ANDROGEN RECEPTOR CAG REPEAT POLYMORPHISM AND RISK OF TMPRSS2:ERG POSITIVE PROSTATE CANCER

Sun Yoo1*, Andreas Pettersson1*, Kristina M. Jordahl1, Rosina T. Lis2,3, Sara Lindstrom1, Allison Meisner1, Elizabeth J. Nuttall1, Edward C. Stack2,3, Meir J. Stampfer1,4,5, Peter Kraft1,6, Myles Brown7, Massimo Loda2,3, Edward L. Giovannucci1,4,5, Philip W. Kantoff7,8, Lorelei A. Mucci1,5

1Department of Epidemiology, 4Nutrition, and 6Biostatistics, Harvard School of Public Health, Boston MA
2Departments of Pathology and 7Medical Oncology, Dana-Farber Cancer Institute, Boston MA
3Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
5Channing Division of Network Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston MA
8Lank Center for Genitourinary Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston MA
* These authors contributed equally to the manuscript

Correspondence:
Lorelei Mucci, ScD and Andreas Pettersson, MD, PhD
Department of Epidemiology
Harvard School of Public Health
677 Huntington Avenue, 9th floor
Boston MA 02115
617.432.1732 (tel)
617.566.7805 (fax)
lmucci@hsph.harvard.edu
apetters@hsph.harvard.edu

The authors have no conflicts of interest to disclose.

Running Title: Androgen receptor CAG repeats in TMPRSS2:ERG prostate cancer

Keywords: Prostate cancer, androgen receptor, TMPRSS2:ERG fusion, genetic polymorphism, case-control
Funding: Funding from this project was supported by the Dana-Farber / Harvard Cancer Center Specialized Programs of Research Excellence (SPORE) in Prostate Cancer P50CA090381-08 to P.W Kantoff and E.L. Giovannucci; the National Cancer Institute CA136578 to L.A. Mucci and E.L. Giovannucci, CA141298 to M.J Stampfer, CA097193 to J. M. Gaziano, PO1 CA055075 to W. Willett and E.L. Giovannucci, UM1CA167552 to W. Willett, U01CA098233 to D.J. Hunter, P. Kraft, and S. Lindstrom; L.A. Mucci is supported by the Prostate Cancer Foundation; A. Pettersson is supported by the Swedish Research Council 2009-7309 and the Royal Physiographic Society in Lund, Sweden.

Word count: 2,138  Tables: 2  Supplemental Tables: 1
ABSTRACT

Background: The androgen receptor (AR) is an essential gene in prostate cancer pathogenesis and progression. Genetic variation in AR exists, including a polymorphic CAG repeat sequence that is inversely associated with transcriptional activity. Experimental data suggest that heightened AR activity facilitates formation of TMPRSS2:ERG, a gene fusion present in approximately 50 percent of tumors of prostate cancer patients.

Methods: We undertook a nested case-control study to investigate the hypothesis that shorter CAG repeat length would be associated with prostate cancer risk defined by TMPRSS2:ERG status. The study included 291 men with prostate cancer (147 ERG-positive) and 1,221 cancer-free controls. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated using logistic regression.

Results: Median CAG repeat length (Interquartile range) among controls was 22 (20-24). Men with shorter CAG repeats had an increased risk of ERG-positive (OR=1.07 per 1 repeat decrease, 95% CI 1.00-1.14), but not ERG-negative prostate cancer (OR=0.99 per 1 repeat decrease, 95% CI 0.93-1.05).

Conclusions: These data suggest that shorter CAG repeats are specifically associated with development of TMPRSS2:ERG positive prostate cancer.

Impact: Our results provide supportive evidence that androgen signaling underlies the development of prostate tumors that harbor TMPRSS2:ERG. Moreover, these results suggest that TMPRSS2:ERG may represent a unique molecular subtype of prostate cancer with an etiology distinct from TMPRSS2:ERG negative disease.
INTRODUCTION

The androgen receptor (AR) is a nuclear transcription factor that mediates the actions of testosterone and dihydrotestosterone. AR signaling is critical for prostate growth and maintenance. In the context of cancer, altered AR signaling is implicated in prostate cancer development and almost all prostate tumors depend on androgens for growth. Located on the Y chromosome, AR contains multiple genetic variants including the length of a polymorphic CAG repeat sequence in exon 1. This CAG repeat encodes a polyglutamine tract in the terminal domain of the protein (1). Variability in the trinucleotide repeat length has functional consequences and is inversely correlated to transcriptional activity of AR (2).

Shorter versus longer CAG repeats have been associated with higher risk of prostate cancer overall in some but not all studies (3-7). The role of the AR CAG repeat is of interest in the context of TMPRSS2:ERG, the gene fusion present in the tumors of about half of Caucasian prostate cancer patients. This common fusion event involves the androgen-regulated promoter TMPRSS2 and the ETS transcription factor family member ERG (8). Prior data suggest that TMPRSS2:ERG is an early event in prostate cancer development and that fusion positive cancer represent a molecularly distinct subgroup (9). In experimental models, androgen signaling induces spatial proximity of the TMPRSS2 and ERG genomic loci, both located on chromosome 21q22 (10). Subsequent treatment with gamma irradiation, which causes DNA double-strand breaks, leads to formation of the TMPRSS2:ERG gene fusion, suggesting that heightened AR signaling can facilitate TMPRSS2:ERG formation. Supporting this notion, in a case-only study of 40 men with prostate cancer, Bastus et al found suggestive evidence that the number of CAG repeats is lower in TMPRSS2:ERG positive versus negative prostate cancer (11).
Androgen receptor CAG repeats in TMPRSS2:ERG prostate cancer

We sought to extend this work to investigate whether CAG repeat length in AR is related to the risk of TMPRSS2:ERG positive or negative prostate cancer in a nested case-control study among 291 men with prostate cancer and 1,221 cancer-free controls. We further tested whether six single-nucleotide polymorphisms (SNPs) capturing variation across AR are associated with risk of TMPRSS2:ERG positive or negative prostate cancer.

MATERIALS AND METHODS

Study population. We included men from previously conducted, prospective case-control studies nested within the Physicians’ Health Study (PHS) and the Health Professionals Follow-up Study (HPFS) cohorts for whom genotyping and tumor data was available (4). The PHS was initiated in 1982 as a randomized trial of 22,071 U.S. male physicians age 40-84 years who were free of cardiovascular disease and cancer. HPFS is an ongoing prospective study of 51,529 male health professionals age 50-75 years initiated in 1986. In both studies, blood specimens were collected from a subset of the men prior to cancer diagnosis (N=14,500 in PHS; N=18,000 in HPFS) and DNA was extracted from whole blood and stored.

The original case-control studies that measured genetic variants in AR included 1,423 incident prostate cancer cases diagnosed 1982 to 2004 and 1,467 matched controls (4-6). In the PHS samples, we previously reported that shorter CAG repeats were related to higher overall risk of prostate cancer (5) but found no association in the HPFS (6). Of the 1,423 cases, we also had available archival tumor tissue for characterization of TMPRSS2:ERG for 291 men. Because more than 96% of participants in the PHS and HPFS are Caucasian, the study was restricted to men of European Ancestry. The final sample size for this investigation included 291 with prostate cancer and known TMPRSS2:ERG status, and 1,221 cancer-free controls. Written consent had previously been obtained from all participants and the
Institutional Review Boards of Partners HealthCare and the Harvard School of Public Health approved the study.

**Genotyping.** CAG repeat length in AR was determined by PCR, running the amplified fragments on a denaturing polyacrylamide gel with automated fluorescence detection of the fragments and sizing (Genescan) at the Dana-Farber Cancer Institute (5). Data on the CAG repeat length was available for 269 of the prostate cancer cases and 1154 of controls. Six polymorphic AR variants (rs6152, rs962458, rs1204038, rs2361634, rs1337080, rs1337082) were selected to capture the haplotype variation in the study population using the Tagger program. Genotyping was done using the fluorogenic 5'-endonuclease assay (TaqMan) with the ABI Prism 7900 (Applied Biosystems, Foster City, CA) at the Harvard School of Public Health (4). Replicate samples were included in all genotyping assays to assess quality control and showed excellent concordance of genotyping data.

**TMPRSS2:ERG status in tumors.** Presence or absence of TMPRSS2:ERG was characterized on tumor tissue available from a biorepository of archival radical prostatectomy (95%) and transurethral resection of the prostate (TURP, 5%) tumor specimens for men with prostate cancer in PHS and HPFS. For each case, the pathology team reviewed hematoxylin and eosin (H&E) slides to confirm presence of prostate cancer, assign standardized Gleason grade, and to identify areas of tumor for construction of tissue microarrays. Tissue microarrays were constructed by sampling at least three 0.6 mm cores of tumor per case from the dominant nodule or nodule with the highest Gleason pattern. To characterize TMPRSS2:ERG status, we performed immunohistochemistry on 5 micron sections of tissue microarrays to assess ERG protein tumor expression (12), which has been shown to be strongly correlated with fusion status assessed by fluorescent in situ hybridization (FISH) (13). Briefly, 5 micron sections were deparaffinized and treated with citrate buffer for antigen retrieval. We applied ERG
antibody (Clone ID: EPR3864, Epitomics, CA) at 1:100 for 1 hour, followed by the BioGenex SS Multilink secondary antibody and visualized using the DAB substrate kit (Vector Laboratories, CA). We classified tumors as ERG-positive (i.e., carrying the TMPRSS2:ERG gene fusion) if at least one core stained positive for ERG and ERG-negative if all cores stained negative for ERG. Eighty-five percent of the ERG positive cases were positive on all tissue microarray cores. On a subset of cases in the PHS for whom TMPRSS2:ERG status was also assessed by FISH, we saw high agreement (>93%) with immunohistochemical evaluation.

**Statistical analysis.** Unconditional logistic regression analysis was used to calculate odds ratios (ORs) and 95% confidence intervals (CIs) of the association between CAG repeat length (continuous, and categorical: ≤19, 20-21, 22-23, ≥24 - reference) and prostate cancer risk by ERG status, comparing ERG-positive cases to cancer-free controls. Additionally, we assessed associations with the six AR SNPs (binary, minor allele as referent). All analyses were adjusted for age at blood draw (continuous). We used the method described by Altman and Bland (14) to test for the statistical interaction between CAG length (continuous) and prostate cancer risk by ERG tumor status.

**RESULTS**

Clinical characteristics of the 291 prostate cancer cases by ERG status are shown in **Table 1**. The prevalence of TMPRSS2:ERG among the cases was 51%. Mean age at cancer diagnosis was 66 years, and 18% of men had pathologic Gleason 8 – 10 tumors. In this set of cases, the clinical characteristics were similar for men who had ERG-positive compared to ERG-negative prostate cancer. The mean (standard deviation) age at blood draw for the controls was 63.2 (5.0) years. Among cases, the mean follow-up from blood draw to
Androgen receptor CAG repeats in TMPRSS2:ERG prostate cancer

diagnosis was 6.4 years, and was similar for ERG positive (6.5 years) and ERG negative (6.4 years) prostate cancer.

The median (interquartile range, IQR) CAG repeat length among controls was 22 (IQR, 20-24) repeats, compared to 21 repeats (IQR, 20-23) for ERG-positive prostate cancer and 22 repeats (20-24) for ERG-negative disease. Men with shorter AR CAG repeats had a higher risk of developing ERG-positive prostate cancer (OR=1.07 per 1 repeat decrease, 95% CI 1.00-1.14), whereas there was no association between CAG repeat length and risk of ERG-negative prostate cancer (OR=0.99 per 1 repeat decrease, 95% CI 0.93-1.05) (Table 2). The test for heterogeneity of CAG repeat by ERG tumor status was borderline significant, p=0.06. Compared to those with the longest (≥24) CAG repeat length, men in the category of shortest (≤19) CAG repeats had a nonsignificant 45% increased risk of ERG-positive prostate cancer (OR=1.40, 95% CI 0.82-2.39), albeit not statistically significant; the corresponding odds ratio for ERG-negative prostate cancer was 0.96 (95% CI 0.57-1.61). In the case only comparison, the odds ratio for ERG positive vs. negative prostate cancer was 1.08 (95% CI 0.99 – 1.20) per 1 shorter CAG repeat.

We observed no significant association between any of the six polymorphic AR genetic variants (rs962458, rs6152, rs1204038, rs2361634, rs1337080, rs1337082) and risk of either ERG-positive or ERG-negative prostate cancer (Supplemental Table 1).

DISCUSSION

In this integrative patho-epidemiology study, we found shorter germline CAG repeat length in AR to be associated with higher risk of ERG-positive prostate cancer, whereas there was no association between CAG repeat length and risk of ERG-negative prostate cancer. This finding confirms the prior study by Bastus et al (11), who found that mean CAG repeat
Androgen receptor CAG repeats in *TMPRSS2:ERG* prostate cancer

length was shorter in ERG-positive (mean length = 20 repeats) compared to ERG-negative (mean length = 21 repeats) cancer in a study of 40 prostate cancer patients. Moreover, the results are in line with experimental evidence that heightened AR signaling induces *TMPRSS2:ERG* formation (10). One of the proposed mechanisms of fusion formation is that AR signaling induces spatial proximity, leading to colocalization of the 5’ and 3’ ends of *TMPRSS2* and *ERG*, which may increase the probability of a fusion event occurring. In prostate cancer and noncancer cell lines, *TMPRSS2:ERG* formation is androgen dose dependent and may be the result of long-term androgen exposure (11). Thus, if shorter CAG repeat length drives increased transactivation of AR and is a proxy for long-term androgen exposure, there may an increased likelihood for the fusion to occur and fusion positive prostate cancer to result.

We found no association between haplotype tagging SNPs in *AR* and risk of ERG-positive or ERG-negative prostate cancer. This is in line with a prior study that showed no association between these variants and prostate cancer risk overall (4).

There are strengths and limitations of our investigation to consider. This is the first genetic epidemiology study of variants in *AR* and risk of prostate cancer defined by *TMPRSS2:ERG* status. The study integrates data on inherited susceptibility and tumor biomarkers within well-defined and prospective cohorts of men. We comprehensively investigated genetic variation in *AR*, including the CAG repeat polymorphism and common SNPs to capture inherited susceptibility across the gene. This analysis was limited to men for whom tumor tissue was available, primarily radical prostatectomy specimen. Among the prostate cancer patients in the two cohorts who had surgery, there were no differences in clinical features for those for whom we did or did not have tissue available. However, the prostate cancer cases in this study tended to be slightly younger at diagnosis, have lower
PSA levels, and be somewhat less likely to have T3 or higher stage disease than the cases without tissue who had primarily undergone radiation therapy or received androgen deprivation therapy. Although the differences in clinical features are not large, it is key to understand the generalizability of these findings to all men with prostate cancer is needed. Although our study was based on a 5-fold greater number of cases than Bastus et al and also included controls, we are somewhat limited in statistical power and may have not detected small associations with inherited AR variants. Future epidemiological studies may require consortium efforts to investigate further the association between CAG repeats and TMPRSS2:ERG formation in prostate cancer.

The CAG repeat length occurs in a domain critical for full in vivo transcriptional activation activity of the receptor. This polymorphic repeat is inversely associated with an androgen responsive reporter in androgen-dependent prostate cell lines (15). Within the NCI Breast and Prostate Cancer Cohort Consortium (BPC3), longer CAG repeat length was intriguingly associated with higher levels of both testosterone and estradiol in the circulation (4), and the age-related decline in testosterone is partly determined by CAG repeat length (16). Given reductions in AR activity associated with longer repeat length, the elevated hormone levels may represent a compensatory mechanism to achieve a balance in hormone signaling.

In the BPC3 study of 5,777 prostate cancer cases and 6,402 controls (4), there was no association between CAG repeat length and prostate cancer risk overall. Given the differing prevalence of both the fusion and average CAG repeat length across ethnicities/study populations (17, 18), including a lower prevalence of the fusion among Asian (Japanese) populations who also have longer CAG repeats on average, but also a lower prevalence of the fusion among African Americans who have shorted CAG repeats on average, our findings
Androgen receptor CAG repeats in *TMPRSS2:ERG* prostate cancer

may partly explain why the association between CAG repeats in *AR* and total prostate cancer risk varies between studies. Further studies are needed to explore this notion.

In summary, data from this epidemiological study provide supportive evidence that androgen signaling underlies the development of prostate tumors that harbor *TMPRSS2:ERG*. Moreover, these results suggest that *TMPRSS2:ERG* may represent a unique molecular subtype of prostate cancer with an etiology distinct from *TMPRSS2:ERG* negative disease.
Acknowledgements

We are grateful to participants in the PHS and HPFS for their dedicated commitment to these research studies. We would like to acknowledge the contributions of several members of the research team for their efforts: Li Moy, Elizabeth Frost-Hawes, Lauren McLaughlin, and Chungdak Li. The tissue microarrays were constructed by the Tissue Microarray Core Facility at the Dana Farber/Harvard Cancer Center.
Androgen receptor CAG repeats in TMPRSS2:ERG prostate cancer

References

Androgen receptor CAG repeats in TMPRSS2:ERG prostate cancer


Androgen receptor CAG repeats in TMPRSS2:ERG prostate cancer

Table 1. Clinical characteristics among 291 men with prostate cancer by ERG tumor status and controls, Physicians’ Health Study and Health Professionals Follow-up Study.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>ERG-negative</th>
<th>ERG-positive</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>144</td>
<td>147</td>
<td>1,221</td>
</tr>
<tr>
<td>Mean age at diagnosis (SD), years</td>
<td>67.2 (6.9)</td>
<td>65.0 (6.7)</td>
<td></td>
</tr>
<tr>
<td>Mean age at blood draw (SD), years</td>
<td>60.8 (7.9)</td>
<td>58.5 (7.3)</td>
<td>63.2 (5.0)</td>
</tr>
<tr>
<td>Median PSA at diagnosis (ng/ml) (IQR)</td>
<td>7.0 (5.3-11)</td>
<td>7.0 (4.7-10.4)</td>
<td></td>
</tr>
<tr>
<td>Gleason score, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤6</td>
<td>32 (22)</td>
<td>25 (17)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>85 (59)</td>
<td>92 (63)</td>
<td></td>
</tr>
<tr>
<td>≥8</td>
<td>26 (18)</td>
<td>28 (19)</td>
<td></td>
</tr>
<tr>
<td>Clinical tumor stage, N (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1/T2</td>
<td>128 (91)</td>
<td>132 (92)</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>7 (5)</td>
<td>9 (6)</td>
<td></td>
</tr>
<tr>
<td>T4, N1/Nx, M1</td>
<td>5 (4)</td>
<td>2 (1)</td>
<td></td>
</tr>
<tr>
<td>CAG repeat length</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean, range</td>
<td>22 (13 – 31)</td>
<td>21 (12 – 30)</td>
<td>22 (9 – 36)</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>20-24</td>
<td>20-23</td>
<td>20-24</td>
</tr>
</tbody>
</table>
Table 2. Adjusted odds ratios\textsuperscript{a} and 95% confidence intervals (CIs) of CAG repeat length in the androgen receptor (AR) and risk of ERG-positive and ERG-negative prostate cancer, Physicians’ Health Study and Health Professionals Follow-up Study.

<table>
<thead>
<tr>
<th>CAG repeat length</th>
<th>≥24</th>
<th>22-23</th>
<th>20-21</th>
<th>≤19</th>
<th>Per 1 CAG repeat decrease</th>
<th>P-value for trend\textsuperscript{b}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls, N (%)</td>
<td>341 (29)</td>
<td>250 (22)</td>
<td>343 (30)</td>
<td>220 (19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERG+ cases (N, %)</td>
<td>32 (24)</td>
<td>29 (21)</td>
<td>45 (33)</td>
<td>30 (22)</td>
<td>1.07 (1.00-1.14)</td>
<td>0.04</td>
</tr>
<tr>
<td>Odds ratio (95% CI)</td>
<td>REF</td>
<td>1.20 (0.70, 2.06)</td>
<td>1.39 (0.86, 2.26)</td>
<td>1.40 (0.82, 2.39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERG- cases (N, %)</td>
<td>41 (31)</td>
<td>30 (23)</td>
<td>36 (27)</td>
<td>26 (19)</td>
<td>0.99 (0.93-1.05)</td>
<td>0.69</td>
</tr>
<tr>
<td>Odds ratio (95% CI)</td>
<td>REF</td>
<td>0.96 (0.58, 1.58)</td>
<td>0.85 (0.53, 1.36)</td>
<td>0.96 (0.57, 1.61)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a} Adjusted for age at blood draw.

\textsuperscript{b} P-value calculated by modeling CAG repeat length as a continuous variable.
Androgen receptor CAG repeat polymorphism and risk of TMPRSS2:ERG positive prostate cancer

Sun Yoo, Andreas Pettersson, Kristina M. Jordahl, et al.

Cancer Epidemiol Biomarkers Prev Published OnlineFirst June 12, 2014.