Deep learning automated pathology in ex vivo microscopy

Marc Combalia,[1] Sergio Garcia,[1] Josep Malvehy,[1] Susana Puig,[1] Alba Guembe Mülberger,[2] James Browning,[2] Sandra Garcet,[2] James G. Krueger,[2] Samantha R. Lish,[2] Rivka Lax,[2] Jeannie Ren,[2] Mary Stevenson,[3] Nicole Doudican,[3] John A. Carucci,[3] Manu Jain,[4] Kevin White,[5] Jaroslav Rakos,[6] and Daniel S. Gareau[2,6,*]

[1]*Department of Dermatology, Hospital Clinic de Barcelona, Universitat de Barcelona, Barcelona, Spain*
[2]*The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA*
[3]*Ronald O. Pearlman Department of Dermatology, New York University, 550 First Avenue, New York, NY 10016, USA*
[4]*Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA*
[5]*Department of Dermatology, Oregon Health & Science University, 3303 South Bond Avenue, Portland, OR 97239, USA*
[6]*SurgiVance Inc., 310 East 67th Street, New York, NY 10065, USA*
dgareau@rockefeller.edu

Abstract: Standard histopathology is currently the gold standard for assessment of margin status in Mohs surgical removal of skin cancer. Ex vivo confocal microscopy (XVM) is potentially faster, less costly and inherently 3D/digital compared to standard histopathology. Despite these advantages, XVM use is not widespread due, in part, to the need for pathologists to retrain to interpret XVM images. We developed artificial intelligence (AI)-driven XVM pathology by implementing algorithms that render intuitive XVM pathology images identical to standard histopathology and produce automated tumor positivity maps. XVM images have fluorescence labeling of cellular and nuclear biology on the background of endogenous (unstained) reflectance contrast as a grounding counter-contrast. XVM images of 26 surgical excision specimens discarded after Mohs micrographic surgery were used to develop an XVM data pipeline with 4 stages: flattening, colorizing, enhancement and automated diagnosis. The first two stages were novel, deterministic image processing algorithms, and the second two were AI algorithms. Diagnostic sensitivity and specificity were calculated for basal cell carcinoma detection as proof of principal for the XVM image processing pipeline. The resulting diagnostic readouts mimicked the appearance of histopathology and found tumor positivity that required first collapsing the confocal stack to a 2D image optimized for cellular fluorescence contrast, then a dark field-to-bright field colorizing transformation, then either an AI image transformation for visual inspection or an AI diagnostic binary image segmentation of tumor obtaining a diagnostic sensitivity and specificity of 88% and 91% respectively. These results show that video-assisted micrographic XVM pathology could feasibly aid margin status determination in micrographic surgery of skin cancer.

1. Introduction

Skin cancer is more common than all other cancer types combined. Basal cell carcinoma (BCC) is the most common human cancer with incidence exceeding 2,000,000 in the United States each year. As such, BCC is a significant public health burden with regard to morbidity and cost. Though most BCCs are readily treated by surgical resection, a subset grow unchecked, resulting in significant morbidity from massive local tissue damage. Roughly one quarter of

BCC cases are treated by specialty Mohs micrographic surgery, which is a precise, tissue-sparing surgical technique involving progressive removal of cancer-containing skin tissue in staged excisions to ensure complete cancer removal with minimal removal of healthy tissue [1]. There are approximately 2000 Mohs surgeons in the U.S., some of whom perform 500 to 1000 Mohs surgeries per year [2]. Mohs surgery offers the highest cure rate for non-melanoma skin cancer but requires confirmation via frozen section pathology of clear margins at conclusion and micrographic tumor mapping between stages to provide surgical guidance [3].

Current standard processing methods require the preparation of frozen section slides, which can introduce sectioning artifacts and take 60 minutes for a single stage/serial excision. To avoid idle surgical time, patients often wait in crowded waiting rooms between stages while the surgeon moves on to the next patient, forcing surgeons to re-situate and re-anesthetize patients between stages and reacquaint themselves with each case. This takes up to 120 minutes under current workflows. The preparation of physical microscope slides with hematoxylin and eosin (H&E) staining alone costs $30 million/year [4]. Frozen tissue sectioning introduces undesirable ice artifacts, including epidermal vacuolization, splaying of collagen in the dermis, loss of cellular architecture, and tissue shrinkage by 11%, [5–7] leading to uncertainty regarding specimen completeness. Next, the frozen tissue must be sectioned, introducing additional artifacts such as tissue folds, transposed epidermis, and dark bands caused by knife chatter, all of which make delineating tumor from normal tissue more difficult [8]. Additional time is then needed to prepare and stain slides. In >80% of cases, tissue is stained with H&E, [9] although toluidine blue may also be used, [10] and specialized stains are gaining in popularity for select cases [11]. H&E slide creation turns precious tissue, that could be banked, into physical glass slides, requiring storage. Finally, tissue slides must be evaluated by a Mohs surgeon, who is a trained pathologist. Mohs surgeons could augment their training by learning standardized, AI-driven auto-pathology derived from XVM. The combination of procedural steps described above results in a lengthy process, taking as long as 30 minutes for a single excision. Procedures are even longer with multiple specimens when a patient undergoes surgery for two or more tumors on the same day or when multiple specimens must be prepared from the same lesion. In the case of staged surgeries like Mohs surgery, in order to avoid idle surgical time, patients often wait with other patients, all receiving multiplexed surgeries that can lead to surgeon confusion and detract from the level of micrographic precision possible.

Additional time to situate, anesthetize, and prep each patient between stages and the need for the surgeon to reacquaint themselves with each case all limit surgical precision. Because of these time constraints, a typical Mohs patient will wait an extra 2 to 3 hours due to tissue processing time alone, and surgery to remove a 5-stage tumor generally takes 5 hours in real time. Expedited pathology is not only needed to decrease the time a surgical wound is open, facilitating sterility, but advanced pathological imaging modalities are needed to bring the image information content up to a level that can support machine learning needs. Currently, the accuracy of pathology is limited by slide preparation procedures such as freezing and sections, which introduce artifacts, as well as the 2D nature of the slides/images.

Despite synergy between digital imaging and integration with the electronic medical record, few novel diagnostic imaging technologies translate into mainstream clinical medicine due to interpretation difficulties. Clinical and surgical pathologists train visually on slides stained with H&E, with more hardware and software needed to perform digital and 3D analyses. Though standard histopathology slides can be digitally imaged, this is limited to few sections, because of slow processing time. There is a clear need for new rapid, point-of-care pathology solutions that can process digital images for simple, standardized, accurate interpretation, reducing undesirable processing artifacts and perioperative complications, and enhancing the accuracy of surgical outcome [12]. Attempts to digitize an analog process via whole-slide scanning are an inefficient solution. In addition, perioperative complications are strongly correlated with the duration of

surgery [13]. New technologies that reduce the total operative time without compromising outcomes would be valuable to both patients and providers, but confocal microscopy has previously been too complex and visually uninterpretable. These factors complicating adoption previously inhibited clinical translation of confocal microscopy, given that histopathology is already fairly accurate for common skin cancers such as BCC. For visually trained pathologists evaluating common skin cancers, previous confocal technologies demonstrated average sensitivity of 98.6 percent and average specificity of 90.70 percent [14–16], but visual interpretability is needed to drive adoption by all pathologists.

Ex vivo confocal microscopy (XVM), is a revolutionary series of newly invented, miniature, highresolution pathology imaging devices that image near the surface of excised specimens. [17] [18]. Figure 1 shows XVM of a typical Mohs surgical excision margin from this study, using the original XVM [19] for colorizing algorithm. Multimodal, colorized images (e.g. Figure 1) can feed machine learning to produce automated diagnostics. Their contrast includes fluorescent nuclear/immunohistochemical labeling and the endogenous reflectance of fresh (not frozen or fixed) tissues. The power of AI and the richness of multimodal XVM image data enable both easy visual sensory decoding during pathology reading and pathology suggestions generated by the AI, such as tumor positivity maps. Scanning time to acquire XVM image data on a typically sized Mohs specimen ranges from 2 minutes [20] to 17 minutes average, across the data presented in this study.
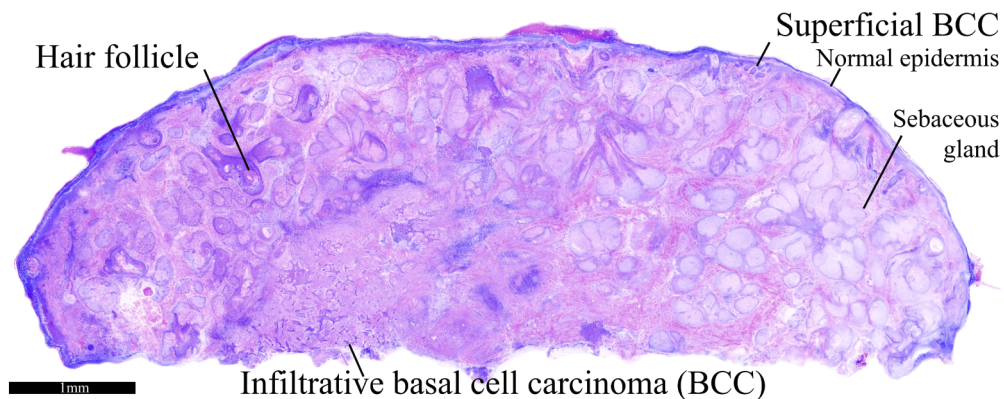

Fig. 1. XVM enables pathology by digitally combining [19] fluorescence imaging with cellular labeling by acridine orange [15] (purple) and endogenous reflectance contrast (pink) to reveal morphological features used in pathology. For a high-resolution version of this image, please see Visualization 1.

XVM may enable point-of-care pathology while enhancing the accurate detection of residual tumor and improving patient outcomes with 1) better margin control via enriched 3D information content and simplified specimen orientation maintenance, decreasing error and better functional outcomes; 2) decreased duration of open surgical wounds, reducing the rate of complications; and 3) a compact, durable, and inexpensive form factor, eliminating the need for bulky and expensive equipment and for tissue transport to a pathology laboratory. We present proof of principle for the two remaining components of clinical translation of XVM: visual transformation of image appearance to resemble standard histology and automated pathological diagnosis on a pixel-by-pixel basis that identifies segments of cancer (BCC in this proof of principal) positivity. The first is a human criteria of reference and the second is a data science connection that will potentially empower medical professionals to utilize AI.

2. Methods

As the most common form of skin cancer, BCCs are a natural choice for automated detection of skin cancer. Pathologists must examine pathology images to diagnose BCC, potentially resulting in delay, error, and inconsistency. To address the need for standardized, expedited diagnosis, we created an automated diagnostic AI to identify BCC in pathology images. We acquired a dataset of BCC XVM images and created gold standard masks using a MATLAB labeler that we created to label ground truth tumor maps on the colorized XVM images. We adapted a neural network image segmentation model to train on the dataset and their corresponding masks, which learns to highlight these nodules in XVM pathology images by predicting a computer-generated, tumor-identifying binary mask.

To stylize XVM images for display like standard H&E staining, we applied a Cycle Consistency Generative Adversarial Network (CycleGAN) to the XVM images. It performed a style transfer for visual interpretation. We also designed an automated diagnostic method to identify BCC that trained a deep neural network image segmentation model, U-Net, to segment BCC nodules via supervised AI on ground truth masks. Figure 2 shows the general outline of our method.
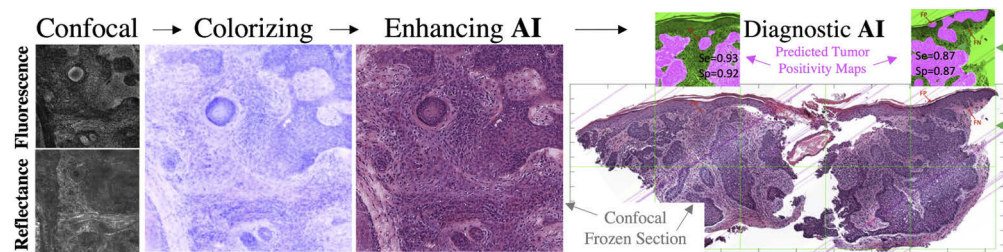
Fig. 2. Data pipeline consisting of 3 methods: (1) reflectance and florescence confocal micrographs are colorized, (2) the XVM image is enhanced using AI and (3) AI is used to generate binary image segmentation that identifies tumor positivity. This data pipeline supports both point scanning and line scanning multimodal (reflectance and fluorescence) images injected at the input and supports both digitally colorized XVM images, digitally-colorized & AI-enhanced XVM images and standard frozen or fixed section digital pathology images for injection into the diagnostic AI. Each of two test patches show that the sensitivity (Se) and specificity (Sp) of BCC detection are high.

Figure 2, which is designed to show the possible variations, includes our image enhancement AI implemented on a colorized confocal XVM image and an illustration of our diagnostic AI on a standard frozen section pathology slide. For the latter, the two tumor positivity maps were each generated using only the single frozen section shown, illustrating excellent performance on a relatively easy task with a small taring set. Each positivity map (one on each of two patches immediately below the map) was a produced by a single training run that used the other 8 square patches as training data and the single patch as a test image. Below, we adapt this approach illustrated here on a single frozen section image with 9 patches to 26 XVM images and 5359 patches.

2.1. Data acquisition and stack collapsing to optimize fluorescence contrast over lateral (xy) dimensions

An RSG4 Confocal Microscope (Caliber ID, Rochester NY), equipped with laser sources of 488nm and 532nm wavelengths, and an objective lens for both reflected and fluorescent light, which it separated and directed to a reflectance detector and a fluorescence detector, respectively. This setup was used to image specimens and resolve morphologic features of BCC and normal skin. Surgical specimens, discarded during surgeries at New York University, were stained with

acridine orange and imaged at The Rockefeller University under IRB approval from New York University and The Rockefeller University using a previously published protocol. [21]

Due to sample surface irregularity, laterally separated points on the tissue block face showed maximum cellular fluorescence contrast at K different Z depths. The stained cells of interest thus formed a 3D manifold within the imaged space, and full 3D images were acquired. To circumvent the need for a pathologist to analyze all images in a stack for a single case, we created a MATLAB-based algorithm to generate a single composite image by combining the highest-contrast lateral surface area elements within a Z-stack, projecting the manifold onto a 2D image. For imaging in optically-turbid tissues like skin, simpler approaches such as maximum intensity or summing lead to poor visual contrast. Various approaches have been previously described for the 2D projection of 3D manifolds – particularly for the study of single layer epithelial tissues – such as StackFocuser [22], PreMosa [23] and Smooth Manifold Extraction [24]. In high-contrast image segments, intensity difference is higher when diagonally shifted by one pixel. By maximizing these intensity differences, regions of maximal contrast from different Z-stack layers are selected and stitched together to laterally form a mosaic with uniform high contrast across its surface. This algorithm thus outputs a single image that is a mosaic selected throughout the input Z-stack. Following a similar sectioning of the 3D stack to the Fiji-based tool PreMosa, our MATLAB-based algorithm selects the Z-sections of interest based solely on local contrast. In high-contrast image segments, intensity difference is higher when diagonally shifted by one pixel. By maximizing these intensity differences, regions of maximal contrast from different Z-stack layers are selected and stitched together to laterally form a mosaic with uniform high contrast across its surface.

Images obtained at different z-planes were each loaded into MATLAB as a 3D matrix, with each pixel's relative intensity described by a 12-bit integer. the image stack – in this case the 3D matrix – was thus created by combining the images as layers in a multidimensional array. The x-y plane was defined by image size, and the number of matrices across the z (the number of layers) equaled the number of loaded files N. N was typically 8 images taken in 5μm increments in the z-direction. The 5μm optical section Z-spacing, which is more sparse than Nyquist sampling, was chosen to keep the overall stack imaging acquisition time short, while ensuring sampling all cells, since skin cells are at least 5μm in size. The average imaging time for N = 8 Z optical sections was 17 minutes.

$$stack = [a_{i,j,k}] = [[a_{i,j,1}], [a_{i,j,2}], \ldots [a_{i,j,K}]] \tag{1}$$

$$Image1 = [a_{i,j,1}],$$
$$Image2 = [a_{i,j,2}]$$
$$\ldots$$
$$ImageN = [a_{i,j,K}]$$

A square surface, i.e. "window," was established of a given side length w_1 = 3 pixels = 1.5 μm that represented a greater size than the a single pixel (so as to ease computational time) and a smaller size than a cell (so as to make the algorithm visualize whole cells continuously). In each lateral window image element of 9 pixels, the same element was evaluated throughout the Z stack to determine which Z position had the best nuclear fluorescence contrast in order to pick that Z location for the final image within that element. The evaluation of fluorescence contrast intensity was performed over a larger window with the same X-Y center as w_1, w_2 = 40 pixels = 20 μm. 20 μm was chosen to evaluate fluorescence contrast because it is at least the half width of the largest (bright) cells in the skin. This window is maximally sensitive to fluorescence contrast, where any given spatial orientation of the window will safely include some bright area inside the cell and some dark area outside the cell.

A triple for loop sampled the XVM in w_1-sized increments across the first (X) and second (Y) dimensions. At each w_1 x w_1 window, w_2 x w_2 x K submatrix was sampled from the image stack. All w_1 x w_1 sampled matrices accounted for the totality of the stack with the exception of the border regions within 20 μm from the outer image border.

The overall processing time on our Windows 10 PC, with an Intel Core i7-8700 CPU running at 3.2 GHz and with 64 GB RAM, for this data transformation was 3-5 minutes for specimens of average size (e.g. Fig. 1).

$$sampled = stack(i : i + w_1, j : j + w_1, k) \qquad (2)$$

was defined iteratively within a triple loop for all three dimensions, using the variables i, j and k;

$$\text{1st dimension loop: for } i = 1 : w_1 : rows - w_1$$

$$\text{2nd dimension loop: for } j = 1 : w_1 : cols - w_1$$

$$\text{3rd dimension loop: for } k = 1 : K$$

Which can be mathematically defined by the following arrays:

$$i = a_n : \{1 + (n - 1)w_1\} \quad n \in ([1, \frac{rows - 1}{w_1}] \cap \mathbb{Z}) \qquad (3)$$

$$j = a_{n'} : \{1 + (n' - 1)w_1\} \qquad n' \in ([1, \frac{cols - 1}{w_1}] \cap \mathbb{Z})$$

$$k = \{1, 2, \ldots K\}$$

The sampled submatrix being:

$$sampled = [b_{i,j,k}] = [[b_{i,j,1}], [b_{i,j,2}], \ldots [b_{i,j,K}]] \qquad (4)$$

Each submatrix sampled from the z-stack underwent a diagonal shift by one element pixel in the xy-plane, as a result of two circular shifts of one pixel in the 1st and 2nd dimension. The function "circshift" was used:

$$shiftedSampled = circshift(sampled, 1, 1) \qquad (5)$$

$$shiftedSampled = circshift(shiftedSampled, 1, 2)$$

$$shiftedSampled = [c_{i,j,k}] = [[c_{i,j,1}], [c_{i,j,2}], \ldots [c_{i,j,K}]]$$

To create a gradient image, the intensity difference was measured for every pixel pair with the same X-Y-Z coordinates between the shifted and non-shifted matrices, and collapsed as a total sum of absolute differences into a matrix. The detection of maximal contrast was gradient-based, following the assumption that high-contrast images present sharper edges and higher contrast. Therefore, the absolute sum of intensity differences would always be higher in higher-contrast areas. The following line of code accounts for the mathematical operation:

$$diffSampled(k) = sum(sum(abs(sampled(k) - shiftedSampled(k)))) \qquad (6)$$

$$diffSampled = \sum_{j=1}^{N_x} \sum_{i=1}^{N_y} |sampled - shiftedSampled| = \sum_{j=1}^{N_x} \sum_{i=1}^{N_y} |[b_{i,j,k}] - [c_{i,j,k}]| = [d_{1,1,k}] \qquad (7)$$

The Z-plane with the highest contrast was identified as the layer with the highest sum value. The "max" function was used within the loop, which returned the maximum value M and the corresponding index I, the layer of the 3D matrix in which M is located. Although fluorescence contrast was quantified using the 20 μm x 20 μm w_2 window, the 1.5 μm x 1.5 μm w_1 window

(with the same x-y center coordinates) was selected as the Z representation for the final image. The final composite image was formed by incorporating the maximal contrast layer of each sampled window into a bi-dimensional matrix.

$$[M, I] = max(diffSampled) \tag{8}$$

$$FlattenedImage(i : i + w_1, j : j + w_1) = stack(i : i + w_1, j : j + w_1, I) = [e_{i,j}] \tag{9}$$

Dark-field fluorescence and reflectance images are converted [19] to bright-field and combined into one fusion image. The fluorescent signal from acridine orange is transformed into a purple color gradient, mimicking the hematoxylin stain. Reflectance signal is colored pink to resemble eosin. After colorization and transformation of the input Z-stack into a single image as described above and shown in Fig. 3(a)-(b), the reflectance artifact from the glass/water interface surrounding the sample was mostly removed by masking (Fig. 3(a) and (c)). The mask to remove the reflectance artifact was generated by thresholding the fluorescence part of the composite image at a very low threshold to clearly delineate areas of tissue, where there is some background fluorescence due to ubiquitous, low-level acridine orange staining and outside this area, there is no florescence at all because the reflecting (artifact) surface is clean of fluorescence completely. Figure 3(c) shows the effect of some pooled fluorescent liquid on the bottom right of the sample where this mask failed to exclude the non-tissue area where some acridine orange had pooled.

2.2. AI for image enhancement

For display, a cycle-consistent generative adversarial network (CycleGAN) [25] was used to perform style transfer from colorized XVM images (domain A) to natural H&E images (domain B) using the technique we previously reported [26].

We trained the style-transfer CycleGAN on 759 XVM patches from several XVM slides and 282 histology patches extracted from a single slide (the one with the best proportion of hematoxylin and eosin stains). The CycleGAN was trained with Adam optimization and a learning rate of 2e - 4. It consisted of two generator and discriminator pairs. The first pair tries to map images from domain A to domain B, while the second pair undergoes the contrary operation. The generators' task is to create images that the discriminators are unable to distinguish from real samples. In this work, we use a ResNet [27] architecture in the generators and a PatchNet [28] in the discriminators.

2.3. AI for diagnosis

We trained the BCC segmentation architecture on 26 XVM whole slide-images (e.g. Figure 1) divided into patches of size 1128 x 1128 pixels (Fig. 4, right) large enough to ensure the inclusion of whole morphology such as epidermis, hair follicle and BCC tumor. Each patch, of 5359 total, was paired with a manually-created binary mask delimiting malignant regions containing BCC. Each patch (red square, Fig. 2(b)) was fed through MATLAB-based display and capture by the engineer (Co-Author Daniel Gareau) in consult with the confocal pathologists (Co-Authors John Carucci and Manu Jain). Tumor labeling was manual and included only the solid body of BCC tumors. When blank voids in the fresh-tissue XVM appeared to resemble "tumor clefting" in standard H&E, these areas just adjacent the tumor bodies were included in the tumor label. The dataset was imbalanced, presenting only 243 patches containing BCC, which represented 4.5 % of the total number of pixels in the study. We provide the image and corresponding label data set as supplementary material to this report (Dataset 1 [29]).

We used the U-Net architecture [30] to segment BCC regions in the XMV images. An EfficientNet-B0 [31], which had been previously pre-trained on ImageNet [32], was used as the encoder path of the network. Data augmentation to reduce over-fitting and improved generalization included rotation, width and height shift, shear, zoom, horizontal flip, and color
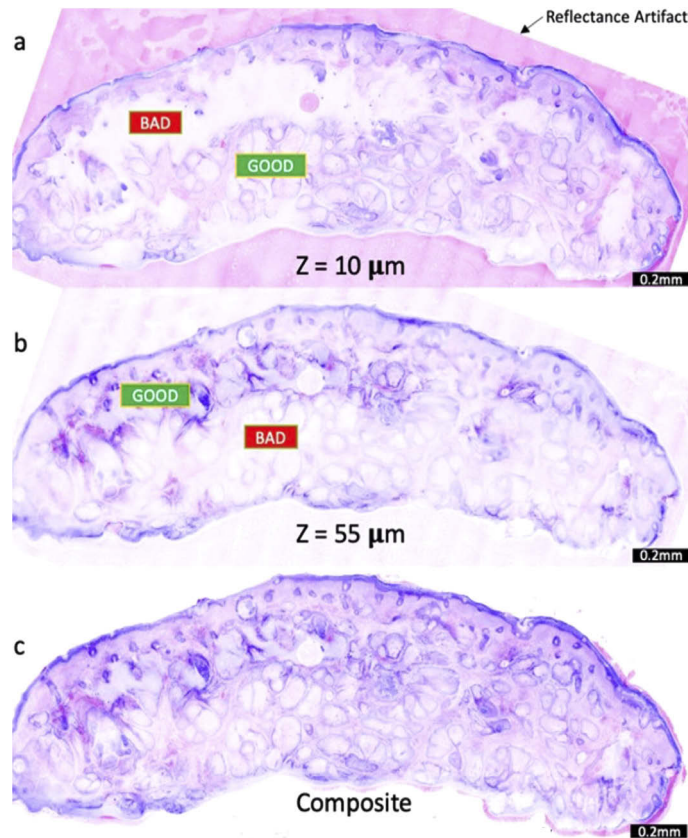
Fig. 3. Visual illustration of the results of image processing (Eqs. 1–9) for collapsing ex vivo microscopy image stacks: (a) and (b) show different optical sections with complementary regions of maximum fluorescence contrast and (c) shows the result of collapsing and reflectance artifact removal.

augmentations. An Adam Optimizer with a learning rate of 1e-4 was used to train the model, reducing the learning rate by a factor of 0.1 when the validation loss stopped improving for more than 5 epochs. Early stopping was used to stop the training loop once the model had converged.

3. Results

We found that style-transformed XVM images (Fig. 5) were strikingly similar to conventional pathology, greatly increasing the rate of image interpretation from a human perspective. This may be worth the computational cost (e.g. 15000 x 10000 px XVM image can be transformed in less than 3 minutes on a NVIDIA Tesla K80 using PyTorch).

The diagnostic performance of the UNET automated pathology segmentation is reported in Table 1 using K-Fold (k=5) cross validation to divide the whole slide images into train and test splits of 80% and 20% images respectively.

XVM is a fast alternative procedure for the processing and imaging of surgically-excised skin tissue for pathological evaluation. Areas of high contrast are combined from images layers taken at different focal Z-planes to form a single image with full coverage (Fig. 3). Dark-field, grey-scale reflectance confocal images are transformed via digital staining to bright-field, mimicking standard histology. The nuclear fluorescent signal is set to a purple color scale and cytoplasmic and
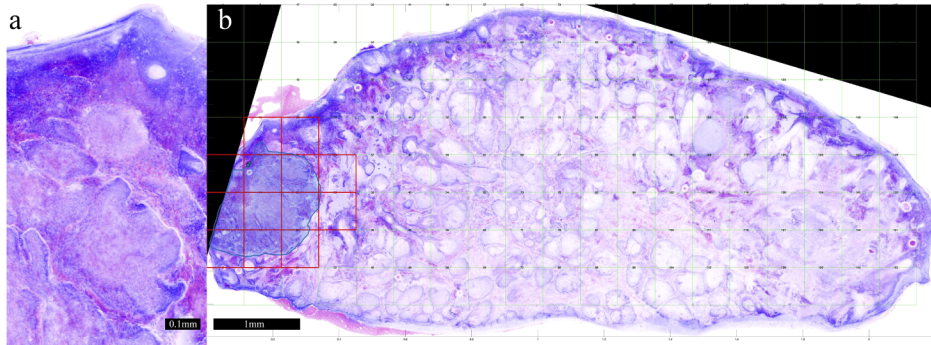
Fig. 4. (a) 10x Viewing enabled tumor (eg. bottom left) identification. Each image was divided into patches (b) where all patches were labeled normal except patches containing the manually delimited tumor region (blue line and shading). Each tumor containing patch (red outline) was manually labeled to create tumor positivity delimiters that contoured to tumor nests.
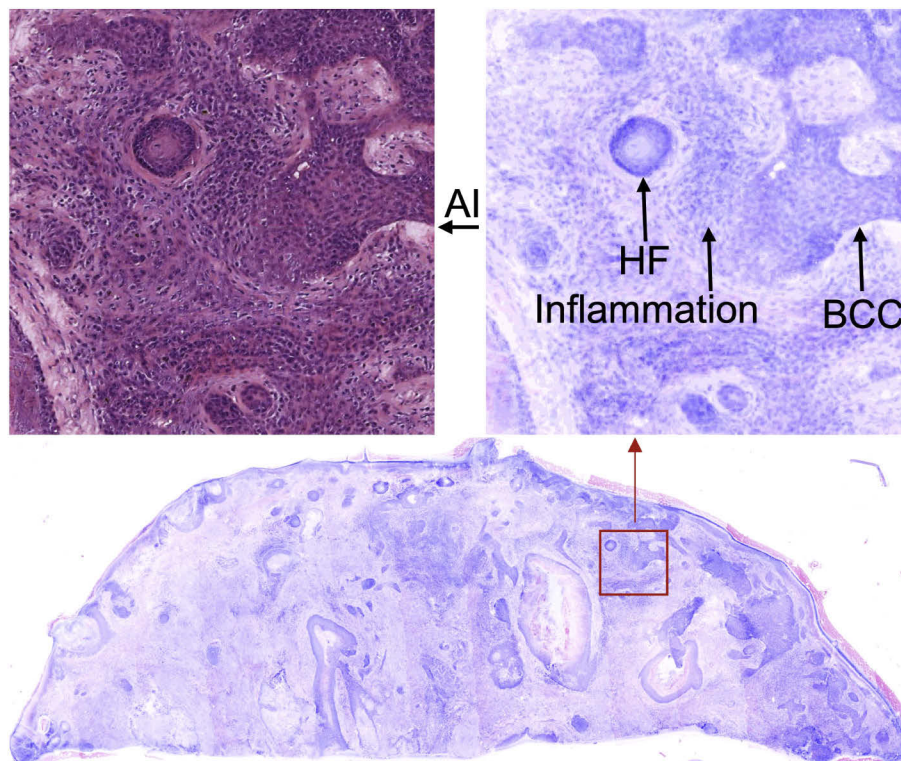
Fig. 5. AI-enhanced XVM image (top left) colorized by CycleGAN as extracted from the original XVM colorized image (bottom) shows a hair follicle (HF), basal cell carcinoma (BCC) and inflammation in the dermis.

Table 1. Pixel sensitivity, specificity and balanced accuracy scores for segmentation model (K-Fold Cross Validation K = 5). A total of 5359 patches from 26 images were divided into 80% dedicated to training and 20% dedicated to testing to produce these statistics. Each patch that contained both tumor and normal tissue produced a single sensitivity and specificity by characterizing each of its pixels as true positive, true negative, false positive or false negative and then following the standard calculation for sensitivity and specificity using the total numbers of those four diagnosis pixel types. The mean values for sensitivity and specificity are shown plus or minus the standard deviation. The balances accuracy was calculated using all patches, even when no tumor positivity was present.

Scoring Metric	Train	Test
Sensitivity	0.98 +/- 0.01	0.878 +/- 0.11
Specificity	0.93 +/- 0.01	0.914 +/- 0.05
Balanced Accuracy	0.96 +/- 0.01	0.887 +/- 0.05

extracellular features from the reflectance signal are colored pink, thus resembling the coloration of hematoxylin and eosin. Reflectance artifacts from outside the tissue sample are significantly reduced by masking. The final image displays cellular resolution for pathological assessment. No additional specialized training is required for analysis of these images by pathologists, as the digital stain provides a H&E appearance that allows them to be analyzed analogously to standard histology, as the CycleGAN architecture for stain transfer is capable of producing realistic H&E-like images.

This is the first step in creating a standardized AI approach for BCC diagnosis. The next steps involve multi-class segmentation, such as differentiation between dermis and background or other skin cancers and model generalizability between tissues and microscope types. This is a surgeon interface that supports rapid pathological assessment and AI diagnostics for pathological features, expediting and standardizing the BCC diagnosis process. The methods described in this paper will undergo clinical validation and their diagnostic accuracy will be tested in future work.

4. Discussion

Collapsing 3D XVM creates a composite image without dropout for rapid margin assessment compared to the time required to thaw, re-embed, re-freeze, recut and stain deeper sections. Thus, digital XVM could provide a distinct time and sensitivity advantage to patients who currently often wait in crowded waiting rooms between stages while the surgeon moves on the next patient, with limited ability to practice recommended social distancing during the current COVID-19 pandemic. Since histopathological processing methods prolong operative times, they decrease the overall patient experience and increase the risk of perioperative complications, including skin infections, bleeding or hematoma, wound dehiscence (disruption of recently repaired wounds), tissue necrosis, and pain. [33] Though Mohs surgeries typically produce clean wounds, these issues are more extreme in wounds in the gastrointestinal tract at an increased risk of becoming infected, where XVM will also likely translate.

The platform can be applied to allow a surgeon in the operating room to obtain pathological consult instantaneously, from an expert located half a world away. Digitization also paves the way to tap into AI algorithms as we have demonstrated with the ability to use our system not only to image but also to diagnose BCC. This platform will serve the surgeon of the future by changing

present day workflow patterns. In Mohs, the surgeon will be able to excise, process, and obtain an instantaneous diagnostic answer about the specimen in real time to decide on further excision vs. wound repair. This will result in increased patient satisfaction, shorter wait times, and decreased risk for infection or bleeding as overall visit times decrease. The XVM platform will lead to an enhanced patient experience, increased rates of cure and decreased rates of surgical morbidity.

4.1.   Translation of XVM

A particularly striking feature of the style-transformed XVM is the nuclear and cellular detail (e.g. Figure 5 top left vs. Figure 5 top right). This raises the question of whether such detail is appropriate for auto-pathology (e.g. Figure 6(c)) or for standard human visual pathology, or both, or neither. Another question is whether such detail increases or decreases the diagnostic accuracy, and that question must be addressed in the context of the ground truth source, whether it be hand labeling as done here or more robust methods like genomics profiling.
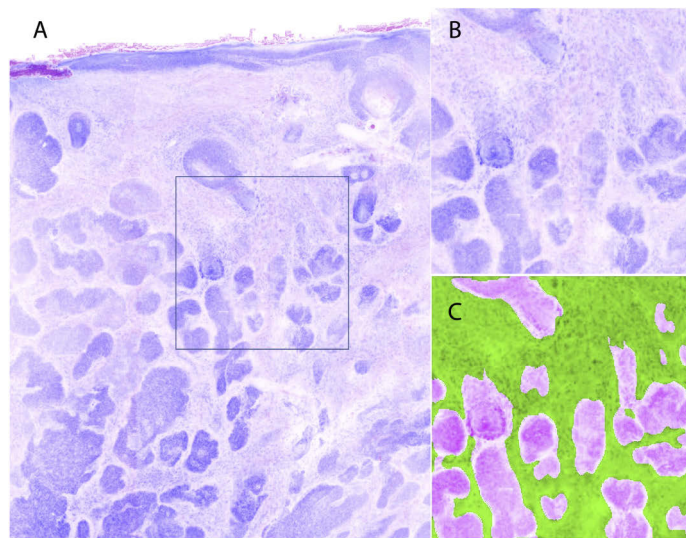
Fig. 6. Application of AI diagnosis to ex vivo microscopy (XVM): (a) XVM of micronodular basal cell carcinoma (b) Zoomed region of the XVM (c) segmentation by the UNET Architecture

On one hand, the AI-added detail does not come from the particular specimen being imaged so it may be unethical to diagnose a case medically, particularly if there is a discrepancy between the phenotype trained on by the AI and the phenotype of the patient.

On the other hand, AI can potentially predict and visualize things through associations. An example might be AI favoring infiltrating leukocyte display when the reflectance shows the collagen patterns (ultra-structural, stromal) resulting from metal metalo-protease remodeling. This could, for instance, correct insufficient fluorescent nuclear stain and rescue readable purple cellular contrast. If accurate, enhanced cellular detail would facilitate diagnosis of tumor for squamous cell carcinoma with poor histologic differentiation [34,35]. Poorly differentiated squamous cell carcinoma (SCC) shows higher rates of margin positivity [36] and higher likelihood of poor outcome in the form of metastatic events and ultimately, disease specific death. Thus, accurate margin interpretation is crucial and yet potentially more difficult in deeply invasive, poorly differentiated SCC. Kinoshita et al. [34] noted cytological features such as a streaming arrangement, a necrotic background, nucleolar enlargement and cannibalism are useful indicators

for the diagnosis of SCC of the breast. Increased resolution and enhanced visualization of cellular detail will render evaluation of subtle nuclear features of SCC.

4.2. Clinical translation

XVM [17,18] is potentially faster, less costly and inherently 3D/digital compared to histopathology as a standardized, medical diagnostic for the 9,500 Americans diagnosed with skin cancer every day [37]. AI helps XVM bridge translational gaps in histologic diagnosis. Reflectance mode confocal microscopy offers a more detailed view of the tumor being studied. This is analogous to a high resolution view offering enhanced visualization via the reflectance mode of XVM, whereby small amounts of tumor (not readily accessible with standard histopathology) can be detected [38] as surgeons dissect their way through normal stroma. In the case of poorly differentiated squamous cell carcinoma, which may evade detection on conventional H&E without additional keratin stains [39], this may be life-saving considering that SCC eventuates in 10,000 deaths annually in the United States [40–42]. The inherent 3D nature of XVM offers advantages regarding complete evaluation of the histologic specimen. In this work we use the 3rd dimension for completeness of the lateral margin surface area and good cellular imaging therein. In future work, landmark detection for all cells in 3D will likely enable more data-rich diagnostic assessment. Thus the two value propositions for clinical translations are that XVM offers to reduce cost, morbidity and mortality within today's existing system and provides an avenue to more advanced digital pathology systems in the future.

Mohs surgery is predicated on evaluation of 100% of the epidermal and deep surface of a tissue specimen [43]. This eliminates false negative evaluations obtained through standard bread loafing routinely performed in pathology laboratories. En face sectioning in the pathology lab puts several degrees of separation between the surgeon and the appointed areas of positivity, which can contribute to greater likelihood of error not encountered in Mohs, since the surgeon excises the tissue, prepares it grossly and maps it prior to having slides made and interpreted by that surgeon with no degrees of separation added. Evaluation of the entire epidermal and deep surface is sometimes difficult and deeper sections may be required (so-called recuts). Here, another advantage of XVM becomes clear, however, several steps are needed to develop XVM into a clinical diagnostic. Figure 7 shows a potential development path.
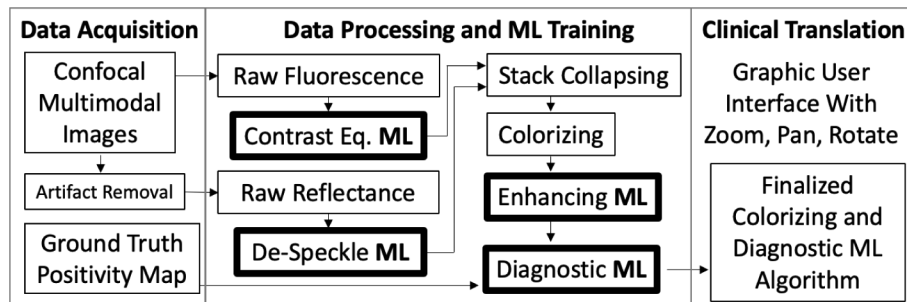
Fig. 7. The workflow and process for optimum XVM clinical translation includes three phases: 1) a data acquisition phase where XVM images and correlating ground truth maps are obtained for a particular disease (e.g. BCC), 2) a data processing and machine learning phase where first the raw confocal images are conditioned by despecking (reflectance) and contrast equalization (fluorescence) and then the machine learning is trained, combining the XVM data and the ground truth positivity maps to form a classifier that is able to output predicted positivity maps given new XVM input, and 3) a user interface that combines the H&E digitally stained image with indications of tumor positivity and enhanced visualization vie zoom, pan and rotate with a touch screen.

Disclosures. Dr. Gareau is an inventor of a patent with claims related to the colorization of confocal images to mimic the appearance of histopathology and holds equity in a company that holds an exclusive license to that patent.

References

1. D. L. Shriner, D. K. McCoy, D. J. Goldberg, J. Wagner, and R. F. Wagner, "Mohs micrographic surgery," J. Am. Acad Dermatol. **39**(1), 79–97 (1998).
2. R. M. Campbell, C. S. Perlis, M. K. Malik, J. Dufresne, and R. G. Raymond, "Characteristics of Mohs practices in the United States: a recall survey of acms surgeons," Dermatol. Surg. **33**(12), 1413–1418 (2007).
3. A. K. El Tal, A. E. Abrou, M. A. Stiff, and D. A. Mehregan, "Immunostaining in Mohs micrographic surgery: a review," Dermatol. Surg. **36**(3), 275–290 (2010).
4. T. L. Group, "The burden of skin disease," Expert Rev. Pharmacoecon Outcomes Res. **9**(3), 271–283 (2009).
5. Q. L. Erickson, T. Clark, K. Larson, and T. Minsue Chen, "Flash freezing of Mohs micrographic surgery tissue can minimize freeze artifact and speed slide preparation," Dermatol. Surg. **37**(4), 503–509 (2011).
6. Y. Rosen and S. C. Ahuja, "Ice crystal distortion of formalin-fixed tissues following freezing," Am. J. Surg. Pathol. **1**(2), 179–182 (1977).
7. E. S. Gardner, W. T. Sumner, and J. L. Cook, "Predictable tissue shrinkage during frozen section histopathologic processing for Mohs micrographic surgery," Dermatol. Surg. **27**, 813–818 (2001).
8. E. B. Desciak and M. E. Maloney, "Artifacts in frozen section preparation," Dermatol. Surg. **26**(5), 500–504 (2000).
9. O. Arnon, R. P. Rapini, A. J. Mamelak, and L. H. Goldberg, "Mohs micrographic surgery: current techniques," Isr. Med. Assoc. J. **12**, 431–435 (2010).
10. D. A. Davis, D. M. Pellowski, and C. William Hanke, "Preparation of frozen sections," Dermatol. Surg. **30**, 1479–1485 (2004).
11. C. J. Miller, J. F. Sobanko, X. Zhu, T. Nunnciato, and C. R. Urban, "Special stains in Mohssurgery," Dermatol. Clin. **29**(2), 273–286 (2011).
12. A. Bennassar, A. Vilata, S. Puig, and J. Malvehy, "Ex vivo fluorescence confocal microscopy for fast evaluation of tumour margins during Mohs surgery," Br. J. Dermatol. **170**(2), 360–365 (2014).
13. B. J. Daley, W. Cecil, P. C. Clarke, J. B. Cofer, and O. D. Guillamondegui, "How slow is too slow? correlation of operative time to complications: an analysis from the Tennessee Surgical Quality Collaborative," J. Am. Coll. Surg. **220**(4), 550–558 (2015).
14. J. K. Karen, D. S. Gareau, S. W. Dusza, M. Tudisco, M. Rajadhyaksha, and K. S. Nehal, "Detection of basal cell carcinomas in mohs excisions with fluorescence confocal mosaicing microscopy," Br. J. Dermatol. **160**(6), 1242–1250 (2009).
15. D. Gareau, A. Bar, N. Snaveley, K. Lee, N. Chen, N. Swanson, E. Simpson, and S. Jacques, "Tri-modal confocal mosaics detect residual invasive squamous cell carcinoma in mohs surgical excisions," J. Biomed. Opt. **17**(6), 066018 (2012).
16. E. W. Mu, J. M. Lewin, M. L. Stevenson, S. A. Meehan, J. A. Carucci, and D. S. Gareau, "Use of digitally stained multimodal confocal mosaic images to screen for nonmelanoma skin cancer," JAMA Dermatol. **152**(12), 1335–1341 (2016).
17. J. Malvehy, J. Perez-Anker, A. Toll, R. Pigem, A. Garcia, L. L. Alos, and S. Puig, "Ex vivo confocal microscopy: revolution in fast pathology in dermatology," Br. J. Dermatol. **183**(6), 1011–1025 (2020).
18. S. Krishnamurthy, J. Q. Brown, N. Iftimia, R. M. Levenson, and M. Rajadhyaksha, "Ex vivo microscopy: a promising next-generation digital microscopy tool for surgical pathology practice," Arch. Pathol. Lab. Med. **143**(9), 1058–1068 (2019).
19. D. S. Gareau, "Feasibility of digitally stained multimodal confocal mosaics to simulate histopathology," J. Biomed. Opt. **14**(3), 034050 (2009).
20. D. S. Gareau, J. G. Krueger, J. E. Hawkes, S. R. Lish, M. P. Dietz, A. G. Mulberger, E. W. Mu, M. L. Stevenson, J. M. Lewin, S. A. Meehan, and J. A. Carucci, "Line scanning, stage scanning confocal microscope (lssscm)," Biomed. Opt. Express **8**(8), 3807–3815 (2017).
21. D. S. Gareau, Y. G. Patel, Y. Li, I. Aranda, A. C. Halpern, K. S. Nehal, and M. Rajadhyaksha, "Confocal mosaicing microscopy in skin excisions: a demonstration of rapid surgical pathology," J. Microsc. **233**(1), 149–159 (2009).
22. C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "Nih image to imagej: 25 years of image analysis," Nat. Methods **9**(7), 671–675 (2012).
23. C. Blasse, S. Saalfeld, R. Etournay, A. Sagner, S. Eaton, and E. W. Myers, "PreMosa: extracting 2D surfaces from 3D microscopy mosaics," Bioinformatics **33**(16), 2563–2569 (2017).
24. A. Shihavuddin, S. Basu, E. Rexhepaj, F. Delestro, N. Menezes, S. M. Sigoillot, E. Del Nery, F. Selimi, N. Spassky, and A. Genovesio, "Smooth 2d manifold extraction from 3d image stack," Nat. Commun. **8**(1), 15554 (2017).
25. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, (2017), pp. 2223–2232.
26. M. Combalia, J. Perez-Anker, A. Garcea-Herrera, L. Alos, V. Vilaplana, F. Marques, S. Puig, and J. Malvehy, "Digitally stained confocal microscopy through deep learning," in *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning*, vol. 102 (2019), pp. 121–129.

27. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2016 (2016), pp. 770–778.

28. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp. 5967–5976 (2016).

29. D. S. Gareau, "Ground truth training images for basal cell carcinoma pathology in ex vivo microscopy: supplemental data," Zenodo, 2021 http://doi.org/10.5281/zenodo.4670596.

30. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, (Springer, 2015), pp. 234–241.

31. M. Tan and Q. V. Le, "Efficientnet: rethinking model scaling for convolutional neural networks," arXiv preprint arXiv:1905.11946 (2019).

32. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, (2009).

33. M. Alam, O. Ibrahim, M. Nodzenski, J. M. Strasswimmer, S. I. Jiang, J. L. Cohen, B. J. Albano, P. Batra, R. Behshad, A. V. Benedetto, C. S. Chan, S. Chilukuri, C. Crocker, H. W. Crystal, A. Dhir, V. A. Faulconer, L. H. Goldberg, C. Goodman, S. S. Greenbaum, E. K. Hale, C. W. Hanke, G. J. Hruza, L. Jacobson, J. Jones, A. Kimyai-Asadi, D. Kouba, J. Lahti, K. Macias, S. J. Miller, E. Monk, T. H. Nguyen, G. Oganesyan, M. Pennie, K. Pontius, W. Posten, J. L. Reichel, T. E. Rohrer, J. A. Rooney, H. T. Tran, E. Poon, D. Bolotin, M. Dubina, N. Pace, N. Kim, W. Disphanurat, U. Kathawalla, R. Kakar, D. P. West, E. Veledar, and S. Yoo, "Adverse events associated with Mohs micrographic surgery: multicenter prospective cohort study of 20,821 cases at 23 centers," JAMA Dermatol. **149**(12), 1378–1385 (2013).

34. M. Kinoshita, Y. Matsuda, T. Arai, Y. Soejima, M. Sawabe, and N. Honma, "Cytological diagnostic clues in poorly differentiated squamous cell carcinomas of the breast: Streaming arrangement, necrotic background, nucleolar enlargement and cannibalism of cancer cells," Cytopathology **29**(1), 22–27 (2018).

35. N. Hirshoren, J. Danne, B. J. Dixon, M. Magarey, S. Kleid, A. Webb, A. Tiong, J. Corry, and D. Gyorki, "Prognostic markers in metastatic cutaneous squamous cell carcinoma of the head and neck," Head & Neck **39**(4), 772–778 (2017).

36. J. Kiely, M. Kostusiak, O. Bloom, and A. Roshan, "Poorly differentiated cutaneous squamous cell carcinomas have high incomplete excision rates with uk minimum recommended pre-determined surgical margins," J. Plast. Reconstr. Aesthetic Surg. **73**(1), 43–52 (2020).

37. H. W. Rogers, M. A. Weinstock, S. R. Feldman, and B. M. Coldiron, "Incidence estimate of nonmelanoma skin cancer (keratinocyte carcinomas) in the u.s. population, 2012," JAMA Dermatol. **151**(10), 1081–1086 (2015).

38. J. Perez-Anker, J. Malvehy, and D. Moreno-Ramirez, "Ex vivo confocal microscopy using fusion mode and digital staining: Changing paradigms in histological diagnosis," Actas Dermosifiliogr **111**(3), 236–242 (2020).

39. K. Sinha, F. Ali, G. Orchard, W. Rickaby, M. Shams, R. Mallipeddi, and R. Patalay, "Use of a novel 1-hour protocol for rapid frozen section immunocytochemistry, in a case of squamous cell carcinoma treated with Mohns micrographic surgery," Clin. Exp. Dermatol. **43**(4), 454–457 (2018).

40. J. B. Carter, M. M. Johnson, T. L. Chua, P. S. Karia, and C. D. Schmults, "Outcomes of primary cutaneous squamous cell carcinoma with perineural invasion: An 11-year cohort study," JAMA Dermatol. **149**(1), 35–42 (2013).

41. C. D. Schmults, P. S. Karia, J. B. Carter, J. Han, and A. A. Qureshi, "Factors predictive of recurrence and death from cutaneous squamous cell carcinoma: A 10-year, single-institution cohort study," JAMA Dermatol. **149**(5), 541–547 (2013).

42. P. S. Karia, J. Han, and C. D. Schmults, "Cutaneous squamous cell carcinoma: Estimated incidence of disease, nodal metastasis, and deaths from disease in the united states, 2012," J. Am. Acad. Dermatol. **68**(6), 957–966 (2013).

43. A. Aslam and S. Z. Aasi, "Frozen-section tissue processing in Mohns surgery," Dermatol. Surg. **45**(2), S57–S69 (2019).