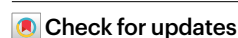


# CODA: quantitative 3D reconstruction of large tissues at cellular resolution

Received: 23 May 2022

Accepted: 14 September 2022

Published online: 24 October 2022



Ashley L. Kiemen<sup>1,11</sup>, Alicia M. Braxton<sup>11</sup>, Mia P. Grahm<sup>1</sup>, Kyu Sang Han<sup>1</sup>, Jaanvi Mahesh Babu<sup>11</sup>, Rebecca Reichel<sup>11</sup>, Ann C. Jiang<sup>2</sup>, Bridgette Kim<sup>3</sup>, Jocelyn Hsu<sup>1</sup>, Falone Amoa<sup>11</sup>, Sashank Reddy<sup>4</sup>, Seung-Mo Hong<sup>5</sup>, Toby C. Cornish<sup>6</sup>, Elizabeth D. Thompson<sup>11</sup>, Peng Huang<sup>7,8</sup>, Laura D. Wood<sup>8,11</sup>, Ralph H. Hruban<sup>8,11</sup>, Denis Wirtz<sup>1,8,9,10,11</sup>✉ and Pei-Hsun Wu<sup>1,10</sup>✉

A central challenge in biology is obtaining high-content, high-resolution information while analyzing tissue samples at volumes relevant to disease progression. We address this here with CODA, a method to reconstruct exceptionally large (up to multicentimeter cubed) tissues at subcellular resolution using serially sectioned hematoxylin and eosin-stained tissue sections. Here we demonstrate CODA's ability to reconstruct three-dimensional (3D) distinct microanatomical structures in pancreas, skin, lung and liver tissues. CODA allows creation of readily quantifiable tissue volumes amenable to biological research. As a testbed, we assess the microanatomy of the human pancreas during tumorigenesis within the branching pancreatic ductal system, labeling ten distinct structures to examine heterogeneity and structural transformation during neoplastic progression. We show that pancreatic precancerous lesions develop into distinct 3D morphological phenotypes and that pancreatic cancer tends to spread far from the bulk tumor along collagen fibers that are highly aligned to the 3D curves of ductal, lobular, vascular and neural structures. Thus, CODA establishes a means to transform broadly the structural study of human diseases through exploration of exhaustively labeled 3D microarchitecture.

The growth of invasive cancer and its spread into microenvironments containing complex vascular, neural, stromal and ductal structures is best understood through accurate three-dimensional (3D) representations<sup>1–3</sup>. Pancreatic ductal adenocarcinoma (PDAC) is one of the deadliest forms of cancer, with a 5-year survival rate of only 10% (refs. <sup>4,5</sup>).

PDAC arises from well-characterized precursor lesions in the pancreatic ducts and has a propensity for metastasis to the liver, lymph nodes and retroperitoneum, often facilitated by vascular and neural invasion<sup>6–8</sup>. These phenomena are classically studied in two dimensions via tissue sectioning and histological staining, where 3D information such as

<sup>1</sup>Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, Baltimore, MD, USA. <sup>2</sup>Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, MD, USA. <sup>3</sup>Department of Mechanical Engineering, The Johns Hopkins University, Baltimore, MD, USA. <sup>4</sup>Department of Plastic and Reconstructive Surgery, The Johns Hopkins University School of Medicine, Baltimore, MD, USA. <sup>5</sup>Department of Pathology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea. <sup>6</sup>Department of Pathology, University of Colorado School of Medicine, Aurora, CO, USA. <sup>7</sup>Department of Biostatistics, The Johns Hopkins University, Baltimore, MD, USA. <sup>8</sup>Department of Oncology, The Johns Hopkins University School of Medicine, Baltimore, MD, USA. <sup>9</sup>Department of Materials Science and Engineering, The Johns Hopkins University, Baltimore, MD, USA. <sup>10</sup>Johns Hopkins Physical Sciences—Oncology Center, The Johns Hopkins University, Baltimore, MD, USA. <sup>11</sup>Present address: Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins University School of Medicine, Baltimore, MD, USA. ✉e-mail: [wirtz@jhu.edu](mailto:wirtz@jhu.edu); [pwu@jhu.edu](mailto:pwu@jhu.edu)

connectivity, morphology and spatial relationships are lost. While many surrogates for studying tumorigenesis have been developed in vitro and in vivo<sup>9–13</sup>, quantitative 3D study of naturally occurring cancers in human tissues, or cancer in situ, is generally lacking.

Recent advances in tissue clearing techniques have been used to explore human diseases in 3D (refs.<sup>14–20</sup>). For example, clearing of human pancreatic samples has been used to study the expression of a limited number of proteins in cancer cells as they invade blood vessels<sup>2</sup> and to enumerate the density of islets of Langerhans to determine the onset of diabetes<sup>21</sup>. However, poor antibody penetration into dense tissues such as PDAC's desmoplastic stroma, long processing times of days to weeks, trade-offs between reconstruction of large volumes and number of structures labeled and longstanding challenges in quantifying complex 3D images hinder the power of tissue clearing techniques<sup>16,19,22</sup>. Reconstruction of serial hematoxylin and eosin (H&E) stained sections using image registration approaches has also been used to study disease in 3D (refs.<sup>23–28</sup>). While use of thinly stained sections avoids the issue of poor antibody penetration seen in study of intact tissues, time-consuming manual annotations and costly immunohistochemical (IHC) labeling and mass spectrometry have been required to identify components in serially sectioned specimens<sup>24,25</sup>.

Here, we introduce CODA: a method for effective 3D reconstruction of large tissues from serially sectioned H&E images. To demonstrate CODA's use in microanatomical research, we explore 3D modes of pancreas tumorigenesis. We analyzed 4,114 H&E sections to reconstruct 13 samples of up to 3.5 cm<sup>3</sup> comprising normal, precancerous and cancerous human pancreas at subcellular resolution. With deep learning semantic segmentation, we label ten distinct cell and tissue types without incorporation of additional stains. The power of CODA is use through visualization of complex pancreatic ductal morphology; characterization of the extent, 3D structure and cellularity of pancreatic precursors; quantification of fiber alignment in a 3D landscape and exploration of structures used by pancreatic cancer to invade far from the bulk tumor.

We present a fully integrated pipeline for labeled, 3D reconstruction of serial tissue images at single cell resolution with detailed comparison to existing methodologies. While previous techniques exist for registration<sup>26,27</sup>, cell detection<sup>29,30</sup> and tissue multilabeling in H&E images<sup>31,32</sup>, we show that our integrated approach allows rapid, consistent reconstruction of serial samples from organs such as pancreas, skin, lung and liver.

## Results

### CODA: 3D reconstruction of serial histological sections

To develop CODA, a method for the 3D reconstruction of serially sectioned tissue, we identified 14 human pancreas samples (designated samples P0–P13) containing normal pancreatic parenchyma, pancreatic parenchyma with precancerous lesions and untreated invasive pancreatic cancer, as detailed in Supplementary Table 1. Sample P0 contains 101 serial images sampled 4 µm apart and was used only to optimize the workflow. Thick formalin-fixed paraffin-embedded samples were sectioned, stained with H&E and digitized at ×20 magnification, providing *x* and *y* (lateral) resolution of 0.5 µm and *z* (axial) resolution of 4 µm (Fig. 1a).

First, the independent serial images were mapped to a common coordinate system using a new image registration approach (Fig. 1b). Images were coarsely aligned using whole field rigid-body registration, followed by an elastic registration approach to account for local tissue warping, similar to previously developed workflows<sup>24–26,33–35</sup>. Briefly, for a pair of images, radon transforms were calculated at discrete angles. The maximum of the cross-correlation of radon transforms of the images yielded registration angle, and the maximum of the cross-correlation of the rotated tissue images yielded translation. Elastic registration was obtained by interpolating a grid of rigid registrations calculated at intervals across the globally registered images.

This method serially aligned the 101 serial histological sections in P0 in 30 minutes (Extended Data Fig. 1, detailed processing time estimates in Supplementary Table 2). To limit accumulation of error due to imperfect tissue sectioning, our algorithm is designed to discard registration to badly deformed tissues (containing large regions of splitting or folding).

Next, we established a high throughput H&E cell detection workflow based on color deconvolution and normalization and a previously established algorithm for particle tracking<sup>36</sup>. The method begins with color deconvolution of the H&E images into H&E channels. Using the hematoxylin channel image, local two-dimensional (2D) intensity maxima are identified as nuclear coordinates. By quantifying nuclear coordinates instead of nuclear boundaries, our technique is capable of rapid cell detection in large serially sectioned samples without the need for training or manual annotations. CODA cell detection delivers a processing time of approximately 90 seconds per whole slide image (Fig. 1c).

We then established a deep learning workflow for semantic segmentation of histological features and used it to identify nine pancreatic cell and tissue components in H&E: normal ductal epithelium, pancreatic cancer precursors, PDAC, smooth muscle, acini, fat, collagen, islets of Langerhans and lymph nodes (Fig. 1d and Extended Data Fig. 2). The pipeline used DeepLab semantic segmentation and a pretrained ResNet50 network<sup>37</sup>, achieved class precision and recall of >90% per sample (Extended Data Fig. 3a), and labeled images to a resolution of 2 µm per pixel in under 3 minutes each (computer specifications in Supplementary Table 2). Our workflow allows segmentation of more pancreas tissues than previously developed 3D methods<sup>2,21,32,38</sup> and is amenable to rapid (roughly 1 day) generation of functional models. As sample collection was staggered, individual deep learning models were created for each sample. To demonstrate the ability of CODA to label additional structures in samples after creation of the first model, a second model was trained on all 13 samples to identify nerves in the pancreas, with precision and recall >90% (Extended Data Fig. 3b).

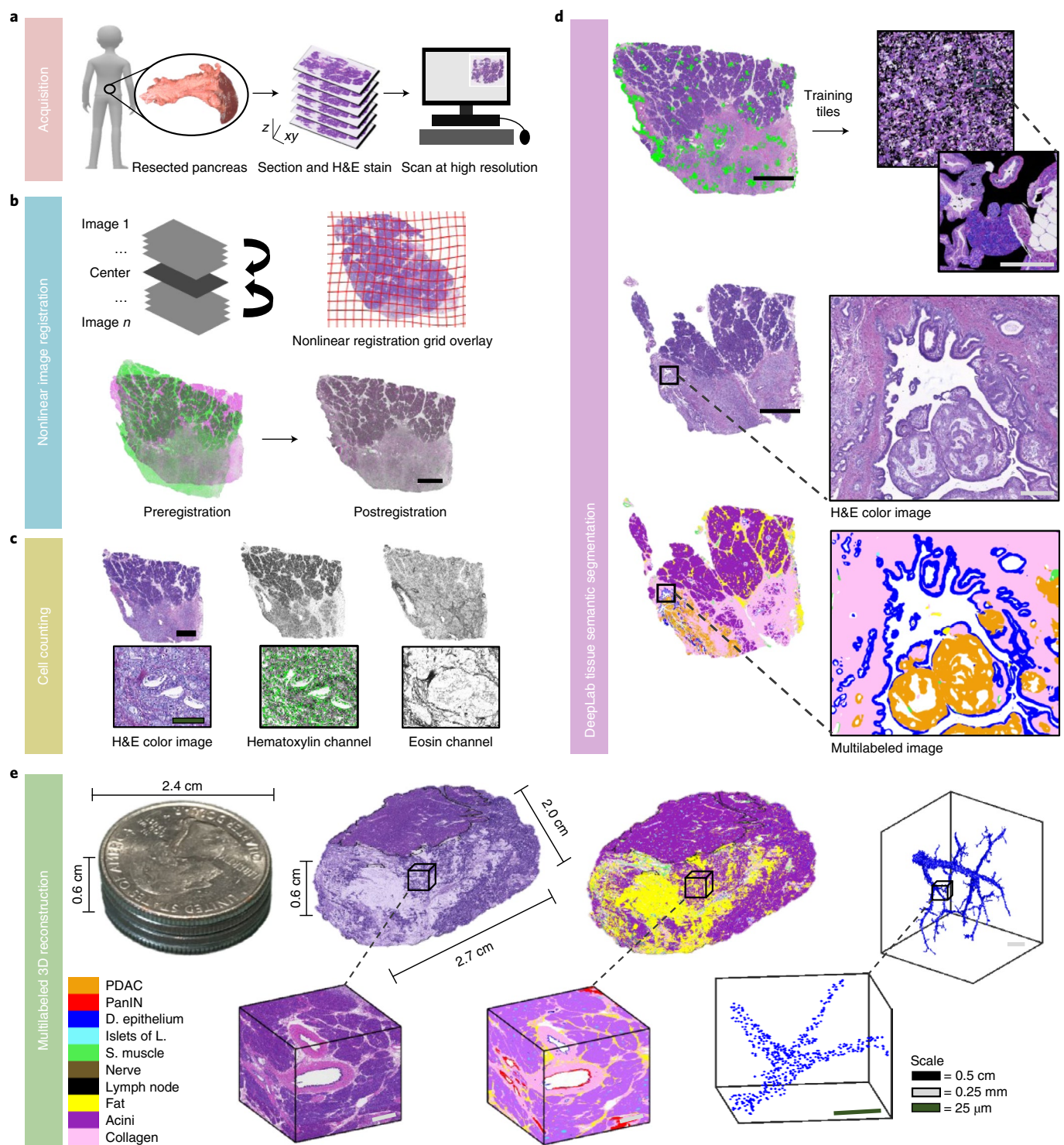
Altogether, CODA facilitates 3D reconstruction of labeled tissues at both tissue and cellular resolution (Fig. 1e).

### Validation of CODA methodology

We compared our registration approach to seven other methods using data in a previously published comparative analysis of tissue registration algorithms<sup>35</sup> and found that CODA registration outperforms the other techniques, particularly in two metrics: limiting the accumulation of error across large samples (accumulated target registration error, ATRE) and maintaining higher pixel correlation between images (root mean squared error (r.m.s.e.)) (Fig. 2a and Supplementary Methods).

To validate cell detection accuracy and compare to pre-existing techniques, five randomly selected 1.5 mm<sup>2</sup> image tiles were manually annotated by two researchers. Manual annotations were compared to CODA cell detection as well as two commonly used approaches<sup>29,30</sup>. CODA cell detection achieved the highest overall accuracy of the three techniques assessed with >90% precision and recall (Extended Data Fig. 4a), and for assessment of samples containing many serial samples, CODA cell detection was on average threefold faster than the other techniques (Extended Data Fig. 4a). In situ diameters of each cell type were measured and incorporated to extrapolate 3D cell counts from cell counts on serial 2D images (Extended Data Fig. 4b).

We additionally assessed the effect of reducing the *z* resolution of the samples by registering a subset of serial images. We found 95% similarity in registration performed with between consecutive sections or sections up to five axial planes apart (Fig. 2b). Further, we found that we maintained 96% accuracy in estimation of cell count and tissue content by interpolating 3D cell count and deep learning labels from sections

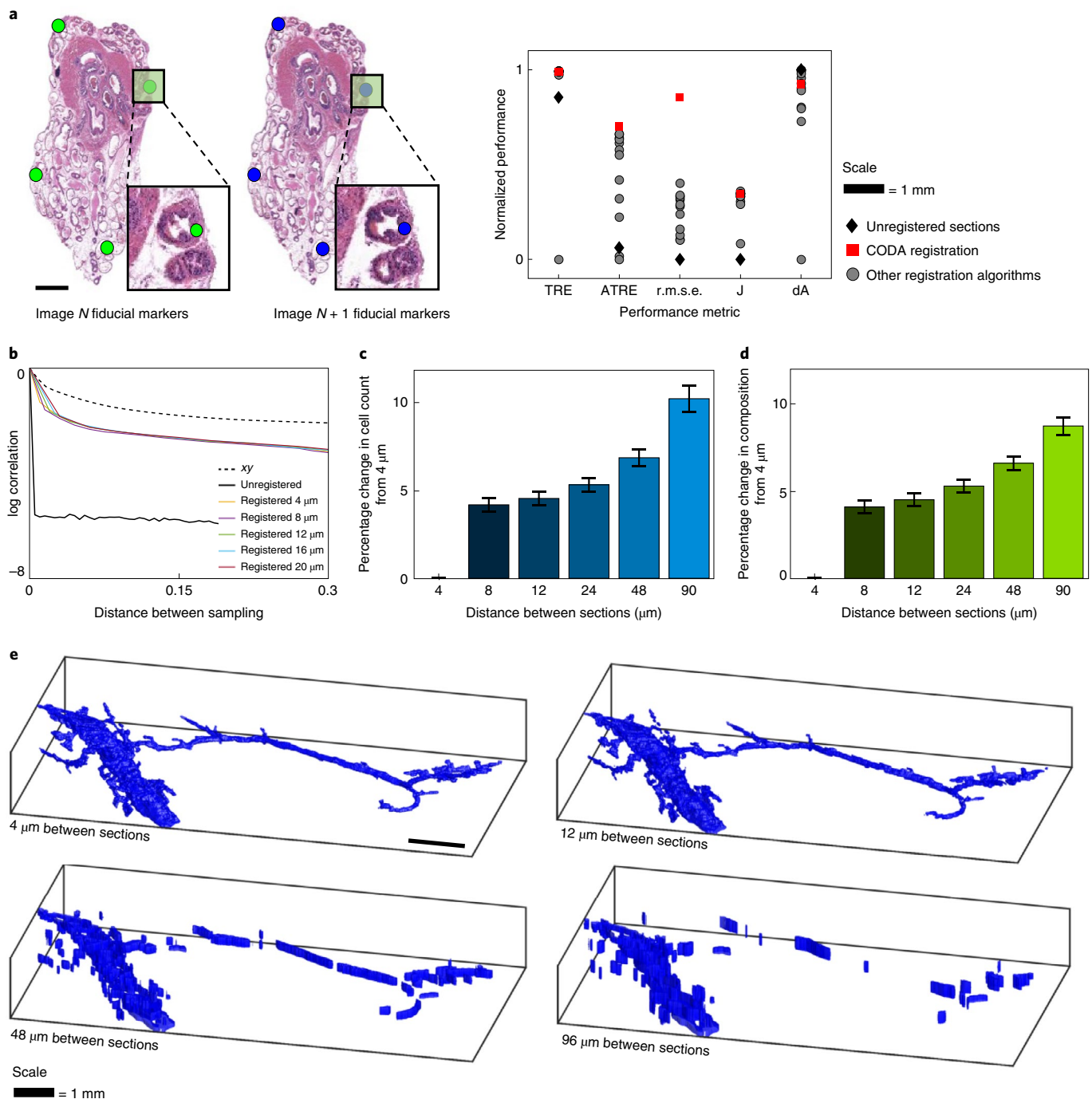


**Fig. 1 | CODA. a**, Human pancreatic tissue was serially sectioned, stained and scanned. **b**, Images were registered using a nonlinear approach to create a digital volume. **c**, Cells were identified using the hematoxylin channel of the H&E images. **d**, Deep learning semantic segmentation models were trained using

randomly overlaid annotations of tissue types. Images are labeled to a resolution of 2  $\mu$ m. **e**, 3D reconstruction of >1,000 serially sectioned pancreas sections. 3D renderings are created at the cm, mm and  $\mu$ m scale at tissue and single cell resolution.

up to three axial planes apart (Fig. 2c,d). This allowed us to improve workflow throughput by processing only one in three serial images in samples P1–P13 for an axial resolution of 12  $\mu$ m. We next confirmed the quality of 3D renderings by creating visualizations of a region of the pancreatic ductal architecture from sample P0 at  $z$  resolutions of 4, 12, 48 and 96  $\mu$ m (Fig. 2e).

Finally, to demonstrate the ability of CODA to reconstruct non-pancreatic structures, we assessed samples of human scalp (Fig. 3a), murine lung (Fig. 3b) and murine liver (Fig. 3c). CODA was used to label six structures in skin: hair follicles, sweat glands, oil glands, epidermis, vasculature and collagen; five structures in lung: bronchioles, alveoli, vasculature, cancer metastases and collagen; and four structures in



**Fig. 2 | Validation of CODA registration and ability to skip z sections.**

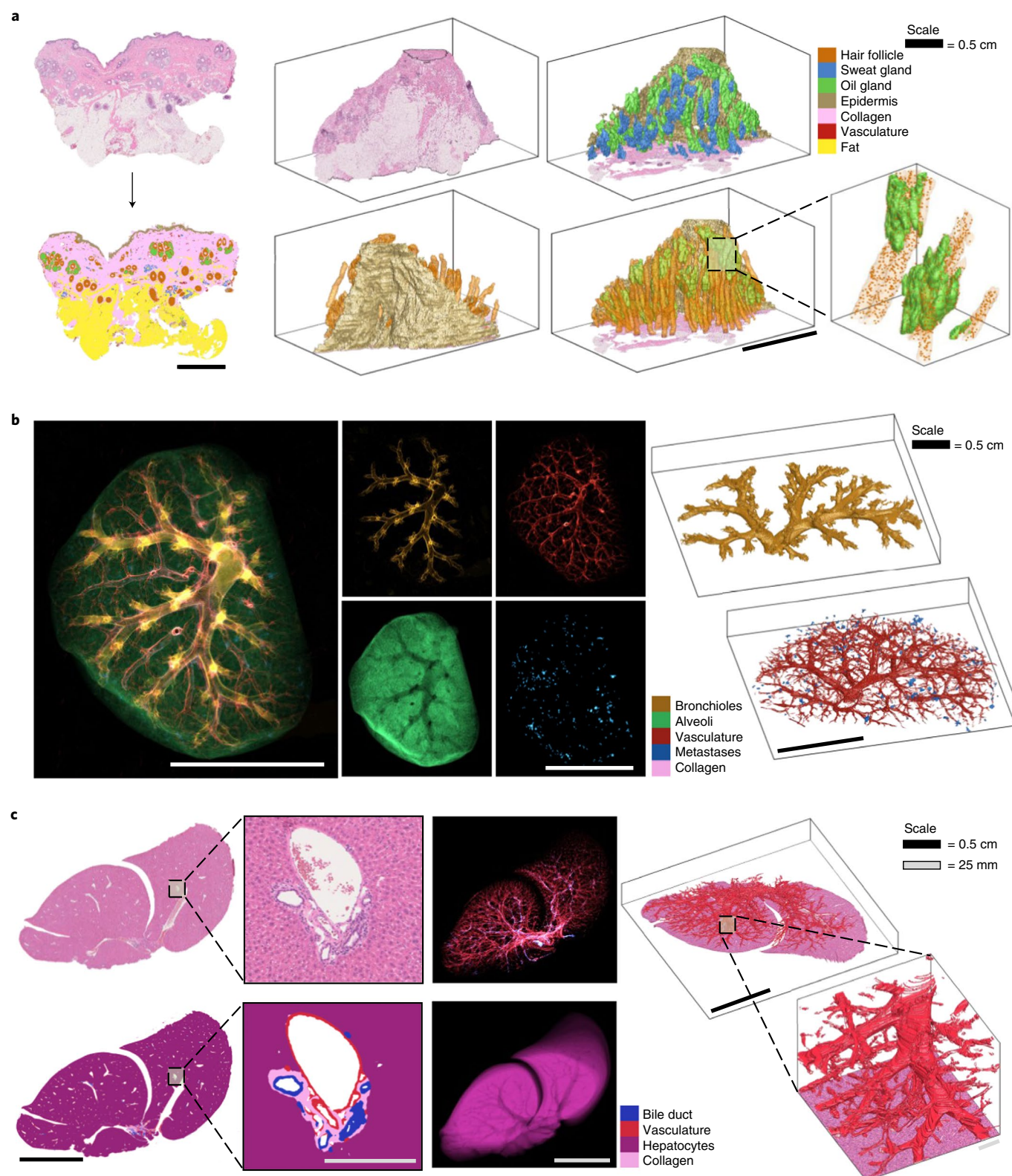
**a**, Sample validation image (from an online dataset first published in ref.<sup>34</sup>) with overlaid fiducial points. Inset shows close up view of a pair of fiducial markers on adjacent sections. Normalized performance metrics: TRE, ATRE, r.m.s.e., Jaccard Index (J) and pre-/postregistration change in area (dA). **b**, Quantification of loss in quality due to reducing the z resolution of serial samples. Calculation of pixel correlation across the z axis (left) shows that >95% correlation is

maintained postregistration when skipping up to four serial sections, or 20  $\mu$ m, between each H&E collected. **c,d**, Calculation of the percentage change in cell count (**c**) and tissue composition (**d**, right) reveals <5% error in 3D cell count and tissue composition extrapolation when skipping up to two serial sections, or 12  $\mu$ m, between each H&E collected. **e**, Validation of 3D rendering quality due to reducing the z resolution of serial samples. Tissues in this study are modeled using a spacing of 12  $\mu$ m between sections (top right rendering).

liver: bile duct, hepatocytes, vasculature and collagen, and to create detailed renderings of a range of tissue microanatomy. Our tissue image registration, cell detection and deep learning tissue labeling algorithms were applicable to the tissues with no changes to the method design. In validation, we found that our tissue labeling algorithm performed with >90% precision and recall for the skin, lung and liver tissues. Therefore, we show that CODA is a tissue agnostic and robust pipeline for 3D reconstruction and quantification of microanatomy.

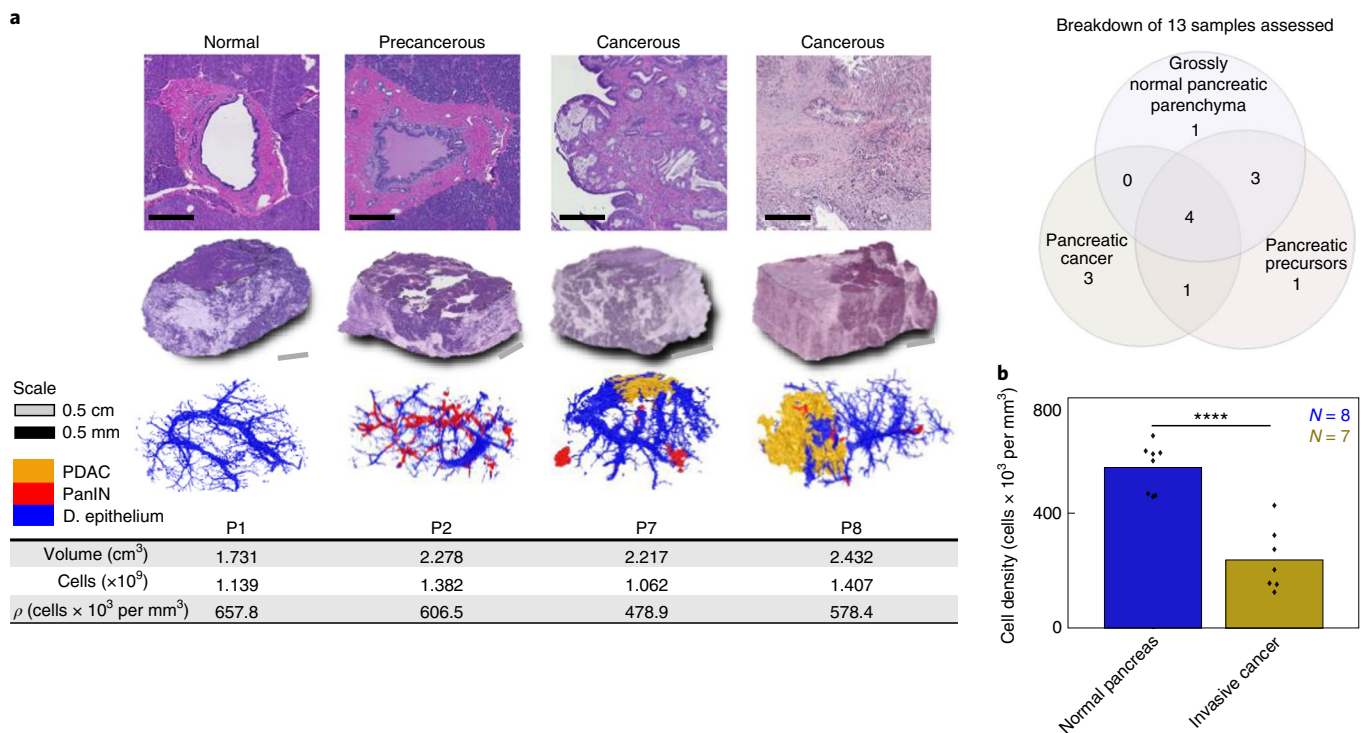
### Exploration of pancreas tumorigenesis in 3D

To demonstrate the use of CODA for biological research, we created multilabeled 3D maps of 13 resected pancreas tissue samples of volumes up to 3.5 cm<sup>3</sup> and containing up to 1.6 billion cells (Fig. 4a). Eight of the samples assessed contained regions of grossly normal pancreatic parenchyma (samples P1–P4, P6–P9), nine contained pancreatic precursor lesions (samples P2–P10) and eight contained regions of invasive pancreatic cancer (samples P6–P13). We created



**Fig. 3 | CODA processing of additional organs. a**, 3D reconstruction of human scalp tissue. Sample H&E and semantically segmented image (far left), visualization of the H&E volume (top left), epidermis, sweat glands and oil glands (top right), external (bottom left) and internal (bottom right) views of epidermis, hair follicles and oil glands, and visualization of single cell resolution (far right). **b**, 3D reconstruction of mouse lung tissue. Z projections of all components

together and individually (left) and 3D renderings of bronchioles (right top) and vasculature and metastases (right bottom). **c**, 3D reconstruction of mouse liver tissue. Sample H&E and semantically segmented image (far left), z projection of vasculature and bile duct (middle top) and hepatocytes (middle bottom) and 3D rendering of vasculature (far right).



**Fig. 4 | Interpatient pancreas analysis from cm scale to single cell resolution.** **a**, Thirteen samples of up to multi-cm scale containing normal, precancerous and cancerous human pancreas were reconstructed. Tissue volumes, cell counts and

cell densities were calculated. **b**, Bulk cell density decreases more than threefold in  $N = 7$  cancerous human pancreas relative to  $N = 8$  grossly normal human pancreas. \*\*\*\* indicates a  $P < 0.0001$  using the Wilcoxon rank sum test.

multiscale renderings that demonstrate the complex, curved architecture of the normal pancreatic ducts and periductal collagen, the surrounding acinar lobules, islets of Langerhans, fat and blood vessels (Supplementary Videos 1–6).

Through quantification of tissue volume and cell count, we investigated the compositional changes to the pancreas during tumorigenesis. We compared the volume and cell composition of tissue components in the samples (Supplementary Table 3). Our results revealed an average 2.3-fold decrease in cell density between healthy regions and invasive cancer regions in the pancreas with a  $P < 10^{-4}$  using the Wilcoxon rank sum test (Fig. 4b). At the extreme end, we found zero acinar tissue and an astonishing 87% collagen composition in sample P11, a 14-fold increase from normal pancreas architecture, emphasizing the scale of atrophy, dense desmoplastic stroma and tissue reorganization brought with pancreatic cancer.

### Microarchitectural properties of pancreatic precancers

Following bulk assessment of the samples, we used CODA to enumerate architectural patterns of pancreatic precursor lesions in 3D. Of the 13 samples analyzed, eight contained pancreatic intraepithelial neoplasia (PanIN) and one contained intraductal papillary mucinous neoplasms (IPMNs). PanIN are clinically defined as mucin-producing epithelial neoplasms residing in ducts  $< 0.5$  cm, with larger neoplasms typically denoted as IPMNs and both involve the complex tubular branches of the pancreatic ducts and ‘bunches of grape-like’ acinar lobules<sup>39</sup>. It is currently not possible to noninvasively detect the smallest of these lesions in the clinic<sup>40–42</sup>.

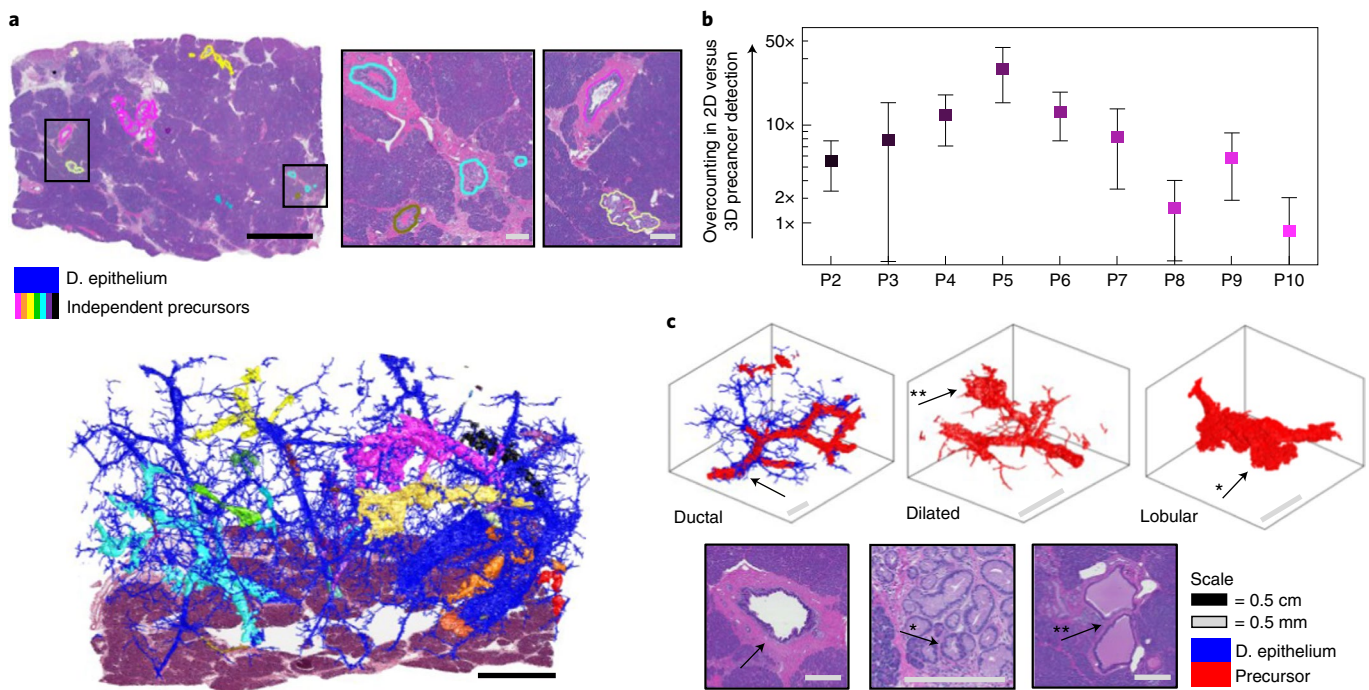
We found that precursors occupy a range of volumes, can be simple or highly branched and may be densely packed yet unconnected in 3D. Using the 3D reconstruction of the ductal system of sample P2, we identified 43 spatially independent precancers in a 2.3 cm<sup>3</sup> sample (Fig. 5a and Supplementary Video 7). In one section, a large precursor was identified in multiple ducts separated by nearly 1 cm and surrounded

by multiple, smaller precursors exemplifying how connectivity is difficult to interpret from 2D alone. In the nine samples containing precursors (samples P2–P10), we compared the number of distinct precursors per section with and without considering 3D connectivity and found that 2D lesion number over-counted the true 3D tumor number per section by as much as a factor of 40, exemplifying the complex 3D connectivity of pancreatic precancers (Fig. 5b). This measurement yielded an average 12.3-fold overcounting in 2D versus 3D with a  $P$  value of  $< 10^{-5}$  using a Wilcoxon rank sum test (Fig. 5b).

While assessing 3D connectivity of the precursors, we identified three distinct 3D structural phenotypes that we term tubular, lobular and dilated (Fig. 5c and Supplementary Video 8). Tubular precancers appeared as ductal, branching structures, dilated precancers appeared as large ballooning of the duct connected to ducts of much smaller diameters and lobular precancers appeared as ‘bunches of grapes’-like connected locules forming a nodule. Review of the corresponding H&E sections by pancreatic pathologists revealed that tubular PanINs resided within pancreatic ducts, dilated PanINs resided within regions of dilated pancreatic ducts and lobular PanINs resided at the terminal junctions of ducts and acinar lobules, involving areas of acinar to ductal metaplasia<sup>43,44</sup>. These phenotypes appear similar to pancreatic precancer phenotypes identified in mice<sup>45</sup>. Notably, 174 of the 265 identified precursors (66%) contained both ductal and lobular morphology, suggesting that extension of precursors between dilated/nondilated pancreatic ducts and acinar lobules is a relatively common occurrence (Extended Data Fig. 2d).

### The role of fiber alignment in pancreatic cancer invasion

Next, we investigated the morphology of invasive pancreatic cancer and the tumor microenvironment in eight large samples. We first focused on the morphology of PDAC at the interface of invasive cancer and adjacent normal tissue in sample P7 to identify patterns of invasion, then enumerated the occurrence of these patterns in all tissues.



**Fig. 5 | Microarchitectural patterns in pancreatic precancers. a**, 43 Spatially independent precancers in sample P2 were color coded and labeled on H&E serial sections (upper panel) and a 3D reconstruction (lower panel). **b**, Number of precancers per 2D section normalized by true 3D precancer number was

calculated for samples containing precancers. **c**, 3D renderings and sample histology illustrate three 3D phenotypes of PanIN observed. Tubular PanIN preserve normal pancreatic ductal morphology, lobular PanIN resemble acinar lobules and dilated PanIN reside in dilated ducts or lobules.

The mass consisted of a region of invasive carcinoma with three prominent protrusions extending into surrounding normal pancreatic tissue (Fig. 6a). The first of these protrusions was invasive cancer extending within the lumen of a vein for at least 4 mm. The second of the protrusions was a >3 mm region of cancer extension along periductal stroma. The third protrusion was a >1 mm focus of perineural invasion. We quantified the occurrence of these phenomena in all eight samples, revealing that all samples (100%) contained regions of venous invasion, seven (87%) contained perineural or neural invasion and five (63%) contained invasion along periductal, perivascular or perilobular stroma. As CODA allows confirmation of 3D findings in high-resolution H&E images, all foci of invasion were further validated via examination of the histology (Extended Data Figs. 5–7).

Finally, we investigated 3D stromal properties at the pancreatic vasculature, ducts and nerves. The alignment of collagen fibers in histological samples of PDAC has been negatively correlated with prognosis<sup>46–48</sup>. However, in previous work using 2D samples of many patients, collagen alignment in the ductal submucosa of normal pancreatic ducts was reported to be low<sup>47,49</sup>. We sought to repeat this measurement to account for the angle of sectioning of the ducts.

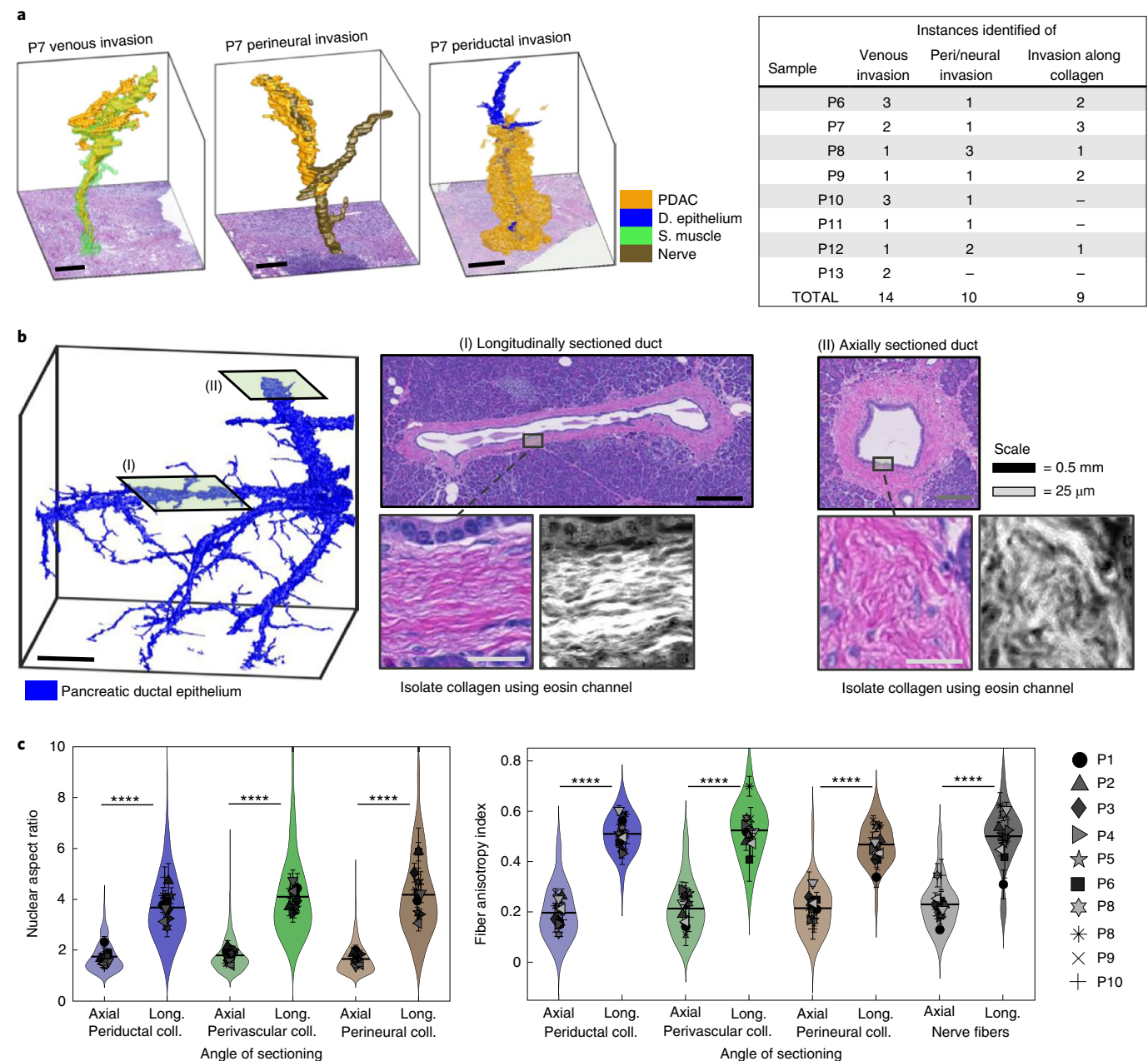
Using our 3D renderings, we identified coordinates where the ducts, blood vessels and nerves were cut at two extremes: perpendicular to the long axis of the structure (axially sectioned) and parallel to the long axis of the structure (longitudinally sectioned) and isolated these regions in H&E (Fig. 6b) to quantify collagen fiber alignment. Our measurements of fiber alignment therefore account for the varying appearance of fibers relative to their orientation to the sectioning blade, allowing more accurate calculation than can be computed from the random plane in a 2D histological section alone. As validation of our fiber alignment measure, we compare our results to measures of alignment of nerve fibers, which are known to be highly aligned in the longitudinal direction<sup>50–52</sup>.

Quantification<sup>53</sup> revealed significantly higher (using the Wilcoxon rank sum test) collagen and nerve fiber alignment and nuclear

aspect ratio in longitudinally compared to axially sectioned structures (Fig. 6c). For nuclear aspect ratio, two independent researchers measure a 2.1-, 2.3- and 2.5-fold change between longitudinally and axially sectioned images for periductal, perivascular and perineural collagen, respectively (all  $P$  values  $<10^{-5}$ ; researchers' measurements are compared in Extended Data Fig. 2e). For fiber alignment, we measure a 2.5-, 2.4-, 2.2- and 2.2-fold change between longitudinally and axially sectioned images for periductal collagen, perivascular collagen, perineural collagen and nerve fibers, respectively (all  $P < 10^{-5}$ ). Contrary to previous studies in 2D, these results indicate that collagen fibers are highly aligned along the longitudinal direction of structures they surround, including ducts, blood vessels and nerves. This is the same direction of alignment as that of the observed cancer protrusions. CODA enables quantification of metrics such as 3D fibrillar alignment that are imprecise when measured from isolated 2D sections.

## Discussion

Here we show that CODA is a powerful complement to tissue clearing and current serial sectioning techniques used to study 3D tissue microarchitecture. Tissue clearing is the most popular current approach to study 3D tissues, wherein intact samples are rendered semitransparent, labeled and imaged using confocal or light-sheet microscopy and have been used to conduct landmark scientific research<sup>14–19,54–56</sup>. However, long wait times of days to weeks between protocol steps, inconsistent antibody penetration, limits on the size of tissues that can be cleared, the number of labels that can be used and longstanding complications in quantification of the rendered 3D datasets represent key challenges in clearing research<sup>22</sup>. Current serial sectioning methods bypass some of the shortcomings of tissue clearing, albeit through introduction of new challenges. The sectioning of tissue causes unpredictable warping, requiring sophisticated registration techniques. Additionally, many serial sectioning methods rely on expensive techniques for labeling including IHC labeling, mass spectrometry and manual annotation<sup>24,25,34</sup>. Although expensive, these 2D labels are easier to



**Fig. 6 | 3D Patterns in pancreatic cancer invasion.** **a**, Occurrence of venous invasion, (peri)neural invasion and invasion along collagen fibers identified in eight samples containing PDAC. Selected 3D reconstructions of pancreatic cancer invasion patterns: invasion along periductal collagen, venous invasion and perineural invasion. **b**, 3D reconstruction of normal ductal epithelium with identified coordinates of (I) longitudinal and (II) axial sectioning. H&E

images extracted from coordinates and eosin channel isolated. **c**, Nuclear aspect ratio and fiber anisotropy index, representing local fiber alignment, of 90 longitudinally (long.) and 90 axially sectioned ducts, blood vessels and nerves from ten patient samples. Nuclear elongation and fiber alignment were significantly higher in longitudinal compared to axial sections. \*\*\*\* indicates  $P < 0.0001$  using the Wilcoxon rank sum test.

quantify than 3D data generated by clearing techniques, as they can take advantage of a plethora of previously developed 2D computational approaches<sup>57–59</sup>.

CODA is a powerful tool that has potential to integrate many current tissue imaging techniques. It incorporates nonlinear image registration and deep learning techniques to create multilabeled tissue volumes using H&E images, which is a relatively inexpensive histological technique. As our results demonstrate that CODA can derive quality 3D reconstructions while skipping at least two intervening sections, future addition of IHC labeling, spatial ‘omics’ and gene expression imaging to the intervening sections will increase the number of labels beyond what is currently achievable. The number of tissue and

molecular phenotypes that CODA can label in the pancreas, skin, liver and lungs has the potential to unlock previously unknown insights into human tissue, health and disease.

In our analysis of pancreatic tumor progression, we identify several findings that are both new and only possible through 3D analysis. We find that many anatomically separate precursor lesions can develop in small or large ducts, and that individual precursors are commonly present both in the pancreatic ducts and in foci of acinar to ductal metaplasia in the acinar lobules. We find invasive cancer cells extending from the central tumor along existing structures such as veins, nerves and peri-ductal, vascular, lobular and neural collagen. Together with previous work, which found highly aligned collagen to be a negative

prognostic factor in PDAC<sup>48</sup>, our identification of cancer cells protruding along aligned fibers suggests that pancreatic cancer cells in situ may invade more easily in regions of aligned collagen and nerve fibers. Overall, there is a need in cancer research for 3D reconstruction techniques that enable the collection of large, quantifiable tissue datasets. We demonstrate that CODA is one such powerful technique.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-022-01650-9>.

## References

- Liebig, C., Ayala, G., Wilks, J. A., Berger, D. H. & Albo, D. Perineural invasion in cancer. *Cancer* **115**, 3379–3391 (2009).
- Hong, S. M. et al. Three-dimensional visualization of cleared human pancreas cancer reveals that sustained epithelial-to-mesenchymal transition is not required for venous invasion. *Mod. Pathol.* **33**, 639–647 (2019).
- Kuett, L. et al. Three-dimensional imaging mass cytometry for highly multiplexed molecular and cellular mapping of tissues and the tumor microenvironment. *Nat. Cancer* **3**, 122–133 (2021).
- Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2021. *CA Cancer J. Clin.* **71**, 7–33 (2021).
- Michaud, D. S. et al. Physical activity, obesity, height, and the risk of pancreatic cancer. *JAMA* **286**, 921–929 (2001).
- Hruban, R. H. et al. Why is pancreatic cancer so deadly? The pathologist's view. *J. Pathol.* **248**, 131–141 (2019).
- Tanaka, M. et al. Meta-analysis of recurrence pattern after resection for pancreatic cancer. *Br. J. Surg.* **106**, 1590–1601 (2019).
- Zhang, J.-F. et al. Influence of perineural invasion on survival and recurrence in patients with resected pancreatic cancer. *Asian Pac. J. Cancer Prev.* **14**, 5133–5139 (2013).
- Huang, L. et al. Ductal pancreatic cancer modeling and drug screening using human pluripotent stem cell- and patient-derived tumor organoids. *Nat. Med.* **21**, 1364–1371 (2015).
- Drost, J. & Clevers, H. Organoids in cancer research. *Nat. Rev. Cancer* **18**, 407–418 (2018).
- Taniuchi, K. et al. Overexpressed P-Cadherin/CDH3 promotes motility of pancreatic cancer cells by interacting with p120ctn and activating Rho-Family GTPases. *Cancer Res.* **65**, 3092–3099 (2005).
- Plentz, R. et al. Inhibition of  $\gamma$ -secretase activity inhibits tumor progression in a mouse model of pancreatic ductal adenocarcinoma. *Gastroenterology* **136**, 1741–1749.e6 (2009).
- Cruz-Monserrate, Z. et al. Detection of pancreatic cancer tumours and precursor lesions by cathepsin E activity in mouse models. *Gut* **61**, 1315–1322 (2012).
- Yang, B. et al. Single-cell phenotyping within transparent intact tissue through whole-body clearing. *Cell* **158**, 945–958 (2014).
- Murakami, T. C. et al. A three-dimensional single-cell-resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing. *Nat. Neurosci.* **21**, 625–637 (2018).
- Susaki, E. A. et al. Versatile whole-organ/body staining and imaging based on electrolyte-gel properties of biological tissues. *Nat. Commun.* **11**, 1982 (2020).
- Zhao, S. et al. Cellular and molecular probing of intact human organs. *Cell* **180**, 796–812.e19 (2020).
- Chung, K. et al. Structural and molecular interrogation of intact biological systems. *Nature* **497**, 332–337 (2013).
- Richardson, D. S. & Lichtman, J. W. Clarifying tissue clearing. *Cell* **162**, 246–257 (2015).
- Xie, W. et al. Prostate cancer risk stratification via nondestructive 3D pathology with deep learning–assisted gland analysis. *Cancer Res.* **82**, 334–345 (2022).
- Hahn, M. et al. Mesoscopic 3D imaging of pancreatic cancer and Langerhans islets based on tissue autofluorescence. *Sci. Rep.* **10**, 18246 (2020).
- Liu, J. T. C. et al. Harnessing non-destructive 3D pathology. *Nat. Biomed. Eng.* **5**, 203–218 (2021).
- Groot, A. Ede et al. Characterization of tumor-associated macrophages in prostate cancer transgenic mouse models. *Prostate* **81**, 629–647 (2021).
- Song, Y., Treanor, D., Bulpitt, A. J. & Magee, D. R. 3D reconstruction of multiple stained histology images. *J. Pathol. Inform.* **4**, 7 (2013).
- Lotz, J. M. et al. Integration of 3D multimodal imaging data of a head and neck cancer and advanced feature recognition. *Biochim. Biophys. Acta: Proteins Proteom.* **1865**, 946–956 (2017).
- Lotz, J. et al. Zooming in: high resolution 3D reconstruction of differently stained histological whole slide images. *Proc. SPIE* **9041**, 16–22 (2014).
- Tempest, N. et al. Histological 3D reconstruction and in vivo lineage tracing of the human endometrium. *J. Pathol.* **251**, 440–451 (2020).
- Rees, J. et al. O36 Investigating clonal expansions in the normal stomach and the 3D architecture of oxyntic gastric glands. *Gut* **70**, A20–A21 (2021).
- Graham, S. et al. Hover-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med. Image Anal.* **58**, 101563 (2019).
- Bankhead, P. et al. QuPath: open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).
- Chan, L. et al. HistoSegNet: semantic segmentation of histological tissue type in whole slide images, in *Proc. International Conference on Computer Vision (ICCV) 2019, Seoul, Korea 10662–10671 (ICCV, 2019)*.
- Ternes, L. et al. VISTA: visual semantic tissue analysis for pancreatic disease quantification in murine cohorts. *Sci. Rep.* **10**, 20904 (2020).
- Magee, D. et al. Histopathology in 3D: from three-dimensional reconstruction to multi-stain and multi-modal analysis. *J. Pathol. Inform.* **6**, 6 (2015).
- Roberts, N. et al. Toward routine use of 3D histopathology as a research tool. *Am. J. Pathol.* **180**, 1835–1842 (2012).
- Kartasalo, K. et al. Comparative analysis of tissue reconstruction algorithms for 3D histology. *Bioinformatics* **34**, 3013 (2018).
- Wu, P. H. et al. High-throughput ballistic injection nanorheology to measure cell mechanics. *Nat. Protoc.* **7**, 155–170 (2012).
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F. & Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation, in *Proc. European Conference on Computer Vision, ECCV 2018 (eds Ferrari, V. et al.) 883–851 (Springer, 2018)*.
- Yoshizawa, T. et al. Three-dimensional analysis of extrahepatic cholangiocarcinoma and tumor budding. *J. Pathol.* **251**, 400–410 (2020).
- Basturk, O. et al. A revised classification system and recommendations from the baltimore consensus meeting for neoplastic precursor lesions in the pancreas. *Am. J. Surg. Pathol.* **39**, 1730 (2015).
- Singhi, A. D., Koay, E. J., Chari, S. T. & Maitra, A. Early detection of pancreatic cancer: opportunities and challenges. *Gastroenterology* **156**, 2024–2040 (2019).
- Hruban, R. H., Maitra, A. & Goggins, M. Update on pancreatic intraepithelial neoplasia. *Int. J. Clin. Exp. Pathol.* **1**, 306 (2008).

42. Canto, M. I. et al. Screening for early pancreatic neoplasia in high-risk individuals: a prospective controlled study. *Clin. Gastroenterol. Hepatol.* **4**, 766–781 (2006).
43. Zhu, L., Shi, G., Schmidt, C. M., Hruban, R. H. & Konieczny, S. F. Acinar cells contribute to the molecular heterogeneity of pancreatic intraepithelial neoplasia. *Am. J. Pathol.* **171**, 263–273 (2007).
44. Morris, J. P. IV, Cano, D. A., Sekine, S., Wang, S. C. & Hebrok, M.  $\beta$ -catenin blocks Kras-dependent reprogramming of acini into pancreatic cancer precursor lesions in mice. *J. Clin. Invest.* **120**, 508–520 (2010).
45. Messal, H. A. et al. Tissue curvature and apicobasal mechanical tension imbalance instruct cancer morphogenesis. *Nature* **566**, 126–130 (2019).
46. Xu, S. et al. The role of collagen in cancer: from bench to bedside. *J. Transl. Med.* **17**, 309 (2019).
47. Puls, T. J., Tan, X., Whittington, C. F. & Voytik-Harbin, S. L. 3D collagen fibrillar microstructure guides pancreatic cancer cell phenotype and serves as a critical design parameter for phenotypic models of EMT. *PLoS ONE* **12**, e0188870 (2017).
48. Drifka, C. R. et al. Highly aligned stromal collagen is a negative prognostic factor following pancreatic ductal adenocarcinoma resection. *Oncotarget* **7**, 76197 (2016).
49. Drifka, C. R. et al. Periductal stromal collagen topology of pancreatic ductal adenocarcinoma differs from that of normal and chronic pancreatitis. *Mod. Pathol.* **28**, 1470–1480 (2015).
50. Sunderland, S. S. The anatomy and physiology of nerve injury. *Muscle Nerve* **13**, 771–784 (1990).
51. Lundborg, G. & Dahlin, L. B. Anatomy, function, and pathophysiology of peripheral nerves and nerve compression. *Hand Clin.* **12**, 185–193 (1996).
52. Axer, H., Axerl, M., Krings, T. & Keyserlingk, D. G. V. Quantitative estimation of 3-D fiber course in gross histological sections of the human brain using polarized light. *J. Neurosci. Methods* **105**, 121–131 (2001).
53. Fraley, S. I. et al. Three-dimensional matrix fiber alignment modulates cell migration and MT1-MMP utility by spatially and temporally directing protrusions. *Sci. Rep.* **5**, 14580 (2015).
54. Rios, A. C. et al. Intracolon plasticity in mammary tumors revealed through large-scale single-cell resolution 3D imaging. *Cancer Cell* **35**, 618–632.e6 (2019).
55. Cuccarese, M. F. et al. Heterogeneity of macrophage infiltration and therapeutic response in lung carcinoma revealed by 3D organ imaging. *Nat. Commun.* **8**, 14293 (2017).
56. Lai, H. M. et al. Antibody stabilization for thermally accelerated deep immunostaining. *Nat. Methods* <https://doi.org/10.1038/s41592-022-01569-1> (2022).
57. Saltz, J. et al. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.* **23**, 181–193.e7 (2018).
58. Lehmann, B. D. et al. Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. *PLoS ONE* **11**, e0157368 (2016).
59. Nirschl, J. J. et al. A deep-learning classifier identifies patients with clinical heart failure using whole-slide images of H&E tissue. *PLoS ONE* **13**, e0192726 (2018).

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

## Methods

### Tissue acquisition and scanning

Formalin-fixed, paraffin-embedded samples were sectioned every 4  $\mu\text{m}$ . Every third tissue section was stained using H&E, with two sections every three held out. All tissues of sample P0 were scanned for validation that skipping two sections maintained registration and reconstruction accuracy. Tissues were scanned at  $\times 20$  using a Hamamatsu Nanozoomer. These studies were approved by the Institutional Review Board of The Johns Hopkins Hospital.

### Image registration

Cases contained series of tissue images scanned at  $\times 20$ , corresponding to approximately 0.5  $\mu\text{m}$  per pixel. Openslide software was used to save reduced size copies of each image, corresponding to 8  $\mu\text{m}$  per pixel using nearest neighbor interpolation<sup>60</sup>. For each sample, the center image was identified as the point of reference ( $\text{image}_n$ ) and global and elastic registration was calculated for all other images in the sample.

We performed registration on greyscale, Gaussian-filtered, downsampled (80  $\mu\text{m}$  per pixel resolution) versions of the high-resolution histological sections. Global registration transformations for a pair of preprocessed tissue images were found through iterative calculation of registration angle and translation via maximization of cross-correlation. Radon transforms of the images taken at discrete angles between 0 and  $359^\circ$  were calculated. The maximum of the cross-correlation of radon transforms of the images yielded registration angle, and the maximum of the cross-correlation of the rotated tissue images yielded translation. Elastic registration was obtained by calculating rigid registration of cropped image tiles at 1.5-mm intervals across the globally registered images at 8  $\mu\text{m}$  per pixel resolution. The resulting local, rigid registration fields were interpolated to the size of the 8  $\mu\text{m}$  per pixel resolution images. Finally, the registration fields were smoothed using a Gaussian filter with standard deviation of 2 pixels to produce a nonlinear, elastic registration transformation.

To account for images with large regions of tissue splitting or folding, rigid global registration was performed to sequentially register each image  $\text{image}_{n\pm m}$  to the three next closest images to center, image  $\text{image}_{n\pm(m+1)}$ , image  $\text{image}_{n\pm(m+2)}$  and image  $\text{image}_{n\pm(m+3)}$ . Quality of each of the three global registrations was assessed by comparing pixel-to-pixel correlation between the moving and each reference image. The registration with the best result was kept and the other two discarded. Thus, if image  $\text{image}_{n\pm(m+1)}$  contained large defects such as tissue splitting or folding, then image  $\text{image}_{n\pm(m+2)}$  would be used as the reference for rigid registration to avoid compound errors. Following global registration, elastic registration was used between the moving image and chosen reference image to create a nonlinear displacement map. This process was repeated for all images in a sample such that all images were elastically registered to the coordinate system of the center image  $\text{image}_n$ .

### Assessment of image registration quality

Quality of image registration within the pancreas image datasets was calculated using pixel-wise Spearman correlation. 'True' biological pixel variation was calculated by correlating pixel intensity along the  $x$  and  $y$  dimensions of single images (longitudinal correlation). It was assumed that 'perfect' registration would result in a similar  $z$  direction (down the image stack) correlation to the  $xy$  correlation, as the  $xy$  correlation represents the variation in pixel intensity in intact tissue. Axial pixel correlation was calculated by correlating pixel intensity along the  $z$  dimension of serial images. Unregistered  $z$  correlation was compared to postglobal registration correlation and postelastic registration correlation to determine improvements to intensity continuity following registration, and postelastic registration was compared to longitudinal correlation to determine how closely our registration results could emulate the true intensity variation between connected tissue.

For each correlation calculation (along the  $xy$  direction, unregistered  $z$  dimension, global registered  $z$  dimension and elastic registered

$z$  dimension) Spearman correlation was calculated for pixels at 4  $\mu\text{m}$  intervals starting at 0  $\mu\text{m}$  apart. Correlation of pixels 0  $\mu\text{m}$  apart is correlation of each pixel to itself (equal to 1). Correlation of pixels 4  $\mu\text{m}$  corresponds to two pixels 4  $\mu\text{m}$  apart in a single image (for the  $xy$  calculation) or one image apart (for the  $z$  calculation). This process was repeated for distances up to 0.3 mm. Additionally, this process was repeated for registration of all images in sample P0, and registration of one in two, one in three, one in four and one in five images in P0 to prove that we maintain >95% correlation when sampling one in every three images per tissue sample.

### Comparison of image registration to existing techniques

CODA registration was applied to a publicly available serial histological sample of 260 mouse prostate images, which was part of a previously published paper comparing the performance of seven registration techniques<sup>35</sup>. The image dataset contained manual annotation of two cells per image each from two different researchers. Performance metrics included pairwise target registration error (TRE), average distance between pairs of fiducial markers; ATRE, estimation of accumulated distortion throughout the stack; r.m.s.e., pairwise comparison of pixel intensities across the stack; Jaccard Index ( $J$ ), pairwise area overlap of consecutive images and dA, change in area of the tissue slides pre- and postregistration. CODA performed similarly to competing techniques in TRE,  $J$  and dA. CODA outperformed all other techniques in ATRE and r.m.s.e., suggesting CODA registration of this sample resulted in less accumulated error than other techniques.

Raw performance metrics are listed in the source data file for Fig. 2a. As the magnitude of various performance metrics varied widely, normalized performance metrics were calculated such that a single graph could concisely express a wide variety of performance parameters. Mean performance metrics were normalized using the following formulas such that they lay within the range of 0–1 and such that higher numbers indicate better performance:

$$\text{TRE}_{\text{Normalized}} = 1 - \text{TRE}_{\text{mean}} / \max(\text{TRE}_{\text{mean}})$$

$$\text{ATRE}_{\text{Normalized}} = 1 - \text{ATRE}_{\text{mean}} / \max(\text{ATRE}_{\text{mean}})$$

$$\text{r.m.s.e.}_{\text{Normalized}} = 1 - \text{r.m.s.e.}_{\text{mean}} / \max(\text{r.m.s.e.}_{\text{mean}})$$

$$\text{r.m.s.e.}_{\text{Normalized}} = 1 - \text{r.m.s.e.}_{\text{mean}} / \max(\text{r.m.s.e.}_{\text{mean}})$$

$$J_{\text{Normalized}} = J_{\text{mean}} / \min(J_{\text{mean}}) - 1$$

$$\text{dA}_{\text{Normalized}} = 1 - |\text{dA}_{\text{mean}}| / \max(|\text{dA}_{\text{mean}}|)$$

(where  $|x|$  denotes absolute value of  $x$ )

### Identification of cells in histological samples

First, the hematoxylin channel of all H&E images was extracted using color deconvolution. Openslide software was used to save reduced size copies of all tissue images, corresponding to 2  $\mu\text{m}$  per pixel using nearest neighbor interpolation<sup>60</sup>. For each image, the tissue region of the image was identified by finding regions of the image with low green channel intensity and high red-green-blue (RGB) standard deviation. Next, RGB channels were converted to optical density. Using kmeans clustering analysis, 100 clusters were identified to represent the optical densities of the image. The most common, blue-favored optical density was chosen to represent the hematoxylin channel and the most common, red-favored optical density was chosen to represent the eosin channel. The background optical density was fixed as the inverse of the average of the H&E optical densities. These three optical densities were used to deconvolve the RGB image into hematoxylin,

eosin and background channel images. Using methods described in ref.<sup>36</sup> the hematoxylin channel images were smoothed, and 2D intensity minima of a designated size and distance from each other were identified as nuclei.

### Validation of cell detection algorithm and comparison to existing techniques

A total of five 1.5 mm<sup>2</sup> regions were randomly extracted from the serial images for validation. For each region, two researchers manually annotated cells using an annotation function built in MATLAB 2021b. Next, CODA cell detection and two popular cell detection algorithms (Hovernet and QuPath)<sup>29,30</sup> were applied to the validation images to automatically generate nuclear coordinates. Automatically generated coordinates were termed true positives if they were within 2 µm of a manually generated coordinate (that was not already paired with another automatically generated coordinate). Then 2 µm was selected as the radius as this was determined to be the average radius of nuclei in the images. Automatically generated coordinates were termed false positives if they were not within 2 µm of a manually generated coordinate (that was not already paired with another automatically generated coordinate). Finally, manually annotated coordinates with no corresponding automatically generated coordinate were termed false negatives. From the true positives, false positives and false negatives, precision and recall were calculated to compare each of the three techniques to both sets of manual annotations.

### Deep learning tissue multilabeling

A deep learning model was created for each case using manual tissue annotations of that sample. Openslide software was used to save reduced size copies of all tissue images, corresponding to 2 µm per pixel using nearest neighbor interpolation<sup>60</sup>. Seven tissue images equally spaced within each sample were extracted. For each of the seven images, we manually annotated 50 examples of each identified tissue subtype using Aperio ImageScope, creating .xml files of annotation coordinates. Annotation coordinates were loaded into MATLAB 2021b using publicly available software and were downsampled to correctly overlay on the 2 µm per pixel tissue images<sup>61</sup>.

To reduce the heterogeneity of the H&E images, the H&E stains of all tissue images in each case were normalized. Using the H&E channel images created for the cell counting analysis and the optical density calculated for a reference H&E image from the same case, we reconstructed RGB images of each tissue type to a chosen optical density. Incorporation of image color normalization allowed us to avoid catastrophic failure of the semantic segmentation on unannotated images with drastically different staining patterns.

Bounding boxes of all annotations were identified and each annotated RGB image region was extracted and saved as a separate image file. A matrix was used to keep track of which bounding box images contained which annotation tissue types. Training images were built through creation of a 9,000 × 9,000 × 3, zero-value RGB image tile. Annotation bounding boxes containing the least represented deep learning class were randomly overlaid on the blank image tile until the tile was >65% full of annotations and such that the number of pixels of each deep learning class was approximately equal. Annotation bounding boxes were randomly augmented via rotation, scaling by a random factor between 0.8–1.2 and hue augmentation by a factor of 0.8–1.2 in each RGB color channel. The 9,000 × 9,000 × 3 image tile was then cut into 324,500 × 500 × 3 images. In total, 20 such large images were built, half with augmentation, to create 6,480 training images and five additional images were built to create 1,620 validation images. Then 324 testing images were created using manual annotations from an image not used for training or validation. This data generation pipeline including the size of the image tile, size of the training tiles and levels of data augmentation was chosen as it gave highest performance during pilot classification of sample PO.

Following dataset creation, a resnet50 network was adapted for DeepLab v.3+ semantic segmentation<sup>35</sup> and trained to a validation patience of 5. If 90% tissue subtype precision and recall was not obtained, additional manual annotations were added to the training and testing images and the process was repeated until desired accuracy was reached. We determined that >90% precision and recall resulted in classified models that generally matched pathological annotation of diseased tissues. Once a satisfactory deep learning model was trained, all tissue images in the sample were semantically segmented to create labeled tissue images with a pixel resolution of 2 µm per pixel.

### Addition of nerve labels to previously deep learning-labeled tissue images

The model design explained above was used to add nerves to the previously labeled pancreas histological images. First, 50 nerve annotations per image were collected on the images used for training of the previous deep learning model. Next, collagen, blood vessel and whitespace annotations from all previous annotation datasets were pooled. All other tissue components (islets, normal ductal epithelium, acini, precancers, cancer and lymph node) were pooled to a fifth class termed 'other tissue'. Collagen and blood vessel annotations were kept as separate classes as the eosin-rich staining on these structures closely resembles the staining pattern on nerves. It was found through training of a tri-class model (nerves, whitespace, other tissue only) that nerves would often be confused with collagen and vascular structures. The five annotation classes were pooled into training tiles as is described above and a semantic segmentation network with >90% precision and recall per class was trained across all 13 pancreas samples. It was calculated that >97% of pixels replaced by the nerve label were previously classified (using the semantic segmentation network that did not contain nerves as a label) as either collagen or vasculature. As this network classified both nerves and 'other tissue components', the nerve classification in this trained model was assumed to supersede the previous classification (thus all pixels labeled as nerves replaced the label for that pixel generated by the previous, ten-class model).

### 3D reconstruction of samples

Multilabeled images created by the deep learning portion of the CODA pipeline were consolidated into a 3D matrix using the H&E image registration results. Similarly, cellular coordinates counted on the unregistered histological sections were consolidated into a 3D cell matrix using the H&E image registration results. 3D renderings of the labeled tissue regions were visualized using the 'patch' and 'isosurface' commands in MATLAB 2021b and using a color scheme with a unique RGB triplet for each tissue subtype. Dimensions of rendered tissues were calculated in .xy using the pixel resolution of the original ×20 scanned histological sections (approximately 0.5 µm per pixel) and using the tissue section spacing (4 µm) in z. The resolution of the 3D renderings was 2 µm per pixel in .xy, the resolution used for image semantic segmentation and 12 µm per pixel in z, as only one in three tissue sections were used in the analysis. Single cells were visualized within the 3D renderings using the 'scatter3' command in MATLAB 2021b. For all calculations performed on the 3D labeled matrices of the tissues, the 3D matrix was subsampled using nearest neighbor interpolation from original voxel dimensions of 2 × 2 × 12 µm<sup>3</sup> per voxel to an isotropic 12 × 12 × 12 µm<sup>3</sup> per voxel.

### Calculation of tissue content, bulk cell density and local cell density

Tissue composition was calculated by counting the total number of voxels in the isotropic 3D matrix corresponding to each tissue subtype and dividing those numbers by the total number of voxels in the tissue region of the 3D matrix. These tissue compositions were formatted into a matrix in MATLAB 2021b and visualized as a heatmap.

Cell density of each tissue subtype was calculated by combining the tissue subtype data in the multilabeled 3D matrix with cell

coordinate data in the cell 3D matrix. Cells at each voxel in the cell 3D matrix corresponded to the tissue subtype label in the multilabeled 3D matrix (for example, a cell is labeled an epithelial cell if the nuclear coordinate was identified in a region labeled as epithelium by the deep learning pipeline). Measurements of nuclear diameter were used to estimate true 3D cell counts from the 2D cell coordinates. Using Aperio ImageScope, 100 nuclei of each tissue subtype were measured for each case. The estimated 3D cell count ( $C_{3D}$ ) of cells counted on serial sections analyzed every three sections was calculated using the formula:

$$C_{3D} = \sum_{\text{images}} \sum_{\text{subtypes}} C_{\text{image}} \frac{3T}{T + D_{\text{subtype}}}$$

where  $C_{\text{image}}$  is the cell count for a given tissue image,  $T$  is the thickness of the histological section and  $D_{\text{subtype}}$  is the measured diameter of a nucleus for a tissue subtype. For each tissue subtype, bulk 3D cell density was calculated by dividing the 3D extrapolated cell count of a particular subtype divided by the total volume of the tissue. Local 3D cell density was calculated by dividing the 3D extrapolated cell count of a particular subtype divided by the volume of that particular tissue subtype.

### Determination of spatially distinct precursor lesions

The 3D multilabeled matrices were used to determine tissue connectivity. Following classification, all objects labeled as pancreatic precancers lesions were visually verified to be precancers through creation of bounding box serial sections. Independent precursors were identified in the 3D multilabeled matrix using the 'bwlabeln' command in MATLAB 2021b. Bwlabeln identifies and labels spatially distinct objects in matrices. We calculated connectivity on both the precancers alone and the precancers plus the normal ductal epithelium. Distinct precancers and cancers identified using bwlabeln could then be quantitatively analyzed or 3D rendered independently from other precancers.

Independent precursor coordinates were used to automatically annotate connected lesions on H&E images of 2  $\mu\text{m}$  per pixel resolution. First, each precursor was assigned a distinct RGB color. Next, for each registered H&E image in the serial sections, the number of distinct precursors appearing on that section was determined. For each independent precursor on the section, voxels defining the precursor in the volume matrix were located. The pixels were dilated and only the outline kept, then rescaled to match the 2  $\mu\text{m}$  per pixel H&E images such that the annotated precursor mask was reformatted to appear as a thick outline overlaid on the precursor region of the H&E section. The outline was overlaid on H&E and the pixels in the H&E image corresponding to the outline were recolored to match the color defining that independent precancer. This was repeated for all precancers in the sample. The same coloring scheme for each precancer was then used in a 3D reconstruction of the sample, allowing users to match precancer histology to the correct 3D reconstructed precancer.

To create the graph in Fig. 5b, the number of precursors present in each sample was calculated. First, we determined the number of lesions present on each 2D section (not considering 3D connectivity). Next, we determined the true number of precursors present on each section when considering 3D connectivity. For each section in which at least one precursor was present, the number of (distinct in 2D space) precursor-classified objects was normalized by the number of (distinct in 3D space) precursor-classified objects that were present on the section. The average and standard deviation of this ratio for each sample was calculated and plotted.

Finally, metrics were performed on each independent precancer to determine 3D morphology. Using the 3D reconstructions and serial bounding boxes of each precancer, we determined 3D phenotype by assessing 3D presentation as well as the location of the precancer in

the pancreatic ducts or pancreatic acinar lobules. Next, cell count was determined by counting the number of cells located in the same voxel coordinates as each defined precursor lesions and corrected using the 3D cell conversion equation listed above. Precursor volume was calculated by summing the number of voxels defining each precancer and converting from voxel to  $\text{mm}^3$  units (1 voxel =  $12 \times 12 \times 12 \times 10^{-9} \text{ mm}^3$ ). Precursor cell density was calculated by dividing cell number per precursor by precursor volume. Precursor primary axis length was determined using the function 'regionprops3' in MATLAB 2021b.

### Calculation of collagen and nerve fiber alignment and nuclear aspect ratio

Using the 3D renderings, we identified three coordinates of axial sectioning and three coordinates of longitudinal sectioning around pancreatic ductal epithelium, blood vessels, and nerves for seven samples containing large regions of normal pancreatic parenchyma (for 42 total images of ducts, nerves and blood vessels each). We located the 2D histological sections using 3D coordinates of the identified regions and cropped the region of interest from the corresponding  $\times 20$  H&E images. We applied the color deconvolution method described above to the cropped  $\times 20$  H&E image to separate the H&E channels. We calculated fiber alignment within selected 2,500  $\mu\text{m}^2$  windows in the eosin channel images using a previously developed method<sup>53</sup>. By measuring fiber alignment within collagen or nerve regions in images of axial or longitudinal sectioning, we can compare the degree of collagen and nerve fiber alignment in axially and longitudinally sectioned regions of the ducts, blood vessels and nerves. An alignment index of one represents completely aligned matrix of fibers and an alignment index of zero represents an isotropic matrix of fibers. We measured the alignment index at two locations of each cropped image.

We next manually measured nuclear aspect ratio of cells within the peri-ductal/vascular/neural space using a script written in MATLAB v.2021b. To confirm the accuracy of the measurements, two scientists measured five randomly selected cells in each image, for a total of 1,260 cells measured. The nuclear aspect ratio measurements between the two researchers were compared in Extended Data Fig. 2e and the differences were shown to be statistically insignificant. Violin plots were constructed from data using code available in the provided reference<sup>62</sup>.

### Construction of z projections

The 3D labeled matrices of each sample were used to construct z projections of each tissue subtype. For each subtype, the pixels of the 3D matrix corresponding to that subtype were summed in the z dimension, creating a projection of the volume on a plane perpendicular to the xy axes. The projections were normalized by their maximum and visualized using the imagesc command in MATLAB 2021b using the same color scheme created for visualization of the 3D tissues.

### Analysis of normal and atrophic pancreatic lobules

For sample P7, the 3D model revealed a large region of acinar atrophy. Using an annotation pipeline written in MATLAB 2021b, registered, serial H&E images were rapidly displayed and manually annotated. In each image, the boundaries of the atrophic lobule and a nearby normal lobule were segmented. These regions were 3D reconstructed and tissue compositions were calculated using the methods described above.

### Confusion matrices

Quality of the deep learning models was visualized using construction of confusion matrices. For each datapoint in the testing dataset, the 'true' label (as manually annotated in H&E) was determined and matched with the 'determined' label (as classified by the deep learning model). A table was constructed to display the number of datapoints corresponding to each true label and their corresponding determined labels, as well as per-class precision and recall and overall model accuracy.

## Statistics and reproducibility

No statistical method was used to predetermine sample size. No data were excluded from the analyses. All statistical analyses were performed using the Wilcoxon rank sum test. All results may be duplicated from the available source data files or the 3D datasets.

## Ethics statement

This retrospective study was approved by the Johns Hopkins School of Medicine institutional review board.

## Reporting summary

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

## Data availability

Due to the extremely large size of the digital files described, data are available upon request from the corresponding author. Source data are provided with this paper.

## Code availability

Code is available on the following GitHub page: <https://github.com/ashleylk/CODA>.

## References

60. Goode, A., Gilbert, B., Harkes, J., Jukic, D. & Satyanarayanan, M. OpenSlide: a vendor-neutral software foundation for digital pathology. *J. Pathol. Inform.* **4**, 27 (2013).
61. Falkena, W. xml2struct v1.8.0.0 (MathWorks, 2020).
62. Hoffmann, H. Simple violin plot using matlab default kernel density estimation (INRES, Univ. Bonn, 2015).

## Acknowledgements

We thank J. Phillip and D. Gilkes for their important feedback in this work. We would additionally like to thank sources of funding for additional projects in our groups: grant nos. NIH/NCI P50 CA62924; NIH/NIDDK K08 DK107781; Sol Goldman Pancreatic Cancer Research Center; Buffone Family Gastrointestinal Cancer Research Fund; Carol S. and Robert M. Long Pancreatic Cancer Research Fund; Allegheny Health Network, Johns Hopkins Cancer Research Fund; American Cancer Society, The Cornelia T. Bailey Foundation Research Scholar grant no. RSG-18-143-01; AACR-Bristol-Myers Squibb Midcareer Female Investigator Grant; Emerson Collective Cancer Research Fund; Robert L. Fine Pancreatic Cancer Research Foundation; Rolfe Pancreatic Cancer Foundation; Joseph C. Monastera Foundation; The Gerald O. Mann Charitable Foundation (H. and A. Wulfstat, Trustees); S. Wojcicki and D. Troper; The Carl and Carol Nale Fund for Pancreatic Cancer Research. The Johns Hopkins University Oncology Tissue Services core used for sectioning and staining is funded by the SKCCC Cancer Center Support grant (CCSG; grant no. P30 CA006973).

Funding came from the National Institutes of Health/National Cancer Institute grant no. U54CA268083. (D.W., P.W. and A.L.K.); National Institutes of Health/National Cancer Institute grant no. U54CA210173. (D.W.); National Institutes of Health/National Institute on Aging grant no. U01AG060903 (D.W.); National Institutes of Health/National Cancer Institute grant no. UG3CA275681 (P.W.); The Sol Goldman Pancreatic Cancer Research Center (A.M.B., L.D.W., F.A., E.D.T., R.H.H., P.W. and D.W.); S. Wojcicki and D. Troper (A.L.K., A.M.B., L.D.W., F.A. and E.D.T.); The Rolfe Foundation for Pancreatic Cancer Research, Allegheny Health Network—Johns Hopkins Cancer Research Fund (A.M.B.); ARCS Foundation, Inc. (A.L.K.); Nanotechnology for Cancer Research T32 Training grant no. 5T32CA153952 (A.L.K.) and an NVIDIA GPU grant (P.W.).

## Author contributions

Conceptualization was done by L.D.W., R.H.H., P.-H.W. and D.W. Image registration methodology was carried out by P.-H.W. and A.L.K. Deep learning methodology was done by A.L.K. 3D reconstruction methodology was done by A.L.K. 3D quantification methodology was devised by P.-H.W. and A.L.K. CODA validation was done by A.C.J., P.-H.W. and A.L.K. Collagen alignment methodology was developed by P.-H.W., K.S.H. and A.L.K. Tissue annotation was done by M.P.G., A.M.B., J.M.B., R.R., F.A., A.L.K., A.C.J., B.K. and J.H. Tissue collection, sectioning and scanning were carried out by S.R., T.C.C. and A.M.B. Pathology consultation was done by S.-M.H., E.D.T., L.D.W. and R.H.H. Biostatistics calculations were done by A.L.K., P.-H.W. and P.H. The original draft was written by A.L.K., P.-H.W. and D.W. Review and editing of the draft were done by D.W., P.-H.W., L.D.W., R.H.H., P.H., M.P.G., A.M.B., J.M.B., R.R., F.A., A.C.J., B.K., J.H., K.S.H., S.M.H., E.D.T., T.C.C., S.R. and A.L.K.

## Competing interests

The authors declare no competing interests.

## Additional information

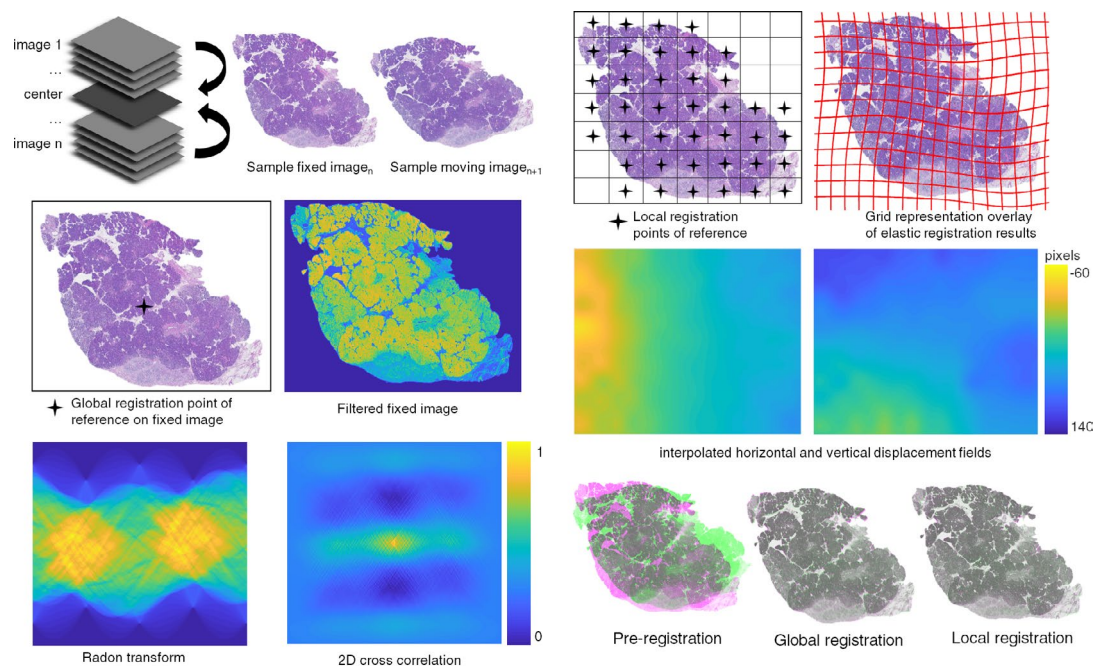
**Extended data** is available for this paper at <https://doi.org/10.1038/s41592-022-01650-9>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41592-022-01650-9>.

**Correspondence and requests for materials** should be addressed to Denis Wirtz or Pei-Hsun Wu.

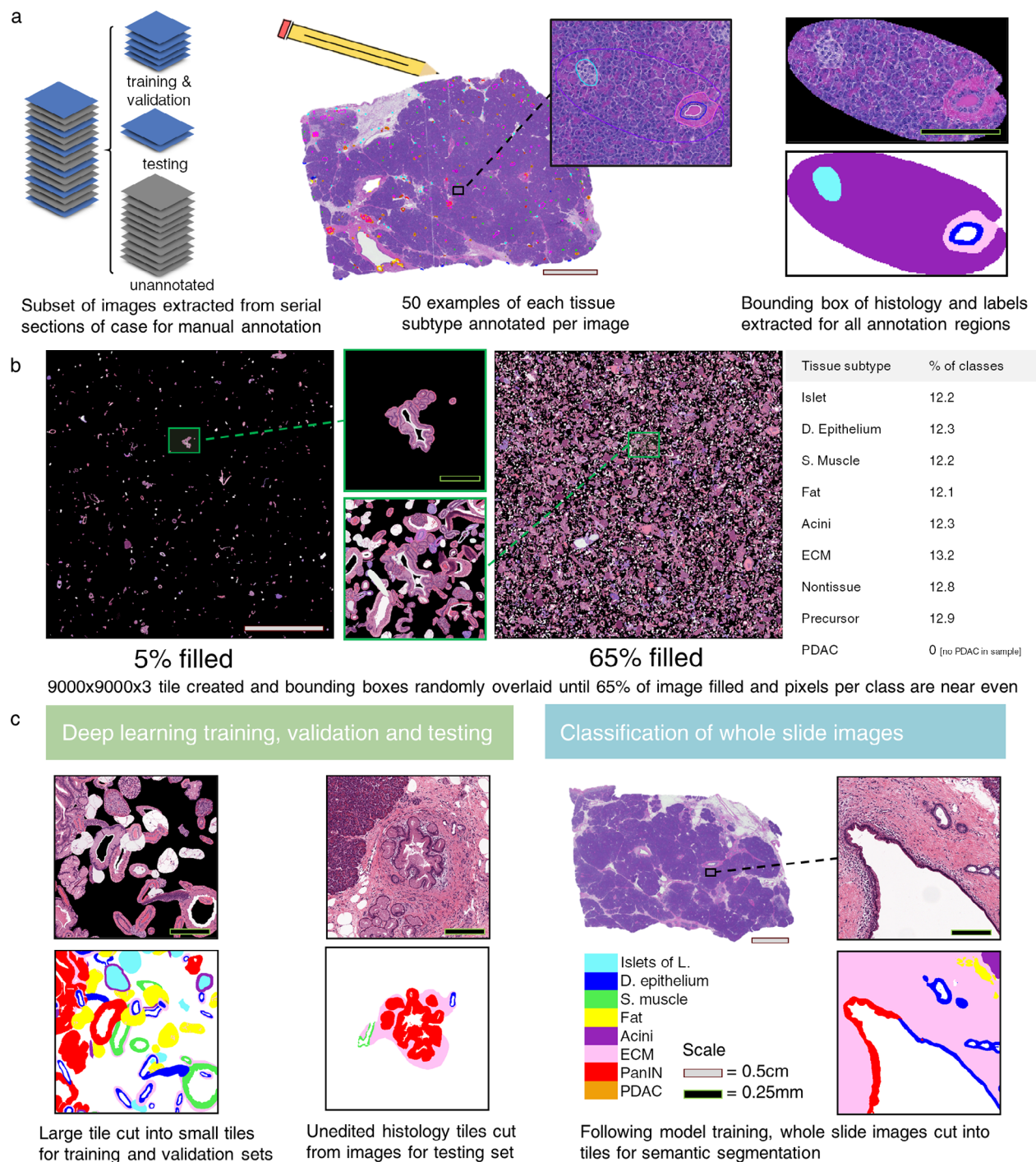
**Peer review information** Primary Handling Editor: Madhura Mukhopadhyay, in collaboration with the *Nature Methods* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



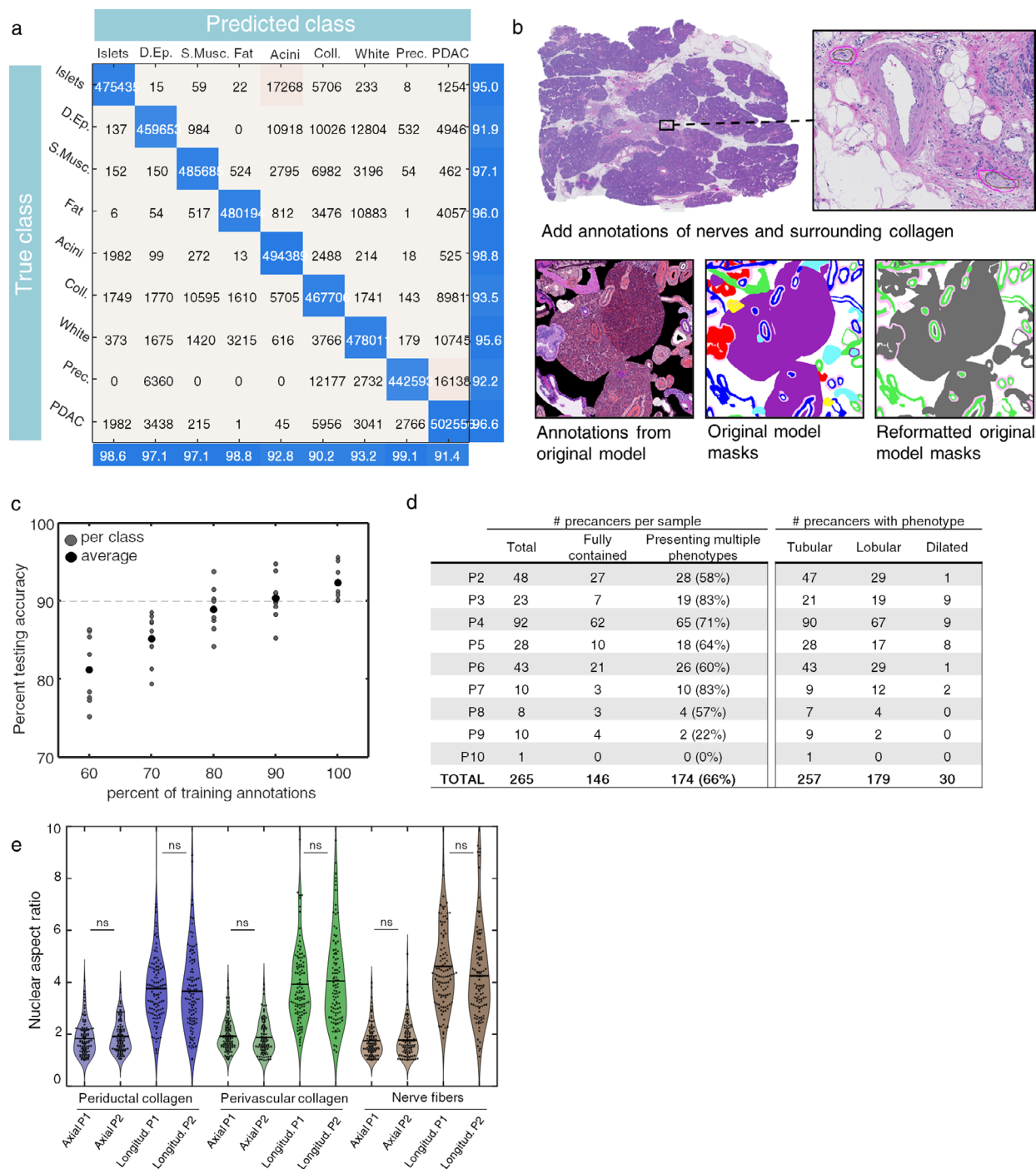
**Extended Data Fig. 1 | Histological image registration sample workflow.** Tissue cases registered with reference at center z-height of sample. Example fixed and moving images shown. Global registration performed with rotational reference at center of fixed image. Fixed and moving images smoothed by conversion to greyscale, removal of non-tissue objects in image, intensity complementing, and Gaussian filtering to reduce pixel-level noise in images. Radon transforms calculated filtered fixed and moving for discrete degrees

0–360. Maximum of 2D cross correlation of radon transforms yields registration angle. Maximum of 2D cross correlation of filtered images yields registration translation. Local registration performed at discrete intervals across fixed image. For each reference point, tiles are cropped from fixed and moving images and coarse registration is performed on tiles. Results from all tiles are interpolated on 2D grids to create nonlinear whole-image displacement fields. Sample overlays of pre and postregistration.



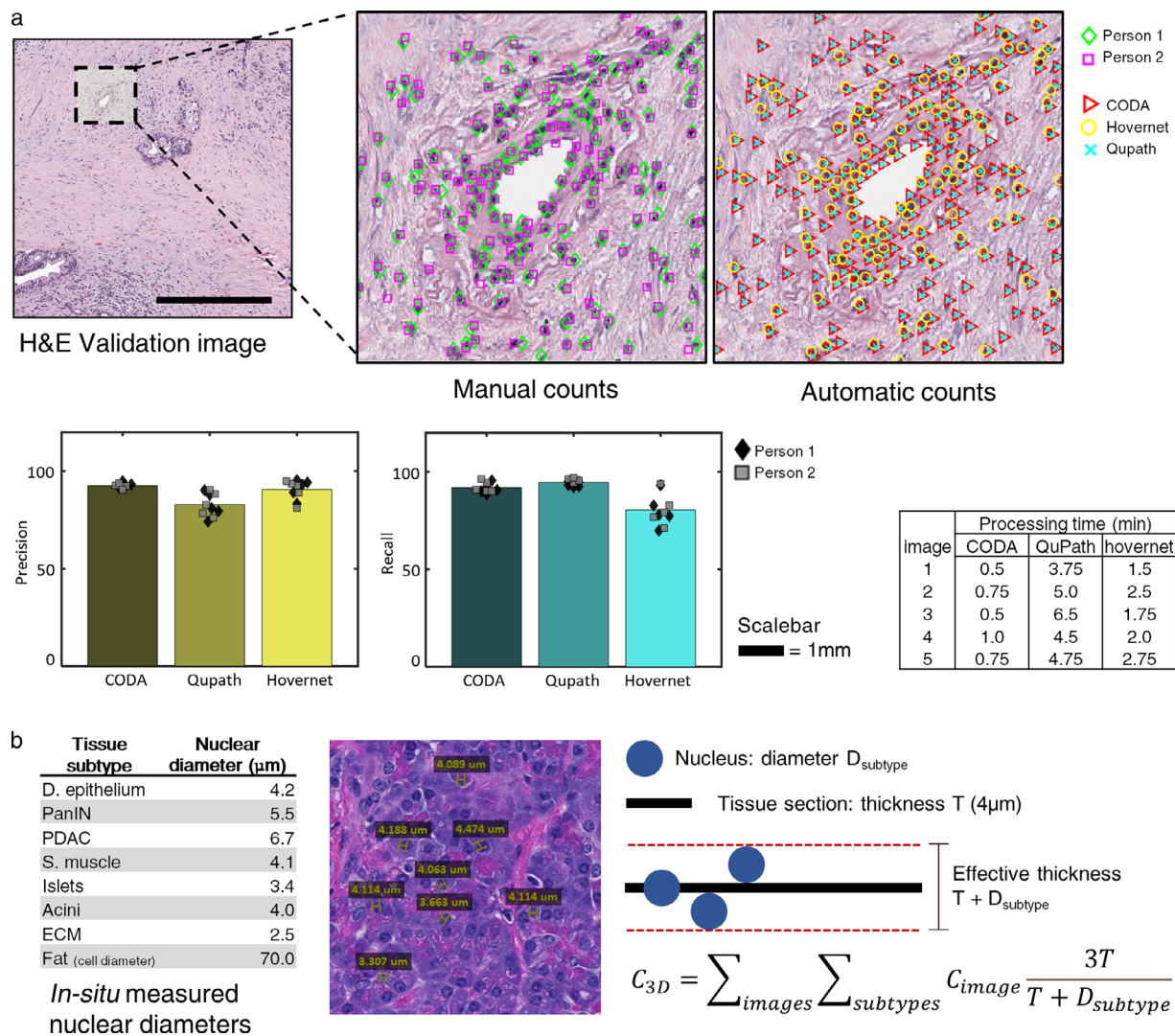
**Extended Data Fig. 2 | Overview of semantic segmentation workflow and training data design.** (a) For each case, a minimum of seven images were extracted from for manual annotation. For each extracted image, minimum 50 examples of each tissue type were annotated, and the annotations cropped from the larger image. (b) To construct training and validation sets, cropped annotations were overlaid on a large image until the image was >65% full and

such that the number of annotations of each type was roughly equal. (c) These large tiles were cut into smaller tiles for training and validation. Additional tiles were created for the testing set where the annotation was not cropped from the image. Testing accuracy was assessed as the percentage of the annotated area of the tile classified correctly. Following model training, the serial images were cropped into tiles and semantically segmented.



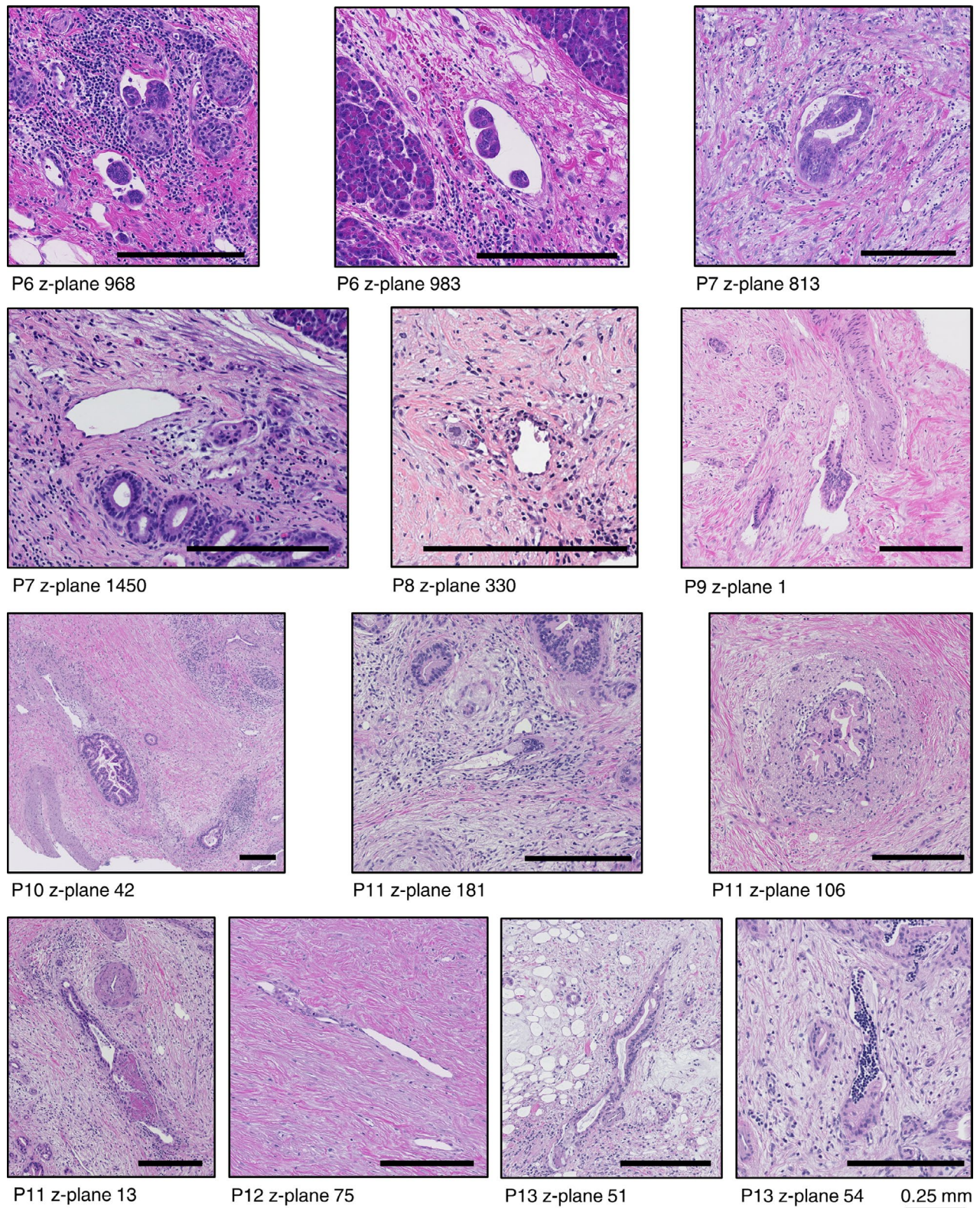
**Extended Data Fig. 3 | Additional methodological supplement.** (a) Sample predicted vs. true outcomes for deep learning models for sample P1 (left) and P8 (right). (b) Workflow for creation of multi-patient semantic segmentation of nerves. Nerve annotations collected from thirteen pancreas samples. Original tissue annotations reformatted to: 1. smooth muscle, 2. collagen, 3. other tissue (islets, normal ducts, acini, precursor, lymph, PDAC), 4. white (whitespace, fat). Nerve annotations combined with original annotations to create a dataset for

nerve recognition in H&E images. (c) Sample P7 average and per class testing accuracy as a function of percent of training annotations used. (d) Incidence of pancreatic phenotypes in eight samples. (e) Comparison of nuclear aspect ratio measurements performed by person 1 and person 2 (N = 150 nuclei per person per condition) show nonsignificant differences between measurements using the Wilcoxon rank sum test.

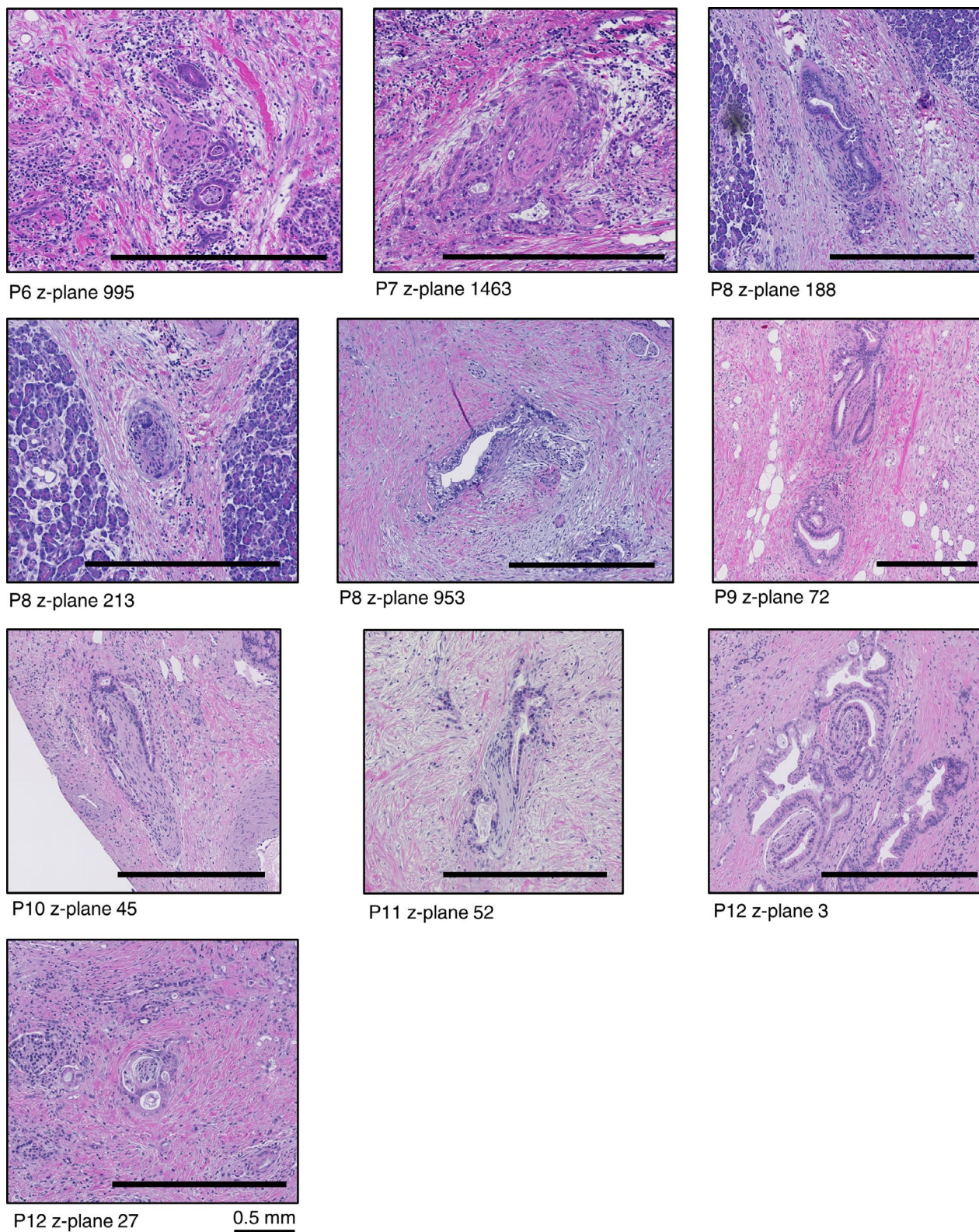


**Extended Data Fig. 4 | Validation of cell count and 2D to 3D cell count extrapolation.** (a) Sample histological section and corresponding color deconvolved hematoxylin channel of image. All cells in five validation images were manually annotated by two persons. Annotations were compared to CODA outputs and outputs from two existing cell counting methods<sup>27,28</sup>. (b) Cell diameters of each tissue subtype were measured using Aperio ImageScope. 2D cell counts were extrapolated to 3D using the formula listed. It was assumed that

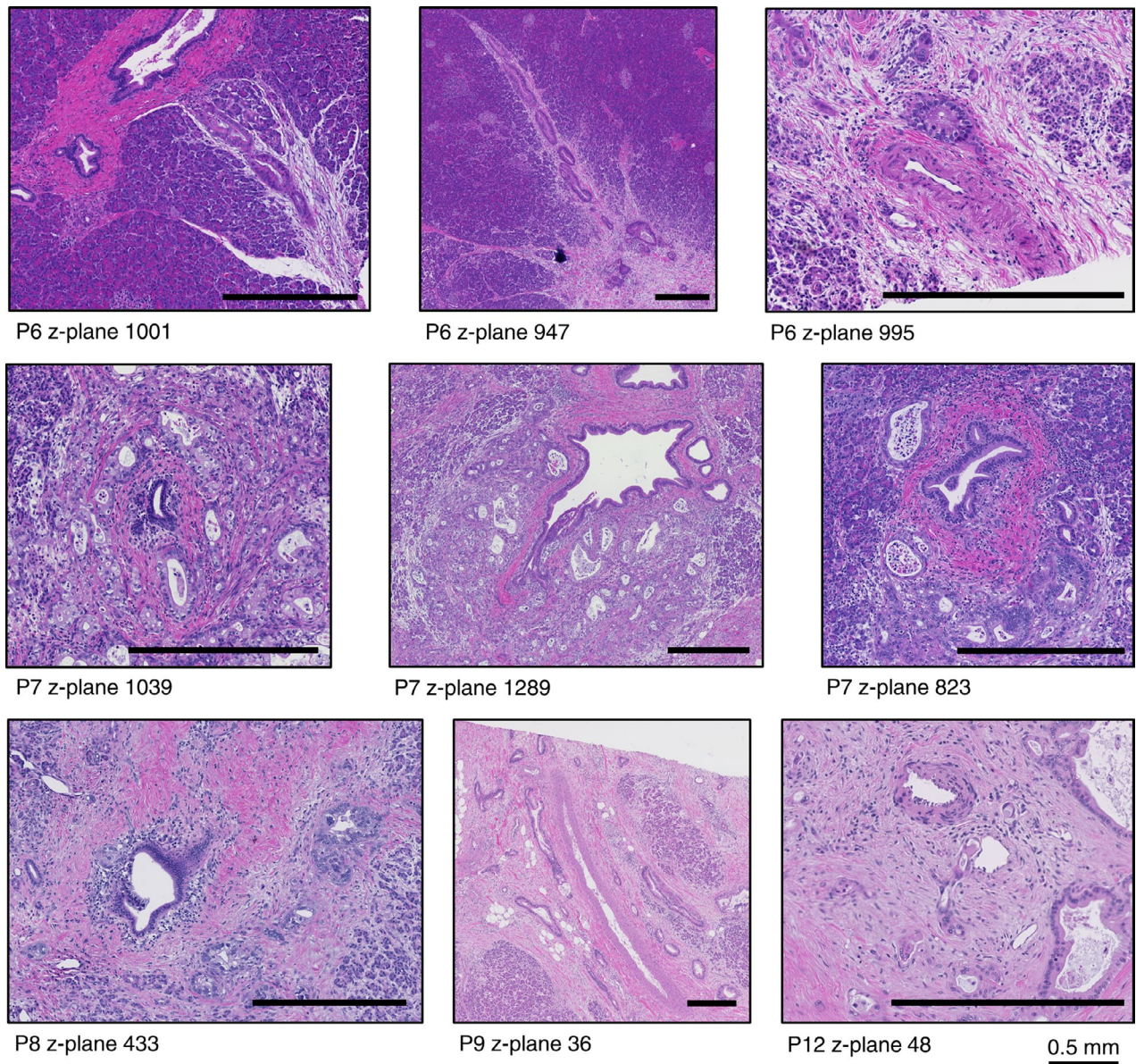
cells could be detected by the algorithm if any part of the nucleus touched the tissue section. Therefore, effective tissue section thickness equals true tissue section thickness plus the diameter of the cell. 3D cell counts were estimated by multiplying 2D cell counts by the true thickness of the tissue section, multiplied by three because two sections were skipped during scanning, divided by the effective thickness of the section.



**Extended Data Fig. 5 | Sample histology of venous invasions identified in samples.** Thirteen distinct venous invasions were identified in eight of the thirteen samples. For each, an H&E image was reviewed to confirm the venous invasion.



**Extended Data Fig. 6 | Sample histology of perineural/neural invasions identified in samples.** Ten distinct neural invasions were identified in seven of the thirteen samples, many containing regions of perineural invasion. For each, an H&E image was reviewed to confirm the structure.



**Extended Data Fig. 7 | Sample histology of invasion along regions of aligned collagen.** Nine distinct regions of invasion along aligned collagen were identified in five of the thirteen samples, including invasion along periductal collagen, invasion along perivascular collagen, and invasion along interlobular collagen. For each, an H&E image was reviewed to confirm the structure.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |                                                                                                                                                                                                                                                            |
|-------------------------------------|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement                                                                                                                                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly                                                                                                                                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>                                                               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of all covariates tested                                                                                                                                                                                                                     |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons                                                                                                                                        |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings                                                                                                                                                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes                                                                                                                                     |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated                                                                                                                                                         |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

### Software and code

Policy information about [availability of computer code](#)

Data collection Aperio Imagescope and MATLAB 2021b were used for data collection in this study.

Data analysis Aperio Imagescope, MATLAB 2021b, and openslide software were used for data analysis in this study

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

all software is available on GitHub (<https://github.com/ashleylk/CODA>). all data is available upon request (dataset is 25TB and cannot be hosted online)

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

|                             |                                                                                                                                                                                                                                                                                                                                                         |
|-----------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Reporting on sex and gender | Of the thirteen pancreatic samples considered in this study, 5 specimens were resected from female (sex) patients, and eight specimens were resected from male (sex) patients. Gender data was not considered in this study. A minimum of 2 specimens of each group (normal, precancer, and cancer) were selected from patients of male and female sex. |
| Population characteristics  | Pancreas tissue was taken from patients who underwent pancreatic resection but who had undergone no prior treatment for pancreatic cancer. Age data was not collected. Pancreas surgery location ranged from 5 samples taken from the pancreatic tail, 2 samples taken from the pancreatic body, and 6 samples taken from the pancreatic head.          |
| Recruitment                 | As this was a retrospective study, patients were not recruited. IRB approval from the Johns Hopkins School of Medicine was obtained to retrospectively study the microarchitecture of normal adjacent pancreas tissue to that previously removed during routine surgery.                                                                                |
| Ethics oversight            | This study was approved by the Johns Hopkins School of Medicine Institutional Review Board                                                                                                                                                                                                                                                              |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

|                 |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|-----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Sample size     | Thirteen large samples of human pancreas were used in this work. Samples were selected from tissues resected at the Johns Hopkins School of Medicine in order to get an even split of tissues containing normal pancreatic parenchyma, pancreatic precancers, and pancreatic cancer, and to get a minimum of five sample regions of each condition. This sample size was determined to be sufficient as the goal of this paper is to demonstrate the power of a novel technique, not to make broad biological statements. |
| Data exclusions | No data was excluded from this work                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| Replication     | Findings may be reproduced using our publicly available software                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Randomization   | Specimens were sorted into normal, precancerous, and cancerous groups. Randomization is not applicable here.                                                                                                                                                                                                                                                                                                                                                                                                              |
| Blinding        | The investigators selected samples consisting of untreated (1) normal pancreas, (2) pancreas containing precursor lesions, and (3) pancreas containing cancer. The investigators had no additional previous knowledge of the specimens.                                                                                                                                                                                                                                                                                   |

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a                                 | Involved in the study                                  |
|-------------------------------------|--------------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern  |

### Methods

| n/a                                 | Involved in the study                           |
|-------------------------------------|-------------------------------------------------|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |