Biological and Genomic Differences of ERG Oncoprotein-Stratified Prostate Cancers from African and Caucasian Americans

email: gpetrovics@cpdr.org

Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc.
Bethesda, MD 20817

Nancy.Jackson@jacksonfoundation.org
Rockville, MD 20852

U. S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

Distribution Unlimited

14. ABSTRACT
There is striking disparity in prostate cancer incidence and mortality for African American men as compared to Caucasian Americans. The objective of this proposal is to determine associations and racial differences in the distribution of key clinico-pathologic patient features and molecular determinants for both ERG positive and ERG negative prostate cancer. It is anticipated that molecular determinants of aggressive prostate cancer in African American men, including somatic mutations and SNPs associated with prostate cancer risk, may associate with ERG fusion negative prostate cancer.

15. SUBJECT TERMS
Prostate cancer, health disparity, ERG oncogene, molecular stratification, germline variants (SNPs), admixture mapping, European and African ancestry, somatic mutations, aggressive cancer, nomograms

16. SECURITY CLASSIFICATION OF:
a. REPORT
U
b. ABSTRACT
U
c. THIS PAGE
U
17. LIMITATION OF ABSTRACT
UU
18. NUMBER OF PAGES
13
19a. NAME OF RESPONSIBLE PERSON
USAMRMC
19b. TELEPHONE NUMBER
301-619-3672
# Table of Contents

1. Introduction................................................................................. 1
2. Keywords...................................................................................... 1
3. Accomplishments........................................................................ 1
4. Impact.......................................................................................... 11
5. Changes/Problems...................................................................... 12
6. Products...................................................................................... 12
7. Participants & Other Collaborating Organizations............... 13
8. Special Reporting Requirements............................................ 14
9. Appendices................................................................................. 14
1. INTRODUCTION:

It is hypothesized that ERG gene fusion status of prostate tumors reflects the underlying biological or genetic differences of prostate cancer (CaP) incidence and/or progression between African American (AA) and Caucasian American (CA) patients. The objective of this proposal is to determine associations and racial differences of key clinico-pathologic features and SNPs for both ERG positive and ERG negative CaP. It is anticipated that molecular determinants of aggressive CaP in AA men include somatic mutations (TMPRSS2-ERG) and germline variants (SNPs).

The objective will be achieved by the following specific aims:

Aim 1: ERG-typing based molecular stratification of AA CaP patients. The goal of this aim is to establish our novel findings of lower ERG frequency in AA than in CA CaP, especially in tumors with high Gleason grade. ERG oncoprotein expression will be evaluated in whole-mounted prostates of 400 AA compared to 200 CA CaP patients.

Aim 2: Determine germline genetic determinants of the somatically acquired TMPRSS2-ERG fusion in AA men. We propose to use admixture mapping as it is particularly well suited for traits that present a sizeable difference in prevalence rates, such as the TMPRSS2-ERG fusion. Ancestry at each point in the genome in AA men will be estimated. Regions in AA genomes that are enriched for European ancestry in cases with the fusion compared to cases without the fusion will be captured. A total of 400 AA individuals with CaP will be genotyped and analyzed by HAPMIX program to infer local ancestry.

Aim 3: Define CaP driver mutations in ERG negative high grade tumors. Recently identified CaP driver mutations present in ERG negative CA-CaP will be directly assayed for, including SPOP mutation and SPINK1 overexpression. The PTEN/AKT pathway, which is often associated with aggressive CaP, will also be tested in this cohort by PTEN expression assay. Finally, we propose that the incorporation of ERG-typing, somatic mutations/markers in ERG-negative CaP, and ERG-type associated SNPs, will complement traditional pathological and clinical feature-based nomograms and lead to improved identification of aggressive CaP in AA patients.

Scope: This study will define the underlying biology and genetics of the ERG positive and ERG negative prostate tumors in AA and CA patients with special focus on the features of ERG negative aggressive CaP in AA patients.

2. KEYWORDS:

Prostate cancer, health disparity, ERG oncogene, molecular stratification, germline variants (SNPs), admixture mapping, European and African ancestry, somatic mutations, aggressive cancer, nomograms

3. ACCOMPLISHMENTS:

o What were the major goals of the project (as stated in the SOW)?

Major Task 1: ERG-typing based molecular stratification of AA CaP patients

Subtasks: ERG oncoprotein expression in 600 whole-mounted prostates from 400 AA compared to 200 CA CaP patients will be evaluated. The specimen cohorts will be identified from the CPDR tissue bank archive with up to 15 years follow-up time, excluding neo-adjuvant treated patients.

- IRB protocol approval
- Selection of AA and CA patient cohorts
- Identification of the archived whole mounted prostate specimens from the CPDR Tissue Bank
- Selection of the best representative blocks (includes index tumor and other tumor foci)
- Sectioning the blocks (10 unstained sections and an H&E stained section from each block)
• IHC with CPDR ERG MAb (clone 9FY)
• ERG IHC reading by pathologist
• Statistical analysis of the data
• Data interpretation, summary of Task 1 for manuscript

Timeline: Months 1-16

Major Task 2: Define germline genetic determinants of the somatically acquired TMPRSS2-ERG fusion in AA men

Subtasks: We propose to use admixture mapping to estimate ancestry at each point in the genome in AA men. Regions in AA genomes that are enriched for European ancestry in cases with the TMPRSS2-ERG fusion, compared to cases without the fusion, will be captured. A total of 400 AA individuals will be genotyped and analyzed by HAPMIX program to infer local ancestry.
• Blood genomic DNA specimens from the 400 AA CaP patients will be prepared (CPDR site).
• The DNA specimens from the 400 individuals will be genotyped on the Illumina Golden Gate genotyping platform.
• The HAPMIX program will be used for the analysis to infer local ancestry.
• Two statistical tests that are both implemented in HAPMIX will be utilized. The case-only admixture association (ADM) and sum of case-control SNP association and case-only admixture association (SUM) statistics will ensure the appropriate null distribution.
• Data interpretation, summary of Task 2 for manuscript

Timeline: Months 6-24

Milestone #1: Submit manuscript on Tasks 1 and 2

Major Task 3: Define CaP driver mutations in ERG negative high grade tumors

Subtasks: Somatic changes including expression (SPINK1), deletion (PTEN) and point mutations (SPOP) will be determined in the ERG negative subset of the 600-patient cohort. ERG-type associated SNPs and somatic markers will be assessed for improvement of prognostic nomogram.
• Unstained sections from the 600 blocks (400 from AA and 200 from CA patients) from Task 1a will be utilized in Task 3.
• Marker genes of pathways in aggressive CaP with ERG negative status will be tested in this cohort.
  a. SPINK1 overexpression will be assayed for by IHC following optimized procedure (Tomlins et al, 2008)
  b. PTEN expression will be determined by IHC assay (Lotan et al, 2011; Chaux et al, 2012)
  c. The stained slides will be read by our GU pathologist, and will also be quantified by specialized image analysis software (Definiens, Parsippany, NJ)
• SPOP mutations reported in CaP with ERG negative status will also be tested (Barbieri et al, 2012)
  a. Tumor areas from the whole mounted prostate tissue sections will be dissected with the ArcturusXT laser capture microdissection (LCM) Instrument (Life Technologies)
  b. DNA will be purification from the microdissected tissue and amplified by Whole Genome Amplification kit (WGA4), as suggested by the manufacturer for the single-cell approach (Sigma-Aldrich)
  c. Standard PCR will be used for targeted enrichment of SPOP exon 6 and exon 7 followed by sequencing.
d. Statistical analysis of the summarized data with clinical and pathological parameters focusing on disease progression will be performed by the biostatistician and the epidemiologist at CPDR

• Finally, ERG-typing, somatic mutations/markers in ERG-negative CaP, and ERG-type associated SNPs will be incorporated into the best available widely used postoperative prognostic nomogram to complement traditional pathological and clinical feature-based nomograms with the goal to improve identification of aggressive CaP in AA patients
  a. All SNPs and gene expression marker candidates (individually and in combinations) will be tested for their significance in multivariate statistical models (Cox analysis) in which the potential markers will be added to standard clinical variables
  b. The postoperative prognostic nomogram with and without a marker candidate will be assessed for improvement of the concordance index.

Timeline: Months 16-36
Milestone #2: Submit manuscript on Task 3

What was accomplished under these goals?

In this reporting period we focused on completion of Major Task 2, as well as the first parts of Major Task 3, as scheduled in the Statement of Work:

Major Task 2: Define germline genetic determinants of the somatically acquired TMPRSS2-ERG fusion in AA men

In this third year of the proposed project we continued to perform various analyses of the data generated under the originally stated Aim2 (Major Task 2). It primarily includes data analysis on 478,299 SNPs from a total of 321 patients obtained after applying initial sample and SNP QC on the datasets.

For genetic association studies in admixture population a crucial aspect to consider is that ancestry differences among the sampled individuals from admixed populations can be a confounder in genetic association studies. Failure to appropriately account for population structure due to ancestry admixture can lead to both spurious association (increased type-I error rates—false positive) as well as reduced power (inflated type-II error rates—false negative). The heterogeneous genomes of individuals from admixed populations, however, may provide advantages over genetic association analyses in homogenous populations, including the possibility of gene mapping by admixture linkage disequilibrium, i.e., admixture mapping.

Association between the SNPs and ERG positive tumors was estimated using logistic regression for all the patients including African American (AA) and Caucasian American (CA).

Figure 1. Q-Q plots showing expected and observed distribution of SNPs in the ERG fusion positive versus fusion negative patient cohort by index tumor (Fusion positive, N=108; Fusion negative, N=180) (left panel) and positive for ERG patient cohort by any tumor foci (Fusion positive, N=158; Fusion negative, N=133) (right panel).
The QQ plot shows the expected distribution of association test statistics (X-axis) across the SNPs compared to the observed values (Y-axis). Any deviation from the X=Y line implies a consistent difference between cases and controls across the whole genome. As expected population structure was found to be a confounder. **High false positive rate** was observed based on logistic regression analysis for all subjects (Figure 1).

Therefore, meta-analysis using GWAMA software tool was performed to account for the difference in population structure and minimize the false positive rate. Principal Component Analysis was performed using the EIGENSTRAT software tool. The approach is widely applied to genome-wide association studies for protection against confounding due to ancestry differences among individuals.

![Figure 2.](image)

Some of the AA (N=17) AA and CA (N=2) subjects were excluded based on <50% of the respective ancestry. The population structure was again found to be a confounder (Figure 2). **This time a high false negative rate** was observed based on meta-analysis accounting for population structure.

**To correct for the confounder** a variance component approach called **Efficient Mixed-Model Association eXpedited (EMMAX)** was used. This approach is based on pair wise relatedness between individuals, using high-density markers to model the phenotype distribution. EMMAX implements linear mixed model approach for association testing, accounting for global population substructures with an empirical covariance matrix.

![Figure 3.](image)

After correcting for population structure (Figure 3), EMMAX analysis identified several SNPs to be associated with the ERG status either by index tumor (Fusion positive, N=103; Fusion negative, N=165) (Figure 4), or by any tumor foci (Fusion positive, N=150; Fusion negative, N=122) patient cohorts (Figure 5).
The three SNPs most significantly (p<10^{-5}) associated with ERG fusion status by index tumor, were \textit{rs6698333, an intron variant of Kruppel-like factor 17 (KLF17) and two SNPs, rs1889877 and rs3798999, in the intron of adhesion G protein-coupled receptor B3 (ADGRB3).}

Krüppel-like factor (KLF) family is highly conserved zinc finger transcription factors that regulate cell proliferation, differentiation, apoptosis, and migration. Reduced-KLF17 in human cancer is involved in TGF-β pathway and p53 pathway. ADGRB3 is a p53 target gene encodes a brain-specific angiogenesis inhibitor and is a member of the secretin receptor family and is involved in angiogenesis. Mutations in ADGRB3 are found to be associated with lung and pancreatic cancers.

\textbf{Figure 4.} Association analysis (EMMAX) of SNPs with ERG positive index tumor (N=103) vs. ERG negative index tumor (N=165). A total of 478,299 SNPs are plotted against their respective positions on the chromosomes.

\textbf{Figure 5.} Association analysis (EMMAX) of SNPs with any tumor foci positive for ERG phenotype in a total of 150 ERG positive and 122 ERG negative cases. A total of 478,299 SNPs are plotted against their respective positions on the chromosomes.
Four SNPs (rs10215144, rs3818136, rs9380660 and rs1792695) were found to be significantly (p<10^{-5}) associated with ERG positive phenotype by any tumor foci positive for ERG fusion. Rs3818136 is a synonymous variant and rs9380660 is downstream variant of TBC1D22B, a GTPase activating protein for Rab family, required to maintain a functional Golgi complex in human cells. Rs10215144 is an intron variant in AGBL3, ATP/GTP binding protein-like protein, and rs1792695 is an intron variant in ncRNA, LOC100505474.

Currently further fine-mapping of the genetic associations by genotype imputation analysis of genome-wide Oncoarray (500,000 SNP) data is being performed using IMPUTE2 approach with the 1000 Genomes reference dataset.

To estimate the ancestral diversity, global ancestry analysis based on admixture mapping was performed on AA cases (Fusion positives=90; Fusion negatives=100) using the ADMIXTURE software. It is based on maximum likelihood estimation of individual ancestries relying on multiple single nucleotide polymorphisms. It was evaluated by comparing the proportion of YRI/CEU alleles across the chromosomes between TMPRSS2-ERG fusion carriers and non-carriers.

The analysis revealed that ERG fusion positive phenotype showed a significant enrichment for CEU alleles in comparison to AA allelic enrichment in ERG negative tumors.

We are currently working on the identification of genes and functional connections linking the defined chromosomal areas and SNPs to ERG fusion generation.

Based on the summary of Tasks 1 and 2 results, we are in the process of writing a manuscript on the identified germline genetic determinants of the somatically acquired TMPRSS2-ERG fusion in AA men.

Major Task 3: Define CaP driver mutations in ERG negative high grade tumors
Subtask: Somatic changes in the expression of SPINK1 will be determined in the ERG negative subset of the patient cohort.

In this reporting period we aimed to describe the expression of SPINK1 in the context of ERG status, which was determined earlier (Major Task 1 in the first reporting period) in higher grade prostate cancer, stratified by race. The Center for Prostate Disease Research database was
queried to identify patients with higher grade disease (Gleason score of 8-10 or 4+3) who underwent radical prostatectomy, and clinical data from over 1,300 patients were evaluated. A total of 63 African Americans (AA) met study criteria and 63 Caucasian Americans (CA) were matched against them. Immunohistochemistry (IHC) was performed to detect SPINK1 protein expression in representative whole-mount prostate specimens (N=126).

The index tumor in CA men was ERG positive in 31 of 63 patients (49%), which was significantly more common than the 10 of 63 patients (16%) in AA men (p<0.0001), as we described in our first report. Multivariable analysis confirmed that ERG was not an independent predictor of BCR in this cohort.

SPINK1 was present in two-thirds of the index tumors in AA patients. AA men were more likely to harbor SPINK1 than CA men and this observation was strongest in the index tumors (65% vs. 49%, P=0.07) (see table below).

<table>
<thead>
<tr>
<th>Molecular Abnormality</th>
<th>Total</th>
<th>CA</th>
<th>AA</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERG+ (Index)</td>
<td>33% (41/126)</td>
<td>49% (31/63)</td>
<td>16% (10/63)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>SPINK1+ (Index)</td>
<td>57% (72/126)</td>
<td>49% (31/63)</td>
<td>65% (41/63)</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Regarding the ERG-negative patients, 33 AA men were SPINK1 positive in the index tumor compared to 13 CA men (a 2.5-times difference). These results are in line with our hypothesis, that in AA men with aggressive (high Gleason grade) ERG negative prostate cancer the frequency of SPINK1 expression is greatly elevated compared to CA men.

In general, SPINK1 positivity was identified in clonal cells, was focal, and was often present in a small percentage of the tumor, usually clustered in the periphery of tumors (Figure 7).

The prevalence of SPINK1 positivity in this study (50-65%) is much greater than previously reported in tissue microarray (TMA) based studies (15-35%). This is likely due to the focal nature of SPINK1 staining, which is fully evaluable in our whole mounted prostate sections, but TMA sampling may often miss a smaller SPINK1 positive area.

SPINK1 is thought to mediate its effects through interaction with epidermal growth factor receptor (EGFR) (Ohmuraya et al, 2001; Ateeq et al, 2011; Ozaki et al, 2009). Previous studies...
have found that EGFR overexpression is prevalent in CaP (Schlomm et al, 1998), and it has a role in the development and progression of CaP (Sherwood et al, 1998; Lorenzo et al, 2003). Ateeq and colleagues used an in vitro model and showed that SPINK1 was associated with increased cell invasion compared with untreated cells while knockdown of SPINK1 inhibited cell proliferation, invasion, and tumor growth in xenograft assays (Ateeq et al, 2011). Use of cetuximab, a monoclonal antibody to EGFR, demonstrated a 60% reduction in tumor growth. Interestingly, overexpression of EGFR in CaP was noted to be significantly more common in AA than CA men (Shuch et al, 2004). The combination of this finding with the overlap of SPINK1 involvement in EGFR pathways, and the high prevalence of SPINK1 in AA men does provide a potential pathway for the higher incidence and more aggressive nature of CaP in AA men. This study underscores, that molecular typing may enhance our understanding of ethnic differences in prostate cancer biology and outcome.

- What opportunities for training and professional development has the project provided? Nothing to report
- How were the results disseminated to communities of interest? Nothing to report
- What do you plan to do during the next reporting period to accomplish the goals?

We received a one year no cost extension for this study. We will focus on the remaining analyses of the identified gene areas under Major Task 2 and submit a manuscript. We will also complete the remainder of Major Task 3 as defined in the SOW, including the evaluation of point mutations in SPOP and deletions in PTEN in the ERG negative subset of the our cohort, to submit as a second manuscript. ERG-type associated SNPs and somatic markers will be assessed for improvement of prognostic nomogram.

4. IMPACT:
- What was the impact on the development of the principal discipline(s) of the project?

Our findings on the predominance of ERG negative high grade prostate cancer in AA men, compared to CA men, impact the area of genetic aspects of racial disparity in prostate cancer. Our unique patient cohort, treated by radical prostatectomy at the Walter Reed National Military Medical Center, is within the equal access DOD healthcare beneficiary system. In this system socio-economic factors influencing disparity are less pronounced leaving genetic factors easier to pinpoint. The finding that the expression of ERG, a major early oncogene in prostate cancer, was significantly (3 times) more likely to be present in the higher grade index tumors of CA men compared to AA men in a tightly matched cohort of 126 patients (Farrell et al, 2014) clearly supports that besides socio-economic factors the somatic genetic events in the prostate tissue may also be different between ethnic groups potentially impacting racial disparity of the disease.

Defining germline genetic determinants of the somatically acquired TMPRSS2-ERG fusion in AA men may impact on stratification of patients based on the relevant SNPs very early, even before prostate cancer is detected. In addition, this knowledge impacts research on understanding the biological mechanism of how the gene fusions are generated.

- What was the impact on other disciplines?
5. CHANGES/PROBLEMS:
   o Changes in approach and reasons for change
     Nothing to report
   o Actual or anticipated problems or delays and actions or plans to resolve them
     A one year no cost extension was requested and recently approved. The reasons for the extension: Delays due to lengthy IRB approval process which resulted also a delay in the hiring of the postdoctoral fellow; the scheduled data analysis in Major Task 2 turned out to be more complex than expected causing some additional delays, cumulatively putting the grant almost one period of performance (period 3) behind (the cost for the remaining tasks is still available in the original budget).
     Nothing to report
   o Changes that had a significant impact on expenditures
     Nothing to report
   o Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents
     Nothing to report
   o Significant changes in use or care of human subjects
     Nothing to report
   o Significant changes in use or care of vertebrate animals
     Nothing to report
   o Significant changes in use of biohazards and/or select agents
     Nothing to report

6. PRODUCTS:
   o Publications, conference papers, and presentations
     Journal publications.
     Abstract will be submitted for AACR or/and AUA meetings.
- **Books or other non-periodical, one-time publications.**
  
  Nothing to report

- **Other publications, conference papers, and presentations.**
  
  Nothing to report

- **Website(s) or other Internet site(s)**
  
  Nothing to report

- **Technologies or techniques**
  
  Nothing to report

- **Inventions, patent applications, and/or licenses**
  
  Nothing to report

- **Other Products**
  
  Nothing to report

### 7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS:

- **What individuals have worked on the project?**

  **Gyorgy Petrovics, PI,** (2.4 person months) The PI provides the overall organization for the execution of the specific aims. He coordinated the selection and processing of patient tissue specimens in close collaboration with Dr. Sesterhenn, and coordinates with Dr. Freedman the genotyping efforts and with Dr. Cullen the data analyses. He closely supervises the Staff Scientist (Dr. Indu Kohaar), Ms. Young’s and Ms. Ravindranath’s experimental work related to this proposal.

  **Matthew Freedman, Qualified Collaborator,** (1.2 person months) Oversees and organizes the genotyping operations in close collaboration with the PI.

  **Denise Young, Histology Technologist,** (1.4 person months) Manages, prepares, and maintains the histologic preparations using state-of-the-art histopathology and molecular pathology procedures pertinent to this proposal under the directions of the PI and Dr. Sesterhenn. Ms. Young performs histological procedures and analytical procedures including tissue sectioning, staining and mounting specimens on slides, immunohistochemistry (IHC) staining of whole mounted prostate sections and optimizing procedures to assure the successful outcome of the proposed experiments.

  **Indu Kohaar, Staff Scientist,** (6.0 person months) Dr. Kohaar has experience in, and performs, mutation and SNP analysis, IHC assays, QRT-PCR experiments and DNA analysis with selected markers for this proposal under the direction of the PI.

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**
Nothing to report

- **What other organizations were involved as partners?**
  Nothing to report

8. **SPECIAL REPORTING REQUIREMENTS:**
   Nothing to report

9. **APPENDICES:**
   Nothing to report