AWARD NUMBER:
W81XWH-13-2-0096

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

PRINCIPAL INVESTIGATOR:
Gyorgy Petrovics, PhD

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

REPORT DATE:
October 2015

TYPE OF REPORT:
Annual

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DA TE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DA TES COVERED *(From - To)*
October 2015	Annual	9-30-2014 - 9-29-2015

4. TITLE AND SUBTITLE
Biological and Genomic Differences of ERG Oncoprotein-Stratified Prostate Cancers from African and Caucasian Americans

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-13-2-0096

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Gyorgy Petrovics. email: gpetrovics@cpdr.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc.
6720-A Rockledge Drive Suite 100
Bethesda, MD 20817

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Distribution Unlimited

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Unlimited

13. SUPPLEMENTARY NOTES
N/a.

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:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	12	

Table of Contents

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:**
   - **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 bio-statistician 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 scheduled in the Statement of Work:

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

We proposed 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, are being captured. A total of **400 individuals are being genotyped and analyzed** by HAPMIX program to infer local ancestry.

Blood genomic DNA specimens from 400 CaP patients were prepared at CPDR. The DNA specimens from the 400 individuals were genotyped on the Illumina Golden Gate genotyping platform using Infinium Oncoarray, a 500K BeadChip kit (Cat No.- WG-355-1002) from Illumina on a genome wide scale (BGI Americas).

The oncoarray contains approximately 500,000 SNPs with genome wide backbone of 275,000 tag SNPs. It includes SNPs covering common ancestry, genetic variants associated with 5 common cancers including breast, colorectal, lung, ovarian and prostate cancers plus SNPs covering quanitative traits, pharmacogenetics, and fine mapping of common cancer susceptibility loci.

First an overall **QC of the data** was performed. 2516 SNPs were excluded while merging three genotype files (corresponding to 3 batches of sample run) due to batch discrepancies issues. The merged file had 496654 SNPs genotype calls from 402 individuals.
The following further QC filters were applied on the dataset:

(1) Additional 14672 SNPs were excluded due to SNP call rate < 0.90
(2) Additional 3683 SNPs were excluded with Hardy–Weinberg Equilibrium Test Statistics P-value < 10^{-8}
(3) 48 samples were excluded due to sample call rate < 0.95 and gender discrepancies
(4) Additional 24 samples were excluded due to unexpected sample contaminations, swaps and duplications as well as unknown familial relationships (plink proportion identity-by-descent (PI_HAT) > 0.15)

The final merged QC filtered file had 478299 SNPs genotype calls from 330 individuals. The expected and observed distribution of SNPs showed strong correlation (Figure 1).

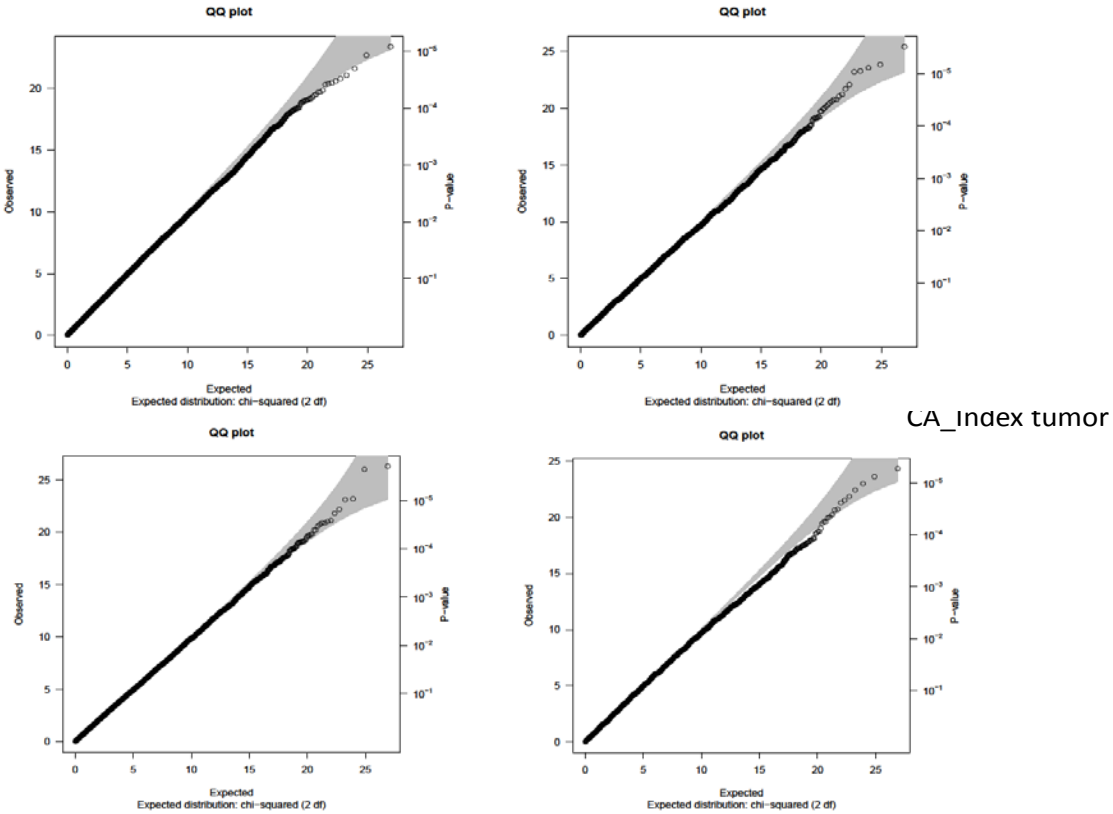

Figure 1. Q–Q plot of the genome-wide association results
Logistic regression (additive model) was used to test the association between genetic variants and ERG positive tumors in the AA (left panel) and CA (right panel) patient cohort. Upper panels: association with ERG positive status for any tumor foci; lower panel: association with ERG positive status for index tumor.

Currently the main **admixture mapping and genome-wide association analysis** is in progress. Examples of admixture ancestry plots along chromosome 13 and chromosome 1 illustrate areas of increased CA (decreased AA) ancestry associating with ERG positive patients (Figure 2).

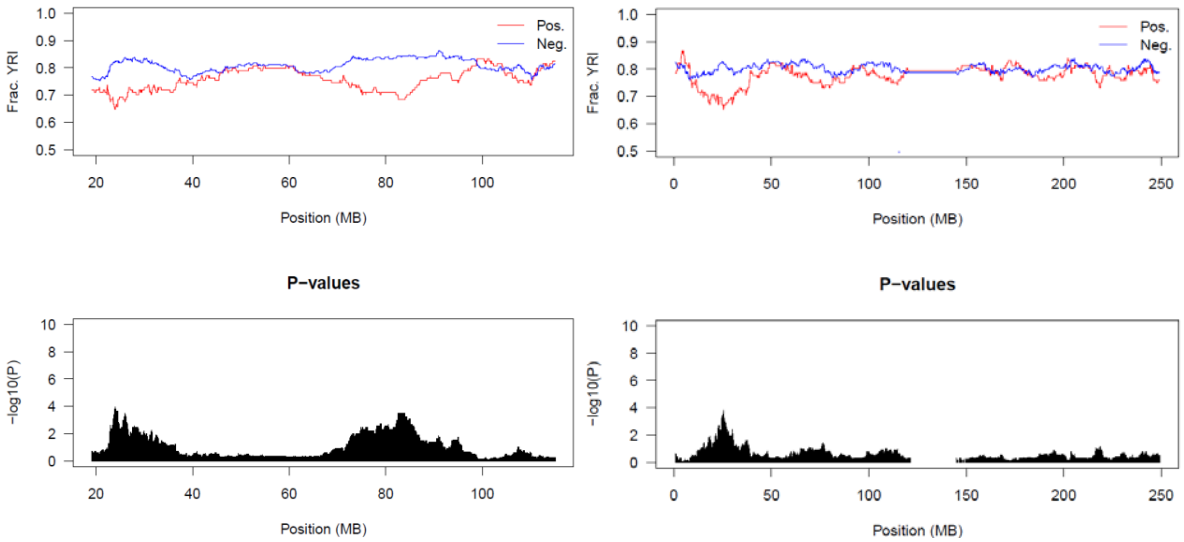

Figure 2. Representative admixture scans for chromosome 13 (left panel) and chromosome 1 (right panel) based on a logistic regression analysis using TMPRSS2-ERG fusion status as

dependent variable and the number of YRI alleles (0, 1, 2) as independent variable in the AA samples. The admixture scan based on the index tumor ERG fusion status includes a total of 203 samples (Fusion positive- 57; Negative=146) for chromosome 13 and 202 samples (Fusion positive- 56; Negative=146) for chromosome 1.

Association of SNPs with ERG status is summarized below.

African American cohort:
Association between SNPs and ERG positive tumors was estimated using logistic regression separately for AA and CA patients by the odds ratio (OR) and 95% confidence interval (CI) using the multivariate logistic regression assuming a trend genetic model.

(1) Association analysis with any tumor foci positive for ERG status
A total of 94 samples were ERG positive and 104 samples were found to be ERG negative for any tumor foci. Additional 83867 SNPs were excluded due to minor allele frequency (MAF) < 0.01. A total of 27 SNPs were found to be significantly associated with ERG phenotype ($p \leq 10^{-5}$).

(2) Association analysis with index tumor positive for ERG status
A total of 53 samples were ERG positive for index tumor and 143 samples were found to be ERG negative for index tumor. A total of 31 SNPs were found to be significantly associated with ERG phenotype ($p \leq 10^{-5}$) (Table 1).

CHR	Tumor Type (ERG positive)	SNP	BP	A1	F_A	F_U	A2	CHISQ	P	OR
2	Index Tumor	rs10173073	80189383	G	0.566	0.3042	A	22.62	1.97E-06	2.984
20	Index Tumor	chr20_62845290_C_INDEL	62845290	D	0.1509	0.02448	I	22.4	2.22E-06	7.086
5	Index Tumor	rs12653109	38089107	T	0.5943	0.3462	C	19.65	9.32E-06	2.767
13	Index Tumor	rs9580448	23416203	A	0.217	0.06294	G	19.6	9.57E-06	4.126
8	Any Tumor	rs2703335	69714947	A	0.2632	0.5795	G	20.7	5.38E-06	0.2591

Table 1. Genetic variants most significantly ($p \leq 10^{-6}$) associated with ERG status in the AA cohort
CHR- Chromosome; SNP- SNP ID; BP- Physical position (base-pair); A1- Minor allele; F_A- Frequency of this allele in cases; F_U- Frequency of minor allele in controls; A2- Major allele name; CHISQ- Basic allelic test chi-square (1df); P- Asymptotic p-value for this test; OR - Estimated odds ratio.

Caucasian American cohort:

(1) Association analysis with any tumor foci positive for ERG status
A total of 68 samples were ERG positive and 34 samples were found to be ERG negative for any tumor foci. Additional 85675 SNPs were excluded due to minor allele frequency (MAF) < 0.01. A total of 29 SNPs were found to be significantly associated with ERG phenotype ($p \leq 10^{-5}$).

(2) Association analysis with index tumor positive for ERG status
A total of 57 samples were ERG positive for index tumor and 44 samples were found to be ERG negative for index tumor. A total of 20 SNPs were found to be significantly associated with ERG phenotype ($p \leq 10^{-5}$) (Table 2).

CHR	Tumor Type (ERG positive)	SNP	BP	A1	F_A	F_U	A2	CHISQ	P	OR
6	Index Tumor	rs3798999	69714947	A	0.2632	0.5795	G	20.7	5.38E-06	0.25
13	Index Tumor	rs282606	1.13E+08	A	0.02679	0.2326	G	20.06	7.49E-06	0.09
4	Any Tumor	rs10013727	1.43E+08	A	0.007353	0.1765	G	21.73	3.14E-06	0.03
6	Any Tumor	rs3818136	37252210	G	0.2647	0.5882	A	20.3	6.62E-06	0.25
6	Any Tumor	rs9380660	37305622	T	0.2794	0.6029	G	20	7.76E-06	0.25
9	Any Tumor	rs12352937	23081882	G	0.4632	0.1471	A	19.72	8.96E-06	5
13	Any Tumor	rs282606	1.13E+08	A	0.04478	0.2576	G	19.68	9.18E-06	0.13

Table 2. Genetic variants most significantly ($p \leq 10^{-6}$) associated with ERG status in CA cohort
CHR- Chromosome; SNP- SNP ID; BP- Physical position (base-pair); A1- Minor allele; F_A- Frequency of this allele in cases; F_U- Frequency of minor allele in controls; A2- Major allele name; CHISQ- Basic allelic test chi-square (1df); P- Asymptotic p-value for this test; OR - Estimated odds ratio.

Genome-wide overview of SNP associations with ERG status is shown by Manhattan plot (Figure 3).

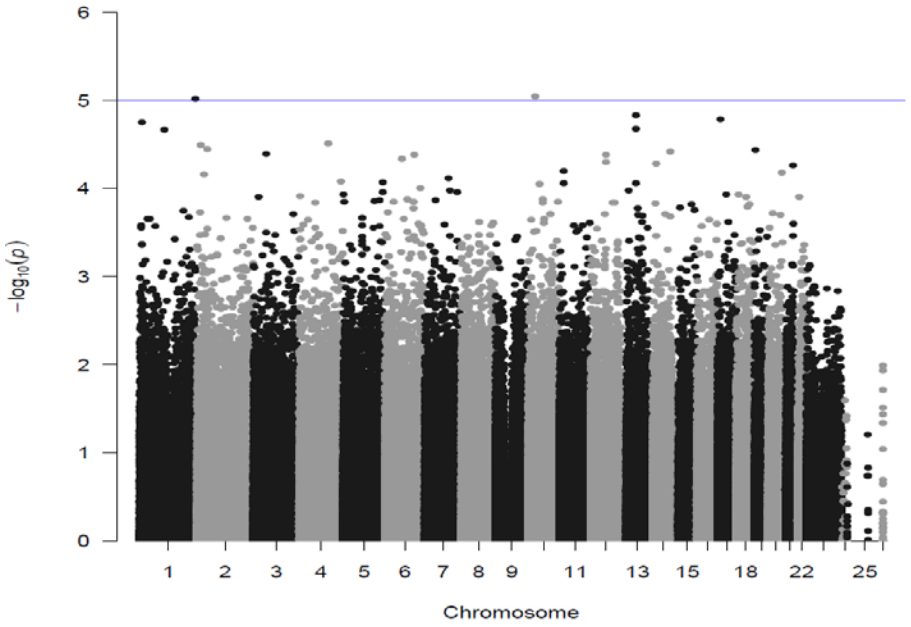

Figure 3. A Manhattan plot showing the -log10 P values from the association analysis of the ERG fusion positive and negative samples in all AA and CA patients. A total of 478,299 SNPs are plotted against their respective positions on the chromosomes.

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

- **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?**

  Focus on the completion of **Major Task 3:** Define CaP driver mutations in ERG negative high grade tumors.

  Subtasks: As defined in the SOW, 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.

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?**

     Nothing to report

   - **What was the impact on technology transfer?**

     Nothing to report

   - **What was the impact on society beyond science and technology?**

     Nothing to report

5. **CHANGES/PROBLEMS:**
   - **Changes in approach and reasons for change**

     Nothing to report

   - **Actual or anticipated problems or delays and actions or plans to resolve them**

     Nothing to report

- **Changes that had a significant impact on expenditures**

  Nothing to report

- **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

  Nothing to report

- **Significant changes in use or care of human subjects**

  Nothing to report

- **Significant changes in use or care of vertebrate animals**

  Nothing to report

- **Significant changes in use of biohazards and/or select agents**

  Nothing to report

6. **PRODUCTS:**
   - **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 postdoctoral fellow's (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 incl uding tissue sectioning, staining and mounting specimens on slides, immunihistochemistry (IHC) staining of whole mounted prostate sections and optimizing procedures to assure the successful outcome of the proposed experiments.

     Indu Kohaar, Postdoctoral Fellow, (6.0 person months) Dr. Kohaar has experience in, and performs, mutation and SNP analysis, IHC assays, QRT-PCR experiments and bDNA 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