Trends in Biotechnology

CellPress OPEN ACCESS

Review

Virtual staining for histology by deep learning

Leena Latonen [1,*,@], Sonja Koivukoski,[1] Umair Khan,[2] and Pekka Ruusuvuori [2,@]

In pathology and biomedical research, histology is the cornerstone method for tissue analysis. Currently, the histological workflow consumes plenty of chemicals, water, and time for staining procedures. Deep learning is now enabling digital replacement of parts of the histological staining procedure. In virtual staining, histological stains are created by training neural networks to produce stained images from an unstained tissue image, or through transferring information from one stain to another. These technical innovations provide more sustainable, rapid, and cost-effective alternatives to traditional histological pipelines, but their development is in an early phase and requires rigorous validation. In this review we cover the basic concepts of virtual staining for histology and provide future insights into the utilization of artificial intelligence (AI)-enabled virtual histology.

Highlights

Deep learning-based virtual staining has the potential to replace chemical staining in histology and to provide more sustainable, rapid, and cost-effective pipelines for histopathology.

Virtual staining will allow multiple outputs of stained images from a single input image of a tissue section, hence enabling virtual multiplexing and single-cell resolution between histological assessments that currently require separate tissue sections.

Deep learning-based virtual staining models require large amounts of training data, and quantitative computational and rigorous histological validation is required for each intended use.

From chemical to virtual staining

Histology (see Glossary) is a key tool in tissue research and clinical pathology. Alterations in tissue are measured in multiple scientific, clinical, and commercial fields by utilizing histopathology, and application areas include clinical disease diagnostics and prognostics, treatment follow-up, drug development, toxicological assessments during R&D in medical and chemical industries, and biomedical and pharmacological research. For decades, histology has depended on the use of chemical stains. Tissue samples are relatively transparent and uninformative to the human eye by themselves, which is why their contrast and specific features and traits are brought forward by staining the different chemical moieties and molecular components. The use of chemicals is, however, both expensive and undesirable in settings with limited resources.

The history of the development and use of histological stains spans centuries. Although some aspects of tissue research have followed methodological advances in microscopy that enable more detailed analyses through increased resolution, histology still mostly utilizes light microscopy-based techniques that are cheap, easily obtainable, and of sufficient resolution to detect tissue features and alterations in routine clinical and research use (Figure 1). Traditionally, the stained samples were observed individually through a light microscope. Currently, digital workflows use a **slide scanner** to generate **whole-slide images (WSIs)** that enable observations using a viewer program on a computer. This has opened the door for further digital solutions to support and facilitate histopathological analysis. The emergence of **deep learning**-based AI [1] has revolutionized computational approaches for data-driven predictive modeling in various fields, including medicine [2]. Histology is no exception, and deep learning-based tools developed for digital slide-based image diagnostics are an expanding field [3]. Recent results have shown that deep-learning-based **virtual staining** of unstained tissue enables tissue examination comparable to chemical staining for basic morphology [4,5], and other work is extending to other staining modalities in histology. In a workflow that now is being enabled through the development of virtual staining methods, the demands for tissues and chemicals will decrease, leading to shorter throughput times and, possibly, increased accuracy (Figure 1).

[1]Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland
[2]Institute of Biomedicine, University of Turku, Turku, Finland

*Correspondence:
leena.latonen@uef.fi (L. Latonen).
@Twitter: @LatonenLeena (L. Latonen) and @PekkaRuusuvuori (P. Ruusuvuori).

© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

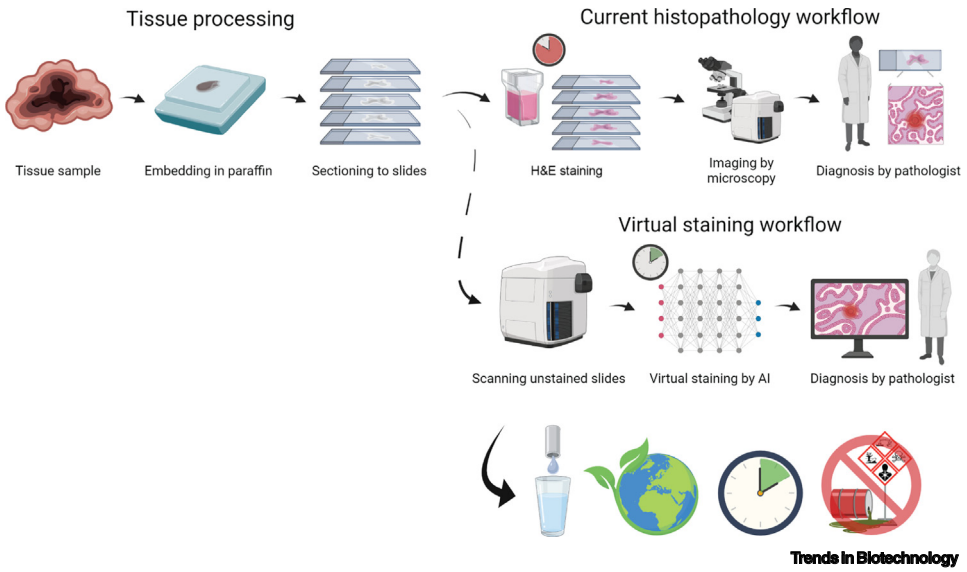

Figure 1. Virtual staining provides a more sustainable alternative to chemical staining-based histology. Although current histology is based on chemically staining tissue sections on slides for traditional or digital microscopy, virtual staining of unstained tissue sections with deep learning-based computational models helps to avoid the use of staining chemicals and decreases the amount of water and time consumed. Abbreviations: AI, artificial intelligence; H&E, hematoxylin and eosin. Figure created with BioRender.

In this review we provide an overview of recent advances in deep learning-enabled virtual histological staining techniques. We cover virtual staining of label-free microscopic images of unstained samples, as well as virtual **stain-to-stain transformation** – the transformation of images of an already stained tissue sample into another type of stain. The term 'virtual staining' is used here to refer to methods in which images containing histological stains are digitally generated using trained deep neural networks. Cultured cells can also be virtually stained, termed *in silico* labeling [6,7] or augmented microscopy [8]. We focus here on the histological virtual staining of tissues, although the basic concepts apply to both types of biological materials. Our aim is to provide a useful introduction not only for those currently utilizing histology but also for scientists who are new to the field, whether from the computational, imaging, or tissue laboratory side.

Basic concepts of traditional and virtual histological staining

Histological tissue samples are stained to bring out different chemical or biological properties of the tissue. Because thin tissue samples lack contrast themselves, the overall tissue structures can be visualized with so-called general morphology stains. The most common and routinely used stain of this type is hematoxylin and eosin (H&E). In addition to general morphology stains, other chemical stains, such as periodic acid–Schiff (PAS) and Masson's trichrome (MTC), are based on reactions with specific chemical moieties in the tissue, thus highlighting particular tissue components. The third class of stains is immunostaining, often utilizing **immunohistochemistry (IHC)** or immunocytochemistry in the clinic, but also through immunofluorescence staining especially in research. Based on specific protein sequences, epitopes, particular proteins in the tissue are recognized with antibodies and made visible with either enzymatically formed precipitates or fluorescent conjugates. Because immunostaining is highly specific for individual proteins or their subtypes, its utilization is dependent on the expression of the epitopes in the tissues under investigation and the availability of suitable antibodies.

Glossary

Deep learning: a subset of machine learning methods based on artificial neural networks with representation learning.
Hallucination: a visual artifact generated by a deep learning model that contains false or misleading information; in the case of histology, the output images contain deep learning-generated artifacts.
High-performance computing: technology that uses server-side computing using clusters of powerful processors, working in parallel, to process massive multidimensional datasets and solve computationally intensive problems.
Histology: the study of the microscopic anatomy of biological tissues.
Image registration: the process of computational alignment of two corresponding images.
Immunohistochemistry (IHC): a laboratory method to selectively identify antigens in a tissue section by the use of antibodies.
Slide scanner: a microscopy device for scanning WSIs.
Stain-to-stain transformation: transformation of images of an already stained tissue sample into a representation of another type of stain.
Supervised learning: a machine learning paradigm that relies on paired class labels and data samples for training.
Unsupervised learning: a machine learning paradigm that does not require labeled input for training.
Virtual staining: a computational reproduction of a histochemical stain.
Whole-slide image (WSI): a single high-resolution digital file created by scanning a complete microscope slide.

Often, several stains are used for each tissue sample to bring out all the necessary elements under investigation in a particular sample (Figure 2). In practice, each chemical stain requires a section of tissue because most stains are permanent and cannot be performed at the same time. It is therefore typical to use adjacent tissue sections for each type of stain and to compare the output on the different sections. Although this is sufficient for many applications, it requires a large amount of tissue material, which can be a limiting factor with small samples such as needle biopsies or scarce research samples. Furthermore, it loses the potential for single-cell resolution where several signals are read from the same tissue section.

For histological staining procedures, the samples are most often fixed and attached to glass slides (Figure 2). For solid tissues, this requires tissue embedding after fixation, followed by cutting thin sections, which can then be stained. The laboratory protocols for both tissue processing and histological staining are highly routine in clinical diagnostics and have remained similar for decades, with the slight added convenience of increased automation of the staining process itself. Histological stains require specified laboratory infrastructure and specially trained staff for a multitude of hands-on steps. In addition, many of the stains contain hazardous chemicals and thus produce hazardous waste, and the staining protocols consume a large amount of water. Further, many of the steps in the process require manual labor, are expensive, and expose the workforce to the chemicals.

The principle of virtual staining entails that the chemical staining process can be avoided either partially or completely by producing the stain signal computationally (Figure 2). This relies on deep learning, where an algorithm is trained to convert an input image without the desired stain

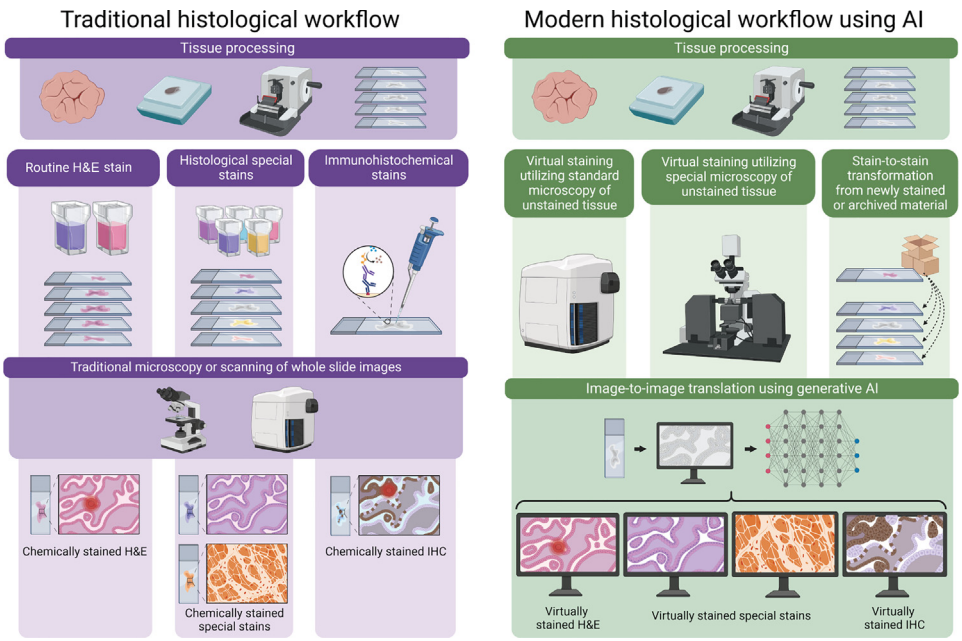

Figure 2. A traditional histological workflow compared to a modern histological workflow using artificial intelligence (AI). Although tissue processing remains the same, virtual staining methods can avoid the need for multiple tissue slides for multiple histological staining methods, including hematoxylin and eosin (H&E), special stains, and immunohistological staining. Traditional microscopy or scanning of individual slides for each stain can be replaced by multiple AI-translated stains of a single tissue section. Abbreviation: IHC, immunohistochemistry. Figure created with BioRender.

to a virtually stained output. Within this concept, there are several options for what is used as an input and what is the desired output, ranging from an image of an unstained, label-free tissue to an image of a tissue with another stain, and the desired output being one or several virtually stained versions of the same input image. Similarly, there are several options for how the input images are obtained in terms of the type of microscopy or spectroscopy (Box 1), and the use of different imaging techniques can contribute to the type and accuracy of the output images and the performance of the virtual staining procedure.

The development of a reliable virtual staining model often involves acquiring and processing a large volume of data and carefully designing and training the neural networks. The workflow necessary to develop a label-free virtual staining or a stain-to-stain transformation model typically consists of image data collection and image preprocessing, as well as network training, computational validation, histological validation, and task-specific validation. For solutions aiming for clinical use, clinical validation is also required.

Computational approaches for virtual staining

Virtual staining as a computational task corresponds to image-to-image transformation. The transformation employed is learned through generative neural network models from training data, where the target image is a chemically stained tissue section image, and the source image is a digital image of an unstained tissue section. Modeling can be done either through **supervised learning**, assuming pixelwise spatial correspondence between the source and target images, or through **unsupervised learning**, where the transformation between source and target images is learned from the content of the images without assuming pixelwise correspondence (Figure 3). Style

Box 1. Imaging modalities for virtual staining of unstained tissue samples

Light microscopy is by far the simplest and most applicable solution for most laboratories because it is cheap and does not rely on specific light sources or detection systems. The equipment can be long-lived and, in principle, the same equipment can be used for virtual staining applications and for traditional microscopy. Brightfield imaging has been trialed for multiple virtual staining methods using label-free imaging, especially for H&E staining [4,5,40]. Most current brightfield scanners are optimized for high-contrast samples, which is currently a challenge for unstained image acquisition. However, the techniques are increasingly available: it is a matter of launching sufficient and suitable equipment onto the market or providing sufficient possibilities for users to tweak settings.

In addition to brightfield imaging, several other techniques have been utilized and tested to enable label-free imaging for virtual staining. Many tissue components such as collagen, elastin, and red blood cells emit autofluorescence when excited with UV or visible light. Hence, autofluorescence can be utilized in assessing tissue morphology without staining the sections. Autofluorescence has been utilized for virtual H&E, MTC, and Jones staining of a variety of tissues [14,27,49,50]. Quantitative phase imaging (QPI) is a relatively new imaging technique for obtaining quantitative measurements of the optical phase delays and their distributions in label-free samples. It has also been utilized for virtual H&E, Jones, and MTC staining [14]. In addition, highly specialized imaging has been trialed for H&E, such as reflectance confocal microscopy [49] and photoacoustic microscopy [51,52].

Specialized imaging techniques that can provide more details in terms of high magnification may provide enhanced recognition and separation of tissue components and pathological variations compared to standard microscopy techniques. Techniques that provide microstructural information in a non-destructive manner can provide added value to current workflows, and even expand the applicability of conventional histology to non-destructive and 3D histological imaging techniques, including X-ray histology, phase contrast micro-computerized tomography (micro-CT), and non-sectioning light microscopy such as light sheet microscopy [53–55]. Even spectroscopy techniques have been utilized for virtual staining of unlabeled tissue, and Fourier transform IR (FTIR) spectroscopy is an example [56].

Although many of the specialized techniques can provide enhanced detection and separation of components in the unlabeled tissue, their utilization requires specialized equipment and may be rate-limiting for collecting large volumes of data for training. Their large-scale utilization is currently also challenged because dedicated equipment is required, resulting in limited access and high investment costs. However, as the imaging technologies advance and the equipment becomes more cost-effective, they also hold promise beyond research use.

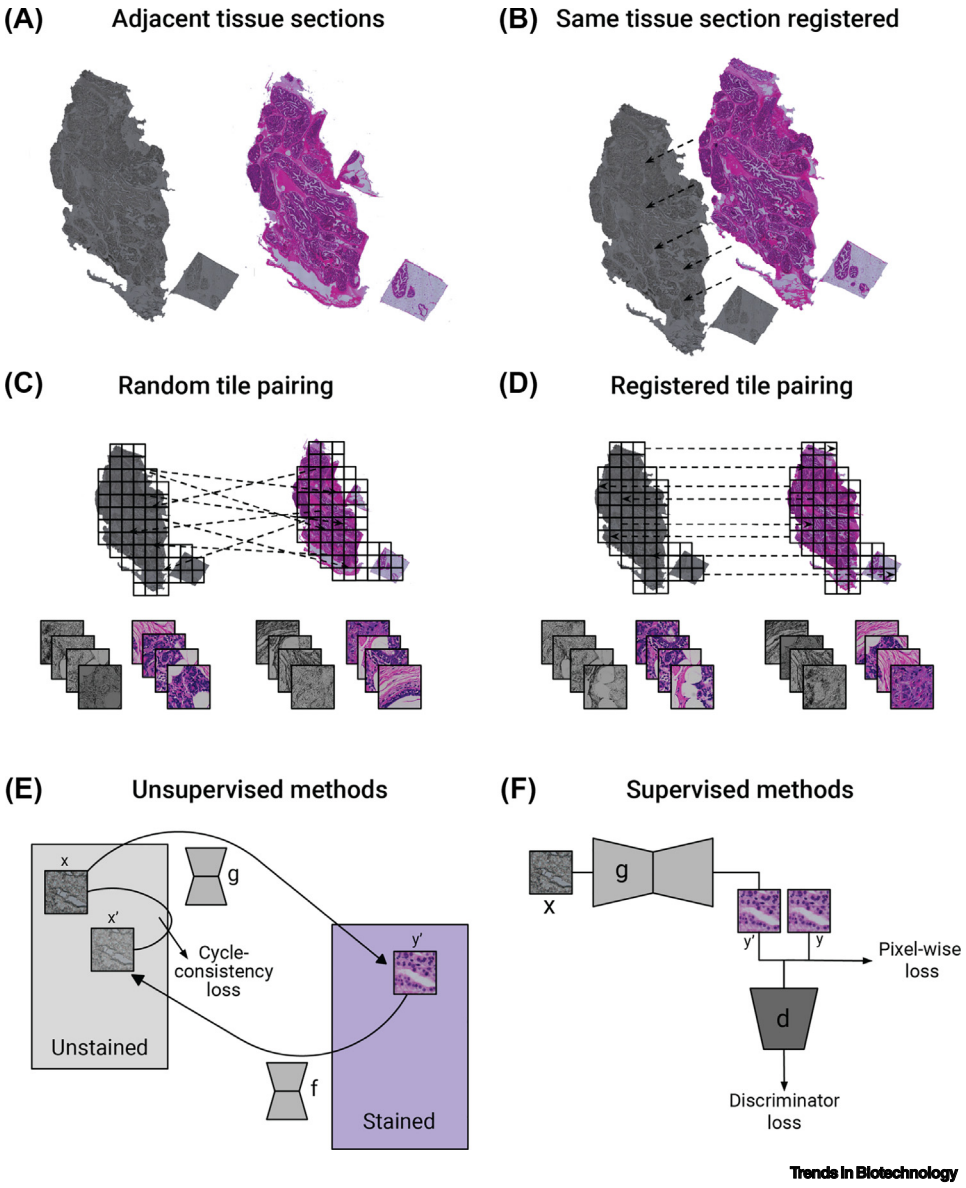

Figure 3. Computational concept of unsupervised and supervised virtual staining methods. The two columns represent the workflow of unsupervised and supervised virtual staining. (A) Adjacent tissue sections can be used to curate unstained–stained tissue image pairs because pixel-to-pixel correspondence is not required between the image pair. (B) The same tissue section is imaged once before staining, and once after, to curate unstained and stained tissue image pairs with pixel-to-pixel correspondence which is achieved through image registration. (C) The tissue images are split into smaller tiles that are randomly paired. (D) The registered tissue images are split into smaller tile pairs, keeping their alignment intact during the process. (E) Unsupervised methods typically use multi-generator and -discriminator generative adversarial network (GAN) models [forward GAN (g), backward GAN (f)], and rely on distribution matching loss functions such as the cycle-consistency loss in CycleGAN to learn the unstained to stained tissue image mapping. (F) Supervised methods use simpler incarnations of GAN such as conditional GAN with a single generator (g) and discriminator (d); in addition to adversarial training, they rely on pixelwise loss functions to learn the unstained to stained tissue image translation.

transfer-based image-to-image transformations without direct spatial correspondence have been successfully applied in virtual staining. The most commonly used method for this purpose is based on a generative adversarial network (GAN), the so-called CycleGAN (Box 2).

Box 2. CycleGAN is the most popular unsupervised deep learning method for virtual staining

CycleGAN is a style transfer-based image-to-image transformation method that does not assume direct spatial, pixelwise correspondence. CycleGAN learns to generate a synthetic image resembling real staining by using two generator and discriminator models. For example, Meng and colleagues [27] used CycleGAN to virtually stain ovarian cancer tissue, which they validated through assessment by specialists in cancer detection. Pradhan and colleagues [57] used a variant of CycleGAN, called Cycle ConditionalGAN, to virtually stain non-linear multimodal (NLM) images consisting of coherent anti-Stokes Raman scattering, two-photon excitation fluorescence, and second-harmonic generation. Through quantitative evaluation metrics they were able to establish the efficacy of virtual staining and showed its potential for diagnostic applications [57]. More recently, Koivukoski and colleagues [4] used CycleGAN to identify the optimal tissue processing and imaging protocols for virtual H&E staining. Although CycleGAN has been used for virtual staining of unstained tissue in several studies, it has been more commonly used for other image-to-image translation applications in histology, wherein the translation task is relatively style-intensive, such as stain-to-stain translation [23,30,35,39] and stain normalization [58–61].

As is commonly the case in machine learning, guiding the process with supervision allows more detailed information to be incorporated about the dependencies between source and target images, and several studies have reported more accurate transformations using supervised methods such as Pix2pix (Box 3). The pixel-level spatial correspondence required by supervised methods necessitates the alignment of different tissue sections with varying rotations, which can be done computationally but is challenging due to variations in the images caused by physical handling of tissue and the effect of chemicals during the staining process, in addition to the inherent differences between the appearance of unstained and stained tissues. Several algorithms for multimodal histological **image registration** between different stains at the WSI level have been published [9] or presented as preliminary data [10], but registration between unstained and stained histology images has gained less attention [11]. The most accurate correspondence can be obtained by training algorithms on images where there are both unstained and chemically stained images of the same sections. After alignment, supervised methods can be utilized for learning the modeling from unstained to virtual staining, and preliminary data suggest that weak registration can also be beneficial for unsupervised methods [12]. Further, an interesting, albeit computationally challenging approach is to simultaneously learn the (staining) style transfer and alignment [13].

The success of training an image-to-image translation model, as in any statistical modeling, is strongly dependent on the quality and representativeness of the training data. Histology represents an infamously complex target because it is rich in spatial variance and subtle nuances in appearance can reflect significant differences in biological meaning. To enable successful translation from source to target, the training material should represent the full range of histological phenotypes present in the material where the model will be applied and used later on. The training dataset

Box 3. Supervised deep learning methods for virtual staining in histology

Guiding the deep learning process with supervision often produces more accurate transformations, which is why supervised methods have also produced successful results for virtual staining in histology. Supervised learning methods assume pixelwise spatial correspondence between source and target images, and therefore require computational registration between the images to match the alignment of images when different tissue sections are used. One of the most commonly used supervised image-to-image translation methods is Pix2pix or conditional GAN [62]. Over time, different variants have been proposed with improvements [63,64]; however, the key idea of adversarial training with aligned image pairs has remained the common denominator. Rivenson and colleagues [25] proposed a conditional GAN-based method called PhaseStain to virtually stain quantitative phase label-free images of human skin, kidney, and liver tissue, and used quantitative methods metrics to establish the high quality of virtually stained tissue images [25]. The clinical potential of conditional GAN-based virtual staining methods has been indicated by some studies wherein the virtually stained tissue images were validated in a downstream analysis task such as tumor detection and localization [22,26]. Khan and colleagues [5] showed that an experimental variant of pix2pix, dubbed DensePix2pix, inspired by DenseUNet [65], helped to alleviate hallucination artifacts and generated high-quality virtually stained tissue images. Zhang and colleagues [29] used class-conditional GANs to develop novel methods for stain blending and the detection of microstructured regions of staining that differ from the primary stain in a tissue image, and these features are only possible in the digital medium.

needs to be relatively large for successful modeling, in practice meaning WSI image pairs in the order of dozens, and preferably hundreds.

Training the models for image-to-image transformations for high-resolution tissue section images is computationally intensive. High-resolution WSIs of histological sections are too large to be applied in one piece, and the training is performed by dividing the image into patches, resulting in a large quantity of image tiles which can number in the millions during the training process for memory efficiency. Computation is carried out in parallel using graphic processors. Leveraging large training datasets requires access to efficient computing resources, preferably a **high-performance computing** environment. After training, using the model for the actual transformation is computationally less demanding. This makes virtual staining an appealing technology because model deployment can be performed widely. Inference using the trained model is applied for image tiles, and, depending on the intended use, the WSIs may or may not be constructed from the tile patches. However, in practice, reconstruction of virtually stained WSIs is likely to suffer from tile-level artifacts unless overlapping, averaging, tile blending, or other computational approaches for mitigating the tiling artifact are applied, as reported by several groups both in published and preliminary data [5,14–17].

Practical considerations for virtual staining approaches

One aim of virtual staining is to provide methods that simplify current tissue processing and imaging procedures, save chemicals, and decrease costs. Although the trained models are likely to provide a viable solution in this respect, the training phase is costly. Deep learning is a data-hungry method, and the amount of data required for training solid models is substantial. More detailed imaging in terms of high magnification and specialized techniques may provide enhanced recognition and separation of tissue components and pathological variations (Box 1), but the most simple and fast imaging techniques are currently the most feasible for both obtaining data for the training phase as well as for fast and simple imaging using the trained models.

Use of unlabeled tissue is the most chemical-sparing option for virtual staining. Training models with this approach is promising because, when an unstained, label-free image and a pair of chemically stained images are used, they can be imaged from the same actual sections, thereby ensuring a high likelihood of pixelwise correspondence between the images and enabling the use of supervised models. However, to build models for this area is tedious: there are practically no existing datasets because unstained imaging has not been widely used in histology before. Hence, all material for training new models needs to be newly produced, involving significant investment in material, scanning, staining, and rescanning of the slides.

Owing to the wide use, accessibility, and affordability of brightfield imaging, virtual staining from unlabeled tissue imaged with visible light microscopy is the most convenient and time- and chemical-sparing option for broad applicability in the near future. However, most high-throughput **slide scanner**s are currently not compatible with scanning of low-contrast samples such as unstained, label-free histology sections, meaning that scanning is currently relatively slow. This is likely to improve because existing hardware technologies can be optimized for scanning of low-contrast samples. For stain-to-stain transformations requiring separate sections for each sample, there is plenty of archived clinical material, especially from H&E-stained sections with other stains from adjacent or near sections, that can be used in training models. Utilizing these for virtual staining is tied to using unsupervised methods, but, given a sufficient amount of training data, well-performing models can be achieved with this approach. In addition to brightfield imaging, several other techniques have been utilized and tested for label-free imaging as a source for virtual staining (Box 1).

How to evaluate the performance of virtual staining

Evaluation of the success of virtual staining requires both computational and histological assessment of model performance. Quantitative evaluation as a computational task can take place through several means. Computational measures of image similarity at the pixelwise level are perhaps the most straightforward method for performance evaluation. Traditional methods including mean square error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) and variants thereof [18,19] enable quantitative evaluation of the correspondence and are simple to implement, whereas in the context of deep generative networks, methods such as Frechet's inception distance [20] enable measurement of the similarity in the latent feature space instead of direct pixelwise comparison. The weakness of these approaches is the lack of weighting towards histologically meaningful attributes in the images – optimizing pixelwise metrics does not guarantee satisfactory histology in the output. Thus, more biomedically meaningful measures have been proposed. For example, quantitative measures of image-derived features, such as cell or nucleus reproduction at the object and pixel level, have been reported [5,16,21].

Histological evaluation is performed through visual examination, where the similarity, applicability, and reliability of the virtual staining process are evaluated by histopathology experts (e.g., [5]) and by using task-specific quantitative metrics such as the accuracy of a diagnostic measure compared to the protocol currently in use (e.g., [22]). Because the goal of interpreting the histological samples is often a subsequent diagnostic task or decision-making process, it is important to consider this task in evaluating the quality and performance of virtual staining. Pathologists often interpret histology using a pattern recognition approach that interprets the structure and appearance of tissue at multiple scales. Pixel-level artifacts are not crucial if the images enable accurate interpretation of histology – and high pixel-level accuracy is not meaningful if the images cannot be used for interpreting tissue histology. Hence, visual evaluation by a histopathology expert with suitable domain expertise is crucial because clinical pathology, veterinary pathology, and various model organisms all have different requirements for the evaluation tasks. For example, in clinical assessment, the suitability of a model should be assessed specifically for the disease, target tissue, and task at hand to ensure that the model is suited for the intended use, for example in finding regions of interest, determining subtypes, or grading pathologies. Furthermore, if virtually stained images are to be used for computational downstream analyses such as feature extraction or learning-based tasks, the suitability of the model outputs should be carefully verified in each context. Moreover, only through visual interpretation is it possible to determine the presence and level of visual artifacts such as **hallucination** by generative AI [4,5]. Similarly to the implementation of virtual staining, the evaluation may also have an additional level if stain transformation is performed at the WSI level rather than on tiles. Reproduction of realistic staining appearance at the tile level does not guarantee success at the WSI level because tile artifacts may easily distinguish virtual staining from real chemical staining. Another aspect worth considering is to ensure that the tiles contain sufficiently large support areas for the model to learn to reproduce (diagnostically) relevant histological structures.

State of the art in replacing chemical stains for histology

H&E

H&E stain is routinely used to display the morphology of tissues, and stains the cell nuclei blue and the cytoplasm and extracellular matrix pink. H&E is by far the most common histological stain, and represents 80% of all stains performed globally [23], and is therefore the most commonly reported stain both as input and output in virtual staining research so far. For example, hyperspectral label-free images of lung tissue have been used to produce virtual H&E staining, and conditional GAN and visual analysis have demonstrated that the method holds promise [24]. Label-free quantitative phase images of human skin, kidney, and liver tissue were used to

generate virtual H&E, and produced virtually stained images of high quality, although the output was sensitive to granular details [25]. Brightfield imaging of histological samples prepared with standard tissue preparation procedures and imaged at 20× magnification can be used to produce high-quality virtual staining of H&E from multiple tissues with pix2pix-based models [4,5] (Figure 4). Although tissue sections of formalin-fixed, paraffin-embedded (FFPE) tissue with sections of standard thickness can be imaged within paraffin, imaging deparaffinized sections without mounting and coverslip seems to be most feasible [4].

Several studies have recently demonstrated the clinical potential of virtually stained H&E images. Rana and colleagues [22] used nonfluorescent imaging and a variant of conditional GAN25 to virtually apply H&E stain to deparaffinized FFPE prostate biopsy images. In addition to direct quantitative evaluation, they tested the output images for tumor segmentation and built an end-to-end deep learning pipeline for the automatic detection and localization of tumors on the virtually stained images [26]. In another study, virtual H&E staining of autofluorescence images of ovarian cancer tissue using CycleGAN produced results gave 93% accuracy in pathologist evaluation of ovarian cancer detection [27].

Histological special stains

Although H&E is the most common histological stain, several other chemical stains are also utilized for research and diagnosis of various diseases. Overall, most clinical histopathology laboratories can perform dozens of different special stains, each having a specific purpose. For example, MTC stains collagen fibers blue, and keratin and muscles red [28]. Virtual MTC staining has been produced from label-free tissue using autofluorescent images and quantitative phase

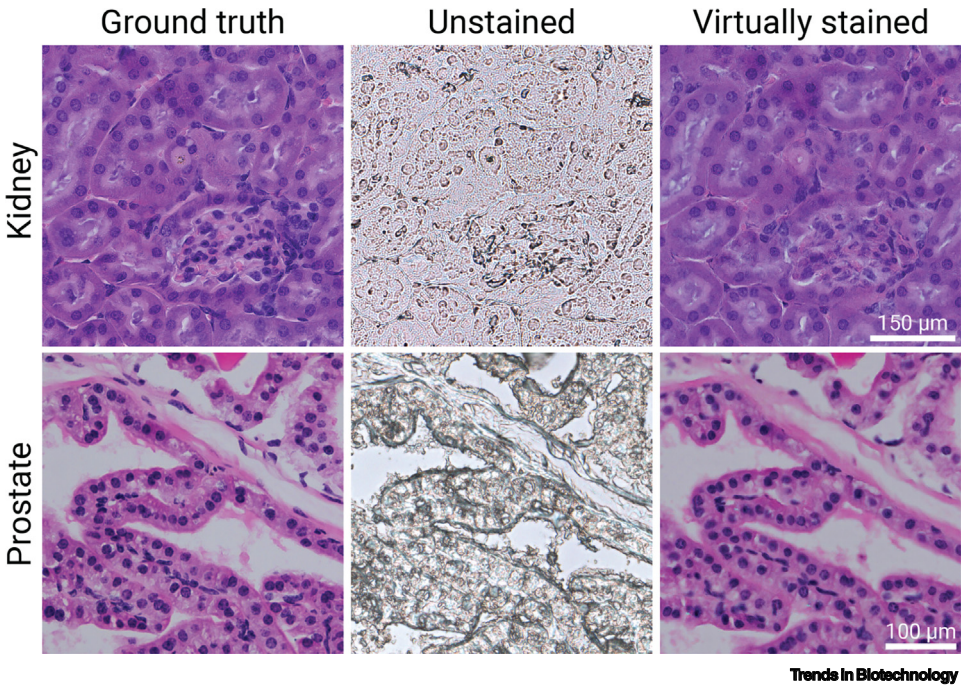

Figure 4. Examples of virtually stained hematoxylin and eosin (H&E) staining. Histological images of preclinical (murine) kidney and prostate tissue. Ground truth: chemically stained H&E. Unstained: tissue imaged with standard brightfield microscopy. Virtually stained: H&E staining produced using the Pix2Pix model as described in [4,5]. The images for each row are generated from a single section. Scale bars are indicated in each row.

imaging from, for example, kidney tissue [14,29]. Preliminary results of H&E to MTC conversion have been reported for samples of nonalcoholic steatohepatitis liver biopsies with CycleGAN [30]. Results for unpaired data using PG-GAN with evaluation by pathologists have been published [31], and for kidney needle core biopsy tissue with supervised learning and pathologist evaluation [23]. Trichrome-resembling virtual staining has also been trialed from H&E-stained slides imaged with dual-mode emission and transmission microscopy using clinical liver and kidney samples [32].

PAS stain labels carbohydrates (e.g., mucins and basement membranes) magenta, and nuclei blue, and is used, for example, to aid the diagnosis of basement membrane disease [28]. Reports of PAS staining in virtual staining mostly involve stain-to-stain translation using kidney tissue. A cascaded deep neural network (DNN) was used to virtually convert autofluorescence images of needle core biopsies to H&E and further to PAS [33]. Stain-to-stain translation of four different stains (Jones H&E, Sirius red, CD68, and CD34) with CycleGAN and StarGAN were further used in segmentation tasks [34]. Two studies have also utilized different IHC stains to produce PAS staining [35,36]. de Haan and colleagues [23] reported stain transformation of needle core biopsy tissue from H&E to PAS using supervised learning and pathologist evaluation.

Immunohistochemical stains

In addition to H&E and special histological stains, tissue research and diagnosis of many diseases relies on the presence or absence of specific proteins, which can be identified by immunological reactions based on antibodies. Antibodies specifically recognize an antigen sequence of a particular amino acid composition and structure, and are thus able to recognize only a specific protein or even a protein isoform. In histology, antibody-based detection is most often assessed with IHC, where a visible precipitate is formed at the antigen-residing sites of the tissue through the use of enzymatic reporters bound to the antibodies. Another method, immunofluorescence, utilizes fluorescence-based detection that requires compatible, specific microscopy for detection, and is common in research applications owing to superior resolution of signal location and easier multiplexing (i.e., assessing several antigens at once).

Setting up virtual staining for antibody-based techniques is a more challenging approach than for chemical stains because the patterns of staining depend on the expression and localization of a single molecular type that can vary for each protein, tissue, and pathology. Hundreds, if not thousands, of IHC markers are already used in the diagnostics of different diseases, and even more are used in research. In principle, IHC requires the most task-specific work in training deep learning models, and the prospects for tissue-agnostic models for IHC patterns are perhaps not as high as for chemical stains. Hence, extra caution should be taken in utilizing IHC-based virtual staining models outside the training data and tissue type.

Virtual IHC staining of label-free tissue reported so far includes IHC autofluorescence images of human epidermal growth factor receptor 2 (HER2) in breast tissue [37]. HER2 is a biomarker used for breast cancer subtyping and guiding diagnostic decisions. Conditional GANs were used for virtual staining, and the results were assessed against chemically stained samples through quantitative analysis and evaluation by three pathologists.

One of the most common targets identified with IHC is Ki-67, a marker of active cellular proliferation that is expressed in the active phases of mitosis and correlates with the aggressiveness of most solid tumors [38]. Virtual Ki-67 staining has been trialed by CycleGAN from H&E staining of neuroendocrine tumors and breast tissue [39], and in mouse lung tissue [40]. Ki-67 IHC has also been used as the input in combination with CD8 IHC using cycleGAN on colorectal carcinoma

metastases in liver tissue to produce fibroblast activation protein and cytokeratin (FAP-CK) staining [41].

Zhang and colleagues [40] reported stain transformation from H&E to several different combinations of IHC stains, including ER/PR/HER2 in breast tissue and Ki-67/CC10/proSPC in mouse lung. Other examples of virtual conversion of chemical stains to IHC include H&E to HER2 IHC with pix2pix [42], H&E to CDX2 and CK818 on colon samples with SC-GAN (structural cycleGAN) [43], and GAN-based stain-translation of PAS to Col3 and CD31 IHC on kidney tissue images [44]. Fluorescent stains have also been trialed as input for virtual IHC staining; for example, Hoechst staining was converted to CD3 and CD8 IHC on samples of clear cell renal cell carcinoma [45].

Considerations and future potential of virtual staining in histology

Lack of standards in the development of virtual staining

Considering the multitude of imaging and computational approaches that can be used for virtual staining, comparison and evaluation of the solutions will soon become challenging. For example, slight modifications in a CycleGAN stain transfer architecture can produce highly distinct results from each other and from the real samples, leading to misleading conclusions being made by visual inspection or pretrained model evaluation [34]. As with any novel informatics field in biomedicine, there will soon be a need for standards both for the development phase of models and for how to report the imaging and tasks to support proper comparison of algorithm performance. Further standardization may be required for virtual staining products related to the uses they are suitable for.

The role of open data in developing AI for virtual staining

In addition to rapid advances in methodology and computing power, the availability of large, open-access datasets is driving the development in AI-based solutions for cutting-edge biomedical applications. The machine learning and AI community has a long history of using benchmark datasets and challenges to attract the interest of the developer community in emerging topics, including those in histopathology AI [10,46–48]. Datasets related to virtual staining remain scarce, and only a few datasets of WSI data from unstained tissue have been published [4,5]. This leaves room for improving the accessibility of the research area to computer scientists and the machine learning community. The availability of data also intimately links to standardization of virtual staining because unstandardized data significantly limit their usability.

Potential in multimodal research

Because the development of different virtual staining models will allow the production of several different staining outputs from a single input histological image, virtual staining will enable enhanced studies of tissue composition, phenotypes, and marker expression. Further, combining virtual staining with spatial molecular measurements, such as spatial transcriptomic or proteomic analyses of the same slides, will enable unprecedented analysis of tissues at single-cell resolution (Figure 5). It will be exciting to study whether, in the future, virtual modeling of such molecular and spatial measurements will become possible, thereby enabling multi-level, advanced multiplexing based only on unstained tissue images.

Potential clinical utility

In the clinic, omitting the chemical staining could save time in the diagnostic decision-making process. Because the time taken is often long when several additional stains are required following the first routine stains, methods that enable several virtual stains to be performed on a single slide have substantial potential for time saving and preventing delays in treatment decisions.

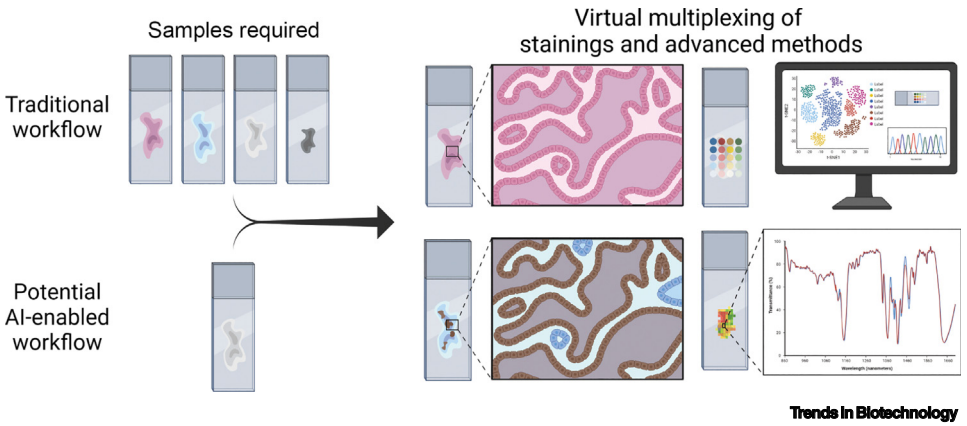

Figure 5. Multimodal virtual staining decreases sample need and allows virtual multiplexing. In traditional histological workflows, most staining and measurement modalities require a tissue section of their own. In the future, virtual staining using artificial intelligence (AI) may enable images to be produced not only of multiple chemical stains of a tissue section but also images of representative readouts of molecular methods such as spatial transcriptomics and spatial proteomics. Figure created with BioRender.

Virtual staining in clinical use could also have a global health impact through making histological assessments more feasible in resource-limited settings, especially in environments with limited supplies of water that is used in large quantities in the traditional laboratory workflow.

As with any methodology, larger clinical use of virtual staining would require retrospective validation and prospective testing, for each task separately, followed by product licensing. The technique currently remains in the development phase, and there are no accepted algorithms for clinical use or trials ongoing to the best of our knowledge. Similarly to other AI-driven tools in digital pathology, standardization is the key challenge to overcome to ensure the clinical utility of replacing standard H&E and other stains.

Concluding remarks

There is a need for alternatives to chemical staining procedures in histology. Deep learning-based virtual staining has emerged as an exciting new field of research and development, and holds promise to provide more sustainable, rapid, accurate, and cost-effective solutions to histopathological analysis. How these can be made fast, reliable, and easy to utilize, keeping in mind the required standards, will remain a challenge for the upcoming years, and several unresolved issues remain (see Outstanding questions).

Given that the early results in the field are encouraging, further research will be necessary to determine how widely virtual staining can be reliably implemented for different chemicals, antibodies, tissues, and scanner environments. The most common chemical stain (H&E) has been most extensively studied for virtual staining. Success in computationally predicting this staining suggests that staining of tissue structure and morphology can be performed virtually. However, more specialized stains may reveal tissue structures that the unstained source images do not contain information about, but that would enable the model to learn unambiguous mapping from source to target. Even though tissue-agnostic models should be our goal, it is possible that not all stains can be virtualized for all tissues, and learning these limitations is a major outstanding question for further research. On the other hand, using AI-based models, it is possible that additional information can be utilized beyond that required for visual interpretation, potentially leading to higher accuracies of interpretation. Given the rapid pace of development of AI, advances in image-to-

Outstanding questions

What are the requirements of model performance that would allow it to adequately replicate the chemical staining for each desired task?

What are the requirements of imaging (method, magnification, image resolution) that would produce virtual staining of sufficient quality?

Which chemical stains can be virtually produced, and what volume of training data is required?

Which antibody-based stains can be reliably produced as virtual staining?

Which are the best or even suitable quantitative evaluation metrics for virtual staining performance in histology?

How should model building and evaluation of virtual staining be standardized for routine use?

image translation as well as translation from other modalities to images will emerge, generating new possibilities for virtual staining and other AI-enabled cross-modality transformations.

Many of the virtual staining efforts reported so far have been performed using relatively small datasets, where the number of tiles may be computationally sufficient, but the number of biological or clinical samples is low and does not sufficiently represent normal variation in tissue. Current progress should thus be taken as a collection of proof-of-principle studies that should be verified with larger and more heterogeneous datasets. In addition to using sufficient amounts of unbiased, real-life data, keys to success will include involving appropriate domain expertise in model building and validation. These strategies would tackle the inherent limitations of the technology engendered by the use of unmatured models based on biased datasets, as well as AI-derived artifacts such as hallucinations. Efforts should be invested in creating open repositories and datasets that are sufficiently large to generate broadly verifiable results and allow efficient model comparisons. International standardization is also warranted, with integral links to current considerations regarding responsible use of AI in medicine.

Although the clinical field requires highly standardized, licensed solutions for visible light microscopy, fundamental research would benefit from easily accessible and broad-range models where images of any size and resolution could be utilized. Open research would be promoted by free-to-use online solutions or add-ins for the image analysis platforms commonly used in the biomedical field. Mobile apps and microscopes enabled with augmented reality solutions would simplify histological assessments and provide access to the technology even in remote and resource-limited settings.

Histochemical stains are designed to help the human eye to evaluate a given tissue section for the presence or absence of pathology. However, if algorithms for the desired tasks can perform equally well on unstained tissue, the visual representations may eventually become unnecessary. This could enable end-to-end solutions that provide diagnostic decisions directly at the imaging stage. If technologies can provide sufficient information for decision support using unfixed tissue (e.g. from surgery), even virtual staining could eventually become unnecessary or merely provide visual interface and serve as a tool for explainability.

Acknowledgments

The research of the authors is supported by the Research Council of Finland (357490 and 359230 to L.L., and 341967 and 359229 to P.R.), European Research Area (ERA) PerMed (L.L. and P.R.), Cancer Foundation Finland (L.L. and P.R.), Emil Aaltosen Säätiö (S.K.), and the University of Turku Graduate School (U.K.).

Declaration of interests

L.L., S.K., and U.K. declare no competing interests. P.R. is a founder and shareholder in Louhi Health Data.

References

1. LeCun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436–444
2. Topol, E.J. (2019) High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* 25, 44–56
3. Niazi, M.K.K. *et al.* (2019) Digital pathology and artificial intelligence. *Lancet Oncol.* 20, e253–e261
4. Koivukoski, S. *et al.* (2023) Unstained tissue imaging and virtual hematoxylin and eosin staining of histologic whole slide images. *Lab. Investig.* 103, 100070
5. Khan, U. *et al.* (2023) The effect of neural network architecture on virtual H&E staining: systematic assessment of histological feasibility. *Patterns* 4, 5
6. Christiansen, E.M. *et al.* (2018) In silico labeling: predicting fluorescent labels in unlabeled images. *Cell* 173, 792–803
7. Imboden, S. *et al.* (2023) Trustworthy in silico cell labeling via ensemble-based image translation. *Biophys. Rep. (N Y)* 3, 100133
8. Wang, Z. *et al.* (2021) Global voxel transformer networks for augmented microscopy. *Nat. Mach. Intell.* 3, 161–171
9. Borovec, J. *et al.* (2020) ANHIR: automatic non-rigid histological image registration challenge. *IEEE Trans. Med. Imaging* 39, 3042–3052
10. Weitz, P. *et al.* (2023) The ACROBAT 2022 challenge: automatic registration of breast cancer tissue. *ArXiv*, Published online May 29, 2023. https://doi.org/10.48550/arXiv.2305.18033
11. Borovec, J. *et al.* (2013) Registration of multiple stained histological sections. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pp. 1034–1037, IEEE

12. Tweel, J.E. *et al.* (2023) Virtual histology with photon absorption remote sensing using a cycle-consistent generative adversarial network with weakly registered pairs. *ArXiv*, Published online June 26, 2023. http://dx.doi.org/10.48550/arXiv.2306.08583
13. Honkamaa, J. *et al.* (2023) Deformation equivariant cross-modality image synthesis with paired non-aligned training data. *Med. Image Anal.* 90, 102940
14. Rivenson, Y. *et al.* (2019) Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat. Biomed. Eng.* 3, 466–477
15. Levy, J.J. *et al.* (2021) A large-scale internal validation study of unsupervised virtual trichrome staining technologies on nonalcoholic steatohepatitis liver biopsies. *Mod. Pathol.* 34, 808–822
16. Kaza, N. *et al.* (2022) Virtual staining, segmentation, and classification of blood smears for label-free hematology analysis. *BME Front.* 2022, 9853606
17. Asaf, M.Z. *et al.* (2023) Dual contrastive learning based image-to-image translation of unstained skin tissue into virtually stained H&E images. *Sci. Rep.* 14, 2335
18. Wang, Z. *et al.* (2004) Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612
19. Wang, Z. *et al.* (2003) Multiscale structural similarity for image quality assessment. In *Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, pp. 1398–1402, IEEE
20. Heusel, M. *et al.* (2017) GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (von Luxburg, U. *et al.*, eds), pp. 6629–6640, Curran Associates
21. McNeil, C. *et al.* (2023) An end-to-end platform for digital pathology using hyperspectral autofluorescence microscopy and deep learning based virtual histology. *Mod. Pathol.* 37, 100377
22. Rana, A. *et al.* (2020) Use of deep learning to develop and analyze computational hematoxylin and eosin staining of prostate core biopsy images for tumor diagnosis. *JAMA Netw. Open* 3, e205111
23. de Haan, K. *et al.* (2021) Deep learning-based transformation of H&E stained tissues into special stains. *Nat. Commun.* 12, 4884
24. Bayramoglu, N. *et al.* (2017) Towards virtual H&E staining of hyperspectral lung histology images using conditional generative adversarial networks. In *IEEE International Conference on Computer Vision Workshops*, pp. 64–71, IEEE
25. Rivenson, Y. *et al.* (2019) PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. *Light Sci. Appl.* 8, 23
26. Bayat, A. *et al.* (2021) Automated end-to-end deep learning framework for classification and tumor localization from native non-stained pathology images. In *Medical Imaging 2021: Image Processing*, pp. 43–54, International Society for Optics and Photonics
27. Meng, X. *et al.* (2021) A computationally virtual histological staining method to ovarian cancer tissue by deep generative adversarial networks. *Comput. Math. Methods Med.* 2021, 4244157
28. Ovalle, W.K. and Nahirney, P.C. (2020) *Netter's Essential Histology: With Correlated Histopathology*, Elsevier
29. Zhang, Y. *et al.* (2020) Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. *Light* 9, 78
30. Levy, J.J. *et al.* (2020) Preliminary evaluation of the utility of deep generative histopathology image translation at a mid-sized NCI cancer center. *BioRxiv*, Published online January 17, 2020. https://doi.org/10.1101/2020.01.07.897801
31. Yan, R. *et al.* (2023) Unpaired virtual histological staining using prior-guided generative adversarial networks. *Comput. Med. Imaging Graph.* 105, 102185
32. Fereidouni, F. *et al.* (2019) Dual-mode emission and transmission microscopy for virtual histochemistry using hematoxylin- and eosin-stained tissue sections. *Biomed. Opt. Express* 10, 6516–6530
33. Yang, X. *et al.* (2022) Virtual stain transfer in histology via cascaded deep neural networks. *ACS Photon.* 9, 3134–3143
34. Vasiljević, J. *et al.* (2022) CycleGAN for virtual stain transfer: is seeing really believing? *Artif. Intell. Med.* 133, 102420
35. Gadermayr, M. *et al.* (2018) Which way round? A study on the performance of stain-translation for segmenting arbitrarily dyed histological images. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (Frangi, A. *et al.*, eds), pp. 165–173, Springer
36. Bouteldja, N. *et al.* (2022) Improving unsupervised stain-to-stain translation using self-supervision and meta- learning. *J. Pathol. Inform.* 13, 100107
37. Bai, B. *et al.* (2022) Label-free virtual HER2 immunohistochemical staining of breast tissue using deep learning. *BME Front.* 2022, 9786242
38. Li, L.T. *et al.* (2015) Ki67 is a promising molecular target in the diagnosis of cancer. *Mol. Med. Rep.* 11, 1566–1572
39. Liu, S. *et al.* (2021) Unpaired stain transfer using pathology-consistent constrained generative adversarial networks. *IEEE Trans. Med. Imaging* 40, 1977–1989
40. Zhang, R. *et al.* (2022) MVFStain: multiple virtual functional stain histopathology images generation based on specific domain mapping. *Med. Image Anal.* 80, 102520
41. Lahiani, A. *et al.* (2019) Virtualization of tissue staining in digital pathology using an unsupervised deep learning approach. In *Digital Pathology: ECDP 2019* (Reyes-Aldasoro, C. *et al.*, eds), pp. 47–55, Springer
42. Liu, S. *et al.* (2022) BCI: breast cancer immunohistochemical image generation through Pyramid Pix2pix. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1814–1823, IEEE
43. Dubey, S. *et al.* (2024) Structural Cycle-GAN for virtual immunohistochemistry staining of gland markers in the colon. In *Machine Learning in Medical Imaging MLMI 2023* (Cao, X. *et al.*, eds), pp. 447–456, Springer
44. Gadermayr, M. *et al.* (2019) Generative adversarial networks for facilitating stain-independent supervised and unsupervised segmentation: a study on kidney histology. *IEEE Trans. Med. Imaging* 38, 2293–2302
45. Wölflein, G. *et al.* (2023) HoechstGAN: virtual lymphocyte staining using generative adversarial networks. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 4986–4996, IEEE
46. Bejnordi, B.E. *et al.* (2017) Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 318, 2199–2210
47. Kumar, N. *et al.* (2019) A multi-organ nucleus segmentation challenge. *IEEE Trans. Med. Imaging* 39, 1380–1391
48. Bulten, W. *et al.* (2022) Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. *Nat. Med.* 28, 154–163
49. Li, X. *et al.* (2021) Unsupervised content-preserving transformation for optical microscopy. *Light Sci. Appl.* 10, 44
50. Picon, A. *et al.* (2021) Autofluorescence image reconstruction and virtual staining for in vivo optical biopsying. *IEEE Access* 9, 32081–32093
51. Kang, L. *et al.* (2021) Deep learning enables ultraviolet photoacoustic microscopy based histological imaging with near real-time virtual staining. *Photoacoustics* 25, 100308
52. Cao, R. *et al.* (2023) Label-free intraoperative histology of bone tissue via deep-learning-assisted ultraviolet photoacoustic microscopy. *Nat. Biomed. Eng.* 7, 124–134
53. Saccomano, M. *et al.* (2018) Synchrotron inline phase contrast μCT enables detailed virtual histology of embedded soft-tissue samples with and without staining. *J. Synchrotron Radiat.* 25, 1153–1161
54. Katsamenis, O.L. *et al.* (2019) X-ray micro-computed tomography for nondestructive three-dimensional (3D) X-ray histology. *Am. J. Pathol.* 189, 1608–1620
55. Bishop, K.W. *et al.* (2024) An end-to-end workflow for nondestructive 3D pathology. *Nat. Protoc.*, Published online January 23, 2024. https://doi.org/10.1038/s41596-023-00934-4
56. Mayerich, D. *et al.* (2015) Stain-less staining for computed histopathology. *Technology (Singap. World Sci.)* 3, 27–31
57. Pradhan, P. *et al.* (2021) Computational tissue staining of nonlinear multimodal imaging using supervised and unsupervised deep learning. *Biomed. Opt. Express* 12, 2280–2298
58. de Bel, T. *et al.* (2018) Stain-transforming cycle-consistent generative adversarial networks for improved segmentation of renal

histopathology. In *Proceedings of the International Conference on Medical Imaging with Deep Learning* (Vol. 102) (Cardoso, M.J. *et al.*, eds), pp. 151–163, MLResearchPress
59. de Bel, T. *et al.* (2021) Residual cyclegan for robust domain transformation of histopathological tissue slides. *Med. Image Anal.* 70, 102004
60. Runz, M. *et al.* (2021) Normalization of HE-stained histological images using cycle consistent generative adversarial networks. *Diagn. Pathol.* 16, 71
61. Baykal Kablan, E. (2023) Regional realness-aware generative adversarial networks for stain normalization. *Neural Comput. Applic.* 35, 17915–17927
62. Isola, P. *et al.* (2017) Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 5967–5976, IEEE
63. Wang, C. *et al.* (2020) Discriminative region proposal adversarial networks for high-quality image-to-image translation. *Int. J. Comput. Vis.* 128, 2366–2385
64. Wang, T.C. *et al.* (2018) High-resolution image synthesis and semantic manipulation with conditional gans. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8798–8807, IEEE
65. Dong, R. *et al.* (2019) DenseU-net-based semantic segmentation of small objects in urban remote sensing images. *IEEE Access* 7, 65347–65356