PACESS: Practical AI-based Cell Extraction and Spatial Statistics for large 3D biological images

George Adams[1,2,†], Floriane S. Tissot[1,2,3,†], Chang Liu[4,†], Chris Brunsdon[5], Ken R. Duffy[6,7,‡], and Cristina Lo Celso[1,2,3,‡]

[1]Department of Life Sciences, Sir Alexander Fleming Building, Imperial College London, SW7 2AZ, London, United Kingdom
[2]Centre for Haematology, Department of Immunology and Inflammation, Imperial College London, W12 0NN, United Kingdom
[3]The Francis Crick Institute, WC2A 3LY, London, United Kingdom
[4]Hamilton Institute, Maynooth University, Maynooth, Co. Kildare, Ireland
[5]National Centre for Geocomputation, Maynooth University, Maynooth, Co. Kildare, Ireland
[6]Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115, USA
[7]Department of Mathematics, Northeastern University, Boston, MA 02115, USA
[†,‡]These authors contributed equally to this work

ABSTRACT

Efficient methodologies to fully extract and analyse large datasets remain the Achilles heels of 3D tissue imaging. Here we present PACESS, a pipeline for large-scale data extraction and spatial statistical analysis from 3D biological images. First, using 3D object detection neural networks trained on annotated 2D data, we identify and classify the location of hundreds of thousands of cells contained in large biological images. Then, we introduce a series of statistical techniques tailored to work with spatial data, resulting in a 3D statistical map of the tissue from which multi-cellular interactions can be clearly understood. As illustration of the power of this new approach, we apply this analysis pipeline to an organ known to have a complex and still poorly understood cellular structure: the bone marrow. The analysis reveals coherent, useful biological information on multiple cell population interactions. This novel and powerful spatial analysis pipeline can be broadly used to unravel complex multi-cellular interaction towards unlocking tissue complexity.

Main

A central tenet of biology and medicine is that within the observed spatial organisation of cells within tissues and organs there exists a set of functional relationships which underpins tissue function. As researchers look to understand more subtle spatially defined phenomena, such as the tumour microenvironment or the stem cell niche, multidimensional imaging (3D or greater) has proven to be invaluable[1–5]. That is because such images provide a single cell resolution view of the spatial organisation of populations *in situ*, resulting in extensive, information-dense data-sets from which researchers can make inferences[1,3–5].

Despite their promise, the extraction and analysis of data from these multidimensional images remains a significant challenge. Imaging tissues at depth almost inevitably results in greater amounts of background noise[2,5,6]. This makes image segmentation by intensity thresholding, which is, historically, the most widely used method for identifying cells-of-interest, much more difficult[7]. That issue is compounded in scenarios where cells are compressed or tightly packed together, such as seen with malignant processes, as boundaries between cells become more difficult to distinguish. Once data are extracted, inferences have typically been made using pairwise hypothesis testing[8,9]. That approach is highly restrictive as it does not facilitate the direct examination of the relationships between cells, making it challenging to formulate definitive conclusions from the underlying data.

To bridge this gap we present a pipeline for extracting and analysing cells from multidemensional data: **PACESS** ('**P**ractical **A**I-based **C**ell-**E**xtraction and **S**patial **S**tatistics') (Fig. 1). This pipeline makes use of convolutional neural-network based object-detection to classify and identify the locations of cells in 3D. Neural networks are the gold-standard in image classification, and object-detection neural networks combine this classification property with an AI-based searching algorithm to enable the accurate identification and localisation of objects within images[10–13]. Although widely used for 2D image analysis, 3D object-detection algorithms have had limited traction within 3D bio-image analysis because of the need to generate sufficiently

large sets of manually annotated 3D samples to train and test models on. To circumvent that difficulty, we introduce an augmented object detection deep neural network trained using 2D data alone, for which images can be rapidly annotated. Annotations are created for each image-layer in the 3D data and the output from multiple layers are automatically combined to identify each cell's location, size and type within the 3D space. Once this spatial data is extracted, we apply a step-wise analytical approach that was originally developed for geographical data[14], but which we have adapted to 3D cellular data. The steps in this process include: generating exploratory statistics that quantitatively assess spatial heterogeneity; identifying regions of abnormally high cellular density; performing hypothesis tests to determine the locality-dependent influence of one cell type's density on another; and, finally, formulating a spatial regression model that quantifies the location-dependent relationships between cell types. The advantage of this approach is that it makes full use of spatial information to provide intuitive and meaningful statistics whose output can be presented as a holistic 'statistical map' of the tissue.

To demonstrate the utility of our approach we have applied it to a series of samples of bone marrow. Bone marrow is a uniquely challenging tissue for histological imaging and quantification[4,5,7]. It has an amorphous structure consisting of a wide-range of different cell types which are tightly packed within a confined cavity[15]. It has a high mineral content, primarily in the form of iron in haem, which creates substantial amounts of background autofluorescence[5,6]. To highlight how robust our extraction method is to background flourescence we demonstrate its use within both optical cleared thick sections of bone marrow and non-cleared intravital microscopy (IVM) images, which have higher levels of background auto-fluorescence[16].

Methods

Experimental animals

For this work we used two types of transgenic reporter mice: mTmG-reporter mice that expressed dTomato within the surface membranes of all cells and a second line that expressed dTomato exclusively within von Willebrand factor (vWF) expressing cells[17], which include megakaryocytes and endothelial cells. The latter transgenic line was used for our intra-vital microscopy (IVM) studies, whilst the former was used to generate the acute myeloid leukemia (AML) cells included in some of the sampls presented. All animal work was performed in accordance with the animal ethics committee (AWERB) at Imperial College London and UK Home Office regulations (ASPA, 1986).

To generate an example of leukaemic infiltrated bone marrow, AML cells were generated from purified granulocyte/monocyte progenitors (GMPs) from the mTmG-reporter mice. These GMPs were transduced with pMSCV-MLL-AF9-GFP-based retroviruses as described in[18,19] and then transplanted into sub-lethally irradiated (conditioned) mice. Approximately ~ 8 weeks post transplantation, these conditioned recipient mice develop highly infiltrated leukaemia. Tomato+, GFP+ cells from these mice are then harvested from bone marrow and spleen and pooled. 100,000 viable AML cells were transplanted through tail vein injection into the secondary, non-conditioned recipient mice[20]. Progressive expansion was observed from day 8-10. Tissues from these mice where harvested once the leukaemic infiltrate approached 15-20% in bone marrow. The percentage of AML infiltration was determined using flow cytometry in which a sample of bone from each mouse was crushed in PBS with 2% fetal bovine serum and then filtered through a 40 μm strainer. Viable cells were distinguished using 4,6-diamidino-2-phenylindole (DAPI, Invitrogen) and AML cells were identified based on dTomato and GFP expression. Of note, GFP expression is low and is lost upon tissue processing for *ex vivo* analyses.

Tissue processing and imaging

The clearing procedure we use was developed to suit our specific microscope system whilst incorporating various advances made in this field[4–6]. The complete steps were: (**1**) Harvested tissues are fixed in 4% PFA for 1-2hrs, then (**2**) placed in 10% EDTA during 15 days for decalcification[3]. After decalcification, (**3**) the bones are embedded in low melting point agarose (sigma A0169) and cut using a Leica T1000 Vibratome at depths of 250μm. After sectioning, (**4**) samples are incubated for 48hrs in a solution made up of 20% CUBIC-1 reagent (urea 25 wt% by weight, Quadrol 25 wt% by weight, Triton X-100 15% by weight) in dH2O) diluted in 80% dH20 at a pH 10[6]. (**5**) Non-specific binding of antibodies is blocked for 24 hours in 1.5% TBS 1% Triton, 20% DMSO, and 10% donkey serum[3]. (**6**) Tissues are then incubated with primary and then secondary antibodies for 48 hours in a solution of 1% TBS, 1% Triton, 200mM sodium sulfide. Nuclei are stained with DAPI for 24 hours. (**7**) Samples are then mounted onto 20x20x0.9mm silicone spacers using an optical clearing solution made up of 1.455g/ml histodenz (sigma) and 40% Methlyacetamide (sigma) diluted in 1% TBS with 4% DABCO (sigma)[4] (SFig.1).

In addition to optically cleared samples, we applied our data extraction process to images generated from intravital microscopy (IVM). In this technique, images of the bone marrow are taken whilst the animal is still alive. This technique allows real-time visualisation of the bone marrow but is subject to higher levels of background auto-fluorescence. The IVM process is fully described in the following references[16,21] and only a brief overview of the procedure is provided here. The calvarium of the animal is exposed whilst the animal is under isoflurane anaesethesia. The position of the animal is then secured using

a headpiece mounting under a confocal microscope. To visualise vasculature, 8mg/ml FITC-dextran was injected prior to imaging.

Imaging was performed using a Zeiss LSM 980 upright confocal microscope equipped with 5 Argon lasers (405, 488, 561, 594 and 639nm), and an Insight (Newport Spectraphysics) 2-photon laser with two excitation lines of which one is fixed and one tunable (1045nm and 680-1300nm respectively). The microscope was equipped with 6 non-descanned external detectors including 2 nose-piece detectors (GaASP). Images were acquired using a 20x, 1.0N.A., water immersion lens with 1.4mm working distance.

Object detection and clustering

A YOLO-V5X model (https://github.com/ultralytics/yolov5) was used as the backbone of the 2D object-detection neural network. More details on this model, and on convolutional neural networks more generally, can be found in the literature[10,22–24]. In this section we provide only a brief overview of the model with a focus on: the manner in which the algorithm identifies objects; and how it was adapted to generate 3D estimates.

The YOLO algorithm works by placing a multitude of boxes within the space of a 2D image and then filtering these boxes based on probability estimates from the model. It is a fully connected neural network which divides a 2D image into $S \times S$ grid of cells into which B bounding boxes are detected. The model identifies a set of box sizes for each class *a priori* using a k-means clustering algorithm run on box sizes observed within the training data. Each bounding box is defined by 5 parameters: the x, y central position, width (w), height (h) and a confidence score, C. This last value, C, is the confidence estimate over the presence, or absence, of an object being within the grid cell. This makes use of the intersection-over-union (IOU) between a predicted bounding box and a ground truth (manually annotated) bounding box. The greater the overlap between the two, the higher the IOU, and greater the confidence in the box. If any object is absent from the grid cell, the probability of the object ($Pr(\text{Object})$) is set to 0. Otherwise it is 1. For the ith bounding box in the jth grid cell, the confidence score, C_{ij} is thus calculated as[24]:

$$C_{ij} = P(\text{Object}_{ij}) \times \text{IOU}$$

In addition to these five parameters a set of conditional class probabilities is calculated. Given K possible classes, this is the probability of the object belonging to any specific k^{th} class: $\Pr(\text{Class}_k|\text{Object})$. A class-specific confidence score (CS_{kij}) is then calculated as a product of C_{ij} and the conditional class probability[24]:

$$\begin{aligned} CS_{ijk} &= P(\text{Object}_{ij}) \times \text{IOU} \times \Pr(\text{Class}_{kij}|\text{Object}_{ij}) \\ &= P(\text{Class}_{ijk}) \times \text{IOU} \end{aligned}$$

The class-specific confidence scores and IOU results are used to *select* bounding boxes through *non-maximum suppression* (NMS)[24,25]. YOLOv5 makes use of *soft*-NMS which is better adapted to overlapping objects[25].

To aggregate the final set of 2D bounding boxes into 3D bounding 'cubes' we ordered the bounding boxes for each class by maximum diameter and mean fluorescence intensity (mFI). For each box within the set of boxes (B) within the kth class, starting from the largest and brightest boxes, a central x, y, z location is calculated, which we call q. We can also determine a maximum diameter for this box, d. From this point q the surrounding cluster of boxes in the z dimension which have a distance from q which is $< d/2$. We call this set of clustered bounding boxes B_c. Within this context $B_c \subset B$ but all the 2D bounding boxes within B_c are assumed to belong to a single cell (cube) surrounding an individual cell. Once identified, B_c is removed from the B. The process is repeated until every 2D box is allocated to a 3D cube.

To apply this model, an $x \times y \times z \times c$ dimensional image, where $c = 3$ for RGB, was divided into a set of $416 \times 416 \times 1 \times 3$ (RGB) tiles. Test/validate/train subsets were selected from random sampling of this tile set. Manual annotation was performed to identify cells of interest within these selected images with a minimum of 500 cells annotated within each cell class. Once trained, the final object detection was performed on the full-set of tiles.

Spatial analysis and modelling

Cells outside the bone marrow were excluded from the object detection dataset. Then, the bone marrow was divided into $\theta(\mu m)^3$ cubes. The midpoint of the lowest plane in the cube (u, v, z) was used as the three-dimensional geographic coordinates, and the number of cells of each type in the cube was recorded. The data in each of these cubes was used as input to the analysis and model.

Moran's I index was used to measure spatial autocorrelation for each cell type[26]. The formula for Moran's I index is

$$I = \frac{n\sum\sum_{i\neq j}\varphi_{ij}(a_i-\bar{a})(a_j-\bar{a})}{(\sum_{i=1}^{n}(a_i-\bar{a})^2)(\sum\sum_{i\neq j}\varphi_{ij})},$$

where n is the total number of cubes, a_i is the number of cells in a particular type at the ith cube, a_j is the number of cells at the jth cube, $\bar{a}$ is the mean of the number of cells at each cube, and φ_{ij} is a spatial weight. The formula for φ_{ij} is

$$\varphi_{ij} = \begin{cases} 1 & \text{if } d_{ij} \leq \theta \\ 0 & \text{if } d_{ij} > \theta \end{cases},$$

where d_{ij} be the Euclidean distance between the centroids of cube i and cube j.

The density-based spatial clustering of applications with noise (DBSCAN) algorithm was the algorithm used to cluster the cells[27]. For cube (u_i, v_i, z_i), $N_i = \{(u_j, v_j, z_j)|d_{ij} \leq \theta\}$ is a set of all neighbouring cubes that are θ μm or less away from the ith cube. The number of cells in N_i is recorded as $\|N_i\|$. When the faces of cubes are connected to each other, they are neighbours, so each cube in this case has six neighbours. If the total number of cells in cube (u_i, v_i, z_i) and its neighbours is greater than 7, the number of cubes being considered, times γ, such as the third quartile of counts, then the ith cube and its neighbours are marked as high-density cubes. Let $\Omega = \{(u_j, v_j, z_j)|\,\|N_i\| \geq 7\gamma, d_{ij} \leq \theta\}$ is the set which includes all high-density cubes. For any cube $(u_i, v_i, z_i) \in \Omega$, $\partial_1 = \{(u_j, v_j, z_j)|d((u_i, v_i, z_i), \Omega) \leq \theta\}$ where ∂_1 includes all high-density cubes close to ith cube, and $d(A, B)$ represents Euclidean distance between the set A and the set B. Then, $\partial_2 = \{(u_j, v_j, z_j)|d(\partial_1, \Omega) \leq \theta\}, \cdots,$ $\partial_{n+1} = \{(u_j, v_j, z_j)|d(\partial_n, \Omega) \leq \theta\}$. When $|\partial_{n+1}| - |\partial_n| = 0$ where $|\partial_n|$ is the number of cubes in the ∂_n, iteration ends and ∂_n is the first cluster which is records as C_1. If $\Omega_1 = \Omega - C_1 = \varnothing$, then there is one cluster. Conversely, any cube $(u_l, v_l, z_l) \in \Omega_1$ are selected. The second cluster C_2 and Ω_2 can be obtained using the same step. When $\Omega_{p+1} = \varnothing, p \in \mathbb{N}^*$, the data has p clusters. In addition, for any cube that does not belong to any cluster, these cubes are in the set C_0.

After cell clustering, the information on all clusters can also be obtained. For the τth cluster ($\tau \in \mathbb{N}^*$ and $\tau \leq p$), $U_\tau = \{(u_\tau, v_\tau, z_\tau)|0 < d((u_\tau, v_\tau, z_\tau), C_\tau) \leq \theta\}$ is a set which contains the cubes in the τth cluster and the cubes θ μm away from the τth cluster, and these cubes do not belong to any other clusters. Hence, the cubes around the τth cluster are $C_\tau^c = \{(u_i, v_i, z_i)|(u_i, v_i, z_i) \in U_\tau, (u_i, v_i, z_i) \notin C_\tau\}$, that C_τ^c.

Permutation tests were used to detect changes in the number of cells in the cluster as well as the number of cells around the cluster[28]. The null hypothesis for the permutation test is that the mean number of cells in the cubes is independent of whether the cubes in C_τ or C_τ^c. Here, $A = (a_{\tau_1}, a_{\tau_2}, \cdots, a_{\tau_{|C_\tau|}}, a_{\tau_{|C_\tau|+1}}, a_{\tau_{|C_\tau|+2}}, \cdots, a_{\tau_{|C_\tau|+|C_\tau^c|}})$ is an ordered observations set, where a_{τ_i} is the number of cells in the ith cube in the C_τ. In the set A, the first $|C_\tau|$ elements are the number of cells in the τth cluster, and the last $|C_\tau^c|$ elements are the number of cells around the cluster. Hence, the real-valued statistic is used under this null hypothesis.

$$M(A) = \frac{\sum_{i=1}^{|C_\tau|} a_i}{|C_\tau|} - \frac{\sum_{j=|C_\tau|+1}^{|C_\tau|+|C_\tau^c|} a_j}{|C_\tau^c|}.$$

Then, a permutation π is created, that can reassign labels to individual datum. A reordered observation set is obtained:

$$A_\pi = (a_{\pi(1)}, a_{\pi(2)}, \cdots, a_{\pi(|C_\tau|)}, a_{\pi(|C_\tau|+1)}, a_{\pi(|C_\tau|+2)}, \cdots, a_{\pi(|C_\tau|+|C_\tau^c|)}).$$

Similarly, a new reordered set also generates statistics $M(A_\pi)$. A collection of permutations Q can be characterised so that reorderings $\{A_\pi\}_{\pi\in Q}$ are equally likely under the null hypothesis. Then, the empirical distribution of $M(A_\pi)_{\pi\in Q}$ is used to compare with $M(A)$. Therefore, the corresponding p-value can be calculated. At a significance level of 0.05, when the p-value is smaller than 0.05, the null hypothesis is rejected.

A geographically weighted regression model (GWR) was used to examine the spatial relationship between explanatory variables and the response variable[14]. The standard GWR model defined for two-dimensional (2D) plane, and this study extends it to three-dimensional (3D) space. The natural extension of the 2D GWR approach to 3D is

$$r(u_i, v_i, z_i) = \beta_0(u_i, v_i, z_i) + \sum_{k=1}^{m} \beta_k(u_i, v_i, z_i) e_k(u_i, v_i, z_i) + \varepsilon(u_i, v_i, z_i),$$

where $r(u_i, v_i, z_i)$ is the response variable at the ith cube, $\beta_0(u_i, v_i, z_i)$ is the intercept in the model, m is the number of explanatory variables, $\beta_k(u_i, v_i, z_i)$ is the coefficient for the kth input variable at the ith cube and $e_k(u_i, v_i, z_i)$ is the kth explanatory variable at the ith cube. In addition, $\varepsilon(u_i, v_i, z_i)$ is the error term at the ith cube.

A weighted least squares method is used to get the coefficients $\hat{\beta}(u_i, v_i, z_i) = (\beta_0(u_i, v_i, z_i), \beta_1(u_i, v_i, z_i), \cdots, \beta_m(u_i, v_i, z_i))^T$. The formula for the coefficients is

$$\hat{\beta}(u_i, v_i, z_i) = (E^T W(u_i, v_i, z_i) E)^{-1} E^T W(u_i, v_i, z_i) r \,,$$

where E is a $\mathbb{R}^{n\times(m+1)}$ matrix that includes 1s for intercept and explanatory variables, r is a $\mathbb{R}^{n\times 1}$ response vector, $W(u_i, v_i, z_i) = \texttt{diag}(w_{i1}, w_{i2}, \ldots, w_{in})$ is the diagonal weighted matrix at position (u_i, v_i, z_i), and it is determined by a kernel function. In this study, the bi-square kernel function is used:

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2 & \text{if } |d_{ij}| < b \\ 0 & \text{otherwise} \end{cases},$$

where b is the bandwidth. The bi-square kernel function reflects that neighbouring points have more influence on the ith cube than distant ones. Its scaling is determined by a bandwidth b that is selected by minimising a corrected version of Akaike Information Criterion (AICc)[29]:

$$\text{AICc}(b) = 2n\ln(\hat{\sigma}) + n\ln(2\pi) + n\frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)} \,,$$

where, $\hat{\sigma}$ is the standard deviation of the residuals, and $\text{tr}(S)$ is the trace of the matrix S, which is called the hat matrix in standard GWR.

Adapting principles from 2D GWR diagnostics, for the 3D GWR the following diagnostic statistic was adopted, the local R^2. Local R^2 can reflect the quality of local models to explain local data. Local R^2 is defined as[14]

$$R^2(u_i, v_i, z_i) = 1 - \frac{\sum_{j=1}^{n} w_{ij}(r(u_j, v_j, z_j) - \hat{r}(u_j, v_j, z_j))^2}{\sum_{j=1}^{n} w_{ij}(r(u_j, v_j, z_j) - \bar{r})^2},$$

where, $\bar{r}$ is the mean of response variable, and $\hat{r}(u_i, v_i, z_i)$ is the fitted $r(u_i, v_i, z_i)$. In addition, w_{ij} is from the weighted matrix.

Software and computational resources

Neural network model training was performed using 4× RTX6000 Nvidia GPUs and 8 CPUs with 96GB RAM available through the high-performance computer cluster available at Imperial College London. Trained neural network models and the spatial models were run on a Dell precision 5560 laptop with 32GB RAM and a NVIDIA T550 4G DDR6 GPU.

Results

Data extraction using deep neural networks

To demonstrate the effectiveness of the methodology introduced here (Fig. 1), two types of 3D biological images were generated: optically cleared 'thick' sections, and images from intravital microscopy (IVM). The latter technique does not require the tissue to cleared or cut, and takes advantage of the natural thinness of the calvarium bone such that the marrow can be visualised *in situ*. The disadvantage of IVM is that it generates images that lack the crispness of clarified samples and is subject to higher levels of auto-fluorescence. The difference between these two types of images can be seen in Fig. 2a, which shows an image of bone marrow calvarium from a dTomato:vWF transgenic animal that has first been visualised with IVM (left part of the image), before being harvested and re-imaged after clarification (right part of the image). Image segmentation using thresholding was performed on both the cleared and IVM image (bottom part of the image) to illustrate the shortcomings of that approach in the presence of overlapping cells where the boundaries of overlapping cells merge, making it difficult to discern individual cells.

IVM performed on vWF:dTomato transgenic mice generated a series of 3D images of the mouse calvarium in which two cell classes, megakaryocytes (MGK) and endothelial cells are both identified using the same fluorophore (Fig. 2b). A 2D YOLO neural network model was trained on a subsample of those data using 7581 manually annotated cells, of which 5409 were used for training, 1284 for validation and 888 for testing. To check the model, we first assessed whether the network was

classifying appropriately by determining which features within the image elicited a response by the neural network using a class activation, or saliency, map[30,31]. The output of this map illustrates how the activation within lower layers of the network respond to the image provided and thus the features that the model has learnt to recognise (example in Fig. 2c). This is an important initial assessment to ensure the model is not being trained on a non-biological variable within the image. We then assessed model accuracy. For both MGK and endothelial cells, the neural network achieves approximately 90% accuracy on the set of labelled test images that were not used during training, Fig. 2d (mean AUC 0.904, MGK AUC 0.894, endothelial cells AUC 0.913). Once satisfied with the interim (2D) model a set of 2D predictions were generated for each layer in the image. These predictions where then used to generate estimates of the 3D locations of cells by identifying neighbouring boxes across different image layers using our extended 3D prediction algorithm (Fig. 2e, orthogonal view panels). Crucially, this extension means that these 3D predictions could be generated without the need for 3D labelled images.

To demonstrate how this procedure performs in the presence of occluding cells, it was used on a sample of clarified bone marrow sternum from a vWF:dTomato transgenic mouse that exhibits densely packed MGK cells, Fig. 3. MGKs are large cells with complex morphology localised adjacent to dense vasculature structure of the sternum bone[7]. Fig. 3a shows a maximum projection image of a clarified sternum from a vWF:dTomato transgenic mouse that was stained for endomucin (green) to visualise blood vessels. This served as an aid for human inspection and was not used as input to the neural network. Such images are a particular challenge for any data-extraction procedure as the large number of overlapping cells confounds segmentation by thresholding and induces difficulties in 3D annotation for training of 3D neural networks. To train and test our network, a total of 6898 cells in 415 2D sub-images of the full image were annotated for the training, validation and testing process (Supplementary Table 1). The network first identifies MGK cells at each individual z depth within the image (Fig. 3b). MGKs are then aggregated across depths (Fig. 3c) to create a 3D prediction as bounding cubes. The cubes approximately surround each identified cell within the sample. Fig. 3d shows a maximum intensity projection (MIP) of the vWF+ cells within the tissue and the estimated location of each cell (white dots) based on the location of the bounding boxes.

Data-extraction in an AML mouse model

The essential promise of dense, 3D data sources is that they will enable understanding of spatial composition of organs. Having established that our neural network approach can automatically identify the location of cells as well as accurately classify their type based on morphology or fluorescence using exclusively 2D training data, we used a more complex set of images with multiple cell types to elucidate the pipeline for spatial analysis. For that illustration, we captured data from a leukaemic mouse model where densely packed acute myeloid leukaemia (AML) cells have been posing a substantial identification challenge, for example making it impossible to identify and track single AML cells found within malignant patches. Moreover, based on published results, AML would be expected to influence the localisation of bone marrow resident cells, making it a suitable model to test the ability to identify spatial relationships between cell types[19,32,33].

Within the images, AML cells and $CD8^+$ T cells were identified by dTomato and AF488 fluorescence respectively. Instead of using a fluorescent reporter, MGK cells were identified based on characteristic morphological features (large size, multilobulated nucleus).

The tissue sample imaged for this experiment had a 15% AML infiltration, as estimated by flow cytometry analysis of other bones from the same animal (Supplementary Fig. S1a and S1b). The neural network model was trained using 18,240 manually annotated cell examples (Supplementary Table 1). The data-extraction procedure identified an order of magnitude more cells than the training set for a total of 163,953 AML cells, 8,362 $CD8^+$ T cells and 2,219 MGK cells. Fig. 4a shows a single cross-section image of the sample on which the cell positions identified from the cell extraction procedure are mapped. Each box represents the transverse plane through bounding cubes. In this sample, although AML cells are present throughout the tissue, there is an area of dense infiltration on either side of the growth plate, extending into both metaphysis and epiphysis. Even in areas of dense infiltration the neural network was capable of distinguishing individual AML cells (Fig. 4a, *i.-vi.*). The model only failed to identify a small number of cells in areas of particularly dense infiltration (Fig. 4a *c.*, *iii.*).

Next, we completed 3D reconstructions of these data within bone marrow cavity space. Fig. 4b shows the locations of the bounding cubes for each extracted cell type of interest (AML cells, $CD8^+$ T cells, and MGK cells) in cross-sectional planes. In Fig. 4c, the central position of each cell within the bone marrow cavity space is shown. These figures quantitatively demonstrate the extent of the AML infiltrate and that $CD8^+$ T cells are largely excluded from the larger dense region of AML. MGK cells, in contrast, are not excluded to the same extent and remain more homogeneously distributed. As there is a substantial increase in the scale of information when moving from a 2D cross-section to a 3D 'thick' section and further interpretation of the data is not possible without a well-developed quantitative framework.

Spatial heterogeneity and automatic identification of areas of high cellular density

In order to visualise and analyse the spatial distribution and relationships between cell types, space was discretised into non-overlapping, adjacent cubes that cover the entire 3D area. The number of cells of each type in each cube was recorded. For meaningful visualisation, the discretisation needs to be sufficiently coarse that some aggregation of cell counts occurs. A cube size of $45(\mu m)^3$ was selected to be sufficiently fine that geographic resolution was retained, but sufficiently coarse that the resulting data could still be computationally assessed without undue burden. With that discretisation, we could report the density of AML cells, T cells and MGKs as a function of their position for a selection of z-depths (Fig. 5a, 5b and 5c).

AML cells were observed to be more prevalent than the other two types of cells, with most of them appearing to be located in a single mass, as previously seen in Fig. 4. This observation highlighted the importance of being able to quantitatively assess the homogeneity of each cell type's spatial distribution, which we achieved through the calculation of a statistic called Moran's I. If Moran's I is positive, cells tend to be aggregated in common areas. If Moran's I is close to zero, cells are distributed randomly in space. When its value is less than zero, cells are more homogeneously dispersed than one would expect from a random process. Moran's I for the AML cells was 0.81, quantitatively substantiating the observation that AML cells were largely concentrated in relatively few patches.

While Moran's I can indicate that cells of a given type are largely co-located, a distinct methodology is needed to identify the regions of high density. One approach is the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm[27]. For these data, when the distance between the cubes was $45\mu m$ or less they were considered to be neighbours, resulting in each cube having seven neighbouring cubes that have a face in common (Supplemental Fig. S2a). If the average number of cells in a cube and its neighbours was more than the third quartile for a single cube, where only a quarter of cubes have more cells than this value, it was considered a dense neighbourhood. Dense neighbourhoods were agglomerated using the DBSCAN algorithm to form contiguous spatial clusters (Supplemental Fig. S2b).

For the AML data, DBSCAN identified 43 distinct clusters, four of which accounted for 58.8%, 3.6%, 1.9% and 1.8% of all AML cells. These clusters are shown Fig. 6a, where the ten largest by cell count are marked in decreasing order. The three largest clusters of AML cells, 1, 2 and 3, were seen to be located close to each other at the edge of the bone marrow. The fourth largest cluster was at the bottom left. In the lower part of Fig. 6a, the AML clusters were much smaller, and most of were located at the boundary of the bone marrow.

For T cells and MGKs, respectively, approximately 70% and 90% of cubes recorded a zero cell count, consistent with these cells being less abundant than AML cells. Moran's I of 0.05 for MGKs indicated these cells had a weak geographical dependency resembling random locations. Moran's I for T cells was 0.34, suggesting some positive spatial clustering but less than found for AML cells.

Spatial density dependencies

Visual inspection of Fig. 5b showed that there were few T cells in areas with high AML cell counts. To statistically assess if the distribution of T cell counts was influenced by areas of high AML density, permutation tests were used to challenge the null hypothesis that the mean number of T cells within each cube is independent of whether the cube is within an AML cluster or in the boundary of cubes surrounding it (Supplemental Fig. S2b).

Clusters for which the hypothesis test was rejected are marked in Fig. 6b. For the four largest clusters, Clusters 1 and 3 showed statistically significant differences with p-values of $< 2.2 \times 10^{-26}$ and 2×10^{-5}, respectively. Of the remaining 39 smaller AML clusters, four p-values were less than 0.05. Cluster 10, which contained 0.5% AML cells, had a p-value of 0.017, while Clusters 9 and 5, which contained 0.05% and 0.9% of all AML cells, respectively, had p-values of 0.021 and 0.03. Finally, Cluster 6, which was found close to Cluster 1 and included 0.6% AML cells, had a marginal p-value of 0.0485. This analysis indicated that areas of high AML cells density influenced the mean number of T cells. In order to quantify the strength of relationships between cell counts of cells of different types, however, a statistical model is needed.

To quantify the spatial relationship between T cell, MGK and AML cell counts, we employed the 3D geographically weighted regression (GWR) model. The objective of a regression model is to determine to what extent the value of response variables, such as counts of one cell type, can be explained in terms of explanatory variables, such as the cell counts of potentially related cell types, elucidating the relationship between the two. In GWR, the additional element is that the relationship can have a geographically varying dependence. Here, the T cell count per cube was treated as the response variable, while location-dependent AML and MGK counts were used as explanatory variables. GWR models are parameterised by a spatial scale, called the optimal bandwidth, that is algorithmically determined by the data and corresponds to the maximum extent of geographic influence. For these data, that value was 453 μm, which corresponds to the width of a little over ten cubes. For models with no spatial component, such as linear regression, the R^2 statistic is a common measure of the quality of the model

description of the data. For the spatial model GWR, the equivalent location-dependent statistic is called the local R^2. In linear regression, the coefficient of an explanatory variable is the best-fit linear multiplier that predicts the response variable given the best offset. Similarly, the GWR coefficients of location-dependent AML and MGK counts informed the multiplicative relationship between the number of T cells in a cube, and the number AML and MGK cells contained in cubes within a $453\mu m$ range. The sign and magnitude of those coefficients captured the nature and strength of the relationship between T cell counts, and AML and MGK counts nearby. By observing how the coefficients change in a spatial context, information was extracted about geographically dependence in those relationships.

For four z-levels, Fig. 7 (a)-(d) plot the GWR coefficient of the AML cell count as a function of spatial location. Compared with the representation in Fig. 6a, locations with zero or negative coefficient largely coincided with the four largest AML clusters. his provided quantitative evidence that areas with high AML density were devoid of T cells, while areas with small AML cluster did not have such a strong effect on T cell density.Most variation occurred in the x-y plane, with little spatial variation in the z direction, which, for this bone, has the smallest extent and almost complete coverage by the large AML clusters. Fig. 7 (e)-(h) indicated a strong positive relationship between MGK and T cells in peripheral areas of the bone marrow. Coefficients in middle areas of the bone marrow cavity were negative, which can be explained by the apparently greater ability of MGKs to resist exclusion by AML cells. Fig. 7 (i)-(m) report the local R^2 values, providing a spatial understanding of how well GWR explains the data. Matching with intuition, the model identified areas of increased AML density where the influence on T cell became prominent.

Taken together, the analysis of these data illustrates how methods from spatial statistics can be adapted to a 3D framework to enable the quantitative evaluation of clustering, the automatic identification of regions of high density, and the statistical assessment of dependencies between cell types.

Discussion

In this study we have described a novel and effective pipeline, PACESS, for extracting meaning and drawing inferences from complex biological images. The method builds on established techniques from machine learning and geospatial statistics and has a number of advantages over current approaches. From the perspective of data-extraction, the technique is robust even in cases with compacted, adjacent cells, which is a major benefit of using neural networks over segmentation based methods[10,34,35]. The technique is also scalable, can reliably identify cells with weak fluorescence intensities, and, once trained, the model can be shared or updated as needed[10]. The approach avoids the need for extensive image processing as images are provided to the model in their raw form. Image transformations, such rotation or artificial introduction of additional noise, are only used in the training phase[13] to improve the neural network's performance when applied to unaltered experimental sample images[35].

As an innovation that makes 3D data extraction practical, we use a neural network on each 2D sample to make predictions that are then grouped into 3D predictions, obviating the need for labelled 3D data. By mapping our final predictions against the baseline image, we can visually observe that estimates are a faithful representation of the underlying cellular data. What is more, the segmentation-based approaches have no quantitative measures of accuracy, interim or not, and rely solely on visual cross-checking[1,4,5].

A known drawback of AI-based approaches is the need to generate manually annotated training data. In this study, models were trained on hundreds of images with thousands of annotated cells. Once trained, however, neural network models can quickly and efficiently identify any number of cells in any number of similar images, unlike manual annotation[8]. Moreover, pre-trained models can be updated with less extensive data-sets through a process called transfer learning[34,35]. This, along with the increasing availability of imaging data, has the potential to greatly expand the availability of 'pre-trained' models to the wider scientific community, thus reducing the burden on researchers to establish extensive training datasets for their particular setting[36]. The dataset presented here is already a pre-trained model and can constitute the starting point for identifying cell types in virtually any tissue.

In previous studies, researchers have assessed whether the distribution of individual cell populations differs significantly from that of randomly distributed dots using hypothesis testing[5,7,9]. In a spatial analysis context, the simulated null must be conditioned to take values only within the observable, imaged, space. Although this method is intuitive, there is no guarantee that the simulated null is physiologically viable. That issue arises as we typically cannot observe all features within the tissue with current technology, which introduces a risk of bias in the findings. In our spatial statistical pipeline, no simulated data is used, and the question of how cells within the bone marrow are related is addressed directly by comparisons between multiple different cell populations.

Our spatial statistics approach provides a reproducible method for quantifying properties of individual cell types, as well as interactions between them. Simple measures quantify the extent of spatial clustering of individual cell types, while geographically-aware clustering methods can then identify regions of high density. Permutation tests provide statistical measures for assessing the relatedness of cell densities in reasons of interest. Geographically Weighted Regression can then quantify the relationships between multiple cells types simultaneously. At its core, it is a form of regression and the output is a series of 3D maps which describe how cellular coefficients vary between different cell populations across space. It identifies areas in which spatial effects, such as those generated by an expanding malignant infiltrate, have the most influence on the presence of other cell types. It adjusts coefficients to take account the presence of cellular interactions. Finally it enables predictions to be made that can be used to provide a measure of confidence in the estimates. In our worked example, we regressed the location-dependent number of $CD8^+$ T cells against the number of AML cells and MGK cells. The interaction between $CD8^+$ T cells and leukaemia has been described previously[37,38]. We observed a largely positive association between these two classes, except in the areas where the dense AML infiltrate generates such a significant spatial effect that it displaces almost all cells, consistent again with previous observations[39]. MGK cells are largely immobile cells and in this model they act as a surrogate for tissue cellularity[7]. For MGK cells there is a markedly different coefficient pattern, with a largely negative coefficient running up the central axis of the bone marrow cavity. This is likely to reflect the increased number of MGK cells in this central 'endothelial' niche[7], and illustrates a single model's ability to characterise multiple interaction types.

The ultimate goal of quantitative 3D imaging is to informatively summarise, in a numerical form, the vast amount of spatial and cellular information present within an image. The methods used for this type of work have been evolving over the course of the last few years[1,4,5,40]. The work presented here provides a framework for the data extraction and analysis of complex 3D biological images. This pipeline builds upon the work that has gone before but expands its utility to include those images which would have proved more challenging to standard methodologies. We also introduced an analytical paradigm to aid in the interpretation of spatial cellular data which is reproducible, scalable, and capable of providing a comprehensive insight of the relationships which exist within the tissue.

Usage Notes

All code used for this study is available at `https://github.com/ga402/PACESS`.

References

1. Acar, M. *et al.* Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. *Nature* **526**, 126–130 (2015).
2. Richardson, D. S. & Lichtman, J. W. Clarifying tissue clearing. *Cell* **162**, 246–257 (2015).
3. Coutu, D. L., Kokkaliaris, K. D., Kunz, L. & Schroeder, T. Three-dimensional map of nonhematopoietic bone and bone-marrow cells and molecules. *Nat. Biotechnol.* **35**, 1202–1210 (2017).
4. Li, W., Germain, R. N. & Gerner, M. Y. Multiplex, quantitative cellular analysis in large tissue volumes with clearing-enhanced 3D microscopy (Ce3D). *Proc. Natl. Acad. Sci. U. S. A.* **114**, E7321–E7330 (2017).
5. Coutu, D. L., Kokkaliaris, K. D., Kunz, L. & Schroeder, T. Multicolor quantitative confocal imaging cytometry. *Nat. Methods* **15**, 39–46 (2018).
6. Matsumoto, K. *et al.* Advanced CUBIC tissue clearing for whole-organ cell profiling. *Nat. Protoc.* **14**, 3506–3537 (2019).
7. Stegner, D. *et al.* Thrombopoiesis is spatially regulated by the bone marrow vasculature. *Nat. Commun.* **8**, 127 (2017).
8. Zhang, J. *et al.* In situ mapping identifies distinct vascular niches for myelopoiesis. *Nature* **590**, 457–462 (2021).
9. Kokkaliaris, K. D. *et al.* Adult blood stem cell localization reflects the abundance of reported bone marrow niche cell types and their combinations. *Blood* **136**, 2296–2307 (2020).
10. Aloysius, N. & Geetha, M. A review on deep convolutional neural networks. In *IEEE ICCSP*, 588–592 (2017).
11. Xu, J. *et al.* Convolutional neural network initialized active contour model with adaptive ellipse fitting for nuclear segmentation on breast histopathological images. *J. Med. Imaging* **6** (2019).
12. Pansombut, T., Wikaisuksakul, S., Khongkraphan, K. & Phon-on, A. Convolutional neural networks for recognition of lymphoblast cell images. *Comput. Intell. Neurosci.* e7519603 (2019).
13. Weigert, M., Schmidt, U., Haase, R., Sugawara, K. & Myers, G. Star-convex polyhedra for 3d object detection and segmentation in microscopy. In *IEEE WACV*, 3655–3662 (2020).

14. Fotheringham, A. S., Brunsdon, C. & Charlton, M. *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships* (John Wiley & Sons, 2003).

15. Morrison, S. J. & Scadden, D. T. The bone marrow niche for haematopoietic stem cells. *Nature* **505**, 327–334 (2014).

16. Lo Celso, C., Lin, C. P. & Scadden, D. T. In vivo imaging of transplanted hematopoietic stem and progenitor cells in mouse calvarium bone marrow. *Nat. Protoc.* **6**, 1–14 (2011).

17. Carrelha, J. *et al.* Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature* **554**, 106–111 (2018).

18. Krivtsov, A. V. *et al.* Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia* **27**, 852–860 (2013).

19. Akinduro, O. *et al.* Proliferation dynamics of acute myeloid leukaemia and haematopoietic progenitors competing for bone marrow space. *Nat. Commun.* **9**, 519 (2018).

20. Hawkins, E. D. *et al.* T-cell acute leukaemia exhibits dynamic interactions with bone marrow microenvironments. *Nature* **538**, 518–522 (2016).

21. Celso, C. L. *et al.* Intravital microscopy reveals fundamental differences in the interaction of stem cells and T acute lymphoblastic leukaemia with the bone marrow microenvironment. *Blood* **128**, 5199–5199 (2016).

22. Shafiee, M. J., Chywl, B., Li, F. & Wong, A. Fast YOLO: A fast you only look once system for real-time embedded object detection in video. *arXiv:1709.05943* (2017).

23. Kharchenko, V. & Chyrka, I. Detection of airplanes on the ground using YOLO neural network. In *IEEE MNET*, 294–297 (2018).

24. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *IEEE CVPR*, 779–788 (2016).

25. Bodla, N., Singh, B., Chellappa, R. & Davis, L. S. Soft-NMS–improving object detection with one line of code. In *IEEE ICCV*, 5561–5569 (2017).

26. Espindola, G., Câmara, G., Reis, I., Bins, L. & Monteiro, A. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation. *Int. J. Remote. Sens.* **27**, 3035–3040 (2006).

27. Schubert, E., Sander, J., Ester, M., Kriegel, H. P. & Xu, X. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM TODS* **42**, 1–21 (2017).

28. Lehmann, E. L. & Lehmann, E. *Testing Statistical Hypotheses*, vol. 2 (Springer, 1986).

29. Gollini, I., Lu, B., Charlton, M., Brunsdon, C. & Harris, P. GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *J. Stat. Softw.* **63**, 1–50 (2015).

30. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. Learning deep features for discriminative localization. In *IEEE CVPR*, 2921–2929 (2016).

31. Arrieta, A. B. *et al.* Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. fusion* **58**, 82–115 (2020).

32. Baccin, C. *et al.* Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. *Nat. Cell Biol* **22**, 38–48 (2020).

33. Lamble, A. J. *et al.* Reversible suppression of T cell function in the bone marrow microenvironment of acute myeloid leukemia. *Proc. Natl. Acad. Sci.* **117**, 14331–14341 (2020).

34. Chanu, M., Lourembam, R. & Thonga, K. A deep learning approach for object detection and instance segmentation using mask RCNN. *J. Adv. Res. Dyn. Control. Syst.* **12**, 95–104 (2020).

35. Lin, T.-Y. *et al.* Feature pyramid networks for object detection. In *IEEE CVPR*, 2117–2125 (2017).

36. Hartley, M. *et al.* The BioImage archive – building a home for life-sciences microscopy data. *J. Mol. Biol.* **434**, 167505 (2022).

37. Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**, 541–550 (2018).

38. Knaus, H. A. *et al.* Signatures of CD8+ T cell dysfunction in AML patients and their reversibility with response to chemotherapy. *JCI Insight* **3**, e120974 (2018).

39. Duarte, D. *et al.* Inhibition of endosteal vascular niche remodeling rescues hematopoietic stem cell loss in AML. *Cell Stem Cell* **22**, 64–77.e6 (2018).

40. Messal, H. A. *et al.* Antigen retrieval and clearing for whole-organ immunofluorescence by FLASH. *Nat. Protoc.* **16**, 239–262 (2021).

Acknowledgements

We thank staff of the core facilities at Imperial college London (CBS facility, FILM) and the Crick Institute (BRF facility) for their valuable help. This work was supported by Cancer Research UK (Programme Foundation Award C36195/A26770 to C.L.C.), the Wellcome trust (Investigator Award 212304/Z/18/Z to C.L.C. and 4i Clinical PhD studentship 203928/Z/16/Z to G.A.), NIHR (Academic Clinical Lectureship to G.A.), C.L.C and K.R.D were supported in part by the Royal Irish Academy - Royal Society International Exchange Program grant IEC\R1\180061 to K.R.D and C.L.C. This publication has emanated from research conducted with the financial support of Science Foundation Ireland under grant number 18/CRT/6049. We also acknowledge the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. The opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NHS, the NIHR, the Department of Health or Science Foundation Ireland.

Author contributions statement

G.A. and C.L.C. conceived the project. G.A., F.T. and C.L.C. refined bone immunofluorescence methods. G.A. and F.T. performed imaging experiments. G.A. labelled data, trained the neural network, and performed the data extraction and classification. C.L, G.A., C.B. and K.R.D. devised the spatial statistical framework. C.L. performed the statistical analysis. All authors contributed to writing the manuscript.

Competing interests

The authors declare no competing financial interests.

Figures

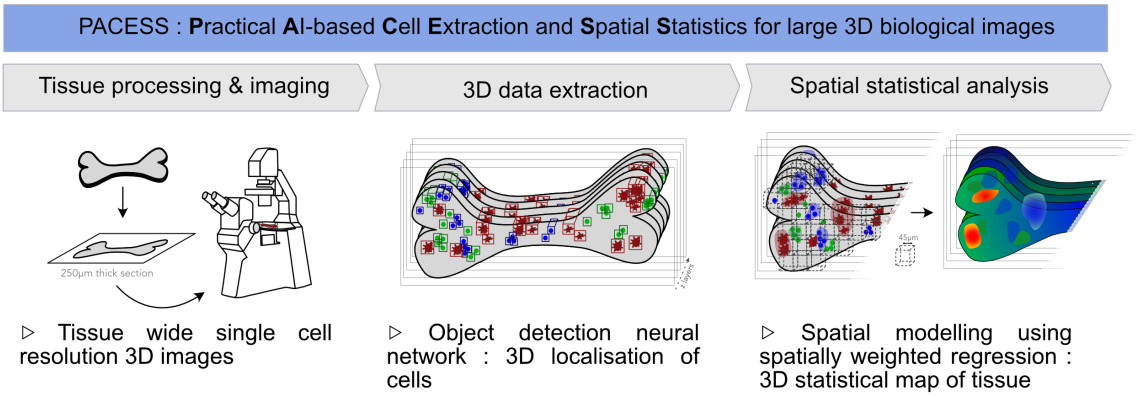

Figure 1. Summary of the PACESS pipeline.
The method consists of three steps: 1) tissue processing and imaging (in-vivo or ex-vivo), 2) 3D data extraction using object detection neural network trained on 2D data and scale to 3D, and 3) spatial statistical analysis consisting in spatial inhomogemeity quantification, automatic identification of areas of high cellular density and geographical weighted regression.

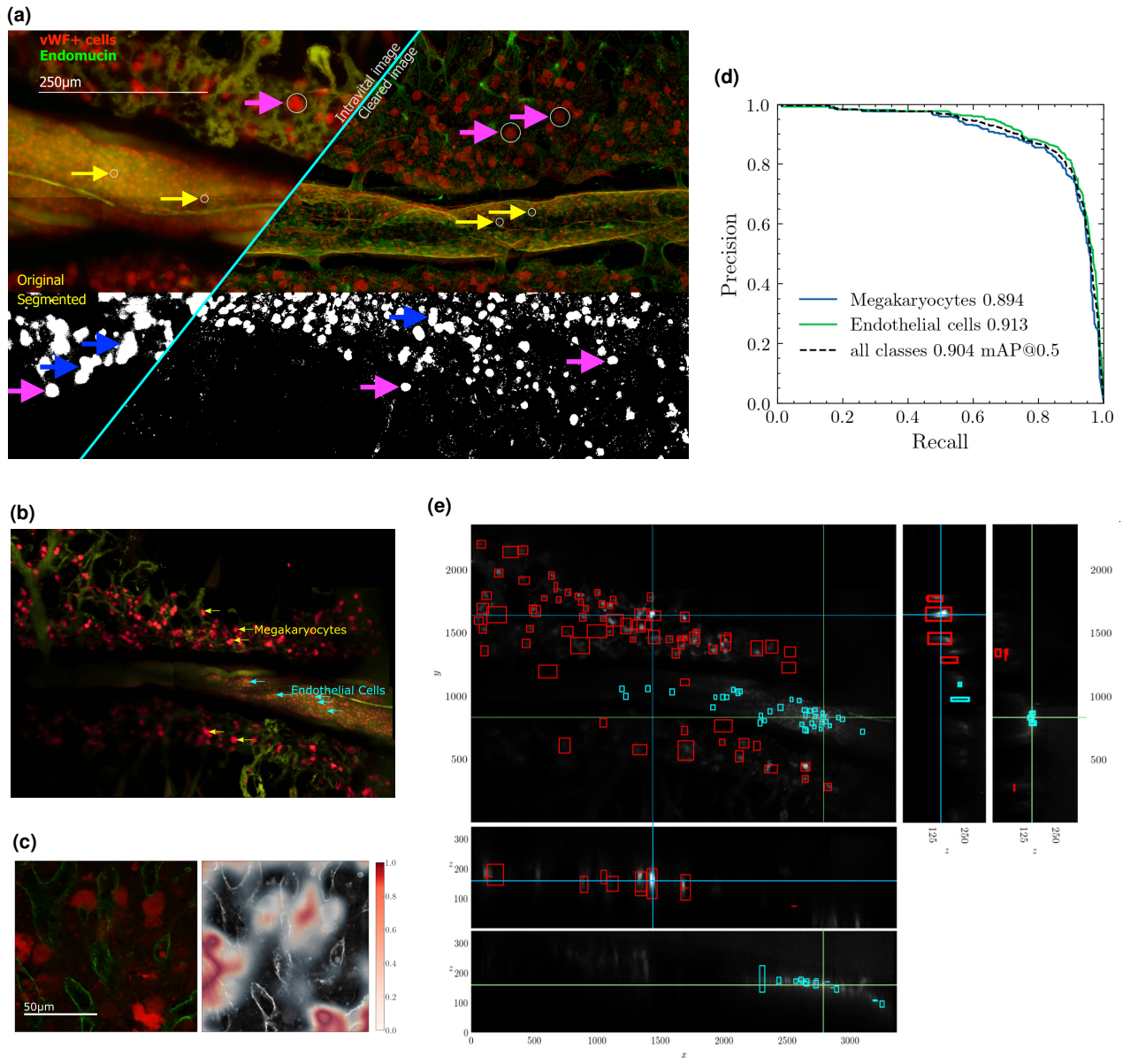

Figure 2. 3D data extraction using a deep neural network trained on 2D image-layer.
(a) Comparison of IVM/cleared images and thresholding segmentation. Mouse calvarium imaged using intravital microscopy (left) then cleared and re-imaged ex-vivo (right). Magenta arrows : megakaryocyte (MGK) cells, yellow arrows: endothelial cells, blue arrows: overlapping cells. The white circles highlight the difference in size between the two cell types. Both cell types are labelled by vWF:dTomato (red). The results of thresholded binary segmentation are shown in the bottom 1/3rd of the image. Within the cleared image, the contours of individual MGK cells are clearly visible (*top right colour image, magenta arrows*). In the IVM image, the boundaries of the cells are less distinct (*top left*). In the thresholded (segmented) image it is not always possible to reliably distinguish individual cells: in both images when cells overlap, or are in close proximity, the cell contours merge (blue arrows). **(b-e)** The use of 2D output to generate 3D reconstructions from an IVM image of the calvarium of a VWF:dTomato transgenic mouse. **(b)** The original image is shown as a maximum projection in the z dimension. A 2D object detection neural network was applied to each layer of the 3D original image. An example saliency map and the precision-recall curve from this interim 2D model are shown in **(c)** and **(d)**, respectively. The final step was to generate a set of 3D predictions, shown in **(e)**. A collection of 3D bounding 'cubes' are shown from each image plane (x, y, z). The red boxes display the 3D results predicting MGK cells, while the smaller blue boxes are predictions of endothelial cells. Cross-sectional lines represent the position of each corresponding plane within each image.

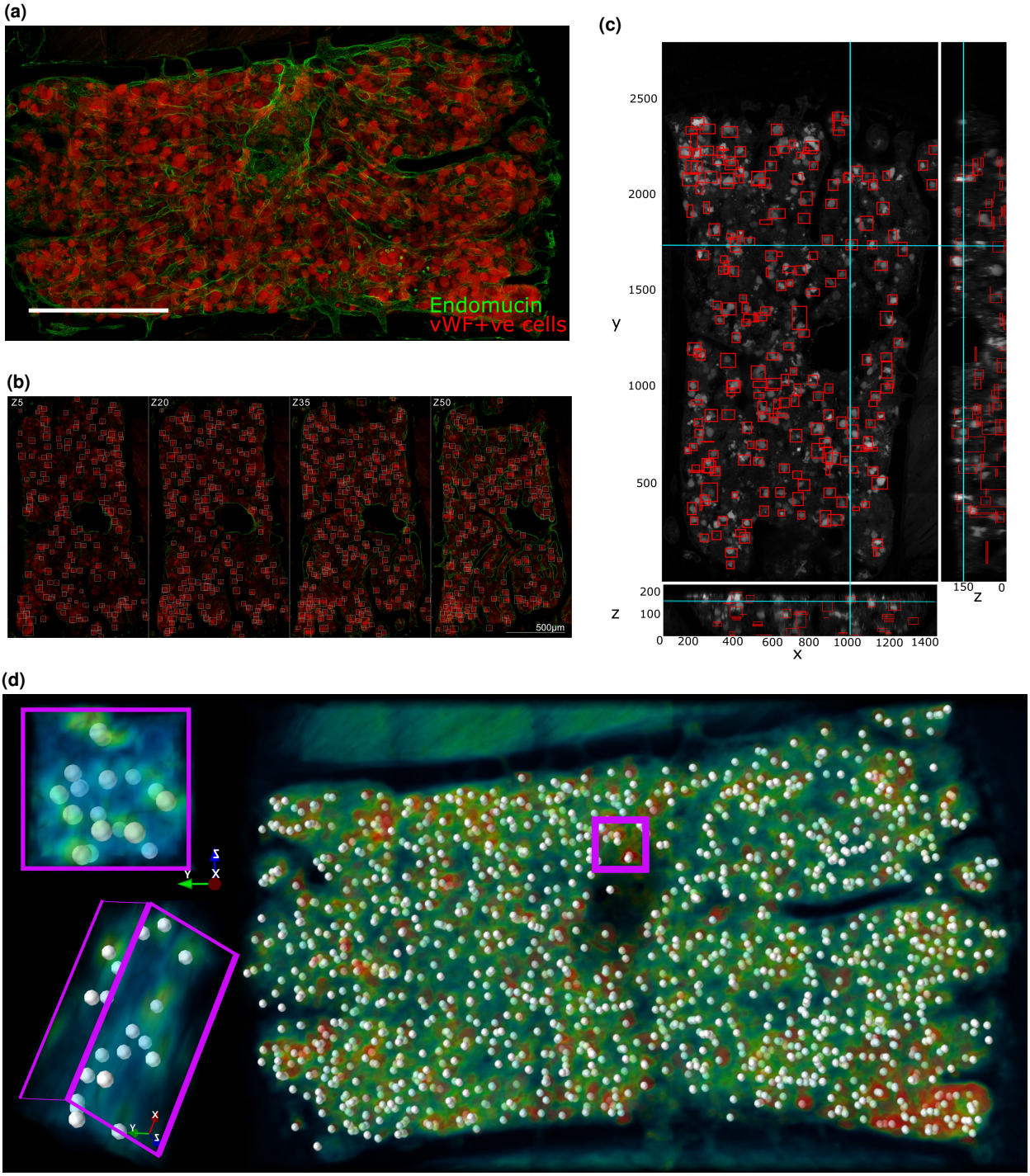

Figure 3. Illustration of the steps in the data extraction.
Example of optically cleared sternum from a vWF:dTomato transgenic mouse. **(a)** Maximum projection of densely packed tissue showing the overlapping nature of cells. Scale bar 400μm. The predictions from the 2D objection detection neural network are then shown in **(b)**, with the final 3D predictions shown in **(c)**. Both **(b)** and **(c)** demonstrate that even overlapping cells are accurately distinguished. **(d)** presents a maximum *intensity* projection upon which the cell locations are shown as white dotes. The *inset* to **(d)** shows a smaller ($50 \times 50 \times 225\mu$m) sections presented in frontal and auxiliary perspectives. These show the close association between the estimates and true positions of cells within the 3D space.

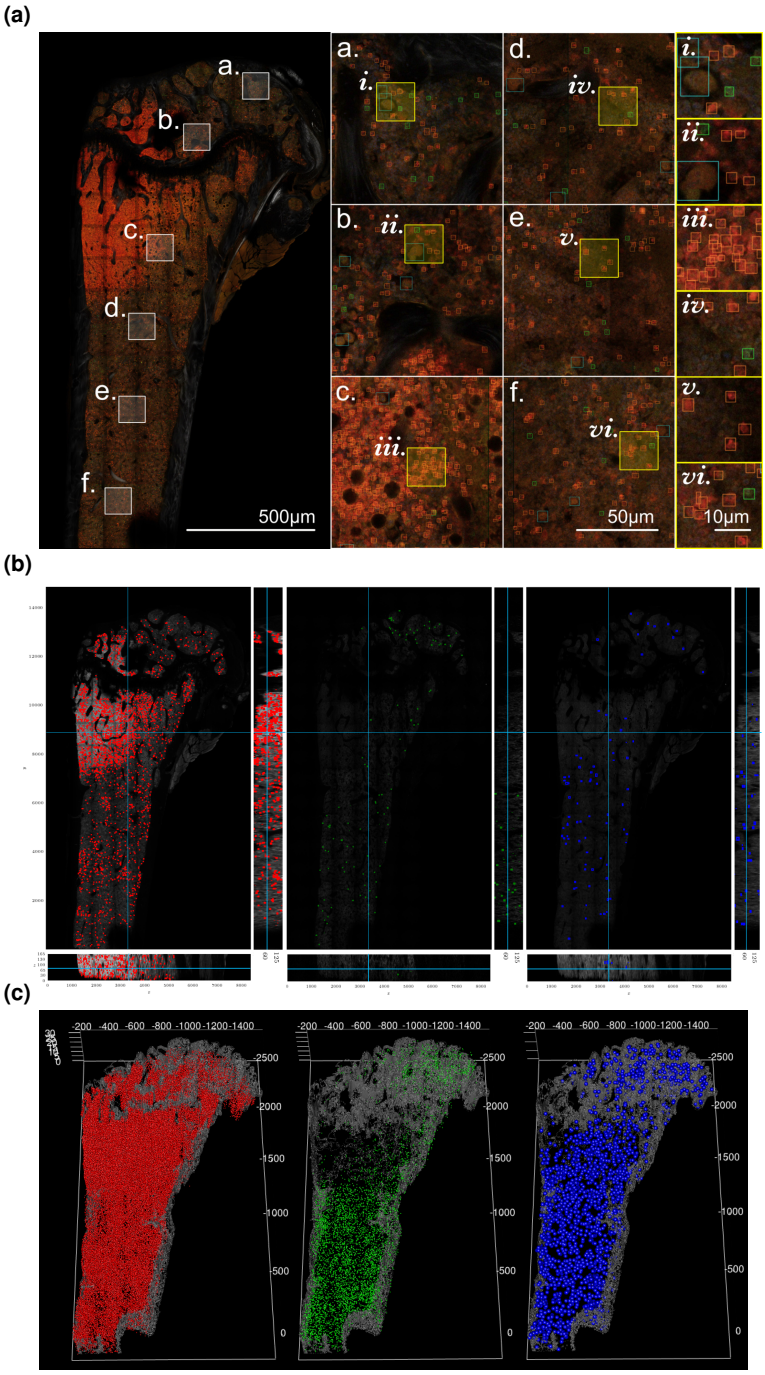

Figure 4. Final prediction of the data-extraction process and 3D reconstruction of leukaemic bone marrow. **(a)**. AML cells (red), CD8$^+$ cells (green) and MGK cells are identified by the model using red, green and blue boxes respectively. MGK cells are identified by their characteristic morphological features. Within each panel, predictions are shown at increasing magnification projected against the original image (far left). The results demonstrate the model accurately identifies cells, even in areas of high density (iii., vi.). **(b)** 3D bounding boxes for AML cells (red), CD8$^+$ T cells (green) and MGK cells (blue), respectively. The locations of the boxes in a set of x, y and z planes is shown. **(c)** The central location of each cell is shown within the 3D contours of the bone marrow cavity space.

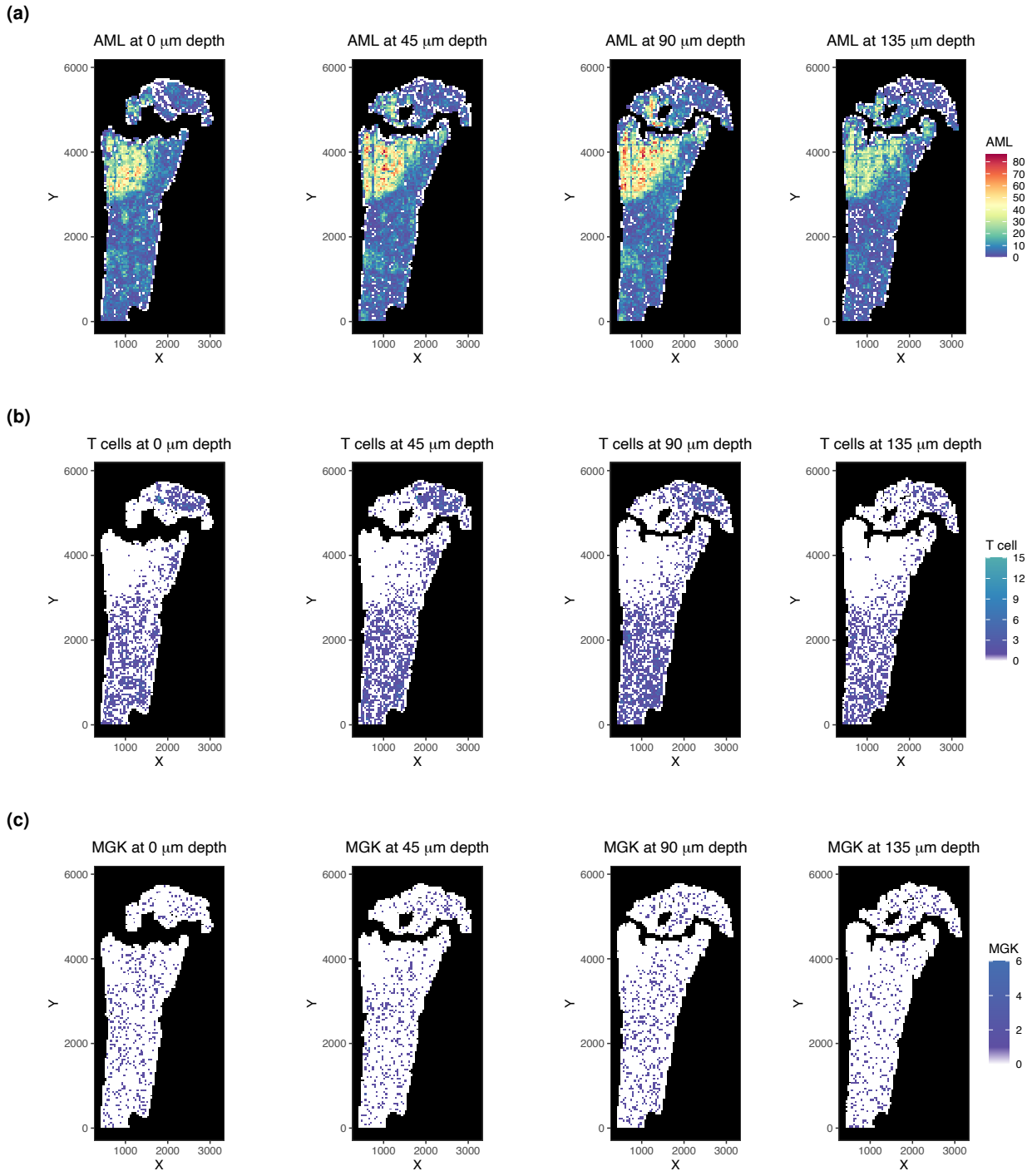

Figure 5. Quantitative 2D projection visualisation.
Quantitative Data are aggregated into $45\mu m^3$ cubes, at a range of z-depths heat maps showing the coordinates and the number of **(a)** AML cells, **(b)** T cells and **(c)** MGK cells. The colour scale indicates the density of cells.

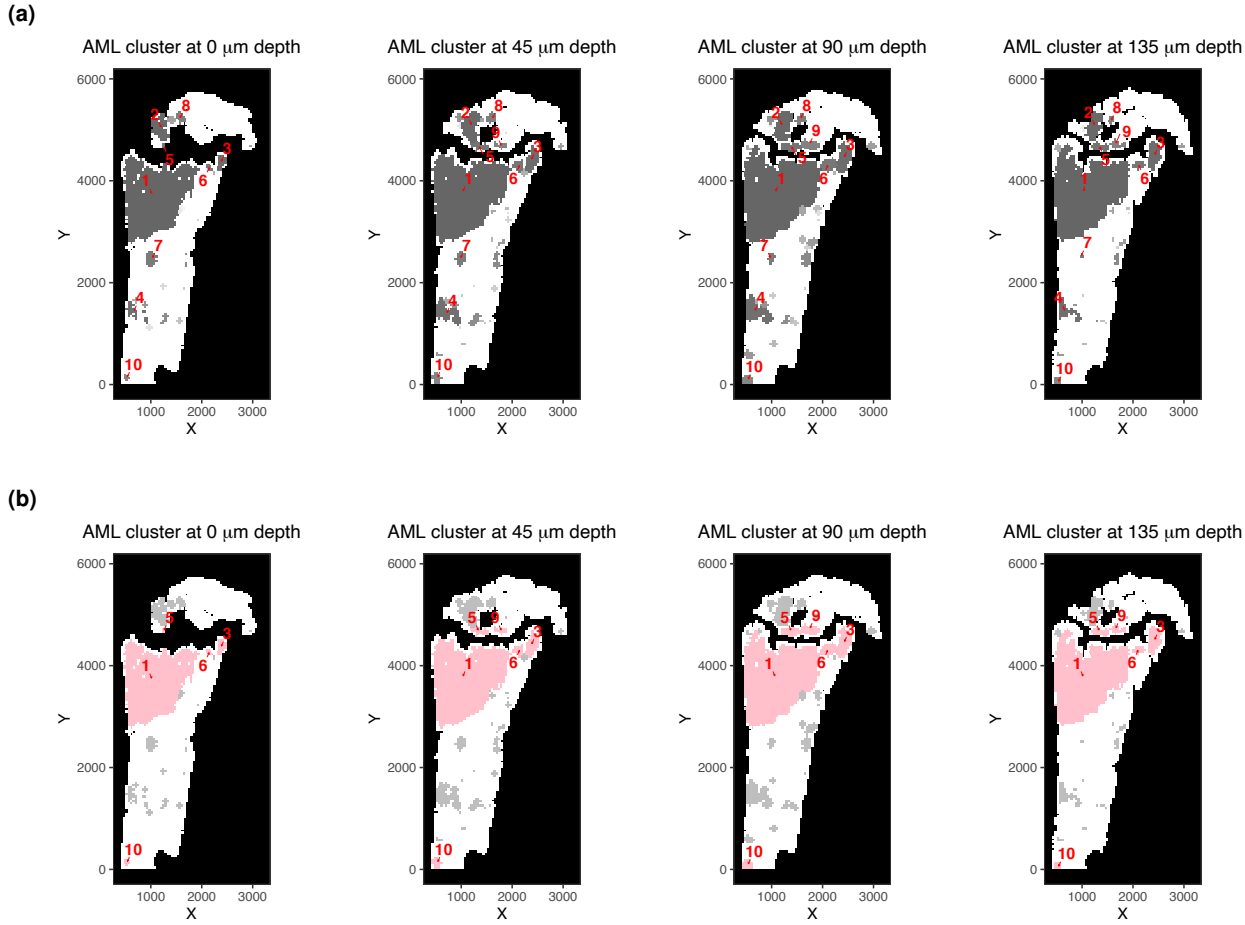

Figure 6. Cluster identification and cell density dependence hypothesis tests.
(a) 2D projection visualizations at different z-depths showing AML clusters identified using DBSCAN. The ten largest clusters are numbered in decreasing order of total AML cell numbers, with lighter colours indicating fewer AML cells in the cluster. **(b)** 2D projection visualizations showing hypothesis test results at different z-depths. The pink areas indicate AML clusters where the mean number of T cells within and around the cluster are different with statistical significance ($p \leq 0.05$), with the value reported for each in the text. Grey areas mark clusters where there is no statistical significance ($p > 0.05$). White areas indicate area where AML cells are not clustered.

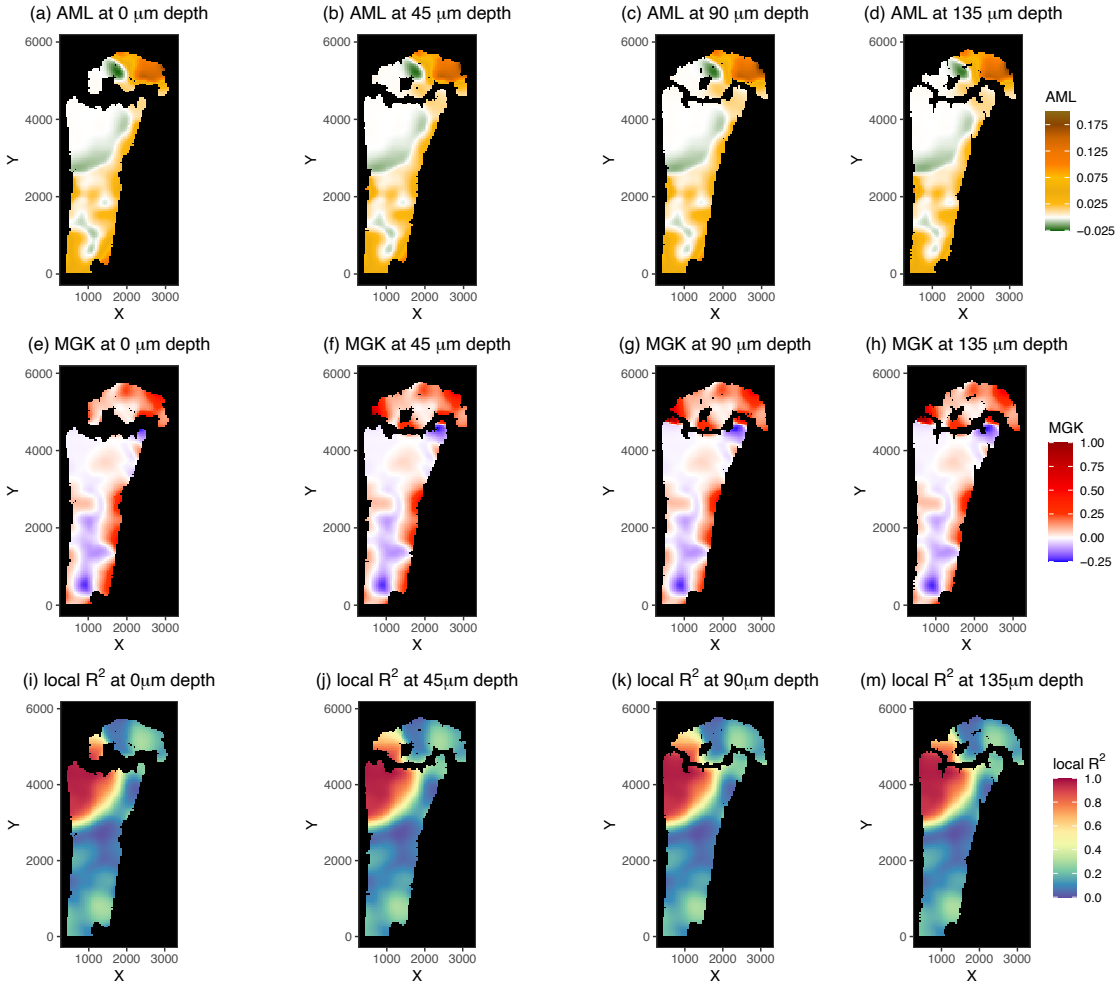

Figure 7. 3D statistical map of leukemic bone.
Projection visualisation for 3D spatial model results of coefficient of acute myeloid leukemia cells **(a-d)** and MGK cells **(e-h)** in different vertical layers. The summary statistic (i-m), local R^2, describes the model quality in each $45\mu m^3$ cube. Note that the coefficient scale of AML cell counts is smaller than the MGK cells one because the maximum number of AML cells in a cube is about 15 times the one of MGK cells, resulting in different coefficient scales.