Title: Prostate cancer risk stratification via non-destructive 3D pathology with deep learning-assisted gland analysis

Running title: Prostate cancer risk stratification via 3D gland analysis

Authors: Weisi Xie[1], Nicholas P. Reder[1,2], Can Koyuncu[3], Patrick Leo[3], Sarah Hawley[4], Hongyi Huang[1], Chenyi Mao[5], Nadia Postupna[2], Soyoung Kang[1], Robert Serafin[1], Gan Gao[1], Qinghua Han[6], Kevin W. Bishop[1,6], Lindsey A. Barner[1], Pingfu Fu[7], Jonathan L. Wright[8], C. Dirk Keene[2], Joshua C. Vaughan[5,9], Andrew Janowczyk[3,10], Adam K. Glaser[1], Anant Madabhushi[3,11], Lawrence D. True[2,8], and Jonathan T.C. Liu[1,2,6*]

[1] Department of Mechanical Engineering, University of Washington, Seattle, WA, USA
[2] Department of Laboratory Medicine & Pathology, University of Washington, Seattle, WA, USA
[3] Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA
[4] Canary Foundation, Palo Alto, CA, USA
[5] Department of Chemistry, University of Washington, Seattle, WA, USA
[6] Department of Bioengineering, University of Washington, Seattle, WA, USA
[7] Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA
[8] Department of Urology, University of Washington, Seattle, WA, USA
[9] Department of Physiology & Biophysics, Seattle, WA, USA
[10] Department of Oncology, Lausanne University Hospital and Lausanne University, Lausanne, Switzerland
[11] Louis Stokes Cleveland Veterans Administration Medical Center, Cleveland, OH, USA

[*] To whom correspondence should be addressed:

Dr. Jonathan T.C. Liu
Box 352600
Seattle, WA 98195
jonliu@uw.edu
206-543-5339 (Office)

Conflict of interest: N.P. Reder., A.K. Glaser, L.D. True, and J.T.C. Liu are co-founders and shareholders of Lightspeed Microscopy Inc. A. Madabhushi is an equity holder in Elucid Bioimaging and in Inspirata Inc. In addition, he has served as a scientific advisory board member for Inspirata Inc, Astrazeneca, Bristol Meyers-Squibb and Merck. Currently, A. Madabhushi serves on the advisory board of Aiforia Inc. He also has sponsored research agreements with Philips, AstraZeneca, Boehringer-Ingelheim and Bristol Meyers-Squibb. A. Madabhushi's technology has been licensed to Elucid Bioimaging. He is also involved in a NIH U24 grant with PathCore Inc, and 3 different R01 grants with Inspirata Inc.

1

2

Abstract

Prostate cancer treatment planning is largely dependent upon examination of core-needle biopsies. The microscopic architecture of the prostate glands forms the basis for prognostic grading by pathologists. Interpretation of these convoluted 3D glandular structures via visual inspection of a limited number of 2D histology sections is often unreliable, which contributes to the under- and over-treatment of patients. To improve risk assessment and treatment decisions, we have developed a workflow for non-destructive 3D pathology and computational analysis of whole prostate biopsies labeled with a rapid and inexpensive fluorescent analog of standard H&E staining. This analysis is based on interpretable glandular features and is facilitated by the development of image-translation-assisted segmentation in 3D (ITAS3D). ITAS3D is a generalizable deep-learning-based strategy that enables tissue microstructures to be volumetrically segmented in an annotation-free and objective (biomarker-based) manner without requiring immunolabeling. As a preliminary demonstration of the translational value of a computational 3D vs. a computational 2D pathology approach, we imaged 300 *ex vivo* biopsies extracted from 50 archived radical prostatectomy specimens, of which 118 biopsies contained cancer. The 3D glandular features in cancer biopsies were superior to corresponding 2D features for risk stratification of low- to intermediate-risk PCa patients based on their clinical biochemical recurrence (BCR) outcomes. The results of this study support the use of computational 3D pathology for guiding the clinical management of prostate cancer.

Significance: An end-to-end pipeline for deep learning-assisted computational 3D histology analysis of whole prostate biopsies shows that non-destructive 3D pathology has the potential to enable superior prognostic stratification of prostate cancer patients.

3

Introduction

Prostate cancer (PCa) is the most common cancer in men and the second leading cause of cancer death for men in the United States (1). Currently, PCa management is largely dependent upon examination of prostate biopsies via 2D histopathology (2), in which a set of core-needle biopsies is formalin-fixed and paraffin-embedded (FFPE) to allow thin sections to be cut, mounted on glass slides, and stained for microscopic analysis. To quantify the aggressiveness of the cancer, the Gleason grading system is used, which relies entirely upon visual interpretation of prostate gland morphology as seen on a few histology slides (thin 2D tissue sections) that only "sample" ~1% of the whole biopsy. Gleason grading of PCa is associated with high levels of interobserver variability (3,4) and is only moderately correlated with outcomes, especially for patients with intermediate-grade PCa (5). This contributes to the undertreatment of patients with aggressive cancer (e.g. with active surveillance) (6), leading to preventable metastasis and death (7), and the overtreatment of patients with indolent cancer (e.g. with surgery or radiation therapy) (8), which can lead to serious side effects, such as incontinence and impotence (9).

Motivated by recent technological advances in optical clearing to render tissue specimens transparent to light (i.e. iDISCO (10), CUBIC (11) etc.) in conjunction with high-throughput 3D light-sheet microscopy, a number of groups have been exploring the value of non-destructive 3D pathology of clinical specimens for diagnostic pathology (12-17). Compared to conventional slide-based histology, non-destructive 3D pathology can achieve vastly greater sampling of large specimens along with volumetric visualization and quantification of diagnostically significant microstructures, all while maintaining intact specimens for downstream molecular assays (18). Like others before us (12,19,20), we hypothesized that 3D vs. 2D pathology datasets could allow for improved characterization of the convoluted glandular structures that

4

pathologists currently rely on for PCa risk stratification. However, since the associated information content of a 3D pathology dataset of a biopsy is >100× larger than a 2D whole-slide image representation (in terms of total number of pixels), computational tools are necessary to analyze these large datasets efficiently and reproducibly for diagnostic and prognostic determinations.

For computational analysis of 3D pathology datasets, a multi-stage pipeline was chosen for classifying patient outcomes based on interpretable "hand-crafted" features (21-23) (i.e., glandular features) rather than an end-to-end deep-learning (DL) strategy for risk classification based directly on the imaging data (24-26). This was motivated by: (1) the attractiveness of an intuitive feature-based approach as an initial strategy to facilitate hypothesis-testing and clinical adoption of an emerging modality in which datasets are currently limited and poorly understood (18,27), and (2) the observation that when case numbers are limited, a hand-crafted feature-based approach can be more reliable than an end-to-end DL classifier (28,29).

Multi-stage feature-based classification approaches rely on the accurate segmentation of morphological structures such as nuclei (30-32), collagen fibers (33,34), vessels (14,35), or in our case, prostate glands (21,36). This is typically achieved in one of two ways: (1) Direct DL-based segmentation methods (37-40) that require manually annotated training datasets, which are especially tedious and difficult to obtain in 3D (**Fig. 1a**) (41); or (2) traditional computer vision (CV) approaches based on intensity and morphology, provided that tissue structures of interest can be stained/labeled with high specificity (**Fig. 1b**) (19,42,43). While immunolabeling can confer a high degree of specificity for traditional CV-based segmentation, it is not an attractive strategy for clinical 3D pathology assays due to the high cost of antibodies to stain

5

large tissue volumes, and the slow diffusion times of antibodies in thick tissues (up to several weeks) (10,44).

To address these challenges, we developed a generalizable annotation-free 3D segmentation method, hereafter referred to as "image-translation-assisted segmentation in 3D (ITAS3D)". In our specific implementation of ITAS3D (**Fig. 1c**), 3D H&E-analog images of prostate tissues are synthetically converted in appearance to mimic 3D immunofluorescence (IF) images of Cytokeratin 8 (CK8) – a low-molecular-weight keratin expressed by the luminal epithelial cells of all prostate glands – thereby facilitating the objective (biomarker-based) segmentation of the glandular epithelium and lumen spaces using traditional CV tools. The deep-learning image-translation algorithm is trained with a generative adversarial network (GAN), which has been previously used for 2D virtual-staining applications (45-47). However, unlike those prior 2D image-translation efforts, we developed a "2.5D" virtual-staining approach based on a specialized GAN that was originally designed to achieve video translation with high spatial continuity between frames, but which has been adapted within our ITAS3D framework to ensure high spatial continuity as a function of depth (see Results) (48).

As a preliminary study to investigate the value of a computational 3D pathology workflow vs. a computational 2D pathology workflow, 300 *ex vivo* biopsies were extracted from archived radical prostatectomy (RP) specimens obtained from 50 patients who underwent surgery over a decade ago. We stained the biopsies with an inexpensive small-molecule (i.e., rapidly diffusing) fluorescent analog of H&E, optically cleared the biopsies with a simple dehydration and solvent-immersion protocol to render them transparent to light, and then used an open-top light-sheet (OTLS) microscopy platform to obtain whole-biopsy 3D pathology datasets. The prostate glandular network was then segmented using ITAS3D, from which 3D glandular features (i.e.,

6

histomorphometric parameters) and corresponding 2D features were extracted from the 118

biopsies that contained PCa. These 3D and 2D features were evaluated for their ability to

stratify patients based on clinical biochemical recurrence (BCR) outcomes, which serve as a

proxy endpoint for aggressive vs. indolent PCa.

Materials and Methods

Key methodological information is provided below for our preliminary clinical demonstration.

The Supplementary Methods contain technical details on the training and validation of ITAS3D,

as well as computational details for the clinical study, such as a description of gland features.

Collection and processing of archived tissue to obtain simulated biopsies for the clinical

study. The following study was approved by the institutional review board (IRB) of the

University of Washington (Study00004980), where research specimens were previously

obtained from patients with informed consent. Archived FFPE prostatectomy specimens were

collected from 50 PCa patients (see **Supplementary Table 1** for clinical data) of which 46

cases were initially graded during post-RP histopathology as having Gleason scores of 3+3, 3+4

or 4+3 (Grade Group 1 – 3). All patients were followed up for at least 5 years post-RP as part of

a prior study (Canary TMA) (49). FFPE tissue blocks were identified from each case

corresponding to the 6 regions of the prostate targeted by urologists when performing standard

sextant and 12-core (2 cores per sextant region) biopsy procedures. The identified FFPE blocks

were first de-paraffinized by heating them at 75°C for 1 hour until the outer paraffin wax was

melted, and then placing them in 65°C xylene for 48 hours. Next, one simulated core-needle

biopsy (~1-mm in width) was cut from each of the 6 deparaffinized blocks (per patient case),

resulting in a total of n = 300 biopsy cores. All simulated biopsies were then fluorescently

labeled with the T&E version of our H&E-analog staining protocol (**Supplementary Fig. 1**).

7

The T&E staining protocol (H&E analog) for the clinical study. Biopsies were first washed in 100% ethanol twice for 1 h each to remove any excess xylene, then treated in 70% ethanol for 1 h to partially re-hydrate the biopsies. Each biopsy was then placed in an individual 0.5ml Eppendorf tube (Cat: 14-282-300, Fisher Scientific), stained for 48 hours in 70% ethanol at pH 4 with a 1:200 dilution of Eosin-Y (Cat: 3801615, Leica Biosystems) and a 1:500 dilution of To-PRO™-3 Iodide (Cat: T3605, Thermo-Fisher) at room temperature with gentle agitation. The biopsies were then dehydrated twice in 100% ethanol for 2 hours. Finally, the biopsies were optically cleared (n = 1.56) by placing them in ethyl cinnamate (Cat: 112372, Sigma-Aldrich) for 8 hours before imaging them with open-top light-sheet (OTLS) microscopy.

Open-top light-sheet (OTLS) microscopy and pre-processing. We utilized a previously developed OTLS microscope (15) to image tissues slices (for training data) and simulated biopsies (for the clinical study). For this study, ethyl cinnamate (n = 1.56) was used as the immersion medium, and a custom-machined HIVEX plate (n=1.55) was used as a multi-biopsy sample holder (12 biopsies per holder). Multi-channel illumination was provided by a four-channel digitally controlled laser package (Skyra, Cobolt Lasers). Tissues were imaged at near-Nyquist sampling of ~0.44 µm/pixel. The volumetric imaging time was approximately 0.5 min per mm^3 of tissue for each wavelength channel. This allowed each biopsy (~1 x 1 x 20 mm), stained with two fluorophores (T&E), to be imaged in ~20 min.

Statistical analysis of the correlation between glandular features and BCR outcomes.
Patient-level glandular features were obtained by averaging the biopsy-level features from all cancer-containing biopsies from a single patient. Patients who experienced BCR within 5 years post-RP are denoted as the "BCR" group, and all other patients are denoted as "non-BCR". BCR was defined here as a rise in serum levels of prostate specific antigen (PSA) to 0.2 ng/ml

8

after 8 weeks post-RP (49). The box plots indicate median values along with interquartile ranges (25% - 75% of the distribution). The whiskers extend to the furthest data points excluding outliers defined as points beyond 1.5× the interquartile range. The *p* values for the BCR group vs. non-BCR group are calculated using the two-sided Mann–Whitney U-test (50). To assess the ability of different 3D and 2D glandular features to distinguish between BCR vs. non-BCR groups, we applied ROC curve analysis, from which an area-under-the-curve (AUC) value could be extracted. The *t*-SNE(51) analyses were performed with 1000 iterations at a learning rate of 100.

To develop multiparameter classifiers to stratify patients based on 5-year BCR outcomes, a least absolute shrinkage and selection operator (LASSO) logistic regression model was developed (22) using the binary 5-year BCR category as the outcome endpoint. LASSO is a regression model that includes a L1 regularization term to avoid overfitting and to identify a subset of features that are most predictive. Here, the optimal LASSO tuning parameter, λ, was determined with 3-fold cross validation (CV), where the dataset was randomly partitioned into three equal-sized groups: two groups to train the model with a specific λ, and one group to test the performance of the model. Along the LASSO regularization path, the λ with the highest R^2 (coefficient of determination) was defined as the optimal λ. Due to the lack of an external validation set, a nested CV schema was used to evaluate the performance of the multivariable models without any bias and data leakage between parameter estimation and validation steps (52). The aforementioned CV used for hyperparameter tuning was performed in each iteration of the outer CV. LASSO regression was applied on the training set of the outer CV once an optimal λ was identified in the inner CV. AUC values were then calculated from the testing group of the outer CV (**Supplementary Fig. 2**). This nested CV was performed 200 times in

9

order to determine an AUC (average and standard deviation). The exact same pipeline was

used to develop multiparameter classifiers based on 3D and 2D features.

Kaplan Meier (KM) analysis was carried out to compare BCR-free survival rates for high-risk vs.

low-risk groups of patients. This analysis utilized a subset of 34 cases for which time-to-

recurrence data is available (see **Supplementary Table 1**). The performance of the models,

either based on 2D or 3D features, was quantified with *p* values (by log-rank test), hazard ratios

(HR) and concordance index (C-index) metrics. For the multiparameter classification model

used for KM analysis, the outer CV (3-fold) in our nested CV schema was replaced by a leave-

one-out approach, where one case was left out of each iteration (50 total iterations) to calculate

the probability of 5-year BCR for that patient (53). The samples were categorized as low- or

high-risk by setting a posterior class probability threshold of 0.5. MATLAB was used for the KM

analysis and all other statistical analysis was performed in Python with the "Scipy" and "Scikit-

learn" packages.

Results

Annotation-free 3D gland segmentation. To segment the 3D glandular network within

prostate biopsies, we first trained a GAN-based image-sequence translation model (see

Supplementary Methods) to convert 3D H&E-analog images into synthetic CK8 IF images,

which can be false colored to resemble chromogenic immunohistochemistry (IHC) (**Fig. 2a**). As

mentioned, the CK8 biomarker is expressed by the luminal epithelial cells of all prostate glands.

The image-translation model is trained in a supervised manner with images from prostate

tissues that are fluorescently tri-labeled with our H&E analog and a CK8-targeted monoclonal

antibody (mAb) **(Supplementary Table 2)**.

10

As shown in step 1 of **Fig. 2b**, for whole-biopsy H&E-to-CK8 conversion, we first sub-divide the 3D biopsy (~ 1 mm × 0.7 mm × 20 mm) datasets into ~1 mm × 0.7 mm × 1 mm (~ 1024 x 700 x 1024 pixel) blocks. Each 3D image block is treated as a 2D image sequence as a function of depth. At each depth level, a synthetic CK8 image is inferred from the H&E-analog image at that level while simultaneously utilizing the images (H&E analog and CK8) from two previous levels to enforce spatial continuity as a function of depth. This "2.5D" image translation method is based on a previously reported "*vid2vid*" method for video translation (time sequences rather than depth sequences) (48) (see **Supplementary Fig. 3** and Supplementary Methods). However, our modified model omits the "coarse-to-fine" training strategy implemented in the original *vid2vid* method because this enables training times to be minimized with negligible performance loss (see **Supplementary Note 1**, **Supplementary Fig. 4** and **Supplementary Video 1**). Once the synthetic CK8 image blocks are generated, they are mosaicked to generate a whole-biopsy CK8 IHC dataset for gland segmentation. In step 2 of **Fig. 2b**, the synthetic CK8 dataset is used to segment the luminal epithelial cell layer ("Epithelium" in **Fig. 2b**) via a thresholding algorithm. The gland-lumen space, which is enclosed by the epithelium layer, can then be segmented by utilizing both the epithelium segmentation mask and the cytoplasmic channel (eosin-analog images). Algorithmic details are provided in the Supplementary Methods and **Supplementary Fig. 5**.

Evaluation of image translation and segmentation. Example 3D prostate gland-segmentation results are shown for benign and cancerous regions in **Fig. 3a**. While the glands can be delineated on the H&E-analog images by a trained observer, automated computational segmentation of the glands remains challenging (54,55). Here we demonstrate that 3D image translation based on H&E-analog inputs results in synthetic-CK8 outputs in which the luminal epithelial cells are labeled with high contrast and spatial precision. We further show that these

11

synthetic-CK8 datasets allow for relatively straightforward segmentation of the gland epithelium, lumen, and surrounding stromal tissue compartments (**Fig. 3a**). Glands from various PCa subtypes are successfully segmented as shown in **Fig. 3a**, including two glandular patterns that are typically associated with low and intermediate risk, respectively: small discrete well-formed glands (Gleason 3) and cribriform glands consisting of epithelial cells interrupted by multiple punched-out lumina (Gleason 4). **Supplementary Video 2** shows depth sequences of an H&E-analog dataset, a synthetic-CK8 dataset, and a segmentation mask of the two volumetric regions shown in **Fig. 3a**. A whole-biopsy 3D segmentation is also depicted in **Supplementary Video 3**.

To demonstrate improved depth-wise continuity with our 2.5D image-translation strategy versus a similar 2D image-translation method (based on the "*pix2pix*" GAN), vertical cross-sectional views of a synthetic-CK8 dataset are shown in **Fig. 3b**. While obvious distortions and discontinuities are seen as a function of depth with 2D image translation, the results of our 2.5D image-sequence translation exhibit optimal continuity with depth. To further illustrate this improved continuity with depth, **Supplementary Video 4** shows a depth sequence of *en face* images (z stack). Abrupt morphological discontinuities between levels are again obvious with 2D translation but absent with the 2.5D translation approach. To quantify the performance of our image-translation method, a 3D structural similarity (SSIM) metric was calculated in which real CK8 IF datasets were used as ground truth. For images generated with 2.5D vs. 2D image translation, the 3D SSIM (averaged over 58 test volumes that were 0.2-mm^3 each) was 0.41 vs. 0.37, reflecting a 12% improvement at a *p* value of 7.8×10^{-6} (two-sided paired t-test). This enhanced image-translation performance facilitates accurate 3D gland segmentations in subsequent steps of our computational pipeline.

12

To assess segmentation performance, ground-truth gland-segmentation datasets were first

generated under the guidance of board-certified genitourinary pathologists (L.D.T. and N.P.R.).

A total of 10 tissue volumes from different patients (512 × 512 × 100 pixels each, representing

0.2-mm^3 of tissue) were manually annotated. We then compared the accuracy of ITAS3D with

that of two common methods: 3D watershed (56) (as a 3D non-DL benchmark) and 2D U-Net

(57) (as a 2D DL benchmark). ITAS3D outperforms the two benchmark methods in terms of

Dice coefficient (58) and 3D Hausdorff distance (59) (**Fig. 3c**). As a visual comparison between

ITAS3D and the two benchmark methods, **Supplementary Video 5** displays image-stack

sequences from three orthogonal perspectives of a representative segmented dataset, where

the higher segmentation accuracy of ITAS3D can be appreciated. Note that a 3D DL-based

benchmark method is not provided since there are currently insufficient 3D-annotated prostate

gland datasets to train an end-to-end 3D DL segmentation model (40,60,61); again, this is one

of the main motivations for developing the annotation-free ITAS3D method.

Clinical validation study: glandular feature extraction and correlation with BCR outcomes.

Due to the slow rate of progression for most PCa cases, an initial clinical study to assess the

prognostic value of 3D vs. 2D glandular features was performed with archived prostatectomy

specimens. Our study consisted of N = 50 PCa patients who were followed up for a minimum of

5 years post-RP as part of the Canary TMA case-cohort study (primarily low- to intermediate-

risk patients) (49). The Canary TMA study was based on a well-curated cohort of PCa patients

in which the primary study endpoints were 5-year BCR outcomes and time to recurrence, which

are also used as endpoints for our validation study. In the original Canary TMA study,

approximately half of the patients experienced BCR within 5 years of RP, making it an ideal

cohort for our study. We randomly selected a subset of 25 cases that had BCR within 5 years of

RP ("BCR" group), and 25 cases that did not have BCR within 5 years of RP ("non-BCR" group).

13

FFPE tissue blocks were identified from each case corresponding to the 6 regions of the

prostate targeted by urologists when performing standard sextant and 12-core (2 cores per

sextant region) biopsy procedures (**Fig. 4a)**. Next, a simulated core-needle biopsy was

extracted from each of the 6 FFPE tissue blocks for each patient (n = 300 total biopsy cores).

The biopsies were deparaffinized, labeled with a fluorescent analog of H&E, optically cleared,

and imaged nondestructively with a recently developed OTLS microscope (15) (see Methods).

Note that we have previously shown that there is no difference in the quality/appearance of our

H&E-analog 3D-pathology datasets whether the tissue is formalin-fixed only vs. FFPE (where

de-paraffinization is necessary) (15-17). Review of the 3D pathology datasets by pathologists

(L.D.T. and N.P.R.) revealed that 118 out of the 300 biopsy cores contained cancer (1 – 5

biopsies per case). The ITAS3D pipeline was applied to all cancer-containing biopsies (see

Supplementary Methods and **Supplementary Fig. 6** for cancer-region annotation). We then

calculated histomorphometric features from the 3D gland segmentations, and from individual 2D

levels from the center region of the biopsy cores, which were then analyzed in terms of their

association with BCR outcomes. For 2D analysis, average values from a total of 3 levels were

calculated, in which the three levels were separated by 20 microns (mimicking clinical practice

at many institutions) as shown in **Fig. 4a**.

We compared multiple 3D and 2D glandular histomorphometric features (see **Supplementary**

Table 3 for a detailed list). For example, the curvature of the boundary between the lumen and

epithelium is a feature that increases as glands become smaller or more irregular, as is often

seen with aggressive PCa (21). This can be quantified in the form of the average surface

curvature of the object in 3D, or the curvature of the object's cross-sectional circumference in

2D (**Fig. 4b**). As another example (**Fig. 4c**), the gland-to-convex-hull ratio (G/H) is defined as

the volume ratio (in 3D) or the area ratio (in 2D) of the gland mask (epithelium + lumen) divided

by the convex hull that circumscribes the gland. This G/H feature is inversely related to the

14

irregularity or "waviness" of the periphery of the gland (at the scale of the gland itself rather than fine surface texture), which is generally expected to increase with aggressive PCa (21). For various 3D and 2D features (**Fig. 4d - 4e** and **Supplementary Table 3**), receiver operating characteristic (ROC) curves were generated to quantify the ability of the features to stratify patients based on 5-year BCR outcomes. When comparing analogous 3D and 2D glandular features, the 3D features largely exhibit an improved correlation with 5-year BCR outcomes in comparison to their 2D counterparts. This is exemplified by the significant *p* values for the 3D features showcased in **Fig. 4b – 4c** (between BCR and non-BCR groups) and higher area-under-the-ROC-curve (AUC) values (**Figs. 4d – 4e**).

We also extracted the 3D skeleton of the lumen network and quantified its branching parameters (skeleton-derived features). Here, a "gland skeleton" is defined as a collection of lines that approximate the center axes of various glands as they propagate in 3D space (similar to a line representation of a vessel network). Example skeleton networks for benign and cancerous glands are shown in **Supplementary Video 6**. Due to the complex 3D branching-tree architecture of the gland-lumen network, there are no straightforward 2D analogs for these skeleton-derived features. In **Fig. 4f,** we show two examples of skeleton-derived features: the average branch length and the variance of the branch lengths. Both features are correlated with BCR outcomes based on *p* values and AUC values (**Fig. 4f – 4g**). Our analysis reveals that aggressive cancers (BCR cases) have shorter branch lengths and a smaller variance in branch lengths, which agrees with prior observations from 2D histology that glandular structures in higher-grade PCa are smaller and more abundant (i.e., less differentiated and varied in size). A histogram of branch lengths (**Supplementary Fig. 7)** demonstrates that the vast majority of branches are < 200-μm long, which suggests that the diameter of standard prostate biopsies (~1-mm) is sufficient for whole-biopsy 3D pathology to quantify PCa branch lengths with reasonable accuracy.

15

To explore the prognostic value of combining multiple glandular features, we used logistic

regression models for feature selection and classification based on 3D vs. 2D features (see

Methods). Brief descriptions and AUC values for the 3D and 2D glandular features involved in

training the multi-parameter models are shown in **Supplementary Table 3**. The ROC curve of a

model that combined 12 non-skeleton 3D features ("3D non-skeleton model") yielded an AUC

value of 0.80 ± 0.05 (average $\pm$ standard deviation; **Fig. 4h**), which is considerably higher than

the AUC value (0.65 ± 0.06) of the model trained with 12 analogous 2D features ("2D model").

By adding 5 skeleton-derived features to the 12 non-skeleton 3D features, a re-trained 3D

multiparameter model ("3D model") yielded a slightly higher AUC value of 0.81 ± 0.05. The

distribution of the 50 cases, based on their glandular features, can be visualized using *t*-

distributed stochastic neighbor embedding (*t*-SNE), where a clearer separation between BCR

and non-BCR cases is evident based on 3D vs. 2D glandular features (**Supplementary Fig. 8**).

Multiparameter classification models based on 3D features alone (non-skeleton) or 2D features

alone were used to divide patients into high- and low-risk groups based on 5-year BCR

outcomes (**Supplementary Table 1**), from which Kaplan-Meier (KM) curves of BCR-free

survival were constructed for a subset of cases in which time-to-recurrence (BCR) data are

available (**Fig. 4i**). Compared to the 2D model, the 3D model is associated with a higher hazard

ratio (HR) and C-index, along with a significant *p* value ($p < 0.05$), suggesting superior

prognostic stratification.

Discussion

As high-resolution biomedical imaging technologies continue to evolve and generate

increasingly larger datasets, computational techniques are needed to derive clinically actionable

information, ideally through explainable approaches that generate new insights and hypotheses.

Interpretable feature-based analysis strategies in digital pathology generally hinge upon

16

obtaining high-quality segmentations of key structural primitives (18,29) (e.g., nuclei, glands, cells, collagen). However, a common bottleneck to achieving accurate segmentations is the need for large amounts of manually annotated datasets (41). In addition to being tedious and difficult to obtain (especially in 3D), such annotations are often performed by one or more individuals who are not representative of all pathologists, thereby introducing an early source of bias. The use of simulated data has been explored to alleviate the need for manual annotations, and has been reported to be effective for training DL-based segmentation models for highly conserved and predictable morphologies (e.g. ellipsoidal nuclei (62,63), or tubular vessel networks (63,64)). However, the 3D glandular networks of prostate tissues are highly irregular and variable, making it challenging to computationally generate simulated datasets. This complex and varied 3D morphology is also in part why 2D Gleason patterns may not be ideal for characterizing prostate glands.

The ITAS3D pipeline is a general approach for the volumetric segmentation of tissue structures (e.g., vasculature / endothelial cells, neurons, collagen fibers, lymphocytes) that can be immunolabeled with high specificity and that are also discernable to a deep-learning model when labeled with small-molecule stains like our H&E analog or similar covalent stains (65). ITAS3D obviates the need for tedious and subjective manual annotations and, once trained, eliminates the requirement for slow/expensive antibody labeling of thick tissues (**Fig. 1**). The 2.5D segmentation approach employed in ITAS3D (i.e., image-sequence translation) offers an attractive compromise between computational speed/simplicity and accuracy for 3D objects that are relatively continuous in space (e.g., prostate glands). Details regarding 2.5D vs. 3D image translation are provided in **Supplementary Note 2 and Supplementary Table 4**. In addition, a video summary of our ITAS3D-enabled gland-segmentation approach for PCa assessment is provided in **Supplementary Video 7**. Note that in this specific implementation of ITAS3D, intermediate images are synthetically generated to mimic an IHC stain (CK8) that is routinely

17

used by genitourinary pathologists. Therefore, it has the added advantage of enabling intuitive troubleshooting and facilitating clinical acceptance of our computational 3D pathology approach. A limitation of our implementation of ITAS3D is that the tissues used to train our deep-learning image-sequence translation modules were predominantly from Gleason pattern 3 and 4 regions, as well as benign regions. The reason for this was that the Canary TMA case-cohort study from which we derived our tissue specimens mainly consisted of low- to intermediate-grade (i.e., pattern 3 and 4) patients. In the future, improved ITAS3D performance over a wider range of PCa grades would be facilitated by a more-diverse set of training specimens.

In this initial clinical study, we have intentionally avoided comparing our method with extant risk classifiers or nomograms that incorporate parameters based on human interpretation of 2D histopathology images (e.g., Kattan (66), CAPRA (67), and Canary-PASS (68)). Rather, our goal has been to demonstrate the basic feasibility and value of 3D pathology by providing a direct comparison of intuitive 3D vs. 2D glandular features analyzed computationally. A direct comparison of computational 3D pathology vs. computational 2D pathology allows us to avoid the challenging variability and subjectivity of human interpretation. A human-observer study would require much-larger patient cohorts and a large panel of pathologist observers to account for inevitable interobserver variabilities. Nonetheless, to encourage the clinical adoption of 3D pathology methods, such large-scale studies will be necessary in the future, including prospective randomized studies on active surveillance vs. curative therapies for low- to intermediate-risk patients (e.g., the PROTECT study (69)), as well as studies to demonstrate the ability of computational 3D pathology to predict the response of individual patients to specific treatments such as androgen deprivation and both neoadjuvant and adjuvant chemotherapy. For completeness, in **Supplementary Table 5**, we provide Gleason scores for the 118 cancerous biopsies based on single-pathologist review (N.P.R.) of whole-biopsy 3D pathology datasets with level-by-level examination of all 2D planes.

18

When trained with large numbers of images/cases and an optimal set of histomorphometric features, computational 2D pathology (based on whole slide images) has already been shown to be highly prognostic (21,70-72). The goal of our study is not to suggest otherwise, but to provide early evidence of the additional prognostic value that 3D pathology can provide. The metrics presented in this study (**Fig. 4**) are intended to be comparative in nature (between computational 3D vs. 2D pathology) rather than regarded as definitive figures from a large prospective study. Our results show clear improvements in risk stratification based on 3D glandular features, both individually and in combination (**Fig. 4b – 4i**). As mentioned, the added prognostic value of 3D pathology is due in part to the significantly increased microscopic sampling of specimens (e.g., whole biopsies vs. sparse tissue sections). In addition, there are a number of advantages of 3D pathology datasets for computational analyses: (1) more-reliable segmentation of tissue structures due to the ability to leverage out-of-plane information (e.g. through continuity constraints); (2) the ability to quantify tissue structures more accurately in 3D while avoiding 2D artifacts (73,74), and (3) the ability to extract novel prognostic features that cannot be derived from 2D tissue sections (e.g. gland-skeleton features).

As future work, ITAS3D can be used for the extraction and analysis of other 3D features (e.g., nuclear features, vascular features, and stromal features) to develop powerful classification models based on multiple morphologic primitives for a variety of tissue types. Annotation-free ITAS3D segmentation results, once available in sufficient quantities, can also be used to train end-to-end DL-based segmentation methods that bypass the image-translation step within ITAS3D. In the context of PCa, studies are underway to identify additional prognostic 3D features based on our unique 3D-pathology datasets. A tiered approach to analyzing PCa glandular features could be useful, such as first identifying broad classes of glandular morphologies (e.g., cribriform glands) and then analyzing class-specific features, as has been

19

suggested in recent studies based on 2D whole slide images of PCa (70). Future studies should also aim to combine computational 3D pathology with patient metadata, such as radiomics, genomics, and electronic health records, to develop holistic decision-support algorithms (18). Nonetheless, as an initial step towards these goals, the results of this study, as enabled by the ITAS3D computational approach, provide the strongest evidence to date in support of the value of computational 3D pathology for clinical decision support, specifically for low- to intermediate-risk PCa patients.

Data availability

Relevant clinical data for this study are provided in **Supplementary Table 1** and **Supplementary Table 5**. Example prostate images for testing ITAS3D codes and models are available in a GitHub repository at https://github.com/WeisiX/ITAS3D. Full 3D prostate imaging datasets, "simulated 2D whole-slide images" extracted from those datasets (3 levels per biopsy), as well as other clinical data are available upon reasonable request and with the establishment of a material-transfer or data-transfer agreement.

Code availability

The Python code for the deep-learning models, and for 3D glandular segmentations based on synthetic-CK8 datasets, are available on GitHub at https://github.com/WeisiX/ITAS3D

Author contributions

W. Xie, N.P. Reder and J.T.C. Liu designed the ITAS3D segmentation pipeline. W. Xie performed tri-labeling, trained the GAN models and implemented the ITAS3D segmentation. W. Xie and C. Koyuncu extracted the glandular features and trained the multi-parameter model with help from P. Leo and A. Janowczyk. S. Hawley provided de-identified clinical data for our study

20

cohort. H&E-analog staining and imaging of biopsies was performed by H. Huang, with help from S. Kang. C. Mao performed thick sectioning of tissue that was used to obtain training datasets. C. Mao and N. Postupna provided guidance for the tri-labeling procedure. R. Serafin performed stitching and false-coloring of the OTLS whole-biopsy images. G. Gao and Q. Han manually annotated 3D datasets as ground-truth for validation studies under the guidance of N.P. Reder and L.D. True. A.K. Glaser designed and built the OTLS microscope. L.D. True obtained tissues for this study. L.D. True, N.P. Reder, and J.T.C. Liu designed the clinical study. J.T.C. Liu supervised and coordinated the project. W. Xie and J.T.C. Liu led the writing of the manuscript. All authors contributed to the manuscript.

21

References

1.   Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin **2021**;71:7-33

2.   Epstein JI. A new contemporary prostate cancer grading system. Ann Pathol **2015**;35:474-6

3.   Ozkan TA, Eruyar AT, Cebeci OO, Memik O, Ozcan L, Kuskonmaz I. Interobserver variability in Gleason histological grading of prostate cancer. Scand J Urol **2016**;50:420-4

4.   Shah RB, Leandro G, Romerocaces G, Bentley J, Yoon J, Mendrinos S*, et al.* Improvement of diagnostic agreement among pathologists in resolving an "atypical glands suspicious for cancer" diagnosis in prostate biopsies using a novel "Disease-Focused Diagnostic Review" quality improvement process. Human pathology **2016**;56:155-62

5.   Kane CJ, Eggener SE, Shindel AW, Andriole GL. Variability in Outcomes for Patients with Intermediate-risk Prostate Cancer (Gleason Score 7, International Society of Urological Pathology Gleason Group 2-3) and Implications for Risk Stratification: A Systematic Review. Eur Urol Focus **2017**;3:487-97

6.   Albertsen PC. Treatment of localized prostate cancer: when is active surveillance appropriate? Nature Reviews Clinical Oncology **2010**;7:394-400

7.   Haffner MC, De Marzo AM, Yegnasubramanian S, Epstein JI, Carter HB. Diagnostic challenges of clonal heterogeneity in prostate cancer. J Clin Oncol **2015**;33:e38-40

8.   Bill-Axelson A, Holmberg L, Garmo H, Rider JR, Taari K, Busch C*, et al.* Radical prostatectomy or watchful waiting in early prostate cancer. New England Journal of Medicine **2014**;370:932-42

9.   Frey AU, Sonksen J, Fode M. Neglected side effects after radical prostatectomy: a systematic review. J Sex Med **2014**;11:374-85

10.   Renier N, Wu Z, Simon DJ, Yang J, Ariel P, Tessier-Lavigne M. iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. Cell **2014**;159:896-910

22

11.   Susaki EA, Tainaka K, Perrin D, Kishino F, Tawara T, Watanabe TM, *et al.* Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis. Cell **2014**;157:726-39

12.   van Royen ME, Verhoef EI, Kweldam CF, van Cappellen WA, Kremers GJ, Houtsmuller AB, *et al.* Three-dimensional microscopic analysis of clinical prostate specimens. Histopathology **2016**;69:985-92

13.   Olson E, Levene MJ, Torres R. Multiphoton microscopy with clearing for three dimensional histology of kidney biopsies. Biomed Opt Express **2016**;7:3089-96

14.   Tanaka N, Kanatani S, Tomer R, Sahlgren C, Kronqvist P, Kaczynska D, *et al.* Whole-tissue biopsy phenotyping of three-dimensional tumours reveals patterns of cancer heterogeneity. Nat Biomed Eng **2017**;1:796-806

15.   Glaser AK, Reder NP, Chen Y, Yin CB, Wei LP, Kang S, *et al.* Multi-immersion open-top light-sheet microscope for high-throughput imaging of cleared tissues. Nat Commun **2019**;10

16.   Reder NP, Glaser AK, McCarty EF, Chen Y, True LD, Liu JTC. Open-Top Light-Sheet Microscopy Image Atlas of Prostate Core Needle Biopsies. Arch Pathol Lab Med **2019**;143:1069-75

17.   Barner LA, Glaser AK, Huang H, True LD, Liu JT. Multi-resolution open-top light-sheet microscopy to enable efficient 3D pathology workflows. Biomedical Optics Express **2020**;11:6605-19

18.   Liu JTC, Glaser AK, Bera K, True LD, Reder NP, Eliceiri KW, *et al.* Harnessing non-destructive 3D pathology. Nat Biomed Eng **2021**;5:203-18

19.   Tolkach Y, Thomann S, Kristiansen G. Three-dimensional reconstruction of prostate cancer architecture with serial immunohistochemical sections: hallmarks of tumour growth, tumour compartmentalisation, and implications for grading and heterogeneity. Histopathology **2018**;72:1051-9

20.   Humphrey PA. Complete Histologic Serial Sectioning of a Prostate-Gland with Adenocarcinoma. Am J Surg Pathol **1993**;17:468-72

21.   Lee G, Sparks R, Ali S, Shih NN, Feldman MD, Spangler E, *et al.* Co-occurring gland angularity in localized subgraphs: predicting biochemical recurrence in intermediate-risk prostate cancer patients. PLoS One **2014**;9:e97954

22.   Leo P, Janowczyk A, Elliott R, Janaki N, Bera K, Shiradkar R, *et al.* Computer extracted gland features from H&E predicts prostate cancer recurrence comparably to a genomic companion diagnostic test: a large multi-site study. npj Precision Oncology **2021**;5:35

23.   Lee G, Singanamalli A, Wang H, Feldman MD, Master SR, Shih NNC, *et al.* Supervised Multi-View Canonical Correlation Analysis (sMVCCA): Integrating Histologic and Proteomic Features for Predicting Recurrent Prostate Cancer. IEEE Transactions on Medical Imaging **2015**;34:284-97

24.   Lu MY, Chen TY, Williamson DFK, Zhao M, Shady M, Lipkova J, *et al.* AI-based pathology predicts origins for cancers of unknown primary. Nature **2021**

25.   Poplin R, Varadarajan AV, Blumer K, Liu Y, McConnell MV, Corrado GS, *et al.* Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nature Biomedical Engineering **2018**;2:158-64

26.   Hollon TC, Pandian B, Adapa AR, Urias E, Save AV, Khalsa SSS, *et al.* Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. Nature medicine **2020**;26:52-8

27.   Leo P, Elliott R, Shih NNC, Gupta S, Feldman M, Madabhushi A. Stable and discriminating features are predictive of cancer presence and Gleason grade in radical prostatectomy specimens: a multi-site study. Sci Rep **2018**;8:14918

23

28. Lin W, Hasenstab K, Moura Cunha G, Schwartzman A. Comparison of handcrafted features and convolutional neural networks for liver MR image adequacy assessment. Sci Rep **2020**;10:20336

29. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol **2019**;16:703-15

30. Chandramouli S, Leo P, Lee G, Elliott R, Davis C, Zhu G*, et al.* Computer Extracted Features from Initial H&E Tissue Biopsies Predict Disease Progression for Prostate Cancer Patients on Active Surveillance. Cancers (Basel) **2020**;12

31. Ali S, Veltri R, Epstein JI, Christudass C, Madabhushi A. Selective invocation of shape priors for deformable segmentation and morphologic classification of prostate cancer tissue microarrays. Computerized Medical Imaging and Graphics **2015**;41:3-13

32. Lefebvre AEYT, Mai D, Kessenbrock K, Lawson DA, Digman MA. Automated segmentation and tracking of mitochondria in live-cell time-lapse images. Nature Methods **2021**;18:1091-+

33. Bhargava HK, Leo P, Elliott R, Janowczyk A, Whitney J, Gupta S*, et al.* Computationally Derived Image Signature of Stromal Morphology Is Prognostic of Prostate Cancer Recurrence Following Prostatectomy in African American Patients. Clin Cancer Res **2020**;26:1915-23

34. Kiemen A, Braxton AM, Grahn MP, Han KS, Babu JM, Reichel R*, et al.* In situ characterization of the 3D microanatomy of the pancreas and pancreatic cancer at single cell resolution. bioRxiv **2020**

35. Arunkumar N, Mohammed MA, Mostafa SA, Ibrahim DA, Rodrigues JJPC, de Albuquerque VHC. Fully automatic model-based segmentation and classification approach for MRI brain tumor using artificial neural networks. Concurr Comp-Pract E **2020**;32

36. Koyuncu CF, Janowczyk A, Lu C, Leo P, Alilou M, Glaser AK*, et al.* Three-dimensional histo-morphometric features from light sheet microscopy images result in improved discrimination of benign from malignant glands in prostate cancer. 2020. International Society for Optics and Photonics. p 113200G.

37. Chen H, Dou Q, Yu L, Qin J, Heng P-A. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. NeuroImage **2018**;170:446-55

38. Bui TD, Shin J, Moon T. 3D densely convolutional networks for volumetric segmentation. arXiv preprint arXiv:170903199 **2017**

39. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. 2016. Springer. p 424-32.

40. Zhu Z, Xia Y, Shen W, Fishman E, Yuille A. A 3D coarse-to-fine framework for volumetric medical image segmentation. 2018. IEEE. p 682-90.

41. van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. Nature Medicine **2021**;27:775-84

42. Di Cataldo S, Ficarra E, Acquaviva A, Macii E. Automated segmentation of tissue images for computerized IHC analysis. Computer methods and programs in biomedicine **2010**;100:1-15

43. Migliozzi D, Nguyen HT, Gijs MAM. Combining fluorescence-based image segmentation and automated microfluidics for ultrafast cell-by-cell assessment of biomarkers for HER2-type breast carcinoma. J Biomed Opt **2018**;24:1-8

44. Lee SS, Bindokas VP, Lingen MW, Kron SJ. Nondestructive, multiplex three-dimensional mapping of immune infiltrates in core needle biopsy. Lab Invest **2019**;99:1400-13

45. Bayramoglu N, Kaakinen M, Eklund L, Heikkilä J. Towards Virtual H&E Staining of Hyperspectral Lung Histology Images Using Conditional Generative Adversarial Networks. 2017 22-29 Oct. 2017. p 64-71.

24

46.   Burlingame EA, McDonnell M, Schau GF, Thibault G, Lanciault C, Morgan T*, et al.*
       SHIFT: speedy histological-to-immunofluorescent translation of a tumor signature
       enabled by deep learning. Sci Rep-Uk **2020**;10
47.   Rivenson Y, Wang H, Wei Z, de Haan K, Zhang Y, Wu Y*, et al.* Virtual histological
       staining of unlabelled tissue-autofluorescence images via deep learning. Nature
       biomedical engineering **2019**;3:466-77
48.   Wang T-C, Liu M-Y, Zhu J-Y, Liu G, Tao A, Kautz J*, et al.* Video-to-video synthesis.
       arXiv preprint arXiv:180806601 **2018**
49.   Hawley S, Fazli L, McKenney JK, Simko J, Troyer D, Nicolas M*, et al.* A model for the
       design and construction of a resource for the validation of prognostic prostate cancer
       biomarkers: the Canary Prostate Cancer Tissue Microarray. Advances in anatomic
       pathology **2013**;20:39
50.   Mann HB, Whitney DR. On a test of whether one of two random variables is
       stochastically larger than the other. The annals of mathematical statistics **1947**:50-60
51.   Van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of machine learning
       research **2008**;9
52.   Dobbin KK, Cesano A, Alvarez J, Hawtin R, Janetzki S, Kirsch I*, et al.* Validation of
       biomarkers to predict response to immunotherapy in cancer: Volume II - clinical
       validation and regulatory considerations. J Immunother Cancer **2016**;4
53.   Stephenson AJ, Scardino PT, Eastham JA, Bianco FJ, Dotan ZA, DiBlasio CJ*, et al.*
       Postoperative nomogram predicting the 10-year probability of prostate cancer
       recurrence after radical prostatectomy. Journal of Clinical Oncology **2005**;23:7005-12
54.   Singh M, Kalaw EM, Giron DM, Chong KT, Tan CL, Lee HK. Gland segmentation in
       prostate histopathological images. J Med Imaging (Bellingham) **2017**;4:027501
55.   Bulten W, Bándi P, Hoven J, van de Loo R, Lotz J, Weiss N*, et al.* Epithelium
       segmentation using deep learning in H&E-stained prostate specimens with
       immunohistochemistry as reference standard. Sci Rep-Uk **2019**;9:1-10
56.   Soille PJ, Ansoult MM. Automated basin delineation from digital elevation models using
       mathematical morphology. Signal Processing **1990**;20:171-82
57.   Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image
       segmentation. 2015. Springer. p 234-41.
58.   Dice LR. Measures of the amount of ecologic association between species. Ecology
       1945;26:297-302
59.   Cignoni P, Rocchini C, Scopigno R. Metro: measuring error on simplified surfaces. 1998.
       Wiley Online Library. p 167-74.
60.   Cirillo MD, Abramian D, Eklund A. Vox2Vox: 3D-GAN for brain tumour segmentation.
       arXiv preprint arXiv:200313653 **2020**
61.   Haft-Javaherian M, Fang LJ, Muse V, Schaffer CB, Nishimura N, Sabuncu MR. Deep
       convolutional neural networks for segmenting 3D in vivo multiphoton images of
       vasculature in Alzheimer disease mouse models. Plos One **2019**;14
62.   Dunn KW, Fu CC, Ho DJ, Lee S, Han S, Salama P*, et al.* DeepSynth: Three-dimensional
       nuclear segmentation of biological images using neural networks trained with synthetic
       data. Sci Rep-Uk **2019**;9
63.   Liu Q, Gaeta IM, Zhao M, Deng R, Jha A, Millis BA*, et al.* Towards Annotation-free
       Instance Segmentation and Tracking with Adversarial Simulations. arXiv preprint
       arXiv:210100567 **2021**
64.   Tetteh G, Efremov V, Forkert ND, Schneider M, Kirschke J, Weber B*, et al.*
       DeepVesselNet: Vessel Segmentation, Centerline Prediction, and Bifurcation Detection
       in 3-D Angiographic Volumes. Front Neurosci-Switz **2020**;14

25

65.   Mao C, Lee MY, Jhan J-R, Halpern AR, Woodworth MA, Glaser AK*, et al.* Feature-rich covalent stains for super-resolution and cleared tissue fluorescence microscopy. Science advances **2020**;6:eaba4542

66.   Kattan MW, Wheeler TM, Scardino PT. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. J Clin Oncol **1999**;17:1499-507

67.   Cooperberg MR, Freedland SJ, Pasta DJ, Elkin EP, Presti JC, Jr., Amling CL*, et al.* Multiinstitutional validation of the UCSF cancer of the prostate risk assessment for prediction of recurrence after radical prostatectomy. Cancer **2006**;107:2384-91

68.   Newcomb LF, Thompson IM, Boyer HD, Brooks JD, Carroll PR, Cooperberg MR*, et al.* Outcomes of active surveillance for clinically localized prostate cancer in the prospective, multi-institutional Canary PASS cohort. The Journal of urology **2016**;195:313-20

69.   Hamdy FC, Donovan JL, Lane JA, Mason M, Metcalfe C, Holding P*, et al.* 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. N Engl J Med **2016**;375:1415-24

70.   Leo P, Chandramouli S, Farré X, Elliott R, Janowczyk A, Bera K*, et al.* Computationally Derived Cribriform Area Index from Prostate Cancer Hematoxylin and Eosin Images Is Associated with Biochemical Recurrence Following Radical Prostatectomy and Is Most Prognostic in Gleason Grade Group 2. European Urology Focus **2021**

71.   Wulczyn E, Nagpal K, Symonds M, Moran M, Plass M, Reihs R*, et al.* Predicting prostate cancer specific-mortality with artificial intelligence-based Gleason grading. Communications Medicine **2021**;1:1-8

72.   Nagpal K, Foote D, Liu Y, Chen PHC, Wulczyn E, Tan F*, et al.* Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer (vol 2, 48, 2019). Npj Digit Med **2019**;2

73.   Glaser AK, Reder NP, Chen Y, McCarty EF, Yin CB, Wei LP*, et al.* Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens. Nature Biomedical Engineering **2017**;1

74.   Epstein JI, Allsbrook Jr WC, Amin MB, Egevad LL, Committee IG. The 2005 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma. The American journal of surgical pathology **2005**;29:1228-42

26

Figure legends

Fig. 1 | General methods for 3D gland segmentation. a, A single-step deep learning (DL) segmentation model can be trained with imaging datasets of tissues labeled with a fluorescent analog of H&E paired with manually annotated ground-truth segmentation masks. While H&E-analog staining is low-cost and rapid, manual annotations are labor-intensive (especially in 3D) and based on subjective human judgements. **b,** By immunolabeling a tissue microstructure with high specificity, 3D segmentations can be achieved with traditional computer-vision (CV) methods without the need for manual annotations. While this is an objective segmentation method based on a chemical biomarker, immunolabeling large intact specimens is expensive and time-consuming due to the slow diffusion of antibodies in thick tissues. **c,** With image-translation-assisted segmentation in 3D (ITAS3D), H&E-analog datasets are computationally transformed in appearance to mimic immunofluorescence datasets, which enables the synthetically labeled tissue structures to be segmented with traditional CV methods. The image-sequence translation model is trained with a generative adversarial network (GAN) based on paired H&E-analog and immunofluorescence datasets. ITAS3D is rapid and low-cost (in terms of staining) as well as annotation-free and objective (i.e., biomarker-based).

Fig. 2 | Image-translation-assisted segmentation in 3D (ITAS3D): a two-step pipeline for annotation-free 3D segmentation of prostate glands. a, In step 1, a 3D microscopy dataset of a specimen, stained with a rapid and inexpensive fluorescent analog of H&E, is converted into a synthetic CK8 immunofluorescence (IF) dataset by using an image-sequence translation model that is trained with paired H&E-analog and real-CK8 IF datasets (tri-labeled tissues). The CK8 biomarker, which is utilized in standard-of-care genitourinary pathology practice, is ubiquitously expressed by the luminal epithelial cells of all prostate glands. In step 2, traditional computer-vision algorithms are applied to the synthetic-CK8 datasets for semantic segmentation of the gland epithelium, lumen, and surrounding stromal regions. **b**, In step 1, a 3D prostate biopsy is sub-divided into overlapping blocks that are each regarded as depth-wise sequences of 2D images. A GAN-trained generator performs image translation sequentially on each 2D level of an image block. The image translation at each level is based on the H&E-analog input at that level while leveraging the H&E-analog and CK8 images from two previous levels to enforce spatial continuity between levels (i.e., a "2.5D" translation method). The synthetic-CK8 image-block outputs are then mosaicked to generate a 3D CK8 dataset of the whole biopsy to assist with gland segmentation. In step 2, the epithelial cell layer (epithelium) is segmented from the synthetic-CK8 dataset with a thresholding-based algorithm. Gland lumen spaces are segmented by filling in the regions enclosed by the epithelia with refinements based on the cytoplasm channel (eosin fluorescence). See Supplementary Methods for details.
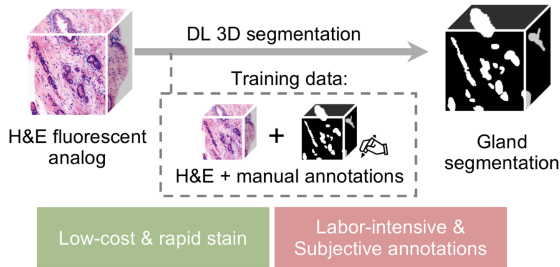
Fig. 3 | Segmentation results with ITAS3D. a, 2D cross-sections are shown (from left to right) of false-colored H&E-analog images, synthetic-CK8 IHC images generated by image-sequence translation, and

27

gland-segmentation masks based on the synthetic-CK8 images (yellow for epithelium, red for lumen, and gray for stroma). The example images are from large 3D datasets containing benign glands (first row) and cancerous glands (second row). Zoom-in views show small discrete well-formed glands (Gleason pattern 3, blue box) and cribriform glands (Gleason pattern 4, red box) in the cancerous region. 3D renderings of gland segmentations, for a benign and cancerous region, are shown on the far right. Scale bar: 100 μm. **b**, Side views of the image sequences (with the depth direction oriented down) of real- and synthetic-CK8 immunofluorescence (IF) images. The 2.5D image-translation results exhibit substantially improved depth-wise continuity compared to the 2D image-translation results. Scale bar: 25 μm. **c,** For quantitative benchmarking, Dice coefficients (larger is better) and 3D Hausdorff distances (smaller is better) are plotted for ITAS3D-based gland segmentations along with two benchmark methods (3D watershed and 2D U-net), as calculated from 10 randomly selected test regions. Violin plots are shown with mean values denoted by a center cross and standard deviations denoted by error bars. For the 3D Hausdorff distance, the vertical axis denotes physical distance (in microns) within the tissue.
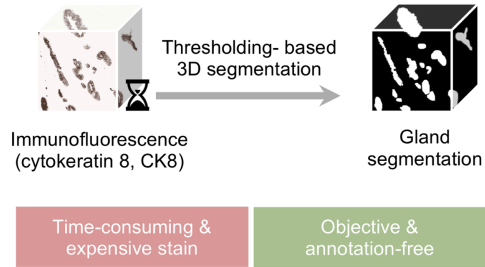
Fig. 4 | Clinical study comparing the performance of 3D vs. 2D glandular features for risk stratification. a, Archived (FFPE) radical prostatectomy (RP) specimens were obtained from a well-curated cohort of 50 patients, from which 300 simulated (*ex vivo*) needle biopsies were extracted (6 biopsies per case, per sextant-biopsy protocol). The biopsies were labeled with a fluorescent analog of H&E staining, optically cleared to render the tissues transparent to light, and then comprehensively imaged in 3D with open-top light-sheet (OTLS) microscopy. Prostate glands were computationally segmented from the resultant 3D biopsy images using the ITAS3D pipeline. 3D glandular features were extracted from tissue volumes containing PCa. 2D glandular features were extracted from 3 levels per volume and averaged. **b,c**, Violin and box plots are shown for two examples of 3D glandular features, along with analogous 2D features, for cases in which biochemical recurrence (BCR) was observed within 5 years of RP ("BCR") and for cases with no BCR within 5 years of RP ("non-BCR"). For both sets of example features, "lumen boundary curvature" in panel **b** and "gland-to-convex hull ratio" (G/H) in panel **c**, the 3D version of the feature shows improved stratification between BCR and non-BCR groups. **d,e**, Receiver-operating-characteristic (ROC) curves also show improved risk stratification with the 3D features vs. corresponding 2D features, with considerably higher area-under-the-curve (AUC) values. **f**, Violin and box plots are shown of representative gland-skeleton features (average branch length and branch length variance), which can only be accurately derived from the 3D pathology datasets, showing significant stratification between BCR and non-BCR groups. **g**, ROC curves are shown, along with AUC values, for average branch length and branch length variance. **h,** ROC curves are shown of various multiparameter models, including those trained with 2D glandular features, 3D glandular features excluding skeleton features, and 3D glandular features including skeleton features. **i**, Kaplan-Meier curves are shown for BCR-free survival, showing that a multiparameter model based on 3D glandular features is better able to

28

stratify patients into low-risk and high-risk groups with significantly different recurrence trajectories ($p = 6.6 \times 10^{-5}$, HR = 11.2, C-index = 0.84).
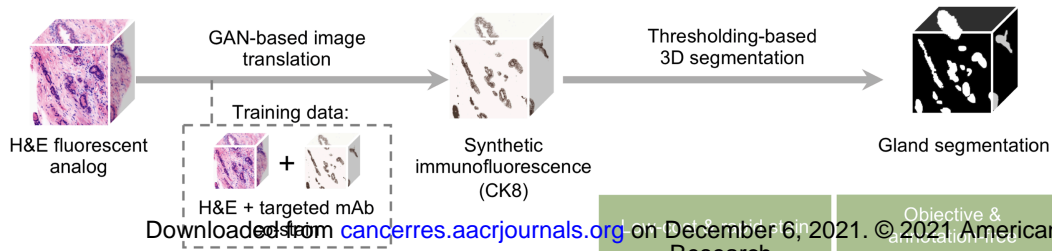
29

a

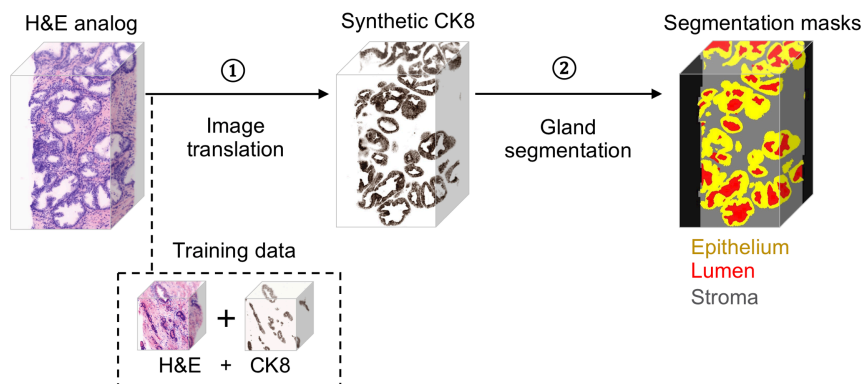**Method 1: Single-step deep-learning (DL) based segmentation**

H&E fluorescent analog → DL 3D segmentation → Gland segmentation

Training data:
H&E + manual annotations

| Low-cost & rapid stain | Labor-intensive & Subjective annotations |

b

Method 2: Immunolabeling-based 3D segmentation

Immunofluorescence (cytokeratin 8, CK8) → Thresholding- based 3D segmentation → Gland segmentation

| Time-consuming & expensive stain | Objective & annotation-free |

c

Annotation-free image-translation-assisted segmentation in 3D (ITAS3D)

H&E fluorescent analog → GAN-based image translation → Synthetic immunofluorescence (CK8) → Thresholding-based 3D segmentation → Gland segmentation

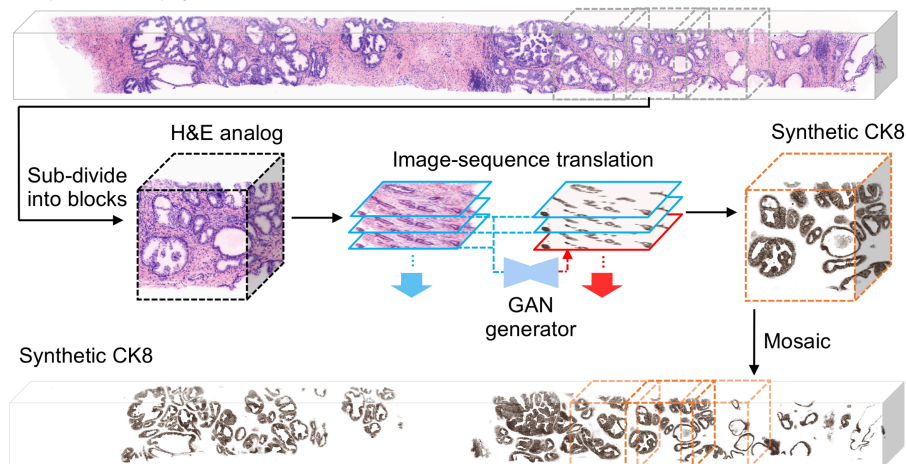Training data:
H&E + targeted mAb

| | Objective & annotation-free |

a

Image-translation-assisted segmentation in 3D (ITAS3D)

H&E analog → ① Image translation → Synthetic CK8 → ② Gland segmentation → Segmentation masks

Training data

H&E + CK8

Epithelium
Lumen
Stroma

b

① Image translation

3D prostate biopsy

Sub-divide into blocks → H&E analog → Image-sequence translation → GAN generator → Synthetic CK8 → Mosaic

Synthetic CK8

② Gland segmentation

Synthetic CK8 → Thresholding → Epithelium

Cytoplasm → Lumen filling → Lumen

3D segmentation mask

a

H&E analog (false-colored) | Synthetic CK8 (false-colored) | Epithelium and lumen segmentation | Gland segmentation (3D rendering)

Benign glands

100 µm

Cancerous glands

25 µm

► Discrete well-formed glands (Gleason 3)
► Cribriform glands (Gleason 4)

b

Real CK8 — 25 µm — ↓ Depth

2D image translation

2.5D image-sequence translation

c

Dice coefficient: ITAS3D 0.890, 3D watershed 0.726, 2D U-Net 0.647

3D Hausdorff distance (µm): ITAS3D 112.7 µm, 3D watershed 144.5 µm, 2D U-Net 143.4 µm

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association for Cancer Research

Prostate cancer risk stratification via non-destructive 3D pathology with deep learning-assisted gland analysis

Weisi Xie, Nicholas P Reder, Can F Koyuncu, et al.

Cancer Res Published OnlineFirst December 1, 2021.

Updated version	Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-21-2843
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2021/12/01/0008-5472.CAN-21-2843.DC1
Author Manuscript	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://cancerres.aacrjournals.org/content/early/2021/11/30/0008-5472.CAN-21-2843. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.