SPECIFIC AIMS

Each year, an estimated 14,000 people in the United States die as a consequence of renal cell carcinoma (RCC). RCC has the greatest associated mortality rate among urologic malignancies. The racial disparities in this disease have been well documented, with African Americans afflicted by RCC having very different disease characteristics and outcomes. Various lines of evidence point to genomic differences as a major contributor to these disparities. First, the incidence of RCC is 30% greater in African Americans than in Caucasians. Second, recent studies show that African Americans present with RCC at significantly younger ages, suggesting a potential hereditary predisposition. Third, racial differences in RCC histological subtypes are significant. Clear-cell RCC (ccRCC) is the most common histological type in all races; however, papillary RCC (pRCC) is three-fold more common in African Americans, accounting for 35-40% of cases. In this proposal, we aim to investigate whether genomic differences can explain these racial disparities. We will focus on mutations associated with the known, canonical drivers of ccRCC and pRCC, *VHL* and *MET*, respectively. We hypothesize that specific germline and somatic alterations in coding regions and associated non-coding elements of these two key driver genes are associated with racial disparities in kidney cancer.

Aim 1: To perform genome sequencing of African Americans with ccRCC to complete a missing aspect of The Cancer Genome Atlas (TCGA). African-American patients are underrepresented in TCGA ccRCC cohorts. To complete this missing aspect of TCGA and to ultimately probe genomic differences in racial disparities in RCC subtypes, we will perform whole-genome sequencing (WGS) on a cohort of African-American patients treated for ccRCC at Yale University. We will carefully match African-American patients that were underrepresented in TCGA ccRCC cohort. This work will generate a combined Yale-TCGA study population that is suitable for comparing genomic alterations according to race and histological subtype.

Aim 2: To identify key genomic variants associated with kidney cancer that exhibit racial disparities. We will first find all genomic annotations associated with the canonical drivers of kidney cancer, *MET* and *VHL*. These annotations include the coding regions and also transcription-factor binding sites (at the promoter and enhancer), and other regulatory elements. We will term the assembled annotations the 'METome' and 'VHLome'. We will then find all somatic and germline variants in these regions and rank them according to both functional impact and recurrence in cancer cohorts. Next, we will re-rank the variants by the extent of racial disparity. For particularly rare variants, we will re-perform our functional and disparity rankings in terms of the burdening of small sub-regions, such as binding sites or exons. This approach will result in a set of variants and burdened sub-regions that score highly both in terms of association with kidney cancer and racial disparities.

Aim 3: To validate specific variants and mutated regions and study clinical and environmental covariates suspected of contributing to kidney cancer racial disparities. In order to validate our discovered variants (Aim 2), we will form an independent cohort of African-American and Caucasian patients treated for ccRCC and pRCC. We will use the Connecticut State Tumor Registry, which includes a substantial number of individuals (n = 11,000) with RCC. This large cohort will allow us to risk-stratify patients by known kidney cancer risk factors and assure statistically rigorous results. We will assess tumor and normal DNA from candidate genomic regions using PCR-mass spectroscopy.

Aim 4: To perform functional characterization of a prioritized, high-confidence list of genetic variants. For a few of the highly-ranked variants in Aim 2 (in terms of association with RCC and racial disparity) that are also validated in Aim 3, we will perform a detailed functional characterization. In particular, we will genotype and gene edit immortalized cell lines using the CRISPR/Cas system to study the prioritized variants. Once matched cell lines are validated, they will allow us to fully interrogate how these variants affect the activity of the VHL/HIF and HGF/MET signaling pathways. We will utilize HIF reporter assays, gene expression, and protein abundance to discern changes in pathway activity. Finally, we will evaluate the effects of these alterations on cell proliferation, invasion, migration, and anchorage-independent growth.

Deliverables: This project will result in concrete data and a number of analytic deliverables: (1) genome sequences, completing of a missing aspect of TCGA, (2) annotation files for the VHLome and METome relating coding and non-coding regions of these important driver genes, (3) a list of variants affecting the VHLome and METome and a ranking of these variants, both in terms of functional impact and racial disparity, and (4) validation results reported as an additional annotation to the table of impactful variants. We will make all data, code, and information available through a project website and through standard repositories such as GitHub.