Supplementary Discussion

Other Approaches to Imaging 3D Tissue Specimens

3D imaging challenges the ability to balance between acquisition time, sensitivity, and most notably, reduction of out-of-focus signal that would otherwise reduce contrast. The latter is required for resolving structures in thick tissue along the axial direction, where the axial resolution depends on wavelength, numerical aperture of the objective lens, axial sampling step size, and pinhole diameter in the case of a confocal. Traditional widefield microscopy can be supplemented with deconvolution to reassign out-of-focus signal back to the focal plane, which requires knowledge of the point spread function and is generally suitable for thin samples[53]. Total internal reflection fluorescence (TIRF) microscopy[54,55] can produce high contrast very thin 3D images when used in conjunction with other techniques such as DNA paint, but is limited to only 100 nm beyond the coverslip and is not suitable for highly multiplexed imaging protocols where the tissue is not mounted to the coverslip. More commonly found in research settings are spinning disk[56] and laser scanning confocal[57,58] microscopes where the latter is much slower due to the need to laser raster but provides finer control over the pinhole diameter, and thus, the size of the optical sectioning and the gain in contrast by rejection of out of focus light.

Confocal microscopes are relatively light inefficient; between a widefield and laser scanning confocal, orders of magnitude far fewer photons from the sample are rejected and never reach the detector[59] although newer micro lenses and sensitive detectors are able to recapture much of this signal. Additionally, sample thickness at any given point is fully illuminated regardless of which focal plane is currently being imaged and this significantly contributes to photobleaching and challenges in downstream quantitative measurements. In specialized research settings, two-photon excitation (TPE) microscopy[60] uses pulsed mode-locked lasers to excite a femto-litre sized volume via the simultaneous absorption of two photons (typically in the near-infrared range). Two-photon lasers can also capitalize on certain collagen types that emit a label-free second harmonic signal[61–63]; the resulting signal has a wavelength that is precisely half of the incident laser wavelength. Since near-infrared wavelengths experience less attenuation and contribute to less phototoxicity, TPE is preferred in tissue and intravital imaging. Furthermore, multiple fluorophores can share similar two-photon excitation spectra, therefore, this creates a throughput advantage in thick specimens. However, these systems are less common than confocal microscopes because they are significantly costlier and require specialist knowledge to operate and maintain.

A computationally-intensive solution to combine high-plex imaging with coarse 3D data has been to register several thin serial sections,[49,64–66] however, the axial resolution of the final dataset is on the order of 5-20 microns and, thus, consists of incomplete cell volumes. A variety of Light Sheet Fluorescence Microscopy (LSFM) and tissue clearing methods[67–69] have been developed to image tissue sections as thick as several mm, with the most recent capable of subcellular resolution, but

not all clearing methods are compatible with FFPE tissue[70] and the total number of fluorophores that has been imaged is up to 5 or fewer depending on the optics[71,72]. While spectral unmixing, spillover compensation, or the use of Raman dyes can increase this number to 8-10 channels[73–75], to date, sequential staining and high-plex high-resolution (sub-micrometer scale) imaging of thick sections has not been demonstrated.

Given all of these considerations, it was not self-evident that confocal microscopy would be suitable for high-plex imaging of thick tissue sections. We are of efforts to create specialized multiplexed deep imaging microscopes, but the our data suggest that an existing Zeiss LSM980 Confocal microscope can be used for CyCIF with few compromises in terms of sensitivity, speed or resolution. We note that the current work does not fully exploit the spectral ummixing capabilities of the LSM980 and that this awaits further development of antibody panels.

Methods

Specimen collection

Specimens for melanoma (MIS and VGP), glioblastoma, lung metastasis, and tonsil were retrieved from the archives of the Department of Pathology at Brigham and Women's Hospital and collected under Institutional Review Board approval (FWA00007071, Protocol IRB18-1363) under a waiver of consent. Three datasets were used for all experiments: two 35 µm serial sections of melanoma (referred to as Dataset 1 (LSP13626) and Dataset 2 (LSP13625)) and a 35 µm section of metastatic melanoma from the NIH Cooperative Human Tissue Network (CHTN) (referred to as metastatic melanoma or dataset 3; LSP22409 / WD-100476). Serous Tubal Intraepithelial Carcinoma (STIC) samples were obtained from University of Pennsylvania. Quantifications are based on Dataset 1. Deep immune cell phenotyping is based on Dataset 3. See **Supplementary Table 2** for clinical metadata regarding this specimen and the related HTAN identifiers. The histopathological regions of interest for each specimen were annotated from a serial section consistent with the work of Nirmal, Vallius, Maliga et al. (2022)[18] by a board-certified pathologist based on melanoma diagnostic criteria. See **Supplementary Table 6** for an index of which figure panels relate to which datasets.

Cyclic immunofluorescence (CyCIF)

The procedure for thick section CyCIF is similar that of standard CyCIF[76] except that additional care is taken during staining steps (see below) A staining plan containing a list of antibodies used can be found in **Supplementary Table 3-5**. Antigen retrieval, staining, and bleaching was performed as described previously.[18]. Due to the fragile nature of thicker samples, extra care was taken during washes, bleaching and decoverslipping. We found that most tissues held up well, but that a subset of skin and melanoma samples disintegrated during antigen retrieval. We have observed this previously with skin and primary melanoma, and 3D imaging showed that these specimens had not fully adhered to the cover slip, instead exhibiting a series of corrugations just above the cover slip with liquid in between. This issue lies in the realm of “pre-analytical variables,” which are common in

histology, and will required additional work to resolve. ,Antibodies for each cycle were diluted in 400 µl of blocking buffer and stained for 8-10 hours to allow for same-day imaging and at room temperature to encourage penetration of antibodies (see below). See **Supplementary Figures 1-11** for the whole slide images of the full dataset for each sample.

Optimization of sample thickness and antibody staining protocol

We sought to determine an ideal tissue thickness for CyCIF imaging using well-characterized tonsil sections. Based on the maximum working distance of most water and oil-immersion lenses (~200 µm) and the thickness of a grade 1.5 coverslip (170 µm), we obtained tonsil sections that were 10 µm, 20 µm, 30 µm, 35 µm, and 40 µm thick (**Supplementary Fig. 13**). These were stained with Hoechst and y-tubulin conjugated in Alexafluor 555. Gamma-tubulin is punctate and can serve as a useful stain for assessing antibody penetration and image aberration. Z-stacks were acquired for each at 103 nm lateral resolution and 230 nm axial resolution with a 40x/1.2W C-Apochromat water immersion objective lens on a Zeiss LSM980 confocal microscope. We observed punctate gamma-tubulin in all thicknesses up to 35 µm tissue thickness, with uniform intensity along the axial axis (**Supplementary Fig. 13a-d**). However, in the 40 µm thickness, gamma-tubulin appeared to significantly diminish in intensity along the axial axis (**Supplementary Figure 13e**). Furthermore, contrast even at the top surface was poorer than thinner samples. This could suggest that standard dewaxing and antigen retrieval protocols are not suitable for thicknesses greater than 35 µm. With Hoechst staining, we also observed severe signal attenuation in the 40 µm. Unlike gamma-tubulin, excitation and emission light penetration may be the issue here, which is well established for short wavelengths.

We then evaluated if certain fluorophores impacted antibody penetration. This is important for CyCIF where the ability to choose different antibody fluorophore combinations is essential. We obtained a primary melanoma and co-stained MART1 conjugated to Alexafluor 647 with other secondary antibodies (Alexafluor 488, Alexafluor 555, Alexafluor 750) (**Supplementary Fig. 14a**) for 8-10 hours at room temperature. We bleached MART1-647 and restained with Alexafluor 647 in a subsequent cycle. **Supplementary Fig. 14b** shows that the MART1 primary conjugate (magenta) penetrated the full thickness of the tissue, as judged by Hoechst staining (turquoise). **Supplementary Fig. 14c-f** shows that all secondary antibodies (magenta) penetrated equally well and showed a similar staining pattern to the MART1 primary conjugate. This demonstrates the ability for secondary antibodies to be used for thick tissue CyCIF. We noted that Alexafluor 750 had lower contrast, which can be attributed to the lower sensitivity of detectors in the near infrared spectrum.

While testing multiple primary conjugated antibodies, we observed antibody penetration issues with some antibody conjugates. Although many immune markers (PD1, CD11c, CD8a, MHC-1, MHC-II; green) exhibited full depth staining, several tumour and stromal markers (αSMA, PCNA, SOX10; red) only stained the top layer of tissue (**Supplementary Fig. 15**). To determine whether the

fluorophore played a role in this, we repeated staining with the same PCNA clone conjugated to Alexafluor 488 or Alexafluor 750. We noticed there that was a difference in staining pattern; the Alexafluor 488 conjugate stained fewer cells (**Supplementary Fig. 16a**) but showed improved staining penetration (**Supplementary Fig. 16b**). For αSMA, we tried a similar strategy, but using a different fluorophore required a different antibody clone. Unlike PCNA, we did not see an improvement in staining penetration of a blood vessel (**Supplementary Fig. 17**). From these data we concluded that antibody penetration is not uniquely dependent on fluorophore or clone but is influenced by multiple factors and that each antibody must therefore be evaluated for its ability to stain a thick section using Z-stacks.

3D image acquisition

Tonsil and melanoma image data were collected on a LSM980 Airyscan 2 (Carl Zeiss) equipped with a 405nm, 488nm, 561nm, 647nm, and 750nm laser lines, and 10x/0.45NA air and 40x/1.3NA oil immersion objective lenses. In ZEN 3.7, a 2D overview scan using the 10x objective lens was used to identify regions of interest for higher resolution imaging at 40x in 3D. Images were sampled at 16-bit at 0.14 microns per pixel in X and Y, and 0.28 microns per pixel in Z for approximately 170 or more optical planes. The pinhole size was set to 35 microns. At both resolutions, a focus surface was used to maintain focus. To increase throughput, bidirectional and fast frame scanning was used. Channels were separated into two tracks: track 1 - Hoechst, Alexafluor 555, and Alexafluor 750 (if present). track 2 - Alexafluor 488 and Alexafluor 647. The emission range for Hoechst, Alexafluor 488, Alexafluor 555, Alexafluor 647, and Alexafluor 750 were 380nm-489nm, 499nm-544nm, 579nm-640nm, 660nm-705nm, and 755nm-900nm respectively.

Type I and II collagen were imaged using Second Harmonic Generation (SHG) in a Stellaris 8 DIVE coupled to an Insight X3 multiphoton laser and running LasX. Images were acquired with a 20x/0.75NA multi-immersion lens and sampled at 0.36 microns laterally and 0.95 microns axially. SHG signal was detected using 4Tune Spectra non-descanned HyD detectors and separated from that of Hoechst 33342 using Fluorescence Lifetime Imaging Microscopy (FLIM).

Microscope slides were secured in a slide holder fitted with a spring-loaded clamp, which correspondingly was secured onto the microscope stage in a plateholder.

3D image processing and registration

All channels acquired on the Zeiss LSM980 were processed using Zeiss ZEN LSM Plus Processing to improve signal-to-noise. Channels were background subtracted by removing a fixed constant grey-level from the background. The first cycle was stitched in ZEN using the Hoechst channel as a reference, and all subsequent cycles were registered to this first stitched cycle. Single-field and stitched 3D datasets were imported using Bioformats in MATLAB (Mathworks). First, the X and Y translations were obtained using max projections of the Hoechst nuclei channel. Following this transformation, subsequent cycles were registered in Z. We found that separating the lateral from axial transformations was more accurate than registering X, Y, and Z in one optimization step. We

then performed histogram equalization with MATLAB's *histeq()* function and fine-tuned image alignment with elastic deformations using MATLAB's *imregdemons()* function. Lastly, all transformations for each cycle were applied to their corresponding channels. Each channel was saved and appended to a TIFF file and visualized in Meshlab, ChimeraX or Imaris 10.0 (Bitplane) as .ims files.

Single-Cell Phenotyping

Manual gating was performed for each marker to differentiate background from true signal. The gates identified for each marker were subsequently used to normalize the single-cell data within a range of 0 to 1, wherein values above 0.5 indicated cells expressing the marker. The scaled data was subsequently used for phenotyping the cells based on known lineage markers as described previously using the SCIMAP Python package (scimap.xyz).[18] See **Supplementary Figure 12** for the detailed marker combinations used to define cell types.

RCN Analysis to Identify Microenvironmental Communities

The Latent Dirichlet allocation (LDA) based recurrent cellular neighbourhood (RCN) was performed using SCIMAP (scimap.xyz)[18] using a k value of 10 (**Extended Data Figure 7)**. The clusters were manually organized into meta-clusters (7 clusters), based on the cellular composition of the clusters. The meta-clusters were also overlaid on the H&E and CyCIF images to validate their characteristics. For instance, RCN1 typically aligned with areas known to be tumour domains, while RCN2 was more closely associated with the epidermis, thereby highlighting the structural elements within the dataset.

Segmentation of individual 3D cells with Cellpose

Individual 3D cells were segmented from the dense tissue volumes using Cellpose (https://github.com/MouseLand/cellpose),[77] a custom gradient tracking approach which aggregates x-y, y-z, x-z 2D slice cell probability and gradient maps predicted by pretrained 2D segmentation models. The full Cellpose segmentation framework, suitable for a wide range of 3D cell imaging data along with in-depth validation and determination of method applicability will be described elsewhere, see below for the Cellpost workflow specific to this project.

Image preprocessing for Cellpose

The 3D volumes were acquired at voxel resolution of 280 x 140 x 140 nm. For each 3D channel image, we resized the x-y slices by half to obtain isotropic voxels. The raw image intensity, I_{raw}^{ch} was then corrected for uneven illumination, $I_{correct}^{ch} = \overline{I_{raw}^{ch}} \frac{I_{raw}^{ch}}{I_{bg}^{ch}}$ where $\overline{I_{raw}^{ch}}$ is the mean image intensity and I_{bg}^{ch} an estimation of the background illumination obtained by downsampling the image by a factor of 8, Gaussian smoothing with sigma = 5 and resizing back to the original image dimensions. $I_{correct}^{ch}$ was then contrast-stretched to a range of 0-1, clipping any intensities less than the 2nd percentile to 0 and any greater than the 99.8th percentile to 1. Cellpose uses a single channel

cytoplasmic and nuclear signal for two-color based cell segmentation. The mean of the intensity-normalized, background-corrected HLA-AB, CD3E, CD11b, and β-actin channels was used as the cytoplasmic signal. DAPI was used as the nucleus signal. Both cytoplasmic and nucleus signals underwent a further round of background correction and contrast stretching as described above before being concatenated to form the input RGB volume image.

Running Cellpose 2D

The RGB volume was input slice-by-slice to Cellpose 2D in three different orientations; x-y, x-z, y-z to obtain three stacks of cell probability and 2D gradients. The performance of Cellpose depends on appropriate setting of the diameter parameter which relates to the size of the cells to be segmented. As the appearance of the cells may vary depending on orientation, we conduct a parameter screen with diameter = [10,100] at increments of 5 using the mid-slice for each orientation. At each diameter we compute the 'sharpness' of the predicted gradient map as the mean of the image variance evaluated over a local 5x5 pixel window in both 'x' and 'y' gradient directions. The diameter maximizing the variance after a moving average smoothing with window size of 3 was used to run Cellpose 2D on the remaining slices in the orientation. The raw cell probability output, P from Cellpose are the inputs to a sigmoid centered at zero, $1/(1+e^{-P})$.This means the probabilities vary predominantly linearly in the range -6 to +6 and this reduces the distinction between foreground and background. Thus, we clip the probabilities to the range [-88.72, 88.72] (to prevent overflow or underflow in float32) and convert back to a normalized probability value in the range 0-1 by evaluating the sigmoid, $1/(1+e^{-P})$. The probabilities from all 3 orientations are combined into one by averaging. Similarly, the 2D gradients are Gaussian smoothed with sigma=1 voxel and combined into a single 3D gradient map. Gradients are then normalized to be unit length. Lastly, we perform 3-level Otsu thresholding on the combined probability map and use the lower threshold to define the foreground binary voxels for gradient tracking.

Aggregating Cellpose 2D predictions

The volume was divided into subvolumes of (256, 512, 512) with 25% overlap. Within each subvolume we run gradient descent with momentum for 200 iterations, momenta, $\mu = 0.98$, step size $\delta = 1$ to propagate the position of foreground pixels towards its final attractor in the 3D gradient map.

$$\left(x_i^{t+1}, y_i^{t+1}, z_i^{t+1}\right) \leftarrow \left(x_i^t, y_i^t, z_i^t\right) + \frac{1}{\delta+\mu}\left(\delta \cdot \nabla\left(x_i^t, y_i^t, z_i^t\right) + \mu \cdot \nabla\left(x_i^{t-1}, y_i^{t-1}, z_i^{t-1}\right)\right)$$

Here (x_i^t, y_i^t, z_i^t) denotes the coordinate of foreground voxel i at iteration number t, μ the momentum ranging from 0-1, δ the step size and ∇ is the gradient map. Nearest neighbor interpolation is used, thus (x_i^t, y_i^t, z_i^t) is always integer valued. Gradient tracking of all subvolumes are conducted in parallel using multiprocessing. The final coordinate positions from all subvolumes are compiled. We then build a volume count map where voxels mapping to the same final coordinate adds +1 to the count. The count map is Gaussian smoothed with sigma=1 and binarized using the mean value as

the threshold. Connected component analysis identifies the unique cell as clusters where foreground voxels have been mapped to the same cell. Transferring this labelling to initial voxel positions $\left(x_i^{t=0}, y_i^{t=0}, z_i^{t=0}\right)$ generates the individual 3D cell segmentations.

Postprocessing 3D cell segmentations

Small individual cell masks (<1000 voxels^3 ≈ 20μm^3) were first removed. We also removed all cell masks that do not agree with the Cellpose predicted 3D gradient map. This is done by computing the 3D heat diffusion gradient map given the computed 3D cell segmentations and computing the mean squared error (MSE) with the input combined Cellpose 3D gradient map for each cell. Cells with MSE > 0.8 were discarded. Cells that are implausibly large, with volume greater than the mean volume $\pm$ 5 standard deviations were also discarded.

For the remainder cells, we run a label propagation[78] to enforce that each segmented cell mask comprises only a single connected component and to denoise the masks. This is done for each cell mask, M_i, by cropping a subvolume, V_i, the size of its bounding box padded isotropically by 25 voxels. Each unique cell region is represented as a positive integer label. Every label in V_i is encoded using a one-hot encoding scheme to create a binary column for each unique label. This generates a label matrix, $L_i \in \mathbb{R}^{N \times p+1}$ for V_i, where N is the total number of voxels and p the number of unique labels in V_i and one additional label for background. We then construct the affinity matrix, A, as a weighted sum ($\alpha = 0.25$) of an affinity matrix based on the intensity difference in the cytoplasmic signal between 8-connected voxel neighbors, $A_{intensity}$, and one based on the connectivity alone, $A_{laplacian}$; $A = \alpha A_{intensiy} + (1-\alpha)A_{laplacian}$. $A_{intensity} = \begin{cases} e^{-D_{intensity}/(2\mu(D_{intensity})^2)} & i \neq j \\ 1 & i = j \end{cases}$ where $D_{intensity}$ is the pairwise absolute difference matrix between two neighboring voxels i and j. $A_{laplacian} = \begin{cases} e^{-D^2_{laplacian}/(2\mu(D_{laplacian})^2)} & i \neq j \\ 1 & i = j \end{cases}$ where $D_{laplacian}$ is the graph Laplacian with a value of 1 if a voxel i is a neighbor of voxel j, and 0 otherwise. $\mu(D)$ denotes the mean value of the entries of matrix D. The iterative label propagation is

$$z \in \mathbb{R}^{N \times p}$$

$$z^{t=0} = \mathbf{0}$$

$$z^{t+1} \leftarrow (1-\gamma)A\,z^t + (\gamma)L,$$

where t is the interation number, $\mathbf{0}$, denotes the empty vector and γ is a 'clamping' factor that controls the extent the original labeling is preserved. We set $\gamma = 0.01$. We run the propagation for 25 iterations. The final z is normalized using the softmax operation and argmax is used to obtain the final labels. The refined cell mask, M_i^{refine} is defined by all voxels where z has the same cell label

i. All postprocessing steps were implemented using parallel multiprocessing iterating over individual cells.

Statistical Tests

All statistical tests were performed using MATLAB's *ttest2* implementation of the two-sample t-test without assuming equal variances and significance value of $p<0.05$.