nature biomedical engineering

ARTICLES

https://doi.org/10.1038/s41551-019-0362-y

Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning

Yair Rivenson[1,2,3,6*], Hongda Wang[1,2,3,6], Zhensong Wei[1], Kevin de Haan[1,2,3], Yibo Zhang[1,2,3], Yichen Wu[1,2,3], Harun Günaydın[1], Jonathan E. Zuckerman[4], Thomas Chong[4], Anthony E. Sisk[4], Lindsey M. Westbrook[4], W. Dean Wallace[4] and Aydogan Ozcan[1,2,3,5*]

The histological analysis of tissue samples, widely used for disease diagnosis, involves lengthy and laborious tissue preparation. Here, we show that a convolutional neural network trained using a generative adversarial-network model can transform wide-field autofluorescence images of unlabelled tissue sections into images that are equivalent to the bright-field images of histologically stained versions of the same samples. A blind comparison, by board-certified pathologists, of this virtual staining method and standard histological staining using microscopic images of human tissue sections of the salivary gland, thyroid, kidney, liver and lung, and involving different types of stain, showed no major discordances. The virtual-staining method bypasses the typically labour-intensive and costly histological staining procedures, and could be used as a blueprint for the virtual staining of tissue images acquired with other label-free imaging modalities.

Microscopic imaging of tissue samples is a fundamental tool that is used for the diagnosis of various diseases and is the workhorse of pathology and biological sciences. The clinically established gold standard image of a tissue section is the result of a laborious process that includes the tissue specimen being formalin-fixed and paraffin-embedded (FFPE), sectioned into thin slices (typically around 2–10 μm), labelled and stained, mounted on a glass slide and microscopically imaged using, for example, a bright-field microscope. All of these steps use multiple reagents and introduce irreversible effects on the tissue. There have been recent efforts to change this workflow using different imaging modalities. One line of work imaged fresh, non-paraffin-embedded tissue samples using nonlinear microscopy methods based on, for example, two-photon fluorescence, second-harmonic generation[1], third-harmonic generation[2] and Raman scattering[3–5]. Another study used a controllable super-continuum source[6] to acquire multi-modal images for the chemical analysis of fresh tissue samples. These methods require the use of ultrafast lasers or super-continuum sources, which might not be readily available in most settings and require relatively long scanning times owing to weaker optical signals. Other microscopy methods for imaging unsectioned tissue samples have recently emerged, that use ultraviolet light excitation on stained samples[7,8], or take advantage of the autofluorescence emission of biological tissue at short wavelengths[9]. In fact, there are unique opportunities to use autofluorescence for imaging tissue samples by making use of the fluorescent light emitted from endogenous fluorophores. It has been demonstrated that such endogenous fluorescence signatures carry useful information that can be mapped to the functional and structural properties of a biological specimen, and have therefore been used extensively for diagnostics and research purposes[9–11]. One of the main focus areas of these efforts has been the spectroscopic investigation of the relationship between different biological molecules and their structural properties under different conditions. Some of these well-characterized biological constituents include vitamins (for example, vitamin A, riboflavin and thiamin), collagen, coenzymes and fatty acids[10].

Although some of the techniques described above have unique capabilities to discriminate between, for example, cell types and sub-cellular components in tissue samples using various contrast mechanisms, pathologists and tumour-classification software packages[12] are, in general, trained for examining histologically stained tissue samples to make diagnostic decisions. Partially motivated by this, some of the above-mentioned techniques were also augmented to create pseudo-haematoxylin and eosin (H&E) images[1,13]. These techniques were based on a linear approximation that relates the fluorescence intensity of an image to the dye concentration per tissue volume by using empirically determined constants that represent the mean spectral response of various dyes embedded in the tissue. These methods also used exogenous staining to enhance the fluorescence signal contrast to create virtual H&E images of tissue samples.

In this study, we demonstrate deep-learning-based virtual histology staining using autofluorescence of unstained tissue that was imaged with a wide-field fluorescence microscope through a standard near-ultraviolet light excitation and emission filter set (see Methods). Virtual staining is performed on a single autofluorescence image of the sample using a deep convolutional neural network (CNN), which is trained using the concept of generative adversarial networks (GAN)[14] to match the bright-field microscopic images of tissue samples after they are labelled with a certain histology stain (Figs. 1, 2 and Supplementary Fig. 1). Therefore, using a CNN, we replace the histological staining and bright-field imaging steps with the output of the trained neural net, which is fed with the autofluorescence image of the unstained tissue. The network inference is fast—taking, for example, 1.9 s mm^{-2} using a desktop

[1]Electrical and Computer Engineering Department, University of California, Los Angeles, Los Angeles, CA, USA. [2]Bioengineering Department, University of California, Los Angeles, Los Angeles, CA, USA. [3]California NanoSystems Institute, University of California, Los Angeles, Los Angeles, CA, USA. [4]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [5]Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [6]These authors contributed equally: Yair Rivenson, Hongda Wang. *e-mail: rivensnonyair@ucla.edu; ozcan@ucla.edu

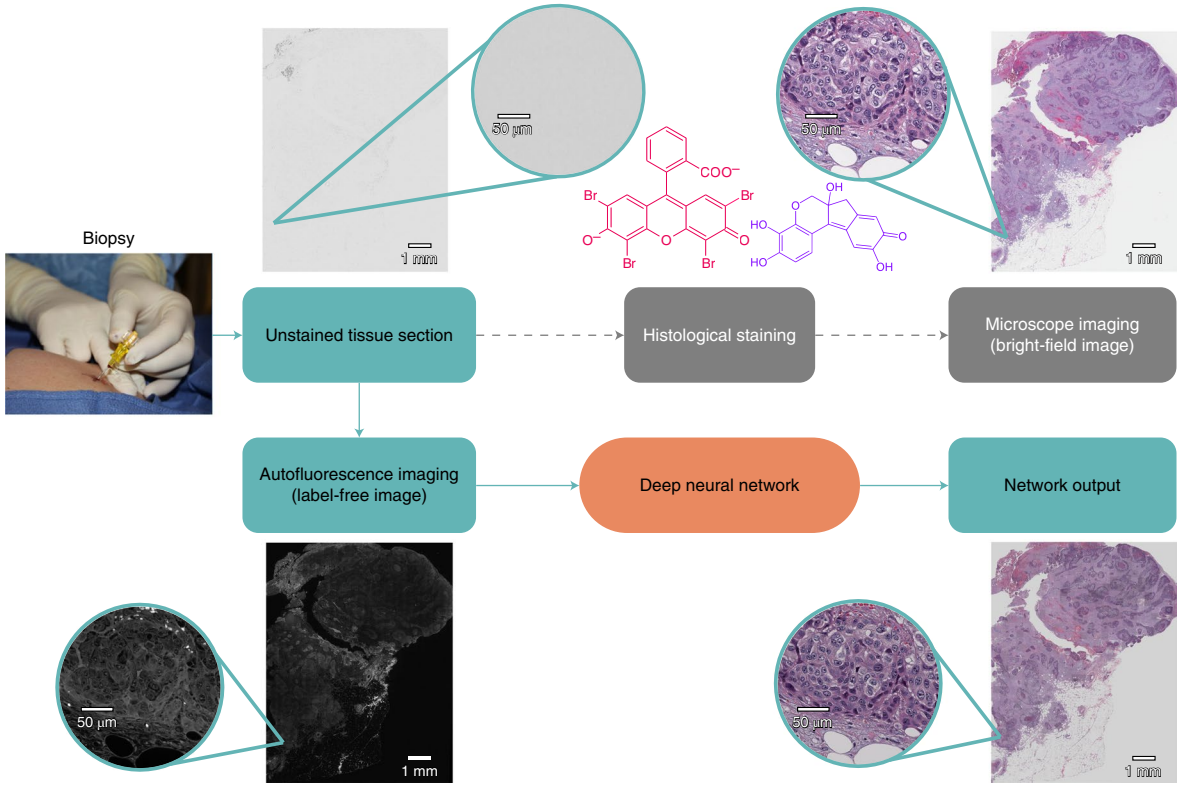


Fig. 1 | Deep-learning-based virtual histology staining using autofluorescence of unstained tissue. The schematic outlines the steps in the standard (top) and virtual (bottom) staining techniques. After training using a GAN, the neural network (orange box) rapidly outputs a virtually stained tissue image (H&E in this case), in response to the input of an autofluorescence image of an unstained tissue section, bypassing the standard histological staining procedure (grey boxes).

computer for a tissue section scanned using a ×20 objective lens—and can be substantially improved by using ever-evolving computing hardware with parallelization capabilities.

We demonstrated this deep-learning-based virtual histology staining method by imaging label-free human tissue samples, including salivary gland, thyroid, kidney, liver and lung. The network output created equivalent images that were well matched to the images of the same samples that were labelled with three different stains, that is, H&E (salivary gland and thyroid), Jones stain (kidney) and Masson's trichrome (liver and lung). Furthermore, the staining efficacy of our approach for whole-slide images (WSIs), corresponding to some of these samples, was blindly evaluated by a group of pathologists. The pathologists were able to recognize histopathological features in images generated with our virtual staining technique, and achieved a high degree of agreement with the histologically stained images of the same samples (see the blind evaluation of the staining efficacy, as part of the Results).

As the input image of the network is captured by a conventional fluorescence microscope with a standard filter set, this approach has transformative potential to use unstained tissue samples for pathology and histology applications, entirely bypassing the histological staining process, saving time and expenditure. As an example, for the histology stains that we learned to virtually stain in this work, each staining procedure of a tissue section takes on average around 45 min (H&E) and 2–3 h (Masson's trichrome and Jones stain), with an estimated cost, including labour, of US$2–5 (H&E)[15,16] and more than US$16–35 (Masson's trichrome and Jones stain)[16,17]. Furthermore, some of these histological staining processes involve time-sensitive steps that require the expert to monitor the process under a microscope, which makes the entire process not only lengthy and relatively costly, but also laborious. The presented method bypasses all these staining steps, and also allows for the preservation of unlabelled tissue sections for later analysis, such as micro-marking of subregions of interest on the unstained tissue specimen that can be used for more advanced immunohistochemical and molecular analysis to facilitate—for example, customized therapies[18,19]. In addition, this deep-learning-based virtual histology staining approach can be broadly applied to other excitation wavelengths or fluorescence filter sets, as well as to other microscopy modalities (such as nonlinear microscopy) that utilize additional endogenous or exogenous contrast mechanisms[1–8]. In our experiments, we used sectioned and fixed tissue samples to provide meaningful comparisons to the results of the standard histological staining process. However, this approach is potentially applicable for use with non-fixed, unsectioned tissue samples, which could make it suitable for use in surgery rooms, or at the site of a biopsy for rapid diagnosis or telepathology applications. Beyond its clinical applications, this method could broadly benefit the histology field and its applications in life-science research and education.

Results

Virtual staining of tissue samples. We demonstrated the presented method using different combinations of tissue sections and stains. After training a deep CNN (see Methods), we blindly tested its inference by feeding it with autofluorescence images of label-free tissue sections that did not overlap with the images used in the training or validation sets. Figure 3 summarizes our results for a salivary gland tissue section, which was virtually stained to match H&E-stained bright-field images of the same sample. These results show that the CNN is capable of transforming an autofluorescence image of a label-free tissue section into a bright-field equivalent image, showing the correct colour scheme that is expected from

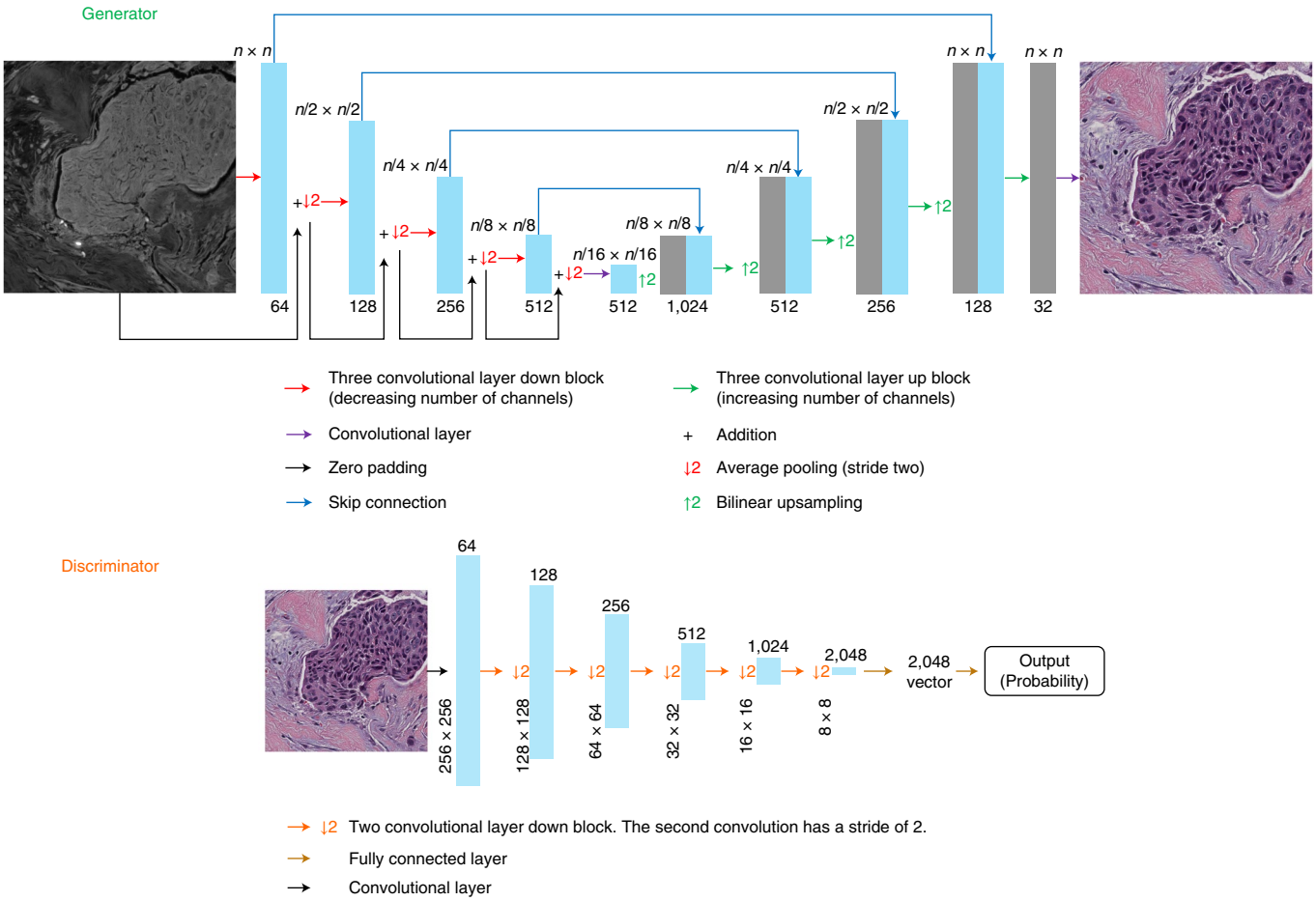


Fig. 2 | Virtual staining GAN architecture. Schematic of the CNN operation. The generator section is used to virtually stain the images. It comprises four 'down blocks', each of which are made up of three convolutional layers that are each followed by an average pooling layer of stride two. The down section is followed by four 'up blocks', which each contain three convolutional layers and are bilinearly upsampled by a factor of two. Skip connections are used to pass data between layers of the same level. The discriminator comprises five down blocks, each of which has two convolutional layers; the second convolutional layer has a stride of two, to reduce the tensor size. The down block reduces the size of the images while increasing the number of channels and is followed by two fully connected layers. The variable *n* represents the number of pixels on the lateral dimensions of each image patch that passes through the network. During the training, a 256 × 256 pixel patch is used; however, during the testing phase larger images can be inferred, as a result of the convolutional nature of the network.

an H&E-stained tissue sample. Evaluation of Fig. 3c,d shows that the H&E-stained images demonstrate a small island of infiltrating tumour cells within subcutaneous fibro-adipose tissue. Note that the nuclear detail—including distinction of nucleoli (arrow) and chromatin texture—is clearly displayed in both panels. Similarly, in Fig. 3g,h, the H&E-stained images demonstrate infiltrating squamous cell carcinoma. The desmoplastic reaction with oedematous myxoid change (indicated by an asterisk) in the adjacent stroma is clearly identifiable in both stains.

Next, we trained our deep network to virtually stain other tissue types with two different 'special' stains, that is, the Jones stain (kidney) and the Masson's trichrome stain (liver and lung). Figures 3 and 4 summarize our results for deep-learning-based virtual staining of these tissue sections, which match very well with the bright-field images of the same samples that were captured after the histological staining process. These results illustrate that the deep network can infer staining patterns of different types of histology staining method used for different tissue types, from a single autofluorescence image of a label-free specimen. In an example of renal cell carcinoma (Fig. 3k,o), the virtual Jones silver stain captures the black staining of extracellular collagen and maintains the visual integrity of the H&E counterstain. The virtual Masson's trichrome staining of liver tissue samples in Fig. 4c correctly reveals the histological features that correspond to hepatocytes, sinusoidal spaces, collagen and fat droplets; this is consistent with the histologic appearance in the bright-field images of the same tissue samples, which were captured after histological staining (Fig. 4d). Similarly, virtual staining of the lung tissue samples shown in Fig. 4c reveals consistently stained histological features that correspond to vessels, collagen and alveolar spaces as they appear in the bright-field images after histological staining (Fig. 4d).

We further compared our virtual staining approach to standard histological staining methods used to diagnose multiple types of condition in various types of FFPE or frozen tissue section. The results are summarized in Supplementary Table 1. The analysis of 15 tissue sections by four board-certified pathologists (who were not aware of our virtual staining technique) demonstrated 100% non-major discordance, which is defined as no clinically significant differences in diagnosis among professional observers. The 'time to diagnosis' varied considerably among observers, ranging from an average of 10 s per image for observer 2 to 276 s per image for observer 3. However, the intra-observer variability was very minor and tended towards a shorter time to diagnosis with our virtual stained slides for all observers except observer 2, who spent equal

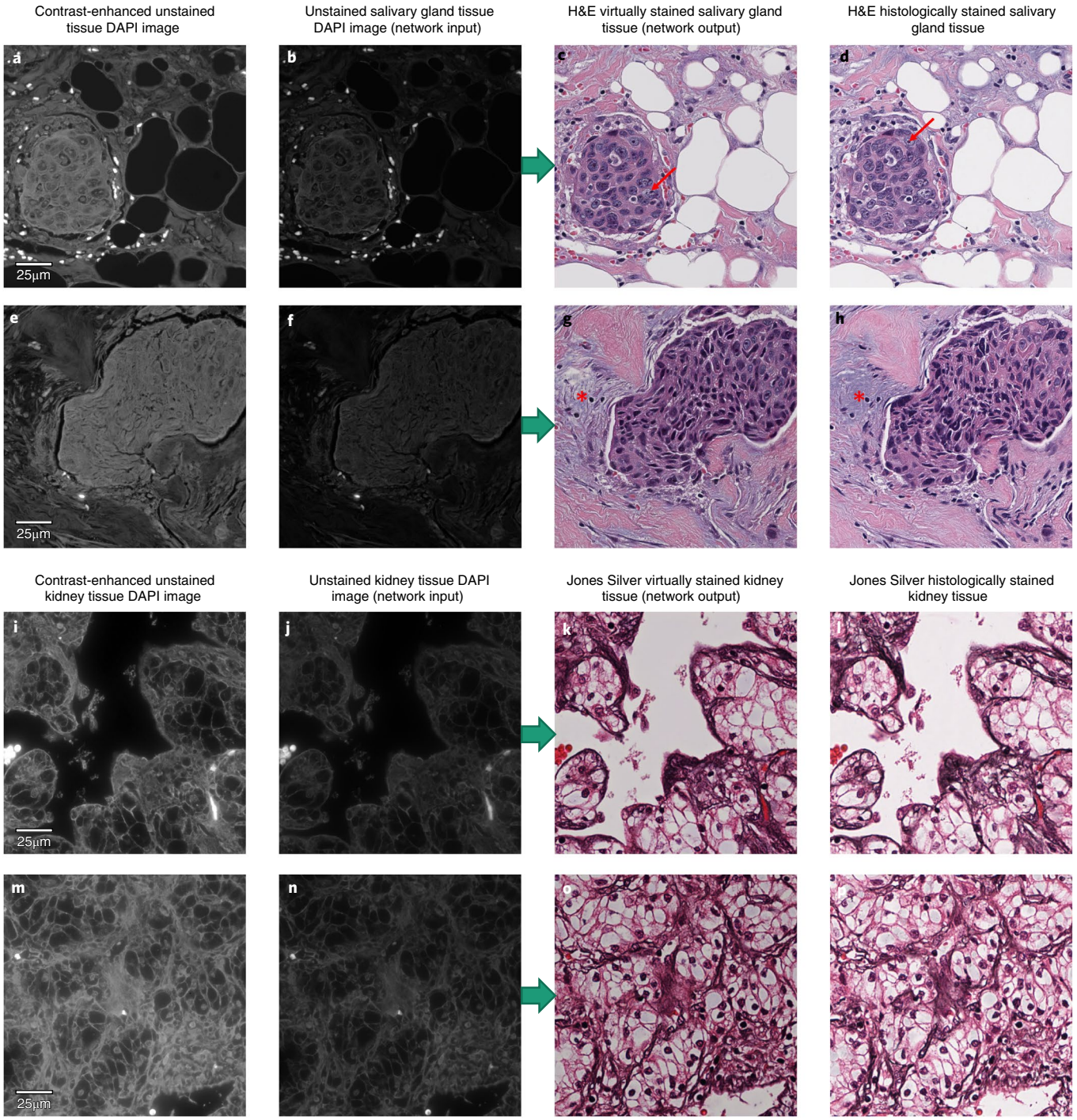


Fig. 3 | Virtual staining results match the H&E- and Jones-stained images. a–**h**, Salivary gland tissue samples that are unstained, or either virtually or histologically stained. **a**,**b**,**e**,**f**, Autofluorescence images of unstained salivary gland tissue sections used as input into the neural network. **c**,**g**, Images showing virtual H&E staining of salivary gland tissue. **d**,**h**, Bright-field images of the salivary gland tissue sections after the histological staining process. Evaluation of both **c** and **d** demonstrates a small island of infiltrating tumour cells within subcutaneous fibro-adipose tissue. Note that the nuclear detail, including distinction of nucleoli (red arrows) and chromatin texture, is clearly displayed in both panels. Similarly, in **g** and **h**, the virtual and histological H&E-stained images demonstrate infiltrating squamous cell carcinoma. The desmoplastic reaction with oedematous myxoid change (marked by asterisks) in the adjacent stroma is clearly identifiable. **i**–**p**, Kidney tissue samples that are unstained, or either virtually or histologically stained. **i**,**j**,**m**,**n**, Autofluorescence images of unstained kidney tissue sections; **j** and **n** show the network input. **k**,**o**, Images showing virtual Jones staining of autofluorescence kidney tissue sections. **l**,**p**, Bright-field images of the same kidney tissue sections after histological staining. Green arrows indicate the staining of individual samples.

time—that is, around 10 s per image—for virtual and histologically stained slides. These results indicate that there is very similar diagnostic utility between the two image modalities.

Blind evaluation of staining efficacy for WSIs. After evaluating the differences in tissue section and staining methods, we tested the ability of our virtual staining approach in the specialized staining

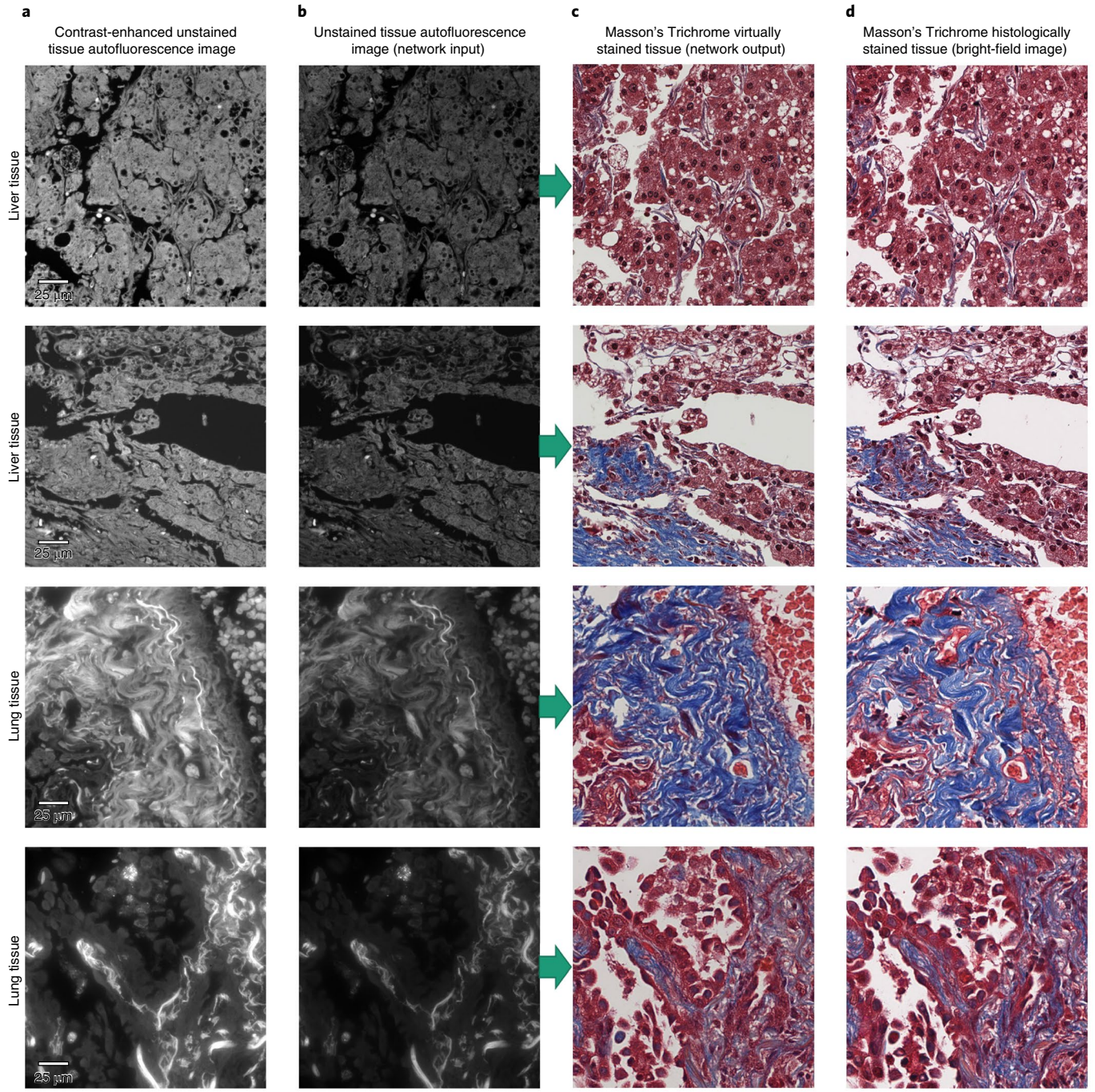


Fig. 4 | Virtual staining results match the Masson's trichrome stain for liver and lung tissue sections. a–d, Liver and lung tissue samples that are unstained (**a**,**b**), or either virtually (**c**) or histologically stained (**d**) with Masson's trichrome. **a**,**b**, Autofluorescence images of unstained liver tissue sections and unstained lung tissue sections. Only the raw images in **b** were used as input to the trained neural network. **c**, Virtual Masson's trichrome staining results (network output) for the same liver and lung tissue samples. **d**, Bright-field images of the same liver and lung tissue sections, after the histological staining process. Green arrows indicate the virtual staining of individual samples by the neural network.

histology workflow. We imaged the autofluorescence distribution of 15 label-free samples of liver tissue sections and 13 label-free tissue sections of kidney, using a ×20/0.75 numerical aperture (NA) objective lens. All liver and kidney tissue sections were obtained from different patients and included both small biopsies and larger resections. All tissue sections were obtained from FFPE tissues that were not coverslipped, as indicated in our virtual staining protocol. After autofluorescence scanning, the tissue sections were histologically stained with Masson's trichrome (4 µm liver tissue sections) and Jones stain (2 µm kidney tissue sections). We then divided the WSIs into training and test sets. For the liver slides cohort, seven WSIs were used to train the virtual staining algorithm and eight WSIs were used for blind testing. For the kidney slides cohort, six WSIs were used to train the algorithm and seven WSIs were used for testing. Three board-certified pathologists were blinded to the staining techniques for each WSI and were asked to apply a grade from 1 to 4 for the quality of the different stains: 4, perfect; 3, very good; 2, acceptable; 1, unacceptable. Moreover, three additional

Table 1 | Blind evaluation of virtual and histological Masson's trichrome staining in liver sections

Tissue number	Pathologist 1				Pathologist 2				Pathologist 3				Average			
	ND	CD	EF	SQ	ND	CD	EF	SQ	ND	CD	EF	SQ	ND	CD	EF	SQ
1 (HS)	3	2	1	1	4	4	3	4	1	1	1	3	**2.67**	2.33	1.67	2.67
1 (VS)	3	3	3	3	3	3	2	3	2	2	3	3	**2.67**	**2.67**	**2.67**	**3.00**
2 (HS)	3	2	4	4	4	4	3	4	1	2	2	2	2.67	**2.67**	3.00	**3.33**
2 (VS)	3	3	4	4	4	3	3	3	2	2	3	3	**3.00**	**2.67**	**3.33**	**3.33**
3 (HS)	3	3	2	2	3	3	4	3	1	1	1	1	**2.33**	**2.33**	**2.33**	**2.00**
3 (VS)	3	2	1	1	3	3	1	4	1	1	1	1	**2.33**	2.00	1.00	**2.00**
4 (HS)	3	2	4	4	3	4	4	4	1	2	1	2	2.33	**2.67**	3.00	3.33
4 (VS)	3	3	4	4	4	3	4	4	2	2	3	3	**3.00**	**2.67**	**3.67**	**3.67**
5 (HS)	3	3	4	4	3	3	2	1	1	3	2	2	2.33	**3.00**	2.67	2.33
5 (VS)	3	2	3	3	3	3	4	2	2	1	3	3	**2.67**	2.00	**3.33**	**2.67**
6 (HS)	3	2	3	3	4	4	4	3	2	2	2	2	**3.00**	**2.67**	**3.00**	**2.67**
6 (VS)	3	3	4	3	4	3	4	3	1	1	1	1	2.67	2.33	**3.00**	2.33
7 (HS)	3	3	4	4	3	4	4	3	2	1	2	2	2.67	**2.67**	**3.33**	**3.00**
7 (VS)	3	2	3	3	4	4	4	3	2	2	3	3	**3.00**	**2.67**	**3.33**	**3.00**
8 (HS)	3	3	4	4	4	4	4	3	1	1	1	1	2.67	**2.67**	3.00	2.67
8 (VS)	3	2	4	4	4	3	4	4	2	2	3	2	**3.00**	2.33	**3.67**	**3.33**

Liver tissue was stained with Masson's trichrome and scored for nuclear detail (ND), cytoplasmic detail (CD), extracellular fibrosis (EF) and overall stain (SQ). The winner (and tied) average scores are in bold. HS, histologically stained sections; VS, virtually stained sections.

board-certified pathologists applied the same score scale (1–4) for specific features for liver samples only: nuclear detail, cytoplasmic detail and extracellular fibrosis. These results are summarized in Table 1 (liver) and Supplementary Table 2 (kidney).

Our data indicate that the pathologists were able to recognize histopathological features using both staining techniques and with a high degree of agreement between the techniques, without a clearly preferred staining technique (virtual versus histological).

Staining standardization. Staining standardization could be an interesting by-product of our virtual staining approach. In other words, the deep network converges to a 'common stain' colourization[20] scheme—which can be observed in Fig. 5—that compares WSIs of histologically and virtually stained liver tissue sections. The variation in the histologically stained liver tissue sections is higher than that of the virtually stained tissue images (Fig. 5). The colourization of the virtual stain is solely the result of its training (that is, the gold standard histological staining used during the training phase) and can be further adjusted to the preferences of pathologists, by retraining the network with a new stain colourization. Such 'improved' training can be created from scratch or accelerated through transfer learning[21]. This potential staining standardization, using deep learning, could remedy the negative effects of human-to-human variations at different stages of the sample preparation[22], create a common ground among different clinical laboratories, enhance the diagnostic workflow for clinicians and assist in the development of new algorithms, such as automatic tissue metastasis detection[12] or grading of different types of cancer.

Transfer learning to other tissue–stain combinations. Using the concept of transfer learning[21], the training procedure for new tissue and/or stain types can converge much faster, while also reaching an improved performance—that is, a better local minimum in the training cost/loss function (see Methods). This means that a pre-learnt CNN model, from a different tissue–stain combination, can be used to initialize the deep network to statistically learn virtual staining of a new combination. Figure 6 demonstrates the favourable attributes of such an approach: a new deep neural network was trained to virtually stain the autofluorescence images of unstained thyroid tissue sections, and it was initialized using the weights and biases of another network that was previously trained for H&E virtual staining of the salivary gland. The evolution of the loss metric as a function of the number of iterations used in the training phase clearly demonstrates that the new thyroid deep network rapidly converges to a lower minimum in comparison to the same network architecture, which was trained from scratch using random initialization. Figure 6 also compares the output images of this thyroid network at different stages of its learning process, which further illustrates the impact of transfer learning to rapidly adapt our approach to new tissue–stain combinations. The network output images, after the training phase with, for example, ≥6,000 iterations, reveal that cell nuclei show irregular contours, nuclear grooves and chromatin pallor, which are suggestive of papillary thyroid carcinoma; cells also show mild to moderate amounts of eosinophilic granular cytoplasm and the fibrovascular core at the network output image shows an increased number of inflammatory cells, including lymphocytes and plasma cells.

Discussion

We demonstrated the ability to virtually stain label-free tissue sections, using a supervised deep-learning technique that uses a single autofluorescence image of the sample as input, captured by a standard fluorescence microscope and filter set. This statistical learning-based method has the potential to restructure the clinical workflow in histopathology, and can benefit from various imaging modalities, such as fluorescence microscopy, nonlinear microscopy, holographic microscopy and optical coherence tomography[23]. It has the potential to provide a digital alternative to the standard practice of histological staining of tissue samples. We used fixed, unstained tissue samples to provide a meaningful comparison to histologically stained tissue samples, which is essential to train the neural network as well as to blindly test the performance of the network output against the clinically approved method. However, the presented deep-learning-based approach is broadly applicable

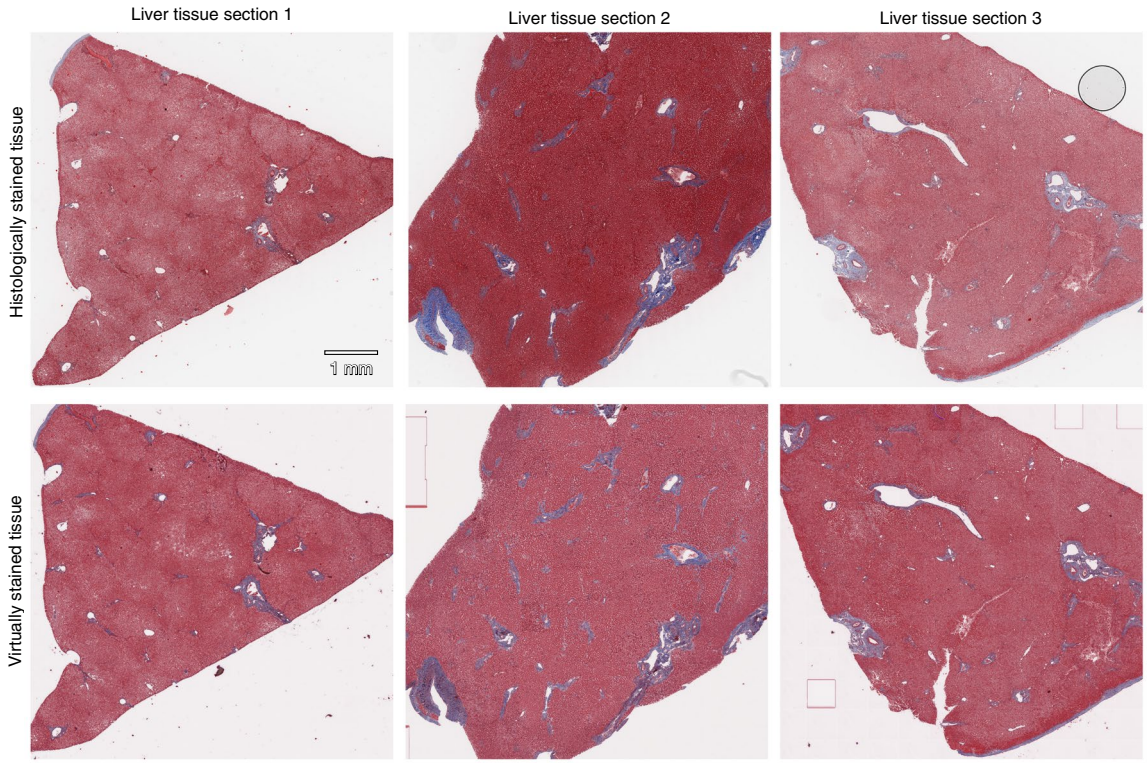


Fig. 5 | Virtual staining reduces staining variability. Staining standardization for whole-slide imaging of liver tissue sections, showing histologically stained and virtually stained images. The virtual staining approach can help to mitigate the staining variability that is part of the histological staining process.

to unsectioned, fresh tissue samples without the use of any labels or stains. After training, the deep network can be used to virtually stain the images of label-free fresh tissue samples, acquired using, for example, ultraviolet light or deep-ultraviolet light excitation, or even nonlinear microscopy modalities. Raman microscopy in particular can provide very rich label-free biochemical signatures that can further enhance the effectiveness of the virtual staining that the neural network learns.

The proposed method can be combined with other excitation wavelengths and/or imaging modalities to enhance its inference performance for different tissue constituents. For example, we attempted to detect melanin on a skin tissue section using virtual H&E staining. However, melanin was not clearly identified in the output of the network, as it presents a weak autofluorescence signal at DAPI excitation and emission wavelengths[24] measured in our system. One potential method to increase the autofluorescence of melanin is to image the samples while they are in an oxidizing solution[25]. As a more practical alternative, here we used an additional autofluorescence channel, originating from, for example, Cy5 filter (excitation 628 nm and emission 692 nm) such that the melanin signal could be enhanced and accurately inferred in our virtual-staining approach. By training the network using both the DAPI and Cy5 autofluorescence channels, the deep network was able to successfully determine where melanin occurs in the sample, as illustrated in Fig. 7. By contrast, when only the DAPI channel was used (Fig. 7a), the network was unable to determine the areas that contained melanin. In other words, the additional autofluorescence information from the Cy5 channel was used by the network to distinguish melanin from the background tissue. We note that the results shown in Fig. 7 were acquired using a lower-resolution objective lens (×10/0.45 NA) for the Cy5 channel, to supplement the high-resolution DAPI scan (×20/0.75 NA), because we hypothesized that most of the necessary information is found in the high-resolution DAPI scan and the additional information (for example, the presence of melanin) can be encoded with the lower-resolution scan. We think that other label-free imaging techniques and/or fluorescence channels could be combined to further enhance the inference of different tissue constituents using the presented deep-learning-based approach, which we leave for future research.

An important part of the training process involves matching autofluorescence images of label-free tissue samples and their corresponding bright-field images after the histological staining process. Note that during the staining process and related steps, some tissue constituents can be lost or deformed in a way that will mislead the loss/cost function in the training phase (an example of this is illustrated in Supplementary Fig. 2). However, this is only a training- and validation-related challenge and does not, in practice, pose any limitations for a well-trained neural network in the virtual staining of label-free tissue samples. To ensure the quality of the training and validation phases and minimize the impact of this challenge on the network's performance, we apply a multi-stage registration process—from global to local registration—in which one of the steps involves training a deep network for the task of enabling high-accuracy local registration. At the end of this initial registration, we set a threshold for an acceptable correlation value between the two sets of images (that is, before and after the histological staining process) and eliminate the non-matching image pairs from our training and validation set to make sure that the network learns the real signal, not the perturbations to the tissue morphology due to the histological staining process. This threshold is tuned for each specific tissue, depending on the correlation between the input and label images. For example, if there is a large misalignment between the input and target images of the registration network, the correlation will be low due to this misalignment. Therefore, an estimated threshold was initially used during the training to make sure that, for example, misaligned or warped images were not used in the training phase. On visual inspection, if it became clear that improperly registered images were making it

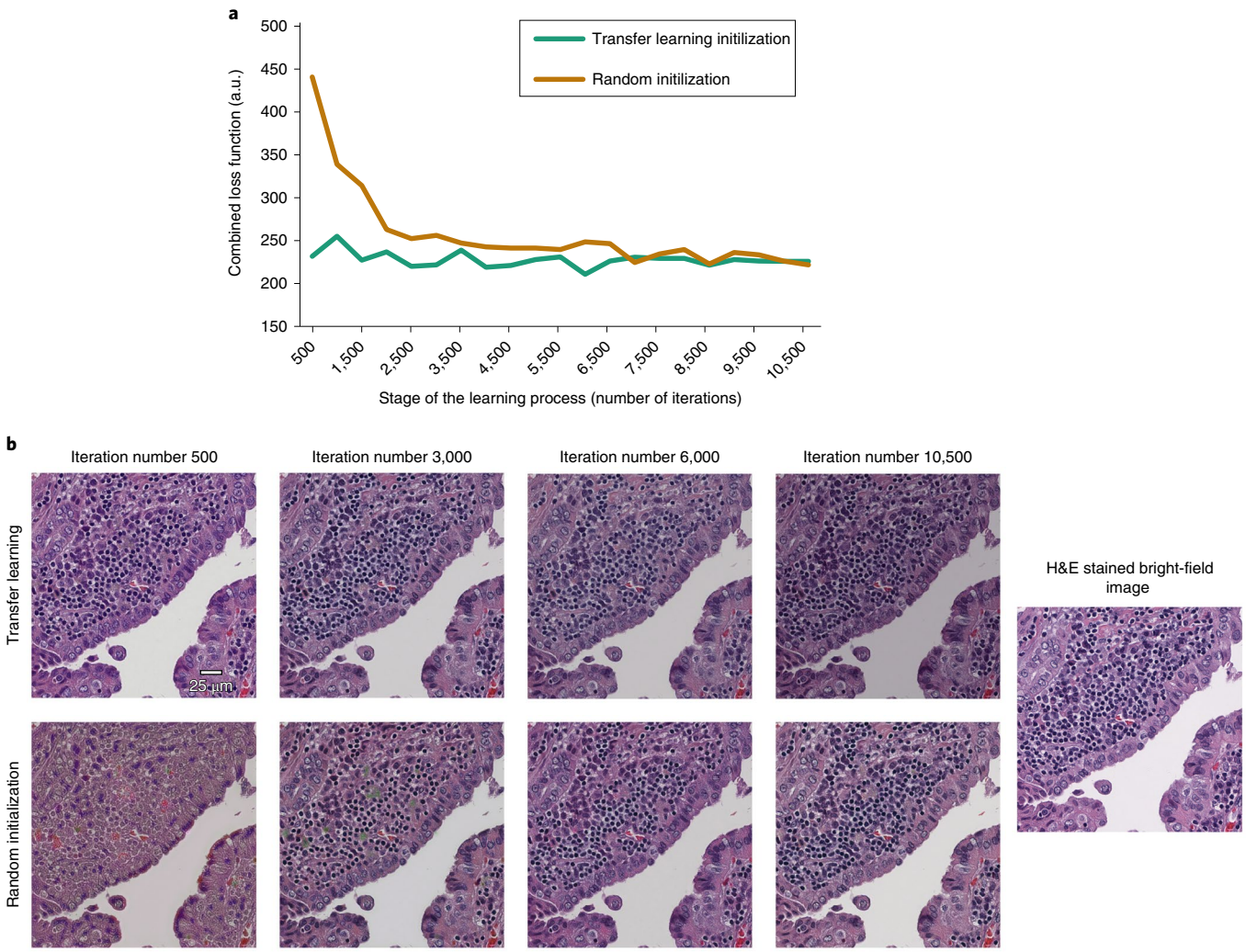


Fig. 6 | Accelerated convergence is achieved using transfer learning. a, Plot showing combined loss function against stage of the learning process. A new deep neural network is initialized using the weights and biases learned from the salivary gland tissue sections (see Fig. 3) to achieve virtual staining of thyroid tissue with H&E. Compared to random initialization, transfer learning enables faster convergence, achieving a lower local minimum. **b**, Network output images of a thyroid tissue section at different stages of the learning process, that is, after 500 iterations, 3,000 iterations, 6,000 iterations and 10,500 iterations (left) and a bright-field image of the same thyroid section stained with H&E. The images are compared to each other to better illustrate the impact of the transfer learning method to translate our approach to new tissue–stain combinations.

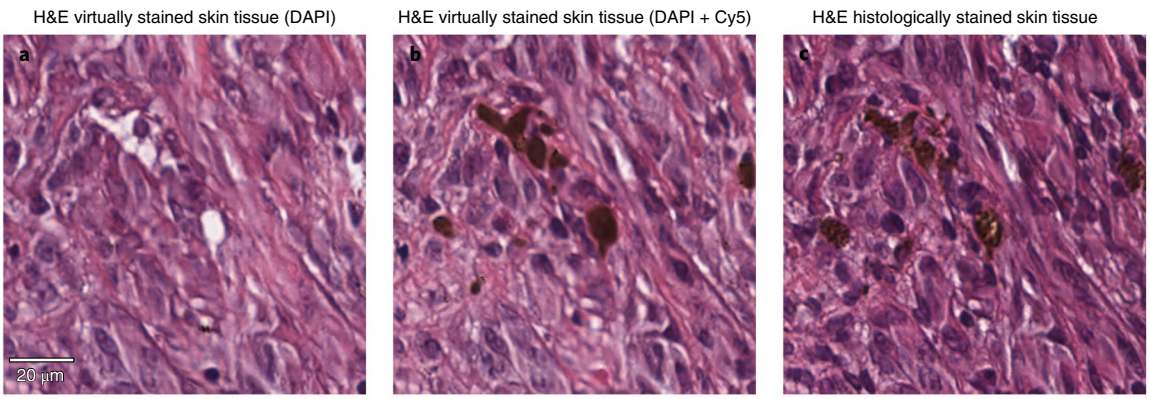


Fig. 7 | Melanin inference using multiple autofluorescence channels. a, Virtually stained skin tissue sample, using the DAPI channel only. **b**,**c**, The same tissue sample, virtually stained using both the DAPI and Cy5 channels (**b**), clearly revealing the melanin (dark-brown) features that are shown in the corresponding histologically stained image (**c**).

past the threshold, the threshold would be raised accordingly. On the other hand, if the threshold excluded too many images it would be lowered as appropriate. However, in our training phase for each tissue–stain model, the addition of poor data is not healthy for the network. The elimination of some training images was therefore not a major concern if there was any doubt regarding the image quality or alignment when visually inspected. As this only involves our training, it is a one-time effort, and we chose to be conservative in our training process to obtain the best blind-testing model for optimized generalization.

Above, we described a methodology to mitigate some of the training challenges that result from the random loss of some tissue features after the histological staining process. In fact, this highlights another motivation to skip the laborious and costly procedures that are involved in histological staining; it will be easier to preserve the local tissue histology in a label-free method, without the need for an expert to handle some of the delicate procedures of the staining process, which sometimes also requires the tissue to be observed under a microscope.

The training phase of our deep neural network takes a considerable amount of time (for example, approximately 13 h for the salivary gland network) using a desktop personal computer. However, this entire process could be substantially accelerated by using dedicated hardware, based on graphics processing units (GPUs). Furthermore, as already emphasized in Fig. 6, transfer learning provides a warm start to the training phase of a new tissue–stain combination, making the entire process substantially faster. Unlike other colour reconstruction or virtual staining approaches[13], once the deep network has been trained, the virtual staining of a new sample is performed in a single, non-iterative manner that does not require a trial-and-error approach or any parameter tuning to achieve the optimal result. Based on its feed-forward and non-iterative architecture, the deep-neural network rapidly outputs a virtually stained image in, for example, 1.9 s mm^{-2} using a dual-GPU desktop computer, for unstained tissue slides scanned using a ×20 objective lens. With further GPU-based acceleration and machine-learning-optimized processors, our approach has the potential to achieve real-time performance, which might especially be useful in the operating room or for in vivo imaging applications.

The virtual staining procedure that is implemented in this work is based on training separate CNNs for each tissue–stain combination. If the CNN is fed with the autofluorescence images of a different tissue–stain combination, it may not perform as desired (see, for example, Supplementary Fig. 3). This, however, is not a limitation, because for histology applications, the tissue and stain type are predetermined for each sample of interest and, therefore, a specific CNN selection for creating a virtually stained image from an autofluorescence image of the unlabelled sample does not require any additional information or resources. A more general CNN model can be trained for multiple tissue–stain combinations by, for example, increasing the number of trained parameters in the model[26], at the cost of a possible increase in the training and inference times. Using a similar strategy, another avenue to explore in future work is the potential of our approach to perform multiple virtual stains on the same unlabelled tissue type.

It is important to note that, like in any other imaging method that is based on automatic sample scanning, parts of the sample field of view (FOV) can be compromised owing to artefacts in the sample preparation process, such as dust or other particles that lie on top of the sample, in addition to tissue folding and cracks, among other artefacts. In Supplementary Fig. 4, we present examples of such artefacts, which result in aberrations in our virtual staining. Further development of auto-focusing algorithms that can learn to reject such artefacts during the imaging stage could minimize their occurrence in the final image.

As a next step, a wide-scale, randomized, clinical study would be needed to validate the diagnostic accuracy of the network output images against the clinical gold standard, which will be important to better understand potential biases in the output images of the network. A substantial advantage of the approach presented here is that it is quite flexible; it can accommodate feedback to statistically mend its performance if a diagnostic failure is detected through a clinical comparison, by penalizing such failures as they are caught. This iterative training and transfer learning cycle, based on clinical evaluations of the performance of the network output, will help us to optimize the robustness and clinical impact of our approach. In this sense, the process bears resemblance to the design phase of a histological stain, where—through trial and error—the stain is optimized to provide desired contrast to specific histological features.

Our virtual-staining approach could also be used for microguiding molecular analysis at the unstained-tissue level, by locally identifying regions of interest on the basis of virtual staining, and by using this information to guide subsequent analysis of the tissue, for example, microimmunohistochemistry or sequencing[18]. This type of virtual microguidance on an unlabelled tissue sample might facilitate the high-throughput identification of disease subtypes and the development of customized therapies for patients[27].

We would also like to note that although we demonstrated the virtual-staining method on the basis of a contrast mechanism that originates from tissue autofluorescence with a single excitation band, other contrast-generating methods to virtually stain label-free tissue samples could be explored, including, for example, multiple excitation and emission wavelengths, as well as other imaging modalities such as polarization imaging, quantitative phase microscopy, optical coherence tomography and perhaps combinations of these modalities.

Methods

Sample preparation. The FFPE tissue sections (2 µm thick) were deparaffinized using xylene and mounted on a standard glass slide using Cytoseal (Thermo-Fisher Scientific); a coverslip was then placed (Fisherfinest, 24x50-1, Fisher Scientific). Following the initial autofluorescence imaging process (using a DAPI excitation and emission filter set) of the unlabelled tissue sample, the slide was put into xylene for approximately 48 h or until the coverslip could be removed without damaging the tissue. Once the coverslip was removed, the slide was dipped (approximately 30 dips) in absolute alcohol, 95% alcohol and then washed in deionized water for around 1 min. This step was followed by the corresponding staining procedures for H&E, Masson's trichrome or Jones stains. This tissue-processing path is only used for the training and validation of the approach and is not needed after the network has been trained. To test our method, we used different tissue–stain combinations: the salivary gland and thyroid tissue sections were stained with H&E, kidney tissue sections were stained with Jones stain, and the liver and lung tissue sections were stained with Masson's trichrome. For the WSI staining efficacy evaluation study, the liver tissue sections were 4 µm thick and the kidney tissue sections were 2 µm thick. In the WSI study, the FFPE tissue sections were not coverslipped during the autofluorescence imaging stage. Following autofluorescence imaging, the tissue samples were histologically stained as described above (Masson's trichrome for the liver and Jones for the kidney tissue sections). The unstained frozen samples were prepared by embedding the tissue section in O.C.T. (Tissue-Tek, Sakura Finetek) and dipped in 2-methylbutane with dry ice. The frozen section was then cut into 4-µm-thick sections and put in a freezer until it was imaged. Following the imaging process, the tissue section was washed with 70% alcohol, H&E stained and coverslipped. The samples were obtained from the Translational Pathology Core Laboratory and were prepared by the Histology Laboratory at UCLA. The kidney tissue sections of diabetic and non-diabetic patients were obtained under IRB 18-001029 (UCLA). All the tissue sections were obtained after the deidentification of patient-related information, and were prepared from existing (that is, archived) specimens. Therefore, this work did not interfere with standard practices of care or sample-collection procedures; the original tissue specimens were archived before this work, and were not collected specifically for our research.

Data acquisition. The label-free tissue-autofluorescence images were captured using a conventional fluorescence microscope (IX83, Olympus) equipped with a motorized stage. The image acquisition process was controlled by MetaMorph microscope automation software (Molecular Devices). The unstained tissue samples were excited with near ultraviolet light and imaged using a DAPI filter cube (OSFI3-DAPI-5060C, excitation wavelength 377 nm/50 nm bandwidth, emission wavelength 447 nm/60 nm bandwidth) with a ×40/0.95 NA objective lens

(Olympus UPLSAPO 40X2/0.95 NA, working distance (WD) 0.18) or ×20/0.75 NA objective lens (Olympus UPLSAPO 20X/0.75 NA, WD 0.65). For melanin inference, we also acquired autofluorescence images of the samples using a Cy5 filter cube (CY5-4040C-OFX, excitation wavelength 628 nm/40 nm bandwidth, emission wavelength 692 nm/40 nm bandwidth) with a ×10/0.4 NA objective lens (Olympus UPLSAPO 10X2). Each autofluorescence image was captured with a scientific CMOS sensor (ORCA-flash4.0 v. 2, Hamamatsu Photonics) with an exposure time of around 50–500 ms for the DAPI channel and approximately 3 s for the Cy5 channel (due to its lower NA). The bright-field images that were used for training and validation were acquired using a slide scanner microscope (Aperio AT, Leica Biosystems) with a ×20/0.75 NA objective (Plan Apo), equipped with a ×2 magnification adaptor.

Image pre-processing and alignment. As our deep neural network aims to learn a statistical transformation between an autofluorescence image of an unstained tissue and a bright-field image of the same tissue sample after histological staining, it is of critical importance to accurately match the FOV of the input and target images. An overall scheme describing the global and local image registration process is shown in Supplementary Fig. 5, which was implemented in MATLAB (MathWorks). The first step in this process was to find candidate features for matching unstained autofluorescence images and stained bright-field images. For this, each autofluorescence image (2,048 × 2,048 pixels) was downsampled to match the effective pixel size of the bright-field microscope images. This resulted in a 1,351 × 1,351-pixel unstained autofluorescent tissue image, which was contrast-enhanced by saturating the bottom 1% and the top 1% of all the pixel values, and contrast-reversed to better represent the colour map of the greyscale-converted WSI (see Supplementary Fig. 5). Then, a normalized correlation-score matrix was calculated by correlating each of the 1,351 × 1,351-pixel patches with the corresponding patch of the same size, extracted from the whole-slide grey-scale image. The entry in this matrix with the highest score represents the most likely matched FOV between the two imaging modalities. Using this information (which defines a pair of coordinates), we cropped the matched FOV of the original whole-slide bright-field image to create target images. After this FOV-matching procedure, the autofluorescence and bright-field microscope images were coarsely matched. However, they were still not accurately registered at the individual-pixel level, owing to the slight mismatch in the sample placement during the two different microscopic imaging experiments (autofluorescence, followed by bright-field), which randomly causes a slight rotation angle (for example, around 1–2°) between the input and target images of the same sample.

The second part of our input-target matching process involves a global-registration step[28] that corrects for this slight rotation angle between the autofluorescence and bright-field images. This was done by extracting feature vectors (descriptors) and their corresponding locations from the image pairs, and matching the features using the extracted descriptors[29]. Then, a transformation matrix that corresponds to the matched pairs was found using the *M*-estimator sample consensus algorithm[30], which is a variant of the random sample consensus (RANSAC) algorithm[31]. Finally, the angle-corrected image was obtained by applying this transformation matrix to the original bright-field microscope image patch. After the application of this rotation, the images were further cropped by 100 pixels (50 pixels on each side) to accommodate for undefined pixel values at the image borders, due to the rotation angle correction.

Next, a neural network was used to learn the transformation between the roughly matched images. This network used the same structure as the network described below and in Fig. 2. A low number of iterations was used so that the network only learns colour mapping, and not any spatial transformations between the input and label images. The autofluorescence images were passed through this network and used to perform local feature registration, using an elastic image registration algorithm. This algorithm matches the local features of both sets of images (autofluorescence versus bright-field) by hierarchically matching the corresponding blocks from large to small (see Supplementary Fig. 5). The calculated transformation map from this step was then applied to each bright-field image patch[32].

At the end of these registration steps, the autofluorescence image patches and their corresponding bright-field tissue image patches were accurately matched to each other and were used as input and label pairs for the deep neural network training phase, allowing the network to solely focus on, and learn the problem of, virtual histological staining.

For the images obtained with a ×20 objective lens that were used for generating Table 1 and Supplementary Table 2, a similar process was used. Instead of downsampling the autofluorescence images, the bright-field microscope images were downsampled to 75.85% of their original size so that they matched the lower magnification images. Furthermore, to create WSIs using these ×20 images, additional shading correction and normalization techniques were applied. Before being fed into the network, each FOV was normalized by subtracting the mean value across the entire slide and dividing it by the standard deviation between pixel values. This normalizes the network input within each slide and between slides. Finally, shading correction was applied to each image to account for the lower relative intensity measured at the edges of each FOV.

Deep neural network architecture, training and validation. In this study, we used a GAN[14] architecture to learn the transformation from a label-free unstained autofluorescence input image to the corresponding bright-field image of the histologically stained sample. A standard CNN-based training algorithm learns to minimize the loss/cost function between the network's output and the target label. Thus, the choice of this loss function is a critical component of the deep network design. For instance, simply choosing an ℓ_2−norm (mean squared error) penalty as a cost function will tend to generate blurry results[33,34], because the network averages a weighted probability of all the plausible results; therefore, additional regularization terms[35] are generally needed to guide the network to preserve the desired sharp sample features in the network's output. GANs avoid this problem by learning a criterion that aims to accurately classify whether the deep network's output image is real or fake (that is, correct in its virtual staining or wrong). As a result, output images that are inconsistent with the desired labels are not tolerated, which makes the loss function adaptive to the data and the desired task at hand. To achieve this goal, the GAN training procedure involves the training of two different networks, as shown in Fig. 2 and Supplementary Fig. 1. First, a generator network, which in our case aims to learn the statistical transformation between the unstained autofluorescence input images and the corresponding bright-field images of the same samples, after the histological staining process; and second, a discriminator network that learns how to discriminate between a true bright-field image of a stained tissue section and the generator network's output image. Ultimately, the desired result of this training process is a generator that transforms an unstained autofluorescence input image into an image that is indistinguishable from the stained bright-field image of the same sample. For this task, we defined the loss functions of the generator and discriminator as such:

$$\begin{aligned}\ell_{\text{generator}} &= \text{MSE}\{z_{\text{label}}, z_{\text{output}}\} \\ &\quad + \lambda \times \text{TV}\{z_{\text{output}}\} + \alpha \times (1 - D(z_{\text{output}}))^2 \\ \ell_{\text{discrimnator}} &= D(z_{\text{output}})^2 + (1 - D(z_{\text{label}}))^2\end{aligned} \tag{1}$$

where D refers to the discriminator network output, z_{label} denotes the bright-field image of the histologically stained tissue and z_{output} denotes the output of the generator network. The generator loss function balances the pixel-wise mean-squared error (MSE) of the generator network output image with respect to its label, the total variation (TV) operator of the output image and the discriminator network prediction of the output image, using regularization parameters (λ and α) that are empirically set to different values, which accommodates for approximately 2% and 20% of the pixel-wise MSE loss and the combined generator loss ($\ell_{\text{generator}}$), respectively. The TV operator of an image z is defined as:

$$\text{TV}(z) = \sum_p \sum_q \sqrt{(z_{p+1,q} - z_{p,q})^2 + (z_{p,q+1} - z_{p,q})^2} \tag{2}$$

where p and q are pixel indices. On the basis of equation (1), the discriminator attempts to minimize the output loss, while maximizing the probability of correctly classifying the real label (that is, the bright-field image of the histologically stained tissue). Ideally, the discriminator network would aim to achieve $D(z_{\text{label}}) = 1$ and $D(z_{\text{output}}) = 0$, but if the generator is successfully trained by the GAN, $D(z_{\text{output}})$ would ideally converge to 0.5.

The deep neural network architecture of the generator follows the design of U-net[36], and is detailed in Fig. 2. The U-net architecture is well-suited for our application, because it is capable of learning features at different scales without increasing the depth of the network. Each level of the U-net downsamples the input and learns the features that act on a larger scale than that of the previous layer. This allows the network to infer small features within each cell as well as the overall structure of the tissue samples. An input image is processed by the network in a multiscale fashion, using downsampling and upsampling paths, helping the network to learn the virtual staining task at various different scales. The downsampling path consists of four individual steps, with each step containing one residual block[37], each of which maps a feature map x_k into feature map x_{k+1}:

$$\begin{aligned}x_{k+1} &= x_k \\ &\quad + \text{LReLU} \\ &\quad [\text{CONV}_{k_3}\{\text{LReLU}[\text{CONV}_{k_2}\{\text{LReLU}[\text{CONV}_{k_1}\{x_k\}]\}]\}]\end{aligned} \tag{3}$$

where CONV{.} is the convolution operator that includes the bias terms, k_1, k_2 and k_3 denote the serial numbers of the convolution layers, and LReLU[.] is the nonlinear activation function (that is, a leaky rectified linear unit) that we used throughout the entire network, defined as:

$$\text{LReLU}(x) = \begin{cases} x & \text{for } x > 0 \\ 0.1x & \text{otherwise} \end{cases} \tag{4}$$

When training the networks for WSIs, an additional batch-normalization layer was added before each LReLU activation to allow for faster training and improve

its stability. This addition particularly improved sections of the tissue where the contrast in the autofluorescence images was particularly low. The number of input channels for each level in the downsampling path was set to: 1, 64, 128, 256, while the number of the output channels in the downsampling path was set to: 64, 128, 256, 512. To avoid the dimension mismatch for each block[35], we zero-padded x_k to match the number of the channels in x_{k+1} The connection between each downsampling level was a 2×2 average pooling layer with a stride of 2 pixels that downsampled the feature maps by a factor of 4 (twofold in each direction). Following the output of the fourth downsampling block, another convolutional layer maintained the number of the feature maps at 512, before connecting it to the upsampling path. The upsampling path consisted of four, symmetrical, upsampling steps, with each step containing one convolutional block. The convolutional block operation, which maps feature map y_k into feature map y_{k+1}, is given by:

$$y_{k+1} = \mathrm{LReLU}[\mathrm{CONV}_{k_6} \{\mathrm{LReLU}[\mathrm{CONV}_{k_5}\{\mathrm{LReLU}[\mathrm{CONV}_{k_4}\{\mathrm{CONCAT}(x_{k+1}, \mathrm{US}\{y_k\})\}]\}]\}] \quad (5)$$

where CONCAT(.) is the concatenation between two feature maps that merges the number of channels, US{.} is the upsampling operator, and k_4, k_5 and k_6, denote the serial numbers of the convolution layers. Similar to the downsampling path, batch normalization was added for the WSI training phase. The number of the input channels for each level in the upsampling path was set to 1,024, 512, 256, 128 and the number of the output channels for each level in the upsampling path was set to 256, 128, 64, 32. The last layer is a convolutional layer mapping 32 channels into three channels, represented by the YCbCr colour map[38]. Both the generator and the discriminator networks were trained with a patch size of 256×256 pixels.

The discriminator network, summarized in Fig. 2, receives three input channels that correspond to the YCbCr colour space of an input image. This input is then transformed into a 64-channel representation using a convolutional layer, which is followed by 5 blocks of the following operator:

$$z_{k+1} = \mathrm{LReLU}[\mathrm{CONV}_{k_2}\{\mathrm{LReLU}[\mathrm{CONV}_{k_1}\{z_k\}]\}] \quad (6)$$

where k_1 and k_2 denote the serial numbers of the convolutional layer. The number of channels for each layer was 3, 64, 64, 128, 128, 256, 256, 512, 512, 1,024, 1,024, 2,048. The next layer was an average pooling layer with a filter size that was equal to the patch size (256×256), which results in a vector with 2,048 entries. The output of this average pooling layer was then fed into two fully connected layers with the following structure:

$$z_{k+1} = \mathrm{FC}[\mathrm{LReLU}[\mathrm{FC}\{z_k\}]] \quad (7)$$

where FC is the fully connected layer, with learnable weights and biases. The first fully connected layer outputs a vector with 2,048 entries, while the second one outputs a scalar value. This scalar value was used as an input into a sigmoid activation function $D(z) = 1/(1+\exp(-z))$ that calculates the probability (between 0 and 1) of the discriminator network input to be real (genuine) or fake, that is, ideally $D(z_{\mathrm{label}}) = 1$

The convolution kernels throughout the GAN were set to be 3×3. These kernels were randomly initialized by using a truncated normal distribution[39] with a standard deviation of 0.05 and a mean of 0; all of the network biases were initialized as 0. The learnable parameters were updated through the training stage of the deep network using an adaptive moment estimation (Adam) optimizer[40] with a learning rate of 1×10^{-4} for the generator network and 1×10^{-5} for the discriminator network. Also, for each iteration of the discriminator, there were four iterations of the generator network, to avoid training stagnation following a potential overfit of the discriminator network to the labels. We used a batch size of 10 in our training.

Once all of the FOVs had passed through the network, the WSIs were stitched together using the Fiji[41] Grid/Collection stitching plugin[42]. This plugin calculates the exact overlap between each tile and linearly blends them into a single large image. Overall, the inference and stitching took approximately 5 min cm^{-2} and 30 s cm^{-2}, respectively, and can be substantially improved using hardware and software advancements. Before being shown to the pathologists, sections that were out of focus or had major aberrations (due to, for example, dust particles) in either the autofluorescence or bright-field images were cropped out. Finally, the images were exported to the Zoomify[43] format (designed to enable viewing of large images using a standard web browser) and uploaded to the GIGAmacro website[44] for easy access and viewing by the pathologists. We used six board-certified pathologists: four of them for the study in Supplementary Table 1, three for the study in Table 1 and three for the study in Supplementary Table 2, assigned according to expertise.

Implementation details. The other implementation details, including the number of trained patches, the number of epochs and the training times, are shown in Supplementary Table 3. The virtual staining network was implemented using Python version 3.5.0. The GAN was implemented using TensorFlow version 1.4.0. Other Python libraries used were os, time, tqdm, the Python Imaging Library (PIL), SciPy, glob, ops, sys and numpy. We implemented the software on a desktop computer with a Core i7-7700K CPU at 4.2 GHz (Intel) and 64 GB of RAM, running a Windows 10 operating system (Microsoft). Network training and testing were performed using dual GeForce GTX 1080Ti GPUs (NVidia).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability

The deep-learning models used in this work employ standard libraries and scripts that are publicly available in TensorFlow. The trained network models for Masson's trichrome stain (liver) and Jones stain (kidney), alongside sample test-image data are available through a Fiji-based plugin at https://github.com/whd0121/ImageJ-VirtualStain (Fiji can be downloaded at: https://imagej.net/Fiji/Downloads). The Fiji Grid/Collection stitching plugin was used to perform FOVs stitching. The inference (testing) software has been adapted to Fiji. MATLAB was used for the shading correction as well as the registration steps (coarse matching, global registration and local registration). Python based on the TensorFlow library was used to implement both the initial CNN used for image registration as well as the CNN used to produce the final virtually stained images. Our custom training codes are proprietary (and managed by the UCLA Office of Intellectual Property) and are not publicly available.

Data availability

The authors declare that all data supporting the results in this study are available within the paper and the Supplementary Information.

Received: 28 March 2018; Accepted: 28 January 2019;
Published online: 04 March 2019

References

1. Tao, Y. K. et al. Assessment of breast pathologies using nonlinear microscopy. *Proc. Natl Acad. Sci. USA* **111**, 15304–15309 (2014).
2. Witte, S. et al. Label-free live brain imaging and targeted patching with third-harmonic generation microscopy. *Proc. Natl Acad. Sci. USA* **108**, 5970–5975 (2011).
3. Ji, M. et al. Rapid, label-free detection of brain tumors with stimulated Raman scattering microscopy. *Sci. Transl. Med.* **5**, 201ra119 (2013).
4. Lu, F.-K. et al. Label-free DNA imaging in vivo with stimulated Raman scattering microscopy. *Proc. Natl Acad. Sci. USA* **112**, 11624–11629 (2015).
5. Orringer, D. A. et al. Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy. *Nat. Biomed. Eng.* **1**, 0027 (2017).
6. Tu, H. et al. Stain-free histopathology by programmable supercontinuum pulses. *Nat. Photon.* **10**, 534–540 (2016).
7. Fereidouni, F. et al. Microscopy with ultraviolet surface excitation for rapid slide-free histology. *Nat. Biomed. Eng.* **1**, 957–966 (2017).
8. Glaser, A. K. et al. Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens. *Nat. Biomed. Eng.* **1**, 0084 (2017).
9. Jamme, F. et al. Deep UV autofluorescence microscopy for cell biology and tissue histology. *Biol. Cell* **105**, 277–288 (2013).
10. Monici, M. Cell and tissue autofluorescence research and diagnostic applications. *Biotechnol. Annu. Rev.* **11**, 227–256 (2005).
11. Croce, A. C. & Bottiroli, G. Autofluorescence spectroscopy and imaging: a tool for biomedical research and diagnosis. *Eur. J. Histochem.* **58**, 2461 (2014).
12. Liu, Y. et al. Detecting cancer metastases on gigapixel pathology images. Preprint at https://arxiv.org/abs/1703.02442 (2017).
13. Giacomelli, M. G. et al. Virtual hematoxylin and eosin transillumination microscopy using epi-fluorescence imaging. *PLoS ONE* **11**, e0159337 (2016).
14. Goodfellow, I. et al. in *Advances in Neural Information Processing Systems 27* (eds Ghahramani, Z. et al.) 2672–2680 (Curran Associates, Inc., New York, 2014).
15. *Histology Laboratory: Price List Effective June 1, 2017* (Miller school of Medicine, accessed 23 March 2018); http://cpl.med.miami.edu/pathology-research/histology-laboratory/price-list
16. *Pathology & Laboratory Medicine: Fee Schedule* (Weill Cornell Medicine, accessed 23 March 2018); https://pathology.weill.cornell.edu/research/translational-research-services/fee-schedule
17. *Research Histology: Rates* (UC Davis Health, accessed 26 March 2018); http://www.ucdmc.ucdavis.edu/pathology/research/research_labs/histology/rates.html
18. Cree, I. A. et al. Guidance for laboratories performing molecular pathology for cancer patients. *J. Clin. Pathol.* **67**, 923–931 (2014).
19. Patel, P. G. et al. Preparation of formalin-fixed paraffin-embedded tissue cores for both RNA and DNA extraction. *J. Vis. Exp.* **21**, 54299 (2016).
20. Cho, H., Lim, S., Choi, G. & Min, H. Neural stain-style transfer learning using GAN for histopathological images. Preprint at https://arxiv.org/abs/1710.08543 (2017).

21. Hamel, P., Davies, M. E. P., Yoshii, K. & Goto, M. Transfer learning in MIR: sharing learned latent representations for music audio classification and similarity. In *Proc. 14th International Conference on Music Information Retrieval* (ISMR, 2013).
22. Badano, A. et al. Consistency and standardization of color in medical imaging: a consensus report. *J. Digit. Imaging* **28**, 41–52 (2015).
23. Vakoc, B. J. et al. Three-dimensional microscopy of the tumor microenvironment in vivo using optical frequency domain imaging. *Nat. Med.* **15**, 1219–1223 (2009).
24. Kozikowski, S., Wolfram, L. & Alfano, R. Fluorescence spectroscopy of eumelanins. *IEEE J. Quant. Electron.* **20**, 1379–1382 (1984).
25. Elleder, M. & Borovanský, J. Autofluorescence of melanins induced by ultraviolet radiation and near ultraviolet light. A histochemical and biochemical study. *Histochem. J.* **33**, 273–281 (2001).
26. Rivenson, Y., Zhang, Y., Gunaydin, H., Teng, D. & Ozcan, A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci. Appl.* **7**, 17141 (2018).
27. Lovchik, R. D., Kaigala, G. V., Georgiadis, M. & Delamarche, E. Micro-immunohistochemistry using a microfluidic probe. *Lab Chip* **12**, 1040–1043 (2012).
28. *Register Multimodal MRI Images* (Mathworks, 2018); https://www.mathworks.com/help/images/registering-multimodal-mri-images.html
29. Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**, 91–110 (2004).
30. Torr, P. H. S. & Zisserman, A. MLESAC: a new robust estimator with application to estimating image geometry. *Comput. Vis. Image Underst.* **78**, 138–156 (2000).
31. Hartley, R. & Zisserman, A. *Multiple View Geometry in Computer Vision* (Cambridge Univ. Press, Cambridge, 2003).
32. Rivenson, Y. et al. Deep learning enhanced mobile-phone microscopy. *ACS Photon.* **5**, 2354–2364 (2018).
33. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. & Efros, A. A. Context encoders: feature learning by inpainting. Preprint at https://arxiv.org/abs/1604.07379 (2016).
34. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition* 5967–5976 (IEEE, 2017).
35. Rivenson, Y. et al. Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).
36. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. Preprint at https://arxiv.org/abs/1505.04597 (2015).
37. He, K., Zhang, X., Ren, S. & Sun, J. in *Computer Vision – ECCV 2016* (eds. Leibe, B. et al.) 630–645 (Springer International Publishing, Basel, 2016)
38. *rgb2ycbcr* (Mathworks, 2018); https://www.mathworks.com/help/images/ref/rgb2ycbcr.html
39. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, 2016).
40. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2014).
41. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
42. Preibisch, S., Saalfeld, S. & Tomancak, P. Globally optimal stitching of tiled 3D microscopic image acquisitions. *Bioinformatics* **25**, 1463–1465 (2009).
43. Zoomify—Zoomable Web Images! (Zoomify Inc., 2018); http://zoomify.com/
44. GIGAmacro: Exploring Small Things in a Big Way (Four Chambers Studio LLC, 2018); https://viewer.gigamacro.com/

Acknowledgements

The Ozcan Research Group at UCLA acknowledges the support of the NSF Engineering Research Center (PATHS-UP), the Army Research Office, the NSF CBET Division Biophotonics Program, the National Institutes of Health (NIH, R21EB023115), HHMI, Vodafone Americas Foundation, the Mary Kay Foundation, and the Steven & Alexandra Cohen Foundation. The authors also acknowledge the Translational Pathology Core Laboratory and the Histology Laboratory at UCLA for their assistance with the sample preparation and staining. The authors acknowledge the time and effort of S. French, B.D. Cone, A. Nobori and C.M. Lee of the UCLA Department of Pathology and Laboratory Medicine for their evaluations; the assistance of R. Gao in preparing the ImageJ plugin and of R. Suh at the UCLA Department of Radiology for his help with Fig. 1.

Author contributions

Y.R. and A.O. conceived the research, H.W. and Y.R. conducted the experiments, and Y.R., Z.W., K.d.H., H.G., Y.Z. and H.W. processed the data. W.D.W. directed the clinical aspects of the research. J.E.Z., T.C., A.E.S. and L.M.W. performed diagnosis and stain efficacy assessment on the virtual and histologically stained slides. Y.R., H.W., Z.W., K.d.H., Y.Z., W.D.W. and A.O. prepared the manuscript and all authors contributed to the manuscript. A.O. supervised the research.

Competing interests

A.O., Y.R., H.W. and Z.W. have applied for a patent (US application number: 62651005) related to the work reported in this manuscript.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41551-019-0362-y.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to Y.R. or A.O.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature biomedical engineering

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41551-019-0362-y

In the format provided by the authors and unedited.

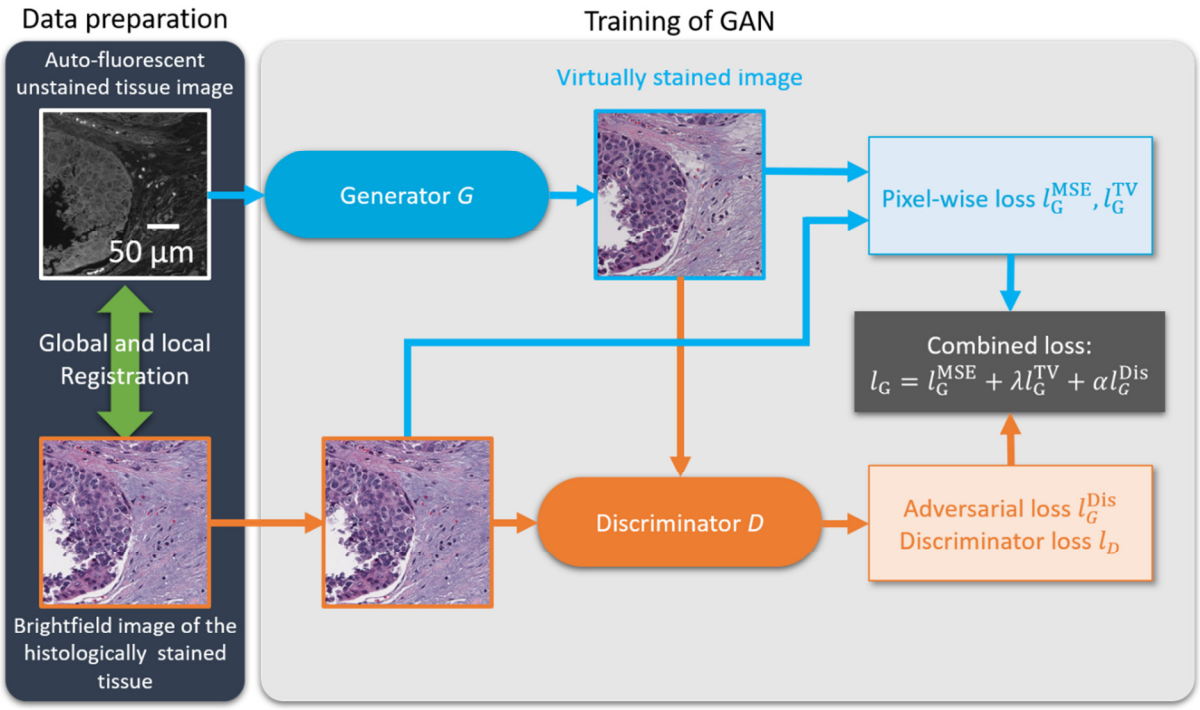
Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning

Yair Rivenson[1,2,3,6*], Hongda Wang[1,2,3,6], Zhensong Wei[1], Kevin de Haan[1,2,3], Yibo Zhang[1,2,3], Yichen Wu[1,2,3], Harun Günaydın[1], Jonathan E. Zuckerman[4], Thomas Chong[4], Anthony E. Sisk[4], Lindsey M. Westbrook[4], W. Dean Wallace[4] and Aydogan Ozcan[1,2,3,5*]

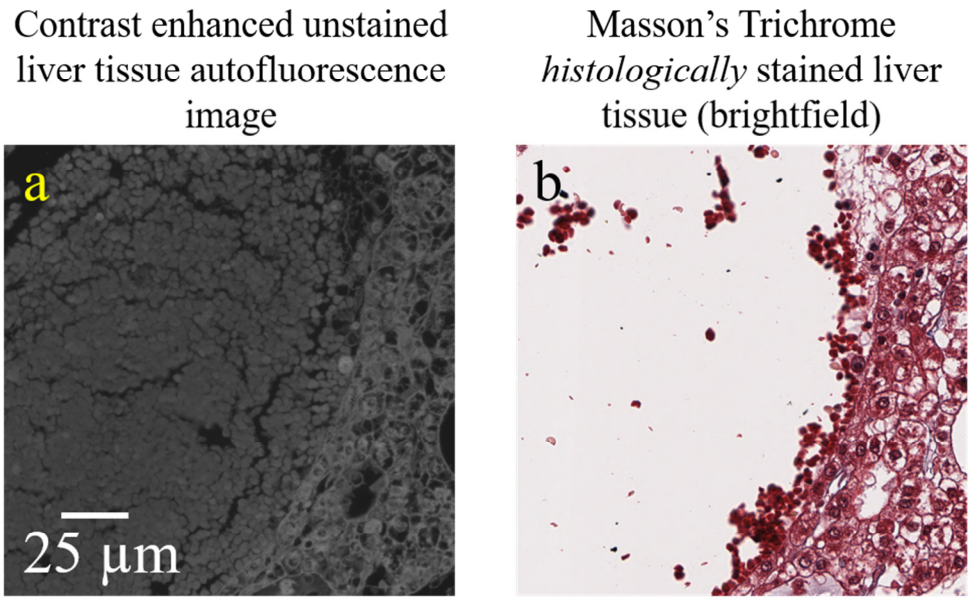
[1]Electrical and Computer Engineering Department, University of California, Los Angeles, Los Angeles, CA, USA. [2]Bioengineering Department, University of California, Los Angeles, Los Angeles, CA, USA. [3]California NanoSystems Institute, University of California, Los Angeles, Los Angeles, CA, USA. [4]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [5]Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [6]These authors contributed equally: Yair Rivenson, Hongda Wang. *e-mail: rivensnonyair@ucla.edu; ozcan@ucla.edu

Table of contents

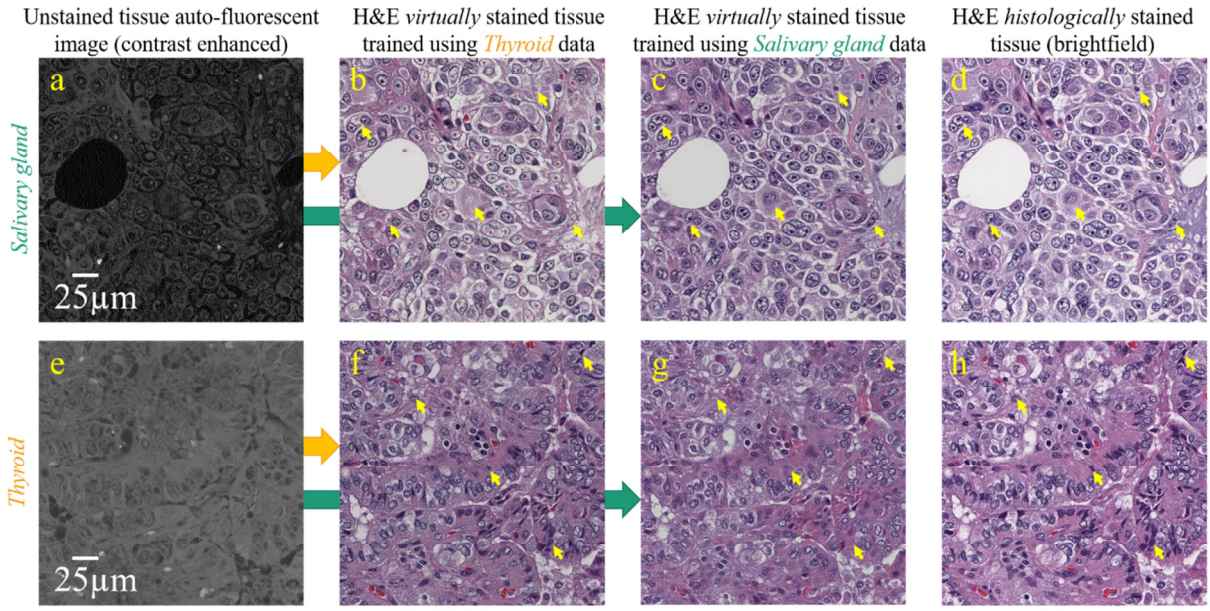
SUPPLEMENTARY FIGURES AND TABLES



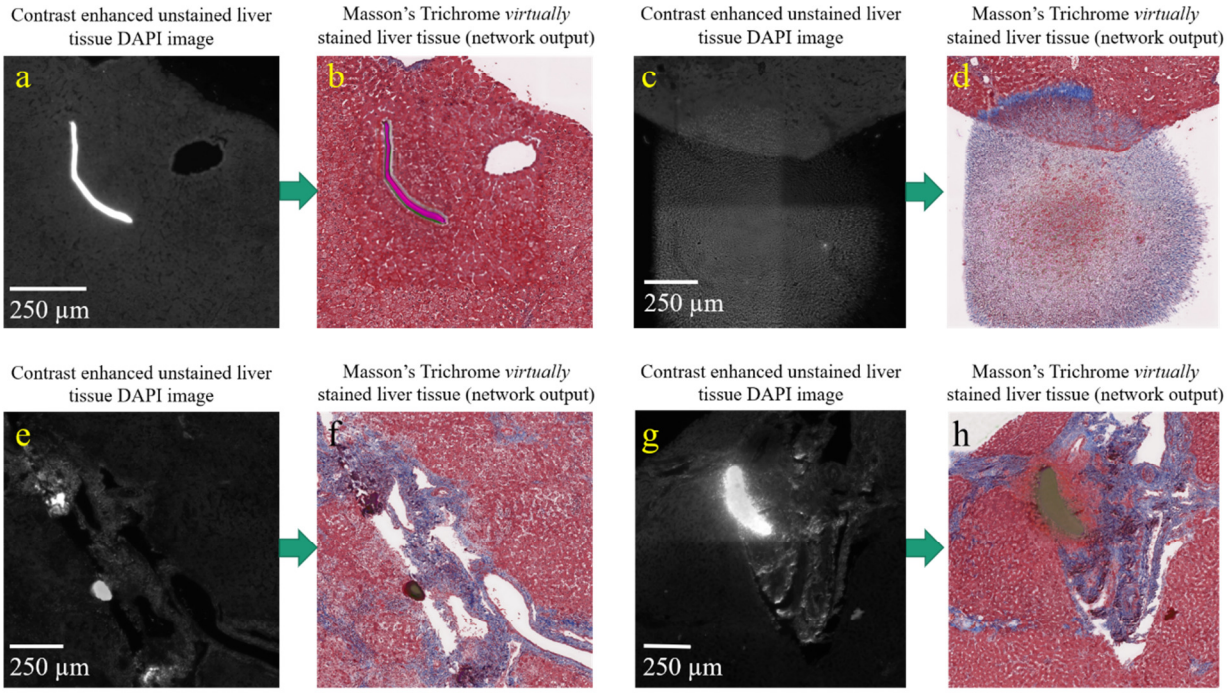
Supplementary Fig. S1. Training process of the virtual staining neural network using a GAN.



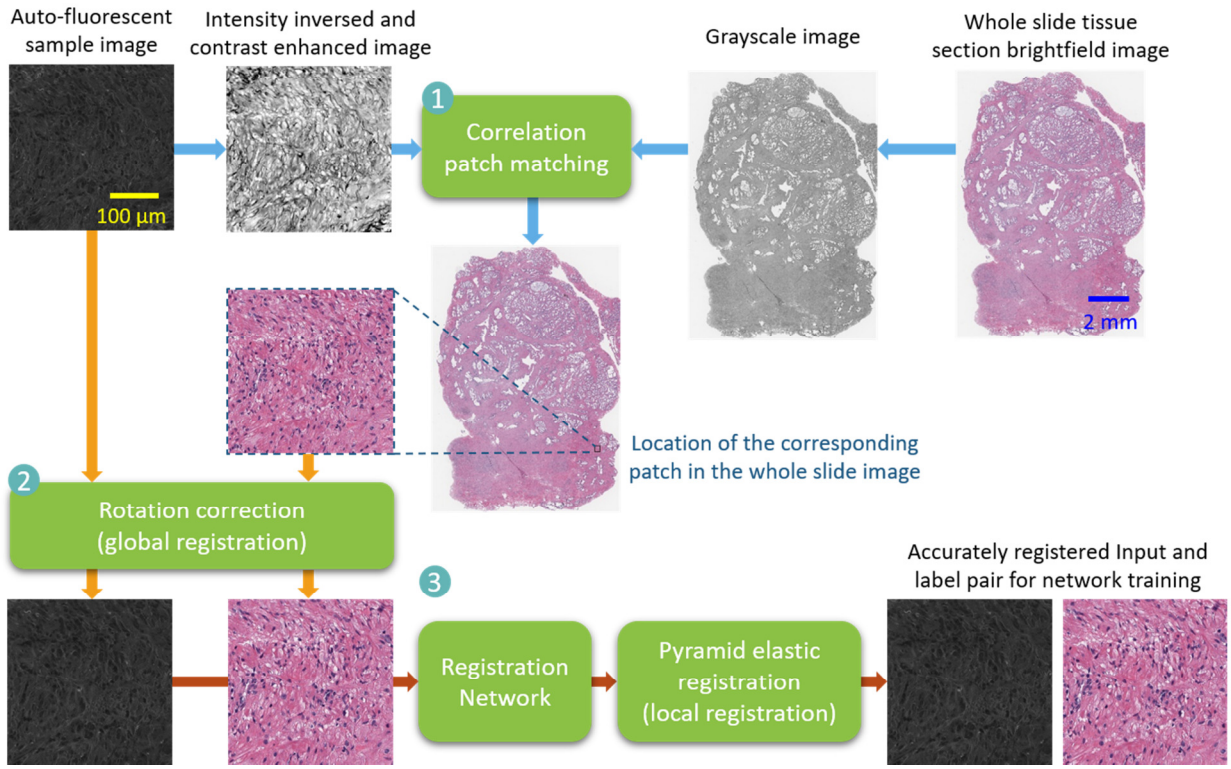
Supplementary Fig. S2. Structural changes caused during the histological staining process. (a) The autofluorescence image of an unstained liver tissue section (contrast enhanced). (b) Following the histological staining procedure, morphological changes are introduced as shown in this bright-field image of the same sample.



Supplementary Fig. S3. **Cross testing of the networks that were trained for specific tissue-stain combinations**. Autofluorescence images of unstained salivary gland and thyroid tissue sections are shown in (a) and (e), respectively. (b) Result of cross-testing the *salivary gland* autofluorescence image on a network which was trained to virtually H&E stain *thyroid* tissue. (c) Result of cross-testing the *salivary gland* autofluorescence image on a network which was trained to virtually H&E stain *salivary gland* tissue. (d) The bright-field image of the same *salivary gland* tissue, after the histological staining (H&E). (e-g) are the same as in (a-c) except for an unstained *thyroid* tissue as the network input. (h) The bright-field image of the same *thyroid* tissue after the histological staining (H&E). Note the higher fidelity between the staining techniques when the virtual stain was matched with the appropriate training. The arrows point to subtle differences in the evaluation of small nuclei between the virtual staining techniques. In all areas, the histologically stained slide is more closely aligned to the virtual staining method trained on the same tissue type.



Supplementary Fig. S4. Examples of virtual staining aberrations. Sample preparation related issues, such as highly fluorescent dust particles as shown in panels (a), (e) and (g), or focusing errors (c) that randomly occur during the microscope scanning procedure, lead to incorrect inference.



Supplementary Fig. S5. Autofluorescence and bright-field image registration. The field-of-view matching and registration process of the autofluorescence images of unstained tissue samples with respect to the bright-field images of the same samples, after the histological staining process, is illustrated.

Serial number	Tissue, fixation, type of stain	Pathologist #	Histochemically / Virtually stained	Diagnosis	Time to diagnose
1	Ovary, Frozen section, H&E	1	VS	Adenocarcinoma	30 sec
		2	VS	Borderline serous tumor	15 sec
		3	HS	Mucinous adenocarcinoma	10 min
		4	HS	Adenocarcinoma, endometrioid	2 min
2	Ovary, Frozen section, H&E	1	VS	Benign ovary	10 sec
		2	VS	Benign ovary	10 sec
		3	HS	Normal ovary with corpus luteal cyst	15 min
		4	HS	Normal	1 min
3	Salivary Gland, FFPE, H&E	1	VS	Benign salivary glands with mild chronic inflammation	10 sec
		2	VS	Benign parotid tissue	5 sec
		3	HS	Normal salivary gland	1 min
		4	HS	No histopathologic abnormality	1 min
4	Salivary Gland, Frozen section, H&E	1	HS	Pleomorphic adenoma	5 sec
		2	HS	Pleomorphic adenoma	5 sec
		3	VS	Pleomorphic adenoma	3 min
		4	VS	Pleomorphic adenoma	2 sec
5	Salivary Gland, FFPE, H&E	1	HS	Mucoepidermoid carcinoma, low grade	5 sec
		2	HS	Salivary duct carcinoma	5 sec
		3	VS	Mucoepidermoid carcinoma	10 min
		4	VS	Mucoepidermoid Carcinoma	10 sec
6	Breast, FFPE, H&E	1	VS	Invasive ductal carcinoma and DCIS	15 sec
		2	VS	Ductal carcinoma	10 sec
		3	HS	Invasive ductal carcinoma with DCIS	2 min
		4	HS	Invasive carcinoma	1 minute
7	Skin, FFPE, H&E	1	HS	Malignant melanoma	30 sec
		2	HS	melanoma	30 sec
		3	VS	Melanoma	5 min
		4	VS	Melanoma	1 min
8	Prostate, FFPE, H&E	1	HS	Prostatic adenocarcinoma 3+4	1 min
		2	HS	Prostatic adenocarcinoma 4+3	5 sec
		3	VS	Prostatic adenocarcinoma, Gleason pattern 3+4	5 min
		4	VS	HG-PIN with cribiforming vs carcinoma	5 min
9	Liver, FFPE, Masson's trichrome	1	VS	Benign liver with mild steatosis	10 sec
		2	VS	Benign liver with steatosis	5 sec
		3	HS	Hepatosteatosis, predominantly macrovesicular	3 min
		4	HS	Minimal steatosis, no fibrosis	5 min
10	Liver, FFPE, Masson's trichrome	1	HS	Benign liver with bridging fibrosis	10 sec
		2	HS	Benign liver, bridging fibrosis	5 sec
		3	VS	Moderate cirrhosis	1 min
		4	VS	Mild portal inflammation, focal bridging fibrosis (Stage 2-3)	5 minutes

11	Salivary Gland, FFPE, H&E	1	VS	Carcinoma	5 sec
		2	VS	Intraductal ca	20 sec
		3	HS	Poorly differentiated carcinoma	1 min
		4	HS	Low-grade salivary gland neoplasm	1 minute
12	Salivary Gland, FFPE, H&E	1	HS	Adenocarcinoma	5 sec
		2	HS	Salivary duct carcinoma	5 sec
		3	VS	Salivary duct carcinoma	2 min
		4	VS	Low-grade salivary gland neoplasm	1 minute
13	Thyroid, FFPE, H&E	1	VS	Papillary thyroid carcinoma, tall cell type	10 sec
		2	VS	Papillary thyroid ca, tall cell	20 sec
		3	HS	Papillary thyroid carcinoma, tall cell variant	5 min
		4	HS	PTC	10 sec
14	Thyroid, FFPE, H&E	1	HS	Papillary thyroid carcinoma	5 sec
		2	HS	Medullary ca	5 sec
		3	VS	Papillary thyroid carcinoma, oncocytic variant	7 min
		4	VS	PTC	10 sec
15	Thyroid, FFPE, H&E	1	VS	Papillary thyroid carcinoma	5 sec
		2	VS	Papillary thyroid ca	5 sec
		3	HS	Papillary thyroid carcinoma	1 min
		4	HS	PTC	10 sec

Supplementary Table 1. Pathology validation study of virtual vs. histochemical staining. HS: histological staining; VS: virtual staining.

Tissue #	Pathologist 1			Pathologist 2			Pathologist 3			Average		
	ND	CD	SQ	ND	CD	SQ	ND	CD	SQ	ND	CD	SQ
1 – HS	3	3	3	2	2	4	2	2	2	2.33	2.33	3.00
1 - VS	2	3	3	3	3	4	3	3	3	**2.67**	**3.00**	**3.33**
2 – HS	2	4	4	3	3	2	1	1	2	**2.00**	**2.67**	2.67
2 - VS	2	3	4	3	3	3	1	2	3	**2.00**	**2.67**	**3.33**
3 – HS	2	3	3	3	3	2	2	3	4	**2.33**	**3.00**	**3.00**
3 - VS	2	3	3	3	3	3	1	2	3	2.00	2.67	**3.00**
4 – HS	3	3	3	2	2	2	1	2	3	**2.00**	**2.33**	**2.67**
4 - VS	3	3	3	2	2	3	1	2	2	**2.00**	**2.33**	**2.67**
5 – HS	3	3	2	3	3	1	3	3	3	3.00	**3.00**	2.00
5 - VS	3	3	2	4	3	4	3	3	4	**3.33**	**3.00**	**3.33**
6 – HS	2	3	3	3	3	1	2	2	2	**2.33**	**2.67**	2.00
6 - VS	2	2	3	2	2	2	2	2	2	2.00	2.00	**2.33**
7 – HS	3	3	2	3	2	2	3	3	3	3.00	**2.67**	**2.33**
7 - VS	3	3	2	4	3	1	3	2	3	**3.33**	**2.67**	2.00

Supplementary Table 2. **Blind evaluation of virtual and histological Jones staining of kidney tissue sections**. Evaluation of nuclear detail (ND), cytoplasmic detail (CD) and overall stain quality (SQ) score. 4 = perfect, 3 = very good, 2 = acceptable, 1 = unacceptable. The winner (and tied) average scores are bolded. HS: histological staining; VS: virtual staining.

Virtual staining network	# of training patches	# of epochs	Training time (hours)
Salivary gland (H&E)	2768	26	13.046
Thyroid (H&E)	8336	8	12.445
Thyroid (H&E, transfer learning)	8336	4	7.107
Liver (Masson's Trichrome)	3840	26	18.384
Lung (Masson's Trichrome)	9162	10	16.602
Kidney (Jones stain)	4905	8	7.16
Liver (Masson's Trichrome, WSI)	211475	3	39.64
Kidney (Jones stain, WSI)	59344	14	57.05
Ovary 1	4738	84	37.21
Ovary 2	11123	14	37.41
Salivary Gland - 1	4417	65	24.61
Salivary Gland – 2	2652	90	23.9
Salivary Gland – 3	13262	24	30.58
Breast	67188	4	24.85
Skin	2566	124	27.02
Skin (DAPI+CY5)	2566	124	29.62
Prostate	677	472	30.27

Supplementary Table 3. **Training details for different tissue/stain combinations**.