Transcriptomes of prostate cancer with TMPRSS2:ERG and other ETS fusions

HHS Public Access

Author manuscript

*Mol Cancer Res*. Author manuscript; available in PMC 2023 July 03.

Published in final edited form as:

*Mol Cancer Res*. 2023 January 03; 21(1): 14–23. doi:10.1158/1541-7786.MCR-22-0446.

# Transcriptomes of prostate cancer with *TMPRSS2:ERG* and other ETS fusions

**Konrad H. Stopsack**[1], **Xiaofeng A. Su**[2], **J. Bailey Vaselkiv**[1], **Rebecca E. Graff**[3,4,5], **Ericka M. Ebot**[1], **Andreas Pettersson**[6], **Rosina T. Lis**[7], **Michelangelo Fiorentino**[1,8], **Massimo Loda**[10], **Kathryn L. Penney**[1,9], **Tamara L. Lotan**[11,*], **Lorelei A. Mucci**[1,*]

[1] Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA

[2] David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA

[3] Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, USA.

[4] Division of Research, Kaiser Permanente Northern California, Oakland, CA

[5] Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA

[6] Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

[7] Department of Pathology and Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA

[8] Pathology Unit, Addarii Institute, S. Orsola-Malpighi Hospital, Bologna, Italy

[9] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

[10] Department of Pathology, Weill Cornell Medical College, New York, NY

[11] Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD

## Abstract

The most common somatic event in primary prostate cancer is a fusion between the androgen-related *TMPRSS2* gene and the *ERG* oncogene. Tumors with these fusions, which occur early in carcinogenesis, have a distinctive etiology. A smaller subset of other tumors harbor fusions between *TMPRSS2* and members of the ETS transcription factor family other than *ERG*. To assess the genomic similarity of tumors with non-*ERG* ETS fusions and those with fusions involving *ERG*, this study derived a transcriptomic signature of non-*ERG* ETS fusions and assessed this signature and *ERG*-related gene expression in 1,050 men with primary prostate cancer from three independent population-based and hospital-based studies. While non-*ERG* ETS fusions involving

*Correspondence:* Konrad Stopsack, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115, stopsack@hsph.harvard.edu, (617) 432-2916.

*Equal contributions

*ETV1*, *ETV4*, *ETV5*, or *FLI1* were individually rare, they jointly accounted for one in seven prostate tumors. Genes differentially regulated between non-*ERG* ETS tumors and tumors without ETS fusions showed similar differential expression when *ERG* tumors and tumors without ETS fusions were compared (differences explained: $R^2$ 69–77%), including ETS-related androgen receptor (AR) target genes. Differences appeared to result from similarities among ETS tumors rather than similarities among non-ETS tumors. Gene sets associated with *ERG* fusions were consistent with gene sets associated with non-*ERG* ETS fusions, including fatty acid and amino acid metabolism, an observation that was robust across cohorts.

**Implications:** Considering ETS fusions jointly may be useful for etiologic studies on prostate cancer, given that the transcriptome is profoundly impacted by *ERG* and non-*ERG* ETS fusions in a largely similar fashion, most notably genes regulating metabolic pathways.

### Keywords

prostate cancer; ETS fusions; *TMPRSS2*; *ERG*; transcriptome

## Introduction

The *TMPRSS2:ERG* gene fusion is the most common somatic event in primary prostate cancer, occurring in ~50% of tumors in men of European ancestry (1). It involves the fusion of androgen-regulated *TMPRSS2* and the oncogene *ERG,* a member of the ETS family of transcription factors (2,3). The fusion occurs less frequently in tumors from men of African and Asian ancestry than of European ancestry (1). A number of nonmodifiable and modifiable risk factors are associated specifically with *TMPRSS2:ERG*-positive versus negative prostate cancer, suggesting that this fusion is a distinct etiologic subtype of prostate cancer (4–10). ERG-positive tumors have higher protein expression of insulin receptor (IR), insulin-like growth factor 1 receptor (IGF-1R), adiponectin receptor 2 (AdipoR2), and fatty acid synthase (FASN) (11), which suggests specific metabolic sensitivities.

Fusion events of *TMPRSS2* with members of the ETS gene family other than *ERG* also occur in prostate cancer, although with a lower prevalence (12,13). The extent to which these somatic events have similar causes and consequences as compared to *TMPRSS2:ERG* is unclear.

In the present study, we used the transcriptome as a global measure to address three questions regarding molecular subtypes defined by ETS fusions in prostate cancer. First, we assessed to what extent tumors with fusions involving *ERG* were transcriptionally similar to or distinct from those with ETS fusions other than *ERG*. Second, we quantified to what extent *ERG*-associated genes from a published ETS transcriptomic signature (14) and associated gene sets could be validated across cohorts with different inclusion criteria and approaches to transcriptome profiling. Third, given the differences in prevalence of the fusion by race and ancestry, and racial disparities in prostate cancer incidence and mortality rates (15), we examined how transcriptomic profiles of ETS fusions differed between Black and White men.

## Methods

### The Cancer Genome Atlas (TCGA)

Analyses contrasting ETS subtypes and assessments of expression differences by race were done in the TCGA prostate adenocarcinoma (PRAD) study of primary prostate cancer (16), which was based on fresh-frozen tissue from radical prostatectomy. RNA sequencing was performed using Illumina HiSeq, mapping to 20,180 unique gene symbols, and quantified reads per kilobase of transcript. *ERG*, *ETV1*, *ETV4*, and *FLI1* gene fusions were identified through the FusionSeq algorithm using RNA sequencing data. TCGA prostate cancer data were retrieved via cBioPortal on February 19, 2019.

### Health Professionals Follow-up Study (HPFS) and Physicians' Health Study (PHS)

Analyses validating ERG-related gene expression were done in a transcriptome substudy of participants from the HPFS and PHS cohorts with prostate cancer. HPFS is an ongoing cohort study of 51,529 health professionals followed since 1986 (17). PHS I and II began in 1982 and 1997 respectively as randomized primary prevention trials of aspirin and supplements among 29,067 physicians (18,19). Incident prostate cancer diagnoses during prospective follow-up of these two studies were confirmed by review of medical records, pathology reports, and re-review of tumor specimens. The study was conducted in in accordance with the U.S. Common Rule, and the study protocol was approved by the institutional review boards at Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. The current study examines tumors with available ERG immunohistochemistry and RNA expression profiling.

Archival formalin-fixed paraffin-embedded prostate tissue specimens from radical prostatectomy or transurethral resection of the prostate retrieved from treating hospitals were centrally re-reviewed. All available samples were included on tissue microarrays, using at least three 0.6 mm cores of tumor per case. Tissue microarrays were characterized for ERG protein expression, using immunohistochemistry as a proxy for the *TMPRSS2:ERG* fusion (20). Endothelial ERG staining served as a positive internal control, and tumors were considered to have an ERG fusion when staining was positive in at least one core.

RNA expression profiling was done in tumors selected using cumulative incidence case-control sampling of lethal cases (participants who developed metastases/died from prostate cancer at any time) and non-lethal controls (participants with prostate cancer but no metastases >8 years after diagnosis). Profiling used the GeneChip Human Gene 1.0 ST microarray (Affymetrix) with a resulting 20,254 unique gene symbols (Gene Expression Omnibus, GSE79021) (21,22).

### Johns Hopkins Hospital prostatectomy cohort (Hopkins)

Analyses validating *ERG*- and non-*ERG* ETS-related gene expression and of differences by race were also done among tumors from men with intermediate/high-risk primary prostate cancer who were part of two partially overlapping radical prostatectomy studies at Johns Hopkins Hospital, oversampled for adverse clinical outcomes (23–25).

Formalin-fixed paraffin-embedded prostate tissue specimens were assayed for *TMPRSS2:ERG* fusions in a comparable fashion as HPFS/PHS, using ERG immunohistochemistry with the same primary antibody (26). Endothelial ERG staining served as a positive internal control, and tumors were considered to have an ERG fusion when staining was positive in at least one core. In addition, fusions involving *ETV1*, *ETV4*, and *ETV5* were detected using chromogenic RNA *in-situ* hybridization with the RNAscope FFPE kit 2.5 (Advanced Cell Diagnostics, Hayward, CA), using separately validated probes (27), as previously described (26). Tumors were classified into mutually exclusive groups as non-*ERG* ETS if ERG protein expression was known and negative (26 tumors were excluded from this analysis because of uninterpretable or heterogeneous ERG status between cores) and RNA *in-situ* hybridization was positive for *ETV1*, *ETV4*, or *ETV5*; the remainder was classified as ERG or non-ETS.

RNA expression profiling used the Human Exon 1.0 ST microarrays (Affymetrix) in a Clinical Laboratory Improvement Amendments (CLIA)-certified clinical laboratory (Decipher Biosciences, San Diego, CA), as previously described (24,25), resulting in 46,052 gene symbols.

## Statistical analysis

Regression-based differential gene expression analyses with empirical Bayes adjustment were conducted using the Bioconductor package limma (version 3.44.3) (28). In TCGA, heteroskedasticity in the count data from RNA sequencing were addressed via precision weighting using the voom function.

The primary signature of ETS-related genes was previously developed by Berglund *et al.* (14) and consists of 413 genes (one of which not profiled in TCGA) that were both differentially expressed by ETS status in TCGA and by inferred ETS status in a clinical microarray series of 10,275 tumors. For consistency with this signature, low-abundance transcripts were included in all analyses.

To develop a signature of non-*ERG* ETS tumors in TCGA, genes were identified that were differentially expressed between tumors with non-*ERG* ETS fusions and tumors without ETS fusions of any type at a transcriptome-wide Benjamini-Hochberg false-discovery rate (FDR) of <0.001.

Three types of gene-level comparisons were conducted. First, Berglund *et al.* ETS genes and the non-*ERG* ETS genes identified in the preceding step were compared in their expression between *ERG* and non-*ERG* ETS transcriptomes, each in reference to non-ETS tumors, in TCGA and in the Hopkins cohort (Figure 1). Overall similarity was expressed as variance explained ($R^2$, the square of Pearson correlations). Second, for validation, the ETS genes associated with *ERG* status in TCGA at $p < 0.05$ were tested against ERG status in HPFS/PHS and the Hopkins cohort; no null hypothesis-preserving FDR correction was used because the null hypothesis of no differences by ERG status was unlikely to be true. Third, differences by self-reported race in expression ratios of ETS genes between ETS and non-ETS tumors were explored in tumors from White and Black men in TCGA and the Hopkins cohort, assessing a multiplicative interaction term between race and ETS status in

a regression model that also included main effects for race and ETS status. FDR correction was calculated among the ETS genes tested.

Androgen receptor (AR) transcriptional activity was summarized as a más-o-menos score (29) of a well-characterized AR signature (30). In addition, we assessed the expression of 99 ETS-related AR target genes identified based on ChIP-seq by Berglund *et al.* (14) after excluding genes in their "ETS+/− Up" group that did not differ by ETS status. Similarity in expression profiles of ETS-related AR target genes was compared as described above for the ETS and non-*ERG* ETS genes.

In competitive gene set analyses, gene sets with greater differential expression than other gene sets were identified for comparisons (Figure 1) of *ERG vs.* non-*ERG* tumors (in all three cohorts separately), for ETS *vs.* non-ETS tumors (TCGA, Hopkins), and for non-*ERG* ETS *vs.* non-ETS tumors (TCGA, Hopkins). These analyses were based on the main-effect regression models described above and were performed using the Camera approach (31) implemented in the limma package. Hallmark and KEGG pathway gene sets from the Molecular Signature Database version 6.2 were used (32,33), with FDR correction among all 236 gene sets.

### Data Availability

Microarray gene expression data are available on Gene Expression Omnibus under accession numbers GSE79021, GSE79957, GSE79958, and GSE51066. Additional data from the Johns Hopkins cohort are available on request from the authors; HPFS and PHS prostate cancer data are available through an HPFS project request (https://sites.sph.harvard.edu/hpfs/for-collaborators).

## Results

### Study populations

This study included 332 men with sufficient quality of tumor RNA sequencing-based transcriptome and gene fusion calls from the TCGA study, 379 men with ERG protein and microarray gene expression data in the HPFS and PHS cohorts, and 339 men with ERG immunohistochemistry and *ETV1/4/5* RNA *in-situ* hybridization data from Johns Hopkins Hospital radical prostatectomy cohorts. Most men (TCGA 84%, HPFS/PHS 98%, Hopkins 90%) self-identified as White, and ETS fusions tended to be more common among White men (Table 1). Tumor characteristics by ETS status were similar across studies (Table 1).

In TCGA, 197 tumors had ETS fusions (59%), including 151 at *ERG* (45%) and 46 non-*ERG* ETS fusions (14%) involving *ETV1*, *ETV4*, or *FLI1* (Figure 1A). No *ETV5* fusions were detected.

HPFS and PHS participants tended to be diagnosed at an older age with lower clinical stage, compared to TCGA, but had similar tumor characteristics in terms of Gleason score and prostate-specific antigen (PSA) at diagnosis. 187 tumors (49%) were positive for ERG protein expression (Figure 1A). Other ETS subtypes were not characterized.

Patients from Johns Hopkins Hospital tended to have lower clinical stage (98% T1/T2) than TCGA; other tumor characteristics were largely comparable. 172 tumors had ETS fusions (51%), including 129 that were positive for ERG protein expression (38%) and 43 non-*ERG* ETS fusions (13%) involving *ETV1*, *ETV4*, or *ETV5* (Figure 1A). *FLI1* was not assayed.

## ERG and non-ERG ETS transcriptomes

The Berglund *et al.* ETS signature of 412 genes (Figure 1B–D) was developed in a study population that included TCGA, where 76% of ETS fusions involved ERG. Indeed, 400 of the ETS genes were expressed differentially by *ERG* vs. non-*ERG* status at $p < 0.05$, and the remaining 12 genes had associations with *ERG* status in the same direction as with ETS status in the original study (Supplementary Table 1).

A new signature of non-*ERG* ETS fusions was generated in TCGA by contrasting the 46 tumors with non-*ERG* ETS fusions to the 135 tumors without any ETS fusions. Interestingly, of the 774 genes identified at FDR < 0.001, 164 genes (21%) were among the 412 Berglund ETS genes.

To address the possibility that additional genes among the 774 non-*ERG* ETS genes were reflective of a broader transcriptional difference between ETS and non-ETS tumors, differences in gene expression in tumors with *ERG* fusions and in tumors with non-*ERG* ETS fusions were contrasted against non-ETS tumors in TCGA (Figure 1B, Figure 2) and the Johns Hopkins cohort (Figure 1D, Figure 2). Overall, as compared to non-ETS tumors, *ERG* and non-*ERG* ETS tumors showed broadly similar transcriptional features, both among the 412 Berglund ETS genes (Figure 2A, C) and among the signature of 774 non-*ERG* ETS genes (Figure 2B, D). For the 412 Berglund *et al.* ETS genes, transcriptional differences between *ERG* and non-ETS tumors explained 76% of the variance (95% CI 72–80) in transcriptional differences between non-*ERG* ETS tumors and non-ETS tumors in TCGA (Figure 2A) and 53% in the Hopkins cohort (95% CI 46–60; Figure 2C). Likewise, for the 774 non-*ERG* ETS genes in TCGA, transcriptional differences between *ERG* and non-ETS tumors explained 77% of the variance (95% CI 74–80) in transcriptional differences between non-*ERG* ETS tumors and non-ETS tumors in TCGA (Figure 2B) and 69% in the Hopkins cohort (95% CI 65–72; Figure 2D).

Despite these similarities, some ETS genes (Figure 2A, C) and some non-*ERG* ETS genes (Figure 2B, D) were expressed differently between *ERG* and non-*ERG* ETS tumors, as compared to non-ETS tumors. These included, as expected, *ERG* and *ETV1*, as well as some of the genes among the 774 non-ERG ETS genes that are part of a 15-gene non-*ERG* ETS signature generated by Tomlins *et al.* (2015) to separate such tumors transcriptionally from *ERG* and non-ETS tumors (Figure 2B, D, asterisks).

Comparisons of ETS-related, *ERG*-related and non-*ERG* ETS related gene expression used the same comparison group of non-ETS tumors, raising the possibility that gene expression differences may have been driven by the transcriptome of a subset of non-ETS tumors. Notably, when stratifying the comparison group of non-ETS tumors in TCGA into 37 non-ETS *SPOP*-mutated tumors (11%), the most common known subtype that is mutually exclusive with ETS fusions, and 98 non-ETS non-*SPOP* tumors (30%), the expression

differences in the Berglund *et al.* ETS genes between ETS tumors were highly consistent ($R^2$ 86%, 95% CI 83–88) between these two groups of non-ETS comparators (Supplementary Figure 1).

## ETS-related androgen receptor (AR) transcriptome

In these cohorts of men with primary prostate cancer who were generally not hormonally treated before surgery, the overall transcriptional output of AR appeared comparable between ETS subtypes (Supplementary Figure 2, Supplementary Table 2). Among 99 AR target genes that have been reported to be ETS-dependent (14), overall patterns of expression were similar among *ERG* and non-*ERG* ETS tumors and contrasted with non-ETS tumors (Supplementary Figure 3). For the ETS-related AR target genes, transcriptional differences between *ERG* and non-ETS tumors explained 79% of the variance (95% CI 70–85) in transcriptional differences between non-*ERG* ETS tumors and non-ETS tumors in TCGA and 60% in the Hopkins cohort (95% CI 46–71).

## Replication of the ERG transcriptome across studies

With at least 400 of the 412 ETS genes associated with *ERG* status in TCGA, replication of the *ERG* transcriptome was assessed in HPFS/PHS (Figure 3A). Of the 400 gene symbols, expression of 354 genes (88%) was quantified on the microarray, and 308 of those genes (87%) were associated with ERG status at $p < 0.05$ and in the same direction as in TCGA (Supplementary Table 3). Similarly, in the Hopkins cohort, expression of 374 was quantified on the microarray, and 355 of those genes (95%) were associated with ERG status at $p < 0.05$ and in the same direction as in TCGA (Figure 3B; Supplementary Table 3).

## Differences in the ETS transcriptome by race

To evaluate whether the transcriptome of ETS fusions differed by race, ratios in expression by ETS status of the 412 Berglund *et al.* genes were compared between tumors from White men and Black men in TCGA (Figure 3C; Supplementary Table 4). Expression ratios were largely concordant by race ($n = 43$ Black men, 16 with ETS fusions). Among 36 genes with $p_{interaction}$ by race of < 0.05, not adjusted for multiple comparisons, two genes (*IL5RA*, *DLGAP1*) had directionally consistent race differences in ETS-related gene expression differences at $p_{interaction} < 0.05$ in the Hopkins cohort ($n = 27$ Black men, 6 with ETS fusions; Figure 3D). Underrepresentation of Black men precluded analyses according to ETS subtypes.

## Gene sets associated with ERG, ETS, and non-ERG ETS fusions

Among all Hallmark and KEGG gene sets in TCGA, the 7 gene sets that were more strongly associated with *ERG vs.* non-*ERG* tumors than other gene sets included KEGG pathways for fatty acid metabolism and amino acid metabolism (valine, leucine, and isoleucine degradation; beta-alanine metabolism; Table 2). The gene sets identified when comparing ETS *vs.* non-ETS tumors and when comparing non-*ERG* ETS tumors *vs.* non-ETS tumors in TCGA included 6 of the 7 *ERG* gene sets (all but KEGG Systemic Lupus Erythematosus) with the same directionality, as well as up to 22 additional pathways at FDR < 0.05 (Supplementary Table 5).

The 7 gene sets from the comparison of *ERG vs.* non-*ERG* from TCGA all had the same direction, with varying FDR, in both HPFS/PHS and the Hopkins cohort (Table 2). Similarly, the top-10 gene sets from the comparison of ETS *vs.* non-ETS tumors and from comparing non-*ERG* ETS tumors *vs.* non-ETS tumors were also consistent for all but one gene set in the Hopkins cohort (Table 2, Supplementary Table 5).

## Discussion

In this study, we evaluated to what extent the transcriptome of the common ETS fusions in prostate cancer differs between fusions including the *ERG* oncogene and individually less common fusions involving other members of the ETS gene family. Overall, we observed that *ERG* and non-*ERG* ETS tumors had relatively similar transcriptional differences relative to non-ETS tumors. Differences between ETS and non-ETS tumors dominated, which likely resulted from the ETS fusions rather than from a distinct transcriptional signal from a subset of non-ETS tumors. We found high consistency of *ERG*-associated genes from a published ETS signature, and of *ERG*- and ETS-associated gene sets, upon validation in separate population-based and hospital-based studies. Our results provide additional validation for an ETS-related broad transcriptional program, which suggests that the transcriptomic impact of this family of fusions is profound and that it might be conserved across study populations and tumors with different characteristics. Against the background of lower prevalence of ETS fusions in tumors from Black men than White men, we did not reliably observe differences in the transcriptional program of ETS fusions between tumors from White and Black men.

ETS-associated gene sets included fatty acid and amino acid metabolic pathways. The findings of a robust transcriptional program encompassing metabolic pathways provide further molecular rationale why ERG-positive prostate tumors have been shown to be particularly susceptible to hormonally-acting cancer risk factors, including adiposity (11,4,5) and physical inactivity (10). The study by Berglund *et al.* (14), which additionally included tumors from the GenomeDX GRID database but did not separate *ERG* and non-*ERG* ETS fusions, also identified differences by ETS status that were not apparent in our analysis but that have strong support from other lines of evidence, such as androgen signaling (34,35), cholesterol homeostasis, and estrogen response (36). Our findings corroborate that ETS fusions, including the *TMPRSS2:ERG* gene fusion, mark a molecularly distinct subtype of prostate cancer with prominent transcriptional and metabolic rewiring of amino acid and fatty acid metabolism. Interestingly, transcriptome patterns are aligned with multiple tissue-based metabolomics studies, which found *TMPRSS2:ERG* fusions to be associated with enrichment of amino acids, complex lipids, and fatty acids (37–40).

We additionally asked to what extent ETS fusions of genes other than *ERG* differ in their transcriptomic program from tumors with *ERG* fusions. This question is particularly relevant for population-based studies of cancer etiology, which have so far typically relied on immunohistochemical quantifications of ERG protein expression, as our two validation sets did. Non-*ERG* ETS fusions, like those involving *ETV1*, *ETV4*, *ETV5*, and *FLI1*, are relatively rare, even collectively, which prohibited disentangling their transcriptomes further. This prevalence would also greatly limit the precision of epidemiologic studies on etiologic

heterogeneity and will likely render such studies on non-*ERG* ETS fusions in isolation infeasible for years to come.

Molecular and structural biology evidence suggests that ERG and other non-ERG ETS fusion transcription factors all have a similar DNA binding motif (12,2,13). Almost all ETS fusions involve *TMPRSS2*, which has oncogenic function irrespective of ETS fusions (41). Nevertheless, contributions to neomorphic transcriptional reprogramming may differ between ETS transcription factors. Although ERG, ETV1, ETV4, FLI1 were all shown to interact with BAF chromatin remodeling complexes (Boulay et al. 2017; Sandoval et al. 2018), it is possible that different ETS transcription factors recruit BAF complexes to different genomic loci. In addition, chromosomal locations of ETS genes can also impact transcriptional consequences. *TMPRSS2* and *ERG* are located within proximity of each other on human chromosome 21q22, and the *TMPRSS2:ERG* fusion may cause a loss of the interstitial region containing multiple genes, which promotes prostate cancer progression in mice (42). In contrast, *ETV1*, *ETV4*, and *FLI1* are located on entirely different chromosomes, and thus these non-*ERG* ETS fusions could reasonably have an impact on many other genes. A signature of non-*ERG* ETS tumors was developed by Tomlins et al. (43), and some non-*ERG* ETS genes are regulated differently by *ERG* status (44), such as mRNA expression of *ETV1* itself. While a study suggested that ERG would not interact with *PTEN* loss and AR activation in the same manner as ETV1 (45), other studies have supported such interactions (46–48). Our data were not suggestive of major differences in AR transcriptional output between ETS subtypes, and differences in expression of ETS-related AR target genes between *ERG* and non-ETS tumors were mirrored by similar differences between non-*ERG* ETS and non-ETS tumors.

Our observation of broad similarities in transcriptional patterns of non-*ERG* ETS fusions to tumors with *ERG* fusions was thus not fully predictable. Importantly, this observation suggests that non-*ERG* ETS tumors could reasonably be expected to be susceptible to and influenced by similar etiologic factors as tumors with *ERG* fusions. While the molecular mechanisms through which fusions form may differ (13), how early tumor cell clones can adapt to selective pressure may be influenced by the interaction of their genomic profile and the risk factor milieu. As one example, ETS fusions in both prostate cancer (*TMPRSS2:ERG*) and Ewing sarcoma (*EWS:FLI1*) may sensitize tumors to the insulin/insulin-like growth factor-1 (IGF-1) signaling axis (11,49). Assessing ERG status alone misclassifies tumors with non-*ERG* ETS fusions together with non-ETS tumors as “ERG-negative,” which would be expected to attenuate between-subtype heterogeneity for etiologic factors and prognostic differences between subtypes. Comparisons by ERG status led to similar inference about the top transcriptional pathways but did not identify the additional transcriptome pathways that were detectable when separating non-*ERG* ETS tumors from non-ETS tumors (Table 2; Supplementary Table 5).

An ostensible limitation of the current study are the different approaches to measuring ETS status and gene expression in the three study populations. For example, the higher similarity between *ERG* and non-*ERG* ETS in TCGA (variance explained, 77%) than in the Hopkins cohort (69%) may be the result of larger measurement error from the microarray technology used in the latter study. Further, transcriptomes of tumors are modified by myriad other

factors that we did not consider, including patient characteristics and other DNA alterations associated with *ERG* fusions, such as *PTEN* loss. That most genes and gene sets were validated despite these differences should be considered additional evidence for the broad impact of ETS fusions on the transcriptome. Not all transcriptional differences may be reflected on the protein level, and it should not be inferred from this study that all individual target proteins would follow the overall transcriptional patterns. Protein expression of fatty acid synthase (FASN) was previously shown to be higher in ERG-positive than ERG-negative tumors in a larger sample from HPFS and PHS (11). Further, the small number of tumors from Black men is an obvious limitation. Our lack of detection of statistically significant differences by race in ETS-related transcriptomes should not be interpreted as proof of their absence. Binding intensity of the AR transcription factor at certain targets, including at lipid metabolism genes, may differ between Black and White men (50), and considering ETS fusions with their different prevalence by race could be informative in this regard. However, instead of attempting to identify racial differences with larger sample size, a more informative next step would be to identify what etiologic factors cause prostate cancers with ETS fusions in the first place, which may lead to insights into potential prevention strategies and might ultimately explain their different race-specific incidence rates.

In conclusion, ETS fusions, involving *ERG* and non-*ERG* ETS genes, are a unique molecular subtype in primary prostate cancer, providing a rationale for differential associations of emerging prostate cancer risk factors that may act through hormonal or metabolic effects specifically with tumors harboring ERG fusions. Our results also suggest that non-*ERG* ETS tumors as a group are transcriptionally largely similar to tumors with *ERG* fusions and could be expected to be influenced by similar risk factors as tumors with *ERG* fusions.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank the participants and staff of the HPFS and the PHS for their valuable contributions. In particular, we would like to recognize the contributions of Liza Gazeeva, Siobhan Saint-Surin, Betsy Frost-Hawes, and Eleni Konstantis. We appreciate the support from the Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA. We would further like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. Central registries may also be supported by state agencies, universities, and cancer centers. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming.

This research was funded in part a Harvard–MIT Bridge Project (K.H. Stopsack, X.A. Su, L.A. Mucci), by NIH/NCI grant R01 CA136578 (to L.A. Mucci), R01 CA141298, and the Specialized Programs of Research Excellence program in Prostate Cancer 5P50 CA090381. The HPFS is supported by the NIH (U01 CA167552, to L.A. Mucci). K.H. Stopsack, R.E. Graff, K.L. Penney, T.L. Lotan, and L.A. Mucci are Prostate Cancer Foundation Young Investigators.

*Disclosures:* R.T. Lis consults for Janssen for other projects. T.L. Lotan has received research funding for other projects from Ventana/Roche, DeepBio and AIRAMatrix. L.A. Mucci has received research funding for other projects from Astra Zeneca and Janssen, served as a consultant to Bayer, and serves on the advisory board for Convergent Therapeutics. The remaining authors have no relevant disclosures.

## References


1. Zhou CK, Young D, Yeboah ED, Coburn SB, Tettey Y, Biritwum RB, et al. TMPRSS2:ERG Gene Fusions in Prostate Cancer of West African Men and a Meta-Analysis of Racial Differences. Am J Epidemiol. 2017;186:1352–61. [PubMed: 28633309]
2. Findlay VJ, LaRue AC, Turner DP, Watson PM, Watson DK. Understanding the role of ETS-mediated gene regulation in complex biological processes. Adv Cancer Res. 2013;119:1–61. [PubMed: 23870508]
3. Nicholas TR, Strittmatter BG, Hollenhorst PC. Oncogenic ETS Factors in Prostate Cancer. Adv Exp Med Biol. 2019;1210:409–36. [PubMed: 31900919]
4. Egbers L, Luedeke M, Rinckleb A, Kolb S, Wright JL, Maier C, et al. Obesity and Prostate Cancer Risk According to Tumor TMPRSS2:ERG Gene Fusion Status. Am J Epidemiol. 2015;181:706–13. [PubMed: 25852077]
5. Graff RE, Ahearn TU, Pettersson A, Ebot EM, Gerke T, Penney KL, et al. Height, Obesity, and the Risk of TMPRSS2:ERG-Defined Prostate Cancer. Cancer Epidemiol Biomarkers Prev. 2018;27:193–200. [PubMed: 29167279]
6. Graff RE, Meisner A, Ahearn TU, Fiorentino M, Loda M, Giovannucci EL, et al. Pre-diagnostic circulating sex hormone levels and risk of prostate cancer by ERG tumour protein expression. Br J Cancer. 2016;114:939–44. [PubMed: 26986253]
7. Hashim D, Gonzalez-Feliciano AG, Ahearn TU, Pettersson A, Barber L, Pernar CH, et al. Family history of prostate cancer and the incidence of ERG- and phosphatase and tensin homolog-defined prostate cancer. Int J Cancer. 2020;146:2694–702. [PubMed: 31318977]
8. Luedeke M, Rinckleb AE, FitzGerald LM, Geybels MS, Schleutker J, Eeles RA, et al. Prostate cancer risk regions at 8q24 and 17q24 are differentially associated with somatic TMPRSS2:ERG fusion status. Hum Mol Genet. 2016;25:5490–9. [PubMed: 27798103]
9. Penney KL, Pettersson A, Shui IM, Graff RE, Kraft P, Lis RT, et al. Association of Prostate Cancer Risk Variants with TMPRSS2:ERG Status: Evidence for Distinct Molecular Subtypes. Cancer Epidemiol Biomarkers Prev. 2016;25:745–9. [PubMed: 26941365]
10. Pernar CH, Ebot EM, Pettersson A, Graff RE, Giunchi F, Ahearn TU, et al. A Prospective Study of the Association between Physical Activity and Risk of Prostate Cancer Defined by Clinical Features and TMPRSS2:ERG. Eur Urol. 2019;76:33–40. [PubMed: 30301696]
11. Pettersson A, Lis RT, Meisner A, Flavin R, Stack EC, Fiorentino M, et al. Modification of the association between obesity and lethal prostate cancer by TMPRSS2:ERG. Journal of the National Cancer Institute. 2013;105:1881–90. [PubMed: 24292212]
12. Hollenhorst PC, McIntosh LP, Graves BJ. Genomic and biochemical insights into the specificity of ETS transcription factors. Annu Rev Biochem. 2011;80:437–71. [PubMed: 21548782]
13. Sizemore GM, Pitarresi JR, Balakrishnan S, Ostrowski MC. The ETS family of oncogenic transcription factors in solid tumours. Nat Rev Cancer. 2017;17:337–51. [PubMed: 28450705]
14. Berglund AE, Rounbehler RJ, Gerke T, Awasthi S, Cheng CH, Takhar M, et al. Distinct transcriptional repertoire of the androgen receptor in ETS fusion-negative prostate cancer. Prostate Cancer Prostatic Dis. 2019;22:292–302. [PubMed: 30367117]
15. Chowdhury-Paulino IM, Ericsson C, Vince R, Spratt DE, George DJ, Mucci LA. Racial disparities in prostate cancer among black men: epidemiology and outcomes. Prostate Cancer Prostatic Dis. 2021;
16. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163:1011–25. [PubMed: 26544944]
17. Giovannucci E, Liu Y, Platz EA, Stampfer MJ, Willett WC. Risk factors for prostate cancer incidence and progression in the health professionals follow-up study. International journal of cancer Journal international du cancer. 2007;121:1571–8. [PubMed: 17450530]

18. Gaziano JM, Sesso HD, Christen WG, Bubes V, Smith JP, MacFadyen J, et al. Multivitamins in the prevention of cancer in men: the Physicians' Health Study II randomized controlled trial. JAMA. 2012;308:1871–80. [PubMed: 23162860]
19. Hennekens CH, Buring JE, Manson JE, Stampfer M, Rosner B, Cook NR, et al. Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. N Engl J Med. 1996;334:1145–9. [PubMed: 8602179]
20. Pettersson A, Graff RE, Bauer SR, Pitt MJ, Lis RT, Stack EC, et al. The TMPRSS2:ERG rearrangement, ERG expression, and prostate cancer outcomes: a cohort study and meta-analysis. Cancer Epidemiol Biomarkers Prev. 2012;21:1497–509. [PubMed: 22736790]
21. Tyekucheva S, Martin NE, Stack EC, Wei W, Vathipadiekal V, Waldron L, et al. Comparing Platforms for Messenger RNA Expression Profiling of Archival Formalin-Fixed, Paraffin-Embedded Tissues. J Mol Diagn. 2015;17:374–81. [PubMed: 25937617]
22. Sinnott JA, Peisch SF, Tyekucheva S, Gerke T, Lis R, Rider JR, et al. Prognostic Utility of a New mRNA Expression Signature of Gleason Score. Clin Cancer Res. 2017;23:81–7. [PubMed: 27663590]
23. Faraj SF, Bezerra SM, Yousefi K, Fedor H, Glavaris S, Han M, et al. Clinical Validation of the 2005 ISUP Gleason Grading System in a Cohort of Intermediate and High Risk Men Undergoing Radical Prostatectomy. PLoS One. 2016;11:e0146189. [PubMed: 26731672]
24. Johnson MH, Ross AE, Alshalalfa M, Erho N, Yousefi K, Glavaris S, et al. SPINK1 Defines a Molecular Subtype of Prostate Cancer in Men with More Rapid Progression in an at Risk, Natural History Radical Prostatectomy Cohort. J Urol. 2016;196:1436–44. [PubMed: 27238617]
25. Ross AE, Johnson MH, Yousefi K, Davicioni E, Netto GJ, Marchionni L, et al. Tissue-based Genomics Augments Post-prostatectomy Risk Stratification in a Natural History Cohort of Intermediate- and High-Risk Men. Eur Urol. 2016;69:157–65. [PubMed: 26058959]
26. Torres A, Alshalalfa M, Tomlins SA, Erho N, Gibb EA, Chelliserry J, et al. Comprehensive Determination of Prostate Tumor ETS Gene Status in Clinical Samples Using the CLIA Decipher Assay. J Mol Diagn. 2017;19:475–84. [PubMed: 28341589]
27. Kunju LP, Carskadon S, Siddiqui J, Tomlins SA, Chinnaiyan AM, Palanisamy N. Novel RNA hybridization method for the in situ detection of ETV1, ETV4, and ETV5 gene fusions in prostate cancer. Appl Immunohistochem Mol Morphol. 2014;22:e32–40. [PubMed: 25203299]
28. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47. [PubMed: 25605792]
29. Zhao SD, Parmigiani G, Huttenhower C, Waldron L. Más-o-menos: a simple sign averaging method for discrimination in genomic data analysis. Bioinformatics. 2014;30:3062–9. [PubMed: 25061068]
30. Hieronymus H, Lamb J, Ross KN, Peng XP, Clement C, Rodina A, et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. Cancer Cell. 2006;10:321–30. [PubMed: 17010675]
31. Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. Nucleic Acids Res. 2012;40:e133. [PubMed: 22638577]
32. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27:1739–40. [PubMed: 21546393]
33. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417–25. [PubMed: 26771021]
34. Yu J, Yu J, Mani R-S, Cao Q, Brenner CJ, Cao X, et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. Cancer Cell. 2010;17:443–54. [PubMed: 20478527]
35. Yoo S, Pettersson A, Jordahl KM, Lis RT, Lindstrom S, Meisner A, et al. Androgen Receptor CAG Repeat Polymorphism and Risk of TMPRSS2:ERG-Positive Prostate Cancer. Cancer Epidemiol Biomarkers Prev. 2014;23:2027–31. [PubMed: 24925673]

36. Setlur SR, Mertz KD, Hoshida Y, Demichelis F, Lupien M, Perner S, et al. Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. Journal of the National Cancer Institute. 2008;100:815–25. [PubMed: 18505969]
37. Butler LM, Mah CY, Machiels J, Vincent AD, Irani S, Mutuku SM, et al. Lipidomic Profiling of Clinical Prostate Cancer Reveals Targetable Alterations in Membrane Lipid Composition. Cancer Res. 2021;81:4981–93. [PubMed: 34362796]
38. Dudka I, Thysell E, Lundquist K, Antti H, Iglesias-Gato D, Flores-Morales A, et al. Comprehensive metabolomics analysis of prostate cancer tissue in relation to tumor aggressiveness and TMPRSS2-ERG fusion status. BMC Cancer. 2020;20:437. [PubMed: 32423389]
39. Hansen AF, Sandsmark E, Rye MB, Wright AJ, Bertilsson H, Richardsen E, et al. Presence of TMPRSS2-ERG is associated with alterations of the metabolic profile in human prostate cancer. Oncotarget. 2016;7:42071–85. [PubMed: 27276682]
40. Meller S, Meyer HA, Bethan B, Dietrich D, Maldonado SG, Lein M, et al. Integration of tissue metabolomics, transcriptomics and immunohistochemistry reveals ERG- and gleason score-specific metabolomic alterations in prostate cancer. Oncotarget. 2016;7:1421–38. [PubMed: 26623558]
41. Lucas JM, Heinlein C, Kim T, Hernandez SA, Malik MS, True LD, et al. The androgen-regulated protease TMPRSS2 activates a proteolytic cascade involving components of the tumor microenvironment and promotes prostate cancer metastasis. Cancer Discov. 2014;4:1310–25. [PubMed: 25122198]
42. Linn DE, Penney KL, Bronson RT, Mucci LA, Li Z. Deletion of interstitial genes between TMPRSS2 and ERG promotes prostate cancer progression. Cancer Res. 2016;76:1869–81. [PubMed: 26880803]
43. Tomlins SA, Alshalalfa M, Davicioni E, Erho N, Yousefi K, Zhao S, et al. Characterization of 1577 primary prostate cancers reveals novel biological and clinicopathologic insights into molecular subtypes. Eur Urol. 2015;68:555–67. [PubMed: 25964175]
44. Paulo P, Ribeiro FR, Santos J, Mesquita D, Almeida M, Barros-Silva JD, et al. Molecular subtyping of primary prostate cancer reveals specific and shared target genes of different ETS rearrangements. Neoplasia. 2012;14:600–11. [PubMed: 22904677]
45. Baena E, Shao Z, Linn DE, Glass K, Hamblen MJ, Fujiwara Y, et al. ETV1 directs androgen metabolism and confers aggressive prostate cancer in targeted mice and patients. Genes Dev. 2013;27:683–98. [PubMed: 23512661]
46. Ahearn TU, Pettersson A, Ebot EM, Gerke T, Graff RE, Morais CL, et al. A Prospective Investigation of PTEN Loss and ERG Expression in Lethal Prostate Cancer. J Natl Cancer Inst. 2016;108:djv346. [PubMed: 26615022]
47. Carver BS, Tran J, Gopalan A, Chen Z, Shaikh S, Carracedo A, et al. Aberrant ERG expression cooperates with loss of PTEN to promote cancer progression in the prostate. Nat Genet. 2009;41:619–24. [PubMed: 19396168]
48. Wasmuth EV, Hoover EA, Antar A, Klinge S, Chen Y, Sawyers CL. Modulation of androgen receptor DNA binding activity through direct interaction with the ETS transcription factor ERG. Proc Natl Acad Sci U S A. 2020;117:8584–92. [PubMed: 32220959]
49. Toretsky JA, Kalebic T, Blakesley V, LeRoith D, Helman LJ. The insulin-like growth factor-I receptor is required for EWS/FLI-1 transformation of fibroblasts. J Biol Chem. 1997;272:30822–7. [PubMed: 9388225]
50. Berchuck JE, Adib E, Abou Alaiwi S, Dash AK, Shin JN, Lowder D, et al. The prostate cancer androgen receptor cistrome in African American men associates with upregulation of lipid metabolism and immune response. Cancer Res. 2022;canres.3552.2021.

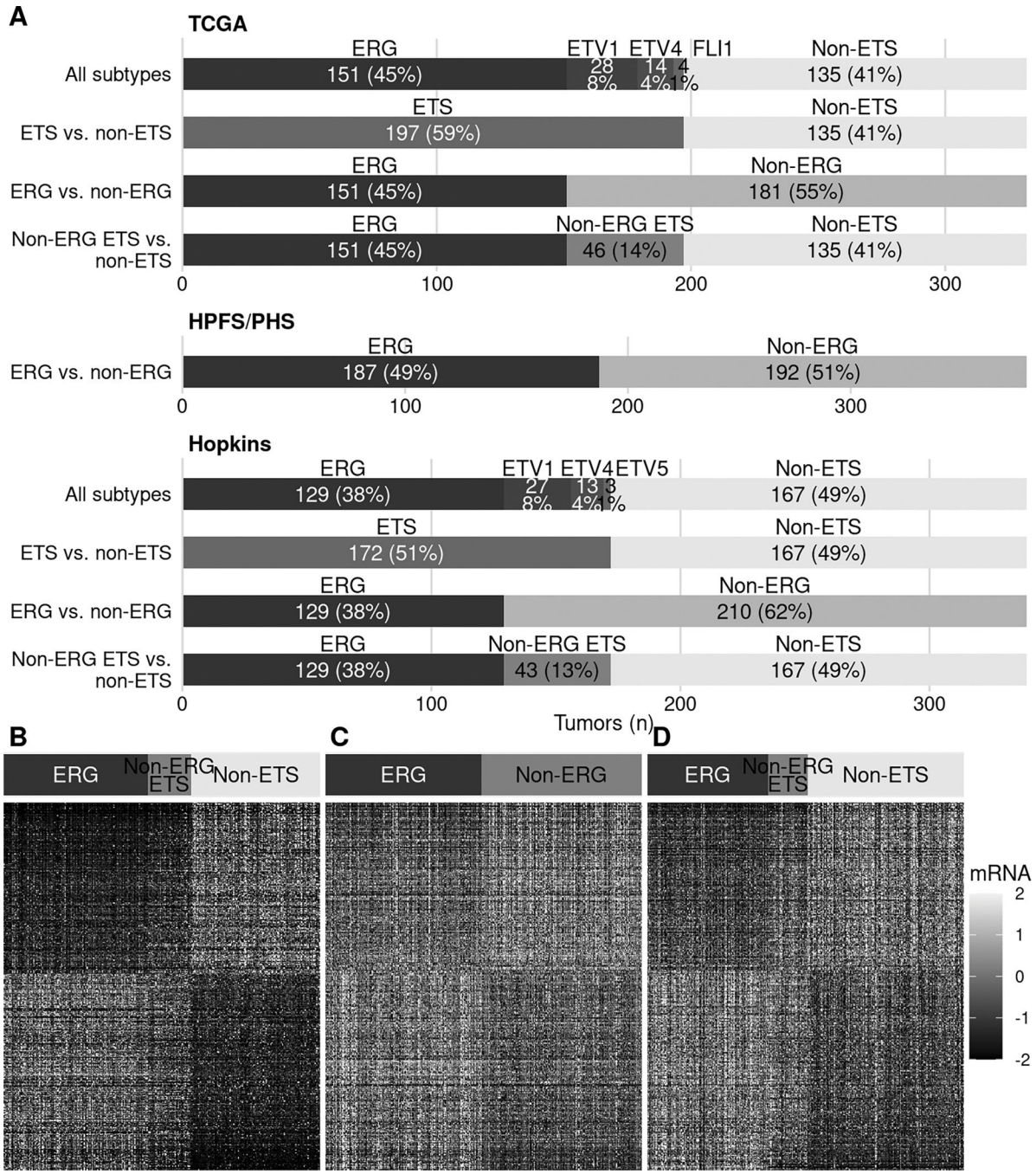


**Figure 1.**

Subtypes of prostate cancer defined by ETS fusions and their combinations in The Cancer Genome Atlas (TCGA), a subset of prostate cancers with gene expression profiling from the Health Professionals Follow-up Study (HPFS) and the Physicians' Health Study (PHS), and the natural history/biochemical recurrence study at Johns Hopkins University (Hopkins). Numbers in the bars are counts of tumors and proportions (**A**). Heatmaps show expression of Berglund et al. ETS genes in TCGA (**B**), HPFS/PHS (**C**), and Hopkins (**D**), capped at the 3rd and 97th within-gene percentiles to reduce influence of outliers and scaled within genes

to [−2, 2]. Tumors are on the *x* axis, sorted by ETS/*ERG* status but without further clustering to avoid overfitting; genes are on the *y* axis, sorted by expression ratios by ETS status in TCGA.

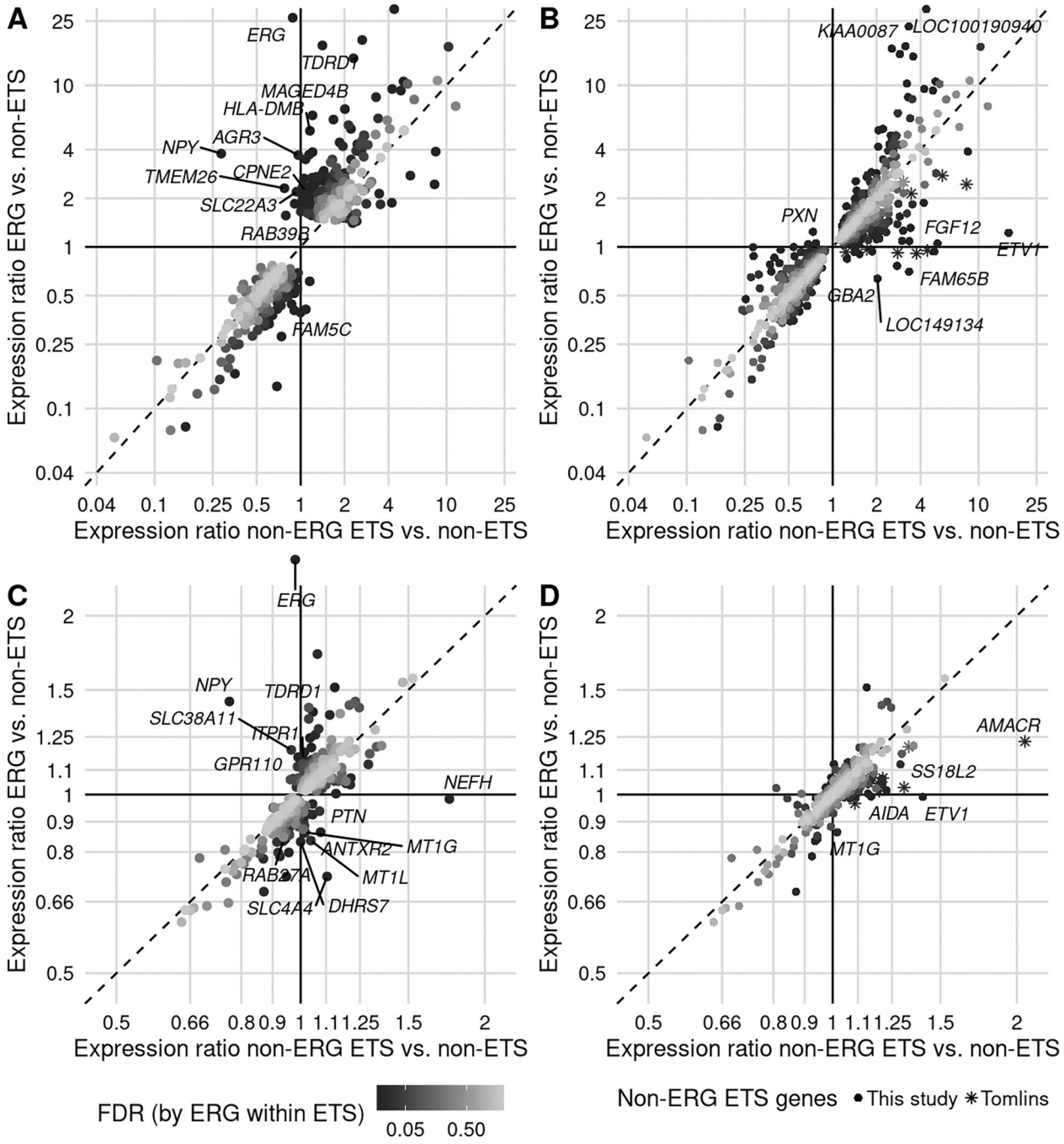


**Figure 2.**

Ratios of gene expression between non-*ERG* ETS tumors *vs.* non-ETS tumors (*x* axis) are compared with ratios of gene expression between *ERG* tumors *vs.* non-ETS tumors (*y* axis) in TCGA (**A**, **B**) and in the Johns Hopkins cohort (**C**, **D**). Genes compared are the 412 Berglund *et al.* ETS genes (**A**, **C**) and the 776 non-ERG ETS genes (**B**, **D**), of which 11 genes (indicated by asterisks) were also part of a non-*ERG* ETS signature by Tomlins *et al.* (2015). The left upper and right lower quadrants contain genes that are regulated in a different direction between the comparisons. Grayscale indicates the false-discovery

rate ($FDR_{int}$) for expression differences between *ERG* and non-*ERG* ETS (*i.e.*, within ETS tumors only). Genes with directionally discordant expression ratios at $FDR_{int} < 0.05$, the 5 genes with the lowest $FDR_{int}$ and concordant directionality, and two outliers by expression ratio (*NEFH* in **C** and *AMACR* in **D**) are labeled.

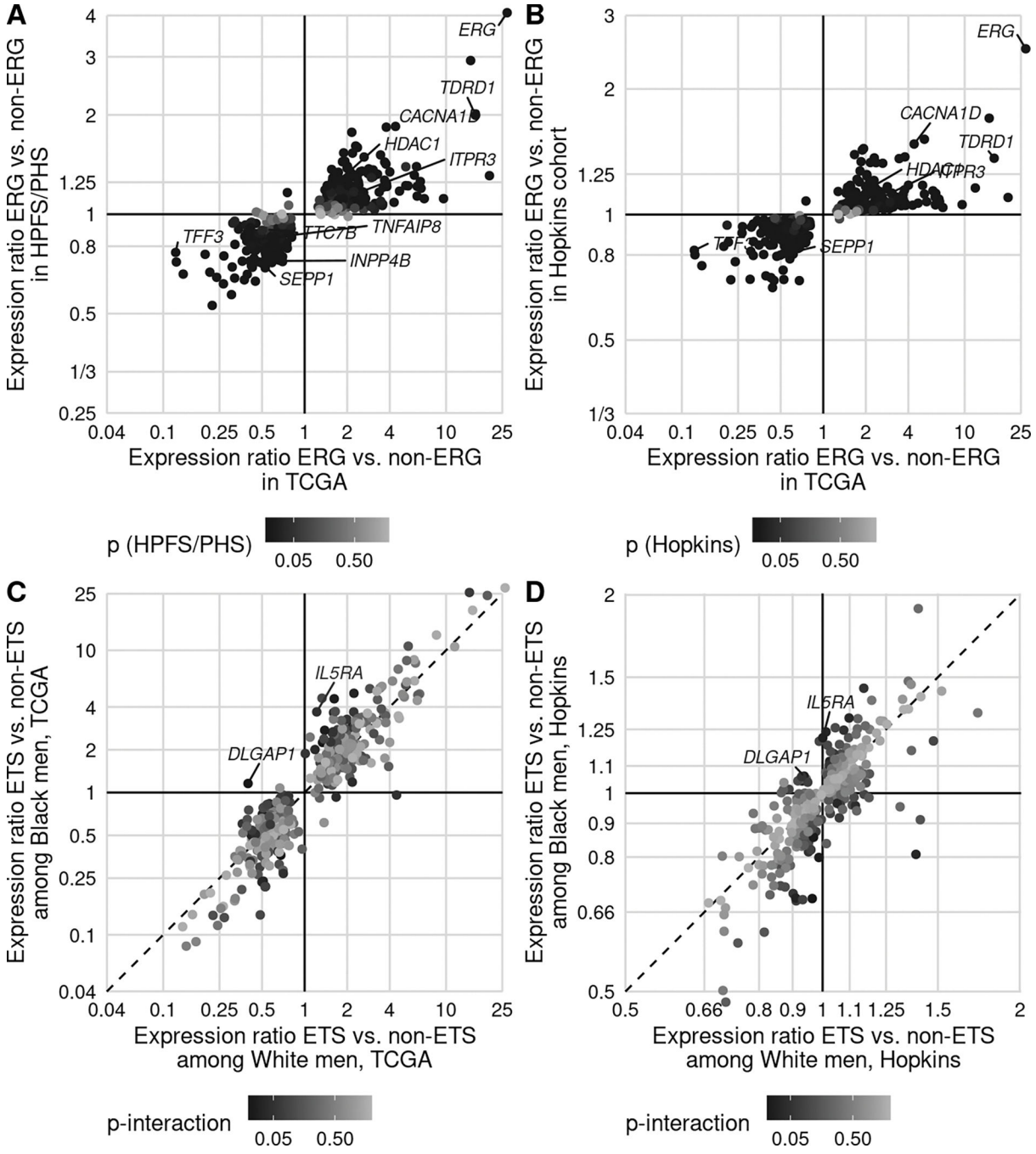


**Figure 3.**

Gene expression ratios between *ERG vs.* non-*ERG* tumors in TCGA one the *x* axis and in HPFS/PHS (**A**) or the Hopkins cohort (**B**) on the *y* axis; note the different scale because of dynamic ranges of RNA sequencing and RNA microarrays. Grayscale indicates *p*-value by ERG status in HPFS/PHS or the Hopkins cohort. Highlighted are genes with the strongest associations (by product of *p*-values in both cohorts per panel). **C** and **D**, Gene expression ratios ETS *vs.* non-ETS tumors in White men (*x* axis) and Black men (*y* axis) in TCGA (**C**) and in the Hopkins cohort (**D**). Grayscale indicates the $FDR_{interaction}$ for race and *ETS*

status. The two genes with directionally concordant ratios and $p_{\text{interaction}} < 0.05$ in both cohorts (but FDR ≥ 0.21 in TCGA or FDR ≥ 0.34 in Hopkins) are highlighted.

**Table 1.**

Characteristics of men with primary prostate cancer and high-quality gene expression data in The Cancer Genome Atlas (TCGA, left; *n* = 332), the Health Professionals Follow-up Study and Physicians' Health Study (HPFS/PHS, middle; *n* = 379), and the natural history/biochemical recurrence study at Johns Hopkins University (Hopkins, right; *n* = 339).[1]

| | TCGA | | | HPFS/PHS | | Hopkins | | |
|---|---|---|---|---|---|---|---|---|
| **Fusion status** | ***ERG*** | **Non-*ERG* ETS** | **Non-ETS** | ***ERG*** | **Non-*ERG*** | ***ERG*** | **Non-*ERG* ETS** | **Non-ETS** |
| ***N*** | 151 | 46 | 135 | 187 | 192 | 129 | 43 | 167 |
| **Age at diagnosis** [years] | 61 (55, 66) | 62 (55, 66) | 62 (57, 66) | 65 (62, 69) | 66 (62, 70) | 60 (55, 64) | 60 (56, 64) | 59 (56, 64) |
| Unknown | 1 | | | | | | | |
| **Self-reported race** | | | | | | | | |
| White | 131 (89%) | 41 (89%) | 98 (77%) | 177 (98%) | 185 (98%) | 122 (95%) | 41 (95%) | 140 (85%) |
| Asian | 4 (3%) | 1 (2%) | 3 (2%) | 0 (0%) | 3 (2%) | | | |
| Black | 12 (8%) | 4 (9%) | 27 (21%) | 1 (1%) | 1 (1%) | 4 (3%) | 2 (5%) | 21 (13%) |
| Other | 0 (0%) | 0 (0%) | 0 (0%) | 2 (1%) | 0 (0%) | 2 (2%) | 0 (0%) | 4 (2%) |
| Unknown | 4 | 0 | 7 | 7 | 3 | 1 | 0 | 2 |
| **PSA at diagnosis** [ng/ml] | 7.2 (5.2, 10.3) | 8.0 (4.8, 14.4) | 8.2 (5.1, 14.6) | 8.0 (5.4, 11.6) | 7.3 (5.4, 13.0) | 7.5 (5.4, 12.6) | 8.9 (6.5, 13.3) | 8.8 (5.7, 13.1) |
| Unknown | 58 | 24 | 64 | 23 | 32 | | | |
| **Gleason grade** | | | | | | | | |
| <7 | 27 (18%) | 8 (17%) | 30 (22%) | 25 (13%) | 33 (17%) | 8 (6.2%) | 3 (7.0%) | 20 (12%) |
| 3+4 | 46 (30%) | 13 (28%) | 42 (31%) | 77 (41%) | 52 (27%) | 53 (41%) | 20 (47%) | 66 (40%) |
| 4+3 | 41 (27%) | 12 (26%) | 25 (19%) | 44 (24%) | 53 (28%) | 32 (25%) | 10 (23%) | 29 (17%) |
| 8 | 16 (11%) | 6 (13%) | 23 (17%) | 19 (10%) | 22 (11%) | 18 (14%) | 4 (9.3%) | 25 (15%) |
| 9–10 | 21 (14%) | 7 (15%) | 15 (11%) | 22 (12%) | 32 (17%) | 18 (14%) | 6 (14%) | 27 (16%) |
| **Clinical stage** | | | | | | | | |
| T1/T2 | 87 (71%) | 17 (47%) | 86 (76%) | 158 (89%) | 167 (90%) | 127 (99%) | 41 (95%) | 164 (98%) |
| T3 | 10 (8%) | 8 (22%) | 9 (8%) | 11 (6%) | 7 (4%) | 1 (1%) | 2 (5%) | 3 (2%) |
| T4/N1 | 25 (20%) | 10 (28%) | 18 (16%) | 1 (1%) | 3 (2%) | | | |
| M1 | 0 (0%) | 1 (3%) | 0 (0%) | 8 (5%) | 9 (4%) | | | |
| Unknown | 29 | 10 | 22 | 9 | 6 | 1 | 0 | 0 |

[1] Data are median (interquartile range) or *n* (%). *ERG* is from *TMPRSS2:ERG* fusion calls based on RNA sequencing (TCGA) or from genomically validated immunohistochmistry for ERG protein expression (HPFS/PHS, Hopkins); non-*ERG* ETS fusions are from RNA sequencing (TCGA) or RNA in-situ hybridization (Hopkins).

**Table 2.**

Gene set analysis for Hallmark gene sets associated with *ERG vs.* non-*ERG* (top), ETS *vs.* non-ETS (middle), and non-*ERG* ETS *vs.* non-ETS status (bottom). All gene sets beyond the top 10 are shown in Supplementary Table 5.

| | | | TCGA | | HPFS/PHS | | Hopkins | |
|---|---|---|---|---|---|---|---|---|
| # | Gene Set | Genes[1] | Direction | FDR[2] | Direction | FDR[2] | Direction | FDR[2] |
| ***ERG vs.* non-*ERG*** [3] | | | | | | | | |
| 1 | KEGG Valine Leucine and Isoleucine Degradation | 44/44/42/44 | Down | 0.009 | Down | <0.001 | Down | <0.001 |
| 2 | Hallmark Myogenesis | 200/200/200/200 | Down | 0.009 | Down | 0.20 | Down | 0.13 |
| 3 | KEGG Oxidative Phosphorylation | 135/116/125/116 | Down | 0.026 | Down | 0.11 | Down | 0.37 |
| 4 | KEGG Systemic Lupus Erythematosus | 140/127/113/121 | Down | 0.026 | Down | 0.039 | Down | 0.55 |
| 5 | KEGG Fatty Acid Metabolism | 42/40/42/40 | Down | 0.039 | Down | 0.039 | Down | 0.003 |
| 6 | KEGG Beta Alanine Metabolism | 22/22/22/22 | Down | 0.048 | Down | 0.001 | Down | 0.002 |
| 7 | KEGG Spliceosome | 128/114/122/112 | Up | 0.026 | Up | 0.82 | Up | 0.86 |
| **ETS *vs.* non-ETS** [4] | | | | | | | | |
| 1 | KEGG Valine Leucine and Isoleucine Degradation | 44/44/--/44 | Down | <0.001 | | | Down | <0.001 |
| 2 | Hallmark Myogenesis | 200/200/--/200 | Down | <0.001 | | | Down | 0.12 |
| 3 | KEGG Fatty Acid Metabolism | 42/40/--/40 | Down | <0.001 | | | Down | 0.008 |
| 4 | KEGG Beta Alanine Metabolism | 22/22/--/22 | Down | 0.002 | | | Down | 0.002 |
| 5 | KEGG Oxidative Phosphorylation | 135/116/--/116 | Down | 0.003 | | | Up | 0.94 |
| 6 | KEGG Glutathione Metabolism | 50/50/--/48 | Down | 0.008 | | | Down | 0.004 |
| 7 | KEGG Drug Metabolism Cytochrome P450 | 72/71/--/69 | Down | 0.019 | | | Down | <0.001 |
| 8 | KEGG Spliceosome | 128/114/--/112 | Up | 0.005 | | | Up | 0.041 |
| 9 | Hallmark MYC Targets V1 | 200/195/--/194 | Up | 0.011 | | | Up | <0.001 |
| 10 | KEGG RNA Degradation | 59/57/--/57 | Up | 0.035 | | | Up | 0.07 |
| **Non-*ERG* ETS *vs.* non-ETS** [5] | | | | | | | | |
| 1 | KEGG Valine Leucine and Isoleucine Degradation | 44/44/--/44 | Down | <0.001 | | | Down | 0.46 |
| 2 | KEGG Fatty Acid Metabolism | 42/40/--/40 | Down | <0.001 | | | Down | 0.35 |
| 3 | KEGG Beta Alanine Metabolism | 22/22/--/22 | Down | <0.001 | | | Down | 0.15 |
| 4 | KEGG Glutathione Metabolism | 50/50/--/48 | Down | <0.001 | | | Down | 0.49 |
| 5 | Hallmark Androgen Response | 101/101/--/100 | Down | <0.001 | | | Down | 0.86 |
| 6 | Hallmark Myogenesis | 200/200/--/200 | Down | <0.001 | | | Down | 0.48 |
| 7 | KEGG Limonene and Pinene Degradation | 10/10/--/10 | Down | 0.003 | | | Down | 0.58 |
| 8 | Hallmark MYC Targets V1 | 200/195/--/194 | Up | 0.009 | | | Up | <0.001 |
| 9 | KEGG Spliceosome | 128/114/--/112 | Up | 0.010 | | | Up | <0.001 |
| 10 | KEGG RNA Degradation | 59/57/--/57 | Up | 0.023 | | | Up | 0.001 |

[1] Total number of genes in this geneset / Genes measured in TCGA with RNA sequencing / Genes measured on HPFS/PHS Affymetrix microarray / Genes measured on Hopkins Affymetrix microarray

[2] False-discovery rate among all 236 gene sets.

[3] All 7 gene sets with FDR < 0.05 in TCGA.

[4] Top 10 gene sets out of 18 gene sets with FDR < 0.05 in TCGA.

[5] Top 10 gene sets out of 28 gene sets with FDR < 0.05 in TCGA.