Virtual staining of label-free tissue in imaging mass spectrometry

Zhang, Yijie; Huang, Luzhe; Pillar, Nir; Li, Yuzhu; Li, Yuhang; Migas, Lukasz G.; Van de Plas, Raf; Spraggins, Jeffrey M.; Ozcan, Aydogan

DOI
10.1126/sciadv.adv0741

Publication date
2025

Document Version
Final published version

Published in
Science Advances

Citation (APA)
Zhang, Y., Huang, L., Pillar, N., Li, Y., Li, Y., Migas, L. G., Van de Plas, R., Spraggins, J. M., & Ozcan, A. (2025). Virtual staining of label-free tissue in imaging mass spectrometry. *Science Advances*, *11*(31), Article eadv0741. https://doi.org/10.1126/sciadv.adv0741

APPLIED SCIENCES AND ENGINEERING

Virtual staining of label-free tissue in imaging mass spectrometry

Yijie Zhang[1,2,3]†, Luzhe Huang[1,2,3]†, Nir Pillar[1,2,3], Yuzhu Li[1,2,3], Yuhang Li[1,2,3], Lukasz G. Migas[4,5], Raf Van de Plas[4,5,6], Jeffrey M. Spraggins[4,6,7,8,9], Aydogan Ozcan[1,2,3,10]*

Imaging mass spectrometry (IMS) enables untargeted, highly multiplexed mapping of molecular species in biological tissue with unparalleled chemical specificity and sensitivity. However, most IMS platforms lack microscopy-level spatial resolution and cellular morphological contrast, necessitating subsequent histochemical staining, microscopic imaging, and advanced image registration to correlate/link molecular distributions with specific tissue features and cell types. We present a diffusion model–based virtual histological staining approach that enhances spatial resolution and digitally introduces cellular morphological contrast into mass spectrometry images of label-free human tissue. Blind testing on human kidney tissue demonstrated that the virtually stained images of label-free samples closely match their histochemically stained counterparts (with periodic acid–Schiff staining), showing high concordance in identifying key renal pathology structures despite using IMS data with 10-fold larger pixel size. Additionally, our approach uses optimized noise sampling during the diffusion model's inference to achieve reliable and repeatable virtual staining. We believe this virtual staining method will open avenues for IMS-based biomedical research.

INTRODUCTION

Imaging mass spectrometry (IMS) (*1*) is a powerful tool for spatial biology, enabling the discovery of intricate relationships between molecular distributions and key tissue features and cell types (*2*–*4*). IMS offers in situ pixel-wise mapping of hundreds to thousands of molecular species with high chemical specificity and sensitivity. By combining mass spectrometric analysis with spatial mapping, IMS has great potential as a discovery platform for biomedical research. Although multiple surface sample approaches have been used for imaging, matrix-assisted laser desorption/ionization (MALDI) is one of the most common IMS technologies due to its applicability to a wide range of biomolecular classes and high–spatial resolution capabilities. Commercially available MALDI IMS platforms can routinely achieve 5- to 30-μm pixel sizes (*5*–*7*), and custom platforms have been developed to enable imaging at cellular resolution with pixel sizes approaching 1 μm (*8*–*10*). Briefly, MALDI IMS is performed by mounting a thin tissue section to a glass slide and coating it with a ultraviolet-absorbing chemical matrix, which assists with the desorption and ionization of endogenous molecules during laser irradiation. The sample stage is scanned from one location to another, recording a mass spectrum at every pixel. An intensity heatmap is then plotted across the measurement region, resulting in an ion image for each molecule detected. MALDI IMS spatial resolution is determined by several factors, including the laser diameter at the sample surface, the precision of the stage movement between the pixels (i.e., pitch), and the extent of molecular delocalization during the sample preparation (*5*, *11*, *12*).

While MALDI IMS provides broad molecular coverage and high chemical specificity, it lacks inherent histological context, making it challenging to directly link molecular profiles to precise cellular features without additional information. The primary reasons for this are (i) the relatively low spatial resolution of IMS and (ii) the absence of color and cellular morphological contrast that experts are familiar with. Both limitations can be addressed by multimodal methods integrating IMS data with optical microscopy, often requiring advanced experimental and computational methods. For example, IMS–microscopy integrative methods have been developed to enhance human interpretability of IMS data through applications such as spatial sharpening of IMS data (*13*), out-of-sample prediction of molecular distributions (*14*), mining of IMS-derived molecular profiles based on microscopy features (*15*), and others (*16*–*18*). Each of these examples requires the collection of complimentary optical microscopy data such as histochemically stained (HS) bright-field or immunofluorescence images. However, in cases where chemical staining is necessary, it can introduce additional experimental complexity, posing a major limitation, as it precludes the preservation of the unlabeled tissue, thereby hindering any further molecular analysis on the same sample. Although serial tissue sections can be used, this requires complicated coregistration processes to account for tissue heterogeneity and cellular differences between adjacent sections. Therefore, a technique capable of predicting high-resolution cellular histopathology cues on the basis of low-resolution IMS data is urgently needed to streamline workflows and improve molecular histology.

Over the past decade, generative artificial intelligence (AI) models have made significant strides, finding wide-ranging applications for vision analysis. In the biomedical domain, one of the most notable uses of these models has been the virtual histological staining of label-free tissues, where generative models are trained to transform

[1]Electrical and Computer Engineering Department, University of California, Los Angeles, CA 90095, USA. [2]Bioengineering Department, University of California, Los Angeles, CA 90095, USA. [3]California NanoSystems Institute (CNSI), University of California, Los Angeles, CA 90095, USA. [4]Mass Spectrometry Research Center, Vanderbilt University, Nashville, TN 37232, USA. [5]Delft Center for Systems and Control, Delft University of Technology, 2628 Delft, Netherlands. [6]Department of Biochemistry, Vanderbilt University, Nashville, TN 37232, USA. [7]Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN 37232, USA. [8]Department of Chemistry, Vanderbilt University, Nashville, TN 37232, USA. [9]Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN 37232, USA. [10]Department of Surgery, University of California, Los Angeles, CA 90095, USA.
*Corresponding author. Email: ozcan@ucla.edu
†These authors contributed equally to this work.

microscopic images of label-free tissues into their HS (*19*–*30*) or fluorescence stained counterpart (*31*–*35*). Most of these approaches rely on label-free imaging modalities that exhibit a spatial resolution comparable to the resolution of the images of the stained tissue captured by a digital histology scanner, which serves as the ground truth (GT) from the perspective of the clinical workflow. Recent efforts have also generated virtual histopathology images using segmentation masks (*36*), domain-specific knowledge, and tissue genomics (*37*).

In this work, we introduce a diffusion model–based virtual histological staining technique that transforms IMS-measured ion images reporting molecular species' distributions in label-free tissue samples into super-resolved bright-field microscopy images, closely matching their HS counterparts [as illustrated in Fig. 1 (A and B)]. Our approach uses an image-conditional diffusion model, underpinned by the Brownian bridge process (Fig. 1C) (*38*, *39*), which integrates the low-resolution conditional input with noise estimation through an attention-based U-Net (Fig. 1, F to H) (*40*) that incorporates the time-step information to reconstruct the high-resolution histological image of the label-free sample. Following a one-time training effort, the diffusion-based virtual staining (VS) model was able to generate bright-field microscopy equivalent periodic acid–Schiff (PAS)–stained images from IMS-measured ion images of label-free human kidney tissue samples (never seen before) despite the fact that the input IMS data had a 10-fold larger pixel size. Quantitative evaluations confirmed that our VS approach effectively overcomes key limitations of IMS for histological interpretation by digitally creating high-resolution histological stain images using low-resolution IMS data without the need for chemical staining and microscopic imaging of stained tissue, also eliminating image registration steps because the

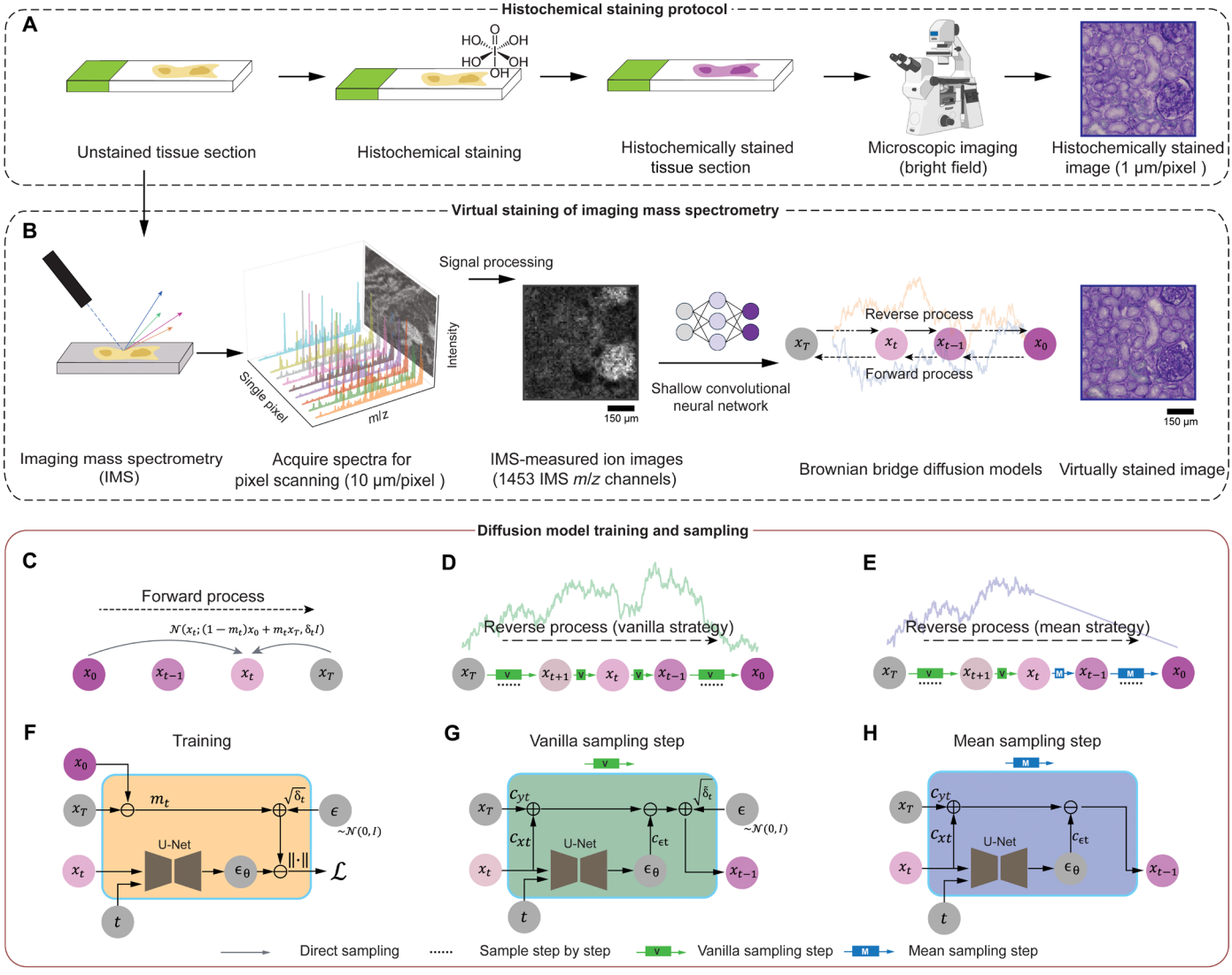

Fig. 1. Diffusion model–based VS of label-free IMS-measured ion images. (**A**) The workflow for histochemical staining of label-free tissue sections and the bright-field optical microscope scanning for digitization. (**B**) The diffusion model–based VS pipeline. The Brownian bridge process is used for both the forward and reverse processes. (**C**) Schematic diagram of the forward process of our Brownian bridge diffusion model (BBDM). (**D** and **E**) Schematic diagrams of the diffusion model reverse process that uses the vanilla and mean sampling strategies. (**F**) Detailed workflow for training the diffusion-based VS model. (**G** and **H**) Workflow for a single vanilla and mean sampling step, used individually or in combination during the reverse sampling process in (D) and (E).

virtually stained images are automatically registered with the IMS data used as the label-free input of the VS model. A board-certified pathologist was able to identify key pathological structures directly from the virtually stained PAS images, demonstrating a high degree of concordance with the features observed in the corresponding HS images and emphasizing the human interpretability of this approach. Furthermore, to mitigate the inherent high variance associated with diffusion models, we used an optimized noise sampling strategy, which eliminates additive random noise in the final stages of the reverse diffusion process. This approach not only quantitatively reduces output variance but also ensures that the pathological features in the diffusion model outputs from different test runs remain consistent and histologically equivalent. In summary, our diffusion-based VS approach overcomes some of the limitations of traditional IMS, including its relatively low spatial resolution and the absence of cellular morphological contrast, also eliminating the need for time-intensive histological staining process and complex image coregistration after the IMS data acquisition. We believe that this technique offers significant benefits for mass spectrometry–based molecular histology and will help accelerate IMS-enabled histological analyses in life sciences.

RESULTS

Virtual histological staining of IMS-measured ion images of label-free tissue using a diffusion model

The dataset used in this study comprises IMS-measured ion images of label-free human kidney tissues and high-resolution bright-field images of the HS versions of the same tissue samples. As shown in Fig. 1B, the IMS data of each tissue sample were acquired using pixel-wise raster scanning with 10-μm lateral spacing. Subsequent data processing selected the most representative channels (*41*), resulting in individual ion images containing 1453 mass/charge ratio (m/z) channels, each with a pixel size of 10 μm. After the IMS data acquisition, the label-free tissue slides were subjected to histopathological PAS staining and digitally imaged using a benchtop bright-field optical microscope. The bright-field images of the HS tissue samples were then registered to their corresponding ion images, which formed our label-free input (IMS) and GT (bright-field and labeled) image pairs. It is worth noting that we also introduced autofluorescence (AF) microscopy imaging acquired both before and following the IMS data acquisition to facilitate the registration between the IMS data and the bright-field microscopy images of HS tissue. Details of the dataset collection and preprocessing are provided in Methods.

The training and sampling process of the Brownian bridge diffusion model (BBDM) (*38*) reflect the two-directional propagation of a Brownian bridge process, as depicted in Fig. 1 (C to E). The forward process starts from $\mathbf{x_0}$, representing the target image domain of the HS bright-field images (GT), and progresses toward $\mathbf{x_T}$, the input image domain, corresponding to lower-resolution IMS images (with 1453 m/z channels), processed through a shallow convolutional neural network for image dimension alignment/match, as shown in Fig. 1B. The mean of the forward process is linearly scheduled from $\mathbf{x_0}$ to $\mathbf{x_T}$, while the variance evolves quadratically over time steps (t). In contrast, the reverse process aims to denoise the input IMS images of label-free tissue step by step, gradually refining the data without directly using the GT image $\mathbf{x_0}$. A U-Net–based denoising neural network (*38*, *40*) is trained to estimate the posterior mean of $\mathbf{x_t}$ based on $\mathbf{x_T}$. As shown in Fig. 1 (F to H), the denoising network is used to consistently estimate the difference between the current state $\mathbf{x_t}$ and the target image $\mathbf{x_0}$ at arbitrary time steps between 0 and T (see Methods). This Brownian bridge framework departs fundamentally from the classic denoising diffusion probabilistic model (DDPM) (*42*). In unconditional DDPM (*42*–*44*), the forward process degrades an initial sample $\mathbf{x_0}$ by adding Gaussian noise at each timestep according to a fixed variance schedule, ultimately producing a pure-noise state $\mathbf{x_T}$. A denoising neural network is then trained to invert this process: beginning from $\mathbf{x_T} \sim \mathcal{N}(0, \mathbf{I})$, it iteratively predicts the noise component and partially removes it, gradually reconstructing $\mathbf{x_0}$. When adapted for conditional image translation, the conditioning signal, whether a text embedding or an image, is concatenated with the noisy image at every forward and reverse step, ensuring that the model retains direct access to the context throughout the diffusion trajectory (*43*–*46*). In contrast, the BBDM reframes the forward process so that the terminal state is the conditional image $\mathbf{x_T}$ itself (the label-free IMS data). Noise is incrementally injected into $\mathbf{x_0}$ (HS, GT image) until it converges to $\mathbf{x_T}$, yielding intermediate states $\mathbf{x_t}$ that stochastically interpolate between $\mathbf{x_0}$ and $\mathbf{x_T}$ along a Brownian bridge. Consequently, in the learning stage, the denoising neural network learns to denoise these blended states. Crucially, for the Brownian bridge reverse process, the initial input is solely the conditional image (label-free IMS data). These forward and reverse processes of the BBDM framework result in more stable image translation outputs, as detailed in prior work (*38*).

To improve the consistency/repeatability of the VS process and minimize stochasticity in the diffusion process-generated images, we implemented a deterministic noise sampling strategy alongside the standard noise sampling method. Specifically, we used a mean sampling strategy (Fig. 1, E and H) that eliminates, after a certain time point is reached, the posterior noise introduced during the vanilla sampling process (Fig. 1, D and G). Detailed sampling algorithms for this mean sampling strategy are provided in Methods. The performance advantages of this deterministic sampling are further evaluated in the "Repeatable diffusion model inference via noise sampling engineering" section.

Following the training phase, the BBDM-based VS model was tested on human kidney tissue excluded from both the training and validation datasets. Figure 2 showcases the VS results generated by the BBDM-based model using IMS data from different patients. The virtually stained outputs, as presented in Fig. 2B, exhibit a good resemblance to the GT bright-field images, despite being generated from 10-fold larger pixel size IMS images, as illustrated in Fig. 2B. Furthermore, a board-certified pathologist annotated key renal structures—glomeruli (denoted as G), as well as proximal (P) and distal (D) convoluted tubules—on both the VS and HS images. These structures are of clinical importance because they are involved in most renal pathological conditions. As depicted in Fig. 2 (B and C), there was very good concordance in the identification of these structures between the VS and HS GT PAS images. This alignment was consistently demonstrated across multiple fields of view (FOVs) of tissue, underscoring the robustness and generalizability of our framework for the VS of low-resolution IMS data.

To quantitatively evaluate the fidelity of the virtually generated high-resolution PAS images in replicating their HS counterparts, we conducted a comparative analysis on a test dataset comprising 201 distinct tissue FOVs, each with 640 × 640 pixels, from 10 patients. This evaluation focused on three critical aspects: (i) image color

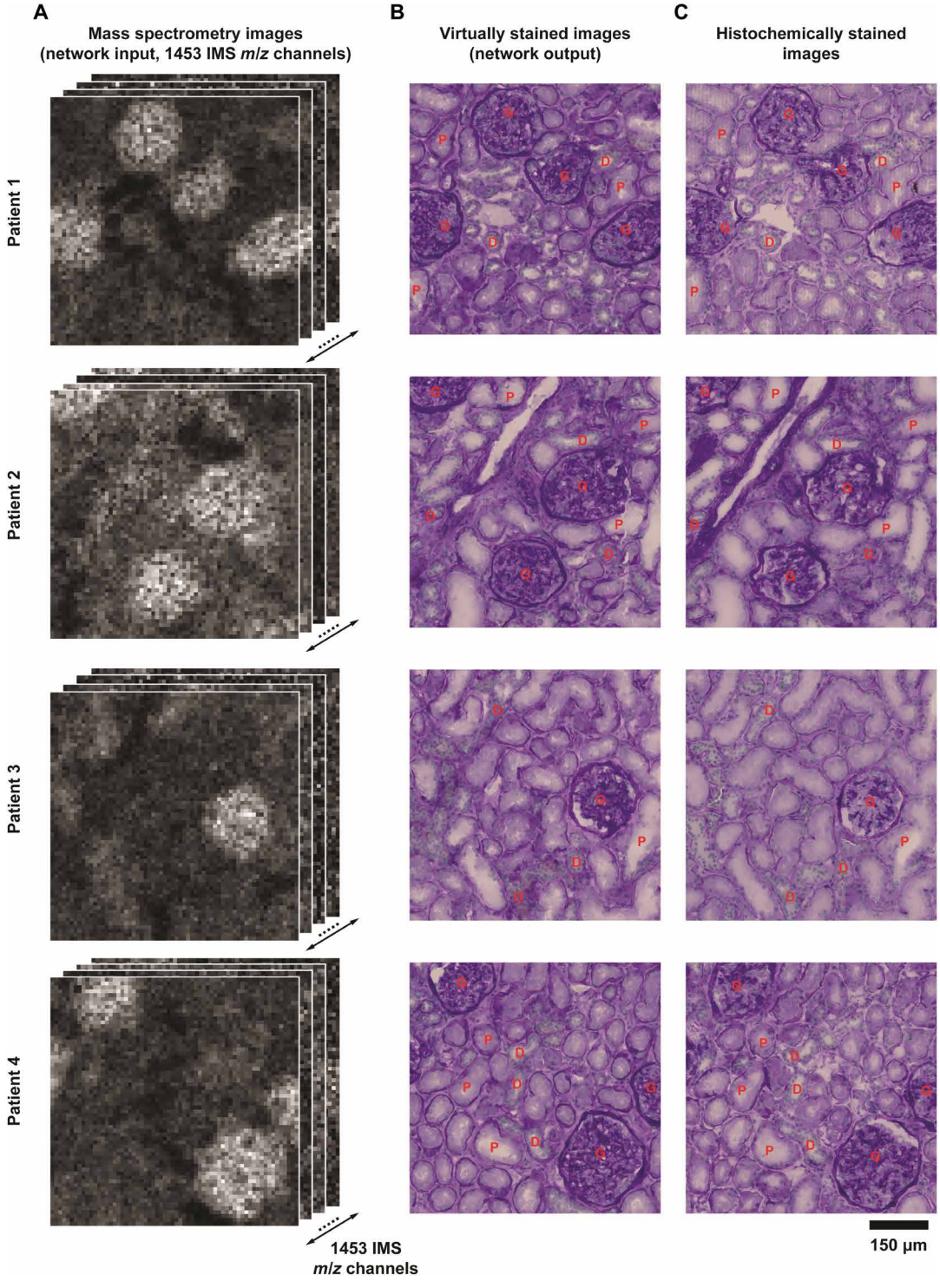

Fig. 2. Visual comparisons between the virtually stained PAS images generated from label-free IMS data of different patients and their HS counterparts. (**A**) Imaging mass spectrometry data of label-free tissue, consisting of 1453 ion (*m*/*z*) channels with a pixel size of 10 μm. (**B**) Virtually stained images digitally generated from the IMS data using our diffusion-based VS model. (**C**) HS GT images. Both the virtually stained and the HS images have a pixel size of 1 μm. The concordant localization of glomeruli (G), proximal convoluted tubules (P), and distal convoluted tubules (D), as annotated by a board-certified pathologist, can be visualized on both the VS and HS images.

distribution; (ii) image contrast (*47*); and (iii) spatial frequency spectrum (detailed in Methods). These quantitative metrics reported in Fig. 3 were selected to evaluate whether the VS images generated from ion images can meaningfully contribute to the interpretation of IMS-measured tissue samples without physically performing the chemical staining and microscopic imaging of stained tissue and, thus, effectively safeguarding the tissue for other assay types and subsequent analysis. While our approach requires multimodal image registration between IMS data and microscopy images during the training phase, which is a one-time effort, this eliminates cumbersome image registration steps during the inference phase. During image inference, the VS images are inherently aligned with their input IMS data, removing the need for additional registration. As depicted in Fig. 3A, the color distribution histograms in the YCbCr color space across all the test FOVs demonstrate strong color agreement between the VS images and the HS bright-field images. This agreement is further supported by the small Hellinger distances (*48*, *49*) calculated between the color histograms of each VS and HS image pair. Moreover, as illustrated in Fig. 3C, the generated VS images achieve image contrast comparable to that of the bright-field HS images.

Furthermore, to showcase the super-resolution capabilities of our VS framework, we conducted a spatial frequency spectrum analysis on the raw single-channel MS images (picked from 1453 m/z channels based on the contrast of glomeruli), network-generated VS images, and their corresponding HS GT images; see Methods for details. This analysis, illustrated in Fig. 3D, includes cross sections of the radially averaged power spectra (*50*, *51*), which demonstrate an excellent match between the spatial frequency spectra of the VS and HS image pairs, as desired. The relative error between the radially averaged power spectra (log scale, normalized) of the VS and the corresponding HS GT images is 6.70 ± 4.84% (means ± SD) across all the frequencies (blue and red curves in Fig. 3D). These results further confirm the effectiveness of the VS output images, which successfully align with the spatial frequency spectra of the high-resolution HS images. Additionally, the relative increase of the radially averaged power spectra (log scale, normalized) of the VS images compared to their lower resolution MS input is 144.23 ± 81.19% (means ± SD) over all the frequencies (blue and green curves in Fig. 3D). These results demonstrate a marked improvement over the spatial frequency spectra of low-resolution MS images, underscoring the diffusion-based VS framework's capacity to enhance spatial resolution significantly.

To further illustrate the effectiveness of the spatial frequency analysis in quantifying the agreement between the virtually stained images generated by our framework and their corresponding paired HS images, we conducted an additional comparison involving two unpaired HS image FOVs with the paired VS/HS images presented in Fig. 3. As demonstrated in fig. S1, there is a clear discrepancy in the radially averaged power spectra between the unpaired HS images and the paired VS/HS images. This result further underscores the effectiveness and specificity of our VS framework in accurately enhancing spatial frequency features to closely match the corresponding HS GT image. Similarly, to validate the significance of the color distribution agreement quantified using the Hellinger distance, we performed an additional comparative analysis involving unpaired HS images and their virtually stained counterparts. Figure S2 presents the calculated Hellinger distances between the color histograms of the paired VS/HS images and those of unpaired VS/HS image pairs across the Y, Cb, and Cr channels. As illustrated in this figure, there is a statistically significant increase in the Hellinger distances for the unpaired VS/HS images compared to the paired VS/HS images, reinforcing the specificity of our VS method in reproducing the color characteristics inherent in histochemical staining.

Reduction analysis on mass spectrometry image channels

The high-resolution image translation capability of VS for IMS data relies on the rich molecular information captured in ion images. To further shed light on this, we trained a series of diffusion-based VS models using different numbers of mass spectrometry (m/z) channels and evaluated their performance on the same test dataset. For the channel selection, we evaluated three distinct strategies. In our primary approach, the IMS channel indices were selected from the top-ranking channels of the sorted list of 1453 channels, prioritized on the basis of signal-to-noise ratio (SNR) values (in descending order). The SNR definition used in this study entails that, for each IMS channel, the mean is divided by the SD across all pixels within that channel. In addition to this SNR-based ranking and channel selection, we explored two alternative channel selection strategies: frequency-based selection and uniform selection. For the frequency-based selection method, we applied a two-dimensional (2D) Fourier transform to the ion image of each channel, ranking the channels by the ratio of the average power in the angular spectrum of the high-frequency components to that of the low-frequency components. We set the half-maximum frequency as the threshold to differentiate between high and low frequencies. For the uniform selection strategy, we sampled the channels at fixed intervals (every 4th, 16th, or 64th channel for reductions of 4-fold, 16-fold, and 64-fold, respectively), starting from the first channel. Using these three distinct selection approaches, we selected channels from the full set of 1453 IMS channels, progressively reducing the total number of channels used per input image from 363 down to 23, corresponding to 4-fold, 16-fold, and 64-fold reductions. These selected IMS channels remained consistent throughout the training and testing of each VS model for a given channel selection strategy.

As shown in Fig. 4A, reducing the number of MS channels from 1453 to 23 resulted in a gradual degradation of the VS performance for all three channel selection strategies, with noticeable losses in critical features such as nuclear morphology. The relationship between the number of IMS channels and VS quality was further quantified using the peak SNR (PSNR) and learned perceptual image patch similarity (LPIPS) (*52*) and Pearson correlation coefficient (PCC). Figure 4B presents the PSNR, LPIPS, and PCCs metrics for the 10 VS models, each trained with a combination of a different number of IMS channels and channel selection approaches, evaluated across a test dataset comprising 201 distinct tissue FOVs from 10 patients, each with 640 × 640 pixels. These results confirmed that as the number of used MS channels increased, the diffusion-based VS models achieved a statistically significantly higher VS fidelity, underscoring the rich molecular information present in MS data and its strong utility for label-free histological staining. Among the three distinct channel selection strategies, the SNR-based selection consistently delivered the best performance compared to the other two channel sampling approaches, further validating its effectiveness in preserving the critical molecular information for high-quality VS with a reduced number of IMS channels.

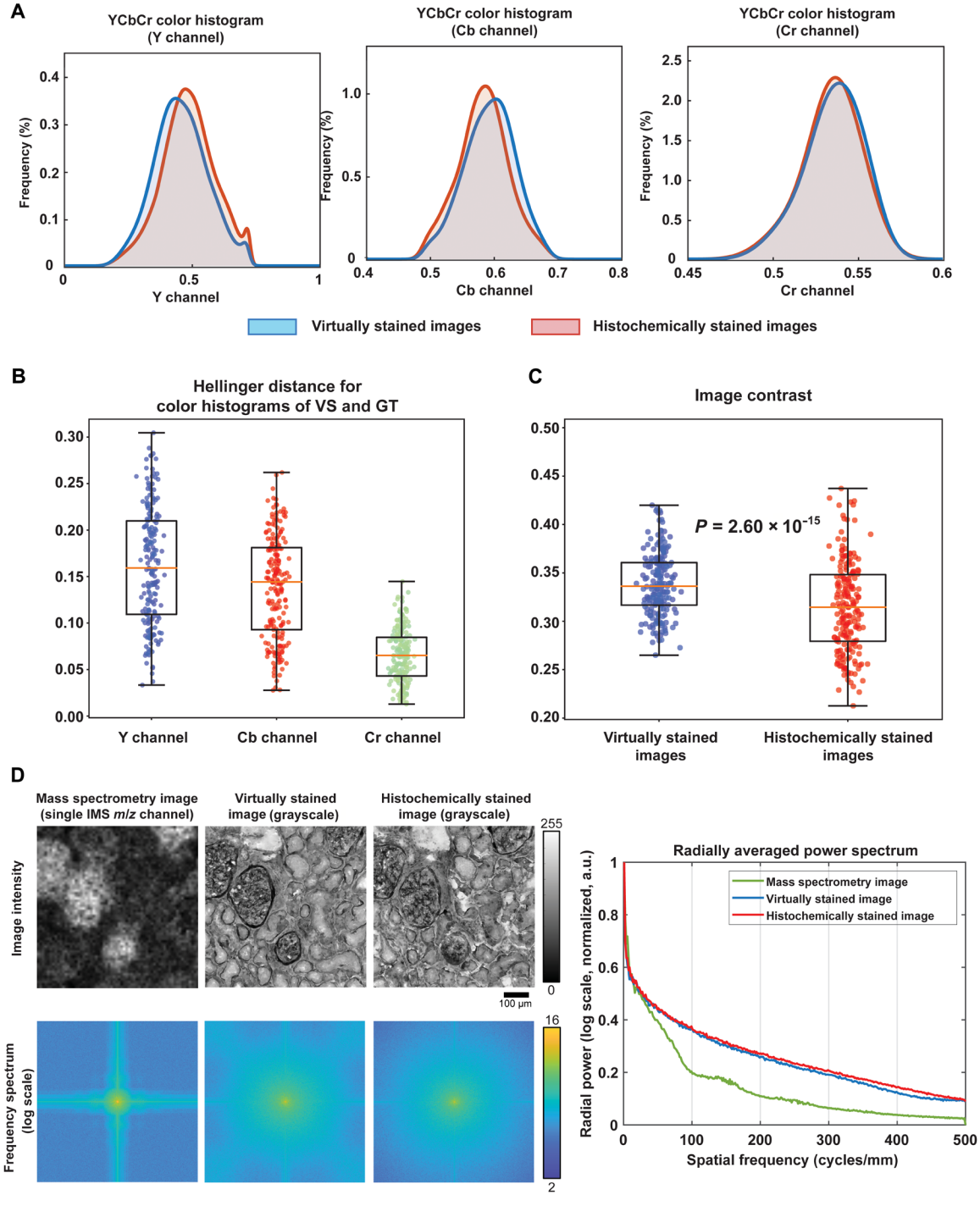

Fig. 3. Quantitative comparisons between virtually stained PAS images generated from label-free IMS data and their HS counterparts. (**A**) The color histogram comparisons of virtually stained PAS images and their HS counterparts in Y, Cb, and Cr channels, calculated separately. (**B**) The Hellinger distances between the color histograms of VS and HS image pairs in Y, Cb, and Cr channels. (**C**) The box plots of image contrast of virtually stained PAS images and their HS counterparts across a test dataset comprising 201 distinct FOVs. A statistically significant improvement of the image contrast values of the VS images over the HS images was determined by a two-tailed paired *t* test ($P = 2.60 \times 10^{-15}$). (**D**) The low-resolution single MS channel image, the corresponding high-resolution grayscale virtually and HS images (top left), and their respective spatial frequency spectra (amplitude, displayed in bottom left). The radially averaged normalized power spectrum [log scale, arbitrary unit (a.u.)] cross section for each case is also presented (right).

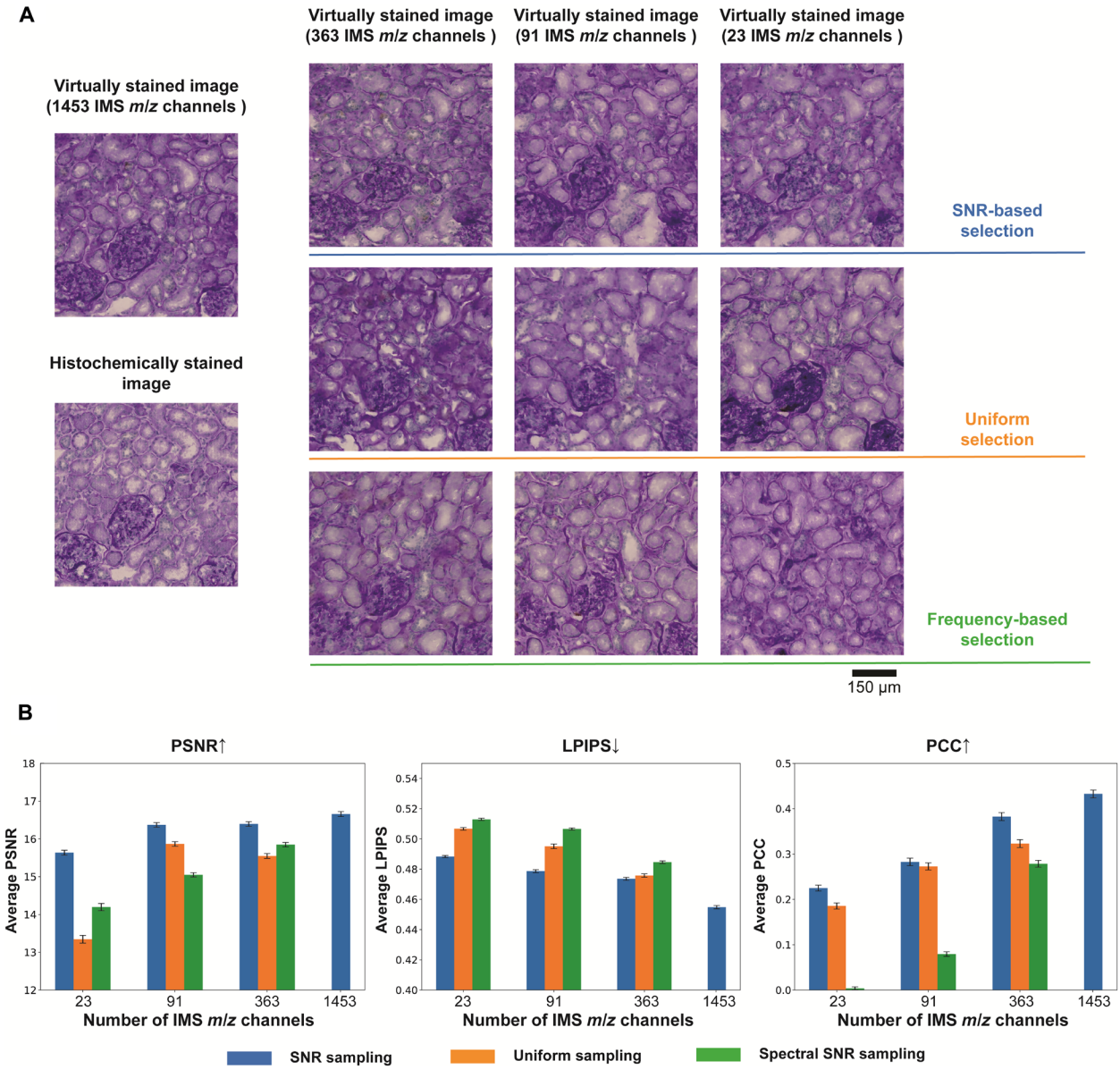

Fig. 4. Comparison of VS performance for diffusion models trained with different numbers of IMS channels and channel selection strategies. (**A**) Visual comparisons of virtually stained PAS images generated from diffusion-based VS models trained using all the IMS channels and three reduced channel sets obtained through three different selection strategies (resulting in nine configurations in total). (**B**) Bar plots displaying the quantitative PSNR, LPIPS, and PCC metrics averaged across testing virtually stained images for all the configurations. Higher PSNR, PCC, and lower LPIPS values are desired.

Repeatable diffusion model inference via noise sampling engineering

Tissue heterogeneity and inherent histochemical staining variability often lead to minor differences between adjacent tissue sections. These spatial changes usually do not influence the overall slide-level diagnosis and are well tolerated by human pathologists in their clinical workflow. However, computational models applied for tissue analysis are frequently misled by these slide-to-slide variations, resulting in lower performance. For our VS model, the primary source of variation in the output VS results comes from the stochastic nature of the noise sampling process during the backward diffusion. To address this, we applied noise sampling process engineering techniques to improve VS consistency and reduce output variance without the need for fine-tuning or transfer learning on the trained model. In addition to the mean sampling strategy illustrated in Fig. 1, we also introduced an alternative "skip sampling" strategy for comparison, which directly estimates $\mathbf{x_0}$ from the denoising network's output.

The rationale behind the mean and skip sampling strategies arises from the significant increase in the additional variance $\tilde{\delta}_t$ during the final stages of the reverse diffusion process (see fig. S3). Both of these sampling strategies, mean and skip, avoid this increasing noise variance in the reverse path after an engineered exit point t_e (see Methods); this effectively reduces stochastic variations (observed

from run to run) in the output of the diffusion model for the same label-free tissue FOV, which is highly desired for VS applications. Detailed descriptions of these engineered noise sampling techniques are provided in Methods and figs. S3 to S5.

To quantitatively assess the effectiveness of these noise sampling strategies, we tested the trained BBDM using the vanilla, mean, and skip sampling strategies, repeating each method five times and calculating the pixel-wise coefficient of variance (CV) across these repetitions of the diffusion-based VS process. Figure 5A shows the CV maps for the three distinct strategies in the YCbCr color channels, demonstrating that both the mean and skip sampling strategies effectively reduce the variance in the sampled VS images compared to the vanilla method. Additionally, we computed the average CV values across all the pixels in the test image FOVs and plotted them

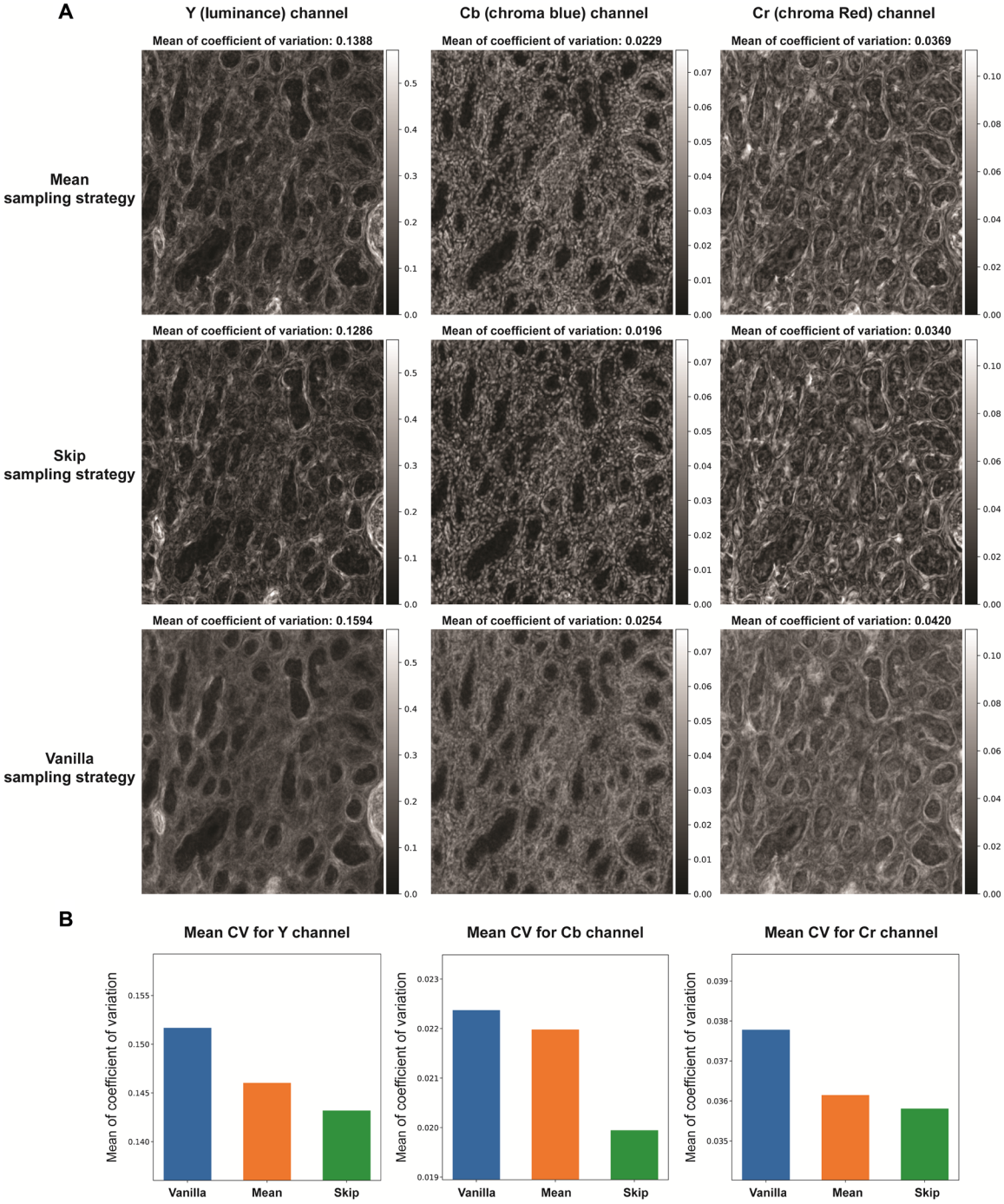

Fig. 5. Comparison of the CV for different noise sampling engineering approaches. (**A**) Visualization of the coefficient of variation (CV) maps for the YCbCr channels of virtually stained images obtained using three different approaches: vanilla, mean, and skip sampling approaches. The CV was calculated using five repetitions of VS inference on the same FOV. (**B**) Plots of the mean CV calculated across all pixels of all test image FOVs for the three sampling approaches.

for each YCbCr channel, as shown in Fig. 5B. Our results further corroborate that the mean and skip sampling strategies are effective in achieving lower output variances, indicating the repeatability of the diffusion-based VS process using these engineered noise sampling techniques.

This comparative analysis in Fig. 5 further reveals that the skip sampling strategy yields a lower average CV value compared to the mean sampling method. However, the mean sampling strategy produces results that exhibit better perceptual similarity to the GT HS PAS images. Additional visual and quantitative comparisons between different noise sampling strategies in diffusion-based VS models are presented in fig. S4. These results confirm that the mean sampling strategy is superior to both the vanilla and skip sampling strategies, achieving a lower average LPIPS as desired. Additionally, we evaluated the performance of an averaging strategy, which involves averaging independent test runs of different inferences for the same tissue FOV. While fig. S4C highlights the advantage of the averaging strategy in reducing pixel-level CV, the resulting images exhibit relatively lower contrast with a significantly worse LPIPS score, falling short of the performance achieved by the mean sampling strategy alone. Therefore, this averaging strategy would only be preferable in scenarios where high consistency is prioritized over image contrast.

The noise sampling strategies demonstrated here can be further optimized. For instance, we evaluated the performance of the mean and skip sampling strategies across eight different exit points (t_e), ranging from 0 (equivalent to the vanilla sampling strategy) to 100, as illustrated in fig. S5. Our findings indicate that both the mean and skip sampling strategies consistently outperformed the vanilla strategy across a range of exit points. Specifically, both strategies achieved optimal LPIPS performance at an exit time point $t_e \sim 10$. Consequently, selecting a well-suited evaluation metric and determining the appropriate exit point are critical for optimizing the performance of diffusion-based VS models, particularly for clinical applications.

DISCUSSION

The effectiveness of the presented label-free VS of IMS-measured ion images, despite the 10-fold larger pixel size of the input IMS data, can be attributed to the capabilities of diffusion models in effectively capturing and modeling complex data distributions (*42*, *43*, *53*). Historically, generative adversarial network (GAN)–based approaches (*54*–*56*) have been a predominant choice in image restoration for biomedical applications, particularly in super-resolution image reconstruction tasks (*51*, *57*, *58*). However, recent advancements (*59*, *60*) in the field have revealed that GANs might struggle with highly challenging image reconstruction problems, especially at extreme super-resolution factors. In contrast, diffusion models have emerged as a superior alternative, consistently producing more realistic and accurate spatial features even at these high super-resolution factors. Moreover, the complexity of multiplexed inference tasks that require simultaneous resolution enhancement and cross-domain image translation has further underscored the limitations of GAN-based techniques. It has been demonstrated that diffusion models outperform GANs in such multiplexed tasks (*30*, *61*), owing to their robustness in generating high-fidelity images. Additionally, the issue of mode collapse (*62*), which is a well-documented limitation of GANs when confronted with sparse or low-quality data, is well mitigated in diffusion models. Diffusion models, in general, exhibit more stable training dynamics, even when faced with significant discrepancies between the input data and GT images, making them more resilient in handling challenging datasets (*63*). For the same super-resolution VS task presented in this work, we also evaluated the cGAN framework (*29*); however, the model collapsed at an early stage and failed to converge effectively, resulting in images containing unacceptable staining artifacts. Taking these factors into account, we can conclude that the success of our VS models in generating high-resolution virtual stains from low-resolution IMS data of label-free tissue samples is a testament to the diffusion models' inherent strengths.

The repeatability and consistency of the IMS-generated virtually stained images are crucial for digital pathology interpretation and were achieved in this work using the mean and skip sampling strategies used during the reverse process of the diffusion model. Another notable achievement of this study is its ability to enable medical experts to directly identify diagnostically relevant renal pathology structures—such as glomeruli, proximal, and distal convoluted tubules—using virtually stained images generated from lower-resolution IMS images. Historically, recognizing these structures directly from IMS images was not feasible. Conventional histology-directed IMS analysis requires both IMS data and bright-field microscopy images of the HS tissue, followed by a complex and time-consuming registration process to link the IMS data with pathology-annotated regions in optical microscopy images of stained tissue. Furthermore, histochemical staining of IMS slides prevents their utilization for genomics/epigenetics analyses and hinders IMS-molecular comparison studies. Our diffusion-based VS framework, however, enables us to bypass these limitations by offering a means of histological interpretation of IMS data for regions of interest within the kidney tissue, immediately after an IMS scan, without the need for histochemical staining. Moreover, once the diffusion model is trained, this technique can be seamlessly integrated into the post-IMS data processing pipeline without requiring any modifications to the existing hardware/setup.

It is also important to note that the quality of IMS-based VS can be further enhanced. In this study, we demonstrated the virtual histological staining of ion images with a pixel size of ~10 μm, as this is a common spatial resolution for MALDI IMS systems. However, with recent advances in IMS technology, achieving higher spatial resolution is now possible (*8*–*10*). For instance, transition mode ion sources allow IMS to achieve 1-μm pixel sizes. Unfortunately, these specialized workflows often inflict significant tissue damage, as the tissue is ablated during the sampling process, making the post-IMS multimodal analysis impossible. By applying our VS framework to these high-resolution IMS images, we not only overcome the post-IMS multimodal analysis challenge of tissue damage but also enhance the fidelity of the virtually stained images, further elevating the precision and quality of the diffusion-model results. Furthermore, although we demonstrated the efficacy of our technique using PAS staining on human kidney samples, this label-free approach can be extended to other types of histochemical stains and various organs. This adaptability is supported by the prior success of various VS techniques (*19*), making it versatile across different staining protocols and tissue types.

We believe that this advancement has the potential to drive further biomedical research, particularly in the study of glomerular and tubular diseases through IMS. Furthermore, this method can augment existing IMS datasets by generating virtually stained images, offering advantages to the broader biomedical research community.

In addition, our reduction analyses on IMS channels further confirm the rich information encoded in IMS data and can potentially reveal the relationship of IMS channels with histochemical staining, which might prospectively facilitate a better understanding of both the virtual and histochemical staining processes. We demonstrated that our VS method remains effective even when using a reduced number of IMS channels, as selected via a SNR-based reduction strategy. This provides several practical advantages. First, computational load and hardware requirements for both model training and inference scale with the number of input channels, thus reducing the channel count enhances the method's accessibility and efficiency. Second, our results suggest that this SNR-based feature reduction makes the method more compatible with high–spatial resolution IMS experiments, such as those performed using MALDI IMS. One of the key limitations to achieving high spatial resolution in MALDI IMS is the trade-off between spatial resolution and molecular sensitivity. As the laser beam diameter is reduced (e.g., via decreased fluence, improved focus, or oversampling), the sampled tissue area—and hence the amount of ablated material—also decreases. This results in lower signal intensities and the potential loss of low-SNR molecules, as they may fall below the detection threshold. In such scenarios, only high-intensity, high-SNR molecules remain detectable. Our findings suggest that, even under these conditions, the VS method could remain viable and effective, further underscoring its applicability.

Study limitations

In our study, we identified a few mismatched areas where our diffusion-based model did not accurately reconstruct uncommon structures, such as an occasional large collecting duct, as illustrated in fig. S7 (A and B). This limitation arises because our training and testing datasets were composed of kidney samples obtained from uninvolved cortical areas during total nephrectomy procedures, which typically lack hypocellular regions. Consequently, the model was not effectively trained to reconstruct such sparse areas, leading to hallucinations in such regions. Expanding the training dataset to include diverse tissue morphologies with pathological changes in future studies could alleviate this issue. Additionally, fig. S7 (C and D) highlights another representative FOV, where both virtually stained and HS images exhibit staining artifacts, particularly within the glomerular regions. These artifacts can be attributed to tissue damage that occurred during tissue processing and the IMS imaging pipeline.

As shown in Fig. 2, the virtually stained images sometimes exhibit minor variations with respect to the HS GT images at the cellular level. These differences primarily arise from the inherently lower spatial resolution of the input IMS data, which have an ~10-fold larger pixel size relative to bright-field microscopy, GT images used in this work. However, such discrepancies generally do not affect diagnostic outcomes at the structure level, as pathologists typically examine multiple serial sections, each of which naturally exhibits minor variations in cellular or tissue morphology due to tissue heterogeneity. Consequently, even with slight misalignments in the virtually stained images compared to the histochemical stains, pathologists and researchers can potentially recognize the tissue features and make accurate diagnoses at the structure level.

However, we also acknowledge that our current VS approach may face limitations in clinical scenarios where precise morphological details, such as the exact shape and geometry of glomeruli, play a critical role in diagnosis (*64*). Additionally, certain single-cell features, particularly rare but diagnostically significant, such as immune cells within the glomeruli, could potentially be obscured or misrepresented by the VS model, due to the necessity of inferring missing spatial information from lower-resolution molecular inputs. To effectively address these limitations in diseases requiring meticulous cellular and geometric detail, using higher-resolution IMS systems with pixel sizes as small as ~1 μm would likely yield substantially improved VS results, as previously discussed in our manuscript.

It is also worth noting that our approach is not intended for primary diagnosis in clinical settings. First, IMS is not a US Food and Drug Administration (FDA)–approved imaging modality for digital pathology. In current digital pathology workflow, the standard FDA-approved whole-slide scanners typically offer much faster imaging speeds and higher spatial resolution compared to IMS systems. Consequently, our IMS-based VS approach is not practical for widespread clinical diagnostic applications in digital pathology workflow but remains a valuable tool for IMS-focused biomedical research. We believe that our IMS-based label-free VS framework will become an essential tool for IMS-driven molecular pathology research, effectively bridging the gap between fundamental IMS investigations and potential clinical diagnostic applications.

METHODS

Sample preparation

Human kidney tissue was surgically removed during a full nephrectomy, and remnant tissue was processed for research purposes by the Cooperative Human Tissue Network at Vanderbilt University Medical Center. Remnant biospecimens were collected in compliance with the Cooperative Human Tissue Network standard protocols and the National Cancer Institute's Best Practices for the procurement of remnant surgical research material. Participants were consented for remnant tissue collection in accordance with Institutional Review Board (IRB) policies. The study received ethical approval from Vanderbilt University's IRB (no. 210190). The study involved kidney specimens from five individual patients. Kidney tissue was flash frozen over an isopentane-dry ice slurry, embedded in carboxymethylcellulose, and stored at −80°C. The tissue was cryosectioned into 10-μm-thick sections using a CM3050 S cryostat (Leica Biosystems, Wetzlar, Germany). The sections were then thaw mounted onto indium tin oxide–coated glass slides (Delta Technologies, Loveland, CO) for IMS analysis or regular glass slides for histological staining. Slides were stored at −80°C and returned to ~20°C within a vacuum desiccator before further processing. To remove endogenous salt for IMS, the section was washed three times with chilled (4°C) 150 mM ammonium formate three times for 45 s each. It was then dried with nitrogen gas to remove excess moisture.

AF microscopy and histochemical staining

To enable coregistration of IMS data with HS images, we introduced AF images as an intermediate modality, to which both IMS and HS images can be registered, as shown in fig. S6A (also see details in the "Multimodal image registration" section). AF microscopy images were acquired on each tissue before IMS analysis using 4′,6-diamidino-2-phenylindole, enhanced green fluorescent protein, and DsRed filters on a Zeiss AxioScan.Z1 slide scanner (Carl Zeiss Microscopy GmbH, Oberkochen, Germany). Additionally, AF

was also collected after IMS data acquisition before matrix removal, enabling visualization of ablation marks created by the MALDI laser (*16*). The resulting images have a pixel size of 0.65 μm.

It is worth noting that acquiring AF microscopy images after the IMS data acquisition but before the chemical matrix is removed offers a clear visualization of the ablation craters for each IMS pixel in the microscopy coordinate space, thereby facilitating precise IMS-microscopy registration. During the MALDI process, the tissue surface, including its chemical matrix layer, is irradiated with a laser that causes the ablation and desorption of material and the ionization of endogenous molecules for mass spectrometry detection. Because the laser is focused on a diameter slightly smaller than that of an IMS pixel to avoid oversampling, this results in an array of distinct ablation marks on the matrix-coated tissue surface. In most experiments, only the chemical matrix layer is ablated during MALDI, while the underlying tissue remains intact.

After the acquisition of the post-IMS AF images, the same unlabeled tissue was stained using a standard PAS staining protocol (*65*). The PAS staining process takes 30 to 60 min for each batch of processed slides. The stained tissue slides were scanned and digitized using a bright-field slide scanner (Leica Biosystems Aperio AT2). The resulting PAS-stained bright-field microscopy images have a pixel size of 0.22 μm.

Imaging mass spectrometry

Samples for IMS analysis were coated with solution (20 mg/ml) of DAN dissolved in tetrahydrofuran using a TM Sprayer M3 (HTX Technologies LLC, Chapel Hill, NC, USA), yielding a coating of 1.67 mg/cm^2 (0.05 ml/hour, four passes, 40°C spray nozzle). MALDI IMS was performed on a prototype timsTOF fleX mass spectrometer (Bruker Daltonik, Bremen, Germany). The ion images were collected in negative ionization mode at 10-μm pixel size with a beam scan set to 6 μm, using 150 laser shots per pixel and 18.6% laser power (30% global attenuator and 62% local laser power) at 10 Hz. Data were acquired in negative ionization quadrupole Time-of-Flight (qTOF) mode, covering an m/z range from 150 to 2000. The acquisition speed of IMS is ~5 hours per slide.

Multimodal image registration

The multimodal registration pipeline is illustrated in fig. S6B. Due to the 10-fold difference in the pixel size and inherent differences in the imaging modalities, direct registration of the HS images with IMS data is not feasible. To address this challenge, we introduced prior-IMS and post-IMS AF images to facilitate the registration process. The pipeline consists of three sequential image registration steps: (i) the first step aligns the HS images with the prior-IMS AF images; (ii) next, the prior-IMS AF images are registered to the post-IMS AF images; and (iii) lastly, the IMS data are registered to the post-IMS AF images. Notably, the third step leverages the laser ablation craters left on the chemical matrix after the IMS acquisition to guide image registration. However, this ablation process causes localized feature loss in the post-IMS AF images, making them unsuitable for direct alignment with the HS images. Therefore, the prior-IMS AF images serve as an intermediary, as they share the same modality as the post-IMS AF images and retain structural features needed for robust image registration. The microscopy modalities were coregistered using the elastix (*66*) framework integrated into wsireg (*67*) software. The post-IMS AF was selected as the target modality to enable integration with IMS, with the PAS-stained bright-field microscopy images becoming the source modalities. The rigid and affine transformations were used because the microscopy modalities were collected on the same tissue section. All registered whole-slide images (WSIs) were stored in the vendor-neutral pyramidal Open Microscopy Environment Tagged Image File Format (OME-TIFF) at a common pixel size (i.e., the PAS image was resampled to the resolution of the prior-IMS AF and post-IMS AF). Furthermore, the MALDI IMS datasets were manually registered to the post-IMS AF images using the laser ablation marks and IMS pixels. The manual registration was performed using IMS Microlink software (*68*), where 8 to 12 fiducial markers were selected in both modalities to estimate the affine transformation.

Data division and preparation

The collected IMS data were exported from the Bruker timsTOF file format (.d) to a custom binary format for ease of access and improved performance. Each pixel/frame contains between 10^4 and 10^5 centroid peaks covering the entire acquisition range, which can be reconstructed into a pseudoprofile mass spectrum using Bruker's SDK (v2.21). The dataset was m/z-aligned using six internally identified peaks (appearing in at least 50% of the pixels) through the msalign library (v0.2.0). This step corrects spectral misalignment (drift along the m/z axis), resulting in increased overlap between spectral features (peaks) across the experiment. Subsequently, the mass axis of the dataset was calibrated using the theoretical masses of the six peaks, achieving a precision of ~±1 ppm. Following the preprocessing steps, normalization correction factors were computed, and a total ion current approach was used for mass spectral and ion image normalization. Subsequently, an average mass spectrum based on all pixels was calculated for each dataset. Because samples from multiple donors were used in this study, an average mass spectrum of all samples was generated and peak-picked, identifying 1453 that were used for further analysis. The resulting average spectrum had a resolving power of ~40,000 at m/z 885.55.

To better match the dimensions of the IMS data and facilitate the transformation from IMS to HS images, we downsampled the HS images to achieve a pixel size of 1 μm, which is 10 times smaller than the pixel size of IMS. The WSIs of IMS and HS images, acquired from four patients, were then segmented into smaller FOV pairs of ~1400 μm by 1400 μm (each IMS image with 140 × 140 pixels and each HS image with 1400 × 1400 pixels), with 10% overlap between neighboring regions. To augment the data for robust model training, the paired WSIs were spatially transformed using a combination of rotation and flipping. The transformed WSIs were segmented as described above. This process generated a training dataset containing 712 paired IMS-PAS microscopic image patches obtained from four deidentified patients. Additionally, 201 paired IMS-PAS microscopic image patches without data augmentation (each FOV with 64 × 64 pixels for IMS and 640 × 640 pixels for the HS image) were reserved for blind testing, obtained from 10 deidentified patient not included in the training set. During each training epoch, the paired image FOVs were further subdivided.

Brownian bridge diffusion process

The processed dataset provided input-output image pairs, where $\mathbf{x_0} \in \mathbb{R}^{H\times W\times 3}$ corresponds to HS image, and $\mathbf{y_0} \in \mathbb{R}^{\frac{H}{10}\times\frac{W}{10}\times 1453}$ represents the label-free IMS-measured ion images of the same tissue FOV. Before training the diffusion model, the IMS dataset $\mathbf{y_0}$ was preprocessed by a shallow neural network g_s, as shown in Fig. 1B.

This shallow network consisted of two convolutional layers followed by a pixel-shuffle (*69*) layer, which upscaled the image dimensions to match the target HS image, i.e., the corresponding GT image. Specifically, the processed output $\mathbf{y} \in \mathbb{R}^{H\times W\times 3}$ has the same dimension as the HS target image. This shallow network can be denoted as

$$\mathbf{y} = g_s(\mathbf{y_0}) \tag{1}$$

Subsequently, the processed label-free image $\mathbf{y}$ and the corresponding HS image $\mathbf{x_0}$ were used as the starting and ending points of the Brownian bridge, respectively, thereby forming an image-conditional diffusion process. In this Brownian bridge diffusion process, $\mathbf{y}$ acts as the image condition input, while $\mathbf{x_0}$ represents the target image. The forward diffusion process of the Brownian bridge with an initial state $\mathbf{x_0}$ and terminal state y is defined as

$$q(\mathbf{x_t}\,|\,\mathbf{x_0}, \mathbf{x_T}=\mathbf{y}) = \mathcal{N}\left[(1-a_t)\mathbf{x_0} + a_t\mathbf{y}, \delta_t\mathbf{I}\right] \tag{2}$$

where a_t and the variance δ_t can be calculated as

$$a_t = \frac{t}{T} \quad \delta_t = \frac{2st(T-t)}{T^2} \tag{3}$$

Here, t refers to the intermediate time step and T is the total sampling steps. T was set to 1000 during both the forward and reverse processes. Factor s here is used to control the sampling diversity and is set to 1 in the forward process. During the forward sampling process, the intermediate state $\mathbf{x_t}$ can be derived from Eq. 2 as

$$\mathbf{x_t} = \mathbf{x_0} + a_t(\mathbf{y}-\mathbf{x_0}) + \sqrt{\delta_t}\epsilon_t, \epsilon_t \sim \mathcal{N}(0, I) \tag{4}$$

As shown in Eq. 4, the target image $\mathbf{x_0}$ can be recovered from any intermediate state $\mathbf{x_t}$ by removing the add-on term $a_t(\mathbf{y}-\mathbf{x_0}) + \sqrt{\delta_t}\epsilon_t$. Therefore, a denoising network ϵ_θ was also trained as part of the BBDM to estimate this add-on term to recover the target image $\mathbf{x_0}$. Because the term $a_t(\mathbf{y}-\mathbf{x_0}) + \sqrt{\delta_t}\epsilon_t$ is a function of time step t, the denoising network takes both the noisy image $\mathbf{x_t}$ and the time step t as input. This network is built upon a U-Net (*40*) architecture, as shown in Fig. 6B. The U-Net structure features two connected parts: a downsampling path and an upsampling path, with a central layer and skip connections linking them. The downsampling path comprises four levels, each consisting of a residual block followed by an attention block, and adjacent levels are connected via 2 × 2 average pooling. In parallel, the upsampling path mirrors this structure but with 2 × 2 nearest-neighbor interpolation connecting the levels. At the core of the U-Net, the middle block incorporates two residual blocks and one attention block, using multihead attention (*70*) for enhanced feature processing. Additionally, the time step t is embedded through a linear layer with a SiLU

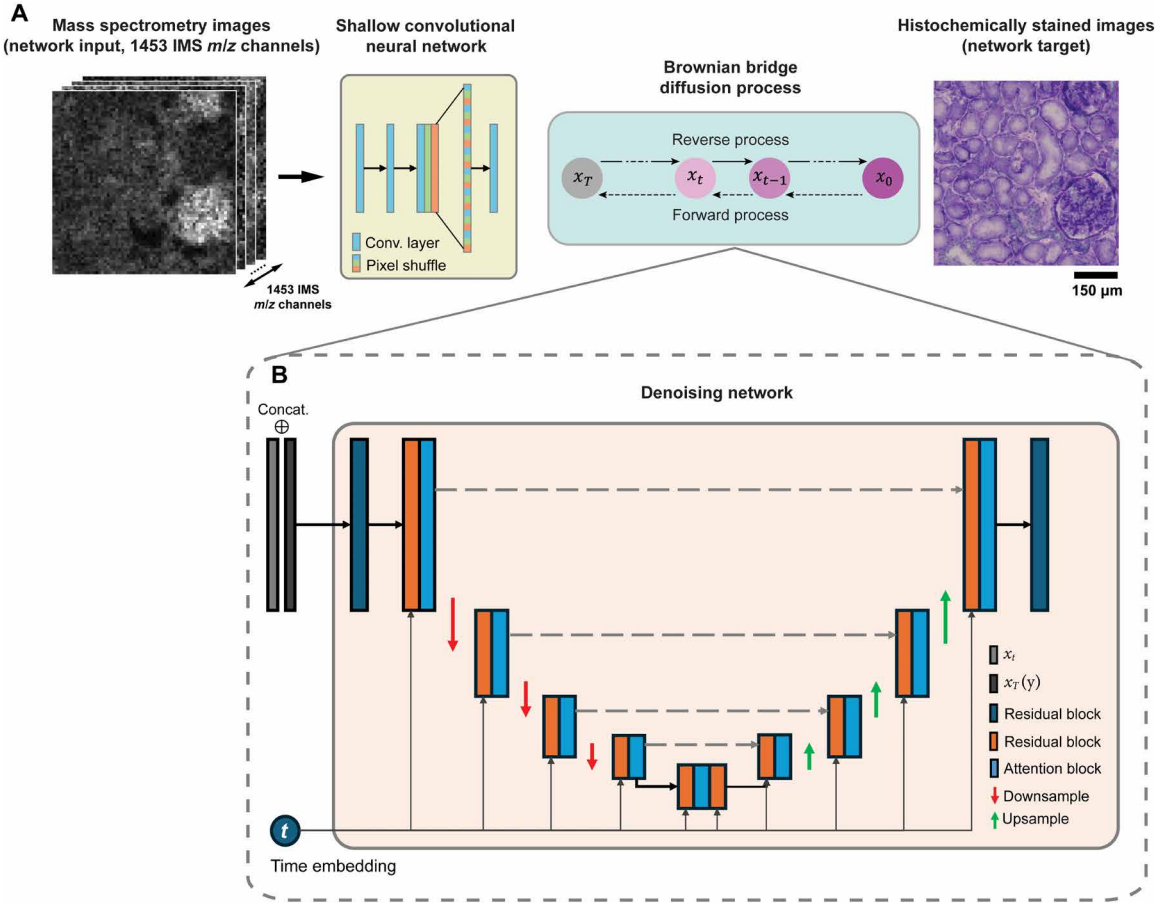

Fig. 6. The network architecture for the diffusion-based VS model. (**A**) The pipeline of the forward and reverse sampling processes. The detailed architecture of the shallow convolutional neural network used for dimension reduction is also presented. (**B**) Detailed architecture of the denoising network used at each step of both the forward and reverse sampling processes.

(*71*, *72*) activation function and integrated into the residual blocks as an added feature.

The loss function of the denoising network ϵ_θ is defined as

$$L = \Sigma_t \gamma_t E_{(x_0,y),\epsilon_t} \| a_t(y - x_0) + \sqrt{\delta_t}\epsilon_t - \epsilon_\theta(x_t, t) \|_2^2 \quad (5)$$

where γ_t is the weight for each t. During the training, γ_t was empirically set as 1 for all the sampling steps, t.

During the blind testing, the reverse sampling process of BBDM starts from the conditional label-free input, $\mathbf{x_T} = \mathbf{y} = g_s(\mathbf{y_0})$. Based on the main idea of the denoising diffusion method (*42*), the reverse process aims to estimate $\mathbf{x_{t-1}}$ from $\mathbf{x_t}$. Hence, the vanilla reverse process can be defined as

$$p(\mathbf{x_{t-1}} \mid \mathbf{x_t}, \mathbf{y}) = \mathcal{N}\left[\mathbf{x_{t-1}}; \boldsymbol{\mu}'_\mathbf{t}(\mathbf{x_t}, \mathbf{y}), \tilde{\delta}_t \mathbf{I}\right] \quad (6)$$

$$\boldsymbol{\mu}'_\mathbf{t}(\mathbf{x_t}, \mathbf{y}) = c_{xt}\mathbf{x_t} + c_{yt}\mathbf{y} - c_{\epsilon t}\epsilon_\theta(\mathbf{x_t}, t) \quad (7)$$

$$\widetilde{\delta}_t = \frac{\delta_{t|t-1}\delta_{t-1}}{\delta_t} \quad (8)$$

where $\boldsymbol{\mu}'_\mathbf{t}(\mathbf{x_t}, \mathbf{y})$ is the estimated mean relying on the trained denoising network, while the $\widetilde{\delta}_t$ is the additional variance added during the sampling process. The expressions of $c_{xt}, c_{yt}, c_{\epsilon t}$, and $\delta_{t|t-1}$ can be formulated as

$$c_{xt} = \frac{\delta_{t-1}}{\delta_t}\frac{(1-a_t)}{(1-a_{t-1})} + \frac{\delta_{t|t-1}}{\delta_t}(1-a_{t-1}) \quad (9)$$

$$c_{yt} = a_{t-1} - a_t\frac{(1-a_t)}{(1-a_{t-1})}\frac{\delta_{t-1}}{\delta_t} \quad (10)$$

$$c_{\epsilon t} = (1-a_{t-1})\frac{\delta_{t|t-1}}{\delta_t} \quad (11)$$

$$\delta_{t|t-1} = \delta_t - \delta_{t-1}\frac{(1-a_t)^2}{(1-a_{t-1})^2} \quad (12)$$

During the blind testing phase, the reverse process of the diffusion model predicts the target image x_0 starting from $\mathbf{x_T} = \mathbf{y}$. The vanilla sampling strategy operates as a Markov chain with learned Gaussian transitions, initiating from $p(\mathbf{x_T}) = \mathcal{N}(\mathbf{x_T}; \mathbf{y}, 0)$. This process can be expressed as

$$p_\theta^{\mathrm{v}} = p(\mathbf{x_T})\prod_{t=1}^{T} p(\mathbf{x_{t-1}} \mid \mathbf{x_t}, \mathbf{y}) \quad (13)$$

Each transition probability $p(\mathbf{x_{t-1}} \mid \mathbf{x_t}, \mathbf{y})$ estimates the intermediate state $\mathbf{x_{t-1}}$ from the previous state $\mathbf{x_t}$ and can be formulated as

$$\mathbf{x_{t-1}} = c_{xt}\mathbf{x_t} + c_{yt}\mathbf{y} - c_{\epsilon t}\epsilon_\theta(\mathbf{x_t}, t) + \sqrt{\tilde{\delta}_t}\mathbf{z},\ \mathbf{z} \sim \mathcal{N}(0, \mathbf{I}) \quad (14)$$

By iterating Eq. 14, each intermediate state is sequentially estimated, starting from $\mathbf{x_T}$ and continuing until the final target image $\mathbf{x_0}$ is obtained, revealing the virtually stained PAS image of the label-free test tissue. This iterative process is referred to as the vanilla sampling strategy and is depicted in Fig. 1 (D and G).

For the mean sampling strategy, shown in Fig. 1 (E and H), the procedure initially mirrors the vanilla sampling approach up until the time step $t_e < T$, where t_e is defined as the exit point in the reverse sampling process. Once the process reaches $t \le t_e$, this strategy estimates $\mathbf{x_{t-1}}$ from $\mathbf{x_t}$ using a mean sampling step as follows

$$p_m(\mathbf{x_{t-1}} \mid \mathbf{x_t}, \mathbf{y}) = \mathcal{N}\left[\mathbf{x_{t-1}}; \boldsymbol{\mu}'_\mathbf{t}(\mathbf{x_t}, \mathbf{y}), 0\right] \quad (15)$$

where the estimation of $\mathbf{x_{t-1}}$ relies solely on $\boldsymbol{\mu}'_\mathbf{t}(\mathbf{x_t}, \mathbf{y})$, without incorporating additional variance $\widetilde{\delta}_t$, i.e., $\mathbf{x_{t-1}} = c_{xt}\mathbf{x_t} + c_{yt}\mathbf{y} - c_{\epsilon t}\epsilon_\theta(\mathbf{x_t}, t)$ for $t \le t_e$. Thus, the mean sampling strategy can be formulated as

$$p_\theta^{\mathrm{m}} = p(\mathbf{x_T})\prod_{t=1}^{t_e} p_m(\mathbf{x_{t-1}} \mid \mathbf{x_t}, \mathbf{y})\prod_{t=t_e}^{T} p(\mathbf{x_{t-1}} \mid \mathbf{x_t}, \mathbf{y}) \quad (16)$$

The skip sampling strategy follows the same vanilla sampling steps up to an exit point $t = t_e$, i.e.

$$p_\theta^{\mathrm{s}} = p(\mathbf{x_T})\prod_{t=t_e}^{T} p(\mathbf{x_{t-1}} \mid \mathbf{x_t}, \mathbf{y}) \quad (17)$$

However, instead of continuing the iterative inference process, this skip sampling method estimates the final virtual stained image directly from the state at $\mathbf{x_{t_e}}$

$$\mathbf{x_0^s} = \mathbf{x_{t_e}} - \epsilon_\theta(\mathbf{x_{t_e}}, t_e) \quad (18)$$

During the training phase, both the shallow convolutional network and the denoising network were trained jointly. The parameters of these networks were optimized using the AdamW optimizer (*73*) with an initial learning rate of 1×10^{-4}. The training process used a batch size of 16 and took ~72 hours to converge. For blind testing, the exit point, t_e was empirically set to 10, where both the mean and skip sampling strategies achieved better performance, as illustrated in fig. S5.

Quantitative performance evaluation metrics

To quantitatively evaluate the performance of PAS VS results for the study reported in Figs. 3 and 4, we used 201 FOVs of virtually stained PAS images together with their corresponding HS images for paired image comparisons. In the quantitative evaluation illustrated in Fig. 3, we conducted a comprehensive comparative analysis using several features: image contrast, histogram distributions in the YCbCr color space, Hellinger distance calculated between the color histograms of the VS and HS image pairs, and spatial frequency spectrum analysis. The image contrast is defined as

$$\mathrm{Contrast} = \frac{A_{90\%} - A_{10\%}}{A_{90\%} + A_{10\%}} \quad (19)$$

where $A_{90\%}$ and $A_{10\%}$ represents the 90th percentile and 10th percentile of the intensity values in image A, respectively. In our case,

image A refers to the grayscale version of the VS or HS image. For the color analysis, the paired VS and HS images were converted from RGB to YCbCr color space to facilitate a detailed comparison of the distributions/histograms in the Y, Cb, and Cr channels, performed separately. The Hellinger distances were calculated between the YCbCr color histograms for each VS and GT image pair as follows

$$\text{Hellinger distance}\left[p(x), q(x)\right] = \sqrt{1 - \sum_{x \in X} \sqrt{p(x)q(x)}} \quad (20)$$

where $p(x)$ and $q(x)$ are the probability distributions of the YCbCr color (approximated using the extracted color distribution points). The Hellinger distance ranges between 0 (identical) and 1 (completely different). As for the spatial frequency spectrum analysis, the single channel raw MS image was bilinearly upsampled by a factor of 10, from 64 × 64 pixels to 640 × 640 pixels, matching the dimensions of the grayscale VS and HS images. The frequency spectrum of each image was obtained by performing a 2D Fourier Transform on the 10× bilinearly upsampled single-channel IMS image (selected on the basis of the contrast of glomeruli), the VS output and the corresponding HS GT image. For this analysis, both the VS and HS images were processed in gray scale. The radially averaged power spectrum was calculated according to Wang *et al.* (*57*) Moreover, the relative error between the radially averaged power spectra (log scale, normalized) of the VS image and their HS GT (represented by blue and red curves in Fig. 3D) is defined as

$$\text{Relative err} = \frac{\left(P_{\text{VS}} - P_{\text{HS}}\right)}{P_{\text{HS}}} \cdot 100\% \quad (21)$$

where P_{VS} and P_{HS} denote the radially averaged power spectra (log scale, normalized) of the VS image and its HS counterpart. Similarly, the relative gain of the radially averaged power spectra (log scale, normalized) of the VS images compared to their lower resolution MS input (blue and green curves in Fig. 3D) is defined as

$$\text{Relative gain} = \frac{\left(P_{\text{VS}} - P_{\text{MS}}\right)}{P_{\text{MS}}} \cdot 100\% \quad (22)$$

where P_{VS} and P_{MS} represent the radially averaged power spectra (log scale, normalized) of the VS image and the corresponding single-channel MS input.

For the evaluations reported in Fig. 4, we used the metrics of PSNR, PCC, and LPIPS. The PSNR is defined on the basis of mean squared error (MSE)

$$\text{MSE} = \frac{1}{MN} \sum_{m}^{M} \sum_{n}^{N} \left[A_{mn} - B_{mn}\right]^2 \quad (23)$$

where A and B present the histochemically and virtually stained bright-field PAS images, respectively. m and n are the pixel indices, and MN denotes the total number of pixels in each image. PSNR can be denoted as

$$\text{PSNR} = 10\log_{10}\left(\frac{\max(A)^2}{\text{MSE}}\right) \quad (24)$$

where $\max(A)$ is the maximum pixel value of the GT HS PAS image.

The PCC was used to quantify the linear correlation between the HS PAS images and the corresponding VS images. PCC is calculated as

$$PCC = \frac{\sum_{m}^{M} \sum_{n}^{N} \left(A_{mn} - \overline{A}\right)\left(B_{mn} - \overline{B}\right)}{\sqrt{\sum_{m}^{M} \sum_{n}^{N} \left(A_{mn} - \overline{A}\right)^2} \sqrt{\sum_{m}^{M} \sum_{n}^{N} \left(B_{mn} - \overline{B}\right)^2}} \quad (25)$$

where $\overline{A}$ and $\overline{B}$ represent the mean pixel value of A and B.

The calculation of the LPIPS metrics used a pretrained VGG network (*74*) to evaluate the learned perceptual similarity between the generated VS images n and their corresponding HS images n_0. These compared image pairs were fed into the pretrained VGG network and their feature stack from L layers can be extracted as $\hat{m}^l, \hat{m}_0^l \in \mathbb{R}^{H_l \times W_l \times C_l}$ for layer l. The LPIPS score can be denoted as

$$d\left(n, n_0\right) = \sum_{l} \frac{1}{H_l W_l} \sum_{h,w} \left\| \left(\hat{m}_{hw}^l - \hat{m}_{0hw}^l\right) \right\|_2^2 \quad (26)$$

The Frechet inception distance [FID; (*75*)] values, reported in fig. S5, are calculated as follows

$$\text{FID} = \left\|\mu - \mu_\omega\right\|^2 + tr\left(\Sigma + \Sigma_\omega - 2\left(\Sigma\Sigma_\omega\right)^{\frac{1}{2}}\right) \quad (27)$$

where $N(\mu, \Sigma)$ represents the multivariate normal distribution estimated from the Inception v3 features (*75*, *76*) of the GT PAS-stained images and $N\left(\mu_\omega, \Sigma_\omega\right)$ represents the distribution estimated from the Inception v3 features of the generated virtually stained images. We used the 36 VS-HS image pairs (each with 640 × 640 pixels) as input to the Inception v3 network to extract 2048-dimensional feature vectors. From these extracted features, the mean vectors $\left(\mu, \mu_\omega\right)$ and covariance matrices $\left(\Sigma, \Sigma_\omega\right)$ were estimated.

The naturalness image quality evaluator [NIQE; (*77*)] values reported in fig. S5, are calculated as

$$D\left(\nu_1, \nu_2, \Sigma_1, \Sigma_2\right) = \sqrt{\left[\left(\nu_1 - \nu_2\right)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} \left(\nu_1 - \nu_2\right)\right]} \quad (28)$$

where ν_1 and Σ_1 represent the mean vector and covariance matrix of our customized reference multivariate Gaussian (MVG) model, which was fitted using 800 GT PAS image patches (each with 640 × 640 pixels) sampled from both the training and testing datasets. Meanwhile, ν_2 and Σ_2 are the mean vector and covariance matrix for the MVG model of a generated virtually stained image. The MATLAB functions fitniqe and niqe were used to fit the reference MVG model and compute the final NIQE values for the virtually stained images (*78*).

Statistical analysis

In Fig. 3, a two-tailed paired t test was conducted to assess whether the image contrast between the virtually stained output and its HS PAS counterpart was statistically equivalent, with a statistical significance level of 0.05. This analysis was performed on 201 paired VS and HS images. A P value greater than 0.05 indicates no statistically significant difference in the image contrast between the VS images and their HS counterparts.

Other implementation details

All network training and testing tasks were conducted on a desktop computer equipped with an Intel Core i9-13900K CPU, 64 GB of memory, and an NVIDIA GeForce RTX 4090 GPU. The code for training the diffusion models was developed in Python 3.9.19 using PyTorch 2.2.1.

Supplementary Materials

This PDF file includes:

Figs. S1 to S7

REFERENCES AND NOTES

1. R. M. Caprioli, T. B. Farmer, J. Gile, Molecular imaging of biological samples: Localization of peptides and proteins using MALDI-TOF MS. *Anal. Chem.* **69**, 4751–4760 (1997).
2. M. E. Colley, A. B. Esselman, C. F. Scott, J. M. Spraggins, High-specificity imaging mass spectrometry. *Annu. Rev. Anal. Chem.* **17**, 1–24 (2024).
3. K. V. Djambazova, J. M. van Ardenne, J. M. Spraggins, Advances in imaging mass spectrometry for biomedical and clinical research. *TrAC Trends Anal. Chem.* **169**, 117344 (2023).
4. K. K. Krestensen, R. M. A. Heeren, B. Balluff, State-of-the-art mass spectrometry imaging applications in biomedical research. *Analyst* **148**, 6161–6187 (2023).
5. J. M. Spraggins, K. V. Djambazova, E. S. Rivera, L. G. Migas, E. K. Neumann, A. Fuetterer, J. Suetering, N. Goedecke, A. Ly, R. Van De Plas, R. M. Caprioli, High-performance molecular imaging with MALDI trapped ion-mobility time-of-flight (timsTOF) mass spectrometry. *Anal. Chem.* **91**, 14552–14560 (2019).
6. E. K. Neumann, L. G. Migas, J. L. Allen, R. M. Caprioli, R. Van De Plas, J. M. Spraggins, Spatial metabolomics of the human kidney using MALDI trapped ion mobility imaging mass spectrometry. *Anal. Chem.* **92**, 13084–13091 (2020).
7. M. J. Taylor, J. K. Lukowski, C. R. Anderton, Spatially resolved mass spectrometry at the single cell: Recent innovations in proteomics and metabolomics. *J. Am. Soc. Mass Spectrom.* **32**, 872–894 (2021).
8. E. C. Spivey, J. C. McMillen, D. J. Ryan, J. M. Spraggins, R. M. Caprioli, Combining MALDI-2 and transmission geometry laser optics to achieve high sensitivity for ultra-high spatial resolution surface analysis. *J. Mass Spectrom.* **54**, 366–370 (2019).
9. M. Niehaus, J. Soltwisch, M. E. Belov, K. Dreisewerd, Transmission-mode MALDI-2 mass spectrometry imaging of cells and tissues at subcellular resolution. *Nat. Methods* **16**, 925–931 (2019).
10. T. Bien, S. Bessler, K. Dreisewerd, J. Soltwisch, Transmission-mode MALDI mass spectrometry imaging of single cells: Optimizing sample preparation protocols. *Anal. Chem.* **93**, 4513–4520 (2021).
11. S. M. Puolitaival, K. E. Burnum, D. S. Cornett, R. M. Caprioli, Solvent-free matrix dry-coating for MALDI imaging of phospholipids. *J. Am. Soc. Mass Spectrom.* **19**, 882–886 (2008).
12. D. M. G. Anderson, K. A. Floyd, S. Barnes, J. M. Clark, J. I. Clark, H. Mchaourab, K. L. Schey, A method to prevent protein delocalization in imaging mass spectrometry of non-adherent tissues: Application to small vertebrate lens imaging. *Anal. Bioanal. Chem.* **407**, 2311–2320 (2015).
13. R. Van de Plas, J. Yang, J. Spraggins, R. M. Caprioli, Image fusion of mass spectrometry and microscopy: A multimodality paradigm for molecular tissue mapping. *Nat. Methods* **12**, 366–372 (2015).
14. L. E. Tideman, L. G. Migas, K. V. Djambazova, N. H. Patterson, R. M. Caprioli, J. M. Spraggins, R. Van de Plas, Automated biomarker candidate discovery in imaging mass spectrometry data through spatially localized Shapley additive explanations. *Anal. Chim. Acta* **1177**, 338522 (2021).
15. N. Verbeeck, J. M. Spraggins, M. J. Murphy, H. Wang, A. Y. Deutch, R. M. Caprioli, R. Van de Plas, Connecting imaging mass spectrometry and magnetic resonance imaging-based anatomical atlases for automated anatomical interpretation and differential analysis. *Biochim. Biophys. Acta Proteins Proteom.* **1865**, 967–977 (2017).
16. N. H. Patterson, M. Tuck, A. Lewis, A. Kaushansky, J. L. Norris, R. Van De Plas, R. M. Caprioli, Next generation histology-directed imaging mass spectrometry driven by autofluorescence microscopy. *Anal. Chem.* **90**, 12404–12413 (2018).
17. W. M. Abdelmoula, M. S. Regan, B. G. C. Lopez, E. C. Randall, S. Lawler, A. C. Mladek, M. O. Nowicki, B. M. Marin, J. N. Agar, K. R. Swanson, T. Kapur, J. N. Sarkaria, W. Wells, N. Y. R. Agar, Automatic 3D nonlinear registration of mass spectrometry imaging and magnetic resonance imaging data. *Anal. Chem.* **91**, 6206–6216 (2019).
18. M. A. Jones, S. H. Cho, N. H. Patterson, R. Van De Plas, J. M. Spraggins, M. R. Boothby, R. M. Caprioli, Discovering new lipidomic features using cell type specific fluorophore expression to provide spatial and biological specificity in a multimodal workflow with MALDI imaging mass spectrometry. *Anal. Chem.* **92**, 7079–7086 (2020).
19. B. Bai, X. Yang, Y. Li, Y. Zhang, N. Pillar, A. Ozcan, Deep learning-enabled virtual histological staining of biological samples. *Light Sci. Appl.* **12**, 57 (2023).
20. Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydın, J. E. Zuckerman, T. Chong, A. E. Sisk, L. M. Westbrook, W. D. Wallace, A. Ozcan, Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat. Biomed. Eng.* **3**, 466–477 (2019).
21. N. Borhani, A. J. Bower, S. A. Boppart, D. Psaltis, Digital staining through the application of deep neural networks to multi-modal multi-photon microscopy. *Biomed. Opt. Express* **10**, 1339–1350 (2019).
22. Y. Zhang, K. de Haan, Y. Rivenson, J. Li, A. Delis, A. Ozcan, Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. *Light Sci. Appl.* **9**, 78 (2020).
23. J. Li, J. Garfinkel, X. Zhang, D. Wu, Y. Zhang, K. De Haan, H. Wang, T. Liu, B. Bai, Y. Rivenson, Biopsy-free in vivo virtual histology of skin using deep learning. *Light Sci. Appl.* **10**, 233 (2021).
24. P. Pradhan, T. Meyer, M. Vieth, A. Stallmach, M. Waldner, M. Schmitt, J. Popp, T. Bocklitz, Computational tissue staining of non-linear multimodal imaging using supervised and unsupervised deep learning. *Biomed. Opt. Express* **12**, 2280–2298 (2021).
25. B. Bai, H. Wang, Y. Li, K. De Haan, F. Colonnese, Y. Wan, J. Zuo, N. B. Doan, X. Zhang, Y. Zhang, J. Li, X. Yang, W. Dong, M. A. Darrow, E. Kamangar, H. S. Lee, Y. Rivenson, A. Ozcan, Label-free virtual HER2 immunohistochemical staining of breast tissue using deep learning. *BME Front.* **2022**, 9786242 (2022).
26. T. M. Abraham, P. C. Costa, C. Filan, Z. Guang, Z. Zhang, S. Neill, J. J. Olson, R. Levenson, F. E. Robles, Label-and slide-free tissue histology using 3D epi-mode quantitative phase imaging and virtual hematoxylin and eosin staining. *Optica* **10**, 1605–1618 (2023).
27. R. Cao, S. D. Nelson, S. Davis, Y. Liang, Y. Luo, Y. Zhang, B. Crawford, L. V. Wang, Label-free intraoperative histology of bone tissue via deep-learning-assisted ultraviolet photoacoustic microscopy. *Nat. Biomed. Eng.* **7**, 124–134 (2023).
28. Y. Li, N. Pillar, J. Li, T. Liu, D. Wu, S. Sun, G. Ma, K. de Haan, L. Huang, Y. Zhang, S. Hamidi, A. Urisman, T. Keidar Haran, W. D. Wallace, J. E. Zuckerman, A. Ozcan, Virtual histological staining of unlabeled autopsy tissue. *Nat. Commun.* **15**, 1684 (2024).
29. X. Yang, B. Bai, Y. Zhang, M. Aydin, Y. Li, S. Y. Selcuk, P. Casteleiro Costa, Z. Guo, G. A. Fishbein, K. Atlan, W. D. Wallace, N. Pillar, A. Ozcan, Virtual birefringence imaging and histological staining of amyloid deposits in label-free tissue using autofluorescence microscopy and deep learning. *Nat. Commun.* **15**, 7978 (2024).
30. Y. Zhang, L. Huang, N. Pillar, Y. Li, H. Chen, A. Ozcan, Super-resolved virtual staining of label-free tissue using diffusion models. arXiv:2410.20073v1 (2024).
31. E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O'Neil, K. Shah, A. K. Lee, P. Goyal, W. Fedus, R. Poplin, A. Esteva, M. Berndl, L. L. Rubin, P. Nelson, S. Finkbeiner, In silico labeling: Predicting fluorescent labels in unlabeled images. *Cell* **173**, 792–803.e19 (2018).
32. C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, G. R. Johnson, Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat. Methods* **15**, 917–920 (2018).
33. M. E. Kandel, Y. R. He, Y. J. Lee, T. H.-Y. Chen, K. M. Sullivan, O. Aydin, M. T. A. Saif, H. Kong, N. Sobh, G. Popescu, Phase imaging with computational specificity (PICS) for measuring dry mass changes in sub-cellular compartments. *Nat. Commun.* **11**, 6256 (2020).
34. A. Somani, A. Ahmed Sekh, I. S. Opstad, Å. Birna Birgisdottir, T. Myrmel, B. Singh Ahluwalia, A. Horsch, K. Agarwal, D. K. Prasad, Virtual labeling of mitochondria in living cells using correlative imaging and physics-guided deep learning. *Biomed. Opt. Express* **13**, 5495–5516 (2022).
35. L. Kreiss, S. Jiang, X. Li, S. Xu, K. C. Zhou, K. C. Lee, A. Mühlberg, K. Kim, A. Chaware, M. Ando, L. Barisoni, S. A. Lee, G. Zheng, K. J. Lafata, O. Friedrich, R. Horstmeyer, Digital staining in optical microscopy using deep learning - a review. *PhotoniX* **4**, 34 (2023).
36. D. Winter, N. Triltsch, M. Rosati, A. Shumilov, Z. Kokaragac, Y. Popov, T. Padel, L. S. Monasor, R. Hill, M. Schick, N. Brieu, Mask-guided cross-image attention for zero-shot in-silico histopathologic image generation with a diffusion model. arXiv:2407.11664 (2024).
37. Z. Li, T. Su, B. Zhang, W. Han, S. Zhang, G. Sun, Y. Cong, X. Chen, J. Qi, Y. Wang, His-MMDM: Multi-domain and multi-omics translation of histopathology images with diffusion models. medRxiv 24310294 [Preprint] (2024). https://doi.org/10.1101/2024.07.11.24310294.
38. B. Li, K. Xue, B. Liu, Y.-K. Lai, "Bbdm: Image-to-image translation with brownian bridge diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2023), pp. 1952–1961; http://openaccess.thecvf.com/content/CVPR2023/html/Li_BBDM_Image-to-Image_Translation_With_Brownian_Bridge_Diffusion_Models_CVPR_2023_paper.html.
39. D. Revuz, M. Yor, *Continuous Martingales and Brownian Motion* (Springer Science & Business Media, 2013), vol. 293; https://books.google.com/books?hl=en&lr=&id=IWjsCAAAQBAJ&oi=fnd&pg=PA1&dq=Revuz,+Daniel%3B+Yor,+Marc+(1999).+Continuous+Martingales+and+Brownian+Motion+(2nd+ed.).+New+York:+Springer-Verlag.&ots=hzrilbNqul&sig=DqvBYBTle9EEUFzeRascv_U70bI.
40. O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), pp. 234–241.

41. A. B. Esselman, N. H. Patterson, L. G. Migas, M. Dufresne, K. V. Djambazova, M. E. Colley, R. Van de Plas, J. M. Spraggins, Microscopy-directed imaging mass spectrometry for rapid high spatial resolution molecular imaging of glomeruli. *J. Am. Soc. Mass Spectrom.* **34**, 1305–1314 (2023).
42. J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **33**, 6840–6851 (2020).
43. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2022), pp. 10684–10695; http://openaccess.thecvf.com/content/CVPR2022/html/Rombach_High-Resolution_Image_Synthesis_With_Latent_Diffusion_Models_CVPR_2022_paper.html.
44. D. Podell, Z. English, K. Lacey, A. Blattmann, T. Dockhorn, J. Müller, J. Penna, R. Rombach, SDXL: Improving latent diffusion models for high-resolution image synthesis. arXiv:2307.01952 (2023).
45. L. Zhang, A. Rao, M. Agrawala, "Adding conditional control to text-to-image diffusion models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (IEEE, 2023), pp. 3836–3847; http://openaccess.thecvf.com/content/ICCV2023/html/Zhang_Adding_Conditional_Control_to_Text-to-Image_Diffusion_Models_ICCV_2023_paper.html.
46. T. Brooks, A. Holynski, A. A. Efros, "Instructpix2pix: Learning to follow image editing instructions," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2023), pp. 18392–18402; http://openaccess.thecvf.com/content/CVPR2023/html/Brooks_InstructPix2Pix_Learning_To_Follow_Image_Editing_Instructions_CVPR_2023_paper.html.
47. A. Janowczyk, R. Zuo, H. Gilmore, M. Feldman, A. Madabhushi, HistoQC: An open-source quality control tool for digital pathology slides. *JCO Clin. Cancer Inform.* **3**, 1–7 (2019).
48. E. Hellinger, Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *J. Für Reine Angew. Math.* **1909**, 210–271 (1909).
49. H. Jeffreys, An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **186**, 453–461 (1946).
50. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, A. Ozcan, Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).
51. K. de Haan, Z. S. Ballard, Y. Rivenson, Y. Wu, A. Ozcan, Resolution enhancement in scanning electron microscopy using deep learning. *Sci. Rep.* **9**, 12050 (2019).
52. R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 586–595; http://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_The_Unreasonable_Effectiveness_CVPR_2018_paper.html.
53. A. Q. Nichol, P. Dhariwal, "Improved denoising diffusion probabilistic models," in *International Conference on Machine Learning* (PMLR, 2021), pp. 8162–8171; https://proceedings.mlr.press/v139/nichol21a.html.
54. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K. Q. Weinberger, Eds. (Curran Associates Inc., 2014), vol. 27; https://doi.org/10.1145/3422622.
55. M. Mirza, Conditional generative adversarial nets. arXiv:1411.1784 (2014).
56. P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2017), pp. 1125–1134.
57. H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L. A. Bentolila, C. Kural, A. Ozcan, Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103–110 (2019).
58. K. de Haan, Y. Rivenson, Y. Wu, A. Ozcan, Deep-learning-based image reconstruction and enhancement in optical microscopy. *Proc. IEEE* **108**, 30–50 (2020).
59. C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, M. Norouzi, Image super-resolution via iterative refinement. *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 4713–4726 (2022).
60. S. Gao, X. Liu, B. Zeng, S. Xu, Y. Li, X. Luo, J. Liu, X. Zhen, B. Zhang, "Implicit diffusion models for continuous super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2023), pp. 10021–10030; http://openaccess.thecvf.com/content/CVPR2023/html/Gao_Implicit_Diffusion_Models_for_Continuous_Super-Resolution_CVPR_2023_paper.html.
61. L. Huang, X. Xiao, S. Li, J. Sun, Y. Huang, A. Ozcan, P. Bogdan, Multi-scale conditional generative modeling for microscopic image restoration. arXiv:2407.05259 (2024).
62. L. Metz, B. Poole, D. Pfau, J. Sohl-Dickstein, Unrolled generative adversarial networks. arXiv:1611.02163 (2017).
63. Z. Wang, H. Zheng, P. He, W. Chen, M. Zhou, Diffusion-GAN: Training GANs with diffusion. arXiv:2206.02262 (2022).
64. M. S. Asghar, A. Denic, A. D. Rule, Morphometric analysis of chronicity on kidney biopsy: A useful prognostic exercise. *Clin. Kidney J.* **17**, sfad226 (2024).
65. E. Neumann, J. Allen, A. Kruse, J. Harvey, M. Brewer, C. Romer, M. De Caestecker, J. Spraggins, PAS staining of fresh frozen or paraffin embedded human kidney tissue, protocols.io (2023); https://dx.doi.org/10.17504/protocols.io.x54v98nkzl3e/v3.
66. S. Klein, M. Staring, K. Murphy, M. A. Viergever, J. P. Pluim, Elastix: A toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* **29**, 196–205 (2009).
67. H. Patterson, NHPatterson/wsireg, GitHub (2022); https://github.com/NHPatterson/wsireg.
68. H. Patterson, NHPatterson/napari-imsmicrolink, GitHub (2022); https://github.com/NHPatterson/napari-imsmicrolink.
69. W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 1874–1883; www.cv-foundation.org/openaccess/content_cvpr_2016/html/Shi_Real-Time_Single_Image_CVPR_2016_paper.html.
70. A. Vaswani, Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, 6000–6010 (2017).
71. D. Hendrycks, K. Gimpel, Gaussian error linear units (gelus). arXiv:1606.08415 (2016).
72. S. Elfwing, E. Uchibe, K. Doya, Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Netw.* **107**, 3–11 (2018).
73. I. Loshchilov, F. Hutter, Decoupled weight decay regularization. arXiv:1711.05101 (2017).
74. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 (2015).
75. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 2818–2826; www.cv-foundation.org/openaccess/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html.
76. M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, GANs trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neural Inf. Process. Syst.* **30**, 6629–6640 (2017).
77. A. Mittal, R. Soundararajan, A. C. Bovik, Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.* **20**, 209–212 (2012).
78. niqe, Naturalness Image Quality Evaluator (NIQE) no-reference image quality score, MATLAB; www.mathworks.com/help/images/ref/niqe.html.
79. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, Auto-fluorescence microscopy data right kidney, HuBMAP Consortium (2024); https://doi.org/10.35079/HBM888.PHHS.299.
80. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, Auto-fluorescence microscopy data left kidney, HuBMAP Consortium (2024); https://doi.org/10.35079/HBM578.WQFJ.765.
81. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, Auto-fluorescence microscopy data right kidney, HuBMAP Consortium (2024); https://doi.org/10.35079/HBM576.CBTX.826.
82. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, Auto-fluorescence microscopy data right kidney, HuBMAP Consortium (2024); https://doi.org/10.35079/HBM353.XGRQ.833.
83. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, Auto-fluorescence microscopy data right kidney, HuBMAP Consortium (2024); https://doi.org/10.35079/HBM836.DHDQ.639.
84. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, MALDI Imaging mass spectrometry data right kidney, HuBMAP Consortium (2023); https://doi.org/10.35079/HBM437.BGCD.226.
85. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, MALDI Imaging mass spectrometry data right kidney, HuBMAP Consortium (2023); https://doi.org/10.35079/HBM644.HLRW.739.
86. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, MALDI Imaging mass spectrometry data right kidney, HuBMAP Consortium (2023); https://doi.org/10.35079/HBM286.ZZNV.536.
87. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, MALDI Imaging mass spectrometry data right kidney, HuBMAP Consortium (2023); https://doi.org/10.35079/HBM946.ZHFC.683.

88. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, MALDI Imaging mass spectrometry data Left kidney, HuBMAP Consortium (2023); https://doi.org/10.35079/HBM794.JRXQ.455.
89. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, Periodic acid-Schiff stained microscopy data right kidney, HuBMAP Consortium (2024); https://doi.org/10.35079/HBM396.XMSB.623.
90. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, Periodic acid-Schiff stained microscopy data right kidney, HuBMAP Consortium (2024); https://doi.org/10.35079/HBM964.CZLM.657.
91. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, Periodic acid-Schiff stained microscopy data right kidney, HuBMAP Consortium (2024); https://doi.org/10.35079/HBM738.MXDD.328.
92. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, Periodic acid-Schiff stained microscopy data right kidney, HuBMAP Consortium (2024); https://doi.org/10.35079/HBM864.MFDZ.983.
93. J. M. Spraggins, A. B. Fogo, D. B. Gutierrez, D. M. G. Anderson, E. K. Neumann, H. Yang, J. L. Allen, L. Tideman, L. Migas, M. E. Colley, M. deCaestecker, M. Brewer, M. A. Farrow, N. H. Patterson, R. Van de Plas, R. C. Harris, R. M. Caprioli, Periodic acid-Schiff stained microscopy data left kidney, HuBMAP Consortium (2024); https://doi.org/10.35079/HBM593.DVGN.469.

Acknowledgments: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We report the use of Grammarly (through MS Word), and we take full responsibility for the content of this paper. **Funding:** The Ozcan Research Group at UCLA acknowledges the support of NIH P41, the National Center for Interventional Biophotonic Technologies (NCIBT). We also acknowledge the NIH Common Fund, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), and the Office of the Director (OD) under Award Numbers U54DK134302 and U01DK133766 (J.M.S. and R.V.d.P.) and, in part, by grant numbers 2021-240339 and 2022-309518 (L.G.M. and R.V.d.P.) from the Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation. **Author contributions:** A.O. conceived the idea. Y.Z. developed the image processing pipeline and prepared the training/validation and test datasets. Y.Z., L.H., and Yuzhu Li, trained/tested the neural networks. Y.Z., L.H., N.P., Yuzhu Li, and Yuhang Li performed the result analysis and statistical studies. L.G.M., R.V.d.P., and J.M.S. contributed optical microscopy and IMS-related data and image registration methods. Y.Z., L.H., and A.O. prepared the manuscript. A.O. supervised the research. **Competing interests:** Y.Z., L.H., and A.O. have a pending patent application filed by UCLA on VS technology. Patent status: pending; coinventors: Y.Z., L.H., and A.O.; application no. 63/712,348; filing date: 25 October 2024. The authors declare that they have no other competing interests. **Data and materials availability:** The AF images (*79*–*83*), IMS data (*84*–*88*), and PAS-stained images (*89*–*93*) can be accessed through the Human Biomolecular Atlas Program (HuBMAP): https://hubmapconsortium.org. Deep learning models reported in this work used standard libraries and scripts that are publicly available in PyTorch. The codes for our VS framework can be found at https://doi.org/10.5281/zenodo.15043291. All other data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 4 December 2024
Accepted 30 June 2025
Published 1 August 2025
10.1126/sciadv.adv0741