***	14q24.3	3.46017	0.00292248
RP11-950C14.7	14q24.3	3.08341	0.00900484	***AHSA1***	14q24.3	3.46017	0.00292248
TMED10	14q24.3	3.08341	0.00900484	***SNORA46***	14q24.3	3.46017	0.00292248
RNU4ATAC14P	14q24.3	3.08341	0.00983602	***ISM2***	14q24.3	3.46017	0.00265441
RP11-293M10.1	14q24.3	4.31838	0.00938947	***SPTLC2***	14q24.3	3.46017	0.00292248
RP11-293M10.2	14q24.3	3.45807	0.00860053	***RN7SL587P***	14q24.3	3.46017	0.00265441
FOS	14q24.3	3.45807	0.00860053	***COX6CP11***	14q24.3	3.46017	0.00321504
RP11-293M10.4	14q24.3	3.45807	0.00860053	***ALKBH1***	14q24.3	3.62921	0.0100171
RP11-293M10.5	14q24.3	3.45807	0.00860053	***RPL21P10***	14q24.3	3.45033	0.00240898
RP11-293M10.6	14q24.3	3.34343	0.00248271	***ZMYND19P1***	14q24.3	3.45033	0.00353406
JDP2	14q24.3	3.45807	0.00938947	***SLIRP***	14q24.3	3.45033	0.00353406
BATF	14q24.3	3.45807	0.00938947	***SNW1***	14q24.3	3.45033	0.00353406
AC007182.6	14q24.3	3.45807	0.00938947	***C14orf178***	14q24.3	3.45033	0.00353406
FLVCR2	14q24.3	3.45807	0.00938947	***AC008372.1***	14q24.3	3.45033	0.00353406
RP11-507E23.1	14q24.3	3.45807	0.00938947	***ADCK1***	14q24.3	3.45033	0.00388166
RNA5SP387	14q24.3	3.45807	0.00938947	***Y_RNA***	14q24.3	3.45033	0.00353406
TTLL5	14q24.3	3.45807	0.00657642	***FRDAP***	14q24.3	3.45033	0.00353406
C14orf1	14q24.3	3.45807	0.00657642	***RNA5SP388***	14q24.3	3.40724	0.0031714
RP11-270M14.4	14q24.3	3.45807	0.0102423	***RP11-332E19.1***	14q24.3	3.2988	0.00237073
RP11-270M14.1	14q24.3	3.45807	0.0102423	***NRXN3***	14q24.3	3.2988	0.00237073
IFT43	14q24.3	3.45807	0.00487443	***RP11-332E19.2***	14q24.3	3.45033	0.00321504
TGFB3	14q24.3	3.45807	0.00487443	***TBC1D5***	3p24.3	3.00537	0.00256871

RP11-270M14.5	14q24.3	3.45807	0.00487443	***AC132807.1***	3p24.3	3.00537	0.00309774
RP11-98L12.2	14q24.3	3.45807	0.00487443	***RARB***	3p24.2	3.04637	0.00010681
GPATCH2L	14q24.3	3.45931	0.0053444	***MIR1269A***	4q13.2	3.02069	0.00745185
RP11-361H10.3	14q24.3	3.45931	0.0053444	***CTD-2593A12.3***	5q35.3	2.97165	0.00973364
RP11-516J2.1	14q24.3	3.45931	0.00197926	***CTD-2593A12.2***	5q35.3	2.97165	0.00973364
RN7SL747P	14q24.3	3.46464	0.00292248	***HLA-DRB6***	6p21.32	3.21283	0.00533388
RP11-187O7.3	14q24.3	3.46464	0.00240898	***HLA-DRB1***	6p21.32	3.20368	0.00305538
CYCSP1	14q24.3	3.46464	0.00265441	***HLA-DQA1***	6p21.32	3.02247	0.0071755
RP11-187O7.1	14q24.3	3.46464	0.00265441	***BCKDHB***	6q14.1	2.89933	1.92E-08
RP11-99E15.3	14q24.3	3.46464	0.00265441	***C6orf118***	6q27	3.16418	0.0114448
RP11-99E15.2	14q24.3	3.46464	0.00265441	***AMZ1***	7p22.3	3.05136	0.0105339
VASH1	14q24.3	3.33063	0.00288084	***PTCH1***	9q22.32	3.05746	0.00533217

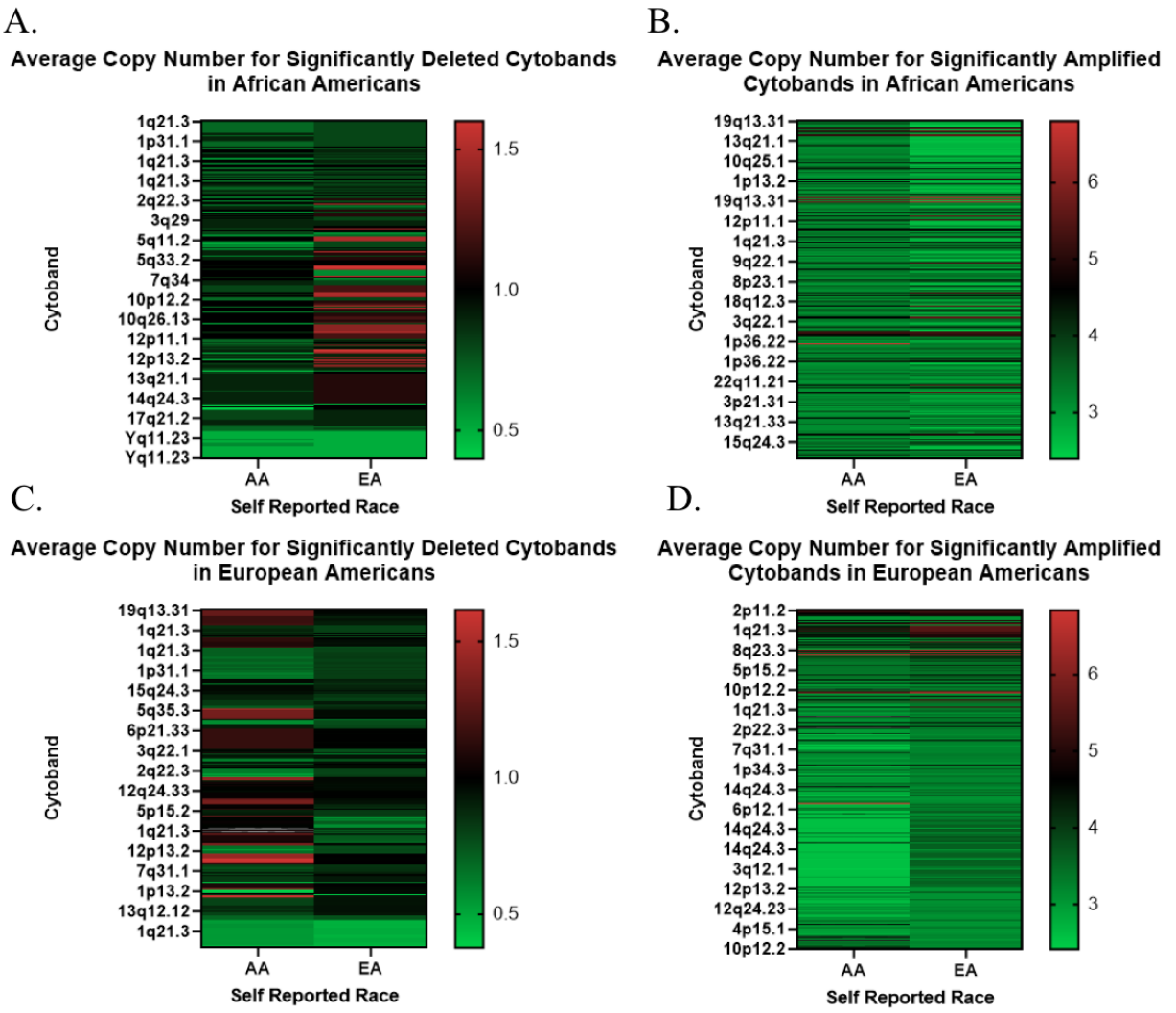


Figure 2. Chromosome region analysis of tumor tissues in AA and EA clear cell renal cell carcinoma patients. Heatmap showing average copy number of significant differential deletions (A) and amplifications (B) in AAs. Heatmap showing average copy number of significant differential deletions (C) and amplifications (D) in EAs.

Table 7. Population-specific Unique Gene SCNVs with Potential Clinical Relevance				
Gene Symbol	**Cytoband**	**Gene Function**	**Disease Association (MIFTS Score)**	***P***
RAB3C	5q11.2	Ras oncogene family member encodes a small GTPase	ovarian cancer (88) breast cancer (97) pancreatic cancer (86) cervical cancer (72)	< 0.05
LRBA	4q31.3	WDL-BEACH-WD (WBW) gene family leads intracellular vesicles to activated receptor complexes; aids in the secretion of immune effector molecules	chronic kidney disease (74) end stage renal disease (54) renal cell carcinoma, nonpapillary (79) adenocarcinoma (63) breast cancer (97)	< 0.01
GSTM1	1p13.3	detoxification of electrophilic compounds (i.e. carcinogens), therapeutic drugs, environmental toxins and products of oxidative stress	chronic kidney disease (74) kidney cancer (60) kidney disease (72) end stage renal disease (54) renal cell carcinoma, nonpapillary (79) clear cell renal cell carcinoma (54)	< 0.001
GSTM2	1p13.3	detoxification of electrophilic compounds (i.e. carcinogens), therapeutic drugs, environmental toxins and products of oxidative stress	adenocarcinoma (63) colon adenocarcinoma (64) bladder cancer (79) prostate cancer (95) colorectal cancer (100)	< 0.05
PTEN	10q23.31	tumor suppressor negatively regulates AKT/PKB signaling pathway	renal cell carcinoma, nonpapillary (79) kidney cancer (60) renal cell carcinoma, nonpapillary (79) clear cell renal cell carcinoma (54) renal cell carcinoma, papillary, 1(79) sarcomatoid renal Cell carcinoma (41) multilocular clear cell renal cell carcinoma (32) chromophil renal cell carcinoma (23) adenocarcinoma (63) kidney disease (72) chronic kidney disease (74) cystic kidney disease (52)	0.221

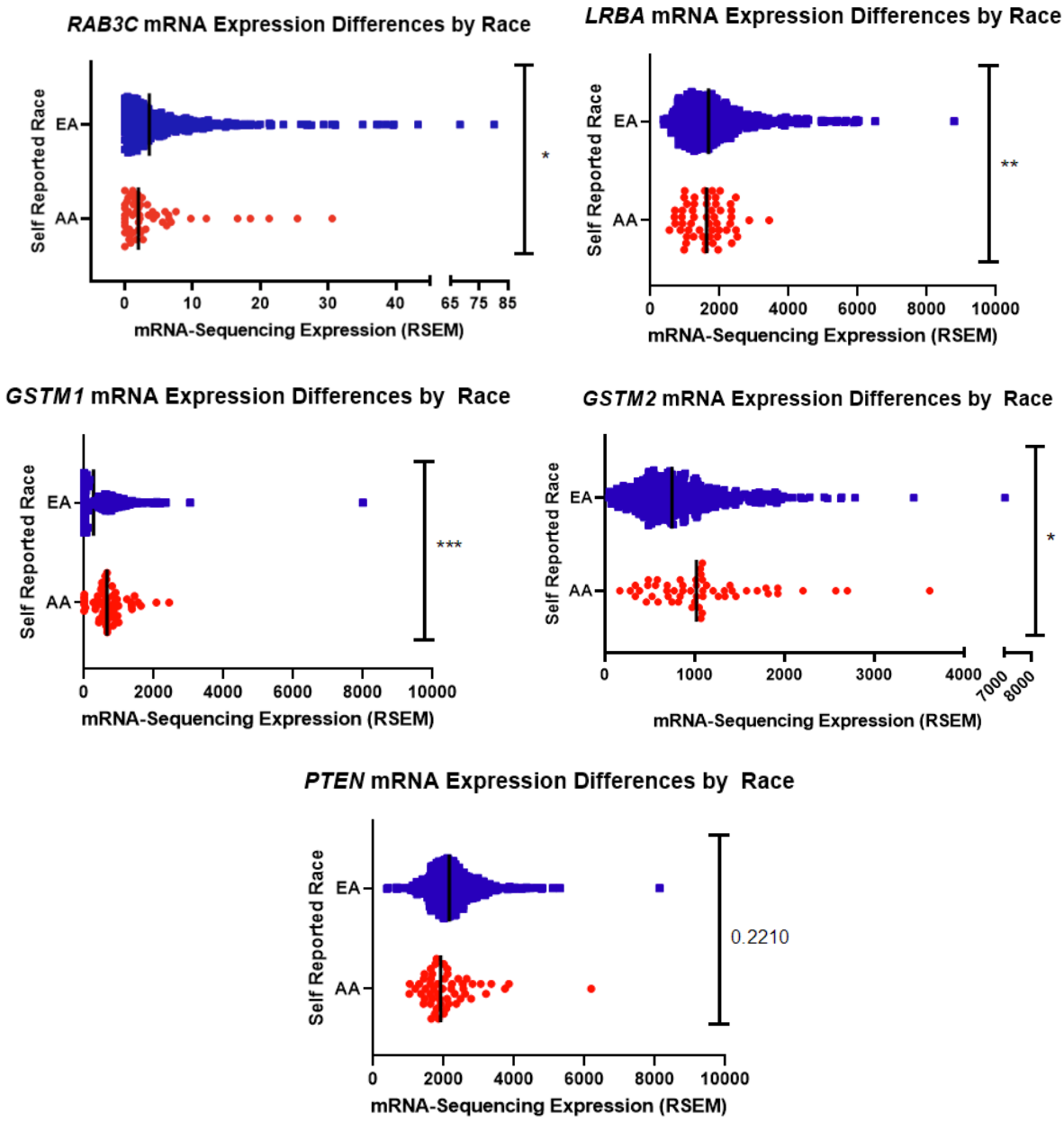
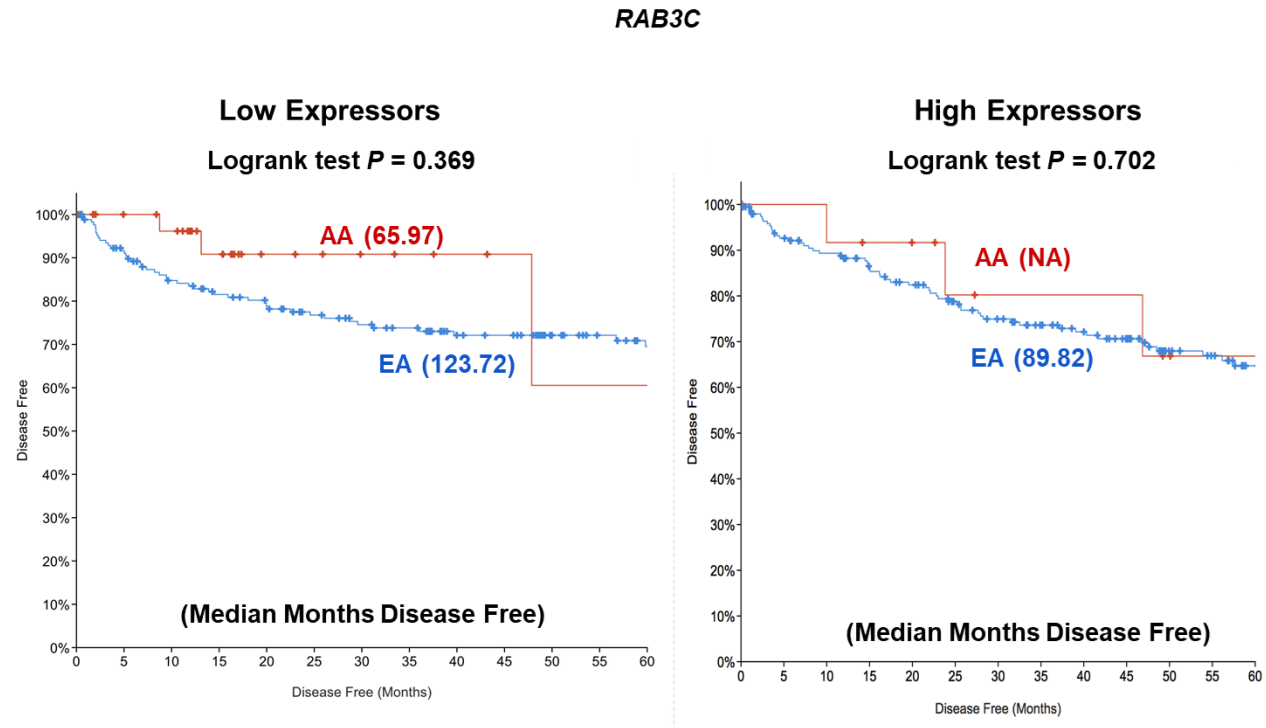
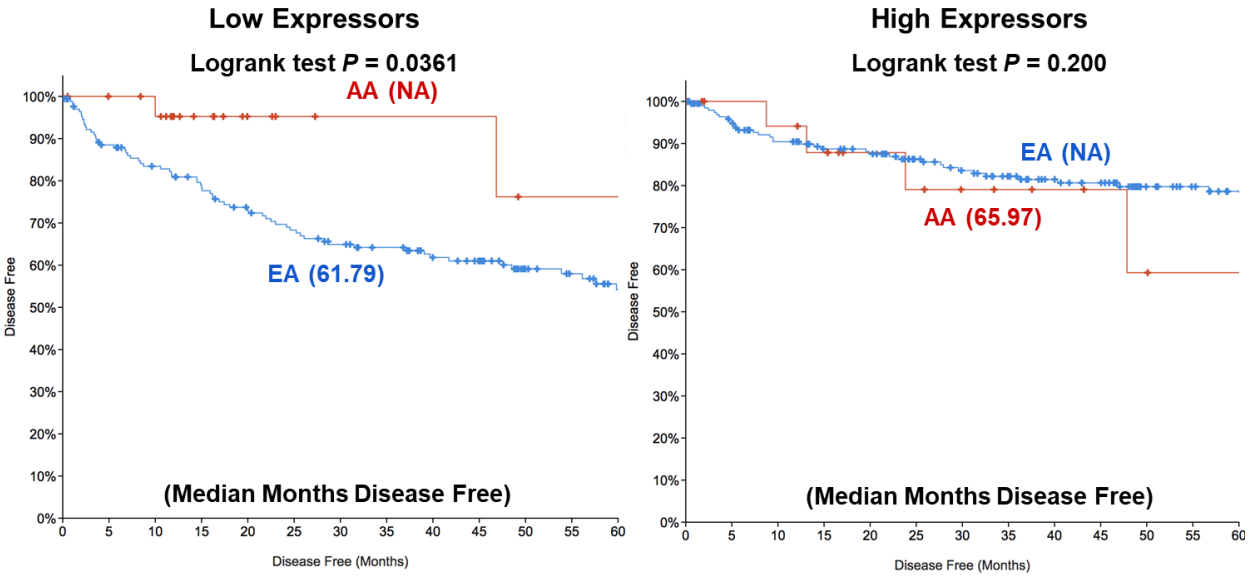


Figure 3. Integrated mRNA expression for unique candidate genes by self-reported race. Scatter plots of *RAB3C, LRBA, GSTM1*, *GSTM2*, and *PTEN* expression patterns in AA and EA ccRCC patients. Tested for significance using unpaired parametric t-tests with Welch corrections. * = <0.05, ** = < 0.01, *** = < 0.001

A

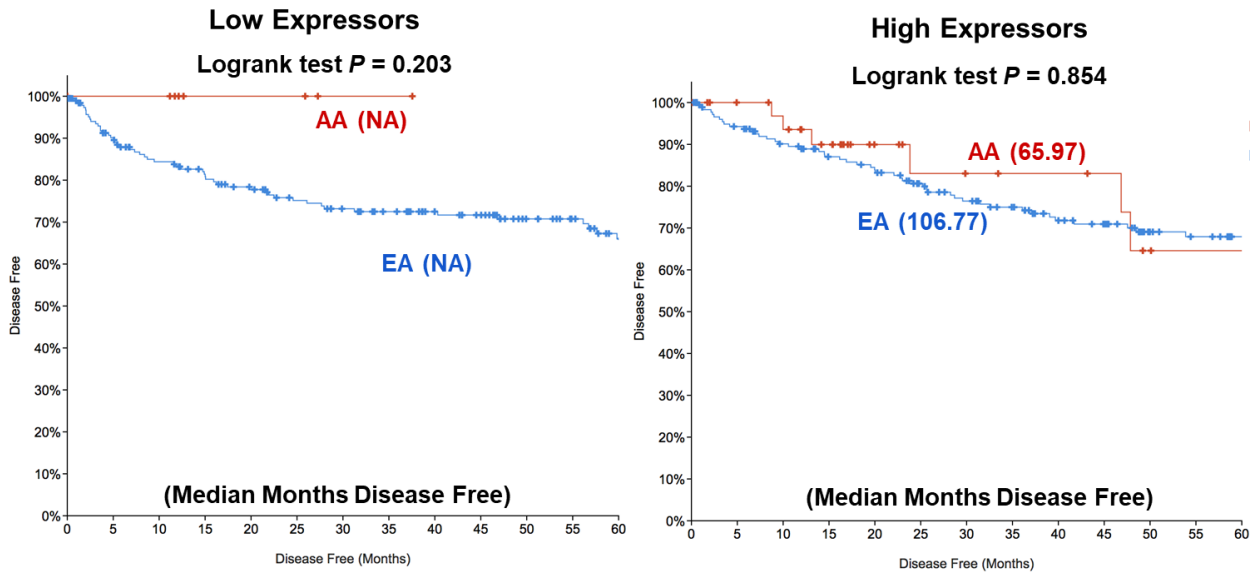

RAB3C
Low Expressors
Logrank test P = 0.369
AA (65.97)
EA (123.72)
(Median Months Disease Free)
Disease Free
Disease Free (Months)
High Expressors
Logrank test P = 0.702
AA (NA)
EA (89.82)
(Median Months Disease Free)
Disease Free
Disease Free (Months)

B


LRBA
Low Expressors
Logrank test P = 0.0361
AA (NA)
EA (61.79)
(Median Months Disease Free)
Disease Free
Disease Free (Months)
High Expressors
Logrank test P = 0.200
EA (NA)
AA (65.97)
(Median Months Disease Free)
Disease Free
Disease Free (Months)

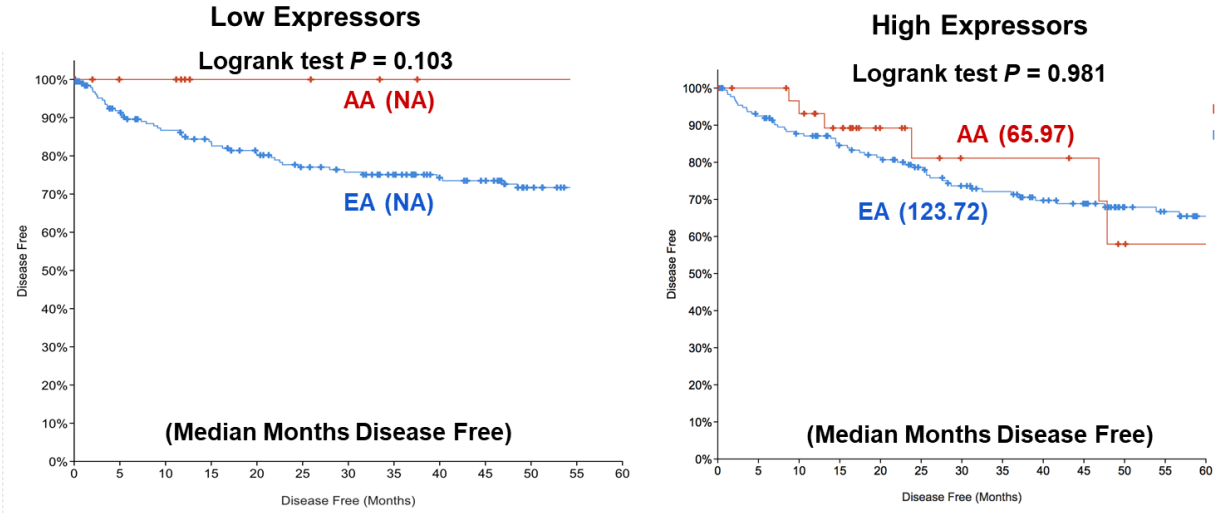
C

GSTM1



D

GSTM2



E

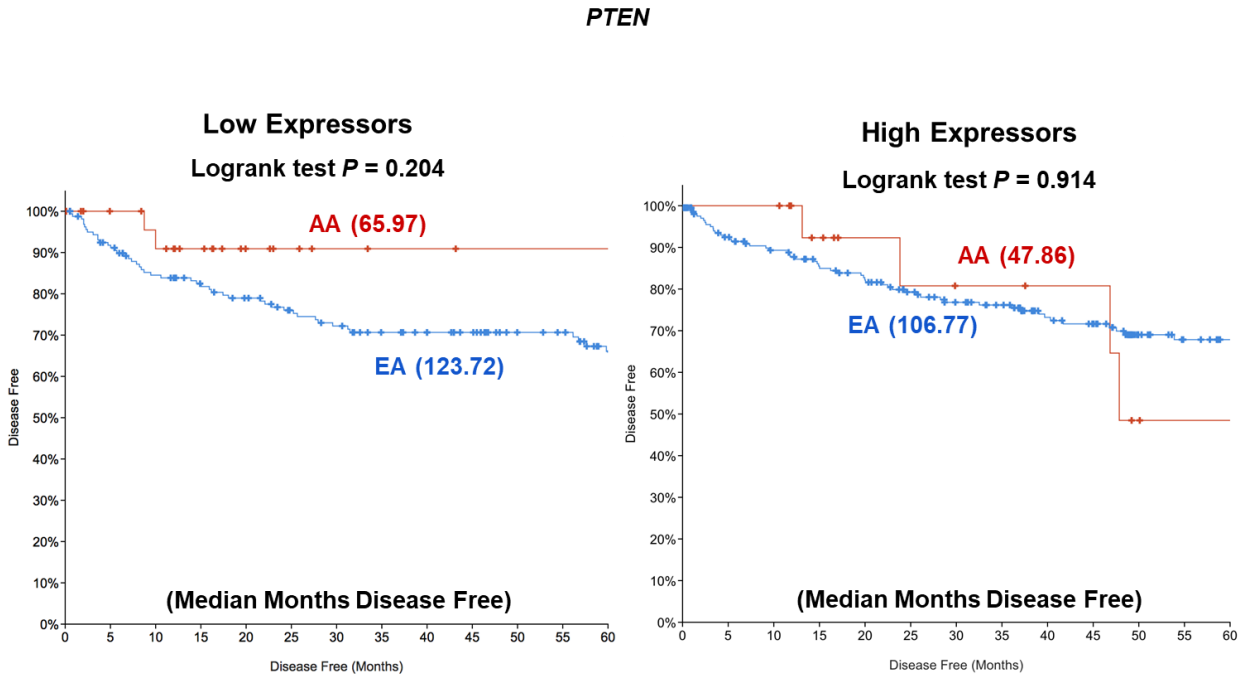


Figure 4. Five-year disease specific survival by self-reported race based on mRNA expression of five candidate genes. Kaplan Meier survival curves based on low and high AA and EA expressors for (A) *RAB3C*, (B) *LRBA*, (C) *GSTM1*, (D) *GSTM2*, and (E) *PTEN*.

Table 8. Stratification of Clear Cell Renal Cell Carcinoma By WAA		
	Self-Reported African Americans (*n* = 55)	**Self-Reported European Americans (*n* = 458)**
Range % WAA	0.22-93.39	0.1-60.28
Range % EURA	3.41-97.91	33.59-99.41
Mean % WAA/ EURA	73.72/ 23.35	1.53/ 94.68
Med % WAA/ EURA	78.63/ 19.14	0.49/ 97.73
< 70% WAA (%)	15 (27.3)	458 (100)
> 70% WAA (%)	40 (72.7)	0 (0)
Med = Median	AA = West African Ancestry	EURA = European Ancestry

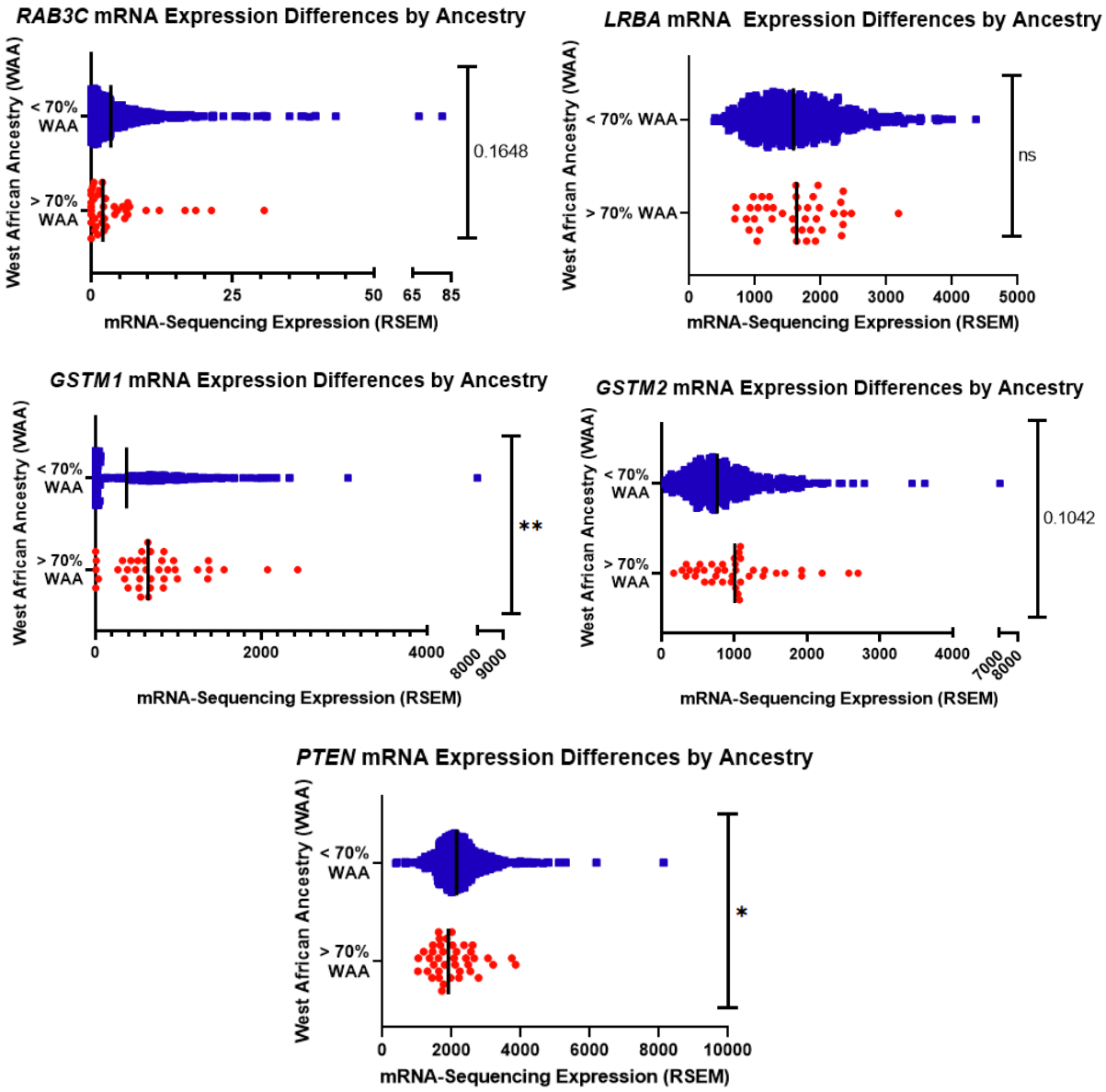


Figure 5. Integrated mRNA expression for unique candidate genes by WAA. Scatter plots of *RAB3C*, *LRBA*, *GSTM1*, *GSTM2*, and *PTEN* expression patterns in AA and EA ccRCC patients. Tested for significance using unpaired parametric t-tests with Welch corrections. * = <0.05, ** = < 0.01

F. References

1. McElroy, J. P., Nelson, M. R., Caillier, S. J. & Oksenberg, J. R. Copy number variation in African Americans. *BMC Genet*10, 15 (2009).
2. Teh, B. T., Farber, L. J. & Furge, K. Molecular Characterization of Renal Cell Carcinoma. in *Renal Cell Carcinoma*(eds. Figlin, R. A., Rathmell, W. K. & Rini, B. I.) 91–111 (Springer US, 2012). doi:10.1007/978-1-4614-2400-0_5.
3. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*499, 43–49 (2013).
4. Massari, F. *et al.*Toward a genome-based treatment landscape for renal cell carcinoma. *Crit. Rev. Oncol. Hematol.*142, 141–152 (2019).
5. Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D. & Mountain, J. L. The Genetic Ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet*96, 37–53 (2015).
6. Tishkoff, S. A. *et al.*The Genetic Structure and History of Africans and African Americans. *Science*324, 1035–1044 (2009).
7. *RAB3C* Gene (Protein Coding). *Gene Cards: The Human Gene Database*, (2020).
8. Fernández-Medarde, A. & Santos, E. Ras in Cancer and Developmental Diseases. *Genes Cancer*2, 344–358 (2011).
9. LRBA. MalaCards: Human Disease Database, Gene Cards Suite, (2020).
10. Stengel, B. Chronic kidney disease and cancer: a troubling connection. *J Nephrol*23, 253–262 (2010).
11. *GSTM1* Gene (Protein Coding). *Gene Cards: The Human Gene Database*, (2020).
12. *GSTM2* Gene (Protein Coding). *Gene Cards: The Human Gene Database*, (2020).
13. Coric, V.M., Simic, T.P., Pekmezovic, T.D., Basta-Jovanovic, G.M., Savic-Radojevic, A.R., Radojevic-Skodric, S.M., Matic, M.G., Suvakov, S.R., Dragicevic, D.P., Radic, T.M., Dzamic, Z.M. & Pljesa-Ercegovac, M.S. GSTM1 genotype is an independent prognostic factor in clear cell renal cell carcinoma *Urologic Oncology35, 409-416 (2017).*
14. *PTEN* Gene (Protein Coding). *Gene Cards: The Human Gene Database*, (2020).
15. Que, Wan-cai, Hong-qiang Qiu, Yu Cheng, Mao-bai Liu, & Chao-yang Wu. PTEN in Kidney Cancer: A Review and Meta-Analysis. *Clinica Chimica Acta*480, (2018).
16. Memmott, R. M. & Dennis, P. A. Akt-dependent and independent mechanisms of mTOR regulation in cancer. *Cell Signal*21, 656–664 (2009).

V. Chapter 2: A pan-renal cell carcinoma analysis of genetic ancestry-associated somatic copy number variation events in African Americans and European Americans

A. Introduction

1. RCC incidence, mortality and survival

Renal Cell Carcinoma (RCC) is the sixth and tenth most common cancer in males and females in the US, respectively[1]. In 2018, there were approximately 65,340 diagnoses and 14,970 deaths due to RCC[1,2]. Between the years 1992 and 2015, incidence was 11.281 per 100,000 people, with males, AAs, and people older than 65 being affected most[1]. Over the same time period, diagnosis rates increased at 2.421 percent per year, until 2008, where rates settled until 2015[1]. The RCC incidence-based mortality rate between 1992 and 2015 was 5.256 per 100,000 people per year, though rates decreased by 2.159 percent per year[1] after increasing and peaking in 2001[1]. Populations most affected by high mortality rates were males, Native Americans/Alaska natives, and patients older than 65[1].

In juxtaposition, the 5-year survival rate in the US between 2011 and 2017 was 76.9 percent[3], making RCC the deadliest urologic cancer[4]. Disease specific survival is dependent on stage at diagnosis and histology. RCC survival rates are 93 percent for localized disease, 74.5 percent for regional disease, and 14.3 for distant disease[3]. By histology, rates are between 55 and 60 percent for clear cell RCC (ccRCC), 80 to 90 percent for papillary RCC (pRCC), and 90 percent for chromophobe RCC (chRCC)[5].

2. RCC biology and molecular features

RCC comprises 85 percent of all primary neoplasms[2]. RCC consists of a heterogeneous set of subtypes, all of which originate from renal tubular epithelial cells[2] and exhibit distinct morphology, genetic changes, and clinical behavior[6]. The three most common subtypes are

ccRCC, pRCC, chRCC. Other rare forms of kidney cancer include transitional cell carcinoma, nephroblastoma, duct carcinoma, renal medullary carcinoma, and urothelial carcinomas[2, 7]. Though RCC can be used to broadly characterize the three major subtypes, distinctions exist which are important for accurate diagnosis and treatment. ccRCC and pRCC both emerge from cells in the proximal convoluted tubules of nephrons, but have differing genetic profiles, while chRCC is derived from intercalated cells in the distal convoluted tubules[5]. However, there may be overlapping morphological structures; clear cells may be present in pRCC and papillary structures may be present in other RCC subtypes[6]. To correct for this, a hybrid subtype is now recognized by the World Health Organization: clear cell papillary RCC[6].

DNA sequencing has been used to study RCC, allowing for the identification of mutated genes across all histologies and unique characteristics to each subtype. The three major RCC subtypes have been associated with mutations to tumor suppressor genes *TP53, PTEN,* and *CDKN2A*[8]. There are known chromosomal abnormalities distinctively associated with each RCC subtype across the genome, helping to distinguish between the various histologies. A total of 91 percent of ccRCC patients have a deletion of chromosome 3p, 45 percent have 14q deletions, and 67 percent have an amplification of 5q[9]. Loss of chromosome 3p results in somatic copy number variation (SCNV) changes to four genes, most commonly associated with ccRCC: *VHL, PBRM1, BAP1,* and *SETD2* and deletion of 14q results in loss of *HIF1A,* which has been linked with aggressive disease[9].

Interestingly, there are overlapping chromosomal regions affected by SCNVs in ccRCC and pRCC patients. Types I and II pRCC have distinctive SCNV profiles. Type I is associated with amplification of chromosome 7 (75-80 percent), 16 (60 percent), and 17 (80 percent)[10]. Type II has a loss of chromosome 3p, 14p, 9p and 22q at a rate of 20 percent and amplifications

of 5q (20 percent), 7 through 16 (30 percent) and 17 (20 percent)[10]. Chromosomes 3p and 5q are deleted and amplified in both histologies. pRCC is associated with changes to genes, including *MET, SETD2, NF2, KDM6A,*and *SMARCB1*[6].

The chRCC subtype is associated with the loss of entire chromosomes 1, 2, 6, 10, 13, and 17 at a rate of 86 percent[11]. Additionally, loss of chromosomes 3, 5, 8, 9, 11, 18, and 21 are seen at a rate of 12 to 58 percent[11]. Genes affected include *TP53* and *PTEN,* which are altered in 30 and 10 percent of cases, respectively[6]. Additionally, changes to *MTOR, NRAS* and *TSC1*or *TSC2* were found in fewer than 5 percent of cases[6]. Interestingly, structural rearrangements in the promoter region of *TERT* have also been related with 10 percent of chRCC cases[6].

3. Need for pan-RCC treatment

RCC histologies are diagnosed using computed tomography, magnetic resonance imaging, or a microscopic examination of Hematoxylin and Eosin stained biopsy slides[5]. Current treatment options based on RCC subtype include surgery, radiation, chemotherapy, immunotherapy, and targeted therapies[12]. Though ccRCC, pRCC, and chRCC have overlapping treatment methods, targeted agents are especially effective in advanced and metastatic ccRCC patients[13]. ccRCC patients with *VHL* mutations, the most commonly mutated gene (52-82%), respond well to Sunitinib and Sorafenib treatment[13]. Interestingly, *VHL* SCNVs (particularly deletions) occur frequently in ccRCC and Type II pRCC tumors[9, 10]. Dysregulated *VHL* causes increased angiogenesis and cell proliferation, by influencing factors such as vascular endothelial growth factor (VEGF) and platelet-derived growth factor (PDGFβ)[13]. Sunitinib and Sorafenib are FDA-approved VEGF and PDGFβ inhibitors for all RCC histologies, however, are less effective in pRCC and chRCC patients. In addition to *VHL* mutation and SCNV changes, there are other overlapping morphological and molecular features across RCC subtypes (see section V.A.2).

4. RCC disparities in AAs and EAs

Genetic ancestry is a molecular measure that explains biological variation in populations better than socially constructed self-reported race. African ancestry, particularly West African ancestry (WAA) has been connected to higher mortality and lower survival in AA cancer patients[14, 15, 16, 17]. Recent pan-cancer work has also shown genetic ancestry is associated with cancer drug resistance[15] and drug metabolism[18]. Exploring genetic ancestry as a potential biological driver of RCC health disparities is an important emerging area of investigation, particularly in a population with high genetic admixture like AAs[19]. Few studies have identified pan-RCC SCNVs[20], with corresponding gene expression profiles, or looked for risk factors by genetic ancestry[21]. This benefit is two-fold: 1) it could allow for more effective treatment across all RCC patients and 2) reveal population-specific pan-RCC biology that can explain RCC outcome differences in AAs and EAs.

5. Hypothesis .

I predict patients with high WAA have distinct SCNV changes that correlate with more aggressive RCC biology and patient survival.

B. Methods

1. Patient clinico-demographics

Patient clinical and demographic data was acquired from cBioPortal (https://www.cbioportal.org) for 513 ccRCC, 257 pRCC, and 62 chRCC patients. Statistical tests were used to identify confounding variables between individuals with high and low WAA across RCC histologies.

2. Building a diverse human reference genome

The International Genome Sample Resource was used to download 1000 Genomes Project copy number data for AA (*n* = 99 ASW) and EA (*n* = 182 CEU) study participants (https://www.internationalgenome.org/category/affy/). ASW refers to individuals of African ancestry that live in the Southwestern United States, whereas CEU individuals are European Americans, specifically Utah residents of the US, with Northern and Western European ancestry. All DNAs were isolated from human lymphocytes. The Coriell Institute for Medical Research conducted Affymetrix Genome-Wide Human SNP 6.0 genotyping arrays as part of the NHGRI Cell line catalog (https://www.internationalgenome.org/category/affy/). PGS 7.0 Batch Effect Removal Tool was utilized to normalize variability due to array scan date and the Create Copy Number Baseline Tool was used to create the .CNMODEL reference file.

3. Differential SCNV analysis by WAA

Controlled Affymetrix Genome-Wide Human SNP 6.0 genotyping array data was downloaded from the National Cancer Institute Genomics Data Commons for self-identified AA and EA ccRCC (*n* = 513 individuals, High WAA *n* = 40, Low WAA *n* = 473), pRCC (*n* = 257 individuals, High WAA *n* = 54, Low WAA *n* = 203), and chRCC (*n* = 62 individuals, High WAA *n* = 4, Low WAA *n* = 58) patients (https://gdc.cancer.gov/access-data). All DNAs were isolated from primary tumors for each patient. Duplicate samples were removed. Array intensity variability was normalized for ccRCC (according to array scan date and sex), pRCC (according to array scan date, age, and stage), and chRCC (according to array scan date, age, and vital status) patients using the PGS 7.0 Batch Effect Removal Tool. The PGS Copy Number Workflow detected amplifications and deletions in comparison to the diverse human reference genome using the Segmentation Algorithm (Segmentation parameters, Minimum genomic

markers: 10, P-value threshold: 0.001, Signal to noise: 0.3; Region Report: Use gender information (Specify Gender), Diploid copy number from: 1.7 to 2.3). Differential genomic segmentation analyses by >70% WAA were analyzed for both amplifications (≥ 2.3 copies, *P* + FDR <0.05) and deletions (≤ 1.7 copy, P + FDR <0.05). Completed Gene Set Analysis on PGS using the following parameters (use Fisher's Exact test, invoke gene ontology browser on the result, restrict analysis to functional groups with more than 5 genes). The GO enrichment analysis used the most recent HGNC nomenclature. Finally, overlapping genes from the copy number segments with significant differences by WAA were identified.

4. Differential SCNV analysis by WAA

Downloaded mRNA-sequencing expression data (RSEM) from cBioPortal (https://www.cbioportal.org) for all candidate genes. PGS 7.0 Batch Effect Removal Tool was utilized to normalize variability due to potential confounding clinical and demographic variables. Overlapping gene expression profiles were compared for patients of ccRCC and pRCC with high and low WAA. Simple linear regression models were run using Graphpad Prism 8 identifying a relationship between ccRCC candidate gene expression and ancestry. A differential expression analysis was conducted on pRCC candidate genes. The resultant three genes with the greatest significance were tested via simple linear regression models.

5. Disease-specific survival based on median gene expression

Graphpad Prism 8 was utilized to graph Kaplan Meier 5-year disease specific survival outcomes for RCC patients with high or low WAA based on expression profiles of significant candidate genes with correlating gene expression. Clinical information was retrieved from cBioPortal (https://www.cbioportal.org) for 513 ccRCC, 257 pRCC, and 62 chRCC patients. Patients were divided into groups of high or low expressors based on median gene expression.

C. Results

1. There were no significant differences by sex in the diverse human reference genome study participants

Most human biology studies use the human reference genome. About 70 percent of the reference genome comes from one EA male, 23% comes from 10 other EA samples, and 7% from over 50 sources[22]. A recent study has proven that individuals with greater amounts of African ancestry can vary widely from the human reference genome[23]. For this reason, a more diverse human reference genome based on the 1000 Genomes Project was used. Of the 281 participants, 99 self-reported as African American (ASW) and 182 as European American (CEU) (Table 1). It is expected that many in the ASW population will have higher WAA than individuals in the CEU population, which would provide a more accurate reference for this study. The created human reference includes more genetic variability and number of samples from individuals of African descent, which better reflects our RCC patient cohorts. There were no significant differences by sex between ASW and CEU study participants.

2. There were significant differences by sex, age, stage, and vital status in the pan-RCC patient cohort

There were significant differences by sex between ccRCC patients with high (n = 40) and low (n = 473) WAA (Table 2). pRCC patients with high (n = 54) and low (n = 203) WAA had significant differences in age and stage (Table 2). Whereas, chRCC patients had significant differences in age and vital status by ancestry (high WAA n = 4; low WAA n = 58). All SCNV and gene expression data were normalized to remove biases from these confounding variables.

3. Self-reported AAs have greater genetic variability across all three RCC subtypes

AAs are an admixed population with a high degree of genetic heterogeneity with African, Native American, and European ancestry due to the Trans-atlantic Slave Trade. The contribution of African and European ancestry differs greatly at the individual level in AAs from the same geographic area in the US. Also, the proportion of African ancestry can vary as much as tenfold in self-reported AA and EA populations[24]. The average genetic ancestry of AAs is ~73 percent from West Africa, primarily from Niger (~71 percent), with varying levels of admixture[24,25]. Therefore, 70 percent WAA was used as a threshold, separating patients into groups based on high and low ancestry. In ccRCC patients that self-identify as AA, a WAA range from 0.22-93.39 percent was observed. By contrast, in ccRCC patients that self-identified as EA, the range was 0.1-60.28 percent WAA (Table 3). AAs had a wider range of WAA. There was a similar trend in pRCC patients, with self-reported AAs having a larger WAA range compared with self-reported EAs (Table 3). In chRCC patients, the range of WAA was comparable between self-identified AAs and EAs (Table 3). The mean WAA for self-reported AAs across all three RCC subtypes was between 73 and 81 percent (Table 3).). The mean WAA for self-reported EAs across all three RCC subtypes was between 0.08 and 1.53 percent (Table 3). The median values across all three subtypes were very similar to the mean in both populations, reflecting symmetrical distribution of our WAA data and minimal skew (Table 3). Although patients self-reported as AA, fifteen ccRCC and 9 pRCC had WAA below 70 percent (Table 3). There are no self-reported EAs with greater than 70 percent WAA (Table 3).

Because AAs have more genetic variability across RCC histologies, a more refined analysis was performed. RCC patients were categorized into quintiles based on WAA (Table 4). Nine ccRCC patients self-identified as AA, while their WAA was recorded between 0 and 20

(two individuals) and 41-60 (seven individuals) percent (Table 4). All self-identified AA pRCC and chRCC patients had greater than 60 percent WAA, with no patients falling within the first three quintiles (Table 4). Because there were only four self-reported chRCC patients, they were excluded from further SCNV analysis due to limited sample size.

4. Comparative genome-wide, chromosomal, and gene SCNV analyses reveal significant differences according to high and low WAA in ccRCC patients

There were 111,766 total amplified and deleted chromosomal segments across the genome, 52,220 of which had ancestral differences (> 70% WAA compared with < 70% WAA) (Figure 1A). At the genome level, some SCNV differences by WAA were clearly visible in size and number (Figure 1B). Next, 25 chromosomal segments on chromosomes 7, 12, and 19 were deemed significant by ancestry (P + FDR <0.05) (Figure 2). A total of 8 different cytobands were involved. Individuals with high WAA had less amplification of the 8 cytobands compared to those with low WAA (Figure 2A, Supplementary Table 1). Whereas, individuals with high WAA had a greater frequency of deletion than their low WAA counterparts (Figure 2B, Supplementary Table 2). There were 12 genes that experienced a SCNV (Supplementary Table 3). Three transcripts were excluded from the analysis for lack of a HGNC ID (*AC003989.4*, *RP11-438N16.1*, and *CTB-133G6.1*) and two genes (*LST3* and *SLCO1B7*) had the same HGNC ID (*SLCO1B7*). Therefore, only eight out of 12 genes had a bonafide HUGO Gene Nomenclature Committee (HGNC) ID. These gene symbols were used for Gene Ontology (GO) enrichment analysis (Supplementary Table 3). Several biological processes, cellular components, and molecular functions were associated with these genes (Figure 3 and Supplementary Table 4).

5. Comparative genome-wide, chromosomal, and gene SCNV analyses reveal significant differences according to high and low WAA in pRCC patients

There were 59,868 total amplified and deleted chromosomal segments across the genome. A total of 31,981 chromosomal segments had ancestral differences (> 70% WAA compared with < 70% WAA) (Figure 4A). At a broad level, some SCNV differences by WAA were clearly visible in size and number (Figure 4B). We found 476 chromosomal segments to be significant by ancestry across 16 chromosomes 1-4, 6-10, 12-14, 16-17, and 20-21 ($P < 0.05$). A total of 103 different cytobands were involved. Amplification patterns across cytobands vary between pRCC patients with high and low WAA (Figure 5A, Supplementary Table 5). Some loci on 17q had less amplification in individuals with high WAA, and cytobands on chromosomes 14, 6, and 10 had increased amplification in patients with high WAA (Figure 5A, Supplementary Table 5). There were fewer 6p and 10q deletions in pRCC patients with high WAA and more deletions at loci 17q21.2 in individuals with low WAA (Figure 5B, Supplementary Table 6). There were 793 genes that experienced a SCNV (Supplementary Table 7). GO enrichment analysis identified top sub-ontologies associated with the candidate genes: fascia adherens, glutathione derivative metabolic process, and glutathione transferase activity (Figure 6).

6. A four gene pan-cancer SCNV signature was identified in ccRCC and pRCC patients

Four genes were identified in both ccRCC and pRCC patients: *ACTR3C*, *LST3*, *SLCO1B3*, and *SLCO1B7*. Gene expression data for both subtypes was only available for *ACTR3C*. There were no large-scale differences according to WAA quintile, with similar median expression between quintiles 41-60 and 61-80, as well as 0-20 and 81-100 percent for ccRCC patients (Figure 7). Similarly, increasing WAA did not correspond with linear increased or

decreased *ACTR3C* expression in pRCC patients (Figure 7). Other genes were studied to explore the relationship between WAA and gene expression in ccRCC and pRCC subtypes.

7. WAA predicts expression the pRCC candidate gene *NOVA1*

In ccRCC patients (504 out of 513), only three candidate genes were considered to be expressed (> 10 reads/individual) (Figure 8A). Simple linear regression analysis showed no significant relationships between expression of *ACTR3C, IMMP2L,* and *SLCO1B3* and WAA (Figure 8B). In pRCC patients (183 of 257), there was expression data for 314 out of 793 candidate genes (Figure 9A). A differential expression analysis by 70% WAA was used to narrow down the list. The simple linear regression model conducted on the three most significantly expressed candidate genes identified a significant relationship between *NOVA1* expression and WAA (Figure 9B). The relationships for *RPL23AP7* and *NOMO3* expression with WAA were not significant (Figure 9B).

8. pRCC patients with low *NOVA1* expression and high WAA had better disease specific survival than those with low *NOVA1* expression and low WAA

Survival analyses was performed for low and high *NOVA1* expressors with pRCC groups by WAA. There were no significant relationships between *NOVA1* expression and WAA (Figure 10). Interestingly, low *NOVA1* expressors with high WAA tended to have better disease-specific survival outcomes compared to patients with low WAA (Figure 10A). There was no relationship between high *NOVA1* expressors and WAA (Figure 10B).

D. Discussion

1. Candidate ccRCC genes have prior cancer-associations

Common fragile sites (CFS) are regions of great genomic instability found in all individuals[26]. CFS have been associated with chromosomal rearrangements in cancers, usually

resulting in deletions of tumor suppressor genes and amplification of oncogenes, oftentimes as a result of SCNVs[26]. CFS instability begins during the pre-cancer stage, suggesting its role in tumorigenesis[26]. There are more than 120 CFS in the human genome, including those that affect SCNVs in ccRCC patients on chromosomes 7, 12, and 19 (https://webs.iiitd.edu.in/raghava/humcfs/chrom.html).

RPH3A is found on 12q24.13, a cytoband with more amplifications in ccRCC patients with low WAA. It is a small effector for RAB3A, a small g-protein, thought to be involved in neurotransmitter release[27]. Small GTPases are a family of GTP-hydrolyzing enzymes that alternate between inactivity (while bound to GDP) and activity (while bound to GTP)[28]. They serve as regulators for cellular processes such as cell differentiation, proliferation, and motility[28]. Prior literature cites that overexpression of certain GTPases can have tumorigenic effects[28]. Specifically, RAB3A has been associated with endocrine tissues, and increased expression has been linked to brain tumors and hepatocellular carcinoma cells[29]. Additionally, upregulation of RAB3A in gliomas has been shown to increase proliferation, metastasis, and drug resistance[29]. RAB3A recruits RPH3A. In combination, these findings suggest that the amplification of *RPH3A* may have oncogenic effects on ccRCC patients with low WAA.

IMMP2L is found on 7q31.1, a cytoband with more amplifications in ccRCC patients with low WAA. Based on previous findings, *IMMP2L* may play an important role in cancer. MalaCards lists breast and prostate cancers are associated with the gene, among other diseases[30]. Through knockdown experiments, *IMMP2L* has been tied to cellular senescence, an irreversible process that contributes to normal aging and tumor suppression[31], resulting in the cell's inability to enter the cell cycle[31]. *IMMP2L* has two substrates: glycerol-3-phosphate dehydrogenase 2 (GPD2), a metabolic enzyme, and apoptosis-inducing factor (AIF), a redox-active

flavoprotein[31,32]. GPD2 has two hypothesized effects on cellular senescence: it increases NAD+ production, because NAD+ has pro-longevity effects, or it encourages mitochondrial phospholipid synthesis by providing the precursor G3P[31]. Loss of phospholipids due to knocked down *IMMP2L* had downstream effects on the catabolism of pyruvate by the tricarboxylic acid cycle, which induced cellular senescence[31]. Amplification of *IMMP2L* could contribute to tumorigenic effects in individuals with low WAA and serve as a future target for novel therapies.

SLCO1B3 is found on 12p12.2, a cytoband with more amplifications in ccRCC patients with low WAA. *SLCO1B3* plays a role in transmembrane transport[33], and is an example of a solute carrier (SLC)[34]. Cancer cells may have increased expression of SLC transporters in order to increase nutrient absorption, especially when competing with normal cells[35]. In addition, SLCs have been identified as drug transporters across cancer cell membranes and could play a role in modifying therapeutic efficacy for existing FDA approved treatments[35]. Specifically, *SLCO1B3* ha contradictory expression patterns. *SLCO1B3* is overexpressed in colonic, lung, breast, prostate, pancreatic, testicular and ovarian cancers[34] and is known to be expressed in the breast, liver, and kidney[34]. By contrast, high expression of *SLCO1B3* was associated with smaller breast tumor size and decreased risk of occurrence[34]. One study looked at various hormone-related cancers and markers of progression. As a future direction, they suggested studying kidney tumors since *SLCO1B3* expression was observed in various normal tissues compared to expression in cancerous tissues[36]. Therefore, further studies on the expression of *SLCO1B3* in RCC patients are needed.

ACTR3C is found on 7q36.1, a cytoband with more amplifications in ccRCC patients with low WAA. *ACTR3C,* an actin related protein, has few known direct associations to cancer. Previous studies have found hypermethylation of the gene associated with keloid development[37,

[38]. However, the association with the Arp2/3 protein complex suggests possible downstream tumorigenic effects[39]. The Arp2/3 complex generates branched actin networks and has been linked with breast, lung, pancreatic, and colorectal cancers, with high expression correlating with aggressive disease and cell invasion[39]. In addition, silencing of the complex has been shown to reduce cell migration in pancreatic cancer[40]. *ACTR3C* amplifications may offer a novel therapeutic target for Arp2/3 protein complex disruption in ccRCC patients. Most of the candidate genes had amplifications in individuals with below 70 percent WAA. Some self-reported AAs fall into this category, and could benefit from more precise treatment using WAA.

2. Candidate pRCC genes have prior cancer-associations

NOVA1 is an alternative splicing regulator on 14q12 (greater gene deletions in patients with low WAA)[41]. *RPL23AP7* is a ribosomal protein on 2q14.1 (greater deletions in patients with high WAA)[42]. *NOMO3* is a Nodal modulator on 16p13.11 (greater deletions in patients with high WAA)[43]. These genes were enriched in glutathione-transferase activity and glutathione-derivative metabolic processes. Glutathione S-transferases (GSTs) are involved in the detoxification of electrophilic metabolites[44]. Interestingly, a large number of carcinogens related to RCC, such as polycyclic aromatic hydrocarbon diol-epoxides and halogen solvents, and chemotherapy agents are metabolized by GSTs[44]. Common changes to GST with associations to cancer include genes *GSTM1*, *GSTT1*, *GSTP1* and *GSTA1*[44]. Additionally, glutathione was recently found to be increased in pRCC patients, serving as a new hallmark of the disease[45]. This supports my SCNV gene-specific findings. Increased glutathione synthesis within pRCC patients is a result of deficient glucose synthesis and inefficient oxidative phosphorylation[45]. This suggests that the detoxification pathway may be less prominent in patients with high WAA and

could offer a novel population-specific treatment to replace these gene products in this population.

NOVA1 enhances full-length *hTERT* splicing, which increases telomerase activity in cancer cells[46]. Telomerase is a ribonucleoprotein complex which maintains telomere length in ~90 percent of cancers, allowing cells to attain unlimited replicative ability[46]. Work is being done to thoroughly understand the mechanistic role of *NOVA1*, in order to exploit its purpose in an anticancer therapy to prevent telomerase activity in cancer cells[46].

Increased expression of *NOMO3* has been found in medullary thyroid cancer[47]. Nodal overexpression has been linked to cancer, specifically RCC, and was found to promote proliferation and invasion, and inhibit apoptosis[48]. Additionally, Nodal expression was also associated with tumor metastasis and poor prognosis in breast cancer[49]. Nodal knockdown was correlated with decreased activity of the vasculogenic mimicry pathway, which promotes tumor growth[49]. Future research should be conducted on the possible prognostic and therapeutic targeting ability of *NOMO3* in RCC based on WAA.

3. Overlapping candidate genes may serve as novel pan-RCC therapeutic targets

Four gene SCNVs were shared across ccRCC and pRCC patients: *ACTR3C*, *SLCO1B3*, *SLCO1B7*, and *LST3* (Alias: *SLCO1B7*). *ACTR3C* has prior cancer associations and SLCs are drug transporters that modifying therapeutic efficacy. All of these genes relate to aggressive tumor biology. Chapter 1 revealed an association with decreased GST activity by race in ccRCC patients. Chapter 2 analyses suggested loss of GST activity was involved in pRCC patients. Glutathione metabolism may also serve as a pan-RCC therapeutic target. Future research should be conducted to identify prognostic and therapeutic effects of these genes because such pan-RCC targets could help to narrow the survival disparity existing in kidney cancer patients.

E. Tables and Figures

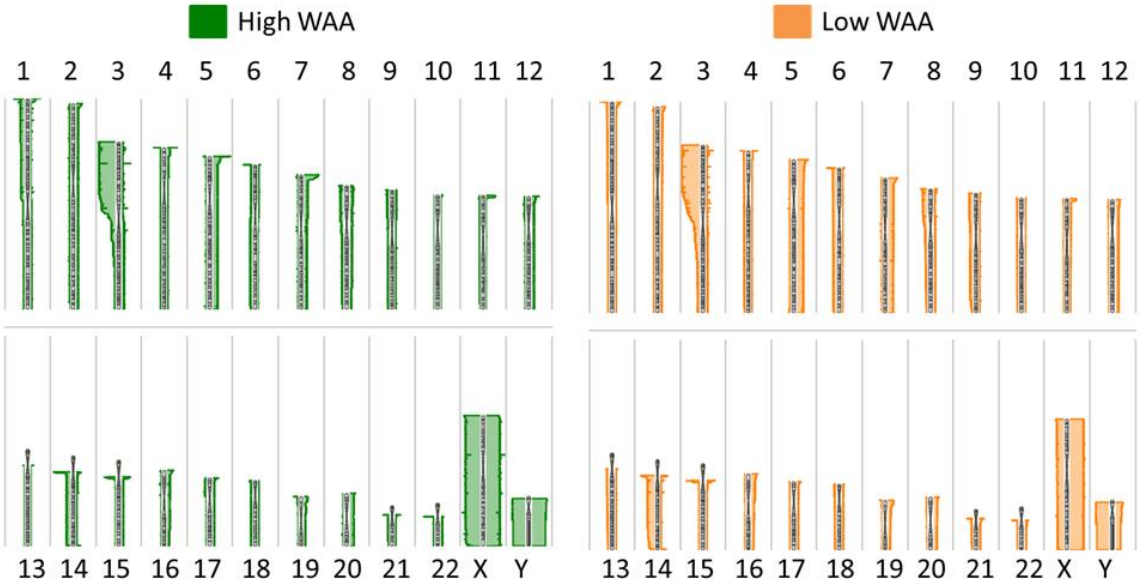
Table 1. Demographic Characteristic of 1000 Genomes Project Study Participants Serving as the Human Reference Genome			
	AA (ASW, *n* = 99)	**EA (CEU, *n* = 182)**	***P***
Sex (%) ^			*0.71*
Female	53 (54)	93 (51)	
Male	46 (46)	89 (49)	
^ Fisher's exact test			

Table 2. Clinical and Demographic Characteristics of Renal Cell Carcinoma Patients in the TCGA Cohort									
	Clear Cell RCC			**Papillary RCC**			**Chromophobe RCC**		
	High WAA (*n* = 40)	**Low WAA (*n* = 473)**	***P***	**High WAA (*n* = 54)**	**Low WAA (*n* = 203)**	***P***	**High WAA (*n* = 4)**	**Low WAA (*n* = 58)**	***P***
Age (years) #			0.205			*0.031*			*0.0431*
Mean	58.5	60.7		58.76	62.6		37.5	51.9	
Range	37-78	26-90		35-79	28-88		26-44	17-86	
Sex (%) ^			*0.024*			0.4958			0.999
Female	21 (53)	160 (34)		17 (32)	54 (27)		1 (25)	23 (40)	
Male	19 (47)	313 (66)		37 (65)	149 (73)		3 (75)	35 (60)	
Stage (%) +			0.064			*0.012*			0.636
I	28 (70)	227 (48)		32 (59)	94 (46)		0 (0)	20 (35)	
II	2 (5)	52 (11)		9 (18)	12 (7)		3 (75)	20 (35)	
III	5 (12)	113 (24)		3 (5)	21 (10)		1 (25)	13 (22)	
IV	5 (12)	78 (16)		3 (5)	6 (3)		0 (0)	5 (8)	
Unknown	0 (0)	3 (1)		7 (13)	70 (34)		0 (0)	0 (0)	
Vital Status (%) ^			0.162			0.403			*0.0004*
Dead	9 (22)	161 (34)		6 (11)	34 (17)		0 (0)	9 (16)	
Alive	31 (78)	309 (65)		48 (89)	169 (83)		4 (100)	49 (84)	
Unknown	0 (0)	3 (1)							
^ Fisher's exact test, + Chi square test, # t test									

Table 3. Stratification of RCC patients by West African Ancestry at a 70 Percent Threshold						
	Clear Cell RCC		**Papillary RCC**		**Chromophobe RCC**	
	Self-Reported African Americans (*n* = 55)	**Self-Reported European Americans (*n* = 458)**	**Self-Reported African Americans (*n* = 63)**	**Self-Reported European Americans (*n* = 194)**	**Self-Reported African Americans (*n* = 4)**	**Self-Reported European Americans (*n* = 58)**
Range % WAA	0.22-93.39	0.1-60.28	62.62-95.41	0.12-9.69	76.44-83.61	0.17-7.89
Range % EURA	3.41-97.91	33.59-99.41	2.01-35.82	89.75-99.24	15.52-21.7	43.73-99.2
Mean % WAA/ EURA	73.72/ 23.35	1.53/ 94.68	79.63/ 18.64	0.08/ 97.73	81.02/ 17.7	1.28/ 95.87
Med % WAA/ EURA	78.63/ 19.14	0.49/ 97.73	81/ 17.51	0.03/ 98.24	82.02/ 16.81	0.51/ 97.78
< 70% WAA (%)	15 (27.3)	458 (100)	9 (14)	194 (100)	0 (0)	58 (100)
> 70% WAA (%)	40 (72.7)	0 (0)	54 (86)	0 (0)	4 (100)	0 (0)
Med = Median		WAA = West African Ancestry			EURA = European Ancestry	

Table 4. Stratification of RCC patients by West African Ancestry Quintiles						
	Clear Cell RCC		**Papillary RCC**		**Chromophobe RCC**	
	Self-Reported African Americans (*n* = 55)	**Self-Reported European Americans (*n* = 458)**	**Self-Reported African Americans (*n* = 63)**	**Self-Reported European Americans (*n* = 194)**	**Self-Reported African Americans (*n* = 4)**	**Self-Reported European Americans (*n* = 58)**
0-20% Ancestry (%)	2 (3.6)	455 (99.4)	0 (0)	194 (100)	0 (0)	57 (99)
21-40% Ancestry (%)	0 (0)	2 (.4)	0 (0)	0 (0)	0 (0)	1 (1)
41-60% Ancestry (%)	7 (12.7)	1 (0.2)	0 (0)	0 (0)	0 (0)	0 (0)
61-80% Ancestry (%)	25 (45.5)	0 (0)	28 (44)	0 (0)	0 (0)	0 (0)
81-100% Ancestry (%)	21 (38.2)	0 (0)	34 (56)	0 (0)	4 (100)	0 (0)

A



B

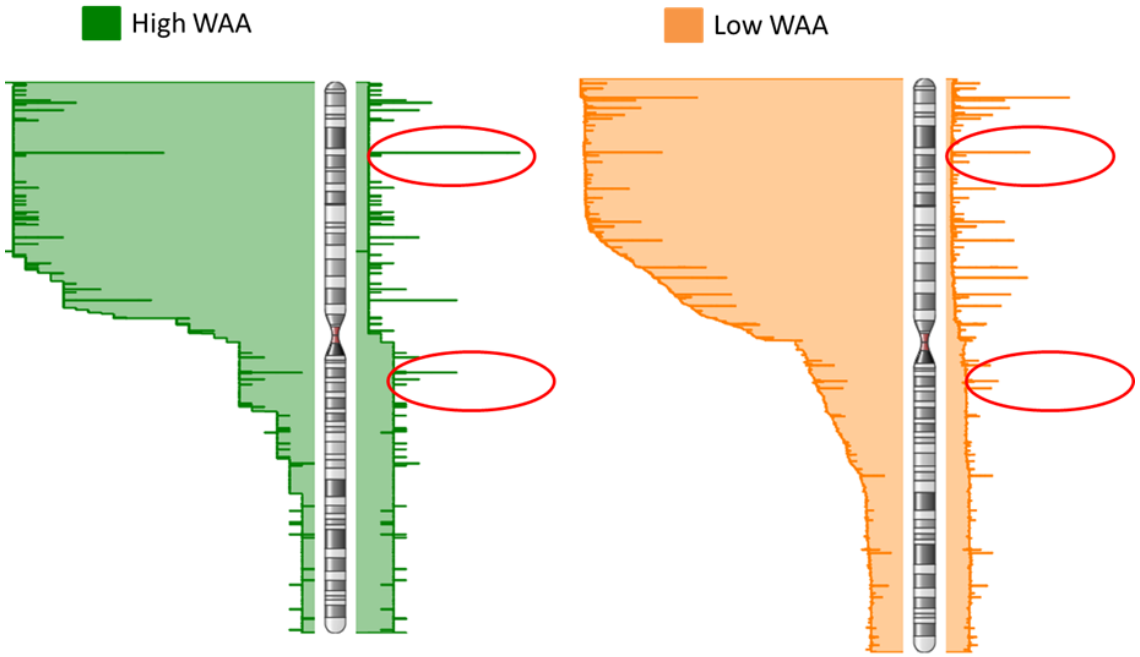


Figure 1. Genome-wide comparison of SCNVs in tumor tissues from ccRCC patients with high and low WAA. (A) Karyogram representing SCNV frequencies across all chromosomes. (B) Chromosome 3 karyogram comparing representative SCNV peak magnitude and number differences by WAA. Left peaks are deletions and right peaks are amplifications.

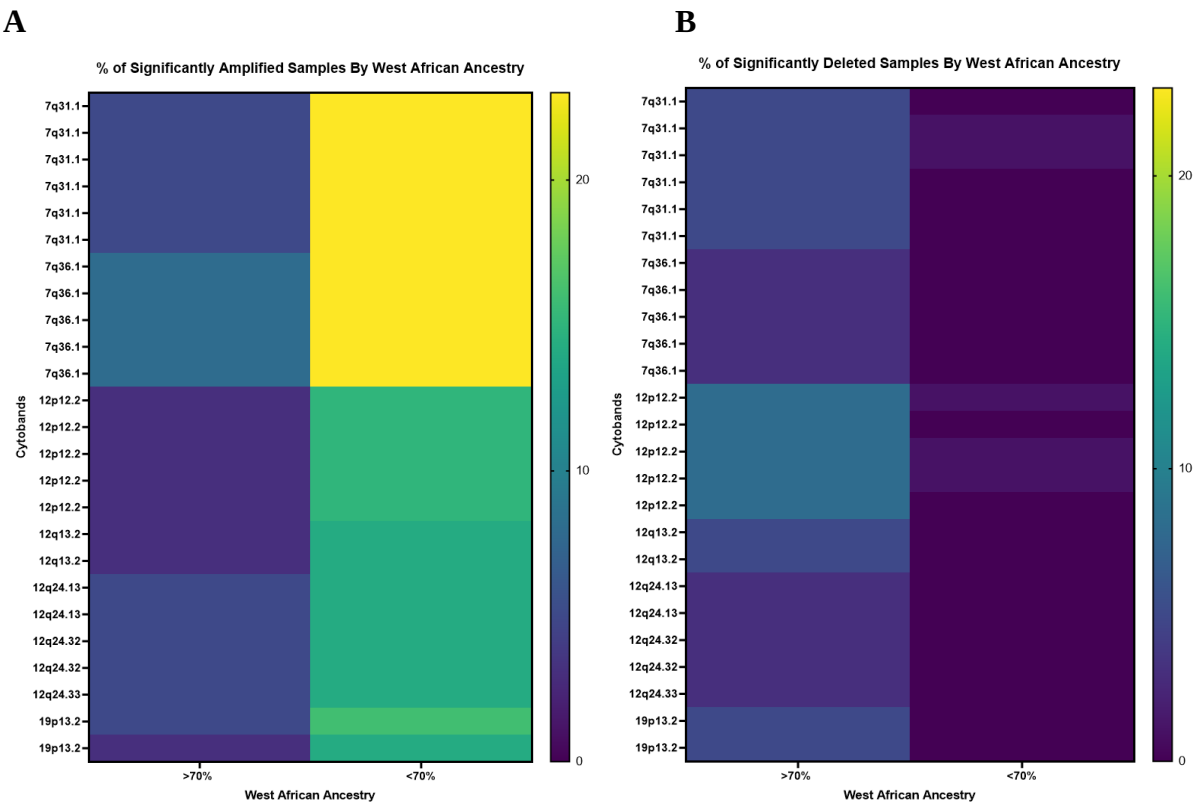


Figure 2. Chromosomal SCNV comparisons in tumor tissues from ccRCC patients with high and low WAA. (A) Heatmap representing amplified cytoband frequencies across three chromosomes. (B) Heatmap representing deleted cytoband frequencies across three chromosomes.

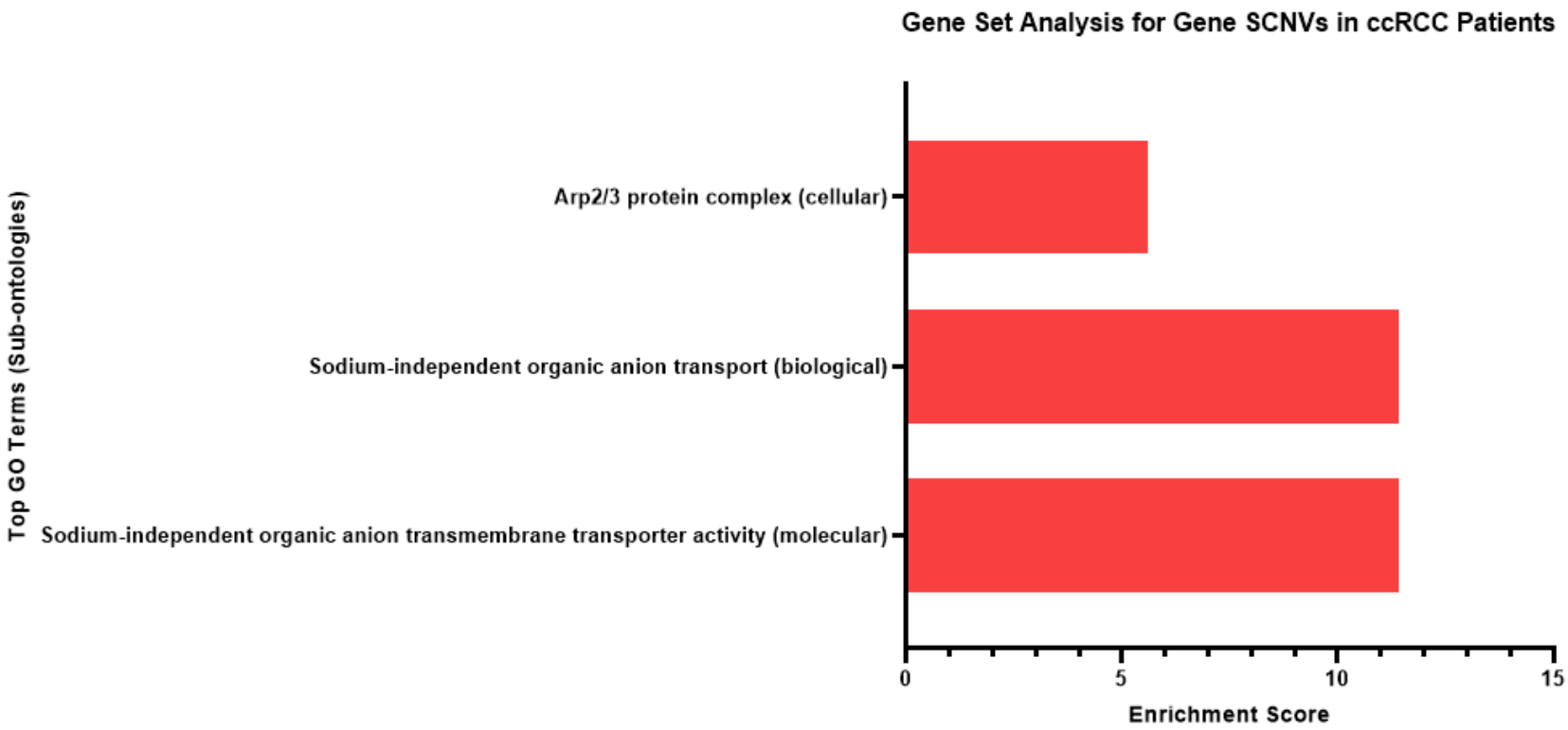
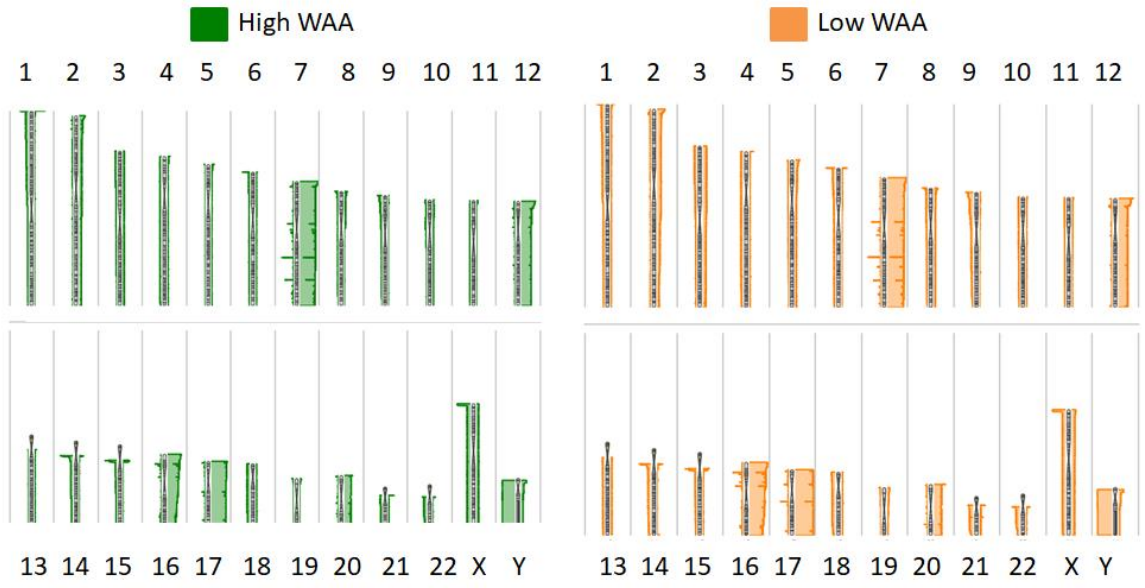


Figure 3. Gene SCNV functional comparisons in tumor tissues from ccRCC patients with high and low WAA. GO terms with the highest enrichment score are shown for cellular, biological, and molecular processes.

A



B

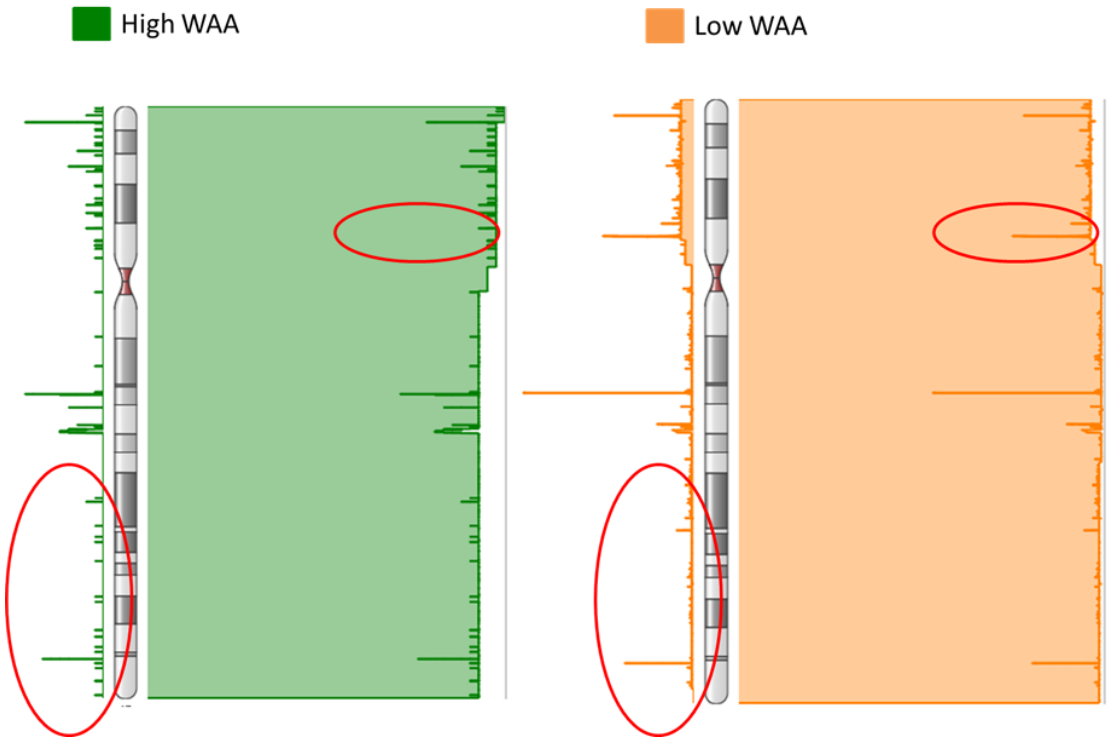


Figure 4. Genome-wide comparison of SCNVs in tumor tissues from pRCC patients with high and low WAA. (A) Karyogram representing SCNV frequencies across all chromosomes. (B) Chromosome 17 karyogram comparing representative SCNV peak magnitude and number differences by WAA. Left peaks are deletions and right peaks are amplifications.

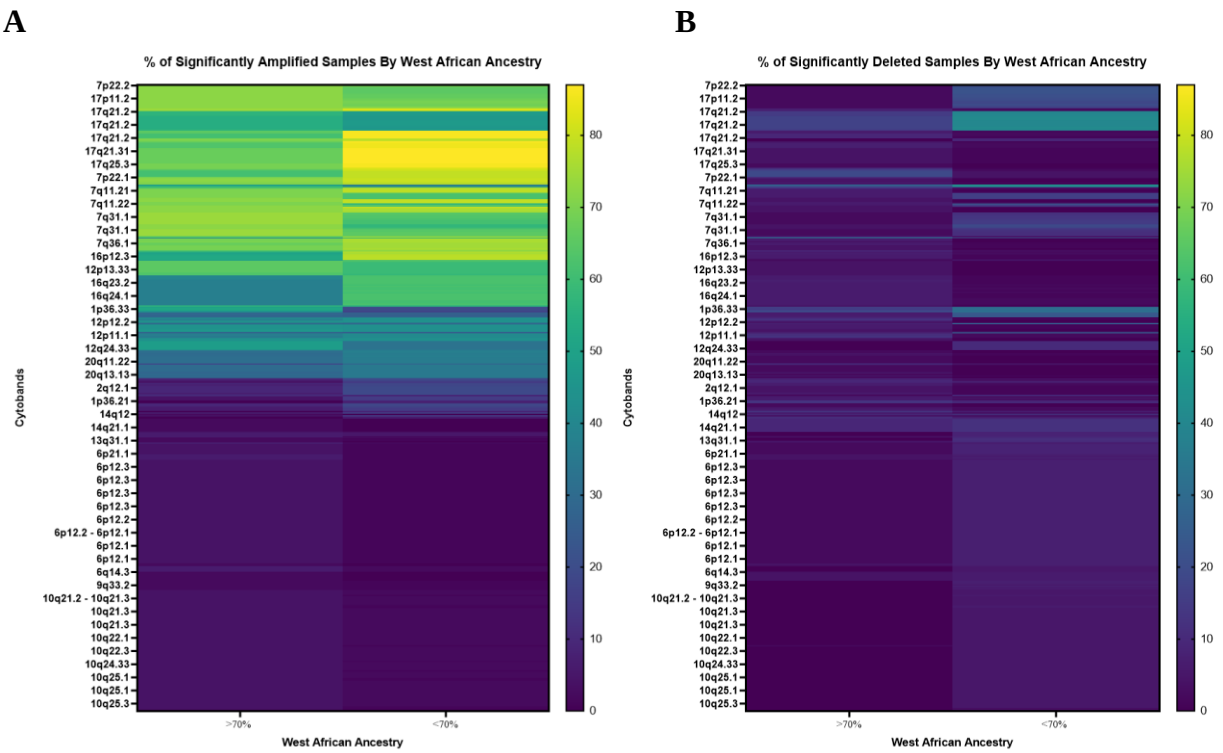


Figure 5. Chromosomal SCNV comparisons in tumor tissues from pRCC patients with high and low WAA. (A) Heatmap representing amplified cytoband frequencies across 16 chromosomes. (B) Heatmap representing deleted cytoband frequencies across 16 chromosomes.

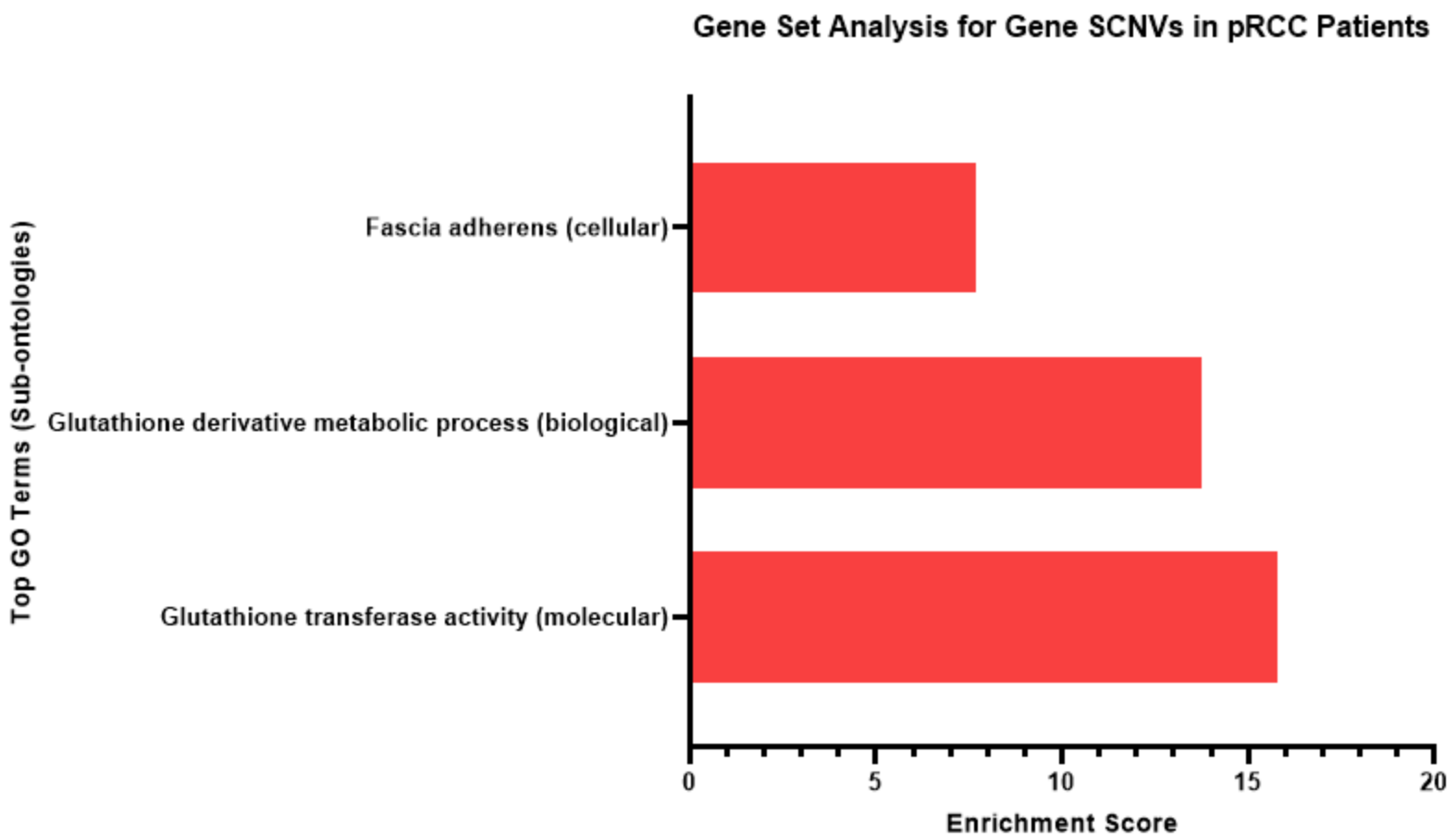


Figure 6. Gene SCNV functional comparisons in tumor tissues from pRCC patients with high and low WAA. GO terms with the highest enrichment score are shown for cellular, biological, and molecular processes.

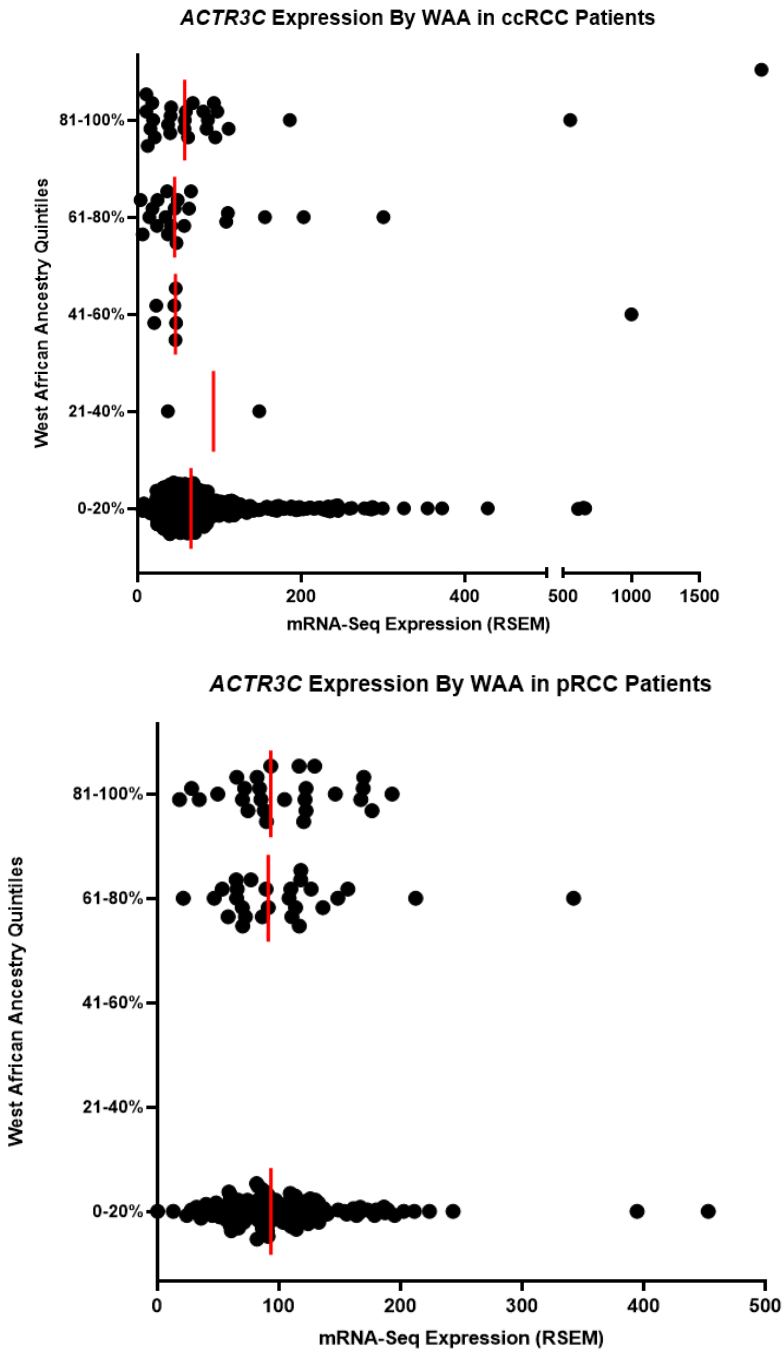
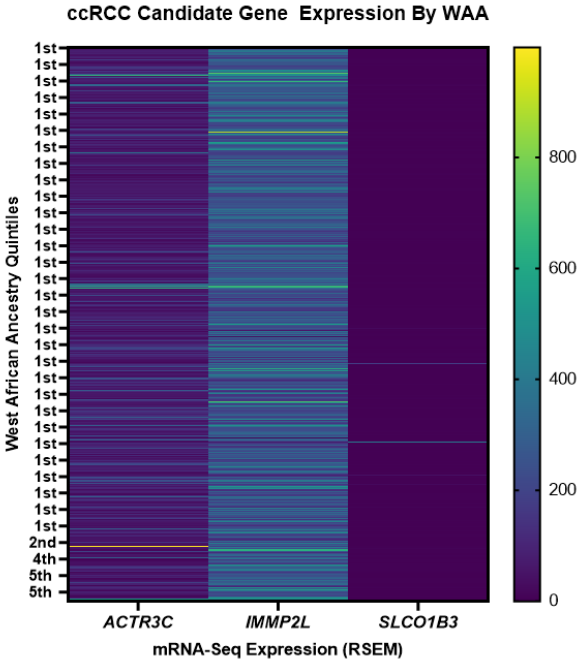


Figure 7. *ACTR3C* expression data by WAA quintiles for both ccRCC and pRCC patients. Scatter plots of *ACTR3C* expression patterns in ccRCC and pRCC patients with varying degrees of WAA.

A



B

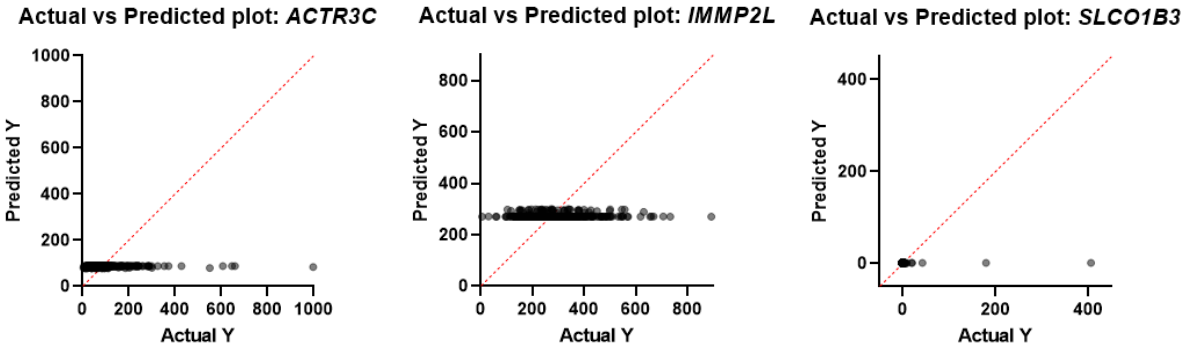
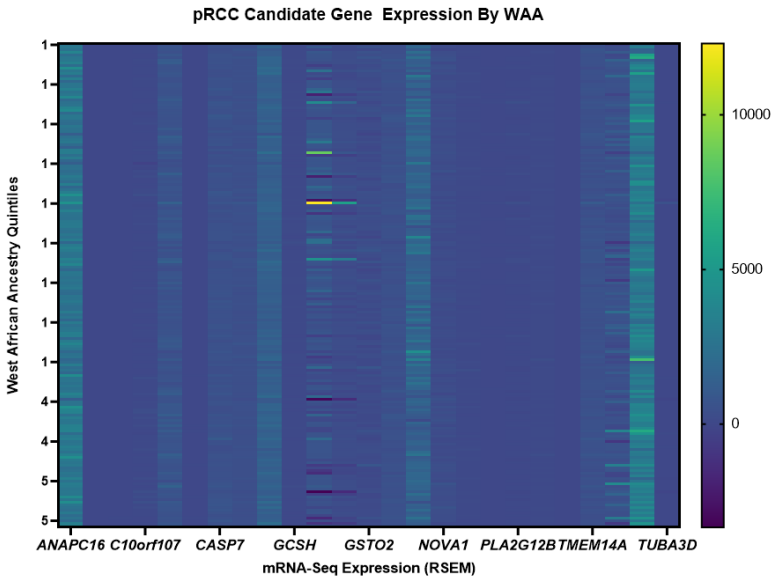
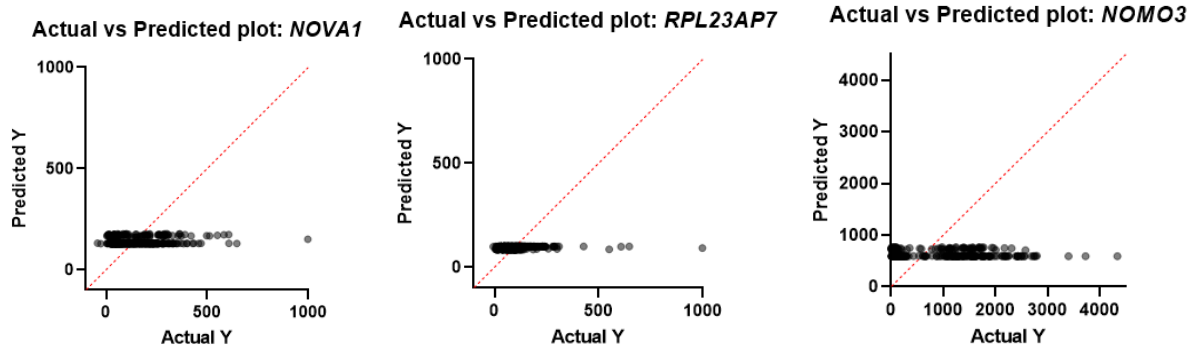


Figure 8. Gene expression patterns for ccRCC patients by WAA quintile. A) Heat map depicts mRNA expression of *ACTR3C*, *IMMP2L*, and *SLCO1B3* for ccRCC patients. B) Simple linear regression models the relationship between WAA and gene expression for three candidate genes.

A



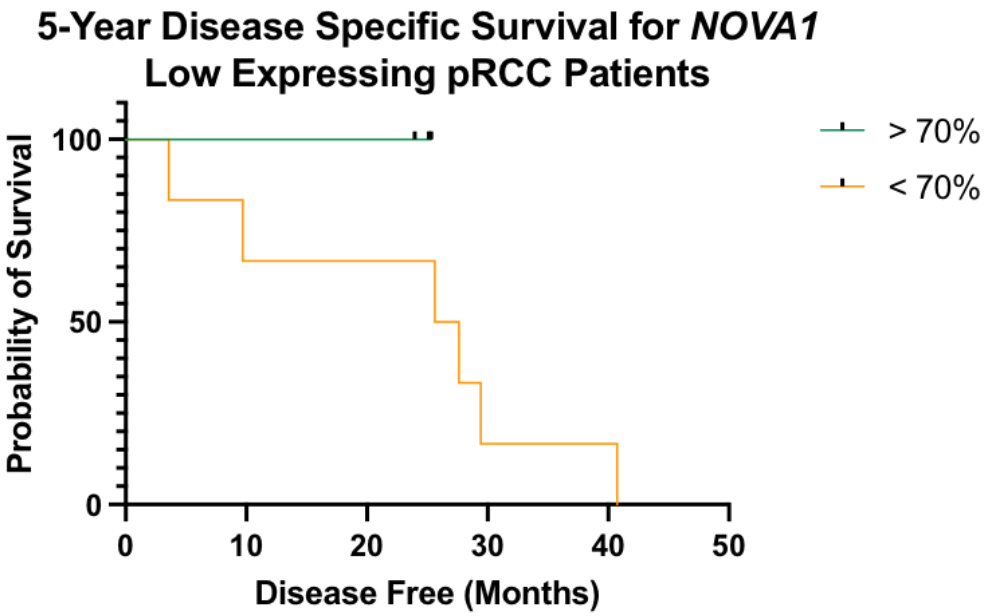
B



***P* = 0.0015** ***P* = 0.1306** ***P* = 0.0954**

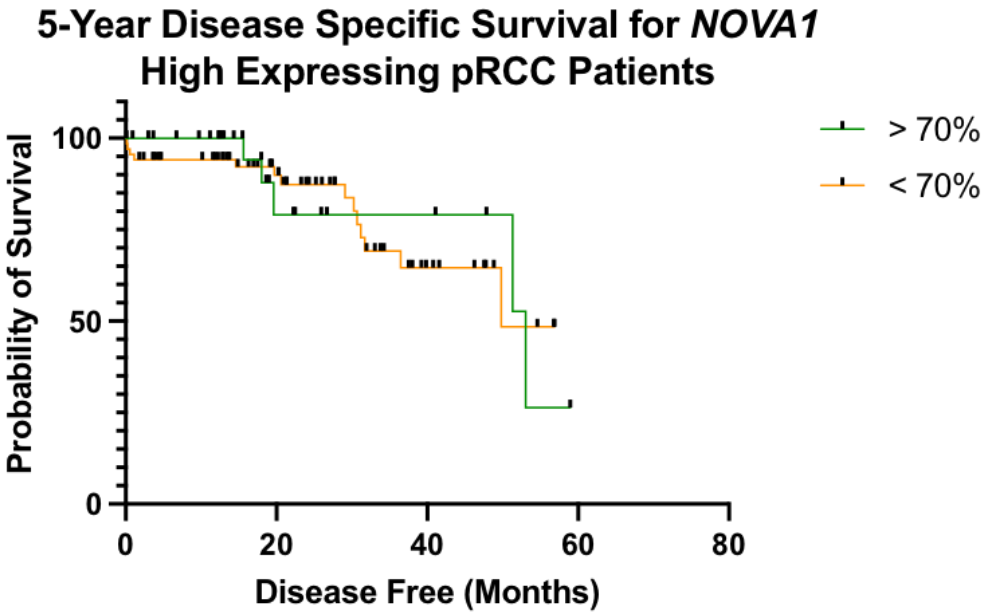
Figure 9. Gene expression patterns for pRCC patients by WAA quintile. A) Heat map depicts mRNA expression of *NOVA1*, *RPL23AP7*, and *NOMO3* for ccRCC patients. B) Simple linear regression models the relationship between WAA and gene expression for three candidate genes.

A.



***P* = 0.2945**

B.



***P* = 0.9460**

Figure 10. Kaplan Meier disease-specific survival outcomes for pRCC patients by WAA. A) Kaplan Meier survival curve for low expressing *NOVA1* pRCC patients. B) Survival outcomes for pRCC high expressors of *NOVA1*.

F. References

1. Saad, A. M. *et al.*Trends in Renal-Cell Carcinoma Incidence and Mortality in the United States in the Last 2 Decades: A SEER-Based Study. *Clin Genitourin Cancer*17, 46-57.e5 (2019).
2. Nabi, S., Kessler, E. R., Bernard, B., Flaig, T. W. & Lam, E. T. Renal cell carcinoma: a review of biology and pathophysiology. *F1000Res*7, (2018).
3. National Cancer Institute. (2018). Kidney Cancer - Surveillance, Epidemiology, and End Results Program - Cancer stat facts. Retrieved October, 2020
4. Padala, S. A. *et al.*Epidemiology of Renal Cell Carcinoma. *World J Oncol*11, 79–87 (2020).
5. Tabibu, S., Vinod, P. K. & Jawahar, C. V. Pan-Renal Cell Carcinoma classification and survival prediction from histopathology images using deep learning. *Scientific Reports*9, 10509 (2019).
6. Liu, Y. J., Houldsworth, J., Emmadi, R., Dyer, L. & Wolff, D. J. Assessing Genomic Copy Number Alterations as Best Practice for Renal Cell Neoplasia: An Evidence-Based Review from the Cancer Genomics Consortium Workgroup. *Cancer Genetics*244, 40–54 (2020).
7. Jia, K., Wu, Y., Huang, J. & Wu, H. Survival-Associated Alternative Splicing Events in Pan-Renal Cell Carcinoma. *Front. Oncol.*9, (2019).
8. Huang, J. J. & Hsieh, J. J. The Pan-Omics Landscape of Renal Cell Carcinoma and Its Implication on Future Clinical Practice. *Kidney Cancer*4, 121–129 (2020).
9. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*499, 43–49 (2013).
10. Massari, F. *et al.*Toward a genome-based treatment landscape for renal cell carcinoma. *Crit. Rev. Oncol. Hematol.*142, 141–152 (2019).
11. Ohashi, R. et al. Classic Chromophobe Renal Cell Carcinoma Incur a Larger Number of Chromosomal Losses Than Seen in the Eosinophilic Subtype. *Cancers*. 11, 1492. 3 Oct. (2019).
12. Renal Cell Cancer Treatment (PDQ®)–Patient Version - National Cancer Institute. (2004).
13. Singer, E. A., Bratslavsky, G., Linehan, W. M. & Srinivasan, R. Targeted therapies for non-clear renal cell carcinoma. *Target Oncol*5, 119–129 (2010).
14. Xicola, R. M. *et al.*Lack of APC somatic mutation is associated with early-onset colorectal cancer in African Americans. *Carcinogenesis*39, 1331–1341 (2018).
15. Lara, O. D. *et al.*Pan-cancer clinical and molecular analysis of racial disparities. *Cancer*126, 800–807 (2020).
16. Ramakodi, M. P. *et al.*Integrative genomic analysis identifies ancestry-related expression quantitative trait loci on DNA polymerase β and supports the association of genetic ancestry with survival disparities in head and neck squamous cell carcinoma. *Cancer*123, 849–860 (2017).

17. Huo, D. *et al.*Comparison of Breast Cancer Molecular Features and Survival by African and European Ancestry in The Cancer Genome Atlas. *JAMA Oncol*3, 1654–1662 (2017).
18. Vantaku, V. *et al.*DNA methylation patterns in bladder tumors of African American patients point to distinct alterations in xenobiotic metabolism. *Carcinogenesis*40, 1332–1340 (2019).
19. Gouveia, M. H. *et al.*Origins, Admixture Dynamics, and Homogenization of the African Gene Pool in the Americas. *Molecular Biology and Evolution*37, 1647–1656 (2020).
20. Ricketts, C. J. *et al.*The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep*23, 313-326.e5 (2018).
21. Purdue, M. P. *et al.*A genome-wide association study of renal cell carcinoma among African Americans. *Cancer Epidemiol Biomarkers Prev*23, 209–214 (2014).
22. Ballouz, S., Dobin, A. & Gillis, J. A. Is it time to change the reference genome? *Genome Biology*20, 159 (2019).
23. Sherman, R. M. *et al.*Author Correction: Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat Genet*51, 364 (2019).
24. Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D. & Mountain, J. L. The Genetic Ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet*96, 37–53 (2015).
25. Tishkoff, S. A. *et al.*The Genetic Structure and History of Africans and African Americans. *Science*324, 1035–1044 (2009)
26. Li, S. & Wu, X. Common fragile sites: protection and repair. *Cell Biosci*10, 29 (2020).
27. *RPH3A* Gene (Protein Coding). *Gene Cards: The Human Gene Database*, (2020).
28. Prieto-Dominguez, N., Parnell, C. & Teng, Y. Drugging the Small GTPase Pathways in Cancer Treatment: Promises and Challenges. *Cells*8, (2019).
29. Raffaniello, R. D. Rab3 proteins and cancer: Exit strategies. *J Cell Biochem*(2021).
30. *MalaCards: Search Human Diseases—IMMP2L*. (n.d.). Retrieved May 17, 2021.
31. Sica, V. & Kroemer, G. IMMP2L: a mitochondrial protease suppressing cellular senescence. *Cell Res*28, 607–608 (2018).
32. Yuan, L. *et al.*Switching off IMMP2L signaling drives senescence via simultaneous metabolic alteration and blockage of cell death. *Cell Res*28, 625–643 (2018).
33. *SLCO1B3* Gene (Protein Coding). *Gene Cards: The Human Gene Database*, (2020).
34. Sutherland, R., Meeson, A. & Lowes, S. Solute transporters and malignancy: establishing the role of uptake transporters in breast cancer and breast cancer metastasis. *Cancer Metastasis Rev*39, 919–932 (2020).
35. Li, Q. & Shu, Y. Role of solute carriers in response to anticancer drugs. *Molecular and Cellular Therapies*2, 15 (2014).
36. Pressler, H., Sissung, T. M., Venzon, D., Price, D. K. & Figg, W. D. Expression of OATP Family Members in Hormone-Related Cancers: Potential Markers of Progression. *PLOS ONE*6, e20372 (2011).

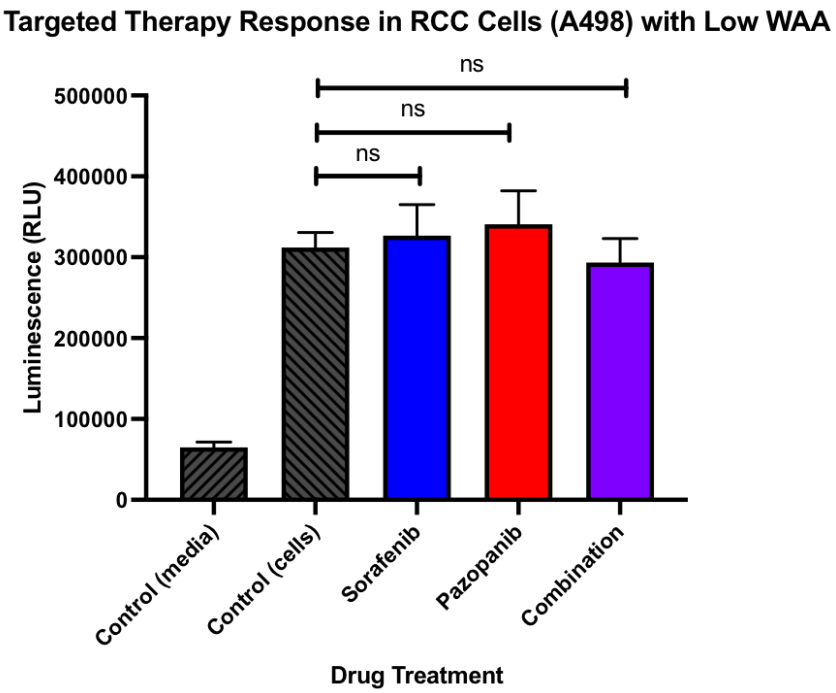
37. Jones, L. R. *et al.*Biological significance of genome-wide DNA methylation profiles in keloids. *Laryngoscope*127, 70–78 (2017).
38. Garcia-Rodriguez, L. *et al.*Causal network analysis of head and neck keloid tissue identifies potential master regulators. *Laryngoscope*126, E319-324 (2016).
39. Molinie, N. & Gautreau, A. The Arp2/3 Regulatory System and Its Deregulation in Cancer. *Physiol Rev*98, 215–238 (2018).
40. Rauhala, H. E., Teppo, S., Niemelä, S. & Kallioniemi, A. Silencing of the ARP2/3 complex disturbs pancreatic cancer cell migration. *Anticancer Res*33, 45–52 (2013).
41. *NOVA1* Gene (Protein Coding). *Gene Cards: The Human Gene Database*, (2020).
42. *RPL23AP7* Gene (Pseudogene). *Gene Cards: The Human Gene Database*, (2020).
43. *NOMO3* Gene (Protein Coding). *Gene Cards: The Human Gene Database*, (2020).
44. Radic, T. M. *et al.*Concomitance of Polymorphisms in Glutathione Transferase Omega Genes Is Associated with Risk of Clear Cell Renal Cell Carcinoma. *The Tohoku Journal of Experimental Medicine*246, 35–44 (2018).
45. Ahmad, A. A. *et al.*Papillary Renal Cell Carcinomas Rewire Glutathione Metabolism and Are Deficient in Both Anabolic Glucose Synthesis and Oxidative Phosphorylation. *Cancers (Basel)*11, (2019).
46. Sayed, M. E. *et al.*NOVA1 directs PTBP1 to hTERT pre-mRNA and promotes telomerase activity in cancer cells. *Oncogene*38, 2937–2952 (2019).
47. Pozdeyev, N. *et al.*Comprehensive Immune Profiling of Medullary Thyroid Cancer. *Thyroid*30, 1263–1279 (2020).
48. Zhang, Z. *et al.*Nodal activates smad and extracellular signal-regulated kinases 1/2 pathways promoting renal cell carcinoma proliferation. *Mol Med Rep*12, 587–594 (2015).
49. Gong, W. *et al.*Nodal signaling promotes vasculogenic mimicry formation in breast cancer via the Smad2/3 pathway. *Oncotarget*7, 70152–70167 (2016).

VI. Conclusions

In Chapter 1, I showed that self-reported AAs and EAs had unique gene deletions and amplifications that correlated with gene expression and patient survival. Specifically, AA ccRCC patients with low candidate gene expression had better 5-year disease specific survival than EAs. If some AA ccRCC patients are high expressors, they may benefit from targeted therapies that will lower expression, improve their survival time, and help reduce racial disparities in survival. Also, using WAA to group ccRCC patients was more sensitive at finding gene expression differences that using self-reported race alone (e.g. *PTEN*).

In Chapter 2, I showed that RCC patients with high and low WAA had distinct gene deletions and amplifications previously associated with key cancer-related pathways. WAA was also seen to predict candidate gene expression. Specifically, pRCC patients with low *NOVA1* expression and high WAA had better 5-year disease specific survival than those with low WAA. Both self-reported race (a social construct) and genetic ancestry (a biological construct) should be considered when categorizing patients and searching for causes of kidney cancer disparities in the clinic.

Future studies should consider the role WAA plays in how RCC patients respond to targeted therapy, since this biological variable has been associated with drug response in a recent study. I performed a preliminary experiment with individual and combination targeted drug treatments in A498, an RCC cell line with low WAA. Combination treatments had decreased luminescence, which corresponded with less viable cells. I would expect to see even further decreased luminescence for a RCC cell line with high WAA (A704), due to its increased sensitivity to the targeted therapies Sorafenib and Pazopanib.



Supplementary Figure 1. Monotherapy and combination therapy drug response in a RCC cell line with low WAA. Cell viability was measured via luminescence using a CellTiter-Glo assay. Error bars reflect technical triplicates.

VII. Acknowledgements

I would like to thank the Lafayette College Department of Biology for the funding and support of my honors thesis, and my thesis committee Dr. James Dearworth, Dr. Eric Ho, and Dr. Tamara Stawicki. Next, I'd like to thank Tri Beta Biology Honors Society for providing me with a research grant to fund my work. Finally, I would like to thank my mentor, Dr. Khadijah A. Mitchell, for her continued guidance throughout this project and for helping me achieve my goals.