Molecular differences in transition zone and peripheral zone prostate tumors

Carcinogenesis, 2015, Vol. 36, No. 6, 632–638

doi:10.1093/carcin/bgv051
Advance Access publication April 13, 2015
Original Manuscript

ORIGINAL MANUSCRIPT

Molecular differences in transition zone and peripheral zone prostate tumors

Jennifer A.Sinnott[1,2,3,†], Jennifer R.Rider[1,3,*,†], Jessica Carlsson[4], Travis Gerke[5], Svitlana Tyekucheva[2,6], Kathryn L.Penney[1,3], Howard D.Sesso[1,7], Massimo Loda[8,9], Katja Fall[1,10], Meir J.Stampfer[1,3], Lorelei A.Mucci[1], Yudi Pawitan[11], Sven-Olof Andersson[4] and Ove Andrén[4]

[1]Department of Epidemiology and [2]Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA, [3]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA, [4]Department of Urology, School of Health and Medical Sciences, University of Orebro, Orebro 701 85, Sweden, [5]Department of Epidemiology, College of Medicine and College of Public Health and Health Professions, University of Florida, Gainesville, FL 32611, USA, [6]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA, [7]Divisions of Preventive Medicine and Aging, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA, [8]Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA, [9]Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA 02115, USA, [10]Department of Clinical Epidemiology and Biostatistics, School of Health and Medical Sciences, Orebro University, Orebro 701 85, Sweden and [11]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm 171 77, Sweden

* To whom correspondence should be addressed. Tel: +617 432 3040; Fax: +617 566 7805; Email: rider@hsph.harvard.edu

†These authors contributed equally to this work.

Abstract

Prostate tumors arise primarily in the peripheral zone (PZ) of the prostate, but 20–30% arise in the transition zone (TZ). Zone of origin may have prognostic value or reflect distinct molecular subtypes; however, it can be difficult to determine in practice. Using whole-genome gene expression, we built a signature of zone using normal tissue from five individuals and found that it successfully classified nine tumors of known zone. Hypothesizing that this signature captures tumor zone of origin, we assessed its relationship with clinical factors among 369 tumors of unknown zone from radical prostatectomies (RPs) and found that tumors that molecularly resembled TZ tumors showed lower mortality ($P = 0.09$) that was explained by lower Gleason scores ($P = 0.009$). We further applied the signature to an earlier study of 88 RP and 333 transurethral resection of the prostate (TURP) tumor samples, also of unknown zone, with gene expression on ~6000 genes. We had observed previously substantial expression differences between RP and TURP specimens, and hypothesized that this might be because RPs capture primarily PZ tumors, whereas TURPs capture more TZ tumors. Our signature distinguished these two groups, with an area under the receiver operating characteristic curve of 87% ($P < 0.0001$). Our findings that zonal differences in normal tissue persist in tumor tissue and that these differences are associated with Gleason score and sample type suggest that subtypes potentially resulting from different etiologic pathways might arise in these zones. Zone of origin may be important to consider in prostate tumor biomarker research.

Introduction

In 1988, McNeal (1) described three glandular regions of the prostate: the peripheral (PZ), the central and the transition zones (TZ), which respectively comprise ~70%, 25% and 5% of the prostatic glandular tissue. The PZ is the site of origin of most carcinomas (70–80%), but some arise in the TZ, and (rarely) in the central zone (2,3). Morphological differences between tumors in these

Received: January 17, 2015; **Revised:** March 30, 2015; **Accepted:** April 5, 2015

Abbreviations

AUC	area under the receiver operating characteristic curve
CI	confidence interval
FFPE	formalin-fixed paraffin embedded
HPFS	Health Professionals Follow-Up Study
PHS	Physicians' Health Study
PSA	prostate-specific antigen
PZ	peripheral zone
RP	radical prostatectomy
RR	risk ratio
TURP	transurethral resection of the prostate
TZ	transition zone

zones have been observed (2), and markers of poor prognosis, such as high Gleason score, are observed at higher rates among PZ cancers (3,4), which are therefore thought to be more aggressive. Distinct tumor subtypes may arise in each zone, potentially through unique etiologic pathways, but this has not been substantially investigated. In this small study, we seek to identify a molecular signature of zone of origin and relate the signature to clinical factors. The identification of molecular differences between zones could pinpoint a key source of variability in tumor prognosis and treatment response, help expose zone-specific risk factors and partially explain difficulties in identifying reproducible tissue biomarkers of prostate cancer prognosis.

Motivation for this study originated from an earlier prostate tumor gene expression study nested in two large cohort studies, the Physicians' Health Study (PHS) and the Swedish Watchful Waiting Study (5–7), in which we saw profound differences in gene expression patterns between the two cohorts. An important distinction between the cohorts was the type of tissue specimen available: the PHS cohort consisted primarily of radical prostatectomies (RPs), whereas the Swedish cohort consisted of men undergoing transurethral resection of the prostate (TURP), a non-curative procedure intended to treat urinary symptoms. The zonal origins of tumors from RPs probably reflect the underlying distribution of tumor zones—i.e. 70–80% PZ, 20–30% TZ. However, because the TZ surrounds the urethra, it is more frequently the location of cancer diagnosed incidentally by TURP (4). We therefore undertook this study primarily to test whether the expression differences between the TURPs and RPs could be due to expression differences between TZ and PZ tissue.

Methods

Study participants

The primary analysis uses data from three groups of men with gene expression assessed on the Affymetrix GeneChip Human Gene 1.0 ST array. One group consists of five Swedish individuals with no signs of prostate cancer who were undergoing surgical treatment for bladder cancer and whose prostates were removed during the surgery. These prostates were evaluated in the same way as prostatectomy specimens using whole mount slides and no cancer was detected. Three 0.7 mm cores were taken from glandular areas of the TZ and PZ of the prostates of these men, and gene expression assessed on these samples was used to build a signature of zone. The second group consists of nine Swedish individuals with prostate cancer treated by RP whose tumors were located entirely in either the TZ or the PZ. Here, whole mount slides 1 cm apart were used to confirm that the prostate cancer was located in only one zone. Three 0.7 mm cores were taken from the area of the tumor with the highest density of tumor cells. The tissue from these men is used for testing whether the signature of zone built in normal tissue can successfully classify the zone of tumors.

The third group of men providing data for the primary analysis consists of 369 predominantly white USA men with prostate cancer with RP tissue. Properties of the signature are investigated among these men who come from two USA-based prospective studies that have been ongoing for over 25 years: the PHS and the Health Professionals Follow-Up Study (HPFS). The PHS began in 1986 enrolling 22071 initially healthy physicians; 7001 additional physicians were added in 1997 (8). The HPFS is a prospective cohort study that enrolled 51529 health professionals in 1986 (9). Men in both studies are followed with regular questionnaires to collect information on diet and lifestyle as well as health outcomes including prostate cancer. When a man is diagnosed with prostate cancer, his medical records and archival tumor tissue are requested from the treating hospital. Among the men with available medical records and tissue, we selected cases using an 'extreme' case-control design in order to facilitate a comparison of lethal and indolent disease. Specifically, if a man died of prostate cancer or developed distant metastases, he was considered a *lethal case*; if he did not die or develop metastases and survived at least 8 years after diagnosis, he was considered a *non-lethal case*. These lethal and non-lethal cases were included in our gene expression study; note that the non-lethal cases are a mixture of men who would have survived without treatment and those who have been cured by treatment. Clinical information such as stage and prostate-specific antigen (PSA) at diagnosis was abstracted from medical records, and Gleason score was determined by a standardized review of archival tissue. Two 0.6 mm cores were extracted from regions of high-density tumor (>80% of cells).

Properties of the signature are further investigated in a study using gene expression measured on an Illumina chip designed with ~6000 genes thought to be relevant to cancer. This study included participants from the PHS as well as the population-based Swedish Watchful Waiting Study, which consisted of men diagnosed with incidental prostate cancer discovered through TURP in the Örebro (1977–94) and South East (1987–99) Health Care Regions of Sweden. No PSA screening programs were then in place, and patients were followed expectantly. Archival TURP specimens with high-density tumor regions (>90% of cells) were identified and included. Participants were also selected for lethal and non-lethal phenotypes, as described previously.

Messenger RNA expression profiling

For men in the primary analysis, we utilized archival formalin-fixed paraffin-embedded (FFPE) specimens and undertook whole-genome gene expression profiling. To conduct this profiling in FFPE tissues, whole transcriptome amplification using the WT-Ovation FFPE System V2 (NuGEN, San Carlos, CA) was paired with microarray technologies using the GeneChip Human Gene 1.0 ST microarray (Affymetrix, Santa Clara, CA). For the expression profiles generated, we regressed out technical variables and then shifted the residuals to have the original mean expression values, and normalized using the robust multi-array average method (10, 11). We mapped gene names to Affymetrix transcript cluster IDs using the NetAffx annotations as implemented in the Bioconductor annotation package pd.hugene.1.0.st.v1; this resulted in 20254 unique named genes. Further details are available elsewhere (12).

For men in the secondary analysis, RNA was extracted from FFPE tumor tissue and four complementary DNA-mediated annealing, selection, extension and ligation assay panels were designed for the Illumina platform (Illumina, San Diego, CA) for the discovery of molecular signatures relevant to cancer, as described previously (5). Mapping the probes to gene names resulted in 6090 named genes. Further details on RNA extraction, sample processing and data normalization may be found elsewhere (5).

Statistical methods

In our primary analysis, we focused only on men with gene expression assessed using the Affymetrix chip, with measurements on 20254 genes. To improve our ability to build a meaningful prediction model, we restricted to genes we considered expressed, which we defined as genes with an expression level of at least 5 in at least one individual. This restriction left us with 8041 genes. We then centered the genes to have mean 0.

To build a model for predicting whether tissue comes from the TZ or the PZ, we used a nearest shrunken centroids classifier to predict zone using tissue from the prostates of patients without prostate cancer. A sample from each zone in each individual was included in the gene expression study, for a total of 10 samples. Matching information was not used in model-building; we felt that this choice is reasonable here because we are using these samples only for model-building and not for

making any inference, and it in fact biases us towards a null model if there are dominant between-subject effects. The nearest shrunken centroids classifier, called Prediction Analysis of Microarrays, classifies individuals according to whether their gene expression patterns are closer to a TZ archetype or a PZ archetype, where these archetypes are defined using the average pattern of gene expression among the samples from each tissue type on a particular subset of genes (13). The classifier determines the optimal subset of genes using cross-validation and shrinks the average values towards 0 for model stability. After building the classifier, we can apply it to new individuals to produce an estimated probability that the tissue is from the TZ; we denote this prediction Prob(TZ). When Prob(TZ) is high, the model predicts that the tissue is from the TZ, and when it is low, the model predicts that the tissue is from the PZ. To determine how well this model succeeds in predicting zone in tumor tissue, we applied it to gene expression assessed in prostatectomy samples from nine men with tumors located entirely within one zone: six with TZ tumors and three with PZ tumors. We assessed prediction accuracy using the area under the receiver operating characteristic curve (AUC).

After developing and evaluating this gene expression model for tumor zone, we applied it to men in the PHS and HPFS with tumor tissue of unspecified zone for whom gene expression on tumor tissue was assessed on the same chip. Although we do not know tumor location or origin for these individuals, we used our zone prediction Prob(TZ) as an approximation for zone. We were then interested in two major questions. First, we wanted to test whether the predicted zone of origin Prob(TZ) was predictive of lethal versus non-lethal disease and other clinical features such as Gleason score, stage and PSA at diagnosis. To relate Prob(TZ) to risk of lethality, we fit a logistic regression model predicting lethal disease from Prob(TZ) (and cohort-specific intercepts). To relate Prob(TZ) to the categorical clinical features, we fit linear models predicting logit-transformed Prob(TZ) (a transformation to yield near-normality) from the clinical factor of interest and cohort-specific intercepts. Second, we were curious whether there might be zone-specific prediction models for lethality. That is, we wondered whether there is a different set of genes that predict lethality among tumors originating in the TZ compared with tumors originating in the PZ. To look at this, we separated the data into two discrete groups: estimated PZ and estimated TZ, which we defined using the samples with known zone: estimated PZ tumors had Prob(TZ) lower than the highest value among known PZ tumors, and estimated TZ had Prob(TZ) higher than the lowest value among known TZ tumors. We built a model predicting lethality among the estimated PZ cases and a second model in the estimated TZ cases. We compared the genes selected in both models to see if they were similar and compared them with the genes in a previously developed 157-gene signature for Gleason score as an exploratory analysis (7).

Finally, we returned to the Illumina gene expression data set where we had observed significant differences between men in two cohorts—the PHS RPs and the Swedish TURPs. We quantified the number of significant differentially expressed genes using *t*-tests to illustrate the strong differences among the 6090 genes available. We reconstructed the model for zone using the normal tissue assessed on Affymetrix, after restricting to genes in the overlap of the 6090 genes available on Illumina and the 8041 genes expressed on the Affymetrix chip. In order to make a model in the Affymetrix data set and apply it to the Illumina data set, we inspected the 65 individuals and 3280 genes in the overlap between the two data sets and decided that if we centered each gene in each platform, we produced sufficiently comparable values of gene expression. We then applied the model to the Illumina gene expression data and assessed how well Prob(TZ) predicted TURP tissue using the AUC.

All analyses were conducted in R (14). This study was approved by the institutional review boards at the Harvard School of Public Health and Partners Health Care, as well as the ethics committee at Örebro University Hospital.

Results

Model-building and testing

To build the model for predicting zone, we considered samples of normal TZ and PZ prostate tissue taken from five patients with bladder cancer and used a nearest shrunken centroids classifier (13), finding that a model with 11 genes exhibited good cross-validated prediction accuracy (Figure 1). Applying this model to a new individual's gene expression values produces an estimated probability that the tissue is from the TZ, which we denote Prob(TZ); a high value suggests the tissue is from the TZ and a low value suggests it is from the PZ.

To determine how well this model predicted zone in tumor tissue, we applied it to nine tumors located exclusively in a single zone: six TZ tumors and three PZ tumors. Clinical characteristics as well as predictions Prob(TZ) for these men are displayed in Table 1. In this small sample, prediction was perfect (AUC = 1, $P = 0.01$). The men with TZ tumors had Prob(TZ) >0.69; the men with PZ tumors had Prob(TZ) below 0.12.

Applying model to tumors with unknown zone

We applied the model to men with tumors of unspecified zone taken from RP samples in the PHS and HPFS with expression assessed on the same chip. Clinical characteristics of these men are displayed in Table 2. The distributions of predictions Prob(TZ) are shown in Supplementary Figure 1, available at *Carcinogenesis* Online. Most of the tumors demonstrate expression profiles similar to PZ tumors, which is consistent with the fact that the majority of tumors originate in the PZ.

We tested whether the predicted zone of origin Prob(TZ) was associated with lethal disease, Gleason score, pathological stage and PSA at diagnosis. The odds ratio of association between Prob(TZ) and lethal disease was 0.43 [95% confidence interval (CI) 0.16–1.14; $P = 0.09$]. Though not statistically significant, this odds ratio is consistent with previous suggestions that TZ tumors are less lethal than PZ tumors. Relationships with clinical factors are displayed in Figure 2. Notably, there is a significant association with Gleason ($P = 0.009$): those with lower Gleason are more like TZ tumors. There is no significant association with stage or PSA at diagnosis.

Next, as an exploratory analysis, we built zone-specific prediction models for lethality. We separated the data into two discrete groups: estimated PZ tumors, with Prob(TZ) $\leq$ 0.12 (the highest value among the known PZ samples), yielding 126 cases with 35 lethal events, and estimated TZ tumors, with

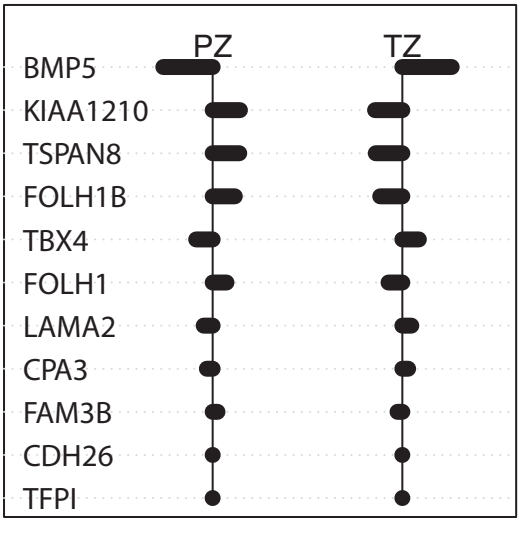

Figure 1. List of the genes in the model built using normal tissue to predict zone (TZ versus PZ) together with the direction and relative importance of the genes in the classifier. Genes with values to the right of the vertical lines are up-regulated in that phenotype.

Table 1. Clinical characteristics of nine patients with known tumor zone used for model testing, together with the model predictions Prob(TZ)

TZ or PZ	Pathological T-stage	Gleason	Initial PSA	Prob(TZ)
TZ	T2	7 (4+3)	5.8	0.98
TZ	T2	7 (4+3)	5.7	0.92
TZ	T2	6 (3+3)	7.9	0.89
TZ	T2	9 (5+4)	7.6	0.88
TZ	T2	6 (3+3)	8.5	0.85
TZ	T2	7 (3+4)	15	0.69
PZ	T2	7 (4+3)	8.5	0.12
PZ	T3	8 (4+4)	6.3	0.12
PZ	T3	8 (4+4)	32	0.01

Table 2. Clinical characteristics of prostate cancer patients with RP tissue in the PHS and HPFS Affymetrix gene expression study

Characteristic	PHS (*N* = 133)	HPFS (*N* = 236)
Years of diagnosis	1987–2005	1986–2004
Gleason score, *N* (%)		
≤6	30 (23)	21 (9)
3+4	47 (35)	91 (38)
4+3	26 (19)	66 (28)
≥8	30 (23)	58 (25)
Lethal cases, *N* (%)	18 (14)	68 (29)
Pathological stage, *N* (%)		
T1/T2	79[a] (59)	141 (60)
T3	48 (36)	81 (34)
T4/N1/M1	6 (5)	14[b] (6)
Age at diagnosis, mean ± SD	65 ± 5.7	65 ± 6.4
PSA at diagnosis, *N* (%)		
<4	15 (12)	18 (9)
4–10	79 (65)	117 (56)
≥10	28 (23)	73 (35)
Missing	11	28

SD, standard deviation.
[a]Two men with missing pathological stage but with clinical stage T1/T2 are included in this category.
[b]One man with missing pathological stage but with clinical stage T4/N1/M1 is included in this category.

Prob(TZ) ≥ 0.69 (the lowest value among the known TZ samples), yielding 46 cases with eight lethal events. The top-ranked genes in these two models are presented in Table 3. Although the numbers were so small that we could not with confidence compare the prediction performance of these models, the model built to predict lethality in the TZ-like cases had different genes than the model built in the PZ-like cases. Nine of the top genes that predict lethality among the PZ cases also appear in a previously described signature for high Gleason score (7), while only two of the top genes that predict lethality among the TZ cases do. However, we cannot rule out that this is due to lack of power to detect a reasonable model, given only eight lethal events.

Applying model to explain differences between RP and TURP

Finally, we returned to our earlier gene expression study where we observed significant differences in expression between men in two cohorts – the PHS RPs and the Swedish TURPs. Clinical covariates are displayed in Table 4. Among 6090 genes on the Illumina chip, 373 (6.1%) demonstrated significant associations with cohort membership after a Bonferroni correction for multiple testing. To test whether this dramatic difference was due to zone, we reconstructed the model predicting zone using the 3280 genes in the overlap between the 8041 genes used in the primary analysis and the genes on the Illumina chip, which reduced the model to six genes: *BMP5*, *TSPAN8*, *FOLH1*, *LAMA2*, *CPA3* and *TFPI*. We evaluated how well the model for TZ predicted TURP tissue in this data, finding excellent prediction (AUC = 0.87, $P < 0.0001$). This suggests that one major contributor to the observed differences in expression between these cohorts is zone. Even after controlling for Prob(TZ), however, a substantial number of genes remain differentially expressed between the cohorts [169 out of 6090 (2.8%), after Bonferroni correction], suggesting that other factors also contribute to the differences in expression between TURP and RP samples.

Discussion

In this study, we investigated whether we could use gene expression to predict the zone in which a tumor originates. Although previous studies have identified expression differences by zone in normal tissue (15,16), we found that a signature for zone-based on normal tissue expression perfectly classified nine tumors located in one zone or the other. Virtually all previous studies define tumor zone as the zone in which 70% of the tumor lies, but we were concerned that this definition may result in misclassification. Thus, our test set consisted of tumors completely contained within a single zone. Because zonal expression differences in benign tissue persisted in tumor tissue, the genes themselves may not only be of interest for prediction, but may also have a role in zone-specific susceptibility to prostate cancer. Among the top genes in our model for zone, *BMP5* (bone morphogenetic protein 5) and *FOLH1*, also known as *PSMA* (prostate-specific membrane antigen) stand out. There has been interest in *BMP* genes in prostate and other cancers because they may influence metastases to bone, which is the most common site of metastasis in prostate cancer (17); *PSMA* has also been implicated in prostate cancer (18).

We further investigated whether the zonal signature was related to clinical features when applied to tumors with unknown zone and found that among 369 such RP samples, tumors that molecularly resembled TZ tumors showed a non-significant trend to lower risk of mortality ($P = 0.09$) largely explained by their generally more favorable Gleason scores ($P = 0.009$), consistent with previous assumptions that TZ tumors tend to have less lethal potential than PZ tumors. The reduced lethality among TZ tumors has been attributed previously to the fact that they are located in an area of the prostate from which it is difficult to spread. Our study suggests that there might also be molecular distinctions. Although identifying zone of origin by eye can be difficult or impossible depending on the tissue type available and the tumor size, a molecular signature such as that proposed here could permit identification of zone of origin for all tumors. With further investigation, this work could have clinical utility, if identification of zone of origin itself proves useful in predicting prognosis or if other prognostic markers vary by zone of origin.

It was with this idea in mind—that the prognostic significance of clinical features or genes could vary depending on where the tumor originated—that we performed the exploratory analysis summarized in Table 3, in which we divided the tumors of unknown zonal origin into those resembling TZ tumors and those resembling PZ tumors, and developed models predicting lethality within those groups. If PZ tumors and TZ tumors progress along similar pathways, we would expect similar genes

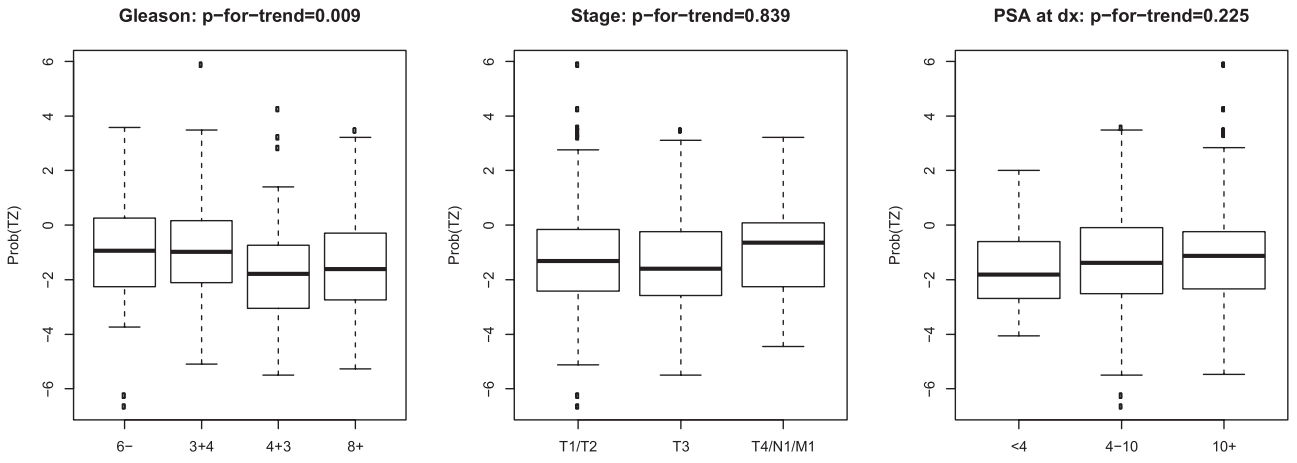

Figure 2. Relationship between logit-transformed Prob(TZ) and clinical factors in the PHS and HPFS cohort. Gleason is categorized in ≤6, 3 + 4, 4 + 3 and ≥8; stage in T1/T2, T3 and T4/N1/M1; and PSA at diagnosis in <4, 4–10 and ≥10.

to appear in models predicting prognosis in these two groups; however, none of the genes selected into the models overlapped. Alternatively, as Gleason is such a strong predictor of disease course, we assumed that the genes in a previously developed signature predicting Gleason score (7) might appear in both zonal models; however, we find that although these Gleason-related genes dominate the model among PZ-like tumors, they are not as prominent in the model among TZ-like tumors. Because the sample sizes (and numbers of lethal events) are limited in these two groups, these results should be interpreted cautiously. However, they support the hypothesis that meaningfully separating tumors based on zone of origin and developing prediction models separately by zone could potentially improve our ability to predict outcomes by illuminating some previously unobserved intertumor heterogeneity.

The zonal signature also demonstrated success discriminating between RP samples and TURP samples, suggesting that some expression differences between these samples may be due to zone. Although zone does appear to be an important contributor, a substantial number of genes unrelated to zone were also differentially expressed. Other factors such as sample processing and fixation may also contribute. This finding has implications for gene expression signature discovery and validation.

When our signature was applied to specimens where tumor zone of origin was unknown, as frequently occurs in practice, our signature produced zone predictions Prob(TZ) on a spectrum between 0 and 1. The predictions at the ends of this spectrum are consistent with patterns from tumors with known zone; however, it is unclear how best to interpret values in the middle of the spectrum. When a tumor originates in one zone and spreads into the other, does its gene expression begin to take on aspects of the new zone? Do there exist additional tumor subtypes that exhibit less 'preference' for a particular zone? Our study is not able to directly address these possibilities. Thus, although we hypothesize that our signature is capturing 'zone of origin,' we cannot rule out that it is only capturing 'zone of current location,' and follow-up studies of tumors involving both zones where zone of origin can be guessed with confidence are needed to distinguish these possibilities.

Our study has several other limitations. We cannot confirm that the gene expression differences are driven by expression in tumor cells or whether they result from contamination by benign tissue. Although we ensured high proportions of tumor cells in our samples, there may be small numbers of normal cells that contribute to the signal we observe in the gene expression. Further studies using microdissection are needed to investigate this issue. Another potential concern is that the signature is simply detecting Gleason score rather than tumor zone. We doubt that the score is directly confounded by Gleason because it was built in benign tissue; moreover, our test set of tumors, while not having matched distributions of Gleason score, at least demonstrate a variety of Gleason scores—6, 6, 7, 7, 7, 9; mean 7 among the TZ samples and 7, 8, 8; mean 7.7 among the PZ samples. However, further signature development and evaluation are certainly needed to determine whether a zonal signature can provide any prognostic information beyond Gleason score. Finally, as mentioned, this study was undertaken as a small study to test whether gene expression differences in TZ and PZ tissue could explain some of the differences observed in Swedish TURP samples and USA RP samples, rather than to fully explore the potential for molecular differences between zones. Thus, although our findings are suggestive, larger studies are needed to identify a robust signature of zone of origin, and to determine its clinical relevance and meaningful cutpoints. Investigations of the signature in cell lines could also help illuminate the signature's significance. In this work, we are putting forward evidence that understanding molecular zonal distinctions might be a fruitful direction to pursue. In particular, we would like to advocate that tumor zone be included among commonly reported clinical features when available because this could provide an avenue for interesting follow-up.

This study has numerous strengths, including the incorporation of comprehensive gene expression data, the use of tumors contained entirely in one zone to reduce misclassification and the ability to investigate signature properties in a large cohort of tumor specimens with unknown zonal origin and long-term follow-up. In addition, we used RP specimens to build the model, removing the potential impact of the TURP procedure itself on gene expression levels.

This study supports our initial hypothesis that zone of origin significantly contributes to molecular differences between TURP- and RP-derived tumors, which may have more widespread implications in epidemiological and clinical studies. First, understanding expression patterns consistent with a particular zone could clarify situations where studies based on TURPs and studies based on RPs produce apparently inconsistent results. For example, there is some evidence that an association may exist between the *TMPRSS2:ERG* fusion and

Table 4. Clinical characteristics of prostate cancer patients with RP tissue in the PHS and TURP tissue in the Swedish Watchful Waiting cohort

Characteristic	PHS (N = 88)	Swedish (N = 333)
Years of diagnosis	1983–2003	1977–98
Gleason score, N (%)		
≤6	10 (11)	102 (31)
3+4	29 (33)	78 (23)
4+3	27 (31)	68 (20)
≥8	22 (25)	85 (26)
Lethal cases, N (%)	17 (19)	152 (46)
Pathological stage, N (%)[a]		
T1/T2	54[b] (61)	
T3	29[c] (33)	
T4/N1/M1	5 (6)	
T1a-N0-MX		87 (26)
T1b-N0-MX		245 (74)
Age at diagnosis, mean ± SD	66±6.3	73±7.0

SD, standard deviation.
[a]One Swedish man missing stage.
[b]Six men with missing pathological stage but with clinical stage T1/T2 are included in this category.
[c]Two men with missing pathological stage but with clinical stage T3 are included in this category.

Table 3. Genes in the zone-specific models for lethality in order of importance, together with indications of direction of association with lethality, and whether the genes belong to the model for zone (Figure 1), or the model for Gleason (7)

Genes in the model for lethality built in estimated PZ cases	Genes in the model for lethality built in estimated TZ cases
AZGP1[a]	*FOLH1B*[b,c]
DPP4[a]	*FOLH1*[b,c]
NRP1[c]	*PRR16*[c]
CPE	*ASPN*[a,c]
MYBPC1[a]	*SRP14*[c]
GREB1[a]	*TMEM60*[c]
CHRNA2[a]	*F2R*[a,c]
KHDRBS3[a,c]	*POLR2K*[c]
TRPM8	*GLO1*[c]
ANTXR2	*HINT2*[c]
ANPEP[a]	*SNORD38B*
PTN[a]	*VCAN*[c]
FAM13C	*LRRIQ1*[c]
RAB27A[a]	*MIR186*[c]

[a]Genes in the signature of Gleason grade (7).
[b]Genes in the TZ/PZ signature.
[c]Genes where up-regulation is associated with increased risk of lethal disease.

survival in TURP cohorts but not RP cohorts. Specifically, in a meta-analysis, men treated by RP with the *TMPRSS2:ERG* fusion were no more likely to have high Gleason score [risk ratio (RR) = 0.85, 95% CI 0.72–1.01, N = 3803] or to die of prostate cancer (RR = 0.99, 95% CI 0.47–2.09, N = 2049) than those without it, whereas among the substantially fewer men diagnosed by TURP, the fusion was associated with higher Gleason (RR = 1.61, 95% CI 1.01–2.57, N = 581) and non-significantly with higher lethality (RR = 1.37, 95% CI 0.53–3.51) (19). It is also thought that the fusion is rare in TZ tumors compared with PZ tumors (20,21). Perhaps among TURP cohorts, the fusion in fact marks tumors of PZ origin, which may have an increased risk of lethality, and thus the moderate association between the fusion and lethality in TURP cohorts is due to confounding by tumor zone. A signature of zone of origin could help control for the effects of zone in such studies. Second, if an association between prognosis and a zonal signature is established in larger studies, such a signature could potentially identify zone of origin and help characterize aggressive potential in TURP and biopsy specimens where only a small portion of the prostate is sampled. Finally, a lack of appreciation for subtypes of prostate cancer with distinct etiologies may contribute to the paucity of established risk factors. Future studies that relate exposures to zone-based subtypes of prostate cancer could identify new zone-specific risk factors.

Supplementary material

Supplementary Figure S1 can be found at http://carcin.oxfordjournals.org/

Funding

The Physicians' Health Study was supported by the National Institutes of Health (NIH) grants CA34944, CA40360, CA097193, HL26490 and HL34595. The Health Professionals Follow-up Study was supported by NIH grants UM1 CA167552 and R01 HL35464. The Swedish Watchful Waiting cohort was supported by the Swedish Cancer Foundation and The Örebro County Council Research Foundation. This work was supported by the NIH grants GM074897 and CA09001 to J.A.S., CA141298 to M.J.S., CA136578 and CA090381 to L.A.M.; by the A. David Mazzone Career Development Award (J.A.S.); by Cancerfonden (CAN 2013/650) and Strategic funding from Örebro University (K.F.), and the Prostate Cancer Foundation (J.R.R. and L.A.M.).

Acknowledgements

We are grateful to Prof. Giovanni Parmigiani for his intellectual contribution to the study. We would like to acknowledge the contributions and collaboration on this project of Dr E.Fox at the Microarray Core Facility at the Dana Farber Cancer Institute; Dr Fox passed away at the end of 2012. We would like to acknowledge the contributions of E.Nuttall and M.Pitt in the gene expression study. We also would like to thank the participants and staff of the Physicians' Health Study and the Health Professionals Follow-up Study, for their valuable contributions, as well as the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA and WY. The authors assume full responsibility for analyses and interpretation of these data.

The study sponsors had no role in the design of the study; the collection, analysis and interpretation of the data; the writing of the manuscript and the decision to submit the manuscript for publication.

Conflict of Interest Statement: None declared.

References

1. McNeal, J.E. (1988) Normal histology of the prostate. Am. J. Surg. Pathol., 12, 619–633.
2. McNeal, J.E. et al. (1988) Zonal distribution of prostatic adenocarcinoma. Correlation with histologic pattern and direction of spread. Am. J. Surg. Pathol., 12, 897–906.
3. Erbersdobler, A. et al. (2002) Tumour grade, proliferation, apoptosis, microvessel density, p53, and bcl-2 in prostate cancers: differences between tumours located in the transition zone and in the peripheral zone. Eur. Urol., 41, 40–46.

4. Greene, D.R. et al. (1991) A comparison of the morphological features of cancer arising in the transition zone and in the peripheral zone of the prostate. J. Urol., 146, 1069–1076.
5. Setlur, S.R. et al. (2008) Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. J. Natl Cancer Inst., 100, 815–825.
6. Sboner, A. et al. (2010) Molecular sampling of prostate cancer: a dilemma for predicting disease progression. BMC Med. Genomics, 3, 8.
7. Penney, K.L. et al. (2011) mRNA expression signature of Gleason grade predicts lethal prostate cancer. J. Clin. Oncol., 29, 2391–2396.
8. Hennekens, C.H. (1989) Final report on the aspirin component of the ongoing Physicians Health Study. N. Engl. J. Med., 321:129–135.
9. Giovannucci, E. et al. (2003) Nutritional predictors of insulin-like growth factor I and their relationships to cancer in men. Cancer Epidemiol. Biomarkers Prev., 12, 84–89.
10. Irizarry, R.A. et al. (2003) Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res., 31, e15.
11. Irizarry, R.A. et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics, 4, 249–264.
12. Penney, K.L. et al. (2015) Association of prostate cancer risk variants with gene expression in normal and tumor tissue. Cancer Epidemiol. Biomarkers Prev., 24, 255–260.
13. Tibshirani, R. et al. (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc. Natl Acad. Sci. USA, 99, 6567–6572.
14. R Core Team. (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
15. Shaikhibrahim, Z. et al. (2012) The peripheral zone of the prostate is more prone to tumor development than the transitional zone: is the ETS family the key? Mol. Med. Rep., 5, 313–316.
16. Der Heul-Nieuwenhuijsen V. et al. (2006) Gene expression profiling of the human prostate zones. BJU Int., 98, 886–897.
17. Bobinac, D. et al. (2005) Expression of bone morphogenetic proteins in human metastatic prostate and breast cancer. Croat. Med. J., 46, 389–396.
18. Kasperzyk, J.L. et al. (2013) Prostate-specific membrane antigen protein expression in tumor tissue and risk of lethal prostate cancer. Cancer Epidemiol. Biomarkers Prev., 22, 2354–2363.
19. Pettersson, A. et al. (2012) The TMPRSS2:ERG rearrangement, ERG expression, and prostate cancer outcomes: a cohort study and meta-analysis. Cancer Epidemiol. Biomarkers Prev., 21, 1497–1509.
20. Falzarano, S.M. et al. (2010) ERG rearrangement is present in a subset of transition zone prostatic tumors. Mod. Pathol., 23, 1499–1506.
21. Guo, C.C. et al. (2009) Prostate cancer of transition zone origin lacks TMPRSS2-ERG gene fusion. Mod. Pathol., 22, 866–871.