3D microscopic imaging with computational image analysis to improve cancer assessment

Weisi Xie

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Jonathan T.C. Liu, Chair

Eric J. Seibel

Lawrence D. True

Program Authorized to Offer Degree:

Mechanical Engineering

University of Washington

Abstract

3D microscopic imaging with computational image analysis to improve cancer assessment

Weisi Xie

Chair of the Supervisory Committee:
Jonathan T.C. Liu
Mechanical Engineering

Non-destructive, slide-free, and comprehensive pathology of clinical specimens can improve clinical workflow efficiency and diagnostic performance. Recent advances in optical-sectioning microscopy and optical clearing techniques enable volumetric imaging of thick tissue specimens with subcellular resolution. In addition, computational analysis of 3D microscopic image datasets can objectively assist human observers with more efficient and accurate pathological analysis. In this dissertation, we explore two emerging optical-sectioning microscopy techniques, microscopy with ultraviolet surface excitation (MUSE) and open-top light-sheet (OTLS) microscopy, along with various computational methods to enhance, synthesize or analyze the 3D pathological images of tissue specimen, to improve clinical assessment of breast cancer and prostate cancer. For accurate intraoperative lumpectomy margin assessment, we present a fully automated MUSE

system that incorporates 3D deconvolution along with a fluorescent analogue of histology stain, facilitating comprehensive pathology of fresh breast specimen surfaces. For rapid diagnosis of prostate cancer, we developed a compact workflow (1Hr2Dx), consisting of fluorescence labeling, tissue clearing, and 3D OTLS microscopy, to diagnose a set of 12 prostate needle cores within an hour of receipt, which can provide patients with a preliminary on-site diagnosis after a biopsy procedure, thereby alleviating anxiety and potentially expediting treatments. To facilitate more reliable cancer risk evaluation, we developed a workflow for non-destructive 3D pathology and computational analysis of whole prostate biopsies that are labeled with a fluorescent analog of H&E staining. The analysis is based upon the development of an annotation-free deep-learning-based volumetric segmentation strategy, namely image-translation-assisted segmentation in 3D (ITAS3D). Based on the glandular features extracted from our 3D gland segmentations, we show that 3D glandular features are superior to 2D features for risk stratification based on patient outcomes.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

In September 2017, I graduated with my bachelor's degree from Zhejiang University and joined Professor Jonathan Liu's lab at the University of Washington to start my journey as a Ph.D. student. During the past 5 years, I have received so much help, support, and care from people around me, including my advisor, my colleagues, the UW staff, my friends, my partner, and my family. Without these people, I wouldn't have made it to my graduation.

Firstly, I would like to thank my advisor, Professor Jonathan Liu, who, from my perspective, is an excellent advisor for graduate students and principal investigator for a lab. In the past five years, I have benefitted tremendously from Jon's patient and skillful guidance in research, presentation, and writing that transformed me into a better researcher and prepared me for my future career. I am also grateful for his advising style that allowed me to thrive in my PhD studies. I appreciate that Jon has trusted me to do scientific research at my own pace while also providing extremely helpful suggestions and encouragement whenever I needed it. I am especially thankful for the support that Jon gave during my last year, as Jon provided sufficient support and space in my work environment when I was setback by health issues to help me adjust in my daily life.

I would also like to thank my supervisory committee members, Prof. Jonathan Liu, Prof. Eric Seibel, Prof. Lawrence True, and Prof. Ruikang Wang, for their valuable comments and suggestions to improve my dissertation.

I am also grateful to all the pathologists that I have worked with, especially Dr. Suzzane Dintzis, Dr. Nicholas Reder, Dr. Lawrence True, and Dr. Funda Vakar-Lopez, who have been remarkably supportive to all our research projects. They provided clinical insights that brought our cross-disciplinary (engineering + clinical) studies to a new level and allowed them to make an impact to the real world.

I would like to thank all the members in the Liu Lab, for being good friends, collaborators, and mentors to me in the past 5 years. Dr. Winston Wang was the first post-doc that I worked with when I first joined the lab. He spent effort and time to train me on optical systems and his previous work laid a solid foundation for the "MUSE" project, which is also the first project that I took lead on in this lab. Dr. Adam Glaser's fantastic work on building highly versatile open-top light-sheet microscopes also provided robust hardware systems for the "1Hr2Dx" project and the "ITAS3D" project. Despite being very busy, Adam has always been friendly and helpful whenever I turned to him for suggestions or help. Dr. Soyoung Kang has always been a great teacher to all the new students, and I'm thankful for her time managing lab business that kept our lab highly organized. Dr. Ye Chen guided me through my first project with open-top light sheet microscopy and shared with me experience and advice as a senior student in the lab. Dr. Chengbo Yin and Dr. Peter Wei, thank you for teaching me many useful lab skills and lab knowledge. My gratitude also goes to Dr. Lindsey Barner, Mr. Robert Serafin, Mr. Kevin Bishop, Ms. Hongyi Huang, Ms. Vanessa Roybal, Ms. Dominie Miyasato, Ms. Kaylene Pang and Mr. Qinghua Han, for always being friendly and helpful to me as great colleagues. Also, many thanks to Mr. Ray Wang, Mr. Andrew Hu, Ms. Elena Braznenok, and Ms. Lydia Lan, for taking over from where I have left off and continuing this exciting new research.

I would like to acknowledge our capable collaborators from across other labs at the University of Washington (Dr. Mark Fauver, Prof. Eric Seibel, Dr. Nadia Postupna, Prof. Dirk Keene, Dr. Chenyi Mao, and Prof. Joshua Vaughan), and at Case Western Reserve University (Prof. Anant

Madabhushi, Dr. Can Koyuncu, Dr. Patrick Leo, Dr. Andrew Janowczyk, and Dr. Pingfu Fu). These collaborators kindly shared their expertise and assisted us in one or more aspects, including optics, fluorescence labeling, tissue processing, and machine-learning analysis, which were an indispensable part for many of our projects.

Last but not least, I cannot express enough gratitude to my beloved family, especially my mom Wei, my aunt Ning, my grandma Xuanci, and my grandpa Guangdao, for always being caring and supportive to me. I am also grateful to have met my partner Bevin (a brilliant researcher himself) who has been patient, considerate, understanding, and believing in me. His company empowers me to go through challenges and multiplies my happiness in life. And finally, I would particularly like to thank my good friends, including Chen, Yaxuan, and Yuanyuan, for their encouragement, help, and precious friendships. I will always remember the happy times that we spent together in Seattle, and I am lucky to have all of you as wonderful friends in my journey of life.

Chapter 1. INTRODUCTION

1.1 CHALLENGES IN CURRENT PATHOLOGY FOR CANCER ASSESSMENT

According to the National Cancer Institute, approximately 39.5% people will be diagnosed with cancer at some point during their lifetimes [1]. Pathology is the precise study of the causes and consequences of a disease, and it has historically played a key role for many aspects of cancer management, including diagnosis, grading/staging, treatment planning, surgical assessment etc. [2] More recently, pathology has also enabled translational research for interpreting disease mechanism or patient stratification [3, 4].

Accurate and rapid pathological analysis of surgically excised specimens is important for a diagnosis for various types of cancer. While imaging tests like CT scans or MRIs can be helpful in detecting abnormal lesions of tissue, alone they cannot reveal the differences between malignant and benign tissue on the scale of tissue micro-structures or cells. For most cancers, the only way to make a diagnosis is to perform a biopsy excision for pathological analysis of tissue and cells, including needle biopsies [5-7], endoscopic biopsies [8], skin biopsies [9], and bone marrow biopsies [10], etc. For prostate cancer (PCa) assessment, sets of core-needle biopsies are often extracted from patient, where a needle of around 1-mm diameter with a cutting tip is used to draw a column of tissue from a suspicious area. These biopsies are then processed by histopathology procedure and examined under a microscope for diagnosis and grading, which will significantly impact treatment decisions.

Pathological assessment of surgical margin is important to ensure adequate surgical excision of the neoplasm to reduce the chance of recurrence. Once cancer is diagnosed with known location of the tumor, a surgery can be performed to remove localized cancer. Surgery is the most traditional yet still one of the most effective ways to treat many types of cancer, given a complete removal of malignant tumor [11, 12]. For instance, a successful lumpectomy surgery to treat localized breast cancer removes the tumor with an adequate amount of surrounding normal tissue, such that negative resection margins are obtained. However, a current challenge for lumpectomy is to ensure a complete resection of tumor, which is highly correlated with the rate of local recurrence [13]. When post-operative pathology reveals that resection margins are positive for carcinoma, an additional re-excision surgery needs to be performed to minimize the chance of recurrence [14]. Unfortunately, the re-excision after lumpectomy can result in the delay of treatments, emotional trauma, and an increased economic burden.

While pathology has been widely used in different aspects of cancer assessment, the current slide-based histopathology workflow has remained unchanged in the past few decades. In many ways, the current histopathology procedure cannot fully satisfy modern clinical needs. During the procedure, the specimens excised from patients are formalin-fixed and paraffin-embedded (FFPE), cut into think tissue sections about 4 μms in thickness, mounted on to glass slides, and stained with histology stains (e.g., hematoxylin-and-eosin, the H&E stain) for visual examination under a standard light microscope. This slide-based pathology workflow is inherently destructive to the tissue, and the consumed tissue are not preserved for any downstream molecular or genomic assays that is commonly requested in the modern-era of precision medicine [15]. In addition, the thin sections cut from a biopsy provides limited tissue sampling, only representing about 1% of the biopsy. The slide-based 2D visualization and analysis of tissue structures and molecular targets can also be misleading and challenging to interpret due to the 3D nature of tissue and cell structures [16-19]. In an effort to address the above-mentioned issues with 2D pathology, in this dissertation we explore the potential of non-destructive 3D pathology to image tissue specimens in a more comprehensively way, which allows for human examination and computational analysis of 3D tissue micro-structures, to improve cancer assessments in multiple clinical scenarios.

1.2 3D PATHOLOGY ENABLED WITH OPTICAL-SECTIONING MICROSCOPY

Our 3D pathology approaches are fundamentally enabled with 3D microscopy imaging technologies. Optical-sectioning microscopes, including confocal microscopy [20-23], non-linear microscopy [24-28], light-sheet microscopes [29-35], microscopy with ultraviolet surface excitation [36-39], structured-illumination microscopy [40-42], photoacoustic microscopy (PAM) [43-46], and optical coherence tomography (OCT) [47-51] are powerful high-throughput microscopy tools that can resolve sub-cellular details in thick tissue specimens related to tumor examination. Various optical clearing techniques have also advanced our ability for 3D imaging of tissue specimens (e.g., CLARITY [52], iDISCO [53], CUBIC [54], ethyl cinnamate clearing [55], etc.). In this dissertation, we discuss the use of two emerging optical-sectioning microscopy techniques, microscopy with ultraviolet surface excitation (MUSE) [38] and open-top light-sheet (OTLS) microscopy [30], for non-destructive 3D pathology of tissue specimens. We show that multi-channel 3D fluorescent imaging using the optical-sectioning microscopes allows for improved patient management for breast cancer and prostate cancer, the most common cancer diagnosed among U.S. women and men [56], respectively.

1.3 COMPUTATIONAL IMAGE ANALYSIS FOR 3D PATHOLOGY

The development of virtualizing and digitalizing whole glass slides (digital pathology) has made possible computational pathology to assist human observers. In the field of pathology, the growing shortage of experienced pathologists and the ever-increasing amount of health data, especially in digital pathology, has made tools, such as computational image analysis, indispensable to increase the workflow efficiency and diagnostic accuracy, and reduce costs in pathological analysis, showing potentials to realize individualized precision medicine for patients [57-59]. Recent research efforts in artificial-intelligence (AI)-based computational analysis on slide-based pathology images, have shown improvement on the efficiency, reproducibility, and reliability of diagnostic and prognostic pathology [58, 60-64]. Additionally, the rapid and robust computational processing and analysis of pathological images has also been demonstrated in other medical-related tasks, such as image quality enhancement [65-68], image synthesis and modality translation [69-73], etc.

Compared to 2D pathology datasets, 3D pathology significantly increases the image data size (e.g., >100× tissue volume is examined in a prostate biopsy), which creates both challenges and opportunities for pathological analysis. One of the major challenges is that the 3D pathology datasets contain vast amount of feature-rich 3D information that is time-consuming and time-consuming for a pathologist to manually read through. Additionally, a 3D view of thick tissue can present much more complex tissue structure than a 2D view of a thin section, and therefore can be less intuitive for manual interpretation. Currently, there is also a lack of well-trained human experts to read multi-channel 3D pathology images which is often in grayscale fluorescence appearances that are not familiar for most pathologists. In general, 3D computational pathology is in need to assist human observers with more accurate and efficient interpretation of 3D tissue images to thoroughly read into the precious 3D information contained in clinical specimens. In our studies [37, 74], we utilize various computational image processing and analysis techniques, such as 3D image deconvolution, surface extraction, false-coloring, image-to-image translation with deep-learning models, semantic 3D segmentation, on our microscopic images to support surgical guidance, rapid diagnosis, and risk stratification of cancer patients.

1.4 OUTLINE FOR THE FOLLOWING CHAPTERS

In Chapter 2, we describe our exploration on microscopy with UV surface excitation (MUSE), a camera-based microscopy that provides superficial optical sectioning of thick excisional specimens, enabled by short penetration depth of UV light in tissue [38]. This convenient and simple principle which makes it suitable for various clinical applications that requires efficient visualization of tissue surfaces. In this study, we combined the use of MUSE for 2-channel 3D imaging of specimen surface with computational image enhancement and visualization techniques, namely 3D deconvolution, surface extraction and false coloring, to achieve comprehensive pathology of fresh breast tissue surfaces that are labeled with a fluorescent analog of H&E stain. This study demonstrated the potential utility of MUSE for intraoperative surgical margin assessment to guide lumpectomy procedures [37].

In Chapter 3 and 4, we focus on 3D pathology for prostate cancer. Prostate cancer prognostication and treatment planning typically depend upon assessing carcinoma grade in biopsies. We utilize an open-top light-sheet (OTLS) microscope [30] along with different staining and clearing protocols to visualize biological features in biopsies in depth. In an OTLS microscope, the illumination and collection objectives are placed below a sample plate that holds a cleared specimen. This open-top architecture enables multiple specimens to be 3D imaged at a high resolution within a single imaging session.

Two projects are based on OTLS imaging of prostate biopsy tissues. In the 1-hour-to-diagnosis (1Hr2Dx) project (in Chapter 3), we developed a less-than-one-hour workflow that consists of fluorescent staining, brief tissue clearing, 3D OTLS imaging (with more extensive sampling than 2D slides), image processing and pathological diagnosis of 12 core-needle biopsies.

This system could provide patients with a preliminary on-site diagnosis within one hour after a biopsy procedure, thereby alleviating anxiety and potentially expediting treatments [75].

In Chapter 4, we discuss the potential of prostate cancer risk stratification with a 3D segmentation pipeline and analysis of glandular features. We developed an annotation-free deep-learning-based strategy to segment out the 3D glandular network of the prostate, namely image-translation-assisted segmentation in 3D (ITAS3D), for cancer risk stratification. Our strategy relies upon synthetic 3D immunofluorescence (IF) enabled by generative adversarial networks (GAN) and a fluorescent analog of H&E, which is fast and inexpensive to label. From the synthetic IF, we can objectively segment out the glands using a relatively straightforward thresholding- and morphology-based algorithm. A preliminary clinical validation of this pipeline is performed with 50 patient cases. Based on the glandular features extracted from our 3D gland segmentations, we show that 3D glandular features are superior to 2D features for prognostic risk stratification based on biochemical recurrence (BCR) outcomes. This pipeline is also potentially useful for rapidly extracting any clinically valuable 3D tissue structures that can be immunolabeled without the need for tedious manual annotation. Our findings support the value of 3D pathology for managing the low- to intermediate-risk prostate cancer [74].

Chapter 2. MICROSCOPY WITH ULTRAVIOLET SURFACE EXCITATION FOR PATHOLOGY OF LUMPECTOMY MARGINS

2.1 BACKGROUND AND MOTIVATION

Lumpectomy is a breast-conserving procedure used to surgically remove breast carcinoma along with a margin of normal tissue surrounding the tumor. Over the years, studies have shown that patients with invasive breast cancer, treated with lumpectomy and local breast irradiation, exhibit no significant difference in survival compared with those undergoing mastectomy, provided that the entire tumor is removed during lumpectomy surgery [11, 76]. However, if post-operative histopathology reveals the presence of carcinoma at the surgical margin, re-excision procedures, including completion mastectomies, are often performed to minimize the possibility of local recurrence [77]. Re-operation after lumpectomy can result in the delay of adjuvant treatments with increased risks of recurrence [78], emotional trauma [79], as well as an increased economic burden for both the healthcare system and patients. The rate of re-operation after lumpectomy has been calculated to be at least 20% according to recent studies [77, 80, 81]. Therefore, there is an acknowledged need for reliable intraoperative methods to assess the surgical margins of freshly resected lumpectomy specimens. (**Figure 2.1**)

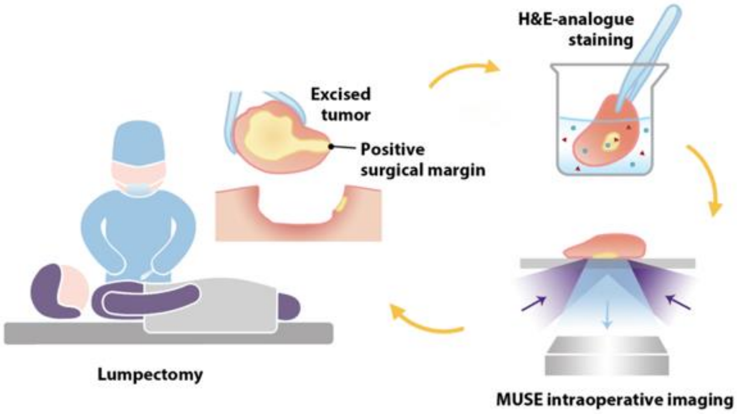

Figure 2.1. Clinical workflow for intraoperative surgical margin assessment by Microscopy with Ultra-violet Surface Excitation (MUSE). During lumpectomy, the freshly resected tissue specimen is immediately stained with a fluorescent analog of H&E staining. The surgical margin surface of the resected specimen is then imaged with MUSE to guide the resection procedure.

Conventional gold-standard histopathology relies upon the microscopic imaging of thin tissue sections (usually ~4-µm thick) mounted on glass slides and stained with hematoxylin and eosin. This lengthy process of tissue fixation, dehydration, paraffin embedding, sectioning and staining requires hours to days (**Figure 2.2**) and is not possible within intraoperative time frames. Additionally, only a few thin cross-sections of the sample are examined in conventional histology, where there is extremely limited sampling of the surgical margin surface [82]. As an alternative, frozen-section histology enables rapid intraoperative assessments to guide certain resection procedures [83], but suffers from inferior image quality and similar sampling limitations as standard histology [84, 85]. Furthermore, frozen sections require the destructive sectioning of tissues that should ideally be preserved for gold-standard post-operative histology. Finally, frozen sectioning introduces freezing artifacts, especially in fatty breast tissues, which negatively impact the quality of downstream post-operative histology [86].

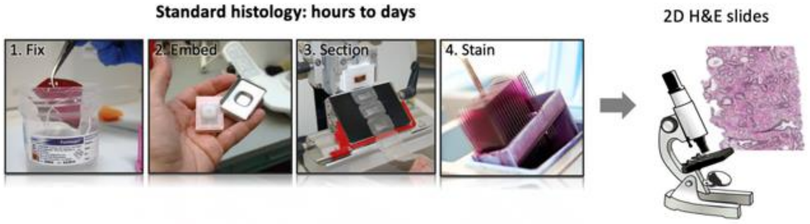

Figure 2.2. The standard histopathology workflow to obtain hematoxylin and eosin (H&E) slides for manual microscopic analysis.

Optical-sectioning microscopy techniques can potentially enable rapid non-destructive imaging of surgical margin surfaces to guide tumor-resection procedures. All the optical sectioning techniques mentioned in Chapter 1 involve trade-offs between speed, simplicity, cost, resolution, and image contrast/depth. In particular, while laser-scanned microscopy techniques such as confocal and nonlinear microscopy can provide exquisite resolution and contrast/depth, even without exogenous contrast agents in the case of nonlinear microscopy, they tend to be more complex, bulky, slow, and expensive. For surface microscopy within a surgical setting, camera-based optical-sectioning techniques such as open-top light-sheet (OTLS) microscopy and structured-illumination microscopy (SIM) have clear benefits regarding speed, and in the case of SIM, clear advantages concerning cost and simplicity. Finally, PAM and OCT can provide label-free high-resolution imaging based on absorption and reflectance contrast, respectively, but are somewhat limited in their ability to provide multi-color imaging of tissue structures that pathologists are accustomed to viewing (e.g., H&E histology).

Here, we explore a recent camera-based superficial optical-sectioning strategy, known as microscopy with ultraviolet surface excitation (MUSE), which is relatively simple, low-cost, and potentially well-suited for intraoperative surgical margin assessment [36, 38, 87]. With MUSE, the use of short-wavelength (285-nm) ultraviolet (UV) light limits the penetration of the illumination light to within ~10 µm beneath the tissue surface, such that a relatively high-contrast “optically sectioned” image of the tissue surface may be obtained rapidly with a 2D detector array. Many conventional visible-excited fluorescent dyes can be excited with UV illumination, with high fluorescence yields in the visible range [38]. Since the collected fluorescence is in the visible range, there is no need for expensive UV-compatible optics. Inexpensive UV LEDs may be used, rather than lasers, to generate sufficient illumination light over a reasonable field of view (typically on the millimeter scale). These features make MUSE particularly cost-effective as compared with other microscopy alternatives.

2.2 A MUSE SYSTEM OPTIMIZED FOR DUAL-CHANNEL FLUORESCENCE IMAGING

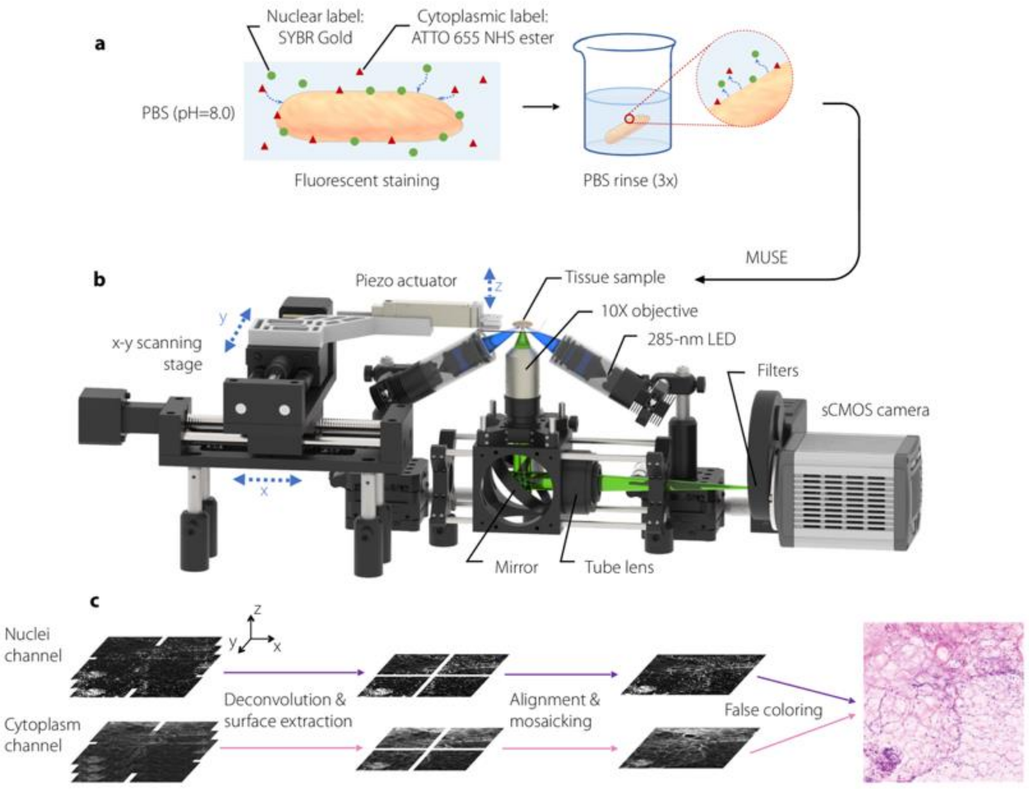

Figure 2.3. Schematic of intraoperative MUSE system. (a) A fluorescent analog of H&E staining is used to label fresh tissues within ~5 minutes. (b) A MUSE system for comprehensive imaging of fresh specimens. Two LEDs (285-nm wavelength) illuminate the specimen surface at

an oblique angle from opposite directions to reduce shadowing artifacts. Fluorescence signal (visible wavelength) is collected and imaged by a 10× (N.A. = 0.3) apochromatic objective and tube lens onto a 2D detector array. A filter wheel is used to image the two-color channels sequentially. Large-area tiled imaging is achieved by scanning the specimen with a motorized XY stage. At each imaged location (lateral image tile), the piezo actuator scans the specimen vertically to obtain a z-stack of images, which allows for 3D deconvolution (to improve resolution and contrast), surface extraction (to mitigate the effects of surface irregularities), and false coloring to mimic the appearance of gold-standard H&E histology, as shown in panel (c).

An optical schematic of our MUSE system is shown in **Figure 2.3b**. UV illumination is provided by two 285-nm LEDs (M285L5, Thorlabs), in which the light is roughly collimated and focused at the specimen through the use of a pair of identical UV-fused-silica lenses (F = 35.0 mm, LA4052-UV Thorlabs). The two UV beams pass through the clean-up filters (FF-1-285/14-25, Semrock) and illuminate the tissue surface at an oblique angle (to avoid being clipped by the large collection objective) from opposite directions to reduce shadowing artifacts. The sample is placed on a UV-fused-silica slide, in which the bottom surface of the tissue sample (illuminated surface) is positioned slightly below the focal plane of the UV beam so that the illumination spot is enlarged and generates a relatively uniform irradiance distribution across the field of view of the collection objective, with an average power of 14 mW (power density ~ 1.55×10^{3} $W\cdot m^{-2}$) incident on the sample. For large-area imaging of tissue specimens, ultrasound gel (Parker Laboratories, Inc.) was placed around the edge of the specimen to reduce tissue deformation caused by dehydration during imaging and to obtain a clearer image of the tissue edge (due to index matching through the gel, rather than air). Visible-wavelength fluorescence signal is collected and imaged by a 10× apochromatic objective (CFI Plan Fluor 10×, NA 0.3, Nikon) and tube lens (ITL200, Thorlabs) onto a 2D detector array (ORCA-Flash4.0 V3 Digital CMOS camera, Hamamatsu) with a lateral resolution of ~1 µm and a field of view of 1.3 mm by 1.3 mm. The axial resolution of the 0.3 NA objective is ~12 µm, which roughly matches the penetration depth of UV light in fresh breast, which is about 10-20 µm. Large-area tiled imaging is achieved by scanning the specimen with a motorized XY stage (ET-50-11, Newmark systems). At each lateral imaging location, the piezo actuator (P-601.4SL, PI) scans the specimen vertically twice to obtain two *z*-stacks of images, in which a motorized filter wheel is used to change the optical filters of 550/40 BP (FB550-40, Thorlabs) for the nuclei channel and 664 LP (LP02-664RU-25, Semrock) for the cytoplasm channel between the two vertical scans. This rapid collection of the two image channels prevents channel co-registration problems due to tissue deformation over time.

Image tiles are collected at a lateral spacing of 1 mm in both the *x* and *y* directions, which represents an optimal balance between imaging speed and mosaicking accuracy. The vertical step size is 5 µm, which satisfies the Nyquist sampling criterion (assuming an axial resolution of ~12 µm). The total *z*-scanning range is 150 µm, which is sufficient to accommodate a majority of the surface irregularities seen at the tissue surface. The integration time for each frame is 50 ms (UV radiant exposure per frame of ~77.5 $J \cdot m^{-2}$), which represents a balance between speed and signal-to-noise ratio (SNR). Large-area image acquisitions are automatically controlled by a custom-developed LabVIEW program.

2.3 UV-EXCITED TWO-COLOR FLUORESCENT ANALOG OF H&E STAINING FOR FRESH TISSUE

To generate images that mimic the appearance of gold-standard H&E histology, which pathologists are accustomed to viewing, a two-color UV-excited fluorescent analog of H&E staining was developed to rapidly stain fresh tissues for MUSE. This method utilizes SYBR Gold (λ_{Em} = 537 nm) as a nuclear stain [88], and ATTO 655 NHS ester (λ_{Em} = 680 nm) as a stromal/cytoplasmic stain (**Figure 2.3.a**). Both of these fluorescent agents can be excited at 285 nm. In our optimized protocol, fresh tissues were stained by submerging them in a 0.825:10,000 v/v solution of SYBR Gold (Thermo Fisher, Cat. No: S11494) and 11.25-µM ATTO 655 NHS ester (Sigma-Aldrich) in 1× Phosphate-Buffered Saline (PBS, Gibco, Cat. No: 10010023) at pH 8.0 for 5 minutes. The stained tissue was then rinsed 3 times in a large volume of 1× PBS (pH 7.4) for 30 seconds per wash, followed by MUSE imaging. For control experiments to compare cytoplasmic/stromal staining approaches, a published MUSE protocol with eosin staining was used [36]. In short, tissues were stained with 200 µg/ml eosin for 2 minutes and then rinsed in 1× PBS.

While eosin is naturally fluorescent and has previously been utilized in fluorescent analogs of H&E staining, it does not bind well to fresh tissues and has often exhibited a "leakage" issue in tissues that are not formalin fixed and dehydrated [36, 89]. We show here that NHS esters can serve as an alternative to eosin for staining fresh, hydrated tissues with improved contrast and microscopic detail in comparison to eosin. MUSE images of mouse kidney stained with ATTO 655 NHS ester and eosin are shown in **Figure 2.4**. As a result of improved tissue binding, the ATTO 655 NHS ester can reduce the pooling effects caused by eosin leakage, and also provides an enhancement in microstructural contrast.

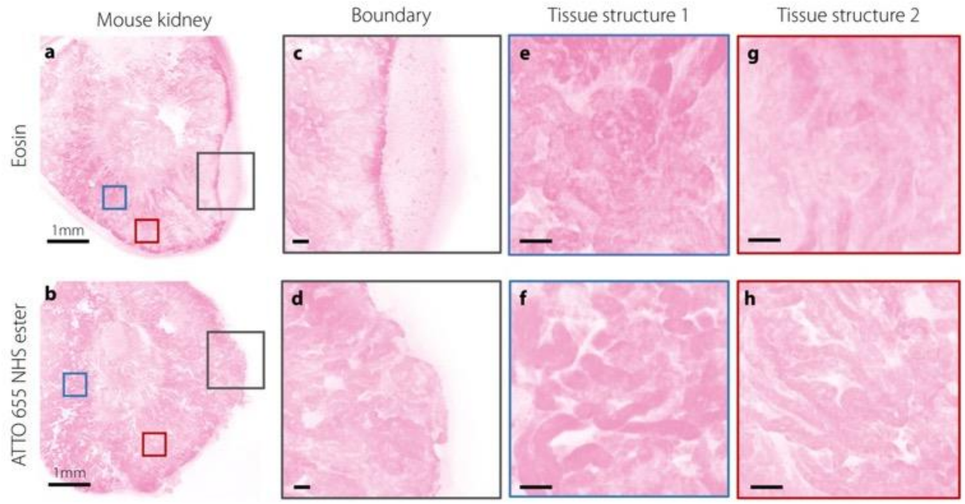

Figure 2.4. A comparison of MUSE image quality with various stromal/cytoplasmic stains. The images show examples of fresh tissues stained with (a) eosin and (b) ATTO 655 NHS ester. With fresh, hydrated specimens, the leakage of eosin, which is poorly bound, results in a pooling effect around the edge of the specimen, as shown in (c), and also reduces image contrast of microstructures (e, g). ATTO 655 NHS ester does not exhibit leaking at the tissue edges (d) and provides improved microstructural contrast (f, h). Scale bar: 1mm (a, b), 100 µm (c-g, d-h).

2.4 IMAGE ENHANCEMENT AND FALSE COLORING FOR MUSE IMAGES

A recent study on MUSE [36] reported that the penetration depth of UV light (i.e., the optical-sectioning thickness) in tissues is approximately 10 - 20 um, which is significantly thicker than conventional histology sections (~4 µm). Thicker sectioning degrades the image quality by introducing more background from out-of-focus light generated above and below the focal plane, which ultimately leads to a reduction in resolution and contrast [90, 91]. To further approach the image quality of gold-standard histology, we collected a *z*-stack of MUSE images at each lateral position and applied 3D deconvolution to improve the resolution and contrast. In addition, a surface extraction algorithm was used to minimize the effects of surface irregularities. With 3D deconvolution, the point spread function (PSF) of the imaging system is used to restore the detected image, including the resolution and signal-to-background ratio, such that the image provides a more faithful representation of the original object [91].

2.4.1 *PSF measurement*

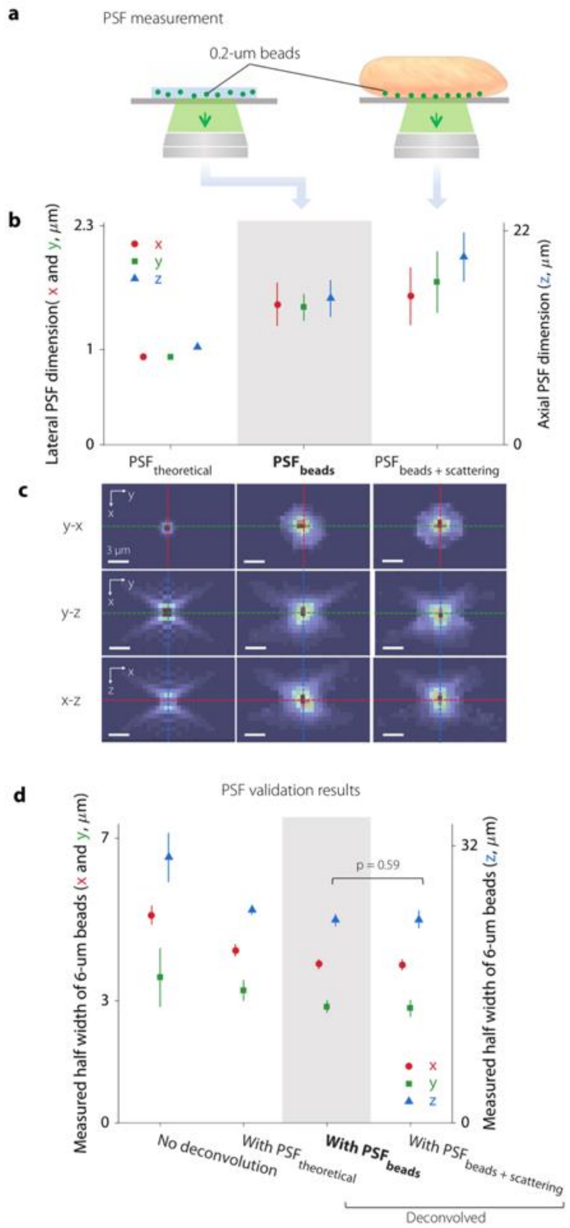

Figure 2.5. Obtaining and validating a point spread function (PSF) for deconvolution. (a) Illustration of two PSF measurement methods, one of which images sub-resolution fluorescent beads (d = ~0.2 µm, emission peak at 520 nm) in an agarose phantom (standard method) and another of which images the same beads at the surface of human breast tissue (which adds the effects of tissue scattering). (b) Results showing the full width at half maximum (FWHM) of the PSF theoretically calculated from system parameters ($\mathrm{PSF}_{\mathrm{theoretical}}$), the average PSF

experimentally measured with beads in an agarose phantom (PSF_{beads}), and the average PSF experimentally measured with beads at the tissue surface ($PSF_{beads+scattering}$). (c) Three orthogonal cross-sectional views of $PSF_{theoretical}$, PSF_{beads} and $PSF_{beads+scattering}$. The colors of the dashed lines correspond to the colors of the data points in (b) and (d) for the x, y and z directions (red, green, and blue, respectively). Scale bar: 3µm. (d) PSF validation results showing the average FWHM of 6-µm beads at a tissue surface without deconvolution, and with deconvolution using three different PSFs. The use of PSF_{beads} for deconvolution yields the best results regarding resolution and variance. While $PSF_{beads+scattering}$ is more accurate, it does not provide improved deconvolution results (as shown by the p-value). (b, d) Left vertical axis: FWHM in the x and y directions (red and green). Right vertical axis: FWHM in z-direction (blue).

Measurements and validation of the point spread function (PSF) of our MUSE system are shown in **Figure 2.5**. All of the experimental data shown in the diagrams are averaged from 8-10 samples with the indicated uncertainties (standard deviation). In the PSF measurement results shown in **Figure 2.5b**, both of the experimentally measured PSFs (PSF_{beads} and $PSF_{beads+scattering}$) show a slightly larger FWHM than the theoretically calculated PSF ($PSF_{theoretical}$). This is due to a variety of minor aberrations introduced by the experimental system. For the two experimental PSFs, the FWHM of $PSF_{beads+scattering}$ is slightly larger than PSF_{beads}. This is expected since beads located at the surface of a scattering tissue (human breast in this study) are influenced by the scattering background, which leads to a slightly enlarged PSF. Theoretically, among the three PSFs we obtained, $PSF_{beads+scattering}$ should most accurately describe the blurring of point objects in tissues that are imaged with MUSE (i.e., due to diffraction, aberrations, and tissue-scattering background). For more intuitive visual interpretation, three orthogonal views of the measured PSFs are shown in **Figure 2.5c**.

To determine which PSF to use for MUSE image deconvolution, images of 6-µm beads at breast tissue surfaces were deconvolved using the three PSFs described above (experimentally averaged PSFs). The FWHM of the 6-µm beads before and after deconvolution is shown in **Figure 2.5d**. The asymmetry in the size of beads in the x and y directions is because the two-sided illumination employed in our system is aligned along the x-direction. The experimental PSFs generate the best improvements in resolution through deconvolution (compared to the theoretical PSF). However, no significant difference is observed between the dimensions of the 6-µm beads deconvolved with the two different experimental PSFs, as shown by the p-value. Furthermore, the FWHM of the beads deconvolved with PSF_{beads} exhibits less variability than that using $PSF_{beads+scattering}$. Therefore, in the rest of this study, we used PSF_{beads} for all deconvolution procedures.

2.4.2 *Image enhancement using 3D deconvolution, surface extraction and false coloring*

Data post-processing consists of the following steps (**Figure 2.3c**): (1) 3D deconvolution is performed for contrast and resolution enhancement. Ten iterations (optimized for speed and image quality) of a Lucy-Richardson algorithm (MATLAB) are applied for each vertical image stack (*z* stack). (2) Surface extraction is performed for surface-irregularity mitigation. For each *z* stack of deconvolved images, we used an open source ImageJ plugin to perform a complex wavelet transform [92] that takes a vertical stack of images and extracts the best focus (tissue surface) for each lateral sub-region (the parameters for the algorithm are tunable and provide a trade-off between extraction quality and speed). (3) The images from the two channels at each lateral position are co-registered (aligned) with a normalized cross-correlation algorithm. (4) The aligned images for each channel are mosaicked by using the Grid/Collection Stitching Plugin in ImageJ [93]. (5) The large-area (mosaicked) images are false-colored (by combining the two image channels) to mimic the appearance of gold-standard H&E histology. Here, we have optimized an H&E false-coloring algorithm that was previously published [94]. In short, localized histogram equalization is applied to each channel, prior to H&E false coloring, to enhance the contrast and consistency of the false coloring across a large-area image.

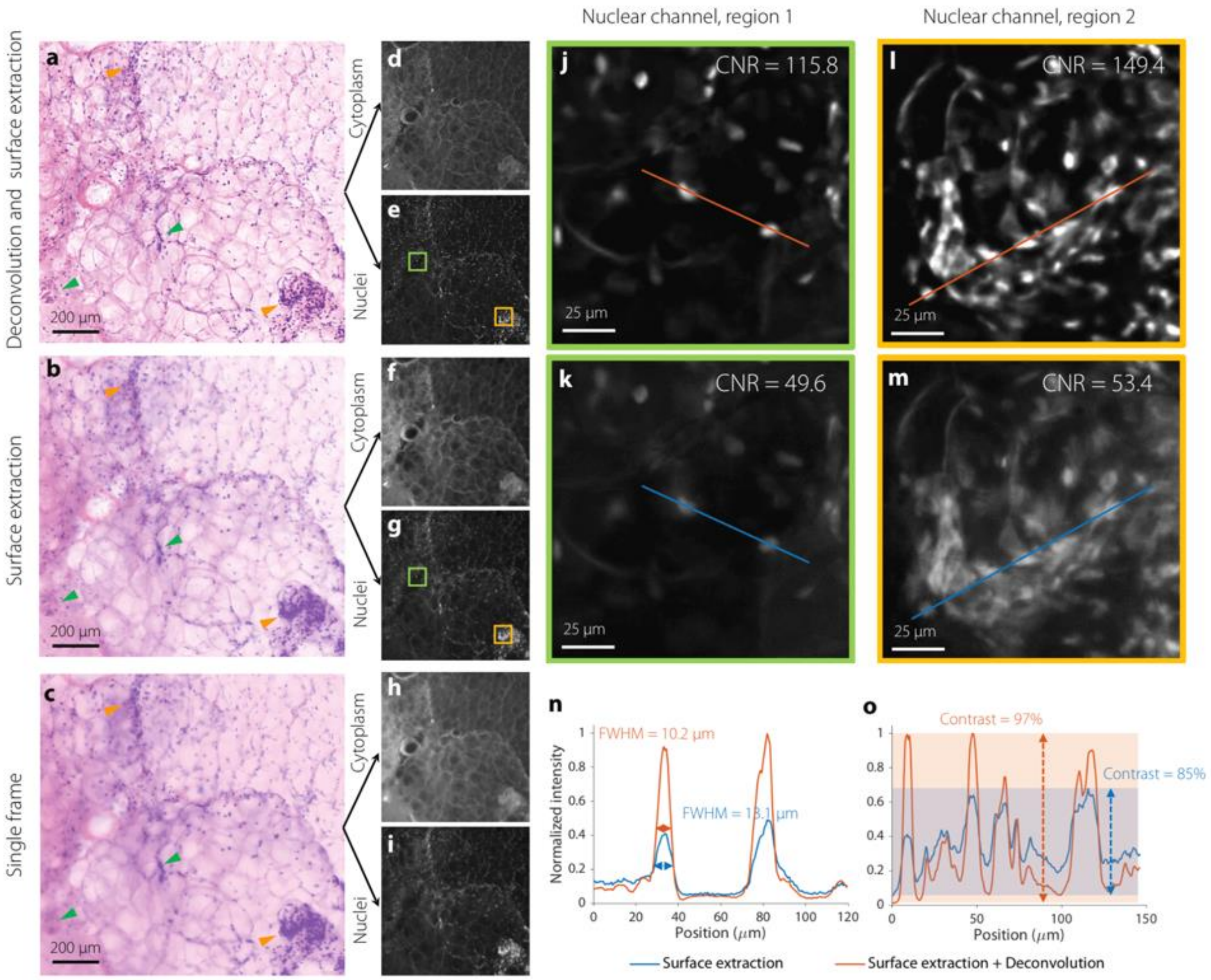

Figure 2.6. A comparison of image quality for MUSE of breast specimens. Results are shown (a) with deconvolution and surface extraction, (b) with only surface extraction, and (c) of a single frame within a z-stack. The green arrows indicate features that are out of focus in a single frame but are brought into focus by surface extraction. The orange arrow points to a cluster of nuclei that are significantly more resolved after deconvolution. These MUSE images are false-colored to mimic H&E histology, based on a nuclear channel (e, g, i) and a stromal/cytoplasmic channel (d, f, h). Two zoomed-in regions from the nuclear channel (j-m) show that deconvolution enables the resolution, contrast, and contrast-to-noise ratio (CNR) to be improved. A line profile from region 1 (n) shows that the FWHM of the nuclei are smaller (more resolved) with deconvolution than

without. A line profile from region 2 (o) shows that the overall image contrast is also improved by deconvolution.

Figure 2.6 shows an image-quality comparison for breast tissues imaged with MUSE. Examples are shown in which deconvolution and surface extraction are both used, surface extraction is used alone, and a single unprocessed frame within a *z* stack. Surface extraction of a *z* stack (150-μm range) brings a number of out-of-focus features (caused by surface irregularities) into focus within a single frame image, as indicated by the green arrows. Deconvolution further improves the image resolution, contrast, and contrast-to-noise ratio, as shown in the zoomed-in nuclear-channel images (**Figure 2.6j-m**) and line profiles (**Figure 2.6n, o**). In summary, the use of 3D deconvolution allows features to be better resolved, as indicated by the yellow arrows in **Figure 2.6a-c**.

2.5 RAPID PATHOLOGY OF FRESH SURGICALLY-EXCISED BREAST SPECIMENS WITH MUSE

We show examples to demonstrate that the image quality of our MUSE approach surpasses that of frozen-section histology, and also compare the image quality of MUSE with archival H&E histology of formalin-fixed paraffin-embedded (FFPE) tissues. Specific differences in appearance will be highlighted as we evaluate the feasibility of this approach for intraoperative surgical margin assessment to guide lumpectomy procedures.

2.5.1 *Comparison with frozen sectioning*

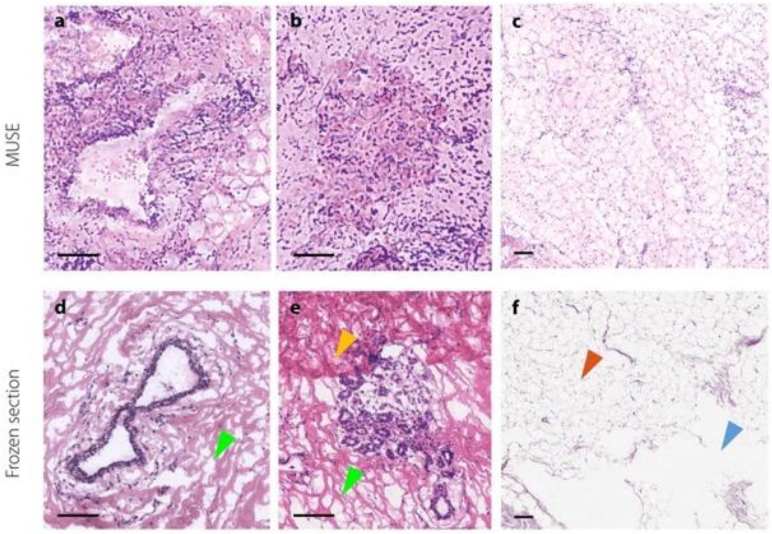

Figure 2.7. Human breast tissue imaged with MUSE in comparison to frozen section histology. Large ducts are shown in MUSE (a) and frozen-section histology (d) images. The frozen sections exhibit optically clear spaces (d and e, as indicated by the green arrow) between collagen fibers, which are freezing artifacts that do not appear in the MUSE images. Lobular units are shown in MUSE (b) and frozen section histology (e) images. The yellow arrow in (e) points to a tissue-fold artifact in the frozen section. Adipose tissue is shown in MUSE (c) and frozen-section histology (f) images. The red arrow in (f) indicates distorted adipose tissue and the blue arrow indicates a tear artifact from sectioning the frozen breast tissue. Scale bar: 100 µm.

The image quality of MUSE vs. frozen sectioning is shown in **Figure 2.7**. In the frozen sections, freezing artifacts result in uneven stromal structure with looser collagen (**Figure 2.7d, e**, green arrow) and distortion of the lipids (**Figure 2.7f**, red arrow). In particular, the adipocytes are highly disrupted and poorly sectioned in the frozen sections (**Figure 2.7f**, blue arrow). Tissue folds are also apparent (**Figure 2.7e**, yellow arrow), which can potentially obscure underlying regions of carcinoma. These diagnosis-impairing issues in the frozen sections are not present in the MUSE images because the tissues are imaged fresh and without physical sectioning.

2.5.2 *MUSE Pathology of human breast tissue compared with H&E histology*

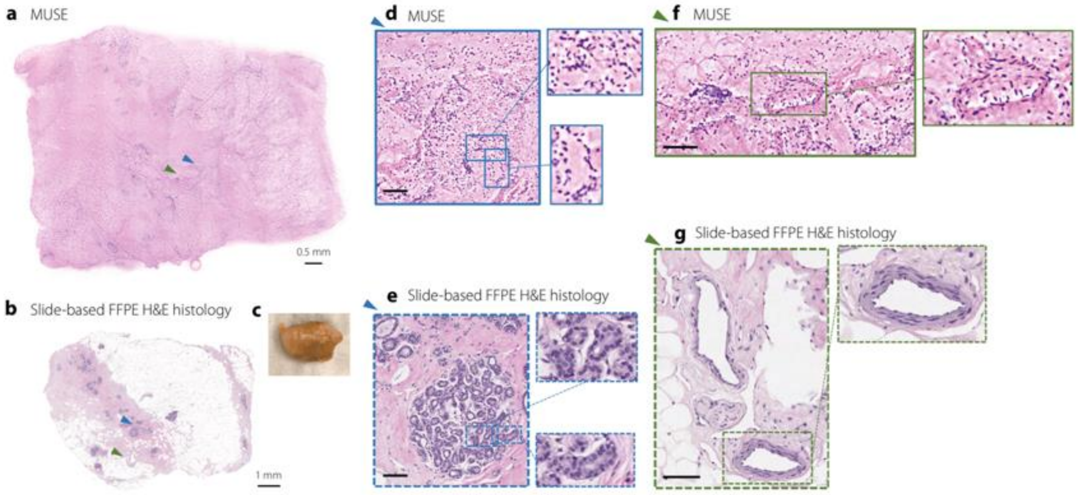

Figure 2.8. MUSE imaging (H&E analog) of benign human breast tissue. After MUSE imaging (a), a fresh benign human breast specimen (9 × 10 × 5 mm), shown in the photo in (c), is submitted for slide-based FFPE H&E histology (b). Images generated by MUSE (d) and slide-based FFPE H&E histology (e) of benign breast lobules are shown with zoomed-in regions highlighting individual acini. Images generated by MUSE (f) and slide-based FFPE H&E histology (g) of breast tissue containing collagen-rich stroma, adipose, and neurovascular bundles, with a zoomed-in region showing a venule. Scale bar: 100 µm (d-g, first level zoom-in**).**

Figure 2.8 and **Figure 2.9** show that our customized MUSE system can enable comprehensive pathology of large-area freshly excised breast specimen surfaces. Image quality is compared to that of gold-standard slide-based FFPE H&E histology. The zoomed-in images and insets show benign breast lobules (**Figure 2.8d-e**), stroma and neurovascular bundles (**Figure 2.8f-g**), adipose tissue (**Figure 2.9e-g**), a benign duct (**Figure 2.9h-j**), and ductal carcinoma in situ (DCIS) (**Figure 2.9k-m**).

Differences exist between MUSE and slide-based FFPE H&E images. For example, the NHS ester staining for MUSE appears to be darker and more spatially distributed than the eosin staining for slide-based FFPE histology; cross sections of individual acini often show optically clear lumens in slide-based FFPE H&E images (**Figure 2.8e**), whereas the acini shown in the MUSE images are filled with eosinophilic material (**Figure 2.8d**); zoomed-in views of the vein (**Figure 2.8f**) and duct (**Figure 2.9h**) in the MUSE image show an oblique perspective with a short segment of the inner walls which are absent in slide-based FFPE H&E images (**Figure 2.8g**, **Figure 2.9i**). This oblique perspective is discussed further in the next few paragraphs. Multiple layers of adipocytes are seen in the MUSE images (**Figure 2.9e**) while only a monolayer of cross-sectioned adipocytes is viewed against a clean background in slide-based FFPE H&E images (**Figure 2.9f**); The epithelial cells filling the lumen of the duct appear less basophilic in the MUSE image (**Figure 2.9k**) than in the slide-based FFPE H&E image (**Figure 2.9l**).

We speculate that the differences in image quality between MUSE and conventional histology are due to three main reasons. Firstly, as stated for the staining comparison in Figure **2.6**, ATTO 655 NHS ester binds covalently to cytoplasmic proteins, with improved tissue binding than eosin. Secondly, the optical sectioning thickness of MUSE for fresh breast tissues is on the order of 10 to 20 microns [36], which is significantly larger than the thickness of a histology section (~4 µm). This increased section thickness of MUSE will result in increased background which may obscure specific finer structures that are seen in the conventional histology sections. Lastly, the surface topology and cellular contents that are present at the surfaces of grossly cut fresh tissue may be different from microtome-sectioned tissues that are formalin-fixed and paraffin-embedded (FFPE). For example, the acini/cells may be cut open during microtome sectioning of FFPE tissues while the surfaces of grossly cut fresh tissue may not contain intra-acinar/intracellular contents. In addition, grossly cut tissue may exhibit irregular and soft surfaces that can deform when placed on the glass slide of the MUSE sample holder, while FFPE tissues are rigidly embedded in wax and sectioned without deformation. Due to the above reasons, for example, in the MUSE image shown in Figure **2.9**.h, a hollow structure appears to be flattened and flayed open at an angle such that the luminal surface on one side of the structure is visible. The relatively large optical-section thickness of MUSE can also contribute to this “oblique perspective,” which is not seen in microtome-sectioned FFPE histology.

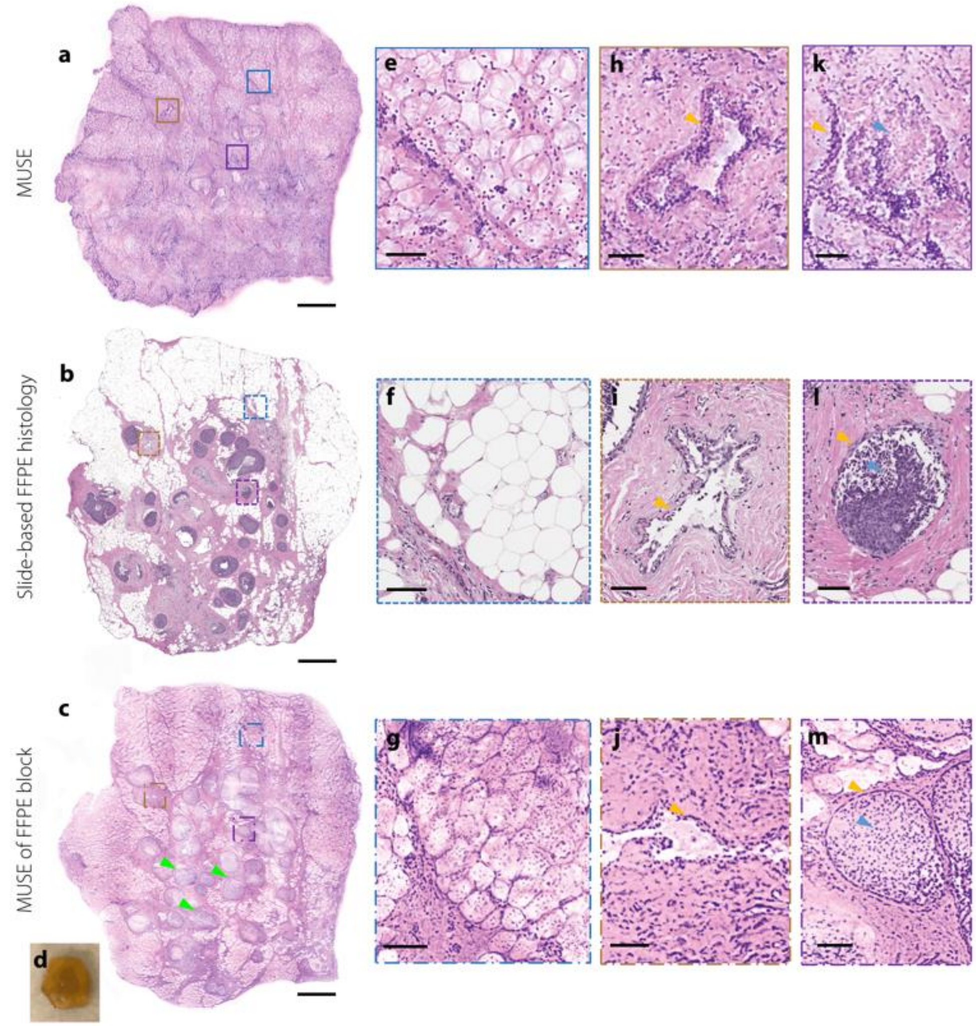

Figure 2.9. Breast tissue with human ductal carcinoma in situ (DCIS, green arrows) imaged with MUSE (a), slide-based FFPE histology (b), and MUSE of an FFPE block face (c). A photo of the unstained tissue is as shown in (d). Zoomed-in features imaged with MUSE (e, h, k), slide-based FFPE histology (f, i, l) and MUSE of an FFPE block face (g, j, m). Adipose tissue is shown in (e-g). A benign duct is shown in (h-j). DCIS is shown in (k-m). Scale bar: 1 mm (a-c), 100 µm (e-m).

To remove the effects of irregular surface topology while maintaining the optical-sectioning thickness of MUSE, we performed MUSE imaging of the cut surfaces of FFPE tissue blocks treated with the same SYBR-Gold-and-ATTO-655-NHS-ester staining protocol (**Figure 2.9c, g, j, m**). With block-face MUSE imaging, since the flat surface topology and exposed intracellular contents are similar to slide-based FFPE histology, the images (**Figure 2.9c, g, j, m**) are more similar to standard histology than MUSE images of fresh tissues (**Figure 2.9a, e, h, k**) are in comparison to standard histology, but with some remaining differences. For example, **Figure 2.9g** shows that, as with MUSE of fresh tissues, MUSE images of FFPE block faces reveal deeper layers of adipose tissue due to a thicker optical sectioning extent (10 - 20 µm) than 4-µm physical sections. However, detailed ductal features in MUSE images of a FFPE block (**Figure 2.9j** and **Figure 2.9m**) more-closely resemble slide-based FFPE H&E images (**Figure 2.9i** and **Figure 2.9l**) than MUSE images of fresh tissues (**Figure 2.9h** and **Figure 2.9k**) in terms of duct walls (yellow arrows) and ductal carcinoma cells (blue arrows). In summary, MUSE images of FFPE block faces support our conjecture that the differences between MUSE images of fresh tissue and slide-based FFPE H&E images are a result of differences in both surface topology and sectioning thickness.

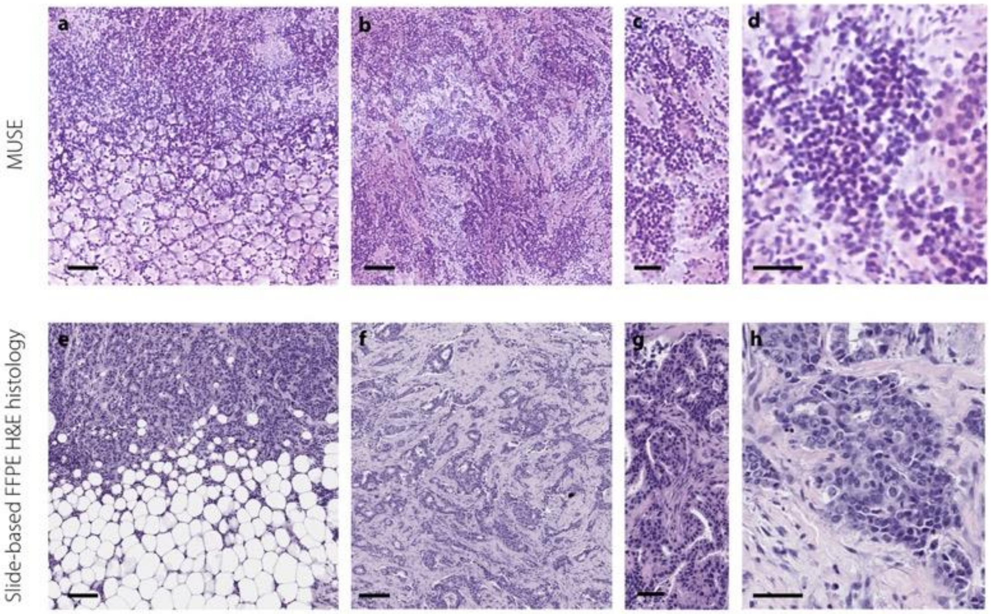

Figure 2.10. Human invasive ductal carcinoma (IDC) imaged with MUSE. Images generated by MUSE (a) and slide-based FFPE H&E histology (e) are shown of IDC that has invaded a region of adipose tissue. Images generated by MUSE (b, c) and slide-based FFPE H&E histology (f, g) of IDC that has invaded a region of fibrous stroma. Images generated by MUSE (d) and slide-based FFPE H&E histology (h) of IDC cells with variable nuclear chromatin structure. Scale bar: 100 µm (a, b, e, f), 50 µm (c, d, g, h).

As a final example, **Figure 2.10** shows MUSE images of invasive ductal carcinoma (IDC) in comparison to slide-based FFPE H&E histology. Note that with similar magnification, the variable nuclear chromatin structures are more observable in slide-based FFPE H&E images than MUSE images. We infer that this can be attributed to the two reasons stated previously: (1) differences in the optical sectioning thickness of MUSE compared with the physical sectioning thickness of slide-based FFPE histology, and (2) differences in surface topology and exposed cellular structures between grossly cut fresh tissues and microtome-sectioned FFPE tissues.

While our current system can achieve imaging speeds that are relevant for intraoperative use, the MUSE system can potentially reach higher speeds to optimize its use during surgery. Currently, the imaging speed is mainly limited by optical detection sensitivity, in which an exposure time of 50 ms/frame is required for sufficient SNR. Potential approaches to improve the imaging speed include: (1) Raising the illumination power to compensate for the lower absorption cross section of the fluorescence contrast agents at 285 nm compared to their visible-wavelength absorption peaks. While autofluorescence background will be increased, this background is negligible at our current settings and should not be a major limiting factor. (2) Increasing the step size in the z (depth) direction (the current z step size is 5 µm), which will reduce the total scanning time per z-stack. Further experiments are needed to determine whether this would degrade the performance of our deconvolution algorithms. (3) Implementing tissue-flattening methods to reduce the topological variation of the tissue surface, such that a smaller depth range is needed to be scanned (for surface extraction). In this study, we are gently flattening the tissue against the sample holder and avoiding excessive pressures to minimize unintended artifacts. Studies are needed to determine if more-aggressive tissue-flattening mechanisms can be employed without degrading image quality.

In the past few years, many exciting studies by other research groups have been done by leveraging our technical development of MUSE for intraoperative imaging: Pocket MUSE [95], FLARE [96], CoreVIEW [97], a quantitative study of MUSE diagnostic accuracy and speed [98], etc.

While there are apparent differences in the images generated by MUSE versus traditional slide-based FFPE histology, these differences may not ultimately be detrimental once pathologists are trained to interpret MUSE images. This work focuses on early technical advances in MUSE microscopy that support its feasibility for rapid pathology of lumpectomy margins. In the future, well-powered clinical studies, with a variety of breast pathologies, are needed to determine if the sensitivity and specificity of carcinoma detection with MUSE can compare favorably with conventional slide-based FFPE histology, and to determine if the increased comprehensiveness of the MUSE imaging approach can improve detection sensitivity. In summary, this study has demonstrated the potential utility of MUSE for intraoperative surgical margin assessment to guide lumpectomy procedures, as well as other surgical oncology applications in the future.

Chapter 3. DIAGNOSING 12 PROSTATE NEEDLE CORES WITHIN AN HOUR OF BIOPSY

3.1 BACKGROUND AND MOTIVATION

Prostate cancer (PCa) is the most common cancer in men and the second leading cause of cancer death for men in the United States [56]. Approximately 268,490 new cases of prostate cancer will be diagnosed in 2022, and 34,500 men will die of prostate cancer in the United States [99]. To diagnose prostate cancer, after a 12-biopsy set is taken from a patient, the standard histopathology procedure (**Figure 2.2**) typically takes at least a day. There is a high degree of anxiety for most patients while awaiting the diagnosis [100, 101]. There are faster procedures, such as "rush" processing of biopsies in a 4-hour, labor-intensive process [102]. Alternatively, frozen sections of each biopsy can be made and diagnosed, a procedure that could take up to 2 hours. Both procedures can compromise the quality of histology and, consequently, diagnostic accuracy [84, 85, 102]. In addition, these procedures, which involve cutting thin sections of tissue, consume irreplaceable biopsy material, compromising its value for subsequent molecular assays.

To address these shortcomings, we developed a novel nondestructive tissue-processing and 3D microscopy method to enable pathologists to rapidly diagnose a set of 12 prostate needle cores. This time-efficient workflow, which we refer to as one-hour-to-diagnosis (1Hr2Dx), is nondestructive, making the tissues fully available for downstream histology and molecular assays (e.g., IHC, sequencing). Furthermore, 1Hr2Dx provides volumetric imaging data, therefore increasing tissue sampling, with potential advantages over conventional "gold-standard" histology. This study is an assessment of the ability of pathologists to diagnose a set of 12 needle core biopsies within an hour of receipt with a high diagnostic accuracy.

3.2 THE 1HR2DX WORKFLOW FACILITATED BY AN OPEN-TOP LIGHT-SHEET MICROSCOPE

3.2.1 *The pipeline procedure*

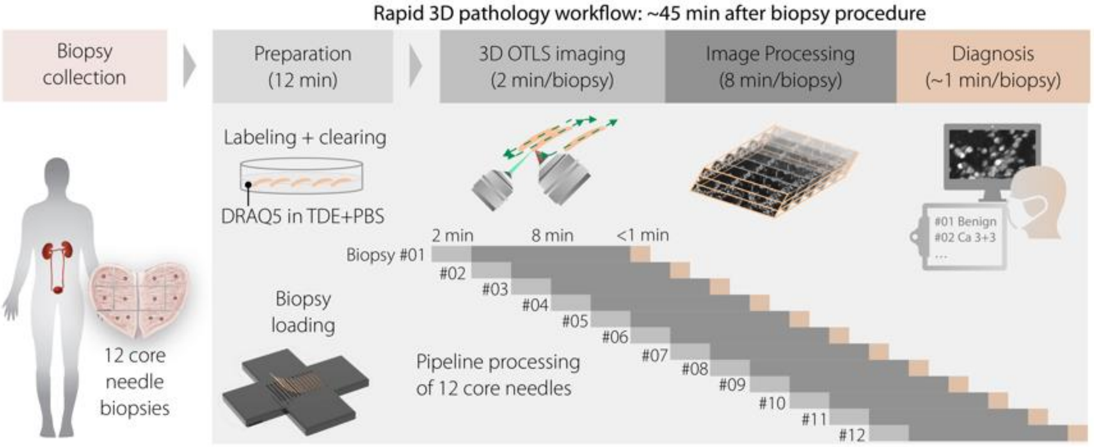

Figure 3.1. Sequence of the 1Hr2Dx workflow. The 12 biopsies are simultaneously cleared and labeled with the nucleic-acid-targeted fluorophore, DRAQ5. The biopsies are then loaded onto a custom 12-biopsy sample holder and imaged in 3D with an open-top light-sheet (OTLS) microscope to a depth of ~80 microns. Imaging (2 min/biopsy) and image processing (8 min/biopsy) of each biopsy occurs sequentially. The pathologist views images of each biopsy as the datasets become available.

As shown in **Figure 3.1**, We developed a pipeline workflow to process and image 12 biopsies as follows. After removal from a patient, the 12-biopsy set were incubated in a solution of 0.1 mM nucleic-acid binding fluorophore DRAQ5 (ThermoFisher, catalog No. 62251) in a mixture of 2,2'-Thiodiethanol (TDE, Sigma-Aldrich, CAS Number 111-48-8) and PBS buffer at room temperature for 5 min. The TDE/PBS ratio was adjusted to achieve a refractive index of 1.46, which matches that of the fused silica plate used as the sample stage holder on our open-top light-sheet (OTLS) microscope [30]. This 5-min staining/clearing procedure provided sufficient optical clearing to allow us to image the biopsies to a depth of >100 µm. After staining/clearing, the biopsies were rinsed 3X in the TDE+PBS buffer (30 s per rinse) and imaged by OTLS microscopy. To minimize turnaround time, we clarified and labeled all 12 biopsies simultaneously and loaded them onto the microscope using a custom-made 12-biopsy holder. Sets of 12 biopsies were imaged and diagnosed in a pipeline manner. In brief, as each biopsy was imaged and became available as a 3D dataset, a pathologist could immediately view and diagnose the biopsy, panning through the range of depths at different magnifications. Subsequent biopsies were continuously imaged and made available for diagnosis.

3.2.2 *Open-top light-sheet microscopy for rapid imaging of a 12-biopsy set*

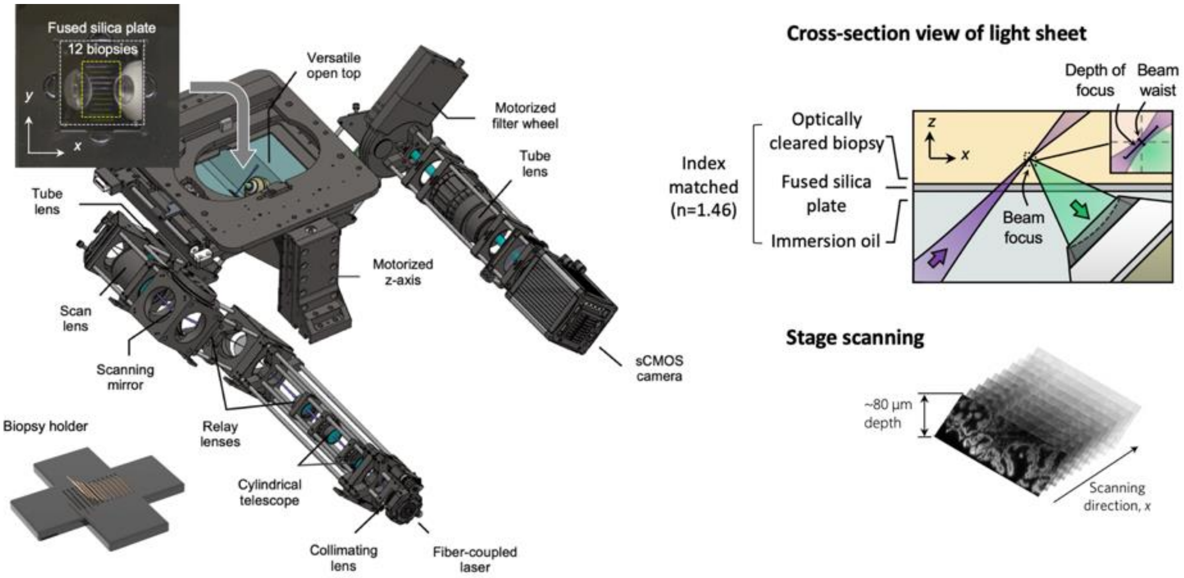

Figure 3.2. Open-top light-sheet microscopy for rapid imaging of biopsies. 12 biopsies can be loaded on the sample plate using a biopsy holder to be automated imaged in a single imaging session. The off-axis illumination light sheet and collected fluorescence travel obliquely through the immersion oil, fused silica plate, and biopsies. Aberrations are minimized by precisely matching the refractive index of all three materials. Lateral stage scanning is used to scan a 3D volume at the surface of tissue.

An in-house multi-immersion open-top light-sheet (OTLS) microscope [30] was used for rapid imaging of biopsies in this study (**Figure 3.2**). In this study, we load 12 biopsies onto the fused silica plate using a custom-designed biopsy holder and image all of them within one automated imaging session. As seen in the cross-section view of the light sheet, the objective immersion media and the biopsies are index matched with the fused-silica sample plate to reduce aberrations. The illumination comes from below the sample plate at a 45-degree angle and creates a thin light-sheet into the optically cleared tissue. The collection path is perpendicular to the illumination and an 2D image of the tissue optically sectioned by the light sheet is collected onto a sCMOS camera chip.

The DRAQ5 nuclear label was excited using a laser line at 638 nm. The OTLS system creates 3D datasets by collecting a series of closely spaced 2D grayscale, fluorescence images and post-processing them into a seamless 3D microscopy dataset. The light sheet is stationary in space, and the samples are "stage scanned" through the light sheet to collect the adjacent light-sheet images. Individual 2D light-sheet image frames were acquired with a lateral field of view of ~0.9 mm at a near-Nyquist sampling pitch of ~0.45 μm/pixel (the lateral resolution was ~0.9 μm). The vertical field of view was set to ~80 μm to match the depth of focus of the illumination light sheet, which was captured in a 2048 × 256-pixel region of the camera. At this size, the camera operated at 800 frames per second. For this study, the camera was operated with 2X binning to decrease the imaging time and dataset size, resulting in a sampling pitch of ~0.9 μm/pixel corresponding to a resolution of ~1.8 μm. The spacing between adjacent light-sheet images was set to be 0.9 μm by adjusting the stage-scanning speed. This level of resolution roughly corresponds to a standard 10× objective on a widefield light microscope.

We sharpened raw OTLS images with a Python-based edge-enhancement algorithm. The images were re-saved as multi-resolution XML/HDF5 files, which enabled them to be quickly opened and viewed at various levels of magnifications using the ImageJ BigStitcher / BigDataViewer plug-in [103]. Pathologists could pan through different depth levels within the 3D datasets (within a range of 0 to ~80 μm).

Images of three ~2-cm long biopsies (the tissue obtaining and processing before 1Hr2Dx procedure is described in section 3.3) produced by our 1Hr2Dx platform are shown in **Figure 3.3**, demonstrating that structures of diagnostic importance are clearly seen.

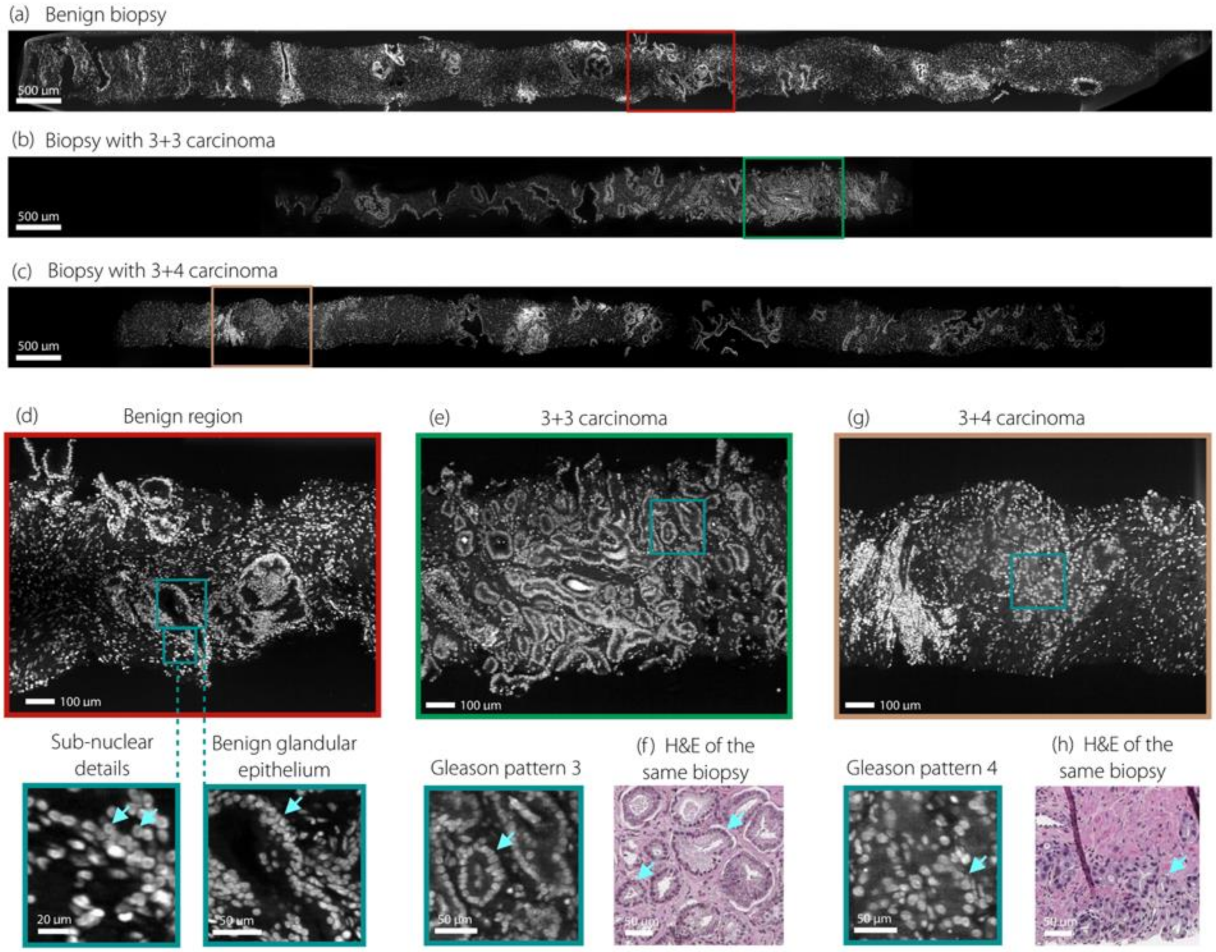

Figure 3.3. Representative 1Hr2Dx images of three biopsies showing regions of benign prostate tissue (a), Gleason score 3+3 adenocarcinoma (b), and Gleason score 3+4 adenocarcinoma (c). Higher-magnification views of these three biopsies show benign glands (d) with the bilayer of epithelial cells that is characteristic of benign glands (arrow), well-formed cancer glands (e) that are characteristic of Gleason pattern 3 carcinoma (arrow), and fused glands (g) that are characteristic of one variant of Gleason pattern 4 carcinoma (arrow). H&E images of Gleason pattern 3 carcinoma (f) and Gleason pattern 4 carcinoma (h) are paired with corresponding 1Hr2Dx images.

To illustrate the importance of increased sampling, we illustrate (**Figure 3.4**) a prostate biopsy with Gleason score 3+3 adenocarcinoma. Note that the small focus of carcinoma is only seen clearly in the image plane at a depth of 70 µm whereas the other image depths reveal only benign glands. Thus, this case would have resulted in a false-negative diagnosis if only a few 2D histology sections were viewed.

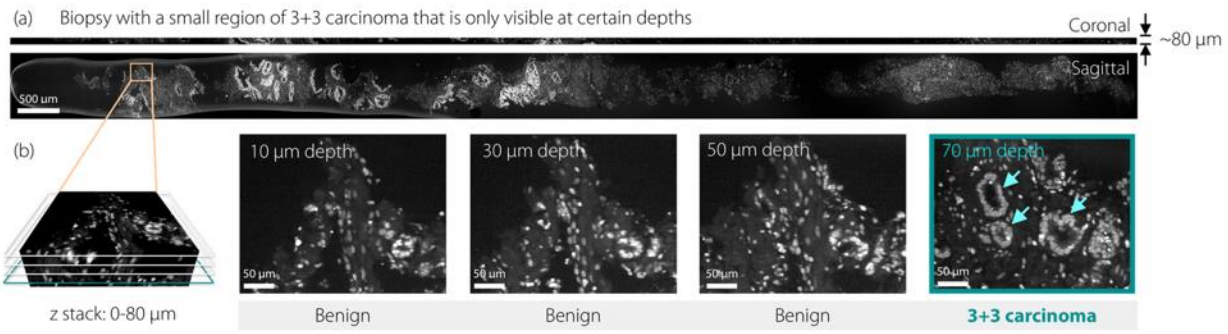

Figure 3.4. 1Hr2Dx images of a biopsy in which a focus of adenocarcinoma is seen only at the 70-µm image level (arrow).

3.3 ASSESSMENT OF ACCURACY IN PATHOLOGICAL DIAGNOSIS USING THE 1HR2DX WORKFLOW

We obtained 44 needle core biopsy specimens which had been collected *ex vivo* from 7 radical prostatectomies. For the studies, these biopsies were fixed in an aqueous solution of phosphate buffered formalin at room temperature for 24 hours, prior to commencing the 1Hr2Dx procedure. After imaging by 1Hr2Dx, all biopsies were embedded in paraffin (FFPE) for standard slide-based H&E histology. Immunohistochemical stains for cytokeratin 5 and cytokeratin 8 were also performed on sections of the biopsies.

Three genitourinary (GU) pathologists independently reviewed both the OTLS-generated gray-scale fluorescence images (nuclear staining only) and the H&E images from the FFPE tissue in a randomized order, categorizing each core as “cancer”, “benign”, or “atypia/suspicious for cancer”. Different identifiers were used for the two histology modalities to mask correspondence between the 1Hr2Dx and H&E images. The total turnaround time (staining/clearing, imaging, post-processing, and diagnosis) was recorded for each biopsy, and for each set of 12 biopsies. To establish a ground truth for biopsies that were diagnosed as “atypia/suspicious for cancer,” we immunohistochemically stained these biopsies with anti-keratin antibodies. Those atypical glands that lacked a basal cell layer (evidenced by absence of antibodies to cytokeratin 5, which is a marker for basal cells that is absent in prostate cancer) were categorized as “cancer”, per standard pathology practice [104].

To calculate the true positive, true negative, false positive and false negative values for the 1Hr2Dx diagnoses, we regarded the diagnoses of H&E-stained sections (conventionally histology) as “ground truth” based on the majority opinion of the three pathologists involved in this study. For any case that had been diagnosed by at least 1 pathologist as “atypia/suspicious for cancer” based on H&E histology, immunohistochemical stains established the ground truth. Sensitivity, specificity, accuracy, and positive- and negative-predictive values were determined for each pathologist as well as for the majority opinion of the pathologists in their diagnosis of the 1Hr2Dx images. Inter-rater agreements for both modalities of tissue processing – 1Hr2Dx and H&E histology - were assessed by the Randolph's free-marginal multi-rater kappa [105].

Table 3.1. Performance of three pathologists in diagnosing the 1Hr2Dx images.

	Pathologist 1	Pathologist 2	Pathologist 3	**Majority diagnosis**
Accuracy	0.93	0.91	0.77	**0.93**
Sensitivity	0.90	0.60	0.90	**0.90**
Specificity	0.94	1.00	0.74	**0.94**
Positive Predictive Value	0.82	1.00	0.50	**0.82**
Negative Predictive Value	0.97	0.89	0.96	**0.97**

Ten of the 44 biopsy cores contained cancer based on conventional H&E and IHC images of FFPE tissue. In general, the 1Hr2Dx diagnoses made by the 3 pathologists agreed with diagnoses based on conventional histology; the majority-opinion diagnoses yielded >0.9 accuracy, sensitivity, specificity, and negative predictive value (**Table 3.1.**). The positive predictive value was lower at 0.82, presumably due to the increased sampling of our 3D pathology methods. For example, in one case, cancer was seen only in the 1Hr2Dx images, but not in the slide-based H&E sections. Of the 3 pathologists, pathologist 1 and pathologist 2 were more accurate in their diagnoses, having an accuracy of 0.93 and 0.91, respectively. Pathologist 3, who had less experience with our 3D pathology technologies, made more false-positive diagnoses, resulting in lower accuracy, positive predictive value, and specificity (**Table 3.1.**). We assessed the inter-rater agreements of diagnosis for each processing method (1Hr2Dx and H&E) using percentage overall agreement and Randolph's free-marginal kappa (**Table 3.2.**). Although the 3 pathologists agreed in their diagnoses of all slide-based H&E histology images, their agreement in diagnoses of 1Hr2Dx images was lower - 82% overall (0.64 kappa value).

Of the majority-opinion 1Hr2Dx diagnoses, two were “false-positive” diagnoses and one was a “false-negative” diagnosis. The three pathologists reviewed these discrepancies between OTLS and H&E images retrospectively and independently. In “false-positive” case #1, the small cluster of atypical glands originally interpreted as “cancer” was re-interpreted as atrophy. Immunohistochemical stains of this biopsy confirmed the diagnosis of “negative”. For “false-positive” case #2, volumetric imaging revealed a “suspicious for cancer” region that was not sampled by the H&E section. In other words, 1Hr2Dx provided superior sampling of the biopsy, and therefore greater sensitivity for detecting cancer than slide-based pathology. Interestingly, the “false-negative” case also resulted from a sampling difference since the range of tissue imaged by 1Hr2Dx did not appear to include the cancer region that was seen in the H&E section.

Table 3.2. Inter-rater agreement of the three pathologists assessed by percent agreement overall and by the Randolph 's free-marginal multi-rater kappa.

	OTLS	H&E
Percent overall agreement	81.82%	100.00%
Free-marginal kappa	0.64	1
95% CI for free-marginal kappa	[0.46, 0.81]	[1.00, 1.00]

The average turn-around time for the 3 pathologists to view the 1Hr2Dx images and diagnose each biopsy was 35 s (range 29 to 47 s). The median time from biopsy retrieval to diagnosis of a set of 12 needle core biopsies was 44.5 min. Based on visual examination, our protocol did not appear to compromise the quality of downstream H&E histology or IHC of formalin-fixed biopsies.

Immunoreactivity of tissue for anti-keratin antibodies did not appear to be affected by 1Hr2Dx for fixed or fresh biopsies.

3.4 DEMONSTRATION OF FEASIBILITY OF 1HR2DX FOR FRESH CORE-NEEDLE BIOPSIES

Additional fresh biopsies were also subjected to the 1Hr2Dx procedure and subsequent H&E histology to assess the robustness of the procedure with unfixed tissue. The images of fresh unfixed biopsies verified that image qualities were comparable to fixed tissue using our 1Hr2Dx method (**Figure 3.5.**). The OTLS image quality of fixed and fresh specimens was comparable.

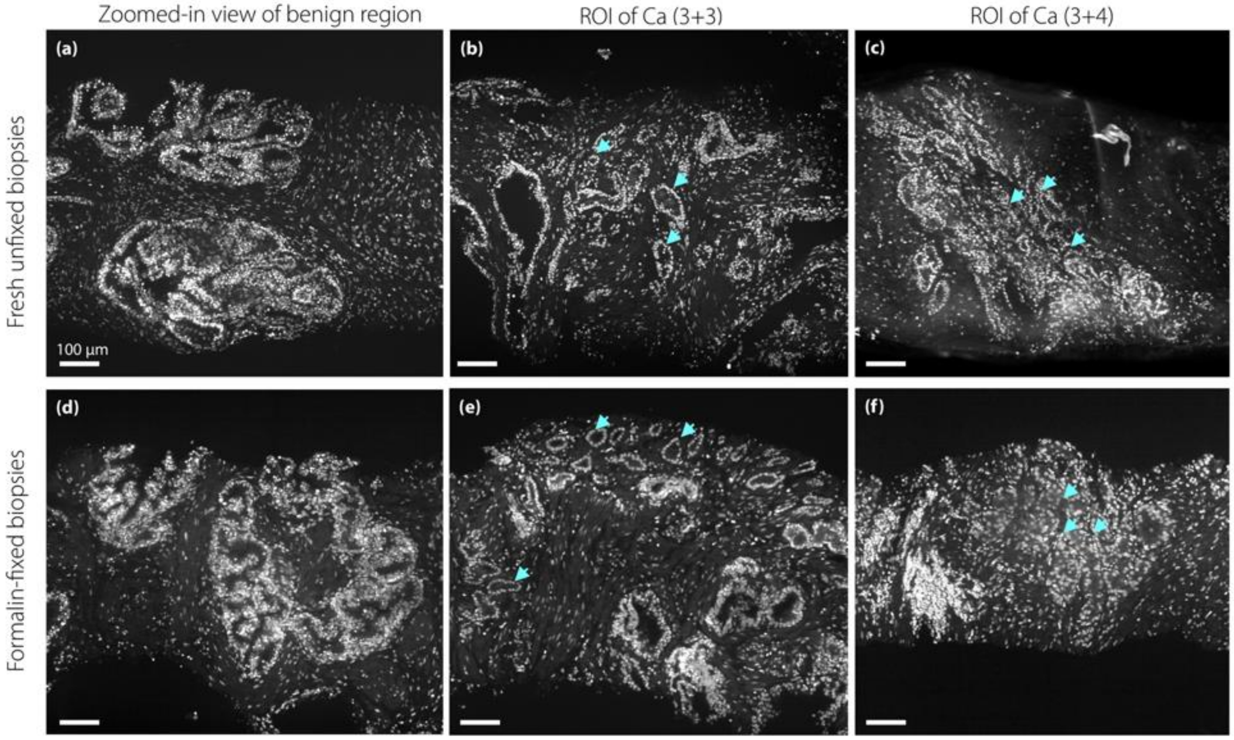

Figure 3.5. Images of fresh unfixed biopsies (a-c) with paired images of formalin-fixed biopsies (d-e). Examples are shown of benign glands (a, d), well-formed Gleason pattern 3 adenocarcinoma

(arrows, b, e), and fused Gleason pattern 4 adenocarcinoma (arrows c, f). The quality of the paired images is similar.

3.5 DISCUSSION

Providing a rapid diagnosis of prostate needle core biopsies plays an important clinical role both in quality of life for patients and for optimizing outcomes through appropriate and timely treatments. In routine pathology practice, sets of 12 prostate needle core biopsies from each patient are batch processed with all specimens received in the pathology laboratory on a given day. H&E-stained sections of these biopsies are available no sooner than the day following receipt of the biopsies. Immunohistochemical stains, which may be performed to clarify the nature of the glands, typically requires at least one additional day.

In this study, we developed and demonstrated the feasibility and accuracy of a nondestructive procedure, which relies on a novel 3D pathology platform, to diagnose a set of 12 prostate needle cores within an hour of receipt. Such a method could potentially provide patients with a preliminary on-site diagnosis after a biopsy procedure, thereby alleviating anxiety and potentially expediting treatments.

In terms of the diagnostic accuracy of 1Hr2Dx, the majority-opinion diagnosis of the 3 pathologists yielded >0.9 accuracy, sensitivity, specificity, and negative predictive value. The positive predictive value for 1Hr2Dx was 0.82, which could be attributed to the increased sampling extent of 1Hr2Dx. Certain discrepancies between pathologists were described in section 3.3, many of which could be attributed to pathologist 3 having less experience with OTLS images than pathologists 1 and 2. Note that while 3D OTLS microscopy can be used to generate H&E-like color images of tissues stained with a fluorescent analog of H&E [30, 32, 106, 107], this 1Hr2Dx study only yielded grayscale images of fluorescently labeled nuclei, for which there is a diagnostic learning curve. Note also that all 3 pathologists agreed in their diagnoses of slide-based H&E histology, for which they are all highly experienced.

Note that previous studies have reported the extent of inter-pathologist variance in diagnosing prostate cancer based on conventional slide-based histology [108-110]. In one study, two pathologists independently reviewed biopsies from 34 patients and disagreed on the presence of cancer in 31 of 407 slides [108]. Another study of 71 prostate biopsies reported similar levels of inter-pathologist variance (kappa 0.36 to 0.59) [109]. These studies, which were all based on routinely processed tissue, provide a goal and context for the performance of our platform.

Improvement can be made to our protocol and instrumentation to address several shortcomings. (1) Our 3D imaging was limited to ~80 microns below the biopsy tissue surface. Given that our OTLS microscope is capable of comprehensively imaging entire 1-mm-diameter biopsies that are well-cleared [30, 107], optimization to our staining/labeling protocols is possible to further increase the 3D sampling extent of each specimen within the one-hour workflow. We expect that this will allow us to identify more small foci of cancer that are missed when relying upon 5-micron-thick histology sections. (2) We noted that in some fresh biopsies imaged with 1Hr2Dx, some epithelial cells detached from the stroma, as can be seen in subsequent slide-based histology images (**Figure 3.6.**). Introducing a brief formalin-fixation step prior to 1Hr2Dx, or within the 1Hr2Dx procedure (so that the tissues are fixed throughout the 1Hr2Dx procedure), could prevent this artifact. (3) To reduce the learning curve for pathologists, and to facilitate clinical adoption, one option is to implement a rapid fluorescent analog of H&E [106] in conjunction with a rapid false-coloring method [111, 112] . (4) Since large 3D image datasets can be time-consuming to visualize and interpret, machine-learning techniques will be explored to segment critical structures [113, 114], such as prostate glands, and also to automatically identify a subset of localized regions of interest or ambiguity for pathologists to manually review [62]. This interesting potential for computer-aided examination of 3D dataset also motivates our effort described in Chapter 4.

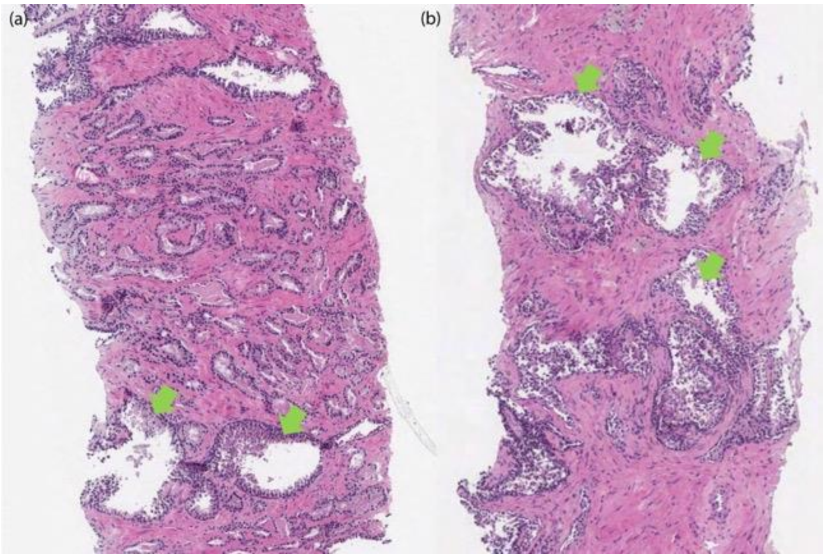

Figure 3.6. The 1Hr2Dx procedure was performed on fresh unfixed biopsies. The biopsies were subsequently fixed in formalin and processed for routine histology (H&E stain). (a-b) Conventional slide-based H&E histology images are shown of these specimens (after 1Hr2Dx). Note that some detachment of epithelial cells is seen, especially in benign glands (green arrows).

Future scaled up studies will enable us to (1) further assess the diagnostic accuracy of our enhanced methods with a broad range of pathologists, (2) fully validate our ability to process and diagnose fresh biopsies with negligible degradation in tissue quality, and (3) assess the accuracy of pathologists to rapidly grade cancer. Finally, large-scale studies will be needed in the future to determine how best to incorporate a 1Hr2Dx procedure into the urology workflow, in which patients are not currently expected to remain on site after their biopsy procedures, and where time is not allocated for an immediate follow-up urologist consultation based on the 1Hr2Dx results.

Chapter 4. PROSTATE CANCER RISK STRATIFICATION VIA NON-DESTRUCTIVE 3D PATHOLOGY WITH DEEP-LEARNING-ASSISTED GLAND SEGMENTATION AND ANALYSIS

4.1 BACKGROUND AND MOTIVATION

Currently, prostate cancer (PCa) management is largely dependent upon examination of prostate biopsies via 2D histopathology [115], 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 [108, 116] and is only moderately correlated with outcomes, especially for patients with intermediate-grade PCa [117]. This contributes to the undertreatment of patients with aggressive cancer (e.g. with active surveillance) [118], leading to preventable metastasis and death [119], and the overtreatment of patients with indolent cancer (e.g. with surgery or radiation therapy) [120], which can lead to serious side effects, such as incontinence and impotence [121].

Motivated by recent technological advances in optical clearing to render tissue specimens transparent to light (i.e. iDISCO [53], CUBIC [54], ECi clearing [55], 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 [16, 29, 30, 107, 122, 123]. 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 the while maintaining intact specimens for downstream molecular assays [19]. Like others before us [122, 124, 125], we hypothesized that 3D vs. 2D pathology datasets could allow for improved characterization of the convoluted glandular structures that 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.

4.2 APPROACHES FOR COMPUTATIONAL ANALYSIS OF 3D PATHOLOGY DATASETS

For computational analysis of 3D pathology datasets, a multi-stage pipeline (**Figure** 4.1) was chosen for classifying patient outcomes based on interpretable "hand-crafted" features [126-128] (i.e., glandular features) rather than an end-to-end deep-learning (DL) strategy for risk classification based directly on the imaging data [129-131]. 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 [19, 132], 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 [60, 133].

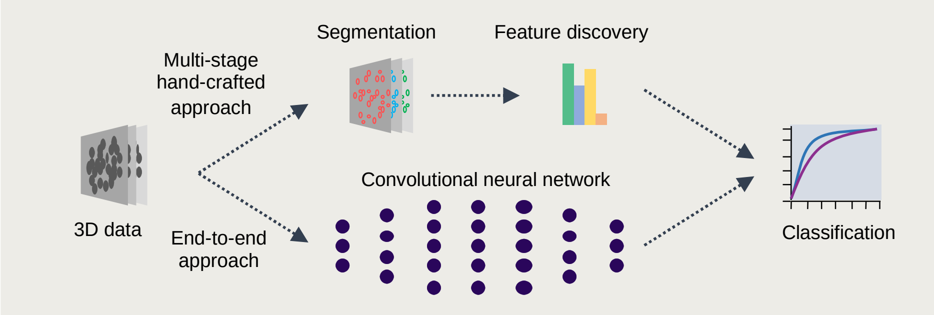

Figure 4.1. Different machine-learning-based analysis paths for clinical decision support. We adopted a multi-stage hand-crafted approach (on the top row) by first segmenting the tissue structure (i.e., prostate glands, etc.), then analyzing tissue features from the segmentation mask to train a decision-support classification model. An alternative end-to-end approach (on the bottom row) can be achieved when a large and reliable clinical dataset is available to train a single-step convolutional neural network to directly predict clinical outcome, etc. [19]

Multi-stage feature-based classification approaches rely on the accurate segmentation of morphological structures such as nuclei or subnuclear features [134-136], collagen fibers [137, 138], vessels [123, 139], or in our case, prostate glands [126, 140]. This is typically achieved in one of two ways: (1) Direct DL-based segmentation methods [141-144] that require manually annotated training datasets, which are especially tedious and difficult to obtain in 3D (**Figure** 4.2**a**) [145]; or (2) traditional computer vision (CV) approaches based on intensity of stains and morphology, provided that tissue structures of interest can be stained/labeled with high specificity (**Figure** 4.2**b**) [124, 146, 147]. 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 large tissue volumes, and the slow diffusion times of antibodies in thick tissues (up to several weeks) [53, 148].

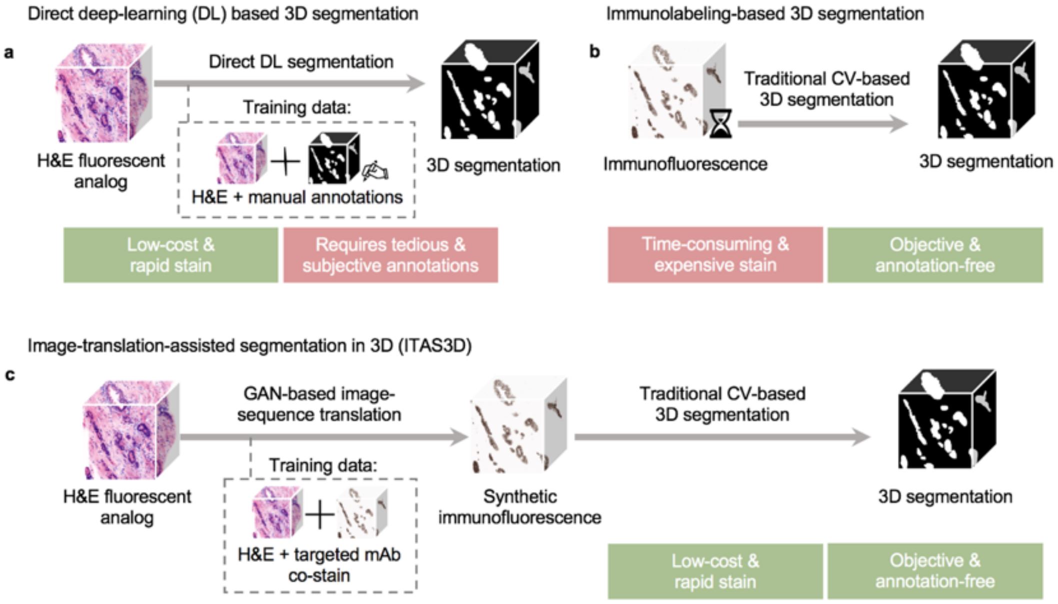

Figure 4.2. General methods for 3D gland segmentation. (a) A single-step deep learning (DL) segmentation model can be trained with imaging datasets of tissues labeled with a fluorescent analog of H&E paired with manually annotated ground-truth segmentation masks. While H&E-analog staining is low-cost 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).

To address these challenges, we developed a generalizable annotation-free 3D segmentation method, hereafter referred to as "image-translation-assisted segmentation in 3D (ITAS3D)" [74]. In our specific implementation of ITAS3D (**Figure** 4.2**c**), 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 cytokeratin 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 [72, 73, 149]. 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 [150].

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., 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.

4.3 EXISTING 2D AND 3D HISTOLOGY SEGMENTATION METHODS

Quantifying the 3D microarchitecture of prostate glands relies upon accurate segmentation of the glandular network from 3D pathology volumes. Existing methods for gland segmentation on H&E histology images usually start off with a color deconvolution or color-value-based feature detection on the pink-and-purple H&E image to separate out the grayscale nuclei and cytoplasm, and then use various morphology-based method accompanied with machine learning techniques to achieve segmentation [114, 151, 152]. End-to-end deep learning models has also been proposed for the same purpose [153-155]. However, existing glandular segmentation methods for H&E histology are based on 2D images. Directly applying 2D segmentation level-by-level on 3D H&E analog images doesn't provide ideal segmentation results due to discontinuity between adjacent 2D frames.

3D deep-learning biomedical segmentation algorithms, such as 3D U-net for confocal images [155], V-net for MRI images [155], a coarse-to-fine 3D segmentation strategy for CT images [144], etc. [141, 142], were developed for various clinical applications. These methods can potentially be applied to 3D pathology images. However, they either require tedious 3D ground-truth annotations that are challenging to obtain for our custom 3D dataset or can only be applied on small tissue volumes due to GPU memory limitations. The limited tissue volume that can be processed at a time can potentially cause severe 3D inconsistency in adjacent small data volumes that are difficult to mitigate when dealing with large 3D data, as is in our case.

4.4 GENERATIVE ADVERSARIAL NETWORKS FOR IMAGE-TO-IMAGE TRANSLATION

To achieve the annotation-free two-step gland segmentation, the first step is to perform a 3D image translation from H&E fluorescent analog stain to cytokeratin 8 (CK8) immunostaining. CK8 stains luminal epithelial cells that line both cancer cells and benign glands. Generative adversarial networks (GANs) [156] provide a generalizable model and loss function to solve different image-translation problems that traditionally would require the design of specialized models and loss functions by practitioners with extensive field-specific expertise. In GANs, the discriminator's goal is to classify images as real or fake (generated), while the generator is trained to synthesize realistic output data that can "fool" the discriminator. In game theory, GANs converges when the discriminator and the generator reach a Nash equilibrium, which is when neither player will not change its action regardless of what the opponent may do, and neither the discriminator nor the generator can improve without a change in the other. Once this adversarial training is done, the generator is taken as the trained model for inference.

As an extension of the GAN structure, conditional GANs (cGANs) are structured such that both the generator and discriminator are conditioned with prior information (in our case, H&E-analog inputs), such that the model can learn how to map inputs (H&E analog) to outputs (synthetic CK8) in a supervised manner. As compared to single-network generative models, such as U-Net [154] and variational autoencoders (VAEs) [157], generative adversarial networks (GAN) generally promote the detailed accuracy and realistic look in image translation by introducing a second network (i.e., discriminator) that is trained against the generative model.

For our task, we treat the 3D image as a z-stack of 2D images, and the initial approach is to perform an image-to-image translation level-by-level. For 2D image translation, *pix2pix* [158] is a conditional generative adversarial network that is trained in a supervised manner (**Figure 4.3**).

This framework directly conditions the generator with input images that are paired with target images. The generator uses a U-Net [154], which uses an encoder-decoder structure with skip connections between mirrored layers in the encoder and decoder stacks for short-cut passing of lower-level details during training. In our implementation, we used a 9-layer U-Net to suit our 1024 px × 1024 px sized images. The discriminator uses a convolutional PatchGAN [159] classifier, which only penalizes structure at the scale of image patches to capture local style statistics. This GAN architecture learns not only the mapping from input image to output image, but also a loss function to train this mapping (through the adversarial training of a discriminator against the generator). During training, the generator tries to minimize a conditional GAN loss against an adversarial discriminator that tries to maximize it:

$$\begin{aligned}\mathcal{L}_{cGAN}(G,D) =&\mathbb{E}_{x,y}[\log D(x,y)]+\\ &\mathbb{E}_{x,z}[\log(1-D(x,G(x,z))]\end{aligned} \tag{4.1}$$

where *x* is input, *y* is output, and *z* is random nose. Meanwhile, the generator is also trying to minimize an L1 distance loss, which measures the pixel-wise error between the generated image and the underlying ground-truth image:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x,z)\|_1]. \tag{4.2}$$

Therefore, the final objective for the *pix2pix* training is:

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_{L1}(G) \tag{4.3}$$

The *pix2pix* model was first introduced for semantic-to-photo or pose-to-photo tasks but are also reported to be useful for biomedical images by many recent studies [69, 70]. Additionally, *pix2pixHD* [160] is a follow-up study on *pix2pix* model but added a course-to-fine training strategy to refine the photo-realistic details in the resultant images.

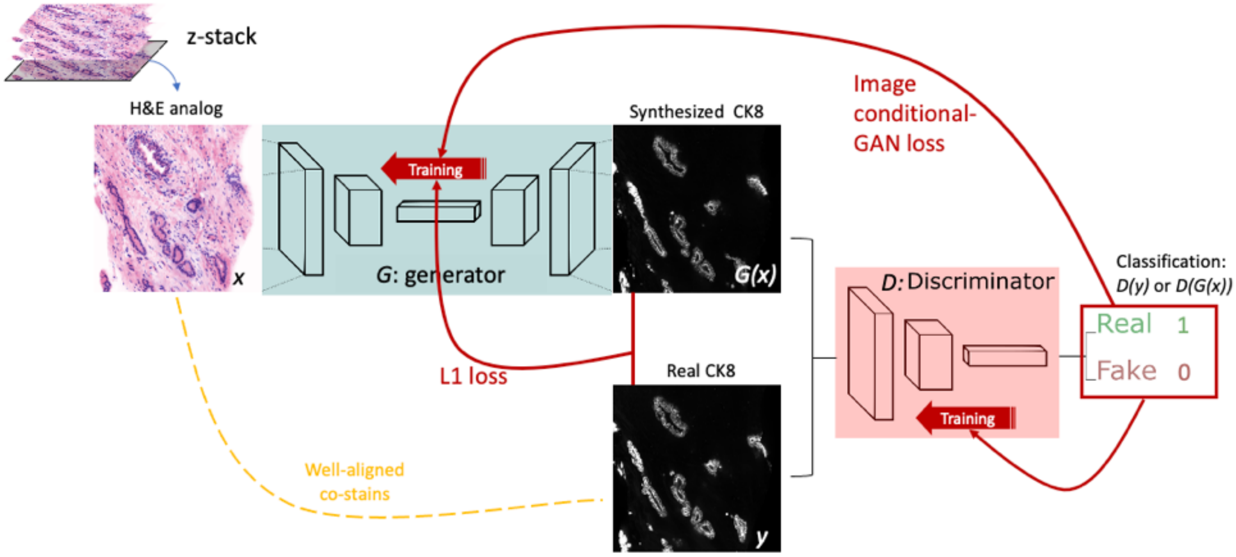

Figure 4.3. Schematic of generative adversarial training for the *pix2pix* model

The problems with performing the image translation level-by-level in a z-stack is the discontinuity across levels. This drawback also increases the difficulty in the successive 3D segmentation step by not providing an accurate 3D synthetic immunofluorescence image. Therefore, we look for a generative solution in 3D. Recent studies have demonstrated 3D deep neural networks to output 3D images [161-164], but a major obstacle is the “out of memory” problem when training an 3D network. For computational analyses, the maximum allowable size of a 3D data cube (volumetric sub region) is typically limited by GPU memory. For instance, with the 3D image-translation method, *vox2vox* [165], which uses a 3D U-Net [143] structure as its generator, a 3D block size that pushes against the computational limits of a NVIDIA GeForce RTX 2080 Ti GPU during the inference phase is ~ 160 × 190 × 130 pixels. With the OTLS microscope used in this study, this corresponds to a physical tissue volume of about 140 × 170 × 110 μm^3. This small block size, which results in many prostate glands being truncated, could lead to errors (e.g., edge and stitching errors) when performing 3D gland segmentations. A common strategy to reduce stitching errors is to include a larger amount of overlap between neighboring sub-volumes, but this would further increase computational times.

As a rough estimate based on our experiences with ITAS3D and published results for *vox2vox* and assuming a 25% overlap ratio between image blocks processed with a 12-GB GPU, approximately 30% of the tissue volume in each biopsy would be redundantly processed using a 2.5D image-translation method (e.g., ITAS3D), whereas more than 100% of each biopsy volume would be redundantly processed using 3D image translation (*vox2vox*) due to the overlap between sub-volumes. This would contribute to increased computational times for a 3D image-translation strategy.

We therefore propose to adapt a video-to-video synthesis model for our 3D images as a workaround to this problem. In general, video translation performs the image conversion frame-by-frame but also takes temporal information in adjacent frames to ensure the temporal continuity. Related works include *vid2vid* [150], *few-shots vid2vid* [166], and several other motion and style translation models [167-169]. We applied this concept to a z-stack in 3D image rather than a t-stack in videos. Although the translation is not done in a full 3D manner, but such a “2.5D translation” will allow us to transfer a 3D image of large volume in its entirety while fitting in GPU memory. In another word, due to the recursive 2D translation in video synthesis models, one could perform a 3D image translation with a certain field of view (significantly larger than 3D models) for a depth in a continuous manner, and this significantly increases the processed volume size limited by GPU memory. We adapted the *vid2vid* framework for our 3D H&E-to-CK8 translation and omitted the coarse-to-fine training strategy in the original paper. As shown in **Figure 4.4**, the sequential generator in the *vid2vid* model synthesizes each CK8 frame through a z-stack iteratively, and each synthesized image depends on a few generated CK8 frames and original H&E analog frames at previous depths as well as the H&E analog frame at the current depth. In this way, the generator utilizes information in previous frames (including an "optical flow" [170] of tissue microstructure changing across levels) to enforce the continuity in depth. Like a basic *pix2pix* model, an image discriminator is trained simultaneously to tell apart real or synthesized current frame to enhance the realistic of the generated image. Meanwhile, a sequence discriminator is trained to ensure a similar "optical flow" in the latest segment of the generated sequence. The generator and the two discriminators are trained in an adversarial fashion.

To be more specific, the sequence generator model consists of two parts. The two-part structure relies on the following assumption: if the optical flow [170] between consecutive levels is known, we can estimate the next level by warping the current level. The Sequence generator F is:

$$F(\tilde{\mathbf{x}}_{t-L}^{t-1}, \mathbf{s}_{t-L}^{t}) = (\mathbf{1} - \tilde{\mathbf{m}}_t) \odot \tilde{\mathbf{w}}_{t-1}(\tilde{\mathbf{x}}_{t-1}) + \tilde{\mathbf{m}}_t \odot \tilde{\mathbf{h}}_t \tag{4.4}$$

Where $\odot$ is the element-wise product operator. The first part corresponds to pixels warped from the previous frame, with $\widetilde{w}_{t-1}$ being the optical flow that is estimated based on the previous levels by a separate pre-trained model, while the second part hallucinates new pixels. The pixel-wise coefficient $\widetilde{m}_t$ is the occlusion mask with continuous value between 0 and 1. By applying this soft mask, the model allows for adding details by gradually blending warped pixels and the newly synthesized pixels. Residual networks were used for constructing $\widetilde{w}_{t-1}$, $\tilde{h}_t$, and $\widetilde{m}_t$ from current level H&E image and synthetic CK8 image, as well as H&E images from a few previous levels.

As mentioned previously, in this image sequence translation architecture, there is also a conditional image discriminator D_I, which should output 1 or 0 for a true or fake pair of H&E and

CK8, and a conditional video discriminator D_v, which should output 1 or 0 for a true or fake pair of optical flow and CK8. The final learning objective function is then given by:

$$\min_{F}\left(\max_{D_I}\mathcal{L}_I(F,D_I)+\max_{D_V}\mathcal{L}_V(F,D_V)\right)+\lambda_W\mathcal{L}_W(F) \tag{4.5}$$

where $\mathcal{L}_1$ is the GAN loss on images defined by the conditional image discriminator D_I, $\mathcal{L}_\mathrm{v}$ is the GAN loss on K consecutive levels defined by D_v, and $\mathcal{L}_\mathrm{w}$ is the flow estimation loss. The image-conditional GAN loss $\mathcal{L}_1$ is given as:

$$E_{\phi_I(\mathbf{x}_1^T,\mathbf{s}_1^T)}[\log D_I(\mathbf{x}_i,\mathbf{s}_i)]+E_{\phi_I(\tilde{\mathbf{x}}_1^T,\mathbf{s}_1^T)}[\log(1-D_I(\tilde{\mathbf{x}}_i,\mathbf{s}_i))] \tag{4.6}$$

where an image sampling operator ϕ_I is used to randomly select a pair of H&E-and-CK8 image from the stack. Similarly, the video GAN loss $\mathcal{L}_\mathrm{v}$ is given by:

$$E_{\phi_V(\mathbf{w}_1^{T-1},\mathbf{x}_1^T,\mathbf{s}_1^T)}[\log D_V(\mathbf{x}_{i-K}^{i-1},\mathbf{w}_{i-K}^{i-2})]+E_{\phi_V(\mathbf{w}_1^{T-1},\tilde{\mathbf{x}}_1^T,\mathbf{s}_1^T)}[\log(1-D_V(\tilde{\mathbf{x}}_{i-K}^{i-1},\mathbf{w}_{i-K}^{i-2}))] \tag{4.7}$$

where a sequence sampling operator ϕ_v is used to randomly retrieve *K* consecutive levels and the corresponding K-1 optical flow estimation images. Lastly, the flow loss $\mathcal{L}_\mathrm{w}$ includes two parts: the endpoint error between the ground truth images and the estimated flow, and the warping loss when the flow warps the previous level to the next level, which is given by:

$$\mathcal{L}_W=\frac{1}{T-1}\sum_{t=1}^{T-1}\left(||\tilde{\mathbf{w}}_t-\mathbf{w}_t\|_1+\|\tilde{\mathbf{w}}_t(\mathbf{x}_t)-\mathbf{x}_{t+1}\|_1\right) \tag{4.8}$$

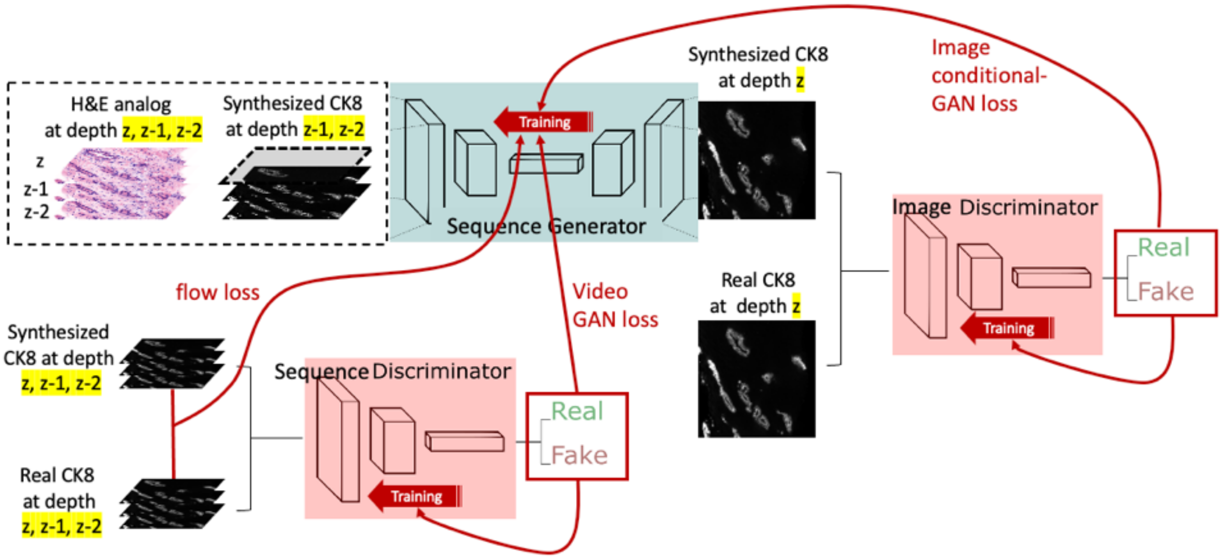

Figure 4.4. Schematic of generative adversarial training of the *vid2vid* model

4.5 AN ANNOTATION-FREE 3D GLAND SEGMENTATION STRATEGY: IMAGE-TRANSLATION-ASSISTED SEGMENTATION IN 3D (ITAS3D)

To segment the 3D glandular network within prostate biopsies, we first trained a GAN-based image-sequence translation model to convert 3D H&E-analog images into synthetic CK8 IF images, which can be false colored to resemble chromogenic immunohistochemistry (IHC) (**Figure 4.5**). 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).

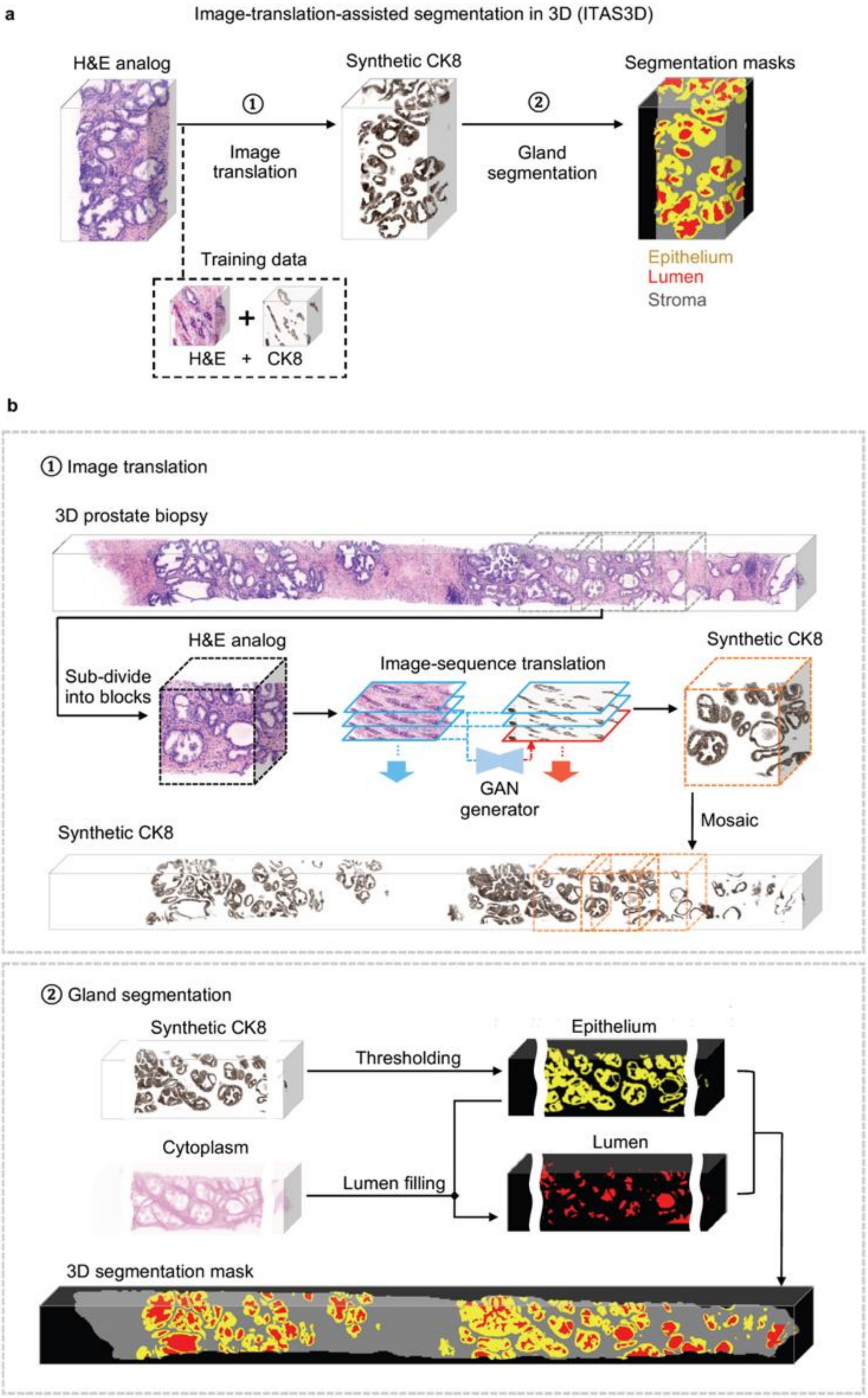

Figure 4.5. Image-translation-assisted segmentation in 3D (ITAS3D): a two-step pipeline for annotation-free 3D segmentation of prostate glands. (a) In step 1, a 3D microscopy dataset of a specimen, stained with a rapid and inexpensive fluorescent analog of H&E, is converted into a

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).

4.6 PREPARATION OF THE DATASETS TO TRAIN AND VALIDATE THE IMAGE TRANSLATION MODELS

4.6.1 *Collection and processing of prostate tissue to obtain training data*

To train an image-sequence translation model, we collected one FFPE block from each of nine radical prostatectomy (RP) specimens archived in an IRB-approved genitourinary biorepository at the University of Washington (UW). Based on the original pathology reports generated from the RP specimens, five specimens were from Gleason Grade Group 1 (GS = 3+3) and four specimens from Gleason Grade Group 2 or 3 (GS = 3+4 and 4+3). The imaging data from this cohort allowed us to train an image-translation model for low- to intermediate-risk PCa (Grade Group 1-3).

For deparaffinization, the FFPE blocks were first heated at 75°C for 1 hour until the outer paraffin wax was melted. The tissue blocks were then treated 2× with 500 ml of 75°C xylene for 24 hours. A hotplate with magnetic stirrer was used to maintain the temperature of the xylene and to promote fluid convection around the specimens. Next, we used a vibratome to cut a 200-µm-thick slice from the surface of each tissue block. This thickness was optimized to balance two factors: 1) providing sufficient 3D context to train and test the image-sequence translation model and 2) allowing for uniform antibody diffusion and staining within the tissue slices. The tissue slices averaged about 1.5 cm × 1 cm in their lateral dimension. Prior to staining, each tissue slice was cut into smaller pieces, measuring approximately 0.5 cm × 0.5 cm × 200 µm, to further promote fluid convection around all specimens during the staining protocol.

4.6.2 *Tissue tri-labeling for generating training datasets*

We performed a tri-labeling protocol (see **Appendix A**) that utilized SYTO™ 85 (Cat: S11366, ThermoFisher) as a nuclear stain (hematoxylin analog), Alexa Fluor™ 488 NHS Ester (Cat: A20000, ThermoFisher) as a cytoplasmic stain (eosin analog) [171], and a CK8-targeted monoclonal antibody (Cat: MA5-14088, ThermoFisher) for immunofluorescence labeling. The arrangement of the three imaging channels considering the excitation and emission spectra is shown in **Figure 4.6**. Our tri-labeling protocol was adapted from the iDISCO protocol [53] with Ethyl Cinnamate clearing [55]. We first labeled the tissue with an Alexa Fluor™ 488 NHS ester (binds to all proteins) prior to performing CK8 immunostaining to prevent the NHS ester from staining the CK8 antibody. This ensures unbiased training of the H&E-to-CK8 translation model. CK8 immunostaining included a primary and secondary antibody staining step. Nuclear staining with SYTO™ 85 was performed afterwards, followed by tissue dehydration in ethanol and optical clearing (index-matching, n = 1.56) with ethyl cinnamate (Cat: 112372, Sigma-Aldrich).

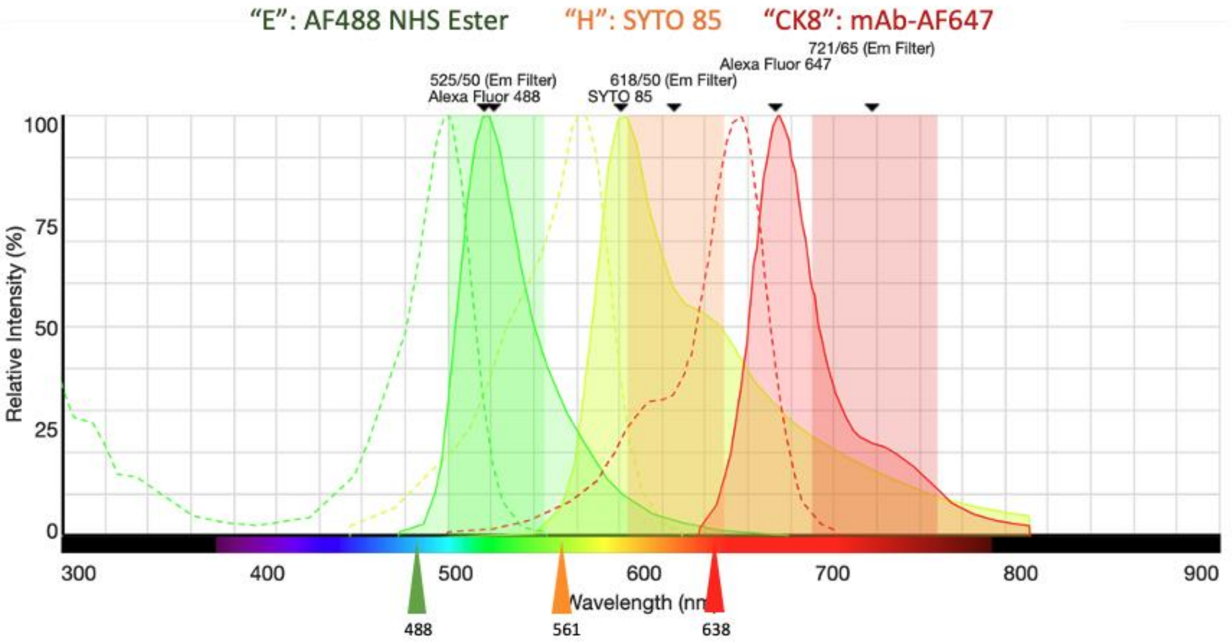

Figure 4.6. Tri-labeling spectra

The fluorescent analog of H&E used for generating training datasets (i.e., SYTO™ 85 + Alexa Fluor™ 488 NHS Ester, "S&N") was slightly different from the H&E analog used for whole-biopsy staining in our inference datasets (i.e., TO-PRO-3 + eosin, "T&E"). The S&N version of our H&E analog was used for generating the training data so that the CK8 immunofluorescence could be placed in the longest-wavelength channel. This was done to ensure that there was negligible crosstalk of the CK8 immunofluorescence into the H&E-analog wavelength channels (for unbiased training). Despite this difference in the H&E-analog staining protocols, our trained model was shown to be applicable to both S&N- and T&E-labeled tissues as inputs. High-fidelity image translation was achieved in both cases with minimal differences in appearance (**Figure 4.7**).

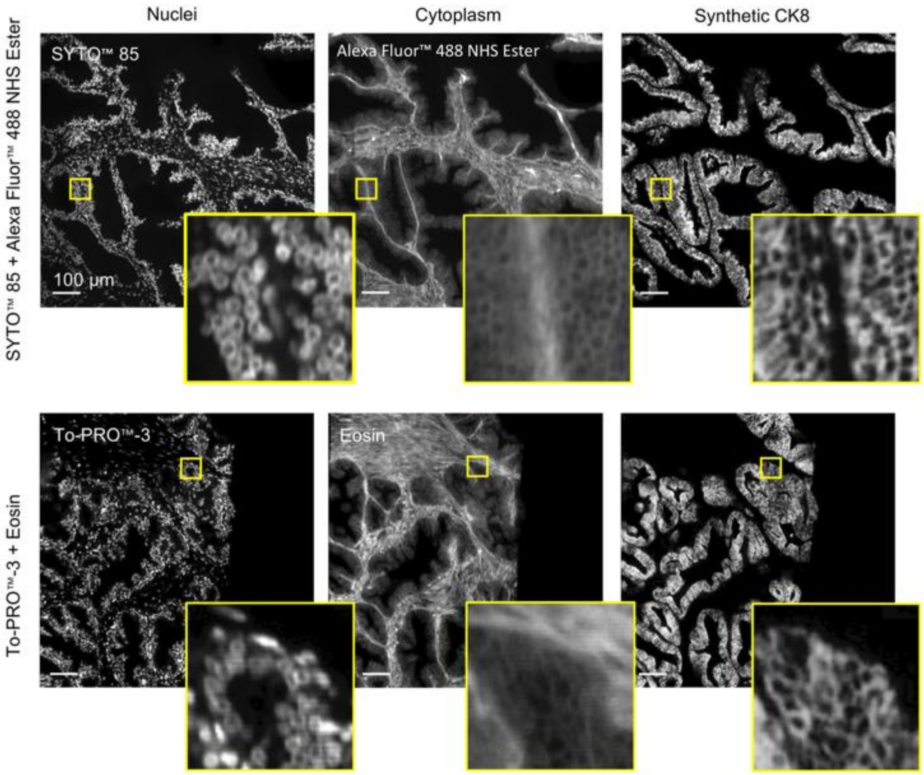

Figure 4.7. Use of a different fluorescent analog of H&E for model training and inference. Although our image-sequence translation model is trained with a H&E analog that is slightly different from the H&E analog used for our inference datasets, the trained model performs comparably well in both cases. The top and bottom rows show image-translation results based on tissues stained with an “S&N” protocol (used for training) versus a “T&E” protocol (used in our clinical studies). Minimal differences are seen in the synthetic-CK8 output images (far right). This is likely due to the highly similar appearance of tissue stained with the two H&E-analog protocols (as shown). Scale bar: 100 μm.

The small-molecule H&E-analog stain was highly reproducible and generated spatially uniform images [30, 107]. Therefore, data screening/cleaning was primarily based on the image quality of the CK8 immunofluorescence channel. In the training and testing datasets, we only included tissue regions with spatially uniform CK8 labeling (based on visual inspection). To further enhance the contrast and uniformity of the CK8 training data (and hence the uniformity of the image-translation outputs), contrast-limited adaptive histogram equalization (CLAHE) [172] and flat-fielding [112] were applied on the CK8 datasets. All the images were down-sampled by 2 times in all 3 dimensions from our original OTLS data, which resulted in a sampling rate around 0.9 μm/pixel.

For the training set, we obtained a total of 1,806 tri-labeled image sequences (each containing 50 levels) as described in the last section. Each level of the image sequence is 1024 px × 1024 px, covering around 1 mm^2 area in tissue. For the validation set, which was used to calculate performance metrics, we set aside 58 image sequences that were not used in the training phase. The H&E-analog channels were used as model inputs, and corresponding co-registered CK8 images was used as model targets (desired outputs) for our GAN-based supervised training strategy. The above tissue processing is shown in **Figure 4.8**.

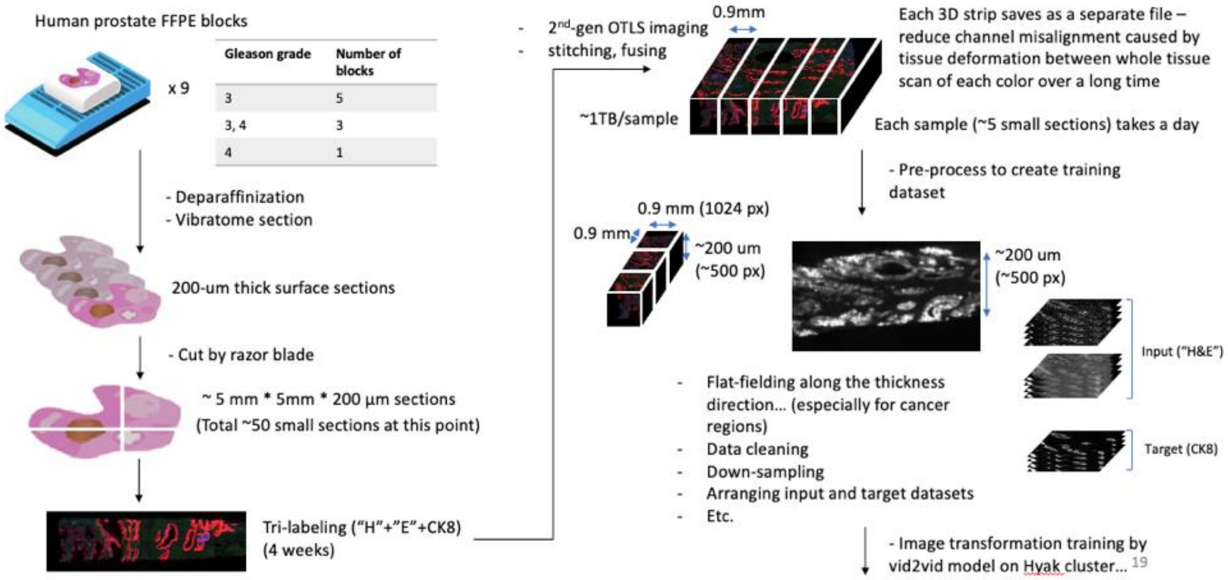

Figure 4.8. Tissue preparation for image translation training and validation data collection

4.7 OPEN-TOP LIGHT-SHEET (OTLS) MICROSCOPY AND PRE-PROCESSING

We utilized a previously developed OTLS microscope [30] 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.

For imaging the 200-µm thick tri-labeled prostate tissue sections (training specimens), the Alexa Fluor™ 488 NHS Ester was excited at 488 nm and imaged through an emission bandpass filter (FF03–525/50–25, Semrock). The nucleic-acid-targeting fluorophore, SYTO™ 85, was excited at 561 nm and imaged through another bandpass filter (FF01–618/50–25, Semrock). CK8 immunofluorescence (Alexa Fluor™ 647 conjugated secondary antibody) was excited at 638 nm and imaged through a third bandpass filter (FF01–721/65–25, Semrock). Raw OTLS images were downsampled by 2× in all three dimensions (which resulted in a sampling pitch of ~0.9 µm/pixel) and fused into a continuous 3D dataset using the BigStitcher plug-in [103] in ImageJ. This was done to alleviate memory requirements and to accelerate image-translation computations. However, this level of downsampling still preserved sufficient sub-cellular details for effective image translation and visual inspection by pathologists. From the fused 3D imaging data, we extracted image volumes measuring 1024 × 1024 × 50 pixels, which were treated as image sequences with 50 depth levels (corresponding to ~45 µm in depth). These image sub-blocks were used for training and testing of the image-translation model.

For imaging the T&E-labeled prostate biopsies, eosin was excited at 488 nm and imaged through an emission bandpass filter (FF03–525/50–25, Semrock). The nucleic-acid-targeted fluorophore, To-PRO™-3 Iodide, was excited at 638 nm laser and imaged through another bandpass filter (FF01–721/65–25, Semrock). The fused datasets were downsampled by 2× in all three dimensions using the BigStitcher plug-in [103] in ImageJ.

4.8 2.5D IMAGE-TO-IMAGE TRANSLATION FOR H&E-TO-CK8 CONVERSION

4.8.1 *Training the image-translation models*

We adapted *pix2pix* [158], a prior implementation of cGANs for single-level 2D image translation, in order to initialize our image-sequence translation workflow. The *pix2pix* framework directly conditions the generator (G_{imge} in **Figure 4.9a**) with input images that are paired with target images. Our adapted generator uses a U-Net [154] architecture with 9 layers of downsampling and up-sampling, and the discriminator (D_{image} in **Figure 4.9a**) uses a convolutional PatchGAN [159] classifier. For *pix2pix* training, from each 50-level image sequence, we selected five levels of 2D images that were 10 pixels apart (~9 µm) to maximize diversity in the training set. This yielded 9030 images in total for *pix2pix* training, each of which was 1024 × 1024 pixels in size. Each 2D image level was treated as an independent input during training. Hyperparameters were set as default values (identical to the original *pix2pix* model) except for those specified in the training script provided in our GitHub repository. The training required 200 epochs for 44 hours with a 12-GB NVIDIA Tesla P100 GPU on a standard node of the UW Hyak high-performance computation (HPC) cluster.

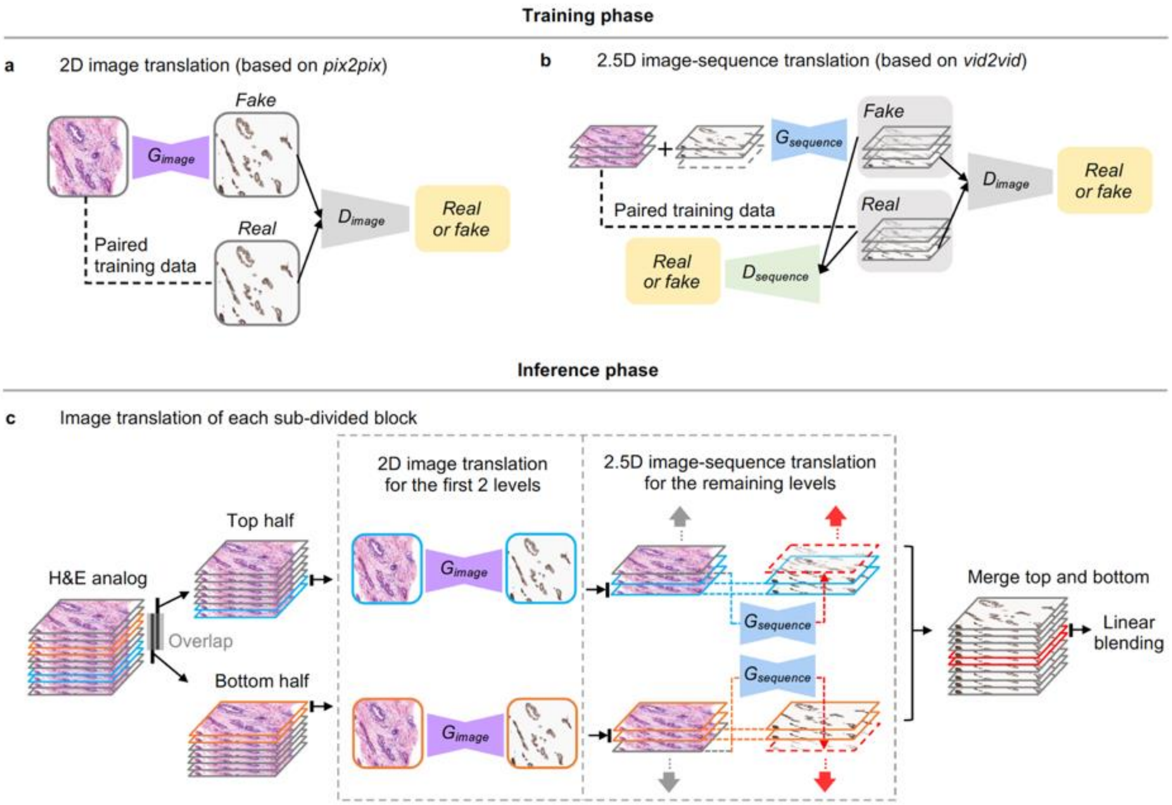

Figure 4.9. Image-sequence translation model training and inference. (a) Image translation in this study is based on previously developed conditional GANs (*pix2pix* for 2D image translation, and *vid2vid* for 2.5D image translation), which are trained with paired imaging data in a fully supervised manner. For 2D image translation, the generator (G_{image}) is trained to translate 2D H&E-analog images into 2D synthetic-CK8 images that cannot be distinguished from real CK8 images by the discriminator (D_{image}), which is adversarially trained to classify between real and fake (synthetic) CK8 images. (b) For 2.5D image translation, a 3D image is regarded as a sequence of 2D images. The generator (G_{sequence}) is trained to produce synthetic CK8 images in a level-by-level manner, conditioned with both an H&E-analog input image at each level, as well as H&E-analog and synthetic-CK8 images at two previous levels to ensure spatial continuity between adjacent images/levels within the sequence. Meanwhile, an image discriminator (D_{image}) and a sequence discriminator (D_{sequence}) are adversarially trained to classify between real or fake (synthetic) 2D images and image sequences, respectively. (c) In the inference phase, to ensure robust image translation, each H&E-analog image block is split into a top and bottom half with overlapping regions. This allows image-sequence translation to be initiated from the center of the biopsy, where image-quality is optimal and tissue-edge artifacts are avoided. For each half, image-sequence

translation is initiated with 2D image translation of the first two levels (in blue for the top half, in orange for the bottom half), which transitions to 2.5D image-sequence translation for the remaining levels. The final synthetic-CK8 dataset is obtained by merging the top and bottom halves of the block with linear blending of the overlapping middle levels (in red).

For "2.5D" image-sequence translation, we adapted a video-translation model, *vid2vid* [150], which performs image conversion frame-by-frame (originally in a "time stack") and utilizes information in adjacent frames to ensure that temporal continuity is maintained. Given that we treat a 3D image as a "z-stack" of 2D images, we can apply this same concept to improve continuity in the depth dimension. As shown in **Figure 4.9b**, the sequence generator ($G_{sequence}$) synthesizes CK8 images in a level-by-level manner by taking into account the synthetic CK8 images generated at two previous levels along with the corresponding H&E analog images at each level (i.e., the current plus two previous levels). Similar to *pix2pix*, a multi-scale PatchGAN architecture is adopted for the image discriminator (D_{image}). In addition, a multi-scale sequence discriminator ($D_{sequence}$) is trained to ensure small- and large-scale consistency in the 3D spatial domain. All 1,806 image sequences (50 levels per sequence) were used for the *vid2vid* training set.

For the *vid2vid* with course-to-fine model, we trained the network at resolution of 256×256, 512×512, 1024×1024 sequentially, each stage added down-sampling and up-sampling blocks at the front and back of the network accordingly, and trained for 20 epochs each, and took 2 months in total. For the *vid2vid* without course-to-fine model, we trained the network at full resolution (1024×1024) for 20 epochs for 1 month. All *vid2vid* training trained with a constant learning rate for the first half, and linear decrease of learning rate down to 0 in the second half of the training process. Because *vid2vid* models depend on previous generated frames, the *vid2vid* testing was done by first using the trained *pix2pix* model to generate the first 2 frames, and then continue generating the rest of frames in the sequence by applying *vid2vid* models. We trained and tested all networks using an NVIDIA Tesla P100 GPU (12 GB memory) on UW Hyak Cluster.

The input for our image translation models consists of three channels: the nuclear channel (analog of hematoxylin), the cytoplasm channel (analog of eosin), and the edge detection map of the cytoplasm channel. Potentially the edge detection map can be waived by replacing with an array of zeros with little reduction in performance. The outputs are single-channel grayscale images of synthetic CK8. Our ITAS3D code for the 2D and 2.5D image translation were modified from the following GitHub repositories: https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix and https://github.com/NVIDIA/vid2vid, with rewriting of data loaders, modification of the U-Net structure used in *pix2pix*, and omission of the original coarse-to-fine training strategy.

4.8.2 *Inferencing phase (step 1 of ITAS3D image translation)*

Whole-biopsy H&E-analog datasets are first sub-divided into blocks of size 1024 × 1024 × 712 pixels with 25% overlap along the biopsy-axis direction (the long axis of the cylindrical biopsy). For level-by-level (2.5D) inference of synthetic CK8 images, each depth level takes into account the synthetic CK8 images generated at two previous levels. Therefore, in order to initiate this process, a 2D translation model (the *pix2pix* model in our case) is required to generate the first two synthetic images within the depth sequence (**Figure 4.9c**). For each image-sequence block, we start the image-sequence translation bi-directionally from the center of the block. This ensures that the initial images are of high-quality and that we can avoid irregularities at the tissue edges (e.g., staining artifacts and surface irregularities at the edges of the biopsies). We split each block into two halves with 34 overlapping levels (**Figure 4.9c**). As mentioned, to initialize the image-sequence translation process, 2D image translation (*pix2pix*) is first performed on the bottom two levels of the top half of the block, and on the top two levels of the bottom half of the block. The image translation of each biopsy (usually consisting of 10-15 blocks with 25% overlap between adjacent blocks) takes ~6 hours with a 12-GB NVIDIA Tesla P100 GPU on a standard node of the UW Hyak HPC cluster. Hyperparameters are set as default values (identical to the original *vid2vid* model) except for those specified in the inference script provided in our GitHub repository. Once image-sequence translation is completed for both halves, the two halves are merged: the first 12 levels in each half are discarded (i.e., they are only used to initiate the image-sequence translation processes), and the remaining 10 overlapping levels are linearly blended (the intensities are smoothly adjusted between the two images so that one image fades out as the other image transitions in). Finally, the synthetic-CK8 image blocks are mosaicked with linear blending using the ImageJ "stitching" plug-in [93].

4.9 IMAGE-TO-IMAGE TRANSLATION RESULTS AND EVALUATION

As shown in step 1 of **Figure 4.5**, 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) [150] (see **Figure 4.9**). 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 **Figure 4.10**).

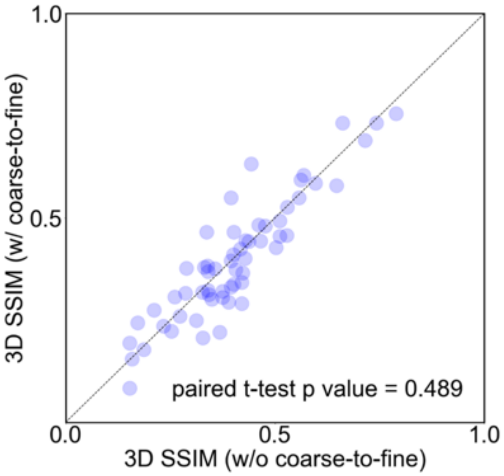

Figure 4.10. A bivariate plot to compare the 3D SSIM metric for synthetic-CK8 images generated with and without coarse-to-fine training. The testing dataset contains 58 tissue volumes, each with a size of ~0.2-mm^3 (1024 × 1024 × 200 pixel). The *p* values (two-sided paired t-test) show no significant differences between the SSIM distributions for synthetic-CK8 images generated by the two models (with and without the coarse-to-fine training strategy).

4.9.1 *Omission of the "coarse-to-fine" training strategy in the vid2vid model*

The coarse-to-fine strategy [160] was a technique employed within the original *vid2vid* model to generate high-resolution videos. In this method, the resolution of the training images was increased incrementally (2× at a time, starting from down-sampled images) while adding resolution-matched residual blocks [173] to the front and back of the generator during the training process until the desired target resolution was reached. With the coarse-to-fine training strategy, image-sequence translation training was performed over 54 epochs, which took 2 months. Without the coarse-to-fine training strategy, only 20 epochs were required, taking about 1 month, but with negligible performance loss (**Figure 4.10**). Both models were trained with a 12-GB NVIDIA Tesla P100 GPU on a standard node at the UW Hyak HPC cluster. Note that if the resources are available, the training can be done with multiple GPUs (e.g., 8 GPUs in the *vid2vid* paper) to potentially shorten the training time to a few days. Hyperparameters were set as default values (identical to the original *vid2vid* model) except for those specified in the training script provided in our GitHub repository. The "coarse-to-fine" training strategy was initially proposed to improve the translation of real-world inputs, such as natural-scenery videos, over a large range of spatial scales [160]. Our speculation is that unlike such videos, our CK8 image translation operates on a more-limited range of micro-scale structures (i.e., close-packed arrangements of luminal epithelial cells surrounding the prostate glands). Therefore, the "coarse-to-fine" method, which significantly increases the training time, does not noticeably improve the model's performance.

4.9.2 *Evaluation of image translation*

While the glands can be delineated on the H&E-analog images by a trained observer, automated computational segmentation of the glands remains challenging [152, 174]. In **Figure 4.11** 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.

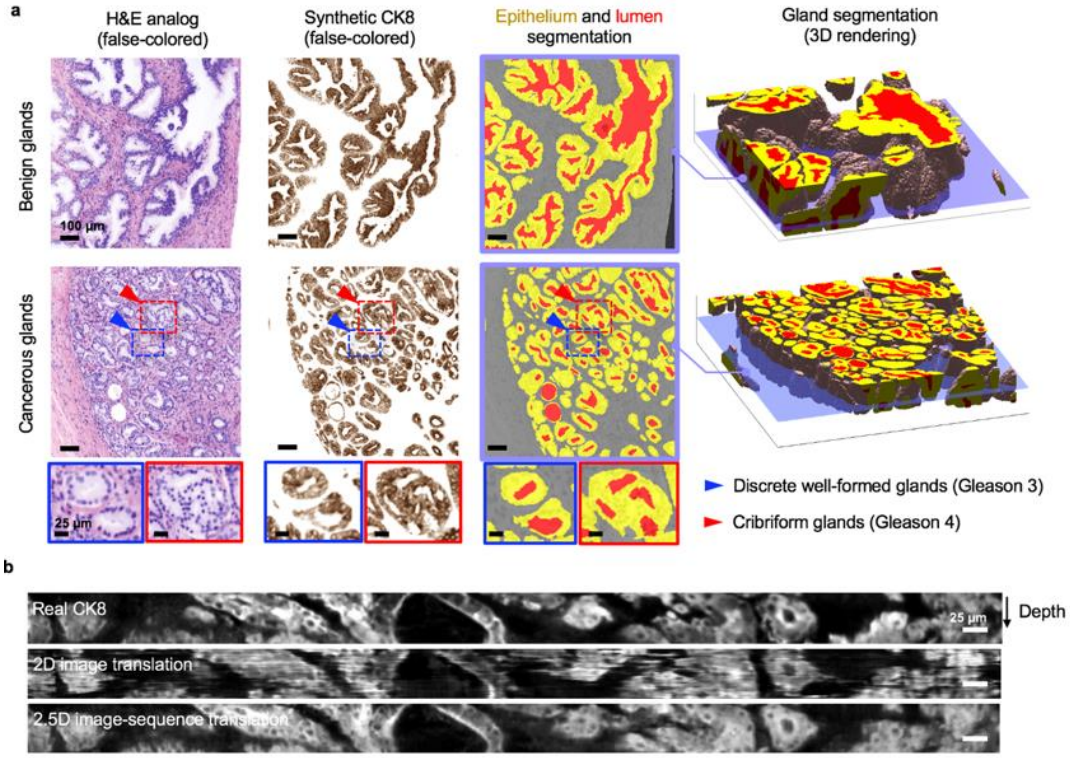

Figure 4.11. Image translation and segmentation results with ITAS3D. (a) 2D cross-sections are shown (from left to right) of false-colored H&E-analog images, synthetic-CK8 IHC images generated by image-sequence translation, and gland-segmentation masks based on the synthetic-CK8 images (yellow for epithelium, 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.

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 **Figure 4.11b**. While obvious distortions and discontinuities are seen as a function of depth with 2D image translation (horizontal distortion on image), the results of our 2.5D image-sequence translation exhibit optimal continuity with depth. 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.

4.10 GLAND SEGMENTATION BASED ON SYNTHETIC-CK8 DATASETS (STEP 2 OF ITAS3D)

Prior to performing gland segmentation, the synthetic-CK8 images are downsampled by 2× in all 3 dimensions (1.8 µm/pixel, which is effectively 4× downsampled from the original resolution of raw images) to reduce memory requirements and computational times while still enabling gland segmentations to be performed at a reasonable spatial resolution. The resulting synthetic-CK8 datasets for each biopsy are approximately 7000 (length) × 512 (width) x 356 (depth) pixels in size.

The overall 3D segmentation workflow is shown in **Figure 4.12**. First, to enhance the contrast of the boundary of the epithelium with respect to the background, a 3×3 edge-sharpening filter is applied to each sagittal 2D level within the 3D dataset (the sagittal view is the typical *en face* view of a biopsy seen by pathologists). Then, a preliminary segmentation mask for the epithelium is obtained via Otsu thresholding. The CK8 biomarker is expressed in the cytoplasm of the luminal epithelial cells. Therefore, in order to fill in the unstained nuclei and to acquire a solid epithelium mask, a binary closing routine (dilation followed by erosion) is performed using a 3×3 structural element with a square connectivity equal to one, followed by a small-hole filling routine applied to the epithelium mask.

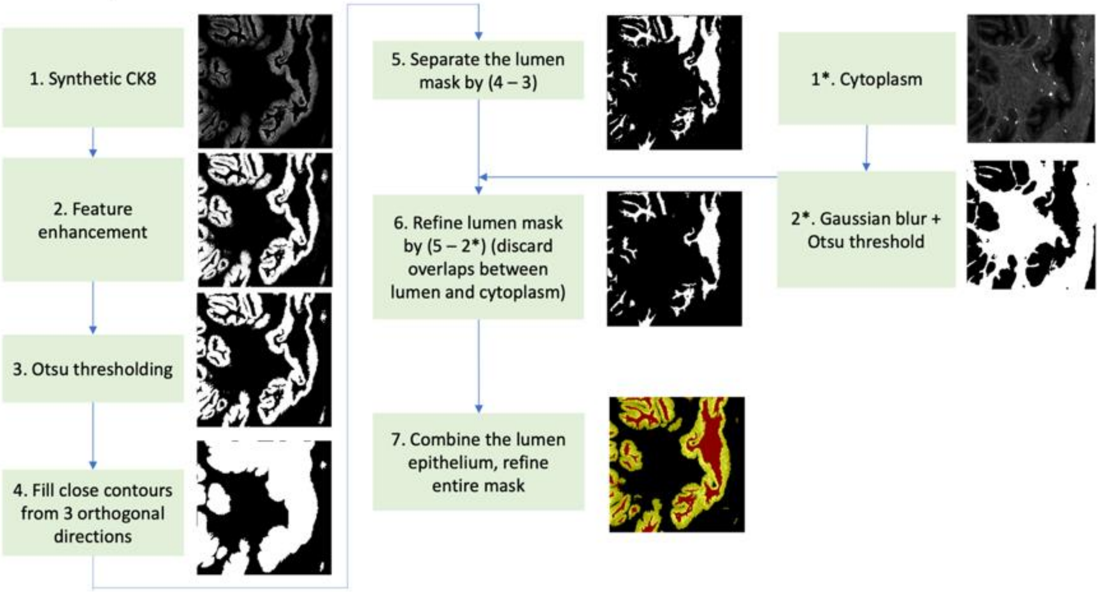

Figure 4.12. A threshold- and morphology-based segmentation algorithm using synthesized CK8 and cytoplasm staining

To segment the lumen, we first fill in any 2D contours enclosed by the epithelium mask in a level-by-level manner along 3 orthogonal directions (see **Figure 4.13**), which are then combined. The key idea is that, although the CK8 signal from the immuno-stained epithelial cells should ideally form a complete 3D outline that encloses the lumen space, in reality, the 3D contour can present gaps due to imperfect staining/thresholding or cut-open glands at biopsy edges. Therefore, an algorithm filling in closed 2D contours from multiple perspectives instead of filling in closed 3D surfaces (which are rare to find due to many small gaps) allow the 3D lumen spaces to identified more correctly. However, this method occasionally introduces false-positive lumen regions that actually correspond to cytoplasm regions. The cytoplasm channel can help to remove these false-lumen regions. An approximate cytoplasm mask is first obtained by applying Gaussian blurring followed by Otsu thresholding on the 2× down-sampled cytoplasm (eosin-analog) images. A binary closing and small-object removal method is then performed to correct for small errors. To extract the false-positive cytoplasm regions, we first subtract the cytoplasm mask from the epithelium mask, and then further separate the real cytoplasm (outside of glands, in the stromal region) from the lumen-overlapping false cytoplasm (inside of glands) with a binary opening (erosion followed by dilation). The number of iterations of the binary opening is empirically determined based on the crowdedness of glands. The false-positive cytoplasm regions inside the glands are identified as having significantly smaller volumes than the real cytoplasm (which we assume should be connected throughout the entire biopsy) and are removed from the final cytoplasm mask. The final lumen mask is refined by subtracting the final cytoplasm mask to discard any overlaps with the actual cytoplasm (stromal) region. We finalize the semantic segmentation of glands by taking the union of the lumen and epithelium masks with some minor adjustments, such as noise removal and filling in of small holes. In addition, a stroma mask is obtained by applying an active contour algorithm [175] to the cytoplasm channel. For each biopsy, the computational time for gland segmentation when starting from a synthetic CK8 dataset is ~30 minutes for a 3.20GHz CPU (Intel® Xeon® Gold 6134) with 64Gb of RAM.

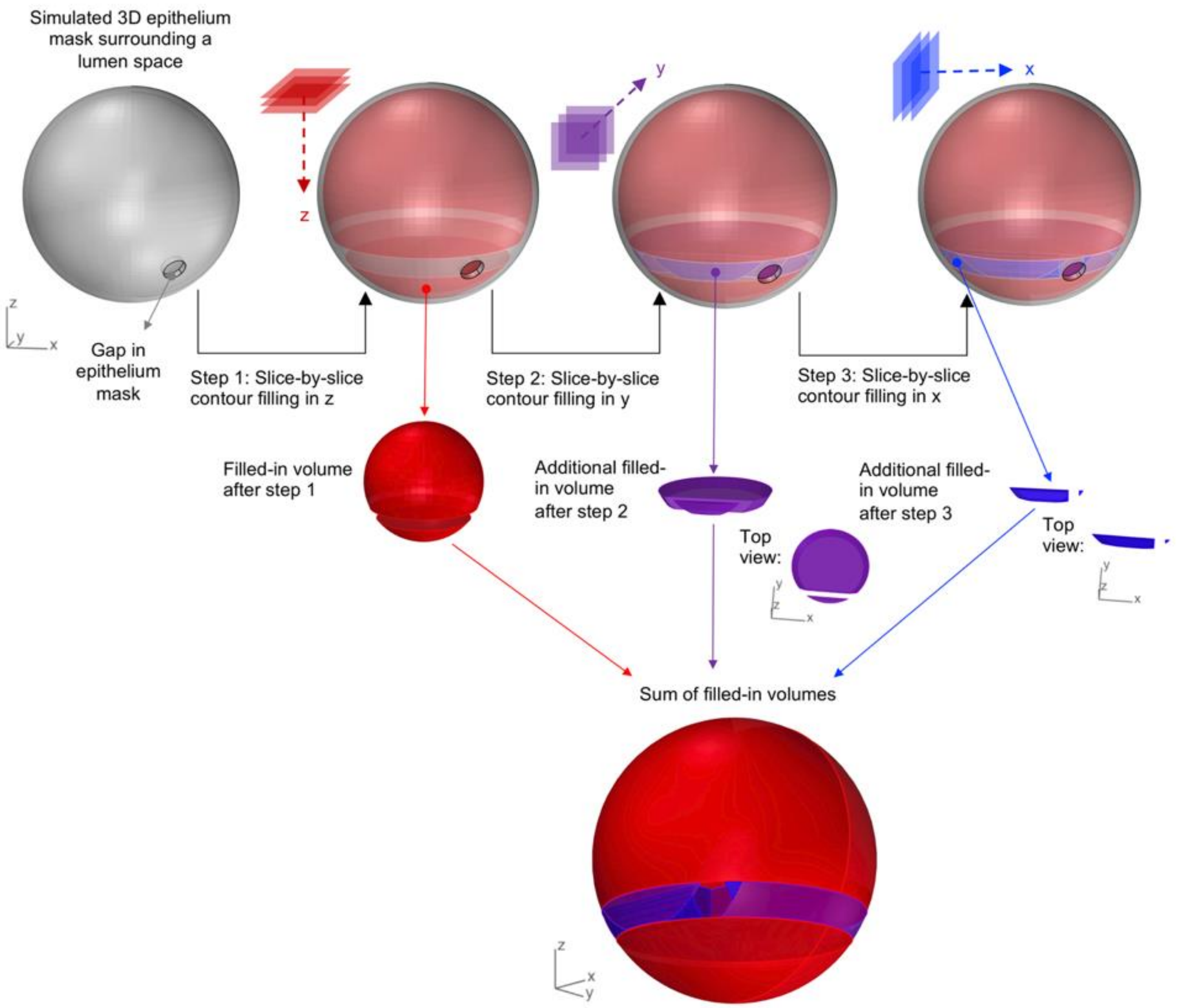

Figure 4.13. Lumen-filling strategy. Epithelial cells should ideally completely enclose all lumen regions. However, due to imperfect labeling, sparse and small gaps in the epithelia occasionally appear. This leads to errors when attempting to identify the lumen spaces with a slice-by-slice contour-filling routine (as shown in step 1). However, by performing slice-by-slice contour filling along 3 orthogonal directions and combining the results, most lumen spaces are accurately filled in. We have found this method to be superior to standard 3D methods for filling in enclosed surfaces.

4.11 RESULTS AND EVALUATION OF 3D GLAND SEGMENTATION

4.11.1 *Gland-segmentation results*

Example 3D prostate gland-segmentation results are shown for benign and cancerous regions in **Figure 4.11a**. While the glands can be delineated on the H&E-analog images by a trained observer, automated computational segmentation of the glands remains challenging [152, 174]. We further show that these synthetic-CK8 datasets allow for relatively straightforward segmentation of the gland epithelium, lumen, and surrounding stromal tissue compartments (**Figure 4.11a**). Glands from various PCa subtypes are successfully segmented as shown, 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).

4.11.2 *Ground-truth annotations to validate gland-segmentation performance*

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.). Ten 0.2-mm^3 regions (512 × 512 × 100 voxel each, representing 0.2-mm^3 of tissue) from different patients were randomly selected from the testing dataset (tri-labeled 3D images not used for training). These regions were manually annotated based on H&E-analog and real-CK8 immunofluorescence images to obtain a ground-truth set of 3D gland segmentations (for both the lumen mask and epithelium mask). These manual annotations were performed using a commercial image-analysis software package, Aivia (Leica Microsystems), under the guidance of a board-certified genitourinary pathologist (N.P.R.).

4.11.3 *Benchmarking the performance of ITAS3D with alternative segmentation methods*

The effectiveness of ITAS3D was compared with two well-known object-segmentation methods. The first benchmarking method utilized traditional 3D image-processing techniques, starting with a watershed method [176] extended to 3D, which was applied on the eosin-analog channel to identify candidate lumen regions only (segmenting the epithelial cells was not successful with this method). Here, the 3D-watershed algorithm was initiated at marker points that were identified with an Otsu thresholding routine applied on the same eosin-analog images. Likewise, nuclei were detected by applying another watershed-based segmentation method on the hematoxylin-analog images [177]. Candidate lumen regions in which the majority of the boundary pixels were not adjacent (within 10 μm) to segmented nuclei were eliminated due to the fact that true lumen regions are always enclosed by epithelial cells. The second benchmarking method was a convolutional neural network, U-Net [154], originally developed for biomedical image segmentation. To train the 2D U-Net model, patches were extracted from 15 ROIs from five biopsies (3 ROIs from each). Manual ground-truth annotations of glands were performed under the guidance of a board-certified pathologist (N.P.R.). Four biopsies were used for training with the remaining biopsy used to validate the accuracy of the model. For each 512 × 512 patch, a split ratio of 4:1 was utilized for training vs. validation [154, 155, 178], which resulted in 1600 patches for training, and 400 for validation. During training, the patches were further augmented via random left or right flipping manipulations, and 90° rotations, to increase the robustness of the model.

We then compared the accuracy of ITAS3D with the two baseline methods, 3D watershed [176] (as a 3D non-DL benchmark) and 2D U-Net [154] (as a 2D DL benchmark). ITAS3D outperforms the two benchmark methods in terms of Dice coefficient [179] and 3D Hausdorff distance [180] (**Figure 4.14)**. 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 [144, 165, 181]; again, this is one of the main motivations for developing the annotation-free ITAS3D method.

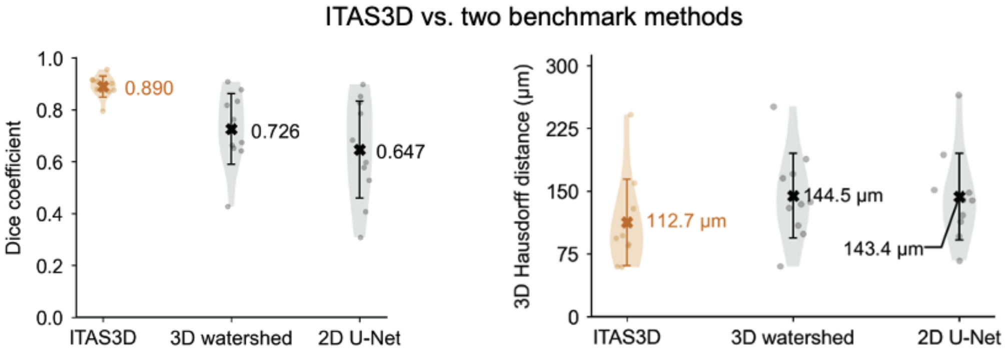

Figure 4.14. Segmentation evaluation against two benchmark methods. For quantitative benchmarking, Dice coefficients (larger is better) and 3D Hausdorff distances (smaller is better) are plotted for ITAS3D-based gland segmentations along with two benchmark methods (3D watershed and 2D U-net), as calculated 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.

4.12 CLINICAL VALIDATION STUDY: GLANDULAR FEATURE EXTRACTION AND CORRELATION WITH BCR OUTCOMES

4.12.1 *Study design and data collection*

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) [182]. 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).

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 (**Figure 4.15)**. 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 [30] (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) [29, 30, 107]. 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. 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 **Figure 4.15**.

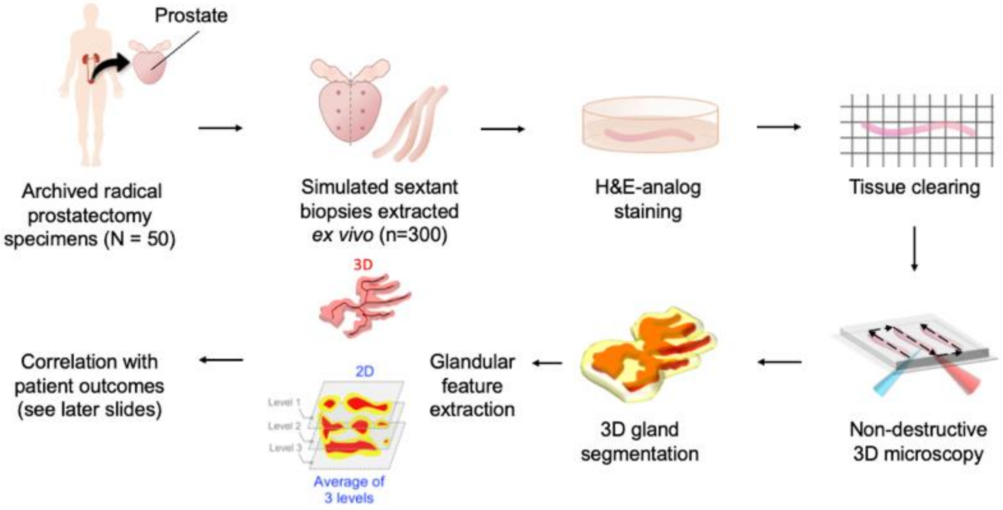

Figure 4.15. Clinical validation workflow. Archived (FFPE) radical prostatectomy (RP) specimens were obtained from a well-curated cohort of 50 patients, from which 300 simulated (*ex vivo*) needle biopsies were extracted (6 biopsies per case, per sextant-biopsy protocol). The biopsies were labeled with 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 from each volume and averaged.

4.12.2 *Cancer region annotations*

Cancer-enriched regions in each biopsy were annotated by recording the coordinates of those regions along the axis of each cylindrically shaped biopsy. This was performed based on a level-by-level visual inspection of the 3D biopsy datasets under the guidance of a board-certified genitourinary pathologist (N.P.R.). We analyzed the glandular features only within the cancerous regions of each biopsy. This was motivated by our finding that histomorphometric features from cancer glands exhibit higher correlation with BCR outcomes than those from adjacent benign glands (**Figure 4.16)**. However, due to the fact that our cancer-region annotations are only performed along the long axis of each cylindrical biopsy (i.e., rough annotations), as well as the observation that benign and cancer glands can be intermixed within certain tissue regions, a more-precise and accurate method to identify individual cancer glands would likely further improve the prognostic value of our methods in the future.

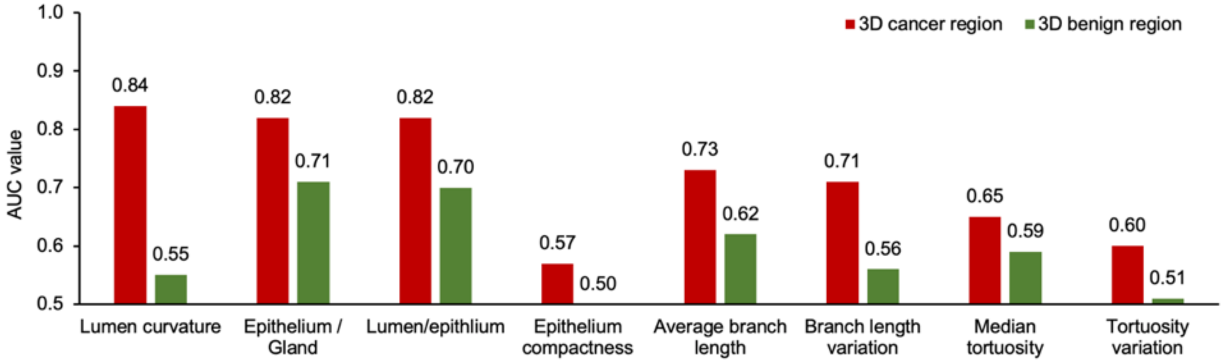

Figure 4.16. Comparing the AUC values for glandular features in cancer and benign regions. Patient-level AUC values for glandular features (4 non-skeleton and 4 skeleton-based features are shown here) are higher for glands in cancerous regions than in adjacent benign regions. We therefore focused our analyses on cancer-enriched regions of each biopsy.

4.12.3 *Feature extraction: non-skeleton glandular features*

From the 3D gland segmentations, 12 non-skeleton glandular features relating to gland size and shape were computed. These features are presented in **Appendix B** along with a brief description and AUC value for each feature. In order to explore the advantages of extending 2D glandular features into 3D, we analyzed features that have direct 2D vs. 3D analogs. The first feature set characterizes the volumetric extent of different tissue compartments (lumen, epithelium, and stroma). The other feature sets characterize the shape of the different compartments, which are derived from a "triangle mesh" that approximates the compartments [183]. Such features include the surface-area-to-volume ratio of an object (in 3D) or circumference-to-area ratio (in 2D); the average surface curvature (in 3D) or boundary curvature (in 2D) of an object [140]; and the gland-to-convex-hull ratio (G/H), defined as the volume ratio (in 3D) or area ratio (in 2D) of the gland mask (epithelium + lumen) to the convex hull that circumscribes the gland. The non-skeleton glandular features were extracted using the "regionprops", "mesh_surface_area", and "ConvexHull" methods from Python's "scikit-image" and "scipy" packages. 3D curvature values were calculated with the "discrete_gaussian_curvature_measure" method from the "trimesh" library using meshes generated by the "marching_cubes" method in "scikit-image".

4.12.4 *Feature extraction: lumen skeleton features*

Extraction of the lumen skeleton (centerline of the lumen) was based on a 3D-thinning algorithm [184]. As mentioned in the main manuscript, meaningful 2D analogs of the 3D skeleton-based features do not exist (**Appendix B**). Before the skeleton extraction, we down-sampled the segmentation mask by 4× and applied a binary erosion (with a spherical kernel of a 2-voxel radius) to prevent false branches from being introduced due to fine surface irregularities. For the skeleton networks, a junction is defined as the point where a lumen splits into two or more lumens. A branch is defined either as a segment between a junction and a junction, a junction and an end point, an end point to an end point, or an isolated loop [185]. We extracted 5 different skeleton features: mean branch length, standard deviation of the branch lengths, median tortuosity (defined as the ratio of branch length to the Euclidean distance between the two end points of the branch), standard deviation of the tortuosity, and branch connectivity. Here, branch connectivity is defined as the ratio of the total number of branches to the total number of connected sets of branches within each biopsy.

4.12.5 *Comparing individual 3D and 2D glandular morphological features in their correlation with patient outcomes*

We compared multiple 3D and 2D glandular histomorphometric features (see **Appendix B** 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 [126]. 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 (**Figure 4.17a**). As another example (**Figure 4.17b**), 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 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 [126]. For various 3D and 2D features (**Figure 4.17a – d** and **Appendix B**), 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 **Figure 4.17a - b** (between BCR and non-BCR groups) and higher area-under-the-ROC-curve (AUC) values (**Figure 4.17c - d**).

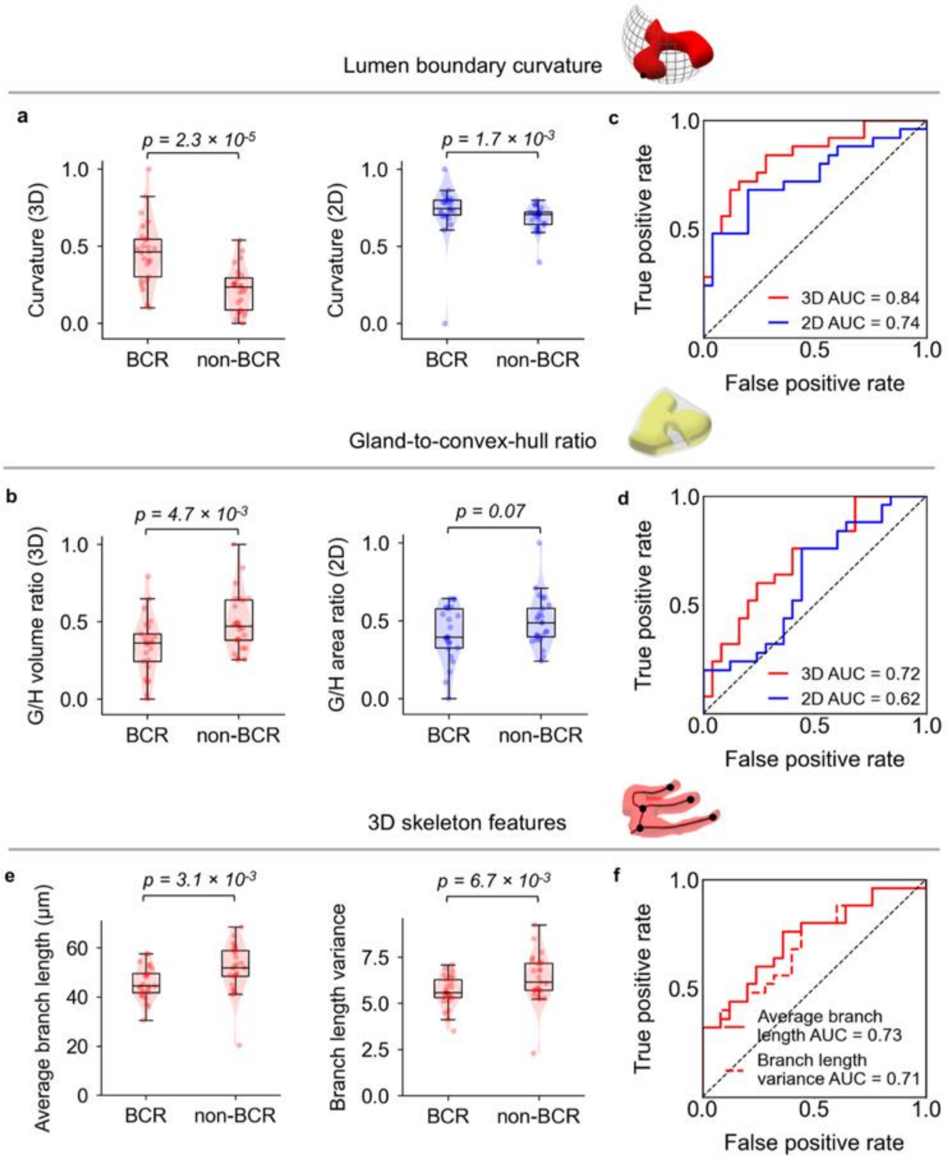

Figure 4.17. Comparing the performance of 3D vs. 2D glandular features for risk stratification. (a-b) Violin and box plots are shown for two examples of 3D glandular features, along with analogous 2D features, for cases in which biochemical recurrence (BCR) was observed within 5 years of RP ("BCR") and for cases with no BCR within 5 years of RP ("non-BCR"). For both sets of example features, "lumen boundary curvature" in panel (a) and "gland-to-convex hull ratio" (G/H) in panel (b), the 3D version of the feature shows improved stratification between BCR and non-BCR groups. (c-d) 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. (e) 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. (f) ROC curves are shown, along with AUC values, for average branch length and branch length variance.

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). 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 **Figure 4.17e,** 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 (**Figure 4.17e – f**). 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 (**Figure 4.18)** 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.

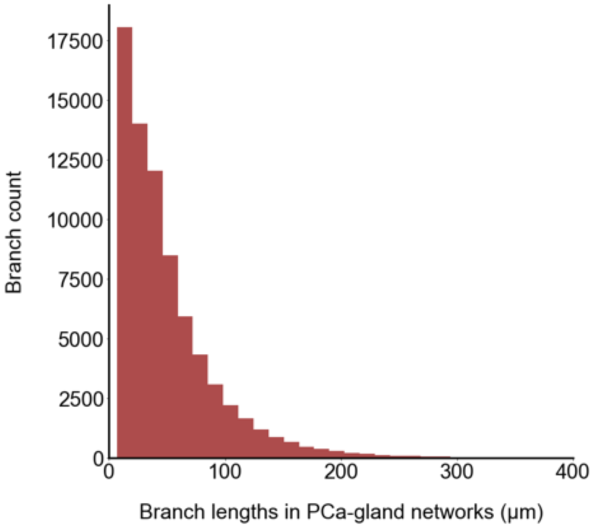

Figure 4.18. Histogram of branch lengths for PCa glands. The diameter of the biopsies (1-mm) imaged in this study is large enough to accurately quantify the majority of branch lengths for PCa glands.

4.12.6 *Combining multiple 3D/2D glandular features for risk stratification*

To explore the prognostic value of combining multiple glandular features, we used linear 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 **Appendix B**. 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 ± standard deviation; **Figure 4.19a**), 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 (**Figure 4.20**). 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 (**Appendix C**), 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 (**Figure 4.19b**). 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.

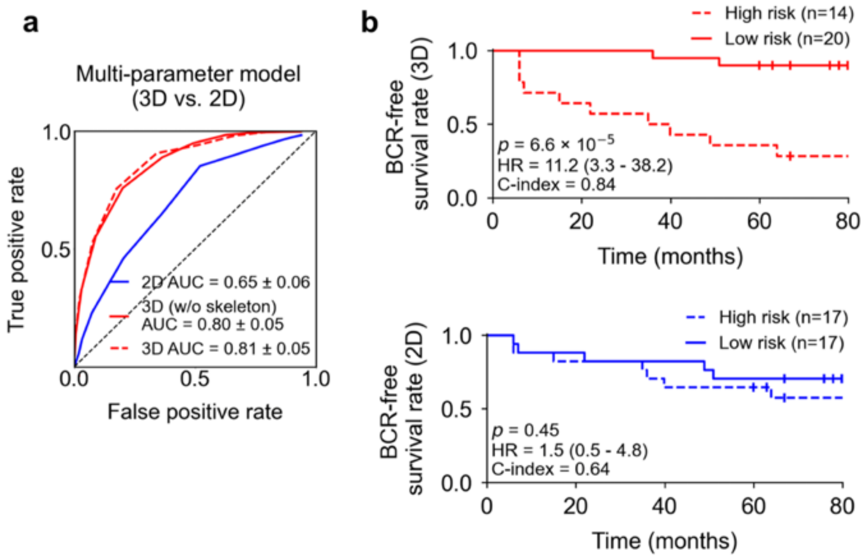

Figure 4.19. Exploring prognosis power of combining multiple 3D / 2D features. (c) ROC curves are shown of various multiparameter models, including those trained with 2D glandular features, 3D glandular features excluding skeleton features, and 3D glandular features including skeleton features. (b) Kaplan-Meier curves are shown for BCR-free survival, showing that a multiparameter model based on 3D glandular features is better able to 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).

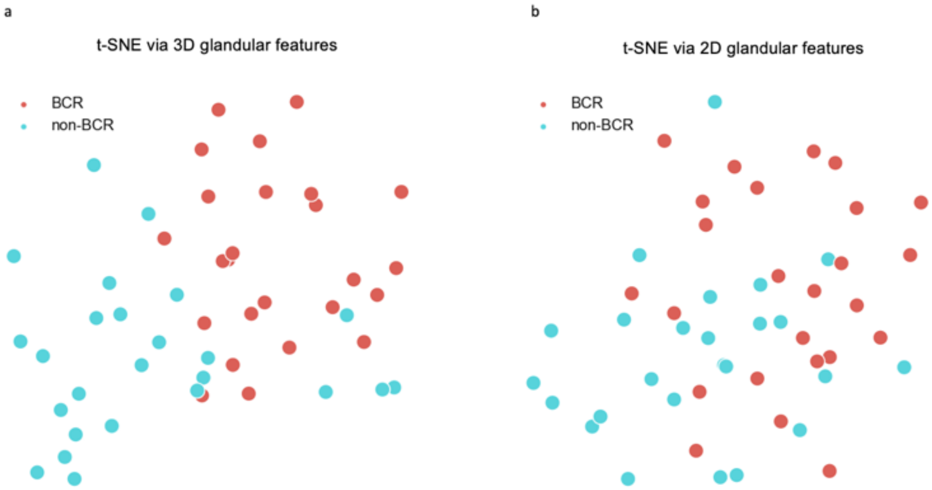

Figure 4.20. Visualization of the separation between BCR and non-BCR groups based on 3D and 2D glandular features. The 25 BCR cases and 25 non-BCR cases were mapped to a 2-dimensional space using *t*-SNE. The separation between the two groups is more evident when the *t*-SNE analysis is based on 3D (a) rather than 2D (b) glandular features.

4.12.7 *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 after 8 weeks post-RP [182]. 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 [186]. 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 [187] 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 [127] using the binary 5-year BCR category as the outcome endpoint. LASSO is a linear 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 [188]. 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 (**Figure 4.21**). This nested CV was performed 200 times in 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.

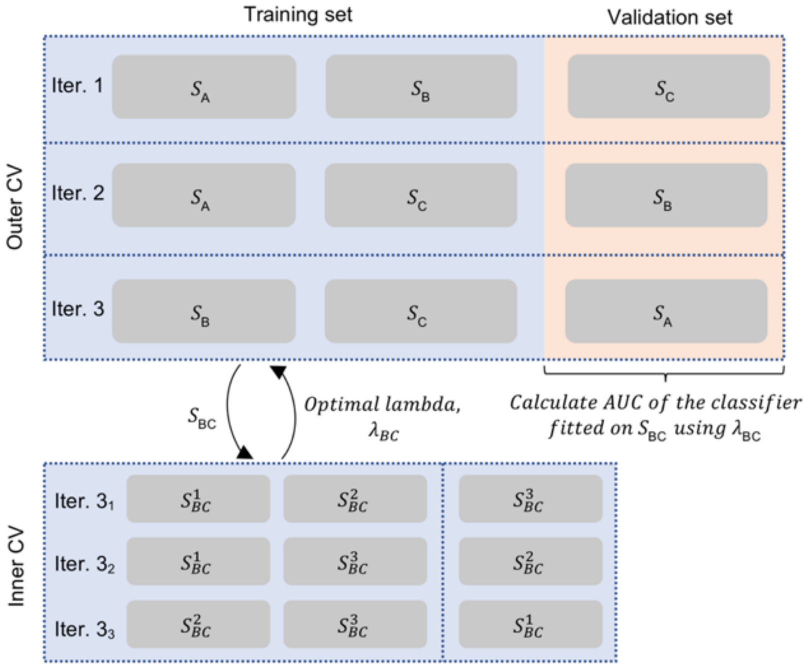

Figure 4.21. Illustration of the model-training and validation schema. A nested 3-fold cross validation (CV) was used for model training and validation. The inner CV was performed at each iteration of the outer CV to determine the optimal model parameter, l. In the outer CV, the model was developed based on the training fold (2/3 of the cases) using the optimal l value. Model performance metrics were then quantified based on the validation fold (remaining 1/3 of cases) to calculate the AUC. The nested CV ensures that there is no overlap between the data used to develop the classification model and the data used to evaluate the performance of the model. The nested CV was performed 200 times, generating 600 AUCs (3 AUCs from each iteration). The average and standard deviation of the AUCs were calculated and compared.

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 **Appendix C**). The performance of the models, either based on

2D or 3D features (non-skeleton features only), 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 [189]. 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.

4.13 DISCUSSION AND FUTURE WORK

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 obtaining high-quality segmentations of key structural primitives [19, 60] (e.g., nuclei, glands, cells, collagen). However, a common bottleneck to achieving accurate segmentations is the need for large amounts of manually annotated datasets [145]. 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 [190, 191], or tubular vessel networks [191, 192]). 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 (code published at https://github.com/WeisiX/ITAS3D) 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 [171]. ITAS3D obviates the need for tedious and subjective manual annotations and, once trained, eliminates the requirement for slow/expensive antibody labeling of thick tissues (**Figure 4.2**). 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). Note that in this specific implementation of ITAS3D, intermediate images are synthetically generated to mimic an IHC stain (CK8) that is routinely 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 [193], CAPRA [194], and Canary-PASS [195]). 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 [196]), 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.

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 [126, 197-199]. 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 (**Figure 4.17** and **Figure 4.19**) 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 (**Figure 4.17** and **Figure 4.19**). 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 [32, 200], 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 suggested in recent studies based on 2D whole slide images of PCa [197]. 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 [19]. 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.

Chapter 5. CONCLUSION AND FUTURE DIRECTIONS

The current histopathology of biopsy and surgical specimens relies on a decade-old traditional workflow that provides limited sampling of tissue and only a 2D view, which results in insufficient reliability of pathological examination. Recent technological advances in high-throughput microscopy and computational image analysis have motivated technical developments for non-destructive 3D pathology, with the hope that non-destructive 3D microscopy can improve current diagnosis and prognosis based on clinical specimens, resulting in superior patient outcomes. On the other hand, the growing shortage of experienced pathologists and the ever-increasing amount of health data, especially in digital pathology, has made tools, such as computational image analysis, indispensable to increase the efficiency and accuracy in pathological analysis. Just in 2021, the FDA approved the first AI-based computational digital pathology solution (Paige Prostate), marking a milestone in the translational research of new computational pathology and indicating the evolution of this field in the years to come. Given the significantly increased amount of complex data in 3D images with new medical image appearances, computational image processing and analysis (with AI and non-AI methods) is in need to improve workflow efficiency and image interpretation reliability, and potentially reduce costs in the clinical adoption of 3D pathology.

In this dissertation, we have discussed the use of 3D optical-sectioning microscopy with computational image processing and analysis for non-destructive 3D pathology of clinical cancer specimens, for the aim of improving the standard of care for breast and prostate cancer, the most common cancer in men and women in the United States, respectively. More specifically, the following three projects has been introduced:

1) For rapid lumpectomy surgical margin assessment, we developed a fluorescent analogue of H&E and a fully automated microscope with UV surface excitation (MUSE) system that incorporates 3D deconvolution and surface extraction to improve image quality, to realize wide-area comprehensive pathology of fresh breast specimen surfaces at a speed that is relevant for intraoperative use.
2) To achieve rapid diagnosis of prostate cancer, a ~45-minute workflow, consisting of fluorescence labeling, tissue clearing, and 3D non-destructive, open-top light-sheet (OTLS) microscopy, has been developed for diagnosis of 12 human prostate needle cores within an hour after biopsy, thereby minimizing patient anxiety.
3) An annotation-free deep learning-assisted segmentation strategy based on image-to-image translation was developed for 3D microscopy images of tissue stained with an H&E analog. We show that non-destructive 3D pathology combined with computational analysis has the potential to enable superior prognostic stratification.

In summary, 3D pathology enabled with 3D microscopy of subcellular resolution and computational image analysis has great potential to improve upon the current slide-based 2D histopathology in terms of speed and reliability. Various 3D microscopy methods have been proposed and demonstrated in recent year, which reveal comprehensive 3D data of large specimens that is rich in biological features and medical information yet challenging and time-consuming for human observers to interpret. These 3D microscopy data are usually very large in their size and contain continuous 3D tissue and cell microstructures that cannot be fully appreciated in 2D histology slides, but vital for our understanding of disease status and progression. The development of computational 3D pathology methods is in need to assist conventional visual analysis. Our study (especially in Chapter 4) is one of first explorations to date for full 3D computational pathology, which shows promising results to advance from 2D to 3D tissue architecture analysis to for cancer management. More follow-up studies are currently in progress. Given multiple previous studies showing 2D nuclear morphological and graphical features are correlated with patient outcome [201, 202], we are interested in prostate cancer prognosis with 3D nuclear features. To establish more powerful prognosis models, other 3D prostate tissue structures that may be indicative of prostate cancer progression include vasculature (based on previous study of vessel biomarkers [203, 204], vessel density [205], lymphovascular invasions [206, 207], etc.) and neurons (based on quantifications of neural microenvironments [208, 209] and perineural invasions [210]). As mentioned, 3D computational pathology is still in its infancy, and many future technical development and large-scale clinical studies are required to develop and demonstrate holistic decision-support algorithms that is suitable for clinical use.

In general, computational pathology aims to increase diagnostic accuracy, improving patient outcome, and reduce costs for patients and clinical systems through global collaboration. With the rapid development of new technology that advances personalized precision medicine, computational pathology became a critical factor to achieve this goal. Though current pathological analysis of specimens is limited to the 2D representation of tissue sections in the slide-based histopathology, numerous 2D computational studies on whole-slide images that incorporated large datasets with high levels of sample diversity have shown promising diagnostic or prognosis results [62, 64, 126, 197, 211]. Our 3D pathology approach is unique in that it is designed without recourse to any established grading system (which relies on 2D slide-based histology ground-truths), and targets directly at optimization of the ultimate clinical goal – the patient outcome. We believe that, empowered by the rapid advancements in 3D machine vision and multi-modality analysis, future developments of non-destructive, comprehensive 3D pathology will provide a biological insight that is more faithful to truth, and therefore support more efficient, objective, and reproducible clinical workflows for oncology management.

BIBLIOGRAPHY

1. *Cancer Statistics*. September 25, 2020; Available from: https://www.cancer.gov/about-cancer/understanding/statistics.
2. Kumar, V., A.K. Abbas, and J.C. Aster, *Robbins and Cotran pathologic basis of disease*. Ninth edition. ed. 2015, Philadelphia, PA: Elsevier/Saunders. xvi, 1391 pages.
3. Carbone, D.P., et al., *First-line nivolumab in stage IV or recurrent non–small-cell lung cancer.* New England Journal of Medicine, 2017. **376**(25): p. 2415-2426.
4. Nagtegaal, I.D., et al., *Pathology is a necessary and informative tool in oncology clinical trials.* Journal of Pathology, 2014. **232**(2): p. 185-189.
5. Grignon, D.J., *Prostate cancer reporting and staging: needle biopsy and radical prostatectomy specimens.* Modern Pathology, 2018. **31**: p. S96-S109.
6. Gokozan, H., et al., *Concurrent Thyroid Fine Needle Aspirations and Core Needle Biopsies: An Institutional Experience.* Modern Pathology, 2022. **35**(Suppl 2): p. 259-260.
7. Goldhoff, P.E., et al., *Fine-Needle Aspiration Biopsy of Liver Lesions Yields Higher Tumor Fraction for Molecular Studies: A Direct Comparison With Concurrent Core Needle Biopsy.* Journal of the National Comprehensive Cancer Network, 2019. **17**(9): p. 1075-1081.
8. Bornschein, J., et al., *Biopsy Sampling in Upper Gastrointestinal Endoscopy: A Survey from 10 Tertiary Referral Centres Across Europe.* Digestive Diseases, 2021. **39**(3): p. 179-189.
9. Nischal, U., K. Nischal, and U. Khopkar, *Techniques of skin biopsy and practical considerations.* J Cutan Aesthet Surg, 2008. **1**(2): p. 107-11.
10. Bain, B.J., *Bone marrow aspiration.* Journal of Clinical Pathology, 2001. **54**(9): p. 657.
11. Fisher, B., et al., *Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer.* New England Journal of Medicine, 2002. **347**(16): p. 1233-1241.
12. Jacobson, J.A., et al., *Ten-year results of a comparison of conservation with mastectomy in the treatment of stage I and II breast cancer.* New England Journal of Medicine, 1995. **332**(14): p. 907-911.
13. Gage, I., et al., *Pathologic margin involvement and the risk of recurrence in patients treated with breast-conserving therapy.* Cancer, 1996. **78**(9): p. 1921-8.
14. Jacobs, L., *Positive margins: the challenge continues for breast surgeons.* Ann Surg Oncol, 2008. **15**(5): p. 1271-2.
15. Makhlouf, H., et al., *Toward Improving Practices for Submission of Diagnostic Tissue Blocks for National Cancer Institute Clinical Trials.* American Journal of Clinical Pathology, 2020. **153**(2): p. 149-155.
16. Olson, E., M.J. Levene, and R. Torres, *Multiphoton microscopy with clearing for three dimensional histology of kidney biopsies.* Biomedical Optics Express, 2016. **7**(8): p. 3089-3096.
17. Paul, D., et al., *Novel 3D analysis of Claudin-5 reveals significant endothelial heterogeneity among CNS microvessels.* Microvascular Research, 2013. **86**: p. 1-10.

18. Torres, R., et al., *Three-Dimensional Morphology by Multiphoton Microscopy with Clearing in a Model of Cisplatin-Induced CKD.* Journal of the American Society of Nephrology, 2016. **27**(4): p. 1102-1112.
19. Liu, J.T.C., et al., *Harnessing non-destructive 3D pathology.* Nat Biomed Eng, 2021. **5**(3): p. 203-218.
20. Brachtel, E.F., et al., *Spectrally encoded confocal microscopy for diagnosing breast cancer in excision and margin specimens.* Laboratory Investigation, 2016. **96**(4): p. 459-467.
21. Flores, E.S., et al., *Intraoperative imaging during Mohs surgery with reflectance confocal microscopy: initial clinical experience.* Journal of Biomedical Optics, 2015. **20**(6).
22. Yin, C.B., et al., *Label-free in vivo pathology of human epithelia with a high-speed handheld dual-axis confocal microscope.* Journal of Biomedical Optics, 2019. **24**(3).
23. Shavlokhova, V., et al., *Detection of oral squamous cell carcinoma with ex vivo fluorescence confocal microscopy: sensitivity and specificity compared to histopathology.* Journal of Biophotonics, 2020.
24. You, S.X., et al., *Intravital imaging by simultaneous label-free autofluorescence-multiharmonic microscopy.* Nature Communications, 2018. **9**.
25. Giacomelli, M.G., et al., *Rapid imaging of surgical breast excisions using direct temporal sampling two photon fluorescent lifetime imaging.* Biomedical optics express, 2015. **6**(11): p. 4317-4325.
26. Giacomelli, M.G., et al., *Multiscale nonlinear microscopy and widefield white light imaging enables rapid histological imaging of surgical specimen margins.* Biomedical optics express, 2018. **9**(5): p. 2457-2475.
27. Wu, Y.C., et al., *Multiview confocal super-resolution microscopy.* Nature, 2021. **600**(7888): p. 279-+.
28. Klioutchnikov, A., et al., *Three-photon head-mounted microscope for imaging deep cortical layers in freely moving rats.* Nature Methods, 2020. **17**(5): p. 509-+.
29. Barner, L.A., et al., *Multi-resolution open-top light-sheet microscopy to enable efficient 3D pathology workflows.* Biomedical Optics Express, 2020. **11**(11): p. 6605-6619.
30. Glaser, A.K., et al., *Multi-immersion open-top light-sheet microscope for high-throughput imaging of cleared tissues.* Nature Communications, 2019. **10**.
31. Glaser, A.K., et al., *A hybrid open-top light-sheet microscope for versatile multi-scale imaging of cleared tissues.* Nature Methods, 2022.
32. Glaser, A.K., et al., *Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens.* Nature Biomedical Engineering, 2017. **1**(7).
33. Merz, S.F., et al., *High-resolution three-dimensional imaging for precise staging in melanoma.* European Journal of Cancer, 2021. **159**: p. 182-193.
34. Yordanov, S., et al., *Single-objective high-resolution confocal light sheet fluorescence microscopy for standard biological sample geometries.* Biomedical Optics Express, 2021. **12**(6): p. 3372-3391.
35. Patel, K.B., et al., *High-speed light-sheet microscopy for the in-situ acquisition of volumetric histological images of living tissue.* Nature Biomedical Engineering, 2022. **6**(5): p. 569-583.
36. Yoshitake, T., et al., *Rapid histopathological imaging of skin and breast cancer surgical specimens using immersion microscopy with ultraviolet surface excitation.* Scientific Reports, 2018. **8**.

37. Xie, W.S., et al., *Microscopy with ultraviolet surface excitation for wide-area pathology of breast surgical margins.* Journal of Biomedical Optics, 2019. **24**(2).
38. Fereidouni, F., et al., *Microscopy with ultraviolet surface excitation for rapid slide-free histology.* Nature Biomedical Engineering, 2017. **1**(12): p. 957-+.
39. Zhang, Y., et al., *High-Throughput, Label-Free and Slide-Free Histological Imaging by Computational Microscopy and Unsupervised Learning.* Advanced Science, 2022. **9**(2).
40. Wang, M., et al., *Gigapixel surface imaging of radical prostatectomy specimens for comprehensive detection of cancer-positive surgical margins using structured illumination microscopy.* Scientific Reports, 2016. **6**.
41. Wang, M., et al., *High-Resolution Rapid Diagnostic Imaging of Whole Prostate Biopsies Using Video-Rate Fluorescence Structured Illumination Microscopy.* Cancer Research, 2015. **75**(19): p. 4032-4041.
42. Zhang, Y., et al., *Speckle illumination microscopy enables slide-free and non-destructive pathology of human lung adenocarcinoma.* bioRxiv, 2021.
43. Wong, T.T.W., et al., *Fast label-free multilayered histology-like imaging of human breast cancer by photoacoustic microscopy.* Science Advances, 2017. **3**(5).
44. Li, X.F., et al., *High-speed label-free ultraviolet photoacoustic microscopy for histology-like imaging of unprocessed biological tissues.* Optics Letters, 2020. **45**(19): p. 5401-5404.
45. Haven, N.J.M., et al., *Reflective objective-based ultraviolet photoacoustic remote sensing virtual histopathology.* Optics Letters, 2020. **45**(2): p. 535-538.
46. Bell, K., et al., *Reflection-mode virtual histology using photoacoustic remote sensing microscopy.* Scientific Reports, 2020. **10**(1).
47. Dornbusch, J.A., et al., *Optical coherence tomography imaging of excised canine apocrine gland anal sac adenocarcinoma tumors.* Veterinary and Comparative Oncology.
48. Martinez-Lapiscina, E.H., et al., *Retinal thickness measured with optical coherence tomography and risk of disability worsening in multiple sclerosis: a cohort study.* Lancet Neurology, 2016. **15**(6): p. 574-584.
49. Schonfeldt-Lecuona, C., et al., *Retinal changes in patients with major depressive disorder - A controlled optical coherence tomography study.* Journal of Affective Disorders, 2018. **227**: p. 665-671.
50. Tang, P.J., et al., *Polarization sensitive optical coherence tomography with single input for imaging depth-resolved collagen organizations.* Light-Science & Applications, 2021. **10**(1).
51. Lu, J., et al., *Application of OCT-Derived Attenuation Coefficient in Acute Burn-Damaged Skin.* Lasers in Surgery and Medicine, 2021. **53**(9): p. 1192-1200.
52. Chung, K., et al., *Structural and molecular interrogation of intact biological systems.* Nature, 2013. **497**(7449): p. 332-+.
53. Renier, N., et al., *iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging.* Cell, 2014. **159**(4): p. 896-910.
54. Susaki, E.A., et al., *Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis.* Cell, 2014. **157**(3): p. 726-39.
55. Klingberg, A., et al., *Fully Automated Evaluation of Total Glomerular Number and Capillary Tuft Size in Nephritic Kidneys Using Lightsheet Microscopy.* Journal of the American Society of Nephrology, 2017. **28**(2): p. 452-459.

56. Siegel, R.L., et al., *Cancer Statistics, 2021.* CA Cancer J Clin, 2021. **71**(1): p. 7-33.
57. Cui, M. and D.Y. Zhang, *Artificial intelligence and computational pathology.* Laboratory Investigation, 2021. **101**(4): p. 412-422.
58. Jung, M.S., et al., *Artificial intelligence system shows performance at the level of uropathologists for the detection and grading of prostate cancer in core needle biopsy: an independent external validation study.* Modern Pathology, 2022.
59. Morales, S., K. Engan, and V. Naranjo, *Artificial intelligence in computational pathology - challenges and future directions.* Digital Signal Processing, 2021. **119**.
60. Bera, K., et al., *Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology.* Nature Reviews Clinical Oncology, 2019. **16**(11): p. 703-715.
61. Hou, L., et al. *Patch-based convolutional neural network for whole slide tissue image classification*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
62. Campanella, G., et al., *Clinical-grade computational pathology using weakly supervised deep learning on whole slide images.* Nature Medicine, 2019. **25**(8): p. 1301-+.
63. Chen, R.J., et al. *Whole Slide Images are 2D Point Clouds: Context-Aware Survival Prediction using Patch-based Graph Convolutional Networks*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2021. Springer.
64. Lu, M.Y., et al., *Data-efficient and weakly supervised computational pathology on whole-slide images.* Nature Biomedical Engineering, 2021. **5**(6): p. 555-+.
65. de Haan, K., et al., *Automated screening of sickle cells using a smartphone-based microscope and deep learning.* Npj Digital Medicine, 2020. **3**(1).
66. Demirel, H. and G. Anbarjafari, *IMAGE Resolution Enhancement by Using Discrete and Stationary Wavelet Decomposition.* Ieee Transactions on Image Processing, 2011. **20**(5): p. 1458-1460.
67. Zhang, L., et al., *Resolution Enhancement for Inversed Synthetic Aperture Radar Imaging Under Low SNR via Improved Compressive Sensing.* Ieee Transactions on Geoscience and Remote Sensing, 2010. **48**(10): p. 3824-3838.
68. Wang, H.D., et al., *Deep learning enables cross-modality super-resolution in fluorescence microscopy.* Nature Methods, 2019. **16**(1): p. 103-+.
69. Choi, H. and D.S. Lee, *Generation of structural MR images from amyloid PET: Application to MR-less quantification.* Journal of Nuclear Medicine, 2018. **59**(7): p. 1111-1117.
70. Olut, S., et al. *Generative adversarial training for mra image synthesis using multi-contrast mri*. in *International Workshop on PRedictive Intelligence In MEdicine*. 2018. Springer.
71. Nie, D., et al. *Medical image synthesis with context-aware generative adversarial networks*. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2017. Springer.
72. Bayramoglu, N., et al. *Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks*. in *Proceedings of the IEEE International Conference on Computer Vision*. 2017.
73. Rivenson, Y., et al., *Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning.* Nature Biomedical Engineering, 2019. **3**(6): p. 466-477.

74. Xie, W.S., et al., *Prostate Cancer Risk Stratification via Nondestructive 3D Pathology with Deep Learning-Assisted Gland Analysis.* Cancer Research, 2022. **82**(2): p. 334-345.
75. Xie, W.S., et al., *Diagnosing 12 prostate needle cores within an hour of biopsy via open-top light-sheet microscopy.* Journal of Biomedical Optics, 2020. **25**(12).
76. Hwang, E.S., et al., *Survival after lumpectomy and mastectomy for early stage invasive breast cancer.* Cancer, 2013. **119**(7): p. 1402-1411.
77. Landercasper, J., et al., *Reasons for Re-Excision After Lumpectomy for Breast Cancer: Insight from the American Society of Breast Surgeons Mastery(SM) Database.* Annals of Surgical Oncology, 2014. **21**(10): p. 3185-3191.
78. Pleijhuis, R.G., et al., *Obtaining Adequate Surgical Margins in Breast-Conserving Therapy for Patients with Early-Stage Breast Cancer: Current Modalities and Future Directions.* Annals of Surgical Oncology, 2009. **16**(10): p. 2717-2730.
79. Waljee, J.F., et al., *Effect of esthetic outcome after breast-conserving surgery on psychosocial functioning and quality of life.* Journal of Clinical Oncology, 2008. **26**(20): p. 3331-3337.
80. Jeevan, R., et al., *Reoperation rates after breast conserving surgery for breast cancer among women in England: retrospective study of hospital episode statistics.* British Medical Journal, 2012. **345**.
81. Huston, T.L., et al., *The influence of additional surgical margins on the total specimen volume excised and the reoperative rate after breast-conserving surgery.* American Journal of Surgery, 2006. **192**(4): p. 509-512.
82. Cabioglu, N., et al., *Role for intraoperative margin assessment in patients undergoing breast-conserving surgery.* Annals of Surgical Oncology, 2007. **14**(4): p. 1458-1471.
83. MacCarty, W.C., *The diagnostic reliability of frozen sections.* American Journal of Pathology, 1929. **5**(4): p. 377-U21.
84. Whitehair, J.G., et al., *The Accuracy of Intraoperative Diagnoses Based on Examination of Frozen-Sections - a Prospective Comparison with Paraffin-Embedded Sections.* Veterinary Surgery, 1993. **22**(4): p. 255-259.
85. Olson, S.M., M. Hussaini, and J.S. Lewis, *Frozen section analysis of margins for head and neck tumor resections: reduction of sampling errors with a third histologic level.* Modern Pathology, 2011. **24**(5): p. 665-670.
86. Schnitt, S.J. and M. Morrow, *Should Intraoperative Frozen Section Evaluation of Breast Lumpectomy Margins Become Routine Practice?* American Journal of Clinical Pathology, 2012. **138**(5): p. 635-638.
87. Qorbani, A., et al., *Microscopy with ultraviolet surface excitation (MUSE): A novel approach to real-time inexpensive slide-free dermatopathology.* Journal of Cutaneous Pathology, 2018. **45**(7): p. 498-503.
88. Tuma, R.S., et al., *Characterization of SYBR gold nucleic acid gel stain: A dye optimized for use with 300-nm ultraviolet transilluminators.* Analytical Biochemistry, 1999. **268**(2): p. 278-288.
89. Elfer, K.N., et al., *DRAQ5 and Eosin ('D&E') as an Analog to Hematoxylin and Eosin for Rapid Fluorescence Histology of Fresh Tissues.* PLOS ONE, 2016. **11**(10): p. e0165530.
90. Zanella, R., et al., *Towards real-time image deconvolution: application to confocal and STED microscopy.* Scientific Reports, 2013. **3**.

91. Bruce, M.A. and M.J. Butte, *Real-time GPU-based 3D Deconvolution.* Optics Express, 2013. **21**(4): p. 4766-4773.
92. Forster, B., et al., *Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images.* Microscopy Research and Technique, 2004. **65**(1-2): p. 33-42.
93. Preibisch, S., S. Saalfeld, and P. Tomancak, *Globally optimal stitching of tiled 3D microscopic image acquisitions.* Bioinformatics, 2009. **25**(11): p. 1463-1465.
94. Giacomelli, M.G., et al., *Virtual Hematoxylin and Eosin Transillumination Microscopy Using Epi-Fluorescence Imaging.* Plos One, 2016. **11**(8).
95. Liu, Y.H., et al., *Pocket MUSE: an affordable, versatile and high-performance fluorescence microscope using a smartphone.* Communications Biology, 2021. **4**(1).
96. Lee, M.Y., et al., *Fluorescent labeling of abundant reactive entities (FLARE) for cleared-tissue and super-resolution microscopy.* Nature Protocols, 2022. **17**(3): p. 819-+.
97. Cooper, D.J., et al., *CoreView: fresh tissue biopsy assessment at the bedside using a millifluidic imaging chip.* Lab on a Chip, 2022. **22**(7): p. 1354-1364.
98. Lu, T.T., et al., *Rapid assessment of breast tumor margins using deep ultraviolet fluorescence scanning microscopy.* Journal of Biomedical Optics, 2020. **25**(12).
99. *Key Statistics for Prostate Cancer*. 2022; Available from: https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html.
100. Dillard, A.J., et al., *Anxiety symptoms prior to a prostate cancer diagnosis: Associations with knowledge and openness to treatment.* British Journal of Health Psychology, 2017. **22**(1): p. 151-168.
101. McQueen, A., *Waiting for a cancer diagnosis.* Cancer Nursing Practice, 2009. **8**(4).
102. Barksdale, S.K., S.A. Oberlender, and R.L. Barnhill, *"Rush" skin biopsy specimens in a tertiary medical center: Diagnostic yield and clinical utility.* Journal of the American Academy of Dermatology, 1998. **38**(4): p. 548-554.
103. Hörl, D., et al., *BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples.* Nature Methods, 2019. **16**(9): p. 870-874.
104. Brawer, M.K., et al., *Keratin immunoreactivity in the benign and neoplastic human prostate.* Cancer Res, 1985. **45**(8): p. 3663-7.
105. Randolph, J.J., *Free-Marginal Multirater Kappa (multirater K [free]): An Alternative to Fleiss' Fixed-Marginal Multirater Kappa.* Online submission, 2005.
106. Chen, Y., et al., *Rapid pathology of lumpectomy margins with open-top light-sheet (OTLS) microscopy.* Biomed Opt Express, 2019. **10**(3): p. 1257-1272.
107. Reder, N.P., et al., *Open-Top Light-Sheet Microscopy Image Atlas of Prostate Core Needle Biopsies.* Archives of Pathology & Laboratory Medicine, 2019. **143**(9): p. 1069-1075.
108. Ozkan, T.A., et al., *Interobserver variability in Gleason histological grading of prostate cancer.* Scand J Urol, 2016. **50**(6): p. 420-424.
109. Shah, R.B., 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.* Hum Pathol, 2016. **56**: p. 155-62.
110. Van der Kwast, T.H., et al., *Variability in diagnostic opinion among pathologists for single small atypical foci in prostate biopsies.* Am J Surg Pathol, 2010. **34**(2): p. 169-77.

111. Giacomelli, M.G., et al., *Virtual Hematoxylin and Eosin Transillumination Microscopy Using Epi-Fluorescence Imaging.* PLoS One, 2016. **11**(8): p. e0159337.
112. Serafin, R., et al., *FalseColor-Python: a rapid intensity-leveling and digital-staining package for fluorescence-based slide-free digital pathology.* bioRxiv, 2020: p. 2020.05.03.074955.
113. Naik, S., et al. *Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology*. in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2008. IEEE.
114. Naik, S., et al. *Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information*. in *MIAAB workshop*. 2007. Citeseer.
115. Epstein, J.I., *A new contemporary prostate cancer grading system.* Ann Pathol, 2015. **35**(6): p. 474-6.
116. Shah, R.B., 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**: p. 155-162.
117. Kane, C.J., et al., *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**(4-5): p. 487-497.
118. Albertsen, P.C., *Treatment of localized prostate cancer: when is active surveillance appropriate?* Nature Reviews Clinical Oncology, 2010. **7**(7): p. 394-400.
119. Haffner, M.C., et al., *Diagnostic challenges of clonal heterogeneity in prostate cancer.* J Clin Oncol, 2015. **33**(7): p. e38-40.
120. Bill-Axelson, A., et al., *Radical prostatectomy or watchful waiting in early prostate cancer.* New England Journal of Medicine, 2014. **370**(10): p. 932-942.
121. Frey, A.U., J. Sonksen, and M. Fode, *Neglected side effects after radical prostatectomy: a systematic review.* J Sex Med, 2014. **11**(2): p. 374-85.
122. van Royen, M.E., et al., *Three-dimensional microscopic analysis of clinical prostate specimens.* Histopathology, 2016. **69**(6): p. 985-992.
123. Tanaka, N., et al., *Whole-tissue biopsy phenotyping of three-dimensional tumours reveals patterns of cancer heterogeneity.* Nat Biomed Eng, 2017. **1**(10): p. 796-806.
124. Tolkach, Y., S. Thomann, and G. Kristiansen, *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**(6): p. 1051-1059.
125. Humphrey, P.A., *Complete Histologic Serial Sectioning of a Prostate-Gland with Adenocarcinoma.* American Journal of Surgical Pathology, 1993. **17**(5): p. 468-472.
126. Lee, G., et al., *Co-occurring gland angularity in localized subgraphs: predicting biochemical recurrence in intermediate-risk prostate cancer patients.* PLoS One, 2014. **9**(5): p. e97954.
127. Leo, P., 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**(1): p. 35.

128. Lee, G., 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**(1): p. 284-297.
129. Lu, M.Y., et al., *AI-based pathology predicts origins for cancers of unknown primary.* Nature, 2021.
130. Poplin, R., et al., *Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning.* Nature Biomedical Engineering, 2018. **2**(3): p. 158-164.
131. Hollon, T.C., et al., *Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks.* Nature medicine, 2020. **26**(1): p. 52-58.
132. Leo, P., et al., *Stable and discriminating features are predictive of cancer presence and Gleason grade in radical prostatectomy specimens: a multi-site study.* Sci Rep, 2018. **8**(1): p. 14918.
133. Lin, W., et al., *Comparison of handcrafted features and convolutional neural networks for liver MR image adequacy assessment.* Sci Rep, 2020. **10**(1): p. 20336.
134. Chandramouli, S., 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**(9).
135. Ali, S., et al., *Selective invocation of shape priors for deformable segmentation and morphologic classification of prostate cancer tissue microarrays.* Computerized Medical Imaging and Graphics, 2015. **41**: p. 3-13.
136. Lefebvre, A.E.Y.T., et al., *Automated segmentation and tracking of mitochondria in live-cell time-lapse images.* Nature Methods, 2021. **18**(9): p. 1091-+.
137. Bhargava, H.K., 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**(8): p. 1915-1923.
138. Kiemen, A., et al., *In situ characterization of the 3D microanatomy of the pancreas and pancreatic cancer at single cell resolution.* bioRxiv, 2020.
139. Arunkumar, N., et al., *Fully automatic model-based segmentation and classification approach for MRI brain tumor using artificial neural networks.* Concurrency and Computation-Practice & Experience, 2020. **32**(1).
140. Koyuncu, C.F., et al. *Three-dimensional histo-morphometric features from light sheet microscopy images result in improved discrimination of benign from malignant glands in prostate cancer*. in *Medical Imaging 2020: Digital Pathology*. 2020. International Society for Optics and Photonics.
141. Chen, H., et al., *VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images.* NeuroImage, 2018. **170**: p. 446-455.
142. Bui, T.D., J. Shin, and T. Moon, *3D densely convolutional networks for volumetric segmentation.* arXiv preprint arXiv:1709.03199, 2017.
143. Çiçek, Ö., et al. *3D U-Net: learning dense volumetric segmentation from sparse annotation*. in *International conference on medical image computing and computer-assisted intervention*. 2016. Springer.
144. Zhu, Z., et al. *A 3D coarse-to-fine framework for volumetric medical image segmentation*. in *2018 International Conference on 3D Vision (3DV)*. 2018. IEEE.
145. van der Laak, J., G. Litjens, and F. Ciompi, *Deep learning in histopathology: the path to the clinic.* Nature Medicine, 2021. **27**(5): p. 775-784.

146. Di Cataldo, S., et al., *Automated segmentation of tissue images for computerized IHC analysis.* Computer methods and programs in biomedicine, 2010. **100**(1): p. 1-15.
147. Migliozzi, D., H.T. Nguyen, and M.A.M. Gijs, *Combining fluorescence-based image segmentation and automated microfluidics for ultrafast cell-by-cell assessment of biomarkers for HER2-type breast carcinoma.* Journal of biomedical optics, 2018. **24**(2): p. 1-8.
148. Lee, S.S., et al., *Nondestructive, multiplex three-dimensional mapping of immune infiltrates in core needle biopsy.* Lab Invest, 2019. **99**(9): p. 1400-1413.
149. Burlingame, E.A., et al., *SHIFT: speedy histological-to-immunofluorescent translation of a tumor signature enabled by deep learning.* Scientific Reports, 2020. **10**(1).
150. Wang, T.-C., et al., *Video-to-video synthesis.* arXiv preprint arXiv:1808.06601, 2018.
151. Peng, Y., et al. *Segmentation of prostatic glands in histology images*. in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2011. IEEE.
152. Singh, M., et al., *Gland segmentation in prostate histopathological images.* J Med Imaging (Bellingham), 2017. **4**(2): p. 027501.
153. Chen, H., et al. *Deep contextual networks for neuronal structure segmentation*. in *AAAI*. 2016.
154. Ronneberger, O., P. Fischer, and T. Brox. *U-net: Convolutional networks for biomedical image segmentation*. in *International Conference on Medical image computing and computer-assisted intervention*. 2015. Springer.
155. Sirinukunwattana, K., et al., *Gland segmentation in colon histology images: The glas challenge contest.* Medical image analysis, 2017. **35**: p. 489-502.
156. Goodfellow, I.J., et al., *Generative adversarial networks.* arXiv preprint arXiv:1406.2661, 2014.
157. Kingma, D.P. and M. Welling, *Auto-encoding variational bayes.* arXiv preprint arXiv:1312.6114, 2013.
158. Isola, P., et al. *Image-to-image translation with conditional adversarial networks*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.
159. Li, C. and M. Wand. *Precomputed real-time texture synthesis with markovian generative adversarial networks*. in *European Conference on Computer Vision*. 2016. Springer.
160. Wang, T.-C., et al. *High-resolution image synthesis and semantic manipulation with conditional gans*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
161. Ounkomol, C., et al., *Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy.* Nature Methods, 2018. **15**(11): p. 917-+.
162. Wu, R., et al. *PQ-NET: A generative part seq2seq network for 3D shapes*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
163. Guan, Y., T. Jahan, and O. van Kaick. *Generalized Autoencoder for Volumetric Shape Generation*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 2020.
164. Liao, Y., et al. *Towards unsupervised learning of generative models for 3d controllable image synthesis*. in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.
165. Cirillo, M.D., D. Abramian, and A. Eklund, *Vox2Vox: 3D-GAN for brain tumour segmentation.* arXiv preprint arXiv:2003.13653, 2020.

166. Wang, T.-C., et al., *Few-shot video-to-video synthesis.* arXiv preprint arXiv:1910.12713, 2019.
167. Chan, C., et al. *Everybody dance now*. in *Proceedings of the IEEE International Conference on Computer Vision*. 2019.
168. Zhou, Y., et al. *Dance dance generation: Motion transfer for internet videos*. in *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2019.
169. Gafni, O., L. Wolf, and Y. Taigman, *Vid2game: Controllable characters extracted from real-world videos.* arXiv preprint arXiv:1904.08379, 2019.
170. Lucas, B.D. and T. Kanade. *An iterative image registration technique with an application to stereo vision*. 1981. Vancouver.
171. Mao, C., et al., *Feature-rich covalent stains for super-resolution and cleared tissue fluorescence microscopy.* Science advances, 2020. **6**(22): p. eaba4542.
172. Reza, A.M., *Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement.* Journal of VLSI signal processing systems for signal, image and video technology, 2004. **38**(1): p. 35-44.
173. He, K., et al. *Deep residual learning for image recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
174. Bulten, W., et al., *Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard.* Scientific reports, 2019. **9**(1): p. 1-10.
175. Kass, M., A. Witkin, and D. Terzopoulos, *Snakes: Active contour models.* International journal of computer vision, 1988. **1**(4): p. 321-331.
176. Soille, P.J. and M.M. Ansoult, *Automated basin delineation from digital elevation models using mathematical morphology.* Signal Processing, 1990. **20**(2): p. 171-182.
177. Xu, H., et al., *Automatic nuclei detection based on generalized laplacian of gaussian filters.* IEEE journal of biomedical and health informatics, 2016. **21**(3): p. 826-837.
178. Zhao, P., et al., *SCAU-Net: Spatial-Channel Attention U-Net for Gland Segmentation.* Frontiers in Bioengineering and Biotechnology, 2020. **8**.
179. Dice, L.R., *Measures of the amount of ecologic association between species.* Ecology, 1945. **26**(3): p. 297-302.
180. Cignoni, P., C. Rocchini, and R. Scopigno. *Metro: measuring error on simplified surfaces*. in *Computer graphics forum*. 1998. Wiley Online Library.
181. Haft-Javaherian, M., et al., *Deep convolutional neural networks for segmenting 3D in vivo multiphoton images of vasculature in Alzheimer disease mouse models.* Plos One, 2019. **14**(3).
182. Hawley, S., 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**(1): p. 39.
183. Lewiner, T., et al., *Efficient implementation of marching cubes' cases with topological guarantees.* Journal of graphics tools, 2003. **8**(2): p. 1-15.
184. Lee, T.-C., R.L. Kashyap, and C.-N. Chu, *Building skeleton models via 3-D medial surface axis thinning algorithms.* CVGIP: Graphical Models and Image Processing, 1994. **56**(6): p. 462-478.

185. Nunez-Iglesias, J., et al., *A new Python library to analyse skeleton images confirms malaria parasite remodelling of the red blood cell membrane skeleton.* PeerJ, 2018. **6**: p. e4312.
186. Mann, H.B. and D.R. Whitney, *On a test of whether one of two random variables is stochastically larger than the other.* The annals of mathematical statistics, 1947: p. 50-60.
187. Van der Maaten, L. and G. Hinton, *Visualizing data using t-SNE.* Journal of machine learning research, 2008. **9**(11).
188. Dobbin, K.K., et al., *Validation of biomarkers to predict response to immunotherapy in cancer: Volume II - clinical validation and regulatory considerations.* Journal for Immunotherapy of Cancer, 2016. **4**.
189. Stephenson, A.J., et al., *Postoperative nomogram predicting the 10-year probability of prostate cancer recurrence after radical prostatectomy.* Journal of Clinical Oncology, 2005. **23**(28): p. 7005-7012.
190. Dunn, K.W., et al., *DeepSynth: Three-dimensional nuclear segmentation of biological images using neural networks trained with synthetic data.* Scientific Reports, 2019. **9**.
191. Liu, Q., et al., *Towards Annotation-free Instance Segmentation and Tracking with Adversarial Simulations.* arXiv preprint arXiv:2101.00567, 2021.
192. Tetteh, G., et al., *DeepVesselNet: Vessel Segmentation, Centerline Prediction, and Bifurcation Detection in 3-D Angiographic Volumes.* Frontiers in Neuroscience, 2020. **14**.
193. Kattan, M.W., T.M. Wheeler, and P.T. Scardino, *Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer.* J Clin Oncol, 1999. **17**(5): p. 1499-507.
194. Cooperberg, M.R., et al., *Multiinstitutional validation of the UCSF cancer of the prostate risk assessment for prediction of recurrence after radical prostatectomy.* Cancer, 2006. **107**(10): p. 2384-91.
195. Newcomb, L.F., 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**(2): p. 313-320.
196. Hamdy, F.C., et al., *10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer.* N Engl J Med, 2016. **375**(15): p. 1415-1424.
197. Leo, P., 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.
198. Wulczyn, E., et al., *Predicting prostate cancer specific-mortality with artificial intelligence-based Gleason grading.* Communications Medicine, 2021. **1**(1): p. 1-8.
199. Nagpal, K., et al., *Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer (vol 2, 48, 2019).* Npj Digital Medicine, 2019. **2**.
200. Epstein, J.I., et al., *The 2005 International Society of Urological Pathology (ISUP) consensus conference on Gleason grading of prostatic carcinoma.* The American journal of surgical pathology, 2005. **29**(9): p. 1228-1242.
201. Gann, P.H., et al., *Development of a Nuclear Morphometric Signature for Prostate Cancer Risk in Negative Biopsies.* Plos One, 2013. **8**(7).

202. Carleton, N.M., et al., *Advances in the computational and molecular understanding of the prostate cancer cell nucleus.* Journal of Cellular Biochemistry, 2018. **119**(9): p. 7127-7142.
203. Wikstrom, P., et al., *Endoglin (CD105) is expressed on immature blood vessels and is a marker for survival in prostate cancer.* Prostate, 2002. **51**(4): p. 268-275.
204. Foroozan, M., et al., *Clinical significance of endothelial cell marker CD34 and mast cell marker CD117 in prostate adenocarcinoma.* Pathol Res Pract, 2017. **213**(6): p. 612-618.
205. Gayed, B.A., et al., *Digoxin inhibits blood vessel density and HIF-1a expression in castration-resistant C4-2 xenograft prostate tumors.* Clin Transl Sci, 2012. **5**(1): p. 39-42.
206. Cheng, L., et al., *Lymphovascular invasion is an independent prognostic factor in prostatic adenocarcinoma.* Journal of Urology, 2005. **174**(6): p. 2181-2185.
207. Shariat, S.F., et al., *Lymphovascular invasion is a pathological feature of biologically aggressive disease in patients treated with radical prostatectomy.* Journal of Urology, 2004. **171**(3): p. 1122-1127.
208. Sigorski, D., et al., *Investigation of Neural Microenvironment in Prostate Cancer in Context of Neural Density, Perineural Invasion, and Neuroendocrine Profile of Tumors.* Frontiers in Oncology, 2021. **11**.
209. Magnon, C., et al., *Autonomic Nerve Development Contributes to Prostate Cancer Progression.* Science, 2013. **341**(6142): p. 143-+.
210. Reeves, F.A., et al., *Prostatic nerve subtypes independently predict biochemical recurrence in prostate cancer.* Journal of Clinical Neuroscience, 2019. **63**: p. 213-219.
211. Bulten, W., et al., *Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge.* Nature Medicine, 2022. **28**(1): p. 154-+.
212. Alilou, M., et al., *Quantitative vessel tortuosity: A potential ct imaging biomarker for distinguishing lung granulomas from adenocarcinomas.* Scientific reports, 2018. **8**(1): p. 1-16.

APPENDIX A: A TRI-LABELING PROTOCOL

Time	Steps
Week 1	
Day 2	Wash samples with PBS (1× phosphate-buffered saline) for 1h, then in 25% methanol (in PBS) for 1h, 50% methanol for 1h, 75% methanol for 1h, and 100% methanol for 1h at room temperature. Samples are then put into 100% methanol for storage.
Day 3	Chill samples at 4°C for 1h, then bleach in 5%H_2O_2 in methanol at 4°C overnight.
Day 4	Wash samples in 75% methanol (in PBS) for 1h, then in 50% methanol for 1h, 25% methanol for 1h, 100% PBS for 1hr at room temperature. For cytoplasm staining, samples are incubated in 5μg/ml Alexa Fluor™ 488 NHS ester (dissolved in pH 5 PBS, where the pH is adjusted with HCl, Cat: A20000, ThermoFisher) at 37°C overnight.
Day 5	Wash samples in PBS/0.2% Triton X-100 for 1h at room temperature twice. Samples are kept in PBS/0.2% Triton X-100 at room temperature.
Week 2	
Day 1	Samples are incubated in PBS/0.2% Triton X-100/20% DMSO/0.3M glycine, at 37°C overnight.
Day 2	Block samples with PBS/0.2% TritonX-100/10% DMSO/6% Donkey Serum/3mM NaN_3, at 37°C overnight.
Day 3	Samples are incubated with the primary antibody (Cytokeratin 8/18 Monoclonal Antibody 5D3, 1:20 diluted, Cat: MA5-14088, ThermoFisher) in PBS/0.2%Tween-20/10μg/ml heparin (PTwH)/5%DMSO/3% Donkey Serum/3mM NaN_3 at 37°C until the next step.
Week 3	
Day 2	Wash samples in PTwH for 15 min, 30 min, 1h and 1h at room temperature. Samples are then kept in PTwH overnight at room temperature.
Day 3	Samples are incubated with the secondary antibody (Alexa Fluor 647 AffiniPure Donkey Anti-Mouse IgG, 1:100 diluted, Cat: 715-605-150, JacksonImmunoResearch) in PTwH/3% Donkey Serum/3mM NaN_3 at 37°C.
Week 4	
Day 1	Samples are incubated in PTwH for 15 min, 30 min, 1h and 1h at room temperature. For nuclear staining, samples are incubated 5uM SYTO™ 85 (in PTwH, Cat: S11366, ThermoFisher) overnight at 37°C.
Day 2	Wash samples with PBS for 1h, then incubate with 25% ethanol (in PBS) for 1h, 50% ethanol for 1h, 75% ethanol for 1h, and 100% ethanol for 1h at room temperature. Samples are then kept in 100% ethanol overnight at room temperature.
Day 3	For optical clearing, samples are incubated in ethyl cinnamate for 30 min 2x at room temperature. Samples are then ready for 3D OTLS microscopy.

APPENDIX B: LIST OF GLANDULAR FEATURES

Feature category	Feature set	3D glandular features (*G: gland* [1]*, E: epithelium, L: lumen, S: stroma)*	3D AUC value	2D glandular features	2D AUC value
Non-skeleton features	Size	*Volume (L) / volume (E)*	0.82	*Area (L) / area (E)*	0.79
		Volume (E) / volume (G)	0.82	*Area (E) / area (G)*	0.78
		Volume (S) / volume (G)	0.67	*Area (S) / area (G)*	0.68
	Compactness	*Surface area (G) / volume (G)*	0.68	*Circumference (G) / area (G)*	0.63
		Surface area (E) / volume (E)	0.57	*Circumference (E) / area (E)*	0.55
		Surface area (L) / volume (L)	0.69	*Circumference (L) / area (L)*	0.75
	Irregularity	*Volume (G) / convex hull volume (G)*	0.72	*Area (G) / convex hull area (G)*	0.62
		Volume (E) / convex hull volume (E)	0.52	*Area (E) / convex hull area (E)*	0.72
		Volume (L) / convex hull volume (L)	0.67	*Area (L) / convex hull area (L)*	0.51
	Boundary Curvature	Average Gaussian curvature (absolute values) of the *G* surface	0.79	Average curvature of the *G* boundary [212].	0.72
		Average Gaussian curvature (absolute values) of the E surface	0.80	Average curvature of the *E* boundary [212].	0.64
		Average Gaussian curvature (absolute values) of the *L* surface	0.84	Average curvature of the *L* boundary [212].	0.74
Skeleton features	Lumen skeleton	Average branch length	0.73	N /A	
		Standard deviation of the branch length	0.71		
		Median tortuosity [2] for all branches	0.65		
		Standard deviation of the tortuosity for all branches	0.6		
		Branch connectivity [3]	0.58		

[1] G = E + L
[2] Tortuosity is defined as: branch length / Euclidian distance between two end points
[3] Branch connectivity is defined as: total number of branches / total number of connected sets of branches

APPENDIX C: RELEVANT CLINICAL PARAMETERS FOR STUDY CASES (N = 50)

Case #	Eligibility	BCR category [1]	Age at diagnosis	Days to recurrence post-RP	Multiparameter model classification based on 2D features [2]	Multiparameter model classification based on 3D (w/o skeleton) features [2]
1	Non-recurrent	0	50-54	N/A	Low risk	Low risk
2	Recurrent (<5 years)	1	65-69	669	Low risk	High risk
3	Recurrent (<5 years)	1	55-59	Data not available	Low risk	Low risk
4	Recurrent (<5 years)	1	55-59	Data not available	High risk	High risk
5	Non-recurrent	0	60-64	N/A	High risk	High risk
6	Recurrent (<5 years)	1	70-74	Data not available	High risk	High risk
7	Recurrent (>5 years)	0	45-49	Data not available	Low risk	Low risk
8	Non-recurrent	0	55-59	N/A	Low risk	Low risk
9	Recurrent (<5 years)	1	70-74	Data not available	High risk	High risk
10	Recurrent (<5 years)	1	45-49	Data not available	High risk	Low risk
11	Recurrent (<5 years)	1	55-59	1491	Low risk	High risk
12	Non-recurrent	0	60-64	N/A	Low risk	Low risk
13	Non-recurrent	0	55-59	N/A	High risk	Low risk
14	Non-recurrent	0	55-59	N/A	Low risk	Low risk
15	Recurrent (>5 years)	0	60-64	1949	High risk	High risk
16	Recurrent (<5 years)	1	50-54	Data not available	High risk	High risk
17	Non-recurrent	0	50-54	N/A	High risk	Low risk
18	Non-recurrent	0	55-59	N/A	High risk	Low risk
19	Recurrent (<5 years)	1	55-59	1553	Low risk	Low risk
20	Non-recurrent	0	60-64	N/A	High risk	Low risk
21	Non-recurrent	0	55-59	N/A	Low risk	Low risk
22	Recurrent (<5 years)	1	70-74	182	Low risk	High risk
23	Non-recurrent	0	70-74	N/A	Low risk	Low risk
24	Recurrent (<5 years)	1	50-54	214	Low risk	High risk
25	Non-recurrent	0	60-64	N/A	High risk	Low risk
26	Non-recurrent	0	65-69	N/A	High risk	High risk
27	Non-recurrent	0	55-59	N/A	Low risk	Low risk
28	Recurrent (<5 years)	1	70-74	1065	High risk	High risk
29	Non-recurrent	0	55-59	N/A	High risk	High risk
30	Non-recurrent	0	55-59	N/A	Low risk	Low risk
31	Non-recurrent	0	50-54	N/A	High risk	Low risk

32	Recurrent (<5 years)	1	55-59	Data not available	Low risk	High risk
33	Recurrent (>5 years)	0	55-59	Data not available	High risk	Low risk
34	Recurrent (<5 years)	1	65-69	457	High risk	High risk
35	Recurrent (<5 years)	1	60-64	Data not available	High risk	High risk
36	Non-recurrent	0	55-59	N/A	Low risk	Low risk
37	Recurrent (<5 years)	1	50-54	1216	High risk	High risk
38	Recurrent (<5 years)	1	55-59	182	High risk	High risk
39	Recurrent (<5 years)	1	45-49	912	High risk	High risk
40	Recurrent (<5 years)	1	60-64	Data not available	High risk	High risk
41	Recurrent (<5 years)	1	65-69	Data not available	High risk	High risk
42	Recurrent (<5 years)	1	60-64	Data not available	Low risk	High risk
43	Recurrent (<5 years)	1	60-64	Data not available	High risk	High risk
44	Recurrent (<5 years)	1	60-64	Data not available	High risk	High risk
45	Recurrent (<5 years)	1	60-64	1096	High risk	Low risk
46	Non-recurrent	0	50-54	N/A	Low risk	High risk
47	Recurrent (<5 years)	1	70-74	Data not available	Low risk	High risk
48	Non-recurrent	0	65-69	N/A	Low risk	Low risk
49	Non-recurrent	0	50-54	N/A	High risk	Low risk
50	Non-recurrent	0	60-64	N/A	Low risk	Low risk

[1] 1 = cases with BCR under 5 years, 0 = all other cases.

[2] This stratification is based on a threshold of 0.5 applied to the multi-parameter classification model’s output (which ranges from 0-1).