Published in final edited form as:

Mod Pathol. 2023 October ; 36(10): 100247. doi:10.1016/j.modpat.2023.100247.

Predicting Prostate Cancer Molecular Subtype with Deep Learning on Histopathologic Images

Eric Erak[1], **Lia DePaula Oliveira**[1], **Adrianna A. Mendes**[1], **Oluwademilade Dairo**[1], **Onur Ertunc**[2], **Ibrahim Kulac**[3], **Javier A. Baena-Del Valle**[4], **Tracy Jones**[1], **Jessica L. Hicks**[1], **Stephanie Glavaris**[1], **Gunes Guner**[5], **Igor Damasceno Vidal**[6], **Mark Markowski**[7], **Claire de la Calle**[8], **Bruce J. Trock**[8], **Avaneesh Meena**[9], **Uttara Joshi**[9], **Chaith Kondragunta**[9], **Saikiran Bonthu**[9], **Nitin Singhal**[9], **Angelo M De Marzo**[1,7,8], **Tamara L. Lotan**[1,7,8]

[1]Department of Pathology, Johns Hopkins University School of Medicine, United States

[2]Department of Pathology, Suleyman Demirel University, Turkey

[3]Koç University School of Medicine, Turkey

[4]Fundacion Santa Fe de Bogota University Hospital, Columbia

[5]Hacettepe University, Turkey

[6]Department of Pathology, University of Alabama School of Medicine, United States

[7]Department of Oncology, Johns Hopkins University School of Medicine, United States

[8]Department of Urology, Johns Hopkins University School of Medicine, United States

[9]AIRA Matrix Private Limited, India

Abstract

Microscopic examination of prostate cancer has failed to reveal a reproducible association between molecular and morphologic features. However, deep learning algorithms trained on hematoxylin and eosin (H&E)-stained whole slide images (WSI) may outperform the human eye and help to screen for clinically-relevant genomic alterations. We created deep learning algorithms to identify prostate tumors with underlying *ERG* fusions or *PTEN* deletions using four stages: A) automated tumor identification; B) feature representation learning; C) classification; and D) explainability map generation. A novel transformer-based hierarchical architecture was trained on a single representative WSI of the dominant tumour nodule from a radical prostatectomy (RP) cohort with known ERG/PTEN status (n=224 and n=205, respectively). Two distinct vision transformer (ViT)-based networks were utilized for feature extraction and a distinct transformer-based model was used for classification. ERG algorithm performance was validated across three RP cohorts, including 64 WSI held out from the pre-training cohort (area under receiver operator characteristic curve or AUC: 0.91), and 248 and 375 WSI from two independent RP cohorts (AUC: 0.86 and 0.89). In addition, we tested ERG algorithm performance in two needle biopsy cohorts comprised of 179 and 148 WSI (AUC: 0.78 and 0.80). Focusing on cases with homogeneous (clonal) PTEN status, PTEN algorithm performance was assessed using 50 WSI held out from the pre-training

To whom correspondence should be addressed: Tamara Lotan, MD, 1550 Orleans Street, Baltimore, MD 21231, tlotan1@jhmi.edu.

cohort (AUC: 0.81) as well as 201 and 337 WSI from two independent RP cohorts (AUC: 0.72 and 0.80) and 151 WSI from a needle biopsy cohort (AUC: 0.75). For explainability, the PTEN algorithm was also applied to 19 WSI with heterogeneous (subclonal) PTEN loss, where the percent tumor area with predicted PTEN loss correlated with that based on immunohistochemistry (r=0.58, p=0.0097). These deep learning algorithms to predict ERG/PTEN status provide proof-of-principle that H&E images can be used to screen for underlying genomic alterations in prostate cancer.

Keywords

Prostate; artificial intelligence; ERG; PTEN; deep-learning

Introduction

In some tumor types, such as renal cell carcinoma, there is a tight correlation between tumor histologic subclass (eg, clear cell renal cell carcinoma) and underlying molecular alteration status (*VHL* gene inactivation). Even within a single histologic tumor subtype, some visually-recognized morphologic features may suggest the presence of a specific genomic finding. For example, pathologists have long recognized that microsatellite-unstable colorectal adenocarcinomas have a prominent inflammatory infiltrate, mucinous differentiation, and absence of necrosis that is tightly correlated to underlying alterations in mismatch repair genes [1,2]. However, in other tumor types, such as prostate cancer, genotype-phenotype correlations as assessed by visual microscopic examination have been rarer, with relatively few exceptions [3], and many are not reproducible across multiple studies [4–6].

Importantly, deep learning algorithms - which are trained on whole slide images (WSI) from hundreds or thousands of individual tumors – can outperform the human eye and improve our ability to screen for underlying molecular alterations using hematoxylin and eosin (H&E)-stained slides. Such algorithms have been developed to augment visual recognition of microsatellite unstable colorectal cancers [7,8], to identify the most commonly mutated genes in lung adenocarcinoma [9] and to predict HER2 status in breast cancer [10]. While sequencing is increasingly recommended for many tumor types to assess for clinically-actionable genomic alterations, it remains expensive, time-consuming, and inaccessible to many patients outside of tertiary care centers. Further, the national shortage of genetic counselors [11] means that germline sequencing cannot be used at scale in common tumor types, and utilization levels remain low in prostate cancer [12]. Thus, deep learning algorithms to screen for genomic alterations using a diagnostic H&E stained slide may provide an inexpensive test to rapidly triage patients who could most benefit from germline or somatic sequencing early in disease progression.

Here, as proof-of-principle, we develop novel vision transformer (ViT)-based deep learning algorithms to identify prostate tumors with two of the most common early genomic alterations, *ERG* gene fusions and *PTEN* gene deletion, using H&E-stained WSI. Across multiple independent radical prostatectomy and biopsy cohorts, these algorithms show

excellent performance and can even predict the spatial pattern of subclonal *PTEN* deletions across a tumor nodule. Given recent promising results in clinical trials of AKT inhibitors for prostate cancer [13,14], deep learning algorithms for alterations such as PTEN loss may be useful to select patients likely to benefit from treatment and will pave the way towards similar screening protocols to select patients who may benefit from germline sequencing or testing for clinically-actionable somatic alterations.

Methods

Patient and tissue cohorts:

Studies were conducted on de-identified human tissues and data under a waiver of consent by the Johns Hopkins Institutional Review Board (IRB). Three previously published Johns Hopkins radical prostatectomy (RP) cohorts were employed for this study: 1) The **Natural History Cohort** is a case-cohort design on the outcome of metastasis in prostate cancer [15] that included 288 cases with known *ERG*-fusion status based on a genetically validated and previously published immunohistochemistry (IHC) assay [16,17]. In addition, 255 cases had known PTEN status, based on a genetically validated IHC assay [18–20] without evidence of heterogeneous (subclonal) loss. 2) The **Case-Cohort** was a non-overlapping case-cohort design comprised of 248 tumors with previously published *ERG*-fusion status based on IHC [21,22]. Similarly, 201 had known and previously published PTEN status, based on genetically validated IHC assay [18–20], without evidence of heterogeneous (subclonal) loss. 3) The **Race Cohort** was a grade-matched cohort comprised of self-identified Black or White men, of whom 375 cases had previously published *ERG*-fusion status determined by IHC [23] and 337 had known PTEN status based on IHC [23] without evidence of heterogeneous (subclonal) loss. Whole slide images from 19 additional RP cases with known heterogeneous (subclonal) PTEN IHC loss on whole slide images from the Natural History Cohort or other studies were included for prediction of the spatial distribution of PTEN loss to validate the PTEN prediction algorithm.

In addition to the radical prostatectomy cohorts, two unpublished needle biopsy cohorts were utilized for this study. 1) The **Radiation Therapy Cohort** included 202 prostate needle biopsies performed at Johns Hopkins from 1991–2010, followed by radiation therapy at our institution and clinical follow-up. A total of 179 patients had adequate tumor volume, known ERG status based on genetically validated IHC assay and 151 had known PTEN status based on genetically validated IHC assay without evidence of heterogeneous (subclonal) loss. 2) **The Active Surveillance Cohort** included 148 needle biopsies performed between 2014–2021 at Johns Hopkins on patients undergoing Active Surveillance with available WSI, >5% tumor involvement by tissue area, and known ERG status by genetically validated IHC assay; this cohort did not have available PTEN IHC data.

Whole slide images:

Cohorts were scanned using the Hamamatsu S360 digital slide scanner at 40x or 20x magnification (Supplementary Table S1).

ERG and PTEN immunohistochemistry staining and scoring:

ERG immunohistochemistry has been described previously and is an excellent surrogate for the presence of underlying *ERG* gene fusions by fluorescence *in situ* hybridization (FISH) [16] or gene expression profiling [17]. Briefly, the protocol utilizes a rabbit monoclonal anti-ERG antibody (Ventana, Roche) with the Optiview secondary detection kit on the Ventana Discovery Ultra. Cases are scored as ERG-positive if any tumor nuclei show staining, and staining is most often homogeneous and clonal within a dominant tumor nodule since this is a relatively early genetic event during tumor development [24]. ERG staining and scoring were performed on tissue microarrays (TMAs) comprised of 3–4 0.6 mm diameter tumor spots for the radical prostatectomy cohorts (the TMA spots were taken from the same donor block subsequently used to generate the WSI H&E, see Figures 1 and 2). ERG staining and scoring was performed on standard histologic slides for the prostate needle biopsies.

PTEN status immunohistochemistry has been described previously and is highly correlated with the underlying presence of *PTEN* gene deletions [18–20]. Briefly, it utilizes a rabbit monoclonal anti-PTEN antibody (D4.3, Cell Signaling Technology) with the Optiview secondary detection kit on the Ventana Discovery Ultra. Cases are scored as having PTEN loss if any tumor cells have complete loss of cytoplasmic and nuclear staining, and staining is often heterogeneous indicating underlying subclonal *PTEN* deletions which most frequently occur subsequent to *ERG* rearrangement [24]. PTEN staining and scoring was performed on TMAs for the radical prostatectomy cohorts comprised of 3–4 0.6 mm diameter tumor spots (the TMA spots were taken from the same donor block subsequently used to generate the WSI H&E assessed by deep learning, see Figures 1 and 2). PTEN staining and scoring was performed on standard histologic slides for the prostate needle biopsies, and for a subset of 19 radical prostatectomy cases with known heterogeneous (subclonal) PTEN status based on TMA scoring (see below).

For 19 tumors at RP with known PTEN heterogeneity (subclonal loss), tumor areas with intact PTEN (red) and PTEN loss (blue) were manually annotated by a trained genitourinary pathologist (EE) on the IHC-stained slide (Figure 2). The pathologist was blinded to the deep learning algorithm prediction at the time of annotation. The sum of all annotated areas with PTEN loss was divided by the total annotated tumor area to derive the percent of the tumor area with PTEN loss by IHC. For each case, this value was compared to the predicted area of tumor with PTEN loss based on the deep learning algorithm using the H&E-stained slide (see below).

Deep learning algorithm:

We present a self-supervised learning-based technique that exploits the hierarchical structure of tumor morphology at several image resolutions. The system includes four stages: 1) Tumor identification; 2) Feature representation learning; 3) Classification; 4) Explainability maps generation (Figure 1).

Tumor Identification: In this study, SegFormer [25] was used as the segmentation network. It is a semantic segmentation network with a multi-layer perceptron (MLP) decoder and a transformer. 110 images of radical prostatectomy from the Natural History Cohort were used

to train the model (these did not include the 64 hold-out samples for ERG validation or the 50 hold-out samples for PTEN validation). The network uses a hierarchical transformer encoder similar to Swin Transformer [26], but without positional encoding. Due to the hierarchical nature of the encoder, the network can produce multi-level, multi-scale features with high-resolution coarse features and low-resolution fine-grained features. The absence of positional encoding enables us to apply the model to patches with varied resolutions without considerably reducing performance, which is one of the most major limitations of transformer-based networks. Its decoder is meant to be lightweight and efficient, creating powerful representations while remaining simple and computationally intensive.

We employ the deep active learning framework proposed by Singhal et al. [27] to reduce the annotation effort of pathologists for this supervised learning technique. On training datasets, the system simulates active learning-based iterative data labelling. In the initial iteration, the model is trained using a small labelled dataset. The trained SegFormer model is then supplied unlabeled images, and a measure of uncertainty is computed for each unlabeled sample. Then, a pathologist annotates the samples with the highest uncertainty (measured uncertainty greater than the criterion) and adds them to the initial training set. In the subsequent iteration, a new set of annotated images is used to retrain the model. This phase is repeated until no samples have an uncertainty measurement that exceeds a predetermined threshold. The framework helped reduce the pathologist's burden for annotating tumor regions on training cases by 60 to 70 percent. A representative comparison of pathologist annotation of tumor-containing areas and the deep learning algorithm can be seen in Figure 2A. Algorithms were run at 20x magnification. At 20x magnification, the discovered tumor region is divided into non-overlapping patches of two different sizes, 256 × 256 and 4096 × 4096. N patches of size 4096 × 4096 are extracted from the detected tumor location. The 256 × 256 patches represent fine-grained features (cell interactions), whereas the 4096 × 4096 patches reflect features at the region level (interactions between cell clusters).

Feature Representation Learning: We pre-train two vision transformers (ViTs) [28] on the tumor patches using the DINO [29] method for learning and extracting qualitative features in an unlabeled, unsupervised manner. For 256 × 256 size patches, a ViT with 12 encoder layers (ViT-S) is pre-trained. The N x 256 × 384 tensor extracted from ViT-S is input to another ViT model with 6 encoder layers (ViT-XS). Supplementary Table S2 lists the length of the feature vector, the number of heads, the depth, and the MLP size for both ViT models. In the classification architecture, the 6 layers from ViT-XS are applied as pre-trained weights. Figure 1B depicts the training process for ViT-S and ViT-XS, respectively.

The ViT-S model is pre-trained on a total of 0.75 million 256X256 patches (from 331 WSI) using DINO for 300 epochs with a batch size of 128 and an initial learning rate of 0.0005 with a cosine scheduler for decay. During ViT-S pre-training, 8 local views of size 96 × 96 and 2 global views of size 224 × 224 are configured for data augmentation in DINO. The ViT-XS model is pre-trained with DINO on a total of 28 thousand N x 256 × 384 tensors (extracted from ViT-S) for 100 epochs with a batch size of 128. During ViT-XS pre-training, 8 local views of size 6 × 6 and 2 global views of size 14 ×14 are configured for data augmentation in DINO.

Feature Extraction & Classification: For training a model to predict WSI-level molecular markers, we use a pre-trained ViT-XS with a 2-layer classification head (Figure 1C). The architecture consists of a vision transformer with 8 encoder layers (ViT-CL), followed by Global Attention Pooling [30] and Multi-Layer Perceptron blocks with WSI-level class labels as the output. The weights of the VIT-XS model are used to initialize the first 6 layers of the classification model, while the latter two layers are initialized randomly. The model is trained on WSIs (with WSI-level class labels of a specific molecular marker) in a supervised manner for 50 epochs with a batch size of 1 and an initial learning rate of 0.0001 with a cosine scheduler for decay. We employed the Multiple-Instance-Learning (MIL) [31] approach for WSI-level classification training to eliminate the issue of varying amounts of patches in a WSI as model input.

Explainability maps generation for PTEN loss prediction: The objective here is to generate a visual heatmap that highlights the image region that corresponds to the class predicted by a model. Using a sliding window technique with a 128 × 128 pixel step size, 1024 × 1024 overlapping regions are extracted from the tumor region identified by the SegFormer model at 20x magnification. On each patch, the ViT-S trained model is employed to generate features. Using the model's inference, an AI score is estimated for each patch's features. A heatmap of a WSI level is generated by piecing together individual regions and superimposing the predicted score. Using a reduced patch size and regions that overlap, we generate a fine-grained explanation map. The heatmap area is quantified so that the proportion of tumor area corresponding to PTEN loss can be calculated.

Implementation Details: All models were trained and tested using the NVIDIA A6000 GPU, and the code was written using PyTorch framework in Python (see Supplementary Table S3 for model size and inference time).

Statistical analysis:

Area under the receiver operator characteristic curve (AUC) was used to benchmark accuracy of the deep learning molecular status prediction algorithms compared to the ground truth molecular status based on immunohistochemistry. AUC was calculated in Python. To mitigate the overfitting and data distribution issues, we employ a k-fold cross validation strategy. The training data is broken down into k subsets. On k-1 folds, a model is trained, and the resulting model is validated using the remaining (held-out) data. This procedure is repeated for each fold. The AUC performance measure reported by k-fold cross-validation is then the average of the values. We use k=5 in all the experiments. For 19 WSI, predicted percent tumor area with PTEN loss based on the deep learning algorithm and actual percent tumor area with PTEN loss based on pathologist-annotated immunostained slides were compared using Spearman correlation (GraphPad).

Results

ERG algorithm performance in radical prostatectomy and needle biopsy samples:

ERG algorithm training utilized 224 H&E-stained WSI from the Natural History radical prostatectomy cohort, of which 46% (104/224) were ERG-positive by

immunohistochemistry (IHC). All ERG-positive cases showed homogeneous (clonal) positivity within the tumor tissue tested, as has previously been described for this early genomic alteration [24,32–34]. Testing was conducted on 64 WSI originally held out from the training cohort, of which 41% (26/64) were ERG-positive, with an AUC on five-fold-cross validation of 0.91±0.034 (Table 1). Performance of these models was further assessed on 248 WSI from radical prostatectomies in the Case Cohort, of which 43% (107/248) were ERG-positive by IHC, with an AUC on five-fold-cross validation of 0.86±0.025. Since the frequency of *ERG* fusion in prostate cancer varies considerably by race [35,36], to ensure algorithm performance in a more racially diverse cohort, we examined performance in 375 WSI from radical prostatectomies in the Race Cohort. Among these cases, 34% (127/375) were ERG-positive, and the algorithm had an AUC on five-fold-cross validation of 0.89±0.007.

Needle biopsies have dramatically less tumor content on a WSI compared to a standard quadrant section of a radical prostatectomy, but diagnostic needle biopsies are the sample type that would be most useful to screen for underlying genomic alteration status in order to select patients who could benefit from confirmatory sequencing. Thus, we next tested the ERG algorithm performance in 179 WSI from the Radiation Therapy needle biopsy cohort, of which 39% (70/179) were ERG positive, achieving an AUC on five-fold-cross validation of 0.78±0.017. Similar performance was seen in a contemporary Active Surveillance needle biopsy cohort, where 40% (59/148) were ERG-positive by IHC, with an AUC on five-fold-cross validation of 0.80±0.029 (Table 1).

PTEN algorithm performance in radical prostatectomy and needle biopsy samples:

Because PTEN deletion is subclonal or heterogeneous in as many as 40% of primary prostate tumors [18–20], we trained PTEN loss deep learning models exclusively on cases with homogeneous PTEN status (intact or loss) by IHC. Training utilized 205 H&E-stained WSI from the Natural History radical prostatectomy cohort, of which 27% (56/205) had homogeneous (clonal) PTEN loss by IHC. Testing was conducted on 50 WSI originally held out from the training cohort, of which 26% (13/50) had homogeneous PTEN loss by IHC, with an AUC on five-fold-cross validation of 0.81±0.043 (Table 2). Performance of these models was further assessed on 201 WSI from radical prostatectomies in the Case Cohort, of which 29% (58/201) had homogeneous PTEN loss by IHC, with an AUC on five-fold-cross validation of 0.72±0.014. Because PTEN deletion is less frequent among African-American compared to European descent patients [36,37], we next examined performance of the PTEN algorithm in 337 WSI from radical prostatectomies in the Race Cohort, of which 17% (56/337) had homogeneous PTEN loss on IHC, with an AUC on five-fold-cross validation of 0.80±0.015. Finally, we tested PTEN algorithm performance in 151 WSI from the Radiation Therapy needle biopsy cohort, of which 19% (28/151) had homogeneous PTEN loss, with an AUC on five-fold-cross validation of 0.75±0.028 (Table 2).

Algorithm explainability via subclonal *PTEN* deletion mapping:

PTEN IHC can provide cellular-level resolution for areas of subclonal *PTEN* gene deletion within a dominant tumor nodule of a radical prostatectomy sample. Having developed the PTEN loss detection algorithm using tumors with homogeneous (clonal) PTEN status based

on IHC staining, we next tested whether the deep learning algorithms could accurately predict PTEN loss in 19 radical prostatectomy cases with known heterogeneous (subclonal) PTEN loss by IHC. For the ground-truth comparison, a pathologist utilized the PTEN IHC slide to annotate areas of PTEN loss. By visual comparison, the predicted and actual PTEN status maps were highly concordant (Figure 2) with some exceptions (Supplementary Figure S1). Finally, the percent tumor area with predicted PTEN loss based on deep learning significantly correlated with the percent tumor area with PTEN loss based on IHC ($r=0.58$, $p=0.0097$). This novel methodology that not only makes the algorithm explainable, but also facilitates the quantification of the percentage of loss in the case of PTEN.

Discussion

Prostate cancer is the most commonly diagnosed solid tumor among men in the US (excluding skin cancer), and as such, it has been a key test case for the development of deep learning algorithms that aid the pathologist in tumor diagnosis and grading. Though the majority of prostate cancer cases are not lethal, some molecular subsets of prostate cancer have prognostic implications, are predictive of therapy response, or may be important to identify for associated risk of underlying germline alterations. PTEN loss, for example, is highly prognostic in localized prostate cancer [18,38,39] and may also be predictive in the metastatic setting as increasingly potent AKT inhibitors are being tested in clinical trials [13,14]. Pathogenic mutations in DNA repair genes - such as those involved in mismatch repair or homologous recombination - are also enriched in lethal prostate cancer [40,41] and make patients eligible for PARP inhibition or immune checkpoint inhibition at the time of metastatic disease [42,43]. These mutations may also be germline in up to half of the time that they occur [40,41], thus it is critical to identify tumors with these alterations as early as possible in the disease course.

Despite recommendations by NCCN that all high risk patients and some subsets of intermediate risk localized prostate cancer patients receive germline sequencing, with both germline and somatic sequencing offered to metastatic patients [44], the national shortage of genetic counselors and expense of sequencing has made this impractical in many care settings [45,46]. The lack of uptake for sequencing in non-academic medical centers may pose an increasing threat to health equity, as precision medicine approaches may not be equally available to all patients [46]. Thus, for a common tumor type like prostate cancer, the development of deep learning algorithms to help triage patients who may particularly benefit from sequencing, could fill an unmet clinical need and reduce the economic burden of sequencing all cases, as well as reducing the burden on the genetic counseling system.

Here, as proof-of-principle that machine learning algorithms can be applied to H&E images of prostate cancer to identify underlying genomic alterations, we leveraged two of the most common genomic changes in primary prostate cancer: *ERG* gene fusions and *PTEN* gene deletion. *ERG (ETS related gene)* is a member of the E-26 transformation-specific family of transcription factors that is involved by a genomic rearrangement resulting in its overexpression in about half of prostate cancers arising in men of European descent. *ERG* rearrangements are an early event during tumorigenesis and are most commonly clonal within a given tumor nodule [24,32–34]. Though not prognostic in surgically treated patients

on its own, ERG expression may have prognostic significance in the context of PTEN loss, where it is associated with improved outcomes [38,47]. In contrast to ERG, loss of *PTEN* function is frequently subclonal in a significant fraction of primary tumors [19,24,32,48], and has been strongly associated with adverse oncologic outcomes in prostate carcinoma [38,39,49], occurring in as many as 40% of cases. Recent trials of Ipatasertib, a potent AKT inhibitor, have suggested that PTEN loss may also be predictive of response to targeted therapy [13,14].

Though both ERG expression and PTEN loss can be assessed in prostate cancer by relatively inexpensive immunohistochemistry assays that are highly concordant with underlying genomic status, potentially obviating the need for a deep learning prediction algorithm, the frequency of these alterations makes them an excellent test case for the development of machine learning algorithms to identify molecular alterations. By visual inspection, prior studies have suggested that some morphologic features may be significantly associated with *ERG* gene rearrangement, including blue-tinged mucin, cribriform growth pattern, macronucleoli, intraductal tumor spread, and signet-ring cell features [4]. However, these features are exceedingly common in primary prostate cancer and are not highly specific for *ERG* alterations. PTEN loss has also be associated with foamy cell change and intraductal carcinoma, though specificity of these features is similarly imperfect [50]. Leveraging thousands of H&E images and an enormous feature space, it is not surprising that artificial intelligence-based algorithms might be able to identify features associated with ERG rearrangements more accurately than the human eye.

To our knowledge, only one prior published study applied deep-learning based algorithms to identify *ERG* fusions in prostate cancer [51]. This study utilized training and validation datasets including 261 and 131 H&E radical prostatectomy images, respectively, with AUCs ranging from 0.82 to 0.85. In the current study, we build on this work to develop novel deep learning models trained on a similarly sized radical prostatectomy dataset with validation in 3 independent radical prostatectomy cohorts comprising 687 tumors, and 327 needle biopsies. Though AUCs ranged from 0.86 to 0.91 in the radical prostatectomy cohorts, performance dropped in the needle biopsy cohorts to an AUC of 0.78 to 0.79, almost certainly due to the four- to six-fold smaller tumor area sampled in the latter. A notable limitation of this study is its single-institution design, and future work will expand our findings to outside cohorts. However, this result provides the first evidence, to our knowledge, that deep learning algorithms for molecular prediction can be used in diagnostic needle biopsies, the most clinically relevant sample type for wide-spread application of this type of screening. In future studies, training on large needle biopsy cohorts may help to develop algorithms with improved performance in these small tissue samples.

Our study is also the first, to our knowledge, to apply deep learning algorithms to identify prostate cancers with PTEN loss. Due to the high frequency of subclonal alterations in PTEN [19,24,32,48], predicting PTEN status based on H&E is particularly challenging. Here, we trained and initially tested algorithms in cases that were highly likely to have clonal PTEN alterations, based on homogeneous PTEN loss or intact PTEN using immunohistochemistry of tissue microarray tumor samples. Though PTEN algorithm performance in this dataset was reasonable, with AUCs ranging from 0.72 to 0.81, the performance was notably less optimal than we saw for ERG algorithms. This may be due

in part to the fact the tissue microarrays (where PTEN status was assessed for all cases) may have undersampled cases with focal subclonal PTEN alterations, leading to mislabeling of the PTEN status. In future studies, the ground truth labeling for PTEN status should be based on IHC assessment on the same standard radical prostatectomy quadrant sections as used for algorithm training and testing. Despite this limitation, the PTEN algorithm developed on clonal cases actually performed reasonably well when applied to WSI of cases with known subclonal PTEN loss in order to test algorithm explainability. By heat mapping PTEN status prediction using the deep learning algorithm across small tumor tiles in these cases, to our knowledge, we are the first to examine explainability of these molecular prediction models. In this analysis, we saw a remarkable association between areas of tumor with PTEN loss based on the immunohistochemistry-stained slide and the deep learning PTEN status prediction based on H&E images. More concretely, the percent tumor area with predicted PTEN loss based on immunohistochemistry correlated significantly with the percent tumor area with PTEN loss based on deep learning.

In summary, we present one of the first proof-of-principle studies to use deep learning models trained on H&E images of prostate cancer to predict underlying molecular subtype. Though we focused on the most common molecular alterations in this first study, future studies will test similar models for rarer molecular subclasses of prostate cancer which may have more immediate clinical relevance, such as those with DNA repair alterations. Such studies will require large multi-institutional cohorts or clinical trials to provide adequate sample numbers and careful correlation with sequencing for evidence of bi-allelic loss where IHC assays are not available. However, once developed, these algorithms may prove instrumental to the inexpensive and rapid triaging of early stage prostate cancers at high risk of harboring a pathogenic alteration in DNA repair genes. With the increasing use of machine learning algorithms in daily pathology practice for tumor grading and staging, similar molecular triage algorithms could provide a quick screen for clinically actionable alterations using the ubiquitously available diagnostic H&E, reducing the burden on our genetic testing and counseling services for this common tumor type, and improving universal access to precision medicine.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Financial support:

This research was supported by a grant from the Prostate Cancer Foundation. Additional funding was from the NCI Cancer Center Support Grant 5P30CA006973-52, NIH/NCI SPORE in Prostate Cancer: P50CA58236, and the NIH/NCI U01 CA196390 for the Molecular and Cellular Characterization of Screen Detected Lesions (MCL), the U.S. Department of Defense Prostate Cancer Research Program (PCRP): W81XWH-18-2-0015.

Conflict of Interest Statement:

AM, UJ, CK, SB and NS are employees of AIRA Matrix; TLL reports research funding from AIRA Matrix for this project, as well as from DeepBio, Roche, and Myriad Genetics for other projects.

References

1. Gologan A. et al. Performance of the revised Bethesda guidelines for identification of colorectal carcinomas with a high level of microsatellite instability. Arch Pathol Lab Med 129, 1390–1397 (2005). [PubMed: 16253017]
2. Halvarsson B, Anderson H, Domanska K, Lindmark G. & Nilbert M. Clinicopathologic factors identify sporadic mismatch repair-defective colon cancers. Am J Clin Pathol 129, 238–244 (2008). [PubMed: 18208804]
3. Mehra R. et al. Highly Recurrent IDH1 Mutations in Prostate Cancer With Psammomatous Calcification. Mod Pathol 36, 100146 (2023).
4. Mosquera JM et al. Morphological features of TMPRSS2-ERG gene fusion prostate cancer. J Pathol 212, 91–101 (2007). [PubMed: 17385188]
5. Risbridger GP et al. Patient-derived xenografts reveal that intraductal carcinoma of the prostate is a prominent pathology in BRCA2 mutation carriers with prostate cancer and correlates with poor prognosis. Eur Urol 67, 496–503 (2015). [PubMed: 25154392]
6. Lozano R. et al. Association between BRCA2 alterations and intraductal and cribriform histologies in prostate cancer. European journal of cancer 147, 74–83 (2021). [PubMed: 33626496]
7. Kather JN et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med 25, 1054–1056 (2019). [PubMed: 31160815]
8. Yamashita R. et al. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. The Lancet. Oncology 22, 132–141 (2021). [PubMed: 33387492]
9. Coudray N. et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med 24, 1559–1567 (2018). [PubMed: 30224757]
10. Farahmand S. et al. Deep learning trained on hematoxylin and eosin tumor region of Interest predicts HER2 status and trastuzumab treatment response in HER2+ breast cancer. Mod Pathol 35, 44–51 (2022). [PubMed: 34493825]
11. Hoskovec JM et al. Projecting the Supply and Demand for Certified Genetic Counselors: a Workforce Study. Journal of genetic counseling 27, 16–20 (2018). [PubMed: 29052810]
12. Aguiar JA et al. Utilization of genetic testing in men with advanced prostate cancer. Prostate 83, 516–523 (2023). [PubMed: 36591888]
13. de Bono JS et al. Randomized Phase II Study Evaluating Akt Blockade with Ipatasertib, in Combination with Abiraterone, in Patients with Metastatic Prostate Cancer with and without PTEN Loss. Clin Cancer Res 25, 928–936 (2019). [PubMed: 30037818]
14. Sweeney C. et al. Ipatasertib plus abiraterone and prednisolone in metastatic castration-resistant prostate cancer (IPATential150): a multicentre, randomised, double-blind, phase 3 trial. Lancet 398, 131–142 (2021). [PubMed: 34246347]
15. Ross AE et al. Tissue-based Genomics Augments Post-prostatectomy Risk Stratification in a Natural History Cohort of Intermediate- and High-Risk Men. Eur Urol 69, 157–165 (2016). [PubMed: 26058959]
16. Chaux A. et al. Immunohistochemistry for ERG expression as a surrogate for TMPRSS2-ERG fusion detection in prostatic adenocarcinomas. Am J Surg Pathol 35, 1014–1020 (2011). [PubMed: 21677539]
17. Torres A. et al. Comprehensive Determination of Prostate Tumor ETS Gene Status in Clinical Samples Using the CLIA Decipher Assay. J Mol Diagn 19, 475–484 (2017). [PubMed: 28341589]
18. Lotan TL et al. PTEN protein loss by immunostaining: analytic validation and prognostic indicator for a high risk surgical cohort of prostate cancer patients. Clin Cancer Res 17, 6563–6573 (2011). [PubMed: 21878536]
19. Lotan TL et al. Analytic validation of a clinical-grade PTEN immunohistochemistry assay in prostate cancer by comparison with PTEN FISH. Mod Pathol 29, 904–914 (2016). [PubMed: 27174589]
20. Lotan TL et al. PTEN loss detection in prostate cancer: comparison of PTEN immunohistochemistry and PTEN FISH in a large retrospective prostatectomy cohort. Oncotarget 8, 65566–65576 (2017). [PubMed: 29029453]

21. Vidal I. et al. GSTP1 positive prostatic adenocarcinomas are more common in Black than White men in the United States. PLoS One 16, e0241934 (2021).
22. Trock BJ et al. Cell Cycle Progression Score, but Not Phosphatase and Tensin Homolog Loss, Is an Independent Prognostic Factor for Metastasis in Intermediate- and High-risk Prostate Cancer in Men Treated With and Without Salvage Radiotherapy. J Urol 208, 1182–1193 (2022). [PubMed: 36006048]
23. Kaur HB et al. Association of tumor-infiltrating T-cell density with molecular subtype, racial ancestry and clinical outcomes in prostate cancer. Mod Pathol 31, 1539–1552 (2018). [PubMed: 29849114]
24. Gumuskaya B. et al. Assessing the order of critical alterations in prostate cancer development and progression by IHC: further evidence that PTEN loss occurs subsequent to ERG gene fusion. Prostate Cancer Prostatic Dis 16, 209–215 (2013). [PubMed: 23545904]
25. Xie E. et al. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. arXiv:2105.15203 (2021). 10.48550/arXiv.2105.15203
26. Liu Z. et al. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 Ieee/Cvf International Conference on Computer Vision (Iccv 2021), 9992–10002 (2021).
27. Singhal N. et al. A deep learning system for prostate cancer diagnosis and grading in whole slide images of core needle biopsies. Scientific reports 12, 3383 (2022). [PubMed: 35233002]
28. Dosovitskiy A. et al. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929 (2020). 10.48550/arXiv.2010.11929
29. Caron M. et al. Emerging Properties in Self-Supervised Vision Transformers. arXiv:2104.14294 (2021). 10.48550/arXiv.2104.14294
30. Seo S, Huang J, Yang H. & Liu Y. in Proceedings of the eleventh ACM conference on recommender systems. 297–305.
31. Shao Z. et al. TransMIL: Transformer based Correlated Multiple Instance Learning for Whole Slide Image Classification. arXiv:2106.00908 (2021). 10.48550/arXiv.2106.00908
32. Bismar TA et al. PTEN genomic deletion is an early event associated with ERG gene rearrangements in prostate cancer. BJU Int 107, 477–485 (2011). [PubMed: 20590547]
33. Mehra R. et al. Heterogeneity of TMPRSS2 gene rearrangements in multifocal prostate adenocarcinoma: molecular evidence for an independent group of diseases. Cancer Res 67, 7991–7995 (2007). [PubMed: 17804708]
34. Barry M, Perner S, Demichelis F. & Rubin MA TMPRSS2-ERG fusion heterogeneity in multifocal prostate cancer: clinical and biologic implications. Urology 70, 630–633 (2007). [PubMed: 17991527]
35. Magi-Galluzzi C. et al. TMPRSS2-ERG gene fusion prevalence and class are significantly different in prostate cancer of Caucasian, African-American and Japanese patients. Prostate 71, 489–497 (2011). [PubMed: 20878952]
36. Khani F. et al. Evidence for molecular differences in prostate cancer between African American and Caucasian men. Clin Cancer Res 20, 4925–4934 (2014). [PubMed: 25056375]
37. Tosoian JJ et al. Prevalence and Prognostic Significance of PTEN Loss in African-American and European-American Men Undergoing Radical Prostatectomy. Eur Urol (2016).
38. Ahearn TU et al. A Prospective Investigation of PTEN Loss and ERG Expression in Lethal Prostate Cancer. J Natl Cancer Inst 108 (2016).
39. Cuzick J. et al. Prognostic value of PTEN loss in men with conservatively managed localised prostate cancer. Br J Cancer 108, 2582–2589 (2013). [PubMed: 23695019]
40. Pritchard CC et al. Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. N Engl J Med 375, 443–453 (2016). [PubMed: 27433846]
41. Robinson D. et al. Integrative clinical genomics of advanced prostate cancer. Cell 161, 1215–1228 (2015). [PubMed: 26000489]
42. de Bono J. et al. Olaparib for Metastatic Castration-Resistant Prostate Cancer. N Engl J Med 382, 2091–2102 (2020). [PubMed: 32343890]
43. Abida W. et al. Rucaparib in Men With Metastatic Castration-Resistant Prostate Cancer Harboring a BRCA1 or BRCA2 Gene Alteration. J Clin Oncol 38, 3763–3772 (2020). [PubMed: 32795228]

44. Mohler JL et al. Prostate Cancer, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw 17, 479–505 (2019). [PubMed: 31085757]
45. Cheng H, Powers J, Schaffer K. & Sartor O. Practical Methods for Integrating Genetic Testing Into Clinical Practice for Advanced Prostate Cancer. American Society of Clinical Oncology educational book. American Society of Clinical Oncology. Meeting 38, 372–381 (2018).
46. Brown LC & Armstrong AJ Germline Testing in Prostate Cancer: Implementation and Disparities of Care. JCO Oncol Pract, OP2200804 (2023).
47. Reid AH et al. Molecular characterisation of ERG, ETV1 and PTEN gene loci identifies patients at low and high risk of death from prostate cancer. Br J Cancer 102, 678–684 (2010). [PubMed: 20104229]
48. Lotan TL et al. PTEN Loss as Determined by Clinical-grade Immunohistochemistry Assay Is Associated with Worse Recurrence-free Survival in Prostate Cancer. Eur Urol Focus 2, 180–188 (2016). [PubMed: 27617307]
49. Yoshimoto M. et al. Absence of TMPRSS2:ERG fusions and PTEN losses in prostate cancer is associated with a favorable outcome. Mod Pathol 21, 1451–1460 (2008). [PubMed: 18500259]
50. Lotan TL et al. Cytoplasmic PTEN protein loss distinguishes intraductal carcinoma of the prostate from high-grade prostatic intraepithelial neoplasia. Mod Pathol 26, 587–603 (2013). [PubMed: 23222491]
51. Dadhania V. et al. Leveraging artificial intelligence to predict ERG gene fusion status in prostate cancer. BMC cancer 22, 494 (2022). [PubMed: 35513774]

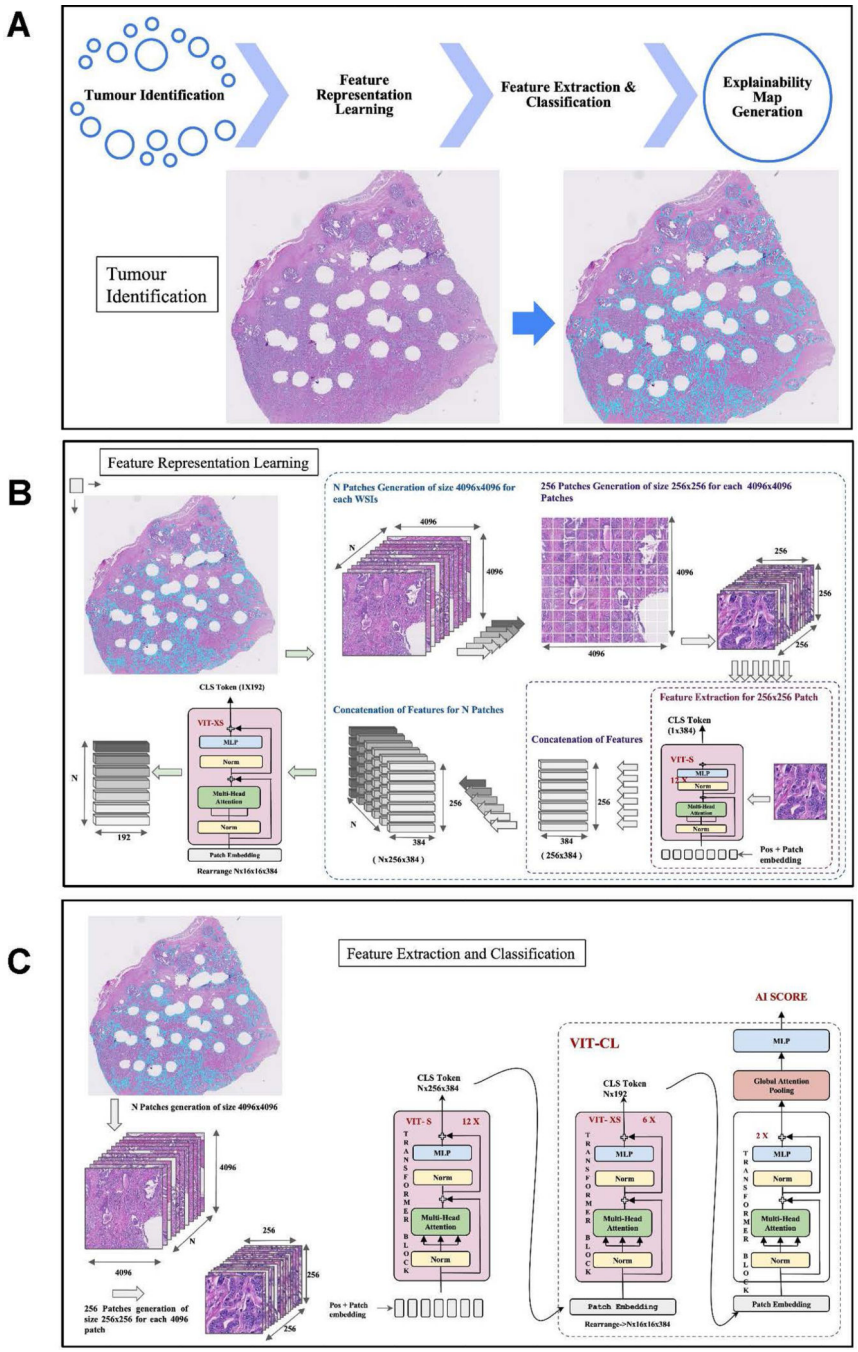

Figure 1: Overview of deep learning framework utilized for molecular classification.

(A) The deep learning methodology includes multiple stages, tumor identification, feature representation learning, feature extraction and classification as well as a final step of explainability map generation. An example of the automated tumor identification algorithm is shown, with tissue microarray (TMA) punch sites visible. **(B)** The feature representation learning comprises a novel hierarchical architecture based on two distinct vision transformer (ViT)-based networks. The first is intended to extract fine-grained characteristics by exploiting the spatial interaction of cells. The second network is trained to reflect region-

level feature extraction, such as cell cluster or gland interactions. **(C)** Feature extraction and classification uses an additional distinct ViT.

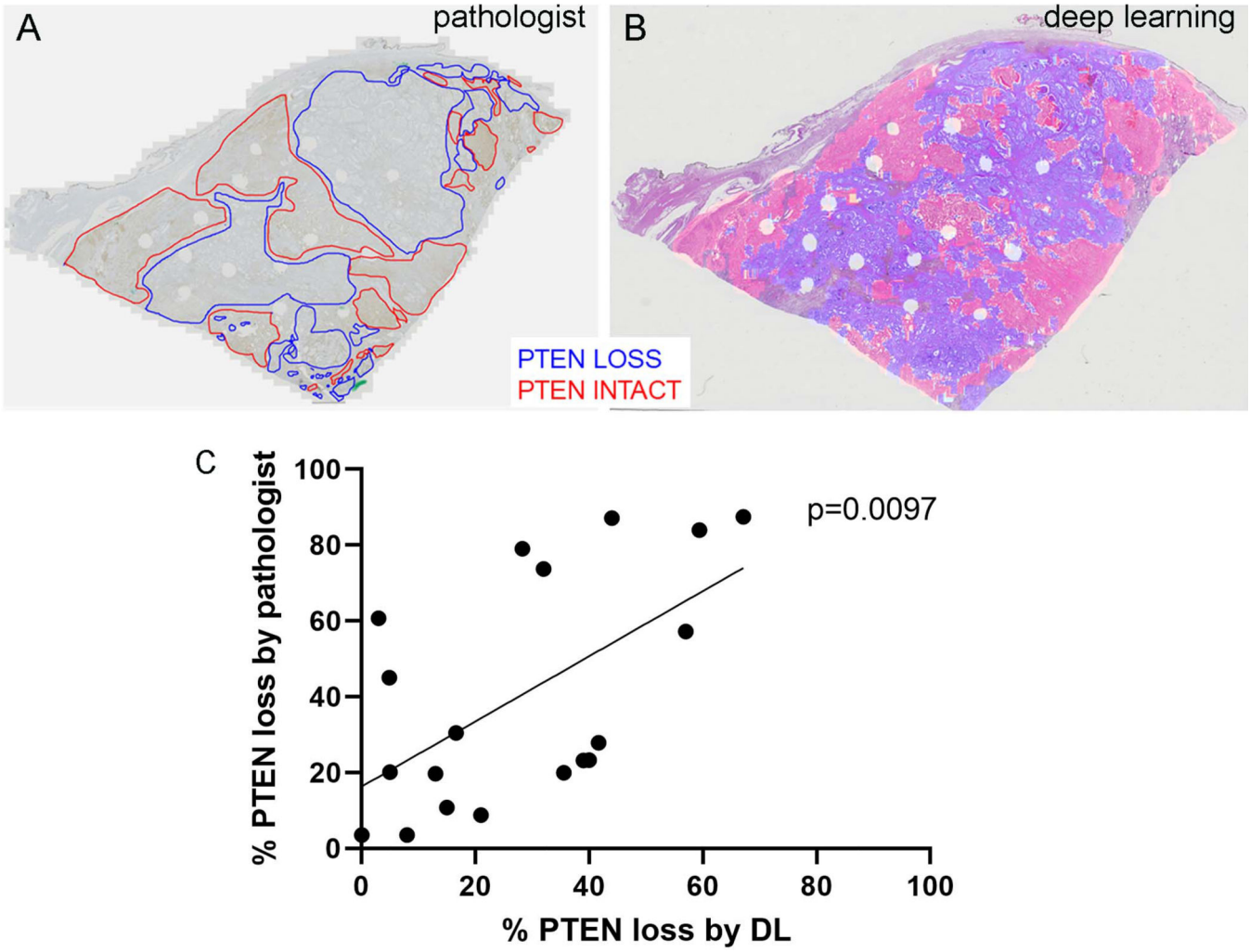

Figure 2: Correlation between pathologist annotations and deep learning algorithm predictions for radical prostatectomy tumor regions with PTEN loss.

(A) Representative PTEN immunostaining slide with pathologist annotations for regions of tumor with PTEN loss (blue) and intact PTEN (red). Punch sites for the TMA construction are visible. **(B)** Corresponding heat maps for deep learning algorithm prediction of tumor regions with PTEN loss based on H&E stained slide. **(C)** Correlation between percent of tumor area with PTEN loss based on pathologist annotation and percent of tumor with PTEN loss based on deep learning (DL) model across 19 whole slide images of radical prostatectomy samples.

Table 1:

Performance of deep learning model for ERG status prediction in validation cohorts

Cohort	# of cases (# of ERG-positive cases)	AUC
Natural History Cohort (RP)	64 (26)	0.91 ± 0.034
Case-Cohort (RP)	248 (107)	0.86 ± 0.025
Race Cohort (RP)	375 (127)	0.89 ± 0.007
Radiation Therapy Cohort (NB)	179 (70)	0.78 ± 0.017
Active Surveillance Cohort (NB)	148 (59)	0.80±0.029

RP: Radical prostatectomy; NB: Needle biopsy

Table 2:

Performance of deep learning model for PTEN status prediction in validation cohorts

Cohort	# of cases (# of cases with homogeneous PTEN loss)	AUC
Natural History Cohort (RP)	50 (13)	0.81±0.043
Case-Cohort (RP)	201 (58)	0.72±0.014
Race Cohort (RP)	337 (56)	0.80±0.015
Radiation Therapy Cohort (NB)	151 (28)	0.75±0.028

RP: Radical prostatectomy; NB: Needle biopsy