# Deep learning enables spatial mapping of the mosaic microenvironment of myeloma bone marrow trephine biopsies

**Yeman B Hagos[1†]; Catherine SY Lecat[2†]**; Dominic Patel[3]; Anna Mikolajczak[2]; Simon P Castillo[1]; Emma J Lyon[2]; Kane Foster[2]; Thien-An Tran[2]; Lydia SH Lee[2]; Manuel Rodriguez-Justo[3]; Kwee L Yong[2*]; Yinyin Yuan[1,4*]

[1]Centre for Evolution and Cancer and Division of Molecular Pathology, The Institute of Cancer Research, London, U.K.

[2]University College London Cancer Institute, Research Department of Haematology, London, U.K.

[3]University College London Cancer Institute, Research Department of Pathology, London, U.K.

[4]Centre for Molecular Pathology, Royal Marsden Hospital, London, U.K.

[†] = joint first authors

* = corresponding authors

**Corresponding authors' contact detail:**

- Prof Yinyin Yuan, Email: yyuan6@mdanderson.org, Tel: (+1)346 722 9360, Address: 1515 Holcombe Blvd, Houston, TX 77030, USA
- Prof Kwee Yong, Email: kwee.yong@ucl.ac.uk, Tel: (+44)20 3447 8028, Address: 72 Huntley St, London WC1E 6DD

**Running Title**: Spatial mapping of bone marrow trephine biopsies

# Conflicts of interest

The funders had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

Y.Y. has received speakers bureau honoraria from Roche and consulted for Merck and Co Inc.

The authors declare that they have no other conflicts of interest.

# Abstract

Bone marrow trephine biopsy is crucial for the diagnosis of multiple myeloma. However, the complexity of bone marrow cellular, morphological, and spatial architecture preserved in trephine samples hinders comprehensive evaluation. To dissect the diverse cellular communities and mosaic tissue habitats, we developed a superpixel-inspired deep learning method (MoSaicNet) that adapts to complex tissue architectures and a cell imbalance aware deep learning pipeline (AwareNet) to enable accurate detection and classification of rare cell types in multiplex immunohistochemistry images. MoSaicNet and AwareNet achieved an area under the curve of >0.98 for tissue and cellular classification on separate test datasets. Application of MoSaicNet and AwareNet enabled investigation of bone heterogeneity and thickness as well as spatial histology analysis of bone marrow trephine samples from monoclonal gammopathies of undetermined significance (MGUS) and from paired newly diagnosed and post-treatment multiple myeloma. The most significant difference between MGUS and newly diagnosed multiple myeloma (NDMM) samples was not related to cell density but to spatial heterogeneity, with reduced spatial proximity of BLIMP1$^+$ tumor cells to CD8$^+$ cells in MGUS compared with NDMM samples. Following treatment of multiple myeloma patients, there was a reduction in the density of BLIMP1$^+$ tumor cells, effector CD8$^+$ T cells, and T regulatory cells, indicative of an altered immune microenvironment. Finally, bone heterogeneity decreased following treatment of MM patients. In summary, deep-learning based spatial mapping of bone marrow trephine biopsies can provide insights into the

cellular topography of the myeloma marrow microenvironment and complement aspirate-based techniques.

**Keywords**: Multiple myeloma; bone marrow trephine; deep learning; spatial analysis

**Significance:** Spatial analysis of bone marrow trephine biopsies using histology, deep learning, and tailored algorithms reveals the bone marrow architectural heterogeneity and evolution during myeloma progression and treatment.



# Introduction

Multiple myeloma (MM) is an incurable hematological malignancy characterized by the uncontrolled proliferation of abnormal plasma cells in the bone marrow (BM) [1][2][3]**.** According to the International Myeloma Working Group, the current diagnosis of MM is based on the demonstration of clonal neoplastic plasma cells and organ dysfunction, of which the most common is bone destruction, which is typically investigated by BM aspirate, trephine biopsy samples, and whole-body non-invasive imaging [4].

Increasingly, there is growing appreciation that myeloma is not driven by malignant plasma cells in isolation, but tumor growth is accompanied by global immune dysregulation in MM [5][6]. These include impaired T cell effector function [7] and antigen presentation [8] and an increase in suppressor cells such as regulatory T cells (Tregs) [9][10][11]. Our previous work showed that MM patients who had high Tregs had shorter progression-free survival [11]. In addition, analysis of $CD4^+$ and $CD8^+$ effectors revealed that a low $CD4^+$ effector to Tregs ratio was an independent predictor of early relapse [11]. However, these studies were based on MM blood/BM aspirates or MM cell lines employing flow cytometry and gene expression analysis, and not using biopsies that preserve the architecture of the BM. Therefore, the spatial relationship between BM cell types in MM has not yet been studied.

Deep learning methods, specifically convolutional neural networks (CNNs), have been shown to accurately identify complex visual patterns in histopathology images without handcrafted features [12][13]. This offers a unique opportunity to harness the cellular and non-cellular mosaic spatial ecology of BM [12][14]. However, the unique tissue integrity and morphology of BM trephine samples are very different from those of solid tumors due to its specialized sampling process and its requirement for decalcification (**Sup. Figure 1A**). The BM also has a highly organized structure, being a specialized hemopoietic and immunological organ. Thus, the BM is one of the priming sites of T cells and contains both rare and abundant cell types (**Sup. Figure 1B**) [15]; the spatial context of cell-to-cell interactions is likely to be crucially important in the development of immunity. Deep learning methods are often sensitive to the biases in the data unless carefully designed. Thus, there are new challenges in the development of reliable automated analysis for BM trephine samples due to possible biases in cell abundance and tissue architecture complexity.

In this study, we propose new deep learning-based image analysis pipelines addressing these challenges: **1)** to dissect the mosaic tissue microenvironment of BM trephine samples (MoSaicNet) and accurately identify immune T and MM plasma cells (AwareNet) on multiplex immunohistochemistry (MIHC) images; **2)** to harness the morphologic features of bone trabeculae in MGUS, diagnostic, and post-treatment MM facilitating new understanding of bone physiology; **3)** to analyse cell density, infiltration pattern and spatial topography of immune T and MM plasma cells facilitating understanding of the cellular topography in the BM niche of MGUS, diagnostic and post-treatment MM samples.

# Materials and methods

## Patients studied

All patients were managed at University College London Hospital (UCLH). BM trephine biopsies from two cohorts of patients were extracted: 11 patients with MGUS and 14 patients with MM. Two patient samples from the MGUS group and four patient samples from the MM group were excluded due to suboptimal tissue samples (small areas of hematopoietic tissue), leaving nine patients with MGUS and

ten patients with MM included in this study. For the second group, we studied newly diagnosed MM (NDMM) patients prior to treatment initiation and also post-treatment, when BM biopsies were taken at 100 days following Autologous Stem Cell Transplant (ASCT). All patients provided written informed consent for this project. Ethical approval was granted by the Health Research Authority, U.K. (Research ethics committee reference: 07/Q0502/17).

Patient characteristics for the MGUS group are shown in **Table 1**. The median age was 61 years, and 56% were male. The majority had IgG MGUS (56%), three had IgA MGUS (33%), and one had kappa light chain MGUS (11%). Five patients (56%) were deemed to have a low risk of MM progression, whilst two (22%) had intermediate risk, and two (22%) had a high risk [16].

The characteristics of the ten patients in the MM group are described in **Table 2**. The median age at MM diagnosis was 56 years, consistent with an age group that would usually proceed with treatment following induction therapy. Six (60%) patients were male, five had IgG disease (50%), and half had standard cytogenetic risk by IMWG criteria. Four patients (40%) had ISS stage I disease, five (50%) had stage II, and one (10%) had stage III [17]. All patients received combination induction therapy with a proteasome inhibitor, cyclophosphamide and dexamethasone, followed by Melphalan 200mg/$m^2$ as a conditioning regimen prior to ASCT.

## Tissue processing

BM samples were collected and processed as per ICSH guidelines [18]. They were first fixed in neutral buffered formalin and then decalcified with formic acid. After decalcification, biopsy specimens were embedded in paraffin wax and cut on a microtome at 2–3µm. Serial sections were cut and mounted on glass slides.

## Immunohistochemistry panel selection

Immune T cells play an active role in the disease's development and progression in MM. In this study, we aimed to analyse the density and the spatial topography of immune T and MM tumor cells in BM trephine biopsies. We chose CD4 and CD8 to

label effector T cells, FOXP3 to represent Tregs [19], and BLIMP1 to stain MM tumor cells [20-22].

The MIHC staining was performed using the fully automated Leica Bond RX$^{m}$ stainer. Each slide was serially stained to identify three different antigens using different membranous or nuclear stains. The details of antibodies used are in **Sup. Table 1**. Two MIHC multiplex panels were used in this study. Panel 1 included T cell markers CD4 and CD8, as well as FOXP3, a transcription factor specifically expressed by CD4$^{+}$ Tregs. Panel 2 comprised CD4, CD8 and BLIMP1. BLIMP1 is a nuclear stain and therefore allowed clear visualization when combined with CD4 and CD8 membranous stains. Staining protocols can be found in (**Sup. Table 2-3**). Stained slides were then scanned using the Hamamatsu Nanozoomer s360 scanner and analysed by the deep learning models.

## Pre-processing of whole slide images

The MIHC whole slide images (WSI) were scanned at 40x magnification with a pixel resolution of 0.23µm/pixel. A representative image has a 40000 x 40000 pixel size at 40x magnification. For efficient image processing, the images were downscaled to 20x magnification and divided into 2000 x 2000 pixel “tiles”.

## MoSaicNet: Segmenting BM trephine components using deep learning and superpixel

The digital image of the BM trephine is a mosaic landscape of blood, bone, cellular tissue, and fat region (**Sup. Figure 1A**). To automatically segment these regions, we developed **MoSaicNet** (**Mo**rphological analysis with **S**uperpixel-based h**ab**itat detection **Net**work) (**Figure 1A**). MoSaicNet contains superpixel extraction and a CNN-based superpixel classifier.

### MoSaicNet training and validation data preparation

To train, validate, and test MoSaicNet, we collected 260 regions of interest from 19 samples (**Sup. Table 4**) annotated by expert pathologists (**Sup. Figure 2A**) from the different regions of the images. The training (47%), validation (31%), and testing

(22%) split was randomly done at the patient level. These annotated regions were extracted from the WSIs and divided into superpixels using the simple linear iterative clustering (SLIC) superpixels algorithm [23] (**Figure 1A**). SLIC groups neighbouring pixels with similar pixel intensity into one superpixel. The shape of the superpixels is controlled by the compactness (C) parameter of the SLIC algorithm. The number of superpixels depends on the size of the images and the parameter *k* (Equation (1)) [23][24]. The parameters *C* and *k* are set by a user to ensure superpixels are capturing homogeneous pixels and bounding to region boundaries in the image under consideration depending on the scenario [23][24]. The number of superpixels (n) was computed using Equation (1).

$$n = \left\lceil \frac{\text{Image area}}{k} \right\rceil \quad (1)$$

Upon visual assessment, superpixels with *k*=2000 and *C*=30 best adhere to the boundaries of tissue and fat regions. This resulted in about 40 x 40 pixel (18.4µm x 18.4µm) sized superpixel regions (**Figure 1A**). After applying SLIC, we generated 69, 884 superpixels from the 260 regions (**Sup. Table 5**). These superpixels belonged to four classes: blood, bone, fat and cellular tissue. Each superpixel was assigned a class of the region it belongs. We implemented and trained a custom-designed convolutional neural network to automatically classify these superpixel regions (**Sup. Methods**).

# AwareNet: attention-based deep convolutional network for cell detection and classification

## Single-cell annotation

To train, validate and test our proposed deep learning-based single-cell detection and classification models, we first collected 8004 single-cell dot annotations on 11 samples by expert pathologists (**Sup. Figure 2A**), using a web-based annotation tool developed in our lab (not published). The annotations belonged to three classes: $CD8^+$ (n = 5103), $FOXP3^-CD4^+$ (n = 2381), and $FOXP3^+CD4^+$ (n = 518). We identified $FOXP3^+$ cells as rare because they represented only 6.5% of all annotated cells, despite histopathologists actively looking for them in the whole tissue instead

of only regions of interest. The training (46%), validation (27%) and test (27%) split was done randomly at the patient level to ensure that cells from the same patients are not included in different categories (**Sup. Table 6**).

## Cell detection and classification

To automatically localize cells in MIHC images, we developed AwareNet (**Figure 1B**). AwareNet is a deep learning method designed to give high attention to rare cell types such as $FOXP3^{+}CD4^{+}$ cells in the case of BM trephine samples. During model training, the attention score was inferred from the relative abundance of each cell type in the training data. A rare cell type was given a larger attention score. The mathematical formulation of attention image generation and usage during model training is detailed in [25].

AwareNet generates a predicted cell nucleus centre probability map image (**Figure 1B**) from which the spatial coordinates of the centre of the cell's nucleus are computed (detailed in **Sup. Methods**). To identify the type of the detected cell, we extracted a 28x28x3 patch centred on the cell nucleus (**Figure 1B**) and applied a custom-designed CNN classifier [25].

### Cell density

Cell density is measured as the number of cells per unit of tissue area ($\mu m^2$). Suppose a given tissue section has $\boldsymbol{N}$ cells and cellular tissue area of $\boldsymbol{A_T}$, cell density is computed using Equation (2).

$$Cell\ density = \frac{\mathrm{N}}{A_T} \qquad (2)$$

### Cells proximity analysis

We investigated the spatial proximity of a pair of cell types (e.g., $BLIMP1^{+}$ MM plasma cells and $CD8^{+}$ T cells) within the BM microenvironment as follows (**Figure 1C**). Consider a tissue section that contains $\boldsymbol{k}$ number of type A cells located at $\{a_i, i \in \{1, 2, 3, \ldots, k\}\}$ and $\boldsymbol{m}$ number of type B cells located at $\{b_j, j \in \{1, 2, 3, \ldots, m\}\}$. Each cell has an (x, y) position. The number of type B cells within a distance $\boldsymbol{r}$ from

type A cell was computed using Equation (3A and B).

$$N_{prox(b \to a)} = \frac{1}{k} \sum_{i=1}^{k} \frac{\sum_{j=1}^{m} \Omega}{\Phi_i} \quad (3A)$$

$$\Omega = \{1, \quad D(a_i, b_j) \leq r \; ; \; 0, \quad otherwise \quad (3B)$$

where *D* is the Euclidean distance function for two cells, $a_i$ and $b_j$. $\mathbf{\Phi_i}$ is a normalizing factor, which is the total number of cells (all types) within ***r*** distance from $\boldsymbol{a_i}$. In BM trephine samples, there is a huge variation in the tissue architecture caused by the prevalence of non-cellular regions such as bone and fat regions (**Sup. Figure 1A**). Moreover, in single-cell based spatial analysis, the density of cells could be a confounding factor. Incorporating $\mathbf{\Phi_i}$ corrects these factors.

## Validation cohort

Bone marrow trephine samples from a separate patient cohort were used to validate this deep learning pipeline. This cohort consisted of 9 NDMM pre- and post-treatment BM samples. Patient characteristics can be found in **Sup. Table 7**. These were collected from 7 different U.K. hospitals (1 from UCLH, 1 Kent & Canterbury Hospital, 2 Sunderland Royal Hospital, 1 Warwick Hospital, 1 Calderdale Royal Hospital, 2 Ninewells Hospital, 1 Huddersfield Royal Infirmary) and were stained with MIHC panel 2 (CD4, CD8 and BLIMP1) using the same staining protocol. A different autostainer of the same model was used. Whole slide images were scanned and underwent color normalization (**Sup. Method**) before analysis to adjust for tissue processing and staining variations.

# Bone density similarity and heterogeneity

To learn the low dimensional representation of bone superpixels, we custom-designed a convolutional auto-encoder (**Sup. Methods, Sup. Figure 2B**). For ease of visualization and applying unsupervised clustering algorithms on the representation of bone superpixels, we applied Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction.

Then, we applied a clustering algorithm to divide the latent representation space into smaller regions. Kmeans and Gaussian Mixture Models (GMM) are the most commonly used clustering algorithms. We applied GMM to detect bone superpixel clusters in the embedding space due to its flexibility to cluster shapes [26]. To determine the number of clusters, we used the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). We used the GMM algorithm and its built-in AIC and BIC methods from the Scikit-Learn python package [27]. A cluster contains superpixels with similar bone density/texture. The clustering enabled us to identify artefact bone superpixels with input from an expert pathologist (MR). These clusters were excluded from further analysis.

To quantify the heterogeneity (**H**) of bone texture within a slide, we computed the maximum variance (**Var**) of the latent representations of all superpixels within the slide using Equation (4).

$$H = \max(Var(Umap\ 1), Var(Umap\ 2)\ ) \quad (4)$$

## Automated machine learning algorithm to quantify bone thickness

The proposed method to quantify bone thickness is shown in **Figure 2A**. We extracted the bone regions from the output of MoSaicNet. To compute bone thickness for a given bone (***B***), first, we applied distance [28], and medial axis transforms [29] as shown in **Figure 2A**. The Distance transform (**DT**) computes the minimum distance from bone pixels to non-bone pixels. The medial axis transform (**MAT**) generates the topological skeleton of the bone, a series of bone pixels that have more than one closest equidistant non-bone pixel. The bone thickness ($\mathbf{B_{thickness}}$) for a given tissue sample was computed as the mean of the mean thicknesses of all bones within the sample using Equation (5).

$$B_{thickness} = \frac{1}{N}\sum_{i=1}^{N}\frac{2\ \sum \mathrm{DT}(B_i)\ \odot\ \mathrm{MAT}(B_i))}{L_i} \quad (5)$$

where **N** is the number of bones in the sample, and $\odot$ is elementwise matrix multiplication. $\mathbf{L_i}$ is the length (number of pixels) of the skeleton of the $i^{th}$ bone, $B_i$.The

distance values on the medial axis of the bone are half the thickness of the bone across its length. Thus, to get the total bone thickness, the distance was multiplied by 2, as shown in Equation (5).

## Spatial analysis

To quantify the degree of clustering or dispersion of cells in BM trephine samples, we used the concept of nearest neighbour distance (NND) and the null hypothesis to identify the infiltration pattern of cells (**Sup. Methods**). NND is the distance from a spatial point to its closest neighbour. Under the null hypothesis, which is complete spatial randomness (CSR), the distribution of NND is normal. (**Figure 2B**). We computed the Z-score to measure the difference between the NND for random distribution of cells and the NND of observed cells pattern. $Z<-1.96$, $Z>1.96$, and $-1.96\leq Z\leq 1.96$ indicate a clustered, dispersed and random distribution of observed cells respectively.

## Statistical analysis

All statistical analyses were carried out using the Python programming language. All correlation values were measured using the non-parametric Spearman test. The p-values were computed using a two-sided unpaired (for MGUS vs NDMM) or paired (for NDMM vs post-treatment), non-parametric Wilcoxon method, considering $p < 0.05$ as significant. Benjamini-Hochberg correction was applied in the case of multiple comparisons to maintain the experiment-wise type I error rate at 0.05.

## Code and data availability

All methods and analyses were implemented in Python. The tested implementation of methods listed above can be found on this Code Ocean link (https://codeocean.com/capsule/0863619/tree/v1) along with documentation explaining how to run the different algorithms. A Docker file containing all the dependencies and a test .ndpi whole slide image is also included in Code Ocean repository. This repository contains an end-to-end analysis of whole slide image comprising of Tiling, superpixel-based tissue classification, cell detection, cell classification, cell counting, bone thickness quantification and cell proximity quantification. In Code Ocean, at test whole slide image is uploaded and pressing the 'Reproducible Run' button at the top right corner will automatically perform the above listed tasks and output will be saved in results folder. The code runs on both local and high-performance clusters using the Docker container. All raw data are available from the corresponding authors upon request.

# Results

## Computational and spatial analysis

Unlike solid tumors, BM trephine sections consist of isolating structural elements over different spatial scales, reflecting a mix of cellular communities and mosaic habitats. To dissect this complex tissue landscape and detect rare cells in MIHC (**Sup. Figure 1**), we specifically designed two deep learning methods, MoSaicNet to dissect the mosaic landscape of BM tissue (**Figure 1A)** and AwareNet to detect and classify cells (**Figure 1B**). First, to dissect the MM tissue into blood, bone, fat, and cellular tissue patches/habitats, a superpixel-based deep learning method was designed to capture the complex landscape (**Figure 1A**). To train and validate MoSaicNet, we collected expert segmentation annotations for 260 regions, which resulted in 69884 superpixels (**Methods**, **Sup. Table 4-5**). Subsequently, we were able to quantify the amount of cellular tissue, which served as an important quality control parameter, to determine whether a slide would be considered for further analysis. With the help of our pathologist, the tissue area threshold was set to 1.1 x

$10^6$ $\mu m^2$. Sections with cellular tissue area less than this threshold were excluded from analysis.

To optimally detect and classify cells within BM trephine samples, that contain both rare (e.g. $FOXP3^+CD4^+$) and abundant cells (**Sup. Figure 1B**). Thus, to optimally detect and classify these cell types, we developed AwareNet [25].

Subsequently, we analysed the BM spatial microenvironment in terms of cell density, cell ratio, cell spatial proximity and clustering, and bone physiology in terms of bone density/texture heterogeneity, and bone thickness (**Figure 1C, Sup. Methods**).

## High accuracy of MoSaicNet classification model

To evaluate the performance of the MoSaicNet classification model, we used 9330 superpixels extracted from separately held manually annotated samples (**Sup. Table 5**). The superpixels belonged to the blood, bone, fat and cellular tissue classes. To measure the classifier's performance, we used accuracy, area under the curve (AUC), precision, recall and F1-score (Sup. Methods). Taking all classes together, the superpixel classifier model achieved an AUC value of 0.99, 95% confidence interval (CI) [0.989, 0.991] (**Sup. Table 8**). Moreover, for each class, the bootstrap mean AUC was >0.984 for all the classes (**Figure 3A** and **Sup. Table 8**). The overall accuracy (unweighted) was 0.937, 95% CI [0.935, 0.94].

Out of the 9330 superpixels, 585 superpixels were misclassified. Out of the 585 misclassified superpixels, 208 tissue superpixels were misclassified as bone, and 122 bone superpixel patches were misclassified as tissue (**Sup. Figure 3A**). This was also evident in the lower precision value for bone class (0.88, 95% CI [0.87, 0.89]), lower recall value for bone class (0.933, 95% CI [0.93, 0.94]) and lower recall value for cellular tissue class (0.932, 95% CI [0.93, 0.94]) (**Sup. Table 8**) compared to other classes. Moreover, 88 tissue superpixels and 29 bone superpixels were misclassified as a fat class, and the precision score for the fat class was 0.933, 95% CI [0.93, 0.94] (**Sup. Table 8**). Areas under precision-recall curves (AUC-PR) were >0.95 across all classes (**Sup. Figure 3B**). A mean F1-score of 0.91 was obtained for the bone class, and for the other classes, the mean F1-score was 0.943. Taking all classes together, an F1-score of 0.94, 95% CI [0.935, 0.945] was obtained (**Sup.**

**Table 8**).

Most of the tissue superpixels misclassified as bone were superpixels with poor tissue quality, non-cellular regions, and bone-bordering areas (**Figure 3B**). Most of the 122 bone superpixels that were misclassified as tissue were a result of background staining of the bordering area (**Figure 3B**).

## Detecting rare cell types with AwareNet

To evaluate the performance of AwareNet, we measured precision, recall, and F1-score on separately held 2131 test single-cell annotations. AwareNet achieved an F1-score of 0.78, a 2% increase compared to U-net [30] and a 1% increase compared to CONCORDe-Net [13]. In particular, AwareNet excels in detecting $FOXP3^{+}CD4^{+}$ cells, which are rare in BM trephines (representing ~7% of the training data) [25].

Taking all three classes together, the single-cell classifier model of AwareNet achieved an AUC value of 0.98, 95% CI [0.977, 0.984] (**Sup. Table 9**). Moreover, for each class, the mean bootstrap AUC value was >0.98, with a minimum AUC 95% CI lower bound of 0.976 for the $CD8^{+}$ class (**Sup. Table 9** and **Figure 3C**). The overall accuracy (unweighted) was 0.965, 95% CI [0.962, 0.969]. Out of the 2131 cells, 74 cells were misclassified (**Sup. Figure 3C**). 11 cells out of 135 $FOXP3^{+}CD4^{+}$ cells were misclassified as $FOXP3^{-}CD4^{+}$ cells, and 12 $FOXP3^{-}CD4^{+}$ cells were misclassified as $FOXP3^{+}CD4^{+}$ cells (**Sup. Figure 3C**). This resulted in precision (0.857, 95% CI [0.83, 0.89]), recall (0.92, 95% CI [0.9, 0.94]), and F1-score (0.887, 95% CI [0.87, 0.91]) for the $FOXP3^{+}CD4^{+}$ class (**Sup. Table 9**). Precision-recall curves are displayed in **Sup. Figure 3D** and the AUC-PR of the rarer cell type, $FOXP3^{+}CD4^{+}$, was 0.82. For the $FOXP3^{-}CD4^{+}$ and $CD8^{+}$ class, the F1-score was 0.956, 95% CI [0.95, 0.96], and 0.98, 95% CI [0.98, 0.98], respectively (**Sup. Table 9**). Moreover, when all classes were combined, the classifier obtained an F1-score of 0.941, 95% CI [0.93, 0.95] (**Sup. Table 9**). The Matthew's correlation coefficient was 0.93 for this panel.

UMAP-based inspection of the misclassified $FOXP3^{-}CD4^{+}$ and $CD8^{+}$ cells revealed that these cells were mainly cells co-expressing both CD8 and CD4 proteins (**Figure**

**3D, Sup. Methods**). These rare cell types have been found in follicular lymphoma [31] and urological cancers [32] but, to the best of our knowledge, they have not been studied in myeloma.

AwareNet was trained on single-cell data from CD4/CD8/FOXP3 panel data and directly applied to both panels, CD4/CD8/FOXP3 and CD4/CD8/BLIMP1. After applying the model to both panels, the numbers of $CD8^+$ and $CD4^+$ cells in both panels were significantly correlated ($r=0.79$, $p=2.97 \times 10^{-7}$ and $r=0.79$, $p=3.43 \times 10^{-7}$ **Figure 3E-F**, respectively), validating the reliability of AwareNet. All cell frequencies from both panels detected by AwareNet can be found in **Sup. Table 10**.

## MoSaicNet reveals changes in bone physiology post-treatment

Using MoSaicNet, we quantified the proportion (%) of blood, bone, fat, and cellular regions in all sections (**Figure 4A**). In the NDMM group, post-treatment trephine samples contained a greater proportion of bone (%bone) when compared with diagnostic samples ($p=0.037$, **Figure 4B**). There was a trend of decrease in %bone with age ($p=0.086$). There was, however, no difference in the %bone between MGUS and NDMM or between male and female patients (**Figure 4C-E**). There was a trend of increase in %fat at post-treatment compared with diagnostic sample pair (**p=0.05, Sup. Figure 4A**) but was not different between MGUS patients and NDMM patients, nor between age or gender (**Sup. Figure 4B-D**).

To investigate the heterogeneity of bone structure in BM samples, we used a convolutional auto-encoder to learn the embedding of 177.6 thousand bone superpixels extracted from nine MGUS (27.8%), ten NDMM (34.4%) and ten post-treatment (37.8%) WSIs (**Sup. Methods**). Bone superpixels were mapped into 32 feature vectors and clustered into 17 groups (**Methods**, **Figure 4F, Sup. Figure 4E-G**). Based on this grouping, there was a positive trend in the similarity of bone superpixels from MGUS to bone superpixels from post-treatment samples compared with bone superpixels from NDMM samples, even though this was not significant ($r=0.4$, $p=0.12$ and $r=-0.13$, $p=0.63$, **Figure 4G)**.

We then asked if the bone texture differed between the patient groups. The intra-and inter-sample bone texture or density heterogeneity in NDMM was significantly higher at diagnosis compared to post-treatment (**Methods,** p=0.0098, **Figure 4H-I**). Moreover, we observe a pattern of increased bone heterogeneity in NDMM samples compared with MGUS samples; however, this was not significant (p=0.086, **Figure 4H, J**). The bone heterogeneity was similar between MGUS and post-treatment samples (**Figure 4H** and p=0.87, **Sup. Figure 4H**).

Furthermore, to analyse bone thickness, we developed an automated image analysis algorithm (**Sup. Method**, **Figure 2A**). The bone thickness of NDMM samples was similar to post-treatment samples (p=0.23, **Figure 4K**) and MGUS (p=0.37, **Figure 4L**). The bone thickness in patients aged ≤58 years (median) was significantly higher compared with that in patients aged >58 years (p=0.018, **Figure 4M**), without variation between gender (p=1.0, **Figure 4N**).

## Decreased $FOXP3^+CD4^+$ and $BLIMP1^+$ cell density post-treatment

When comparing cell density on the NDMM and post-treatment samples, we observed a decrease in both regulatory T ($FOXP3^+CD4^+$), as well as $CD8^+$ T cells following treatment (p=0.0039 and p=0.0039, respectively, **Figure 5A-B**). However, $FOXP3^-CD4^+$ T cell density did not change post-treatment (p=0.77, **Figure 5C**). The $FOXP3^+CD4^+$:$FOXP3^-CD4^+$ ratio is significantly reduced after ACST (p=0.0137, **Figure 5D**), largely due to the reduction in the density of $FOXP3^+CD4^+$ cells post-treatment. However, the $FOXP3^-CD4^+$:$CD8^+$ ratio ($CD4^+$ effector:$CD8^+$ effector cells ratio) and the $FOXP3^+CD4^+$:$CD8^+$ ratio were not different between the two-time points (**Figure 5E, Sup. Figure 5A, respectively**). We defined $FOXP3^-CD4^+$ cells as $CD4^+$ effector T cells and $CD8^+$ cells as $CD8^+$ effector T cells. Tumor burden as measured by $BLIMP1^+$ cells per unit area was significantly reduced post-treatment compared with the paired diagnostic samples (p=0.0134, **Figure 5F**). However, the $CD8^+$:$BLIMP1^+$ and $CD4^+$:$BLIMP1^+$ ratios were not significantly different between the diagnostic and post-treatment pairs (p=0.275 **Figure 5G** and p=0.43, **Sup. Figure 5B**, respectively).

## Increased spatial proximity between BLIMP1$^+$ cells and CD8$^+$ cells in NDMM compared to MGUS

The density and ratio of CD8$^+$, FOXP3$^+$CD4$^+$, and FOXP3$^-$CD4$^+$ cells were not significantly different between MGUS and NDMM (**Figure 5H, Sup. Figure 5C-G**). There was a pattern of increase in BLIMP1$^+$ cells density and BLIMP1$^+$:CD4$^+$ ratio in the NDMM sample compared to MGUS samples, though this was not significant (p=0.08, **Figure 5I**, and p=0.08 **Sup. Figure 5H,** respectively). Furthermore, the ratio of the number of BLIMP1$^+$ cells to CD8$^+$ cells did not differ between MGUS and NDMM (p=0.165, **Figure 5J**). The density of FOXP3$^+$CD4$^+$ cells was significantly correlated with the density of BLIMP1$^+$ cells in the post-treatment (r=0.79, p=0.006, **Sup. Figure 5I**) samples but not in MGUS and NDMM samples (r=0.47, p=0.205 and r=0.20, p=0.58, **Sup. Figure 5I**, respectively). **Figure 5K** and **Figure 5L** are paired pre- and post-treatment BM examples that illustrate a reduction in FOXP3$^+$CD4$^+$, CD8$^+$ and BLIMP1$^+$ cell densities post-treatment.

Next, we asked if the spatial proximity between immune cells and BLIMP1$^+$ plasma cells differed according to disease state and treatment. To demonstrate that the spatial analysis result is not dependent on the distance threshold chosen, cell proximity was calculated for a range of distances with the maximum distance set at the cell-cell communication distance of 250μm (30, 50, 100, 150, 200, 250μm) [33][34]. Cell proximity data was corrected for cell abundance (**Methods** and **Sup. Figure 6A-D**). The number of FOXP3$^+$CD4$^+$ cells in proximity to FOXP3$^-$CD4$^+$ cells decreased at post-treatment compared with the paired diagnostic samples (Benjamini-Hochberg (BH) corrected p=0.023 for r=30-250μm **Sup. Figure 7A**). However, the number of FOXP3$^+$CD4$^+$ cells in proximity to CD8$^+$ cells was not different between NDMM samples and paired post-treatment samples (BH corrected p>0.05 for r=30-250μm **Sup. Figure 7B**). The number of BLIMP1$^+$ cells in proximity to CD8$^+$ and CD4$^+$ cells significantly reduced after treatment (BH corrected p<0.05 for r= 30-250μm, **Figure 6A** and **Sup. Figure 7C,** respectively), indicating a significant change in the immune microenvironment post-treatment. However, the number of FOXP3$^+$CD4$^+$ cells in proximity to FOXP3$^-$CD4$^+$ and CD8$^+$ cells and the number of BLIMP1$^+$ cells in proximity to CD4$^+$ cells was not different between NDMM and MGUS samples (**Sup. Figure 7D-F**). Interestingly, despite similar cell density,

the number of BLIMP1$^+$ cells in proximity to CD8$^+$ cells in MGUS samples was significantly lower than in NDMM samples (BH corrected p=0.036 for r=30-250µm **Figure 6B, C**), which may indicate variability in anti-tumor immune activity in the precursor stage compared with the malignant stage.

## Significant spatial clustering of CD8$^+$ cells in NDMM samples compared with post-treatment

We next asked how cells distribute within the BM tissues; do they display a spatially dispersed or clustered pattern? To identify the spatial pattern of a specific cell type, we compared the observed nearest neighbour distance with the spatial randomness of the cell type within the tissue section (**Sup. Methods**). In most MGUS, NDMM, and post-treatment samples, we observed clustered patterns (Z-score < -1.96) of CD8$^+$, BLIMP1$^+$ and FOXP3$^-$CD4$^+$ cells compared to spatial randomness but not for FOXP3$^+$CD4$^+$ cells (**Figure 6D-H and Sup. Figure 8A-C**). The degree of clustering of CD8$^+$ cells in the NDMM was significantly higher at diagnosis than in post-treatment samples (p=0.027, **Figure 6D**) but not compared to MGUS samples (p=0.514, **Figure 6G**). There was a trend towards increased clustering of BLIMP1$^+$ cells in the NDMM samples compared with their paired post-treatment and with MGUS samples (p=0.065 and p=0.06, **Figure 6B, H**, respectively). The degree of clustering of BLIMP1$^+$ cells in female samples was significantly higher than in male patients (p=0.039, **Figure 6I**) but not different between age groups (**Sup. Figure 8D**).

## High accuracy achieved in the validation cohort

The validation cohort contained 9 NDMM and paired post-treatment BM samples (n=18) obtained from different hospitals and were stained with MIHC panel 2 using a different Leica Bond RX$^m$ autostainer. All samples had a tissue area of above 1.1 x $10^6$ µm$^2$, a threshold set for analysis inclusion. They also underwent color normalization before analysis (**Sup. Figure 9A-B**). To evaluate the performance of our model on this cohort, 4857 single-cell annotations (BLIMP1 = 2330, CD4 = 1589, CD8 = 938) and tissue segmentation (e.g. fat, bone, blood) annotations in 54 regions of interest were made on 10 samples. Despite possible variations from tissue

processing and staining, MoSaicNet was able to achieve an AUC value of 0.97, 95% CI [0.974, 0.978] taking all classes into account (**Sup. Table 11 and Sup. Figure 10**). In addition, each class had a mean AUC of >0.94, reaching an overall accuracy of 0.949, 95% CI [0.946, 0.953].

Of the 4487 superpixels, 227 superpixels were misclassified. Most of the misclassified superpixels were bone being misclassified as blood (65 superpixels), followed by blood being misclassified as tissue (51 superpixels). Taking all classes together, the overall precision value was 0.947, 95% CI [0.942, 0.95], the recall value was 0.938, 95% CI [0.933, 0.942] and the F1-score was 0.942, 95% CI [0.938, 0.945] (**Sup. Table 11**).

When evaluating the performance of AwareNet in the validation cohort, the single-cell classifier achieved an AUC value of 0.987, 95% CI [0.985, 0.988] for $BLIMP1^+$ cells, 0.988, 95% CI [0.986, 0.989] for CD4 and 0.979, 95% CI [0.973, 0.977] for CD8 (**Sup. Figure 11A-C and Sup. Table 12**). The overall accuracy was 0.905, 95% CI [0.901, 0.909]. Of the 4857 cells, 441 cells were misclassified. 192 $CD8^+$ cells were misclassified as $CD4^+$ cells and 103 $BLIMP1^+$ cells were misclassified as $CD4^+$ cells. Nevertheless, high F1-scores were noted across all three cell types: 0.944, 95% CI [0.94, 0.95] for BLIMP1, 0.897, 95% CI [0.89, 0.90] for CD4 and 0.814, 95% CI [0.80, 0.82] for CD8, with a combined F1-score of 0.885, 95% CI [0.88, 0.89] (**Sup. Table 12**). AUC-PR for all cell types were >0.91 and the Matthew's correlation coefficient was 0.85 for this cohort (**Sup. Figure 11D**).

Furthermore, quantitative and spatial analysis of the validation cohort revealed similar findings to the original dataset. As in the original dataset, NDMM samples had significantly higher $BLIMP1^+$ cell density (p=0.004, **Sup. Figure 12A-B, 13A**) than post-treatment samples in the validation cohort. Similarly, $CD4^+$ T cell densities were not significantly different between the two groups (p=0.91, **Sup. Figure 13B**). $CD8^+$ T cell densities also did not differ significantly (p=0.82, **Sup. Figure 13C**), a finding at variance with our discovery cohort, this could be due to the small sample size. Spatial analysis demonstrated significantly lower numbers of $BLIMP1^+$ cells in proximity to $CD4^+$ as well as $CD8^+$ T cells in the post-treatment group, in concordance with the original dataset (BH corrected, p=0.003, r=30-250μm, **Sup. Figure 14A-B**).

## Post-hoc analysis for training dataset sample size calculation

To estimate the sample size needed to train AwareNet and MoSaicNet, we evaluated the performance of these models using different sample sizes and displayed this as learning curves (**Sup. Methods** and **Sup. Figure 15A-B**). For AwareNet, using only 40% of the training data, we achieved an F1-score of 0.973 compared to 0.98 when using 100% of the training data (**Sup. Figure 15A**). Thus, by reducing the number of required annotations by about 60%, AwareNet could achieve comparable performance to the model trained on the whole dataset. For MoSaicNet, the model showed the highest performance when trained on 80% of the data, achieving an F1-score of 0.932 compared to a model trained on 100% of the data, with a gap of about 1% (**Sup. Figure 15B**).

# Discussion

Myeloma, like many other blood cancers, initiates and evolves in the BM. The BM ecological niche is highly organized, where hemopoietic, including immune cells, osteoblasts, osteoclasts, adipocytes, and other cells interact and co-evolve with neoplastic cells [35][36]. The BM milieu and its architectural pattern are, therefore, crucial to the decoding of neoplasm evolution for many blood cancers. Analysis of the intact BM niche has been limited in the past, both due to the difficulty in preserving epitopes and nucleic acid during the processing of BM trephines and the lack of specialized computational methods that are capable of removing sample artefacts and dissecting BM components.

Here, we demonstrate that, through the generation of carefully preserved BM trephine tissue sections and the development of spatial histology methods based on deep learning and spatial statistics, new biological insights on MM neoplastic progression and treatment response can be derived. The spatial architecture of MM BM was interrogated by establishing fully automated computational pipelines to analyse immune cells' spatial topography, bone texture heterogeneity and thickness, in addition to the changes in tumor load and BM components during neoplastic

progression and treatment. Previously, spatial interactions of stromal components in BM using 3D microscopy in a mouse model [37] and spatial interactions of BM adipose tissue and hematopoietic stem cells in rhesus macaques were studied [38]. To the best of our knowledge, this is the first study to use spatial histology based on deep learning to discover spatial cellular topologies and architectural patterns in human BM trephine samples that inform changes in disease status in MM. This is in contrast to the many machine-learning methods available for BM aspirate derived cell suspensions for cell counts and marrow evaluation [18][39]. Methods developed in our study may impact the study of many other diseases by unlocking the potential of deep learning and spatial tissue architecture, thus generating new insights from routine BM trephine samples.

BM trephine tissue is a mosaic landscape of blood, bone, cellular tissue, and fat. To dissect the complex mosaic tissue microenvironment into individual components in MIHC images, MoSaicNet was developed. Instead of a standard application of CNNs to generate patch-level [40] or pixel-level classification [30][41], MoSaicNet can efficiently define the highly irregular tissue component boundary without requiring large amounts of expert annotation training, thus combining the best of two approaches. Patch-based approaches use rigid image patches as units for classification tasks, requiring fewer annotations but cannot generate a detailed mapping of the tissue. In comparison, pixel-based algorithms such as U-Net [30] or Micro-Net [41] generate detailed contours, but such algorithms often require large amounts of training data. MoSaicNet combines a machine learning-based approach, superpixel segmentation, and deep learning classification to efficiently map out the MM BM tissue landscape using superpixels as spatial units, classifying them into cellular components, blood, bone, fat, and background.

Building on MoSaicNet, a new autoencoder-based approach was developed to study bone physiology. This was inspired by the potential role of bone and related cells, such as osteoblasts and osteoclasts, in regulating BM remodelling [14][42] and MM dormancy and proliferation [43]. Autoencoder is an effective method for dimension reduction and denoising. Here we demonstrated its value in bone texture heterogeneity analysis, using feature extraction based on autoencoder and unsupervised clustering of the bone superpixels. We observed that the amount of

bone in the biopsies taken post-treatment was greater than those taken at diagnosis, reflecting the destructive effect of MM tumor cells on bone. The bone density of NDMM samples was also more heterogeneous when compared to matched post-treatment samples, again reflecting an effect of the disease process on bone physiology that occurs in a spatially heterogeneous manner [44]. Moreover, a novel method was developed to study bone thickness using distance transform and topological analysis. In agreement with the bone trabecular surface analysis on lymphoid cancer samples [12], bone% and bone thickness showed a decreasing pattern with ageing but was not different between male and female samples. Taken together, our data indicate that bone analytical methods may be useful for the study of bone degeneration during MM progression and treatment, and bone heterogeneity may be a useful marker for disease activity.

Subsequently, AwareNet, developed specifically to identify rare immune cell types, enabled us to dissect the hematopoietic ecosystem of BM in the context of MM. Deep learning models are often sensitive to class imbalance, resulting in lower accuracy in detecting rare cell types such as $FOXP3^{+}CD4^{+}$ Tregs in our samples. To resolve this, cell segmentation-based spatial cell weighting was proposed [30][45]. AwareNet extends cell segmentation-based spatial cell weighting [30][45] by using cell identification instead of segmentation, which is less costly. Furthermore, giving a higher attention score to rare cell types improved the detection of rare cell types compared to U-Net [30] and CONCORDe-Net [13].

Using AwareNet, we observed a reduction in the density of $BLIMP1^{+}$ tumor cells, and of the immune cell subsets, CD8 and Tregs in post-treatment BM, compared with diagnostic samples from paired NDMM. While the reduction in tumor cell density is expected, the decrease in immune cell subsets may suggest an alteration in immune function, such as anti-tumor responses. Several studies have reported on the changes in frequency or proportion of T cell subsets in post-treatment BM or blood. However, all these studies have hitherto studied BM aspirate samples and assessed immune cell subsets as a percentage of the CD138-negative fraction of mononuclear cells, while our study quantified cell density as a function of tissue surface area. Thus, although we ourselves have reported an increase in $CD8^{+}$ T cells as a fraction of $CD3^{+}$ cells in post-treatment BM aspirates compared to pre-treatment samples

[46], it is not possible to directly compare these data. Regulatory T cells have attracted a great deal of attention in MM, and most studies, including our previous work in BM aspirates, concur in reporting an increased abundance of these cells in MM patients compared with healthy controls [11][47][48]. Hence, our observation in this study of a greater density of Tregs in NDMM samples compared with post-treatment samples is consistent with previous studies [49]. On the other hand, our observation that the density of $CD8^+$ cells falls following treatment may be at odds with studies using aspirate samples, for the reasons described above, as well as variation in sampling time and site, but the actual treatments received, and type of transplant are also likely to influence the results [9][5][6]. Our previous work on BM aspirates found no difference in the actual frequency of Tregs between pre- and post-treatment [46].

Importantly, new insights were derived from the topological analysis between MM plasma cells and immune T cells. In solid tumors such as oestrogen receptor-positive breast [50] and lung tumors [34], spatial scores were found to be more prognostic than cell counts. In MM, however, the spatial relationship of cells and their prognostic value have remained unexplored. Our approaches control for cell abundance and take into account the local tissue architecture and cell distribution. Interestingly, the number of $BLIMP1^+$ cells in spatial proximity with $CD8^+$ cells was significantly greater in diagnostic MM samples compared with MGUS and post-treatment samples. Given reports of tumor-reactive $CD8^+$ T cell populations in MM patients [51], the proximity of $CD8^+$ T cells to tumor cells may represent increased immune activity in MM, and the “homing” of $CD8^+$ T cells to tumor sites. This is consistent with the clustered pattern of $CD8^+$, $CD4^+$ and $BLIMP1^+$ cells in most cases. We observed a dispersed pattern of $FOXP3^+CD4^+$ Tregs. The expansion of Tregs has been found to contribute to the growth, proliferation, and survival of myeloma plasma cells [9]. Thus, the dispersed pattern of Tregs may be a phenotype of expansion, which may promote the invasion and differentiation of MM plasma cells.

Accuracy of a deep learning platform often fails when it is applied to a different set of samples with different sample preparation procedures, introducing technical variation [52]. BM samples in our validation cohort were collected from different hospitals that

may have slightly different tissue processing protocols. They were also stained using a different Leica Bond RX$^{m}$ stainer resulting in staining variation. With the use of a color normalization step, our deep learning model achieved high overall accuracy with an AUC of >0.9. There was also good concordance in the quantitative and spatial findings between the original and the validation cohort. This suggests that our model could potentially be applied to different datasets after image normalization, maintaining a high performance.

Training machine learning models on limited sample size may result in training bias such as overfitting, impacting the model's performance and generalizability [53]. In order to justify our training sample size, we performed post-hoc learning curves to evaluate performance of our models against different sample sizes. AwareNet achieved high F1-score of >0.97 when trained on 40% to 100% of the training data, whereas MoSaicNet showed best performance when trained on 80% of the data with a slight drop in performance when trained on 100% of the data. While having more data is believed to generate a better model, adding more heterogeneous data could confuse the model and lead to a reduction in performance [54]. This could explain the fluctuation of the model performance in MoSaicNet as the sample size increases. Results from these learning curves suggested that we had an adequate amount of data to train our models.

The limitations of this study include the limited number of samples. More samples are needed to capture the full cellular and non-cellular region heterogeneity, and the results should be interpreted with this consideration. Our quantitative and spatial results are likely underpowered, but these are exploratory analyses and as such, there was no pre-specified power or sample size. Finally, the MIHC staining contained three parameters. Our next step will be to apply the computational methods developed in this study to more parameters, allowing us to distinguish more immune cell subsets.

To conclude, we demonstrated how spatial and machine learning methods can be used to dissect the mosaic tissue microenvironment of BM trephine samples (MoSaicNet) and accurately identify immune T and MM plasma cells (AwareNet). Despite the limited sample size, bone trabeculae morphologic and cell spatial

proximity analyses enabled the deep mine of both cellular and non-cellular parts of the BM niche. Future works include: 1) adapting MoSaicNet and AwareNet to routinely available hematoxylin and eosin stain of BM trephine samples to further explore bone remodelling; 2) integrating morphologic and spatial features with molecular features to identify genetic aberrations associated with morphologic or spatial phenotypes in the BM niche; 3) identifying morphologic and spatial features of progressor and non-progression patients with MM precursor conditions [55] to help refine risk models; 4) exploring the association of bone morphologic features and cellular spatial topography features with patients' clinical outcomes such as treatment response and survival. Insights generated from this study warrant further validation and investigation in larger cohorts, which is in progress.

# Authors' contributions

**YBH** and **CL** contributed equally to this work and share first authorship. **YY and KY** conceived and designed the study; **YBH** developed the deep learning pipelines, image analysis and spatial analysis pipelines, and performed the statistical analysis. **EL** provided the trephine biopsies and curated the database. **CL** and **DP** undertook the cutting of trephine tissue sections, MIHC panels optimization and staining. **CL** and **AM** collected the clinical data. **CL, MRJ, DP** and **TAT** performed pathological annotations and BM tissue regions, which were used for model training and evaluation. **MRJ** reviewed the bone trabeculae analysis. **YBH, CL, KY, LL, and YY** wrote the paper. **SC** contributed to the spatial analysis and reviewing the manuscript. **KF** reviewed the manuscript. All authors reviewed and approved the final version.

# Acknowledgement

**YBH** is funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No766030. **YY** acknowledges funding from Cancer Research UK Career Establishment Award (C45982/A21808), Breast Cancer Now (2015NovPR638), Children’s Cancer and Leukaemia Group (CCLGA201906), NIH U54 CA217376 and R01 CA185138, CDMRP Breast Cancer Research Program Award BC132057, CRUK Brain Tumour

Awards (TARGET-GBM), European Commission ITN (H2020-MSCA-ITN-2019), Wellcome Trust (105104/Z/14/Z), and The Royal Marsden/ICR National Institute of Health Research Biomedical Research Centre. This study has also received funding from Blood Cancer UK. Dr Xiaoxi Pan from The Institute of Cancer Research provided expert advice on the color normalization step before analysing the validation cohort.

# Tables:

**Table 1 | Patient characteristics:** MGUS

| Patient characteristics (n=9) | Patient no. (%) |
|---|---|
| **Age at diagnosis** | |
| Median (range) | 61 (54, 89) |
| **Gender** | |
| Male | 5 (56) |
| **Immunoglobulin (Ig) isotype** | |
| IgG | 5 (56) |
| IgA | 3 (33) |
| Light chains only | 1 (11) |
| **Light chain isotype** | |
| Kappa | 5 (56) |
| Lambda | 3 (33) |
| Polytypic | 1 (11) |
| **IMWG Cytogenetics risk** | |
| Standard risk | 5 (56) |
| High risk | 1 (11) |
| Unknown | 3 (33) |
| **Risk categories for progression to MM** | |
| Low | 5 (56) |
| Intermediate | 2 (22) |
| High | 2 (22) |

**Table 2 | Patient characteristics:** Paired diagnostic and post-treatment samples

| **Patient characteristics (n=10)** | Patient no. (%) |
|---|---|
| **Age at diagnosis** | |
| Median (range) | 56 (53, 63) |
| **Gender** | |
| Male | 6 (60) |
| **Immunoglobulin (Ig) isotype** | |
| IgG | 5 (50) |
| IgA | 2 (20) |
| Light chains only | 3 (30) |
| **Light chain isotype** | |
| Kappa | 7 (70) |
| Lambda | 3 (30) |
| **IMWG Cytogenetics risk** | |
| Standard risk | 5 (50) |
| High risk | 5 (50) |
| **IMWG ISS staging** | |
| I | 4 (40) |
| II | 5 (50) |
| III | 1 (10) |
| **PC % in diagnostic BM biopsy** | |
| Median (range) | 70% (13, 80) |
| **Line of therapy at treatment** | |
| 1 | 10 (100) |
| **Induction therapy** | |
| KCD* | 10 (100) |
| **PC % at D100 BM biopsy post-treatment** | |
| Median (range) | 0.5% (0, 10) |

*K = Carfilzomib, C = cyclophosphamide, D = dexamethasone.

# Figures legend

**Figure 1 | Overview of computational deep learning and image processing pipelines for BM MIHC images: A**) MoSaicNet pipeline. The polygons (black) indicate superpixels. MoSaicNet dissects a tissue section into bone, blood, fat, and cellular tissue regions (**Methods**). **B**) AwareNet for attention-based cell detection and classification (**Methods**). The attention image pixel values were generated from the abundance of cell types. An attention image was applied to the objective function during model parameter optimization to regularize the algorithm by assigning high attention to rare cell types. The cell detection algorithm generates a cell probability map. A post-processing algorithm was developed to find the cell nucleus centre, (x, y) location, from the probability map (**Methods**). A patch centred on each cell was extracted and fed to deep learning (DL) based classifier to infer its class. **C**) Spatial and morphological analysis of BM trephine samples. Bone texture and structural heterogeneity were investigated using an auto-encoder-based machine learning method (**Sup. Methods**). We used spatial proximity analysis to study the spatial relations of cells. r = radius. Cell density refers to the number of cells per unit of tissue area.

**Figure 2 | Computational methods for bone thickness analysis and cells infiltration patterns: A**) Image analysis to estimate bone thickness (**Sup. Methods**). Using the same BM sample image as Figure 1A, the bone segmentation (ii) is an output of MoSaicNet (**Methods**), and each bone is displayed in a different color. The color bar shows the pixel intensity of the image in (iii and iv). The pixel intensity on the skeleton indicates half of the bone thickness (**Sup. Methods**). **B**) Cells infiltration pattern analysis using nearest neighbour distance (NND) and the null hypothesis of complete spatial randomness (CSR) (**Sup. Methods**). $Z<-1.96$, $Z>1.96$, and $-1.96 \leq Z \leq 1.96$ indicate a clustered, dispersed, and random distribution of observed cells, respectively. std=standard deviation; μ=mean NND of CSR.

**Figure 3 | Performance evaluation of MoSaicNet and AwareNet deep learning models: A**) The receiver operating characteristic (ROC) curves and AUC values of the MoSaicNet superpixel classifier. The values in brackets indicate the 95% confidence interval. **B**) 2-dimensional mapping of superpixels using MoSaicNet

learned 200-dimensional features after dimensionality reduction by Uniform Manifold Approximation and Projection (UMAP). **C**) The ROC curves and AUC values of single-cell classifier model on separately held test data. The values in brackets indicate the 95% confidence interval. **D**) UMAP features visualization of deep learned features by AwareNet single-cell classifier CNN. **E-F**) Validation of AwareNet model using correlation of density of $CD8^+$ (**E**) and $CD4^+$ cells (**F**) in panel 1 and panel 2.

**Figure 4 | Studying bone physiology using MoSaicNet: A**) Proportion of different compartments of BM trephine digital images. One stacked bar represents a sample. **B-E)** Boxplots showing the difference in %bone between samples from NDMM and post-treatment (**B**), MGUS and NDMM (**C**), different age groups (**D**) (median age=58.0 years), and gender groups (**E**). **F**) Scatter plot showing the number of bone superpixels in 17 clusters from MGUS, NDMM and post-treatment samples. The size of the dots represents the percentage of superpixels. The color represents the number of slides in each cluster. **G**) correlation of percentage of superpixels in each cluster between different patient groups. A point represents a cluster. **H**) Scatter plot of slide-level heterogeneity of bone features measured by features variance (**Sup. Methods**). A point represents a patient/slide. **I,J**) box plots showing differences in bone density heterogeneity between NDMM and post-treatment (**I**), and between MGUS and NDMM (**J**). **K-L**) Boxplots showing the difference in bone thickness between samples from NDMM and post-treatment (**K**), MGUS and NDMM (**L**), and different age groups (median age=58.0 years) (**M**) and gender (**N**).

**Figure 5 | Density of immune T cells and plasma cells in MGUS, NDMM and post-treatment samples:** (**A-G**) Boxplots showing the difference in density of $FOXP3^+CD4^+$ (**A**), the density of $CD8^+$ (**B**), the density of $FOXP3^-CD4^+$ (**C**), $FOXP3^+CD4^+$:$FOXP3^-CD4^+$ ratio(**D**), $FOXP3^+CD4^+$:$CD8^+$ ratio(**E**), density of $BLIMP1^+$ (**F**), and $CD8^+$:$BLIMP1^+$ ratio(**G**) between paired NDMM samples and post-treatment samples (n=10 pairs). **H-J**) Boxplot showing the difference in density of $FOXP3^+CD4^+$(**H**), the density of $BLIMP1^+$ (**I**) cells, and $CD8^+$:$BLIMP1^+$ cells (**J**) between MGUS and NDMM samples (n=19). **K-L**) Sample images showing the reduction of the density of $FOXP3^+CD4^+$ and $CD8^+$ cells (**K**) and $BLIMP1^+$ (**L**) cells at post-treatment compared to paired NDMM samples. The cell density is presented per 1 $mm^2$ tissue area.

**Figure 6 | Spatial neighbourhood of immune and tumor cells: A-B**) and between MGUS and NDMM (**B**). The p* indicate p values after multiple testing correction using the Benjamini-Hochberg method. The points represent the mean and the bars are 95% confidence intervals indicating uncertainty. **C**) Sample images showing an increased number of BLIMP1$^{+}$ cells in the neighbourhood with CD8$^{+}$ on NDMM samples (NDMM example shown here is the same image as Figure 5L) compared with MGUS samples. **D-I)** Clustered or dispersed pattern of immune and tumor cells in BM trephine sample. Boxplots showing the difference in nearest neighbour distance (NND) and Z score between NDMM and post-treatment for CD8$^{+}$ cells (**D**), BLIMP1$^{+}$ cells (**E**), FOXP3$^{-}$CD4$^{+}$ cells (**F**). Boxplots showing the difference in NND and Z score between NDMM and MGUS for CD8$^{+}$ cells (**G**) and BLIMP1$^{+}$ cells (**H**), and between male and female for BLIMP1$^{+}$ cells (**I**). The unit of NND is µm. The Z score shows the significance of the difference between the NND distribution for a given cell type from a complete spatial random distribution and the observed NND (**Sup. Methods**).

# References


[1] van de Donk NWCJ, Pawlyn C, Yong KL. Multiple myeloma. Lancet 2021;397:410–27. https://doi.org/10.1016/S0140-6736(21)00135-5/ATTACHMENT/FAA32D97-1AF7-4D13-B4A2-F55E9888AE92/MMC1.PDF.

[2] Fairfield H, Falank C, Avery L, Reagan MR. Multiple myeloma in the marrow: pathogenesis and treatments. Ann N Y Acad Sci 2016;1364:32. https://doi.org/10.1111/NYAS.13038.

[3] Kumar SK, Rajkumar V, Kyle RA, Van Duin M, Sonneveld P, Mateos MV, et al. Multiple myeloma. Nat Rev Dis Prim 2017 31 2017;3:1–20. https://doi.org/10.1038/nrdp.2017.46.

[4] Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. Lancet Oncol 2014;15:e538–48. https://doi.org/10.1016/S1470-2045(14)70442-5.

[5] Galustian C, Meyer B, Labarthe MC, Dredge K, Klaschka D, Henry J, et al.

The anti-cancer agents lenalidomide and pomalidomide inhibit the proliferation and function of T regulatory cells. Cancer Immunol Immunother 2009;58:1033–45. https://doi.org/10.1007/S00262-008-0620-4.

[6] Kasyapa CS, Sher T, Chanan-Khan AA. Multiple myeloma and immunomodulation: regulating the regulatory cells. Leuk \& Lymphoma 2012;53:1253–4. https://doi.org/10.3109/10428194.2012.670233.

[7] Dosani T, Carlsten M, Maric I, Landgren O. The cellular immune system in myelomagenesis: NK cells and T cells in the development of myeloma [corrected] and their uses in immunotherapies. Blood Cancer J 2015;5:e306. https://doi.org/10.1038/BCJ.2015.32.

[8] Casey M, Nakamura K. The Cancer-Immunity Cycle in Multiple Myeloma. ImmunoTargets Ther 2021;10:247. https://doi.org/10.2147/ITT.S305432.

[9] Hadjiaggelidou C, Katodritou E. Regulatory T-Cells and Multiple Myeloma: Implications in Tumor Immune Biology and Treatment. J Clin Med 2021;10:4588. https://doi.org/10.3390/JCM10194588.

[10] An G, Acharya C, Feng X, Wen K, Zhong M, Zhang L, et al. Osteoclasts promote immune suppressive microenvironment in multiple myeloma: therapeutic implication. Blood 2016;128:1590–603. https://doi.org/10.1182/BLOOD-2016-03-707547.

[11] Alrasheed N, Lee L, Ghorani E, Henry JY, Conde L, Chin M, et al. Marrow-Infiltrating Regulatory T Cells Correlate with the Presence of Dysfunctional CD4 + PD-1 + Cells and Inferior Survival in Patients with Newly Diagnosed Multiple Myeloma. Clin Cancer Res 2020;26:3443–54. https://doi.org/10.1158/1078-0432.CCR-19-1714.

[12] van Eekelen L, Pinckaers H, van den Brand M, Hebeda KM, Litjens G. Using deep learning for quantification of cellularity and cell lineages in bone marrow biopsies and comparison to normal age-related variation. Pathology 2022;54:318–27. https://doi.org/10.1016/J.PATHOL.2021.07.011.

[13] Hagos YB, Narayanan PL, Akarca AU, Marafioti T, Yuan Y. ConCORDe-net: Cell count regularized convolutional neural network for cell detection in multiplex immunohistochemistry images. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2019;11764 LNCS:667–75. https://doi.org/10.1007/978-3-030-32239-7_74/COVER.

[14] Brück OE, Lallukka-Brück SE, Hohtari HR, Ianevski A, Ebeling FT, Kovanen

PE, et al. Machine Learning of Bone Marrow Histopathology Identifies Genetic and Clinical Determinants in Patients with MDS. Blood Cancer Discov 2021;2:238–49. https://doi.org/10.1158/2643-3230.BCD-20-0162.

[15] Duffy D, Perrin H, Abadie V, Benhabiles N, Boissonnas A, Liard C, et al. Neutrophils Transport Antigen from the Dermis to the Bone Marrow, Initiating a Source of Memory CD8+ T Cells. Immunity 2012;37:917–29. https://doi.org/10.1016/J.IMMUNI.2012.07.015.

[16] Maciocia N, Wechalekar A, Yong K. Monoclonal gammopathy of undetermined significance (MGUS) and smoldering myeloma (SMM): a practical guide to management. Hematol Oncol 2017;35:432–9. https://doi.org/10.1002/HON.2345.

[17] Durie DBGM. International Staging System for Multiple Myeloma | The IMF. Int Myeloma Found 2021. https://www.myeloma.org/international-staging-system-iss-reivised-iss-r-iss (accessed August 25, 2022).

[18] Chandradevan R, Aljudi AA, Drumheller BR, Kunananthaseelan N, Amgad M, Gutman DA, et al. Machine-based detection and classification for bone marrow aspirate differential counts: initial development focusing on nonneoplastic cells. Lab Investig 2019 1001 2019;100:98–109. https://doi.org/10.1038/s41374-019-0325-7.

[19] Yagi H, Nomura T, Nakamura K, Yamazaki S, Kitawaki T, Hori S, et al. Crucial role of FOXP3 in the development and function of human CD25+CD4+ regulatory T cells. Int Immunol 2004;16:1643–4656. https://doi.org/10.1093/INTIMM/DXH165.

[20] Shapiro-Shelef M, Lin KI, Savitsky D, Liao J, Calame K. Blimp-1 is required for maintenance of long-lived plasma cells in the bone marrow. J Exp Med 2005;202:1471. https://doi.org/10.1084/JEM.20051611.

[21] Shaffer AL, Lin KI, Kuo TC, Yu X, Hurt EM, Rosenwald A, et al. Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. Immunity 2002;17:51–62. https://doi.org/10.1016/S1074-7613(02)00335-7.

[22] Nutt SL, Taubenheim N, Hasbold J, Corcoran LM, Hodgkin PD. The genetic network controlling plasma cell differentiation. Semin Immunol 2011;23:341–9. https://doi.org/10.1016/J.SMIM.2011.08.010.

[23] Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels

compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell 2012;34:2274–81. https://doi.org/10.1109/TPAMI.2012.120.

[24] Zormpas-Petridis K, Noguera R, Ivankovic DK, Roxanis I, Jamin Y, Yuan Y. SuperHistopath: A Deep Learning Pipeline for Mapping Tumor Heterogeneity on Low-Resolution Whole-Slide Digital Histopathology Images. Front Oncol 2021;10. https://doi.org/10.3389/FONC.2020.586292.

[25] Hagos YB, Lecat CSY, Patel D, Lee L, Tran TA, Justo MR, et al. Cell abundance aware deep learning for cell detection on highly imbalanced pathological data. Proc - Int Symp Biomed Imaging 2021;2021-April:1438–42. https://doi.org/10.48550/arxiv.2102.11677.

[26] Li Y, Dong M, Hua J. A Gaussian Mixture Model to Detect Clusters Embedded in Feature Subspace. Commun Inf Syst 2007;7:337–52. https://doi.org/cis/1211574970.

[27] Pedregosa FABIANPEDREGOSA F, Michel V, Grisel OLIVIERGRISEL O, Blondel M, Prettenhofer P, Weiss R, et al. Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, Pedregosa FABIANPEDREGOSA, F., Michel, V., Grisel OLIVIERGRISEL, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., . J Mach Learn Res 2011;12:2825–30. https://doi.org/10.5555/1953048.

[28] Grevera GJ. Distance Transform Algorithms And Their Implementation And Evaluation. Deform Model 2007:33–60. https://doi.org/10.1007/978-0-387-68413-0_2.

[29] Tsogkas S, Dickinson S. AMAT: Medial Axis Transform for Natural Images. Proc IEEE Int Conf Comput Vis 2017;2017-Octob:2727–36. https://doi.org/10.48550/arxiv.1703.08628.

[30] Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, et al. U-Net: deep learning for cell counting, detection, and morphometry. Nat Methods 2018 161 2018;16:67–70. https://doi.org/10.1038/s41592-018-0261-2.

[31] Hagos YB, Akarca AU, Ramsay A, Rossi RL, Pomplun S, Ngai V, et al. High inter-follicular spatial co-localization of CD8+FOXP3+ with CD4+CD8+ cells predicts favorable outcome in follicular lymphoma. Hematol Oncol 2022. https://doi.org/10.1002/HON.3003.

[32] Bohner P, Chevalier MF, Cesson V, Rodrigues-Dias SC, Dartiguenave F,

Burruni R, et al. Double positive CD4+CD8+ T cells are enriched in urological cancers and favor T helper-2 polarization. Front Immunol 2019;10. https://doi.org/10.3389/FIMMU.2019.00622/FULL.

[33] Francis K, Palsson BO. Effective intercellular communication distances are determined by the relative time constants for cyto/chemokine secretion and diffusion. Proc Natl Acad Sci U S A 1997;94:12258–62. https://doi.org/10.1073/PNAS.94.23.12258/ASSET/9FCEA2F0-1E16-4E00-9F15-60D84AE2B7DB/ASSETS/GRAPHIC/PQ2272845004.JPEG.

[34] AbdulJabbar K, Raza SEA, Rosenthal R, Jamal-Hanjani M, Veeriah S, Akarca A, et al. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. Nat Med 2020 267 2020;26:1054–62. https://doi.org/10.1038/s41591-020-0900-x.

[35] Ghobrial IM, Detappe A, Anderson KC, Steensma DP. The bone-marrow niche in MDS and MGUS: implications for AML and MM. Nat Rev Clin Oncol 2018 154 2018;15:219–33. https://doi.org/10.1038/nrclinonc.2017.197.

[36] Shiozawa Y, Havens AM, Pienta KJ, Taichman RS. The bone marrow niche: habitat to hematopoietic and mesenchymal stem cells, and unwitting host to molecular parasites. Leuk 2008 225 2008;22:941–50. https://doi.org/10.1038/leu.2008.48.

[37] Gomariz A, Helbling PM, Isringhausen S, Suessbier U, Becker A, Boss A, et al. Quantitative spatial analysis of haematopoiesis-regulating stromal cells in the bone marrow microenvironment by 3D microscopy. Nat Commun 2018 91 2018;9:1–15. https://doi.org/10.1038/s41467-018-04770-z.

[38] Robino JJ, Pamir N, Rosario S, Crawford LB, Burwitz BJ, Roberts CT, et al. Spatial and biochemical interactions between bone marrow adipose tissue and hematopoietic stem and progenitor cells in rhesus macaques. Bone 2020;133. https://doi.org/10.1016/J.BONE.2020.115248.

[39] Allegra A, Tonacci A, Sciaccotta R, Genovese S, Musolino C, Pioggia G, et al. Machine Learning and Deep Learning Applications in Multiple Myeloma Diagnosis, Prognosis, and Treatment Selection. Cancers (Basel) 2022;14. https://doi.org/10.3390/CANCERS14030606.

[40] Xu Y, Jia Z, Wang LB, Ai Y, Zhang F, Lai M, et al. Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. BMC Bioinformatics 2017;18:1–17.

https://doi.org/10.1186/S12859-017-1685-X/FIGURES/8.

[41] Raza SEA, Cheung L, Shaban M, Graham S, Epstein D, Pelengaris S, et al. Micro-Net: A unified model for segmentation of various objects in microscopy images. Med Image Anal 2019;52:160–73. https://doi.org/10.1016/J.MEDIA.2018.12.003.

[42] Manier S, Sacco A, Leleu X, Ghobrial IM, Roccaro AM. Bone marrow microenvironment in multiple myeloma progression. J Biomed \& Biotechnol 2012;2012. https://doi.org/10.1155/2012/157496.

[43] Lawson MA, McDonald MM, Kovacic N, Khoo WH, Terry RL, Down J, et al. Osteoclasts control reactivation of dormant myeloma cells by remodelling the endosteal niche. Nat Commun 2015 61 2015;6:1–15. https://doi.org/10.1038/ncomms9983.

[44] Schürch CM, Rasche L, Frauenfeld L, Weinhold N, Fend F. A review on tumor heterogeneity and evolution in multiple myeloma: pathological, radiological, molecular genetics, and clinical integration. Virchows Arch 2020;476:337–51. https://doi.org/10.1007/S00428-019-02725-3.

[45] Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2017;10553 LNCS:240–8. https://doi.org/10.1007/978-3-319-67558-9_28.

[46] Lee L, Alrasheed N, Khandelwal G, Fitzsimons E, Richards H, Wilson W, et al. Increased Immune-Regulatory Receptor Expression on Effector T Cells as Early Indicators of Relapse Following Autologous Stem Cell Transplantation for Multiple Myeloma. Front Immunol 2021;12. https://doi.org/10.3389/FIMMU.2021.618610.

[47] Kawano Y, Zavidij O, Park J, Moschetta M, Kokubun K, Mouhieddine TH, et al. Blocking IFNAR1 inhibits multiple myeloma-driven Treg expansion and immunosuppression. J Clin Invest 2018;128:2487–99. https://doi.org/10.1172/JCI88169.

[48] Zavidij O, Haradhvala NJ, Mouhieddine TH, Sklavenitis-Pistofidis R, Cai S, Reidy M, et al. Single-cell RNA sequencing reveals compromised immune microenvironment in precursor stages of multiple myeloma. Nat Cancer 2020 15 2020;1:493–506. https://doi.org/10.1038/s43018-020-0053-3.

[49] Muthu Raja KR, Rihova L, Zahradova L, Klincova M, Penka M, Hajek R. Increased T regulatory cells are associated with adverse clinical features and predict progression in multiple myeloma. PLoS One 2012;7. https://doi.org/10.1371/JOURNAL.PONE.0047077.

[50] Heindl A, Sestak I, Naidoo K, Cuzick J, Dowsett M, Yuan Y. Relevance of Spatial Heterogeneity of Immune Infiltration for Predicting Risk of Recurrence After Endocrine Therapy of ER+ Breast Cancer. J Natl Cancer Inst 2018;110:166–75. https://doi.org/10.1093/JNCI/DJX137.

[51] Vuckovic S, Bryant CE, Lau KHA, Yang S, Favaloro J, McGuire HM, et al. Inverse relationship between oligoclonal expanded CD69\- TTE and CD69+ TTE cells in bone marrow of multiple myeloma patients. Blood Adv 2020;4:4593–604. https://doi.org/10.1182/BLOODADVANCES.2020002237.

[52] Li D, Bledsoe JR, Zeng Y, Liu W, Hu Y, Bi K, Liang A, Li S. A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals. Nat Commun 2020;11:6004. doi: 10.1038/s41467-020-19817-3.

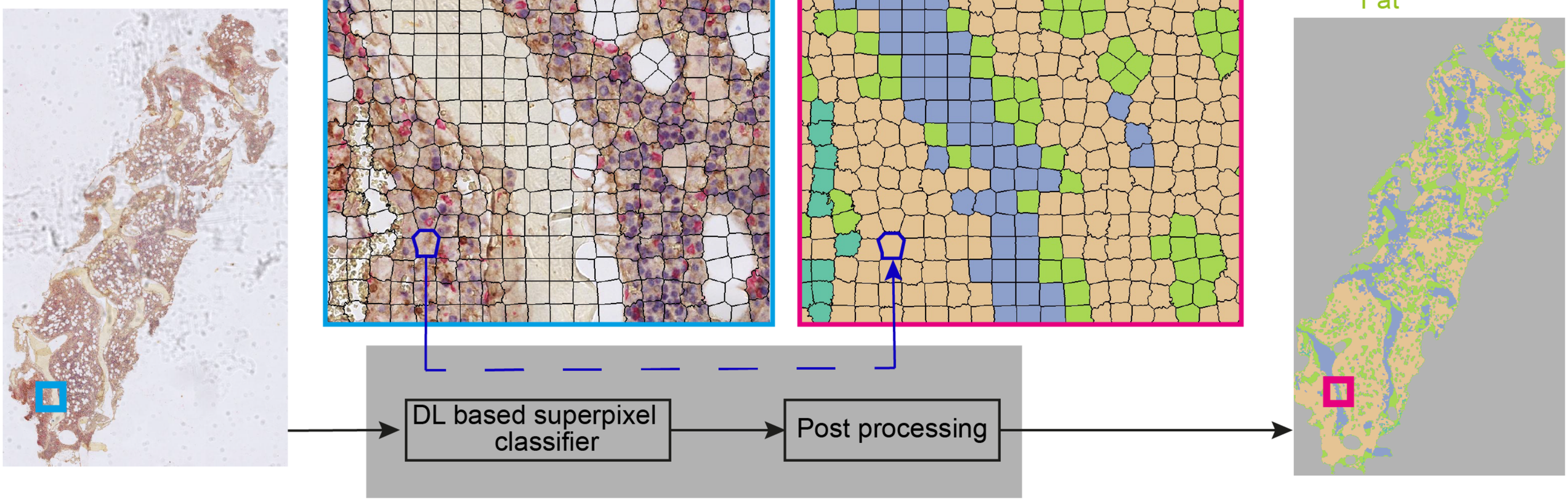
[53] Rajput D, Wang WJ, Chen CC. Evaluation of a decided sample size in machine learning applications. BMC Bioinformatics 2023; 24(1):48. doi: 10.1186/s12859-023-05156-9.

[54] Budach L, Feuerpfeil M, Ihde N, Nathansen A, Noack N, Patzlaff H, et al. The effects of data quality on machine learning performance. arXiv 2022. preprint arXiv:2207.14529.

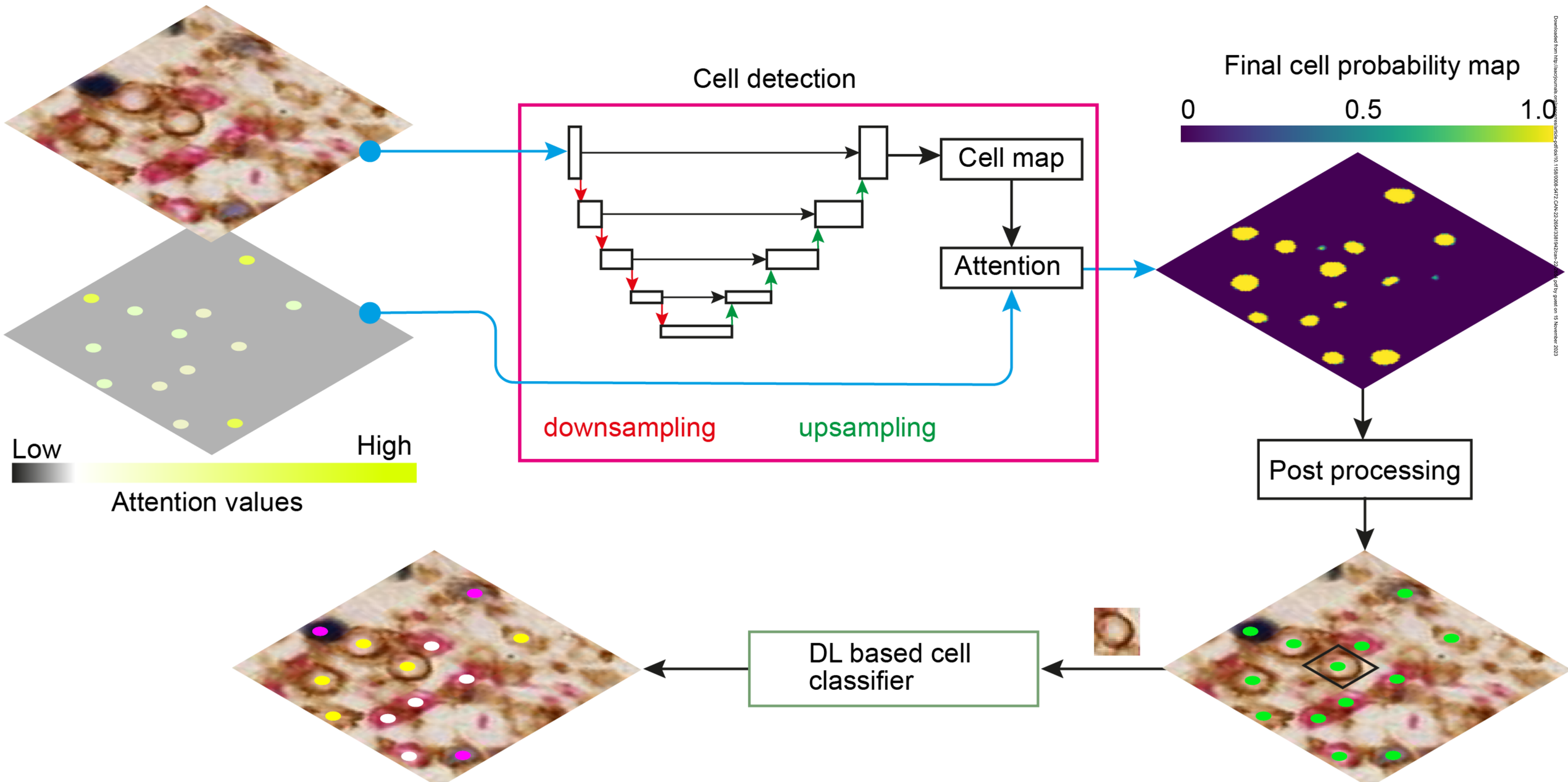
[55] Oben B, Froyen G, Maclachlan KH, Leongamornlert D, Abascal F, Zheng-Lin B, et al. Whole-genome sequencing reveals progressive versus stable myeloma precursor conditions as two distinct entities. Nat Commun 2021 121 2021;12:1–11. https://doi.org/10.1038/s41467-021-22140-0.

# Figure 1

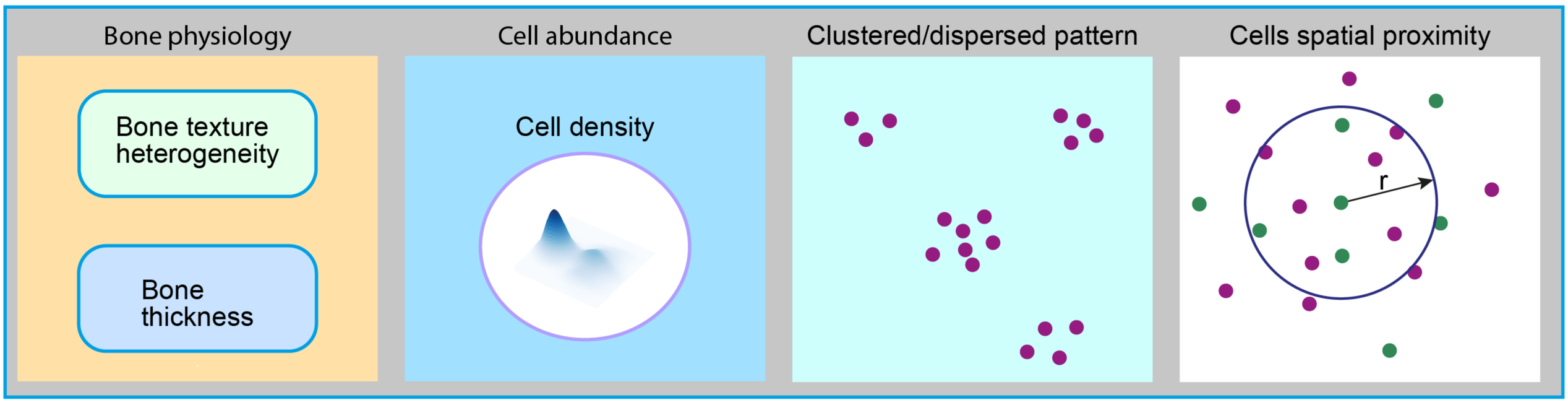
**A** MoSaicNet to dissect BM landscape



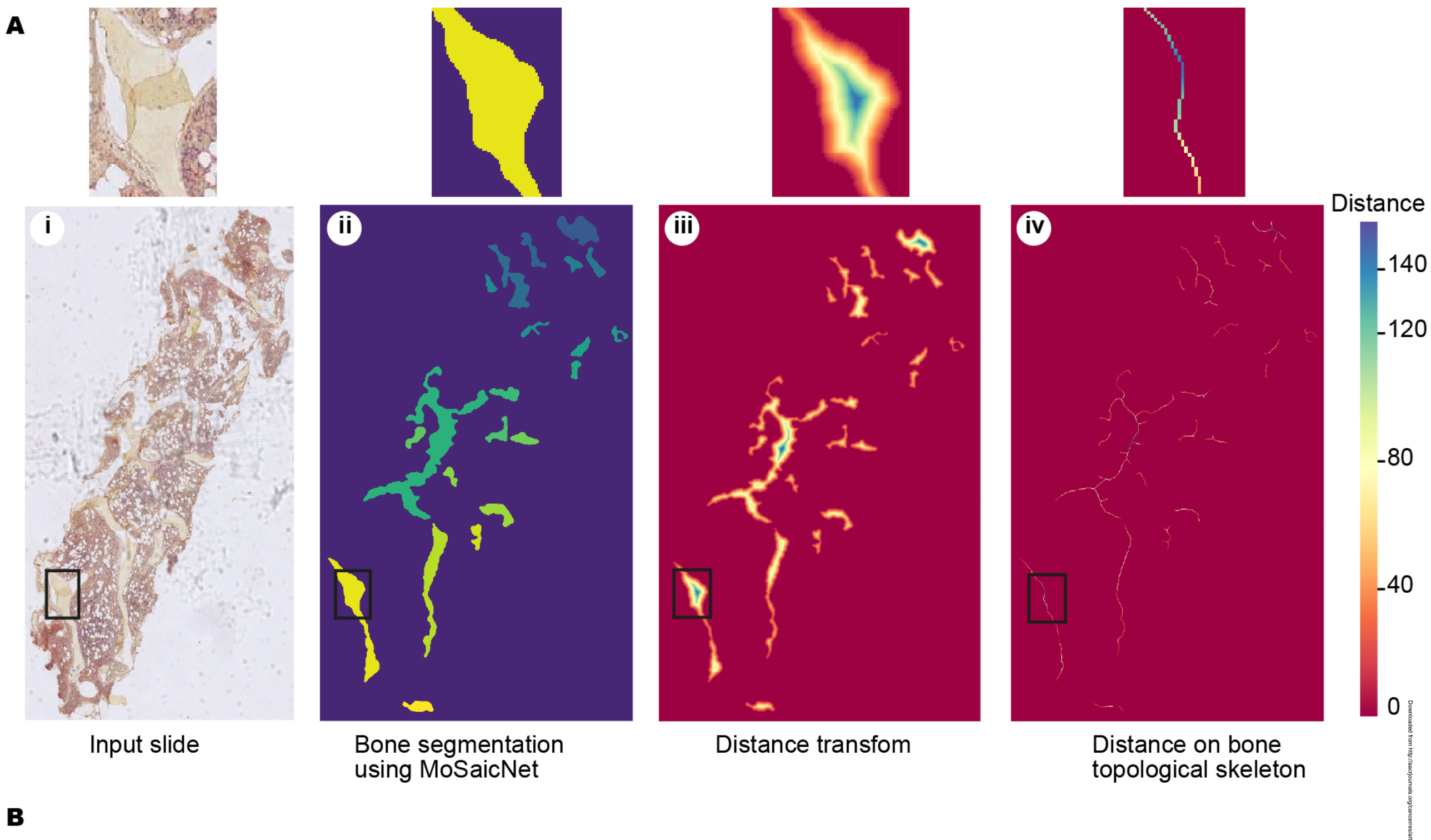

**B** AwareNet for attention based cell detection and classification



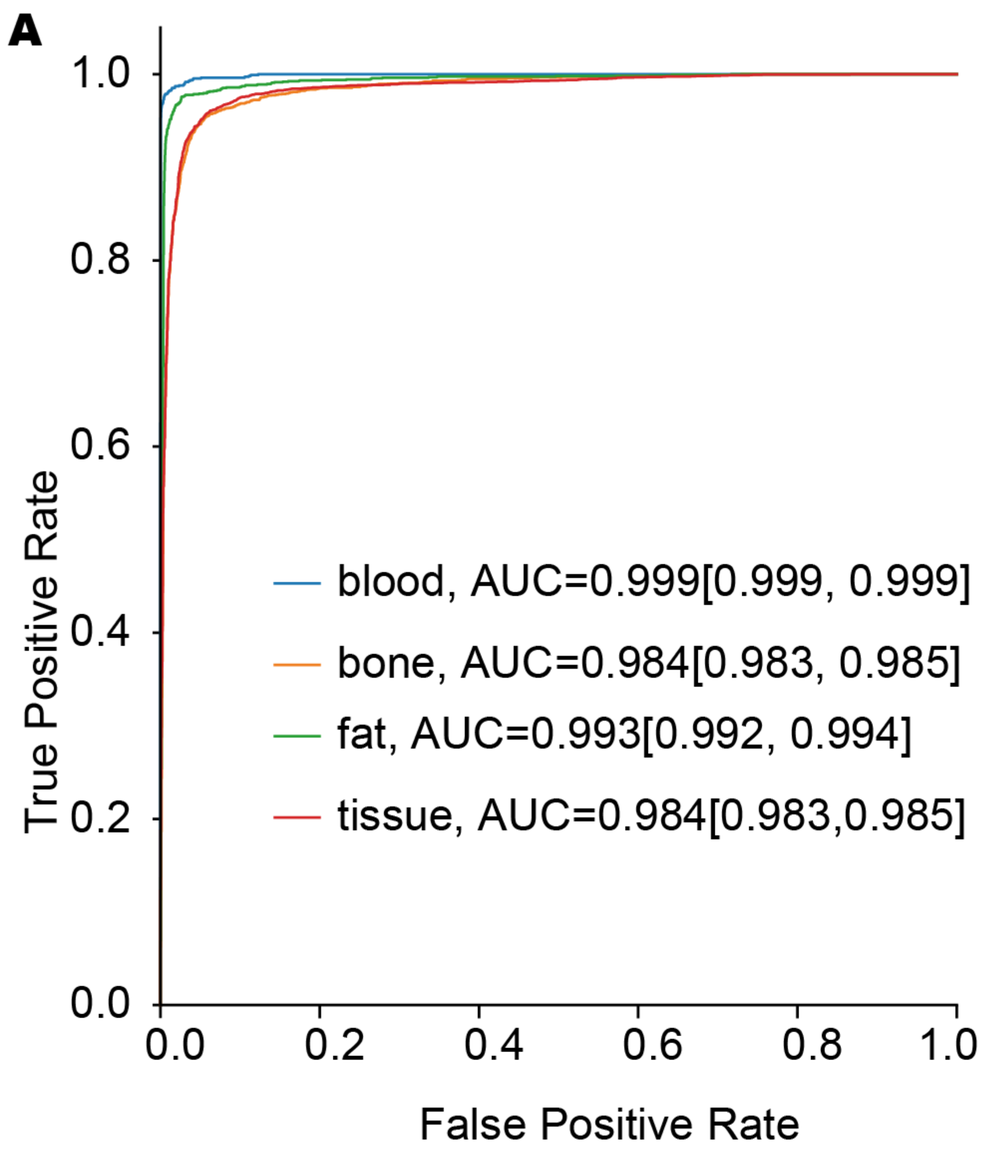

**C** Spatial and morphological analysis of BM microenvironment

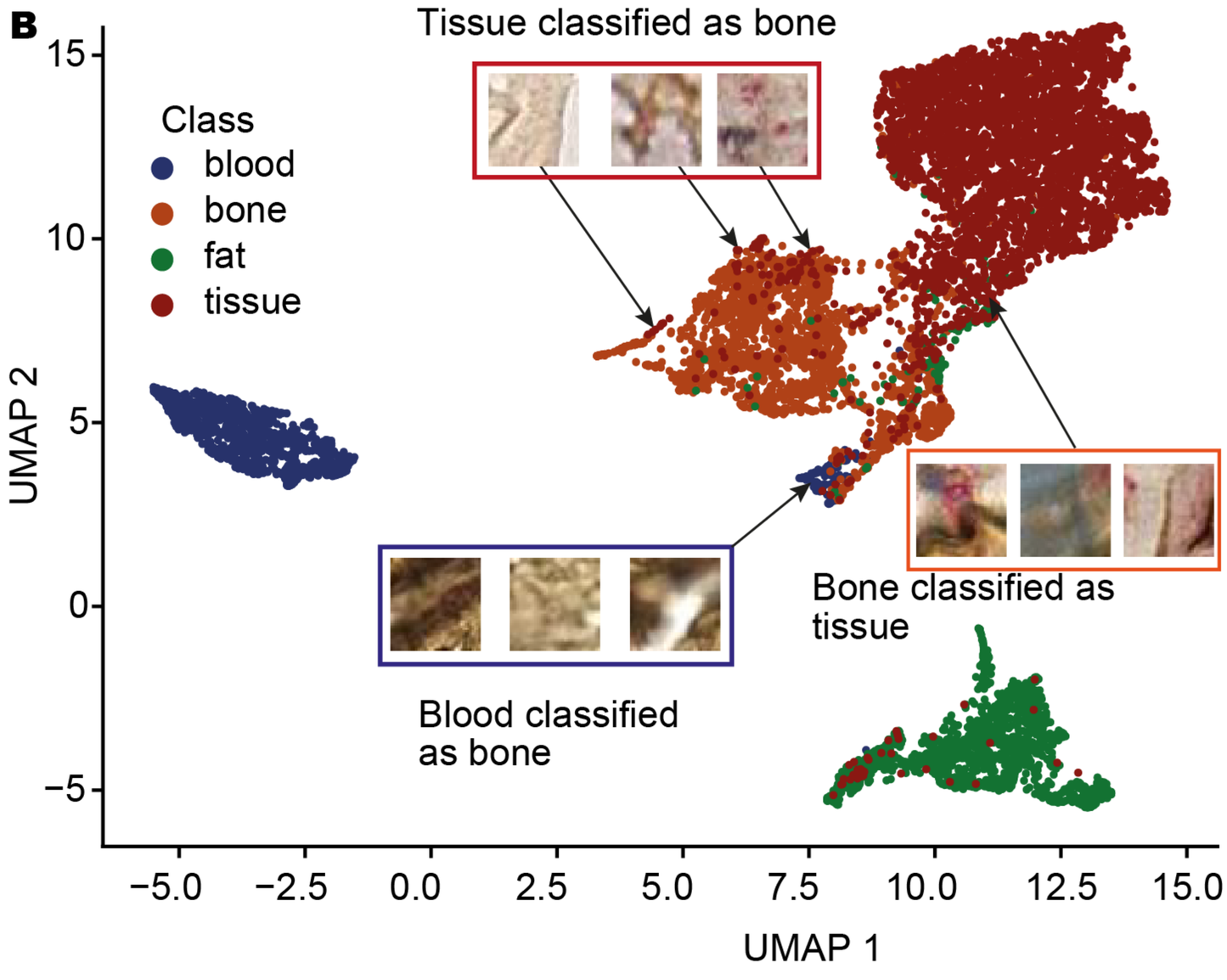


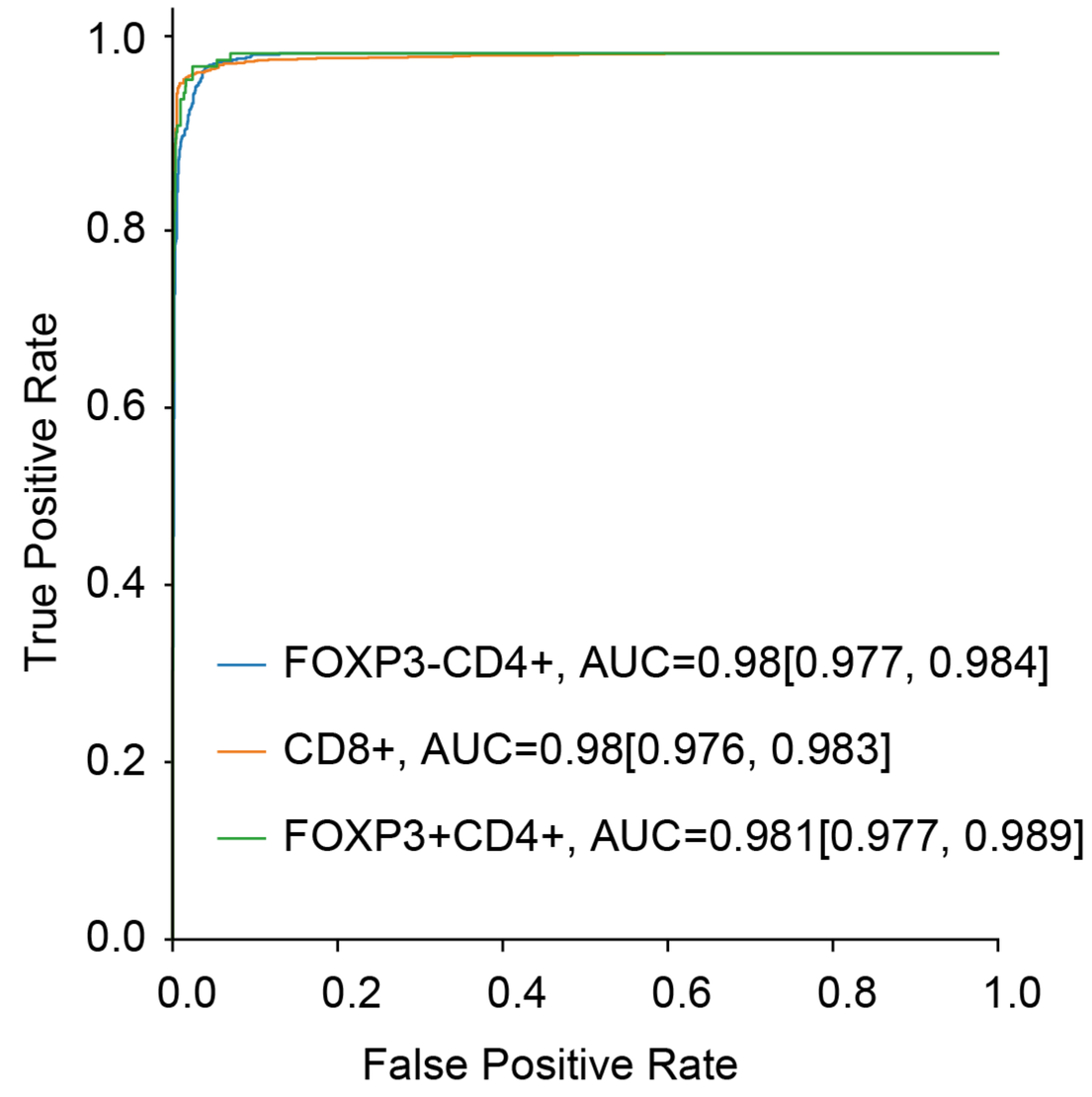
# Figure 2

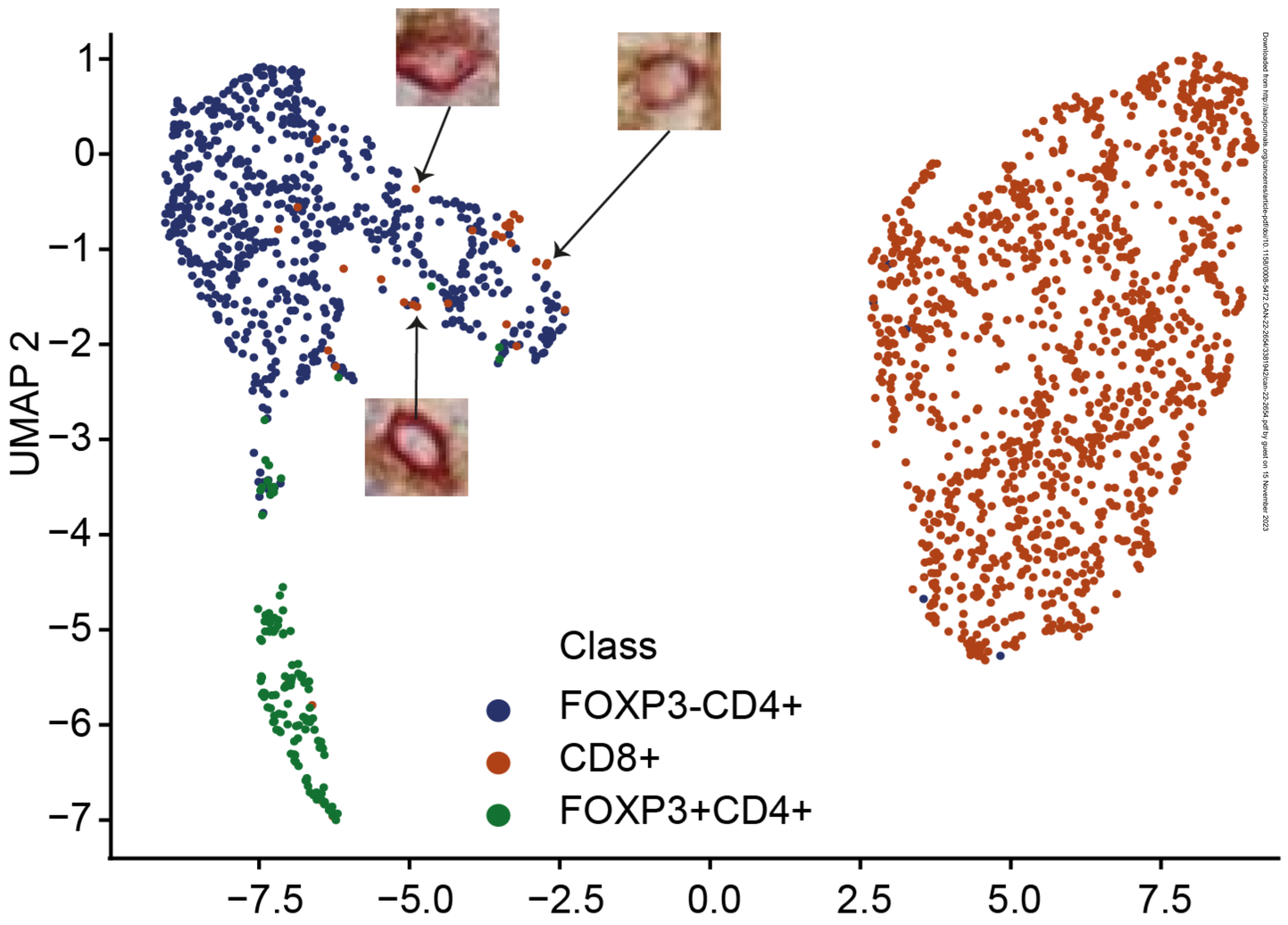


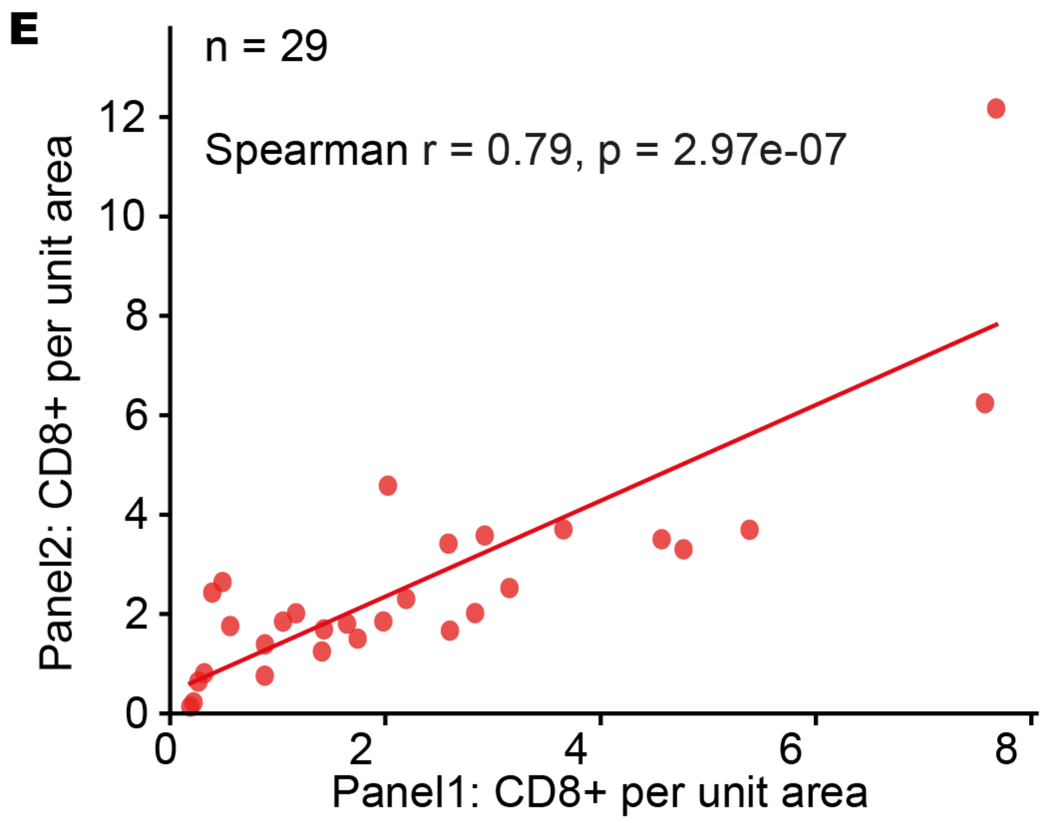
# Figure 3

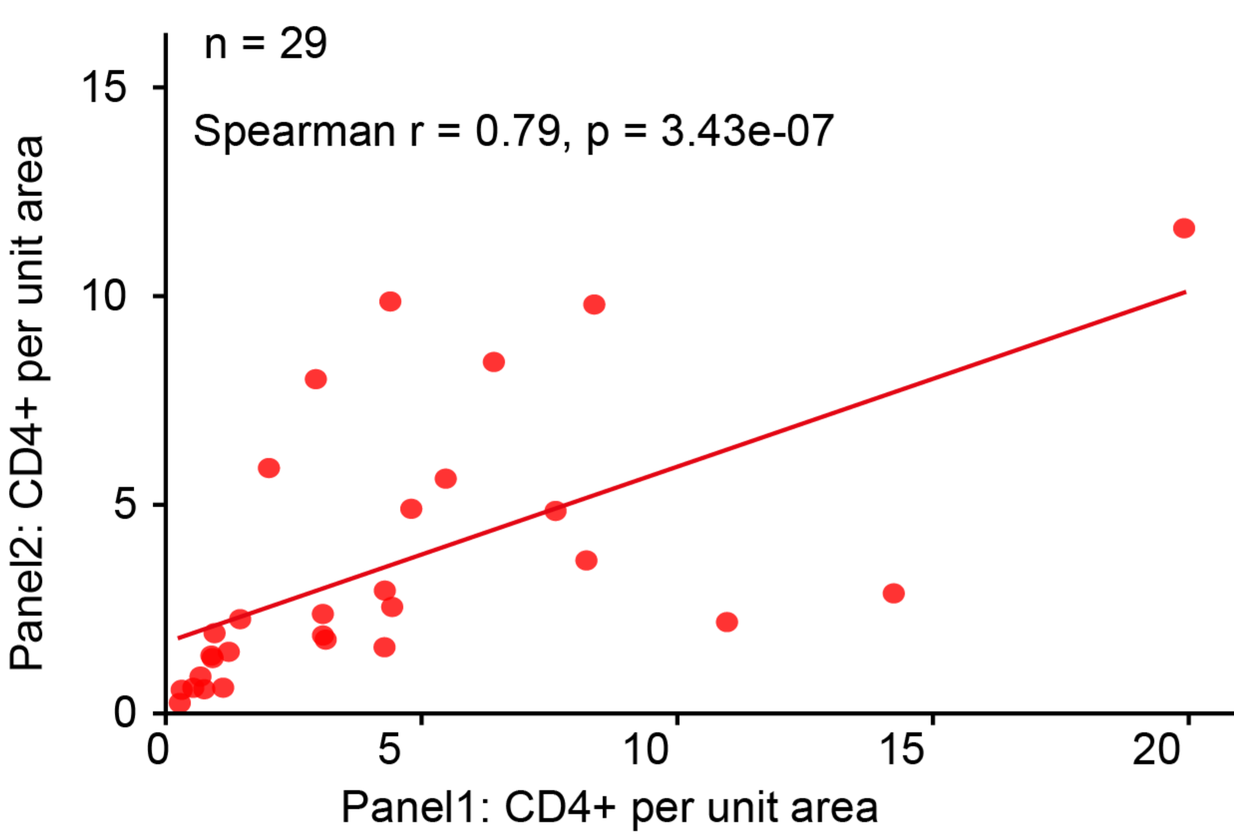

A
True Positive Rate
False Positive Rate
blood, AUC=0.999[0.999, 0.999]
bone, AUC=0.984[0.983, 0.985]
fat, AUC=0.993[0.992, 0.994]
tissue, AUC=0.984[0.983,0.985]


B
Tissue classified as bone
Class
blood
bone
fat
tissue
Bone classified as tissue
Blood classified as bone
UMAP 2
UMAP 1

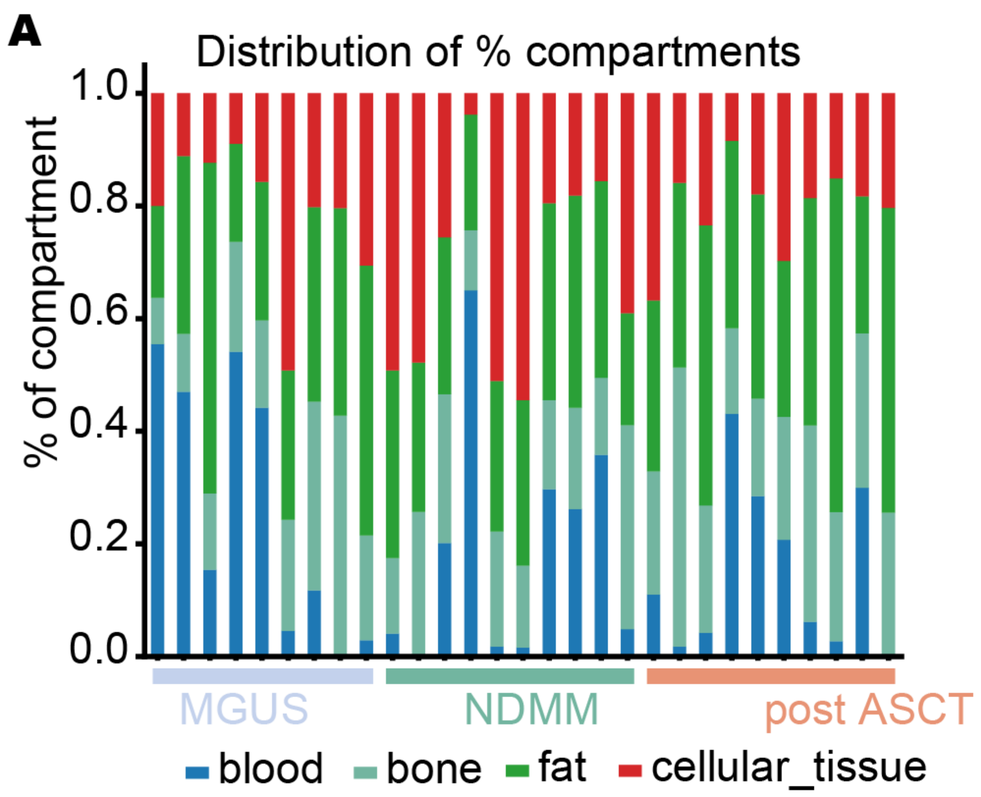

C
True Positive Rate
False Positive Rate
FOXP3-CD4+, AUC=0.98[0.977, 0.984]
CD8+, AUC=0.98[0.976, 0.983]
FOXP3+CD4+, AUC=0.981[0.977, 0.989]

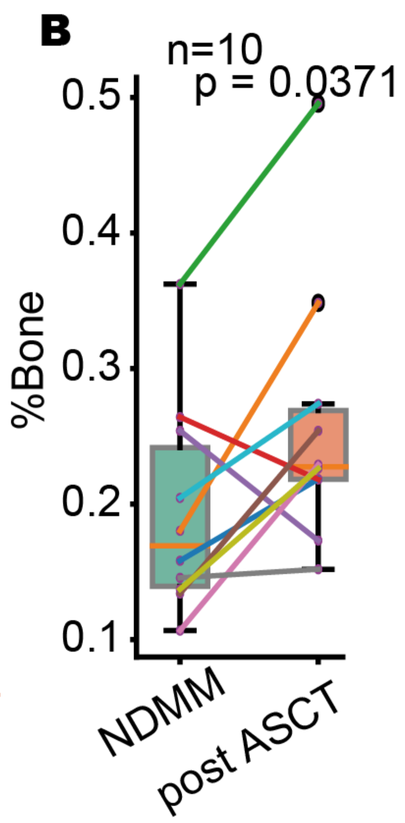

D
UMAP 2
UMAP 1
Class
FOXP3-CD4+
CD8+
FOXP3+CD4+

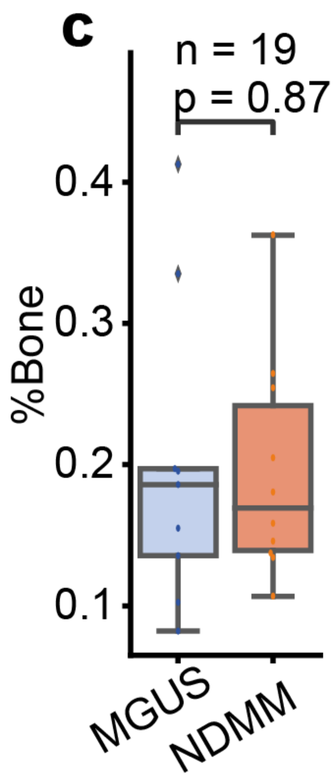

E
n = 29
Spearman r = 0.79, p = 2.97e-07
Panel2: CD8+ per unit area
Panel1: CD8+ per unit area

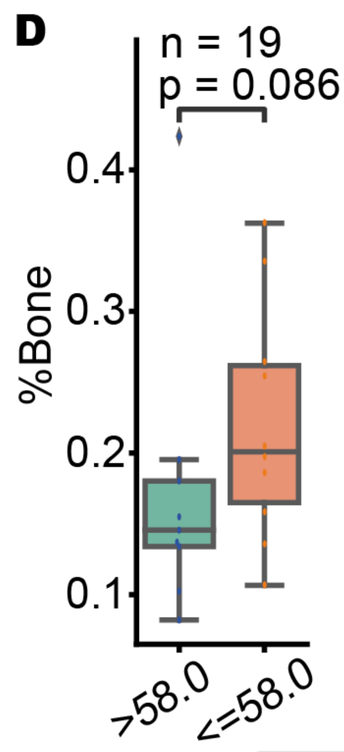

F
n = 29
Spearman r = 0.79, p = 3.43e-07
Panel2: CD4+ per unit area
Panel1: CD4+ per unit area

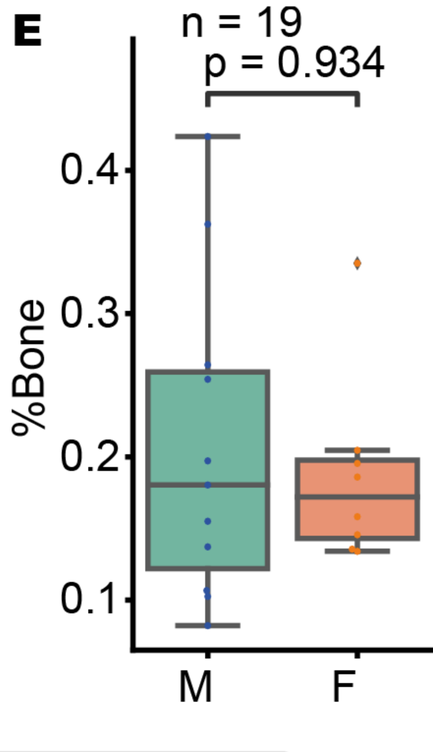
# Figure 4

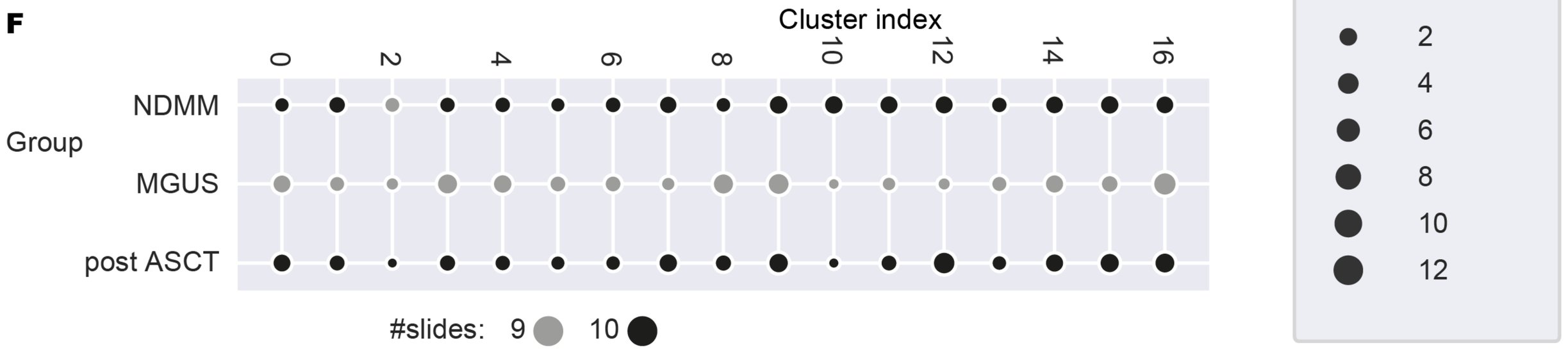


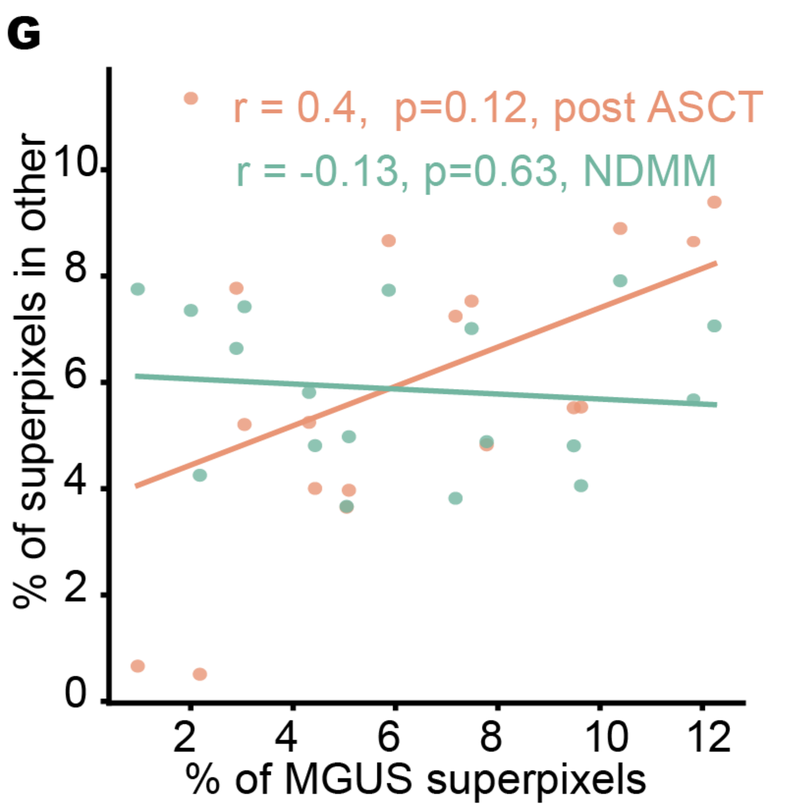



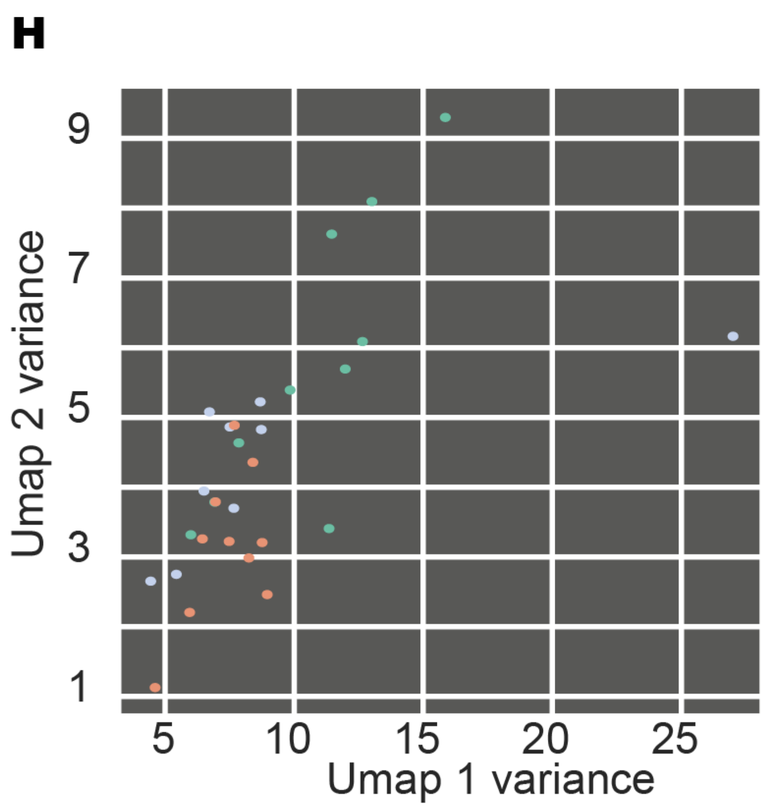



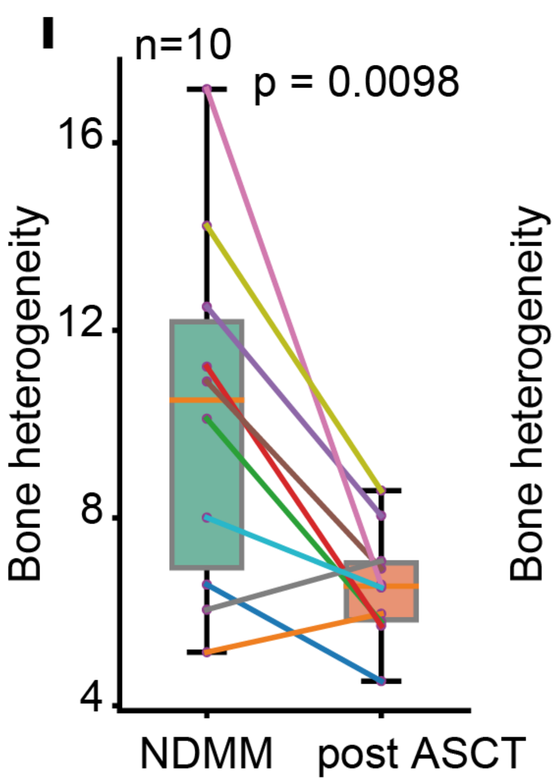



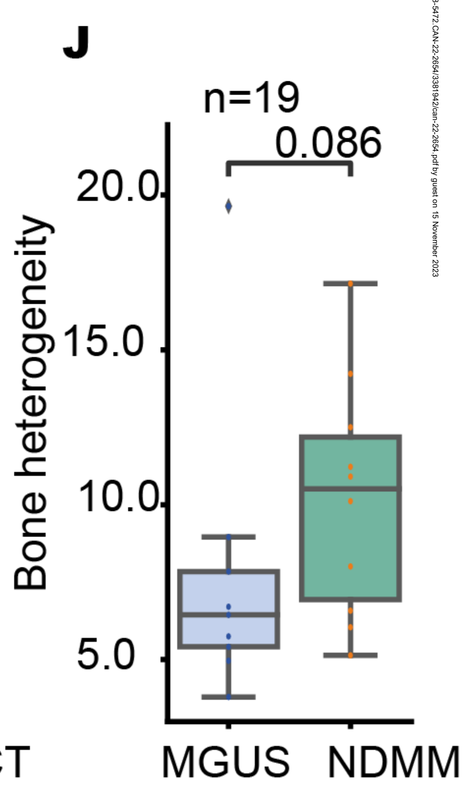



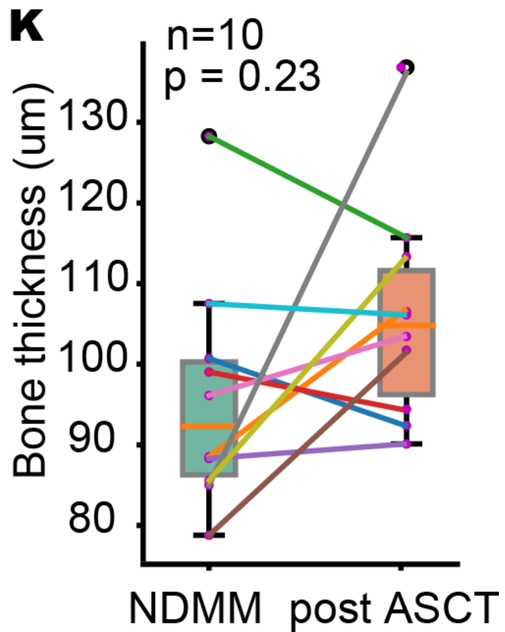



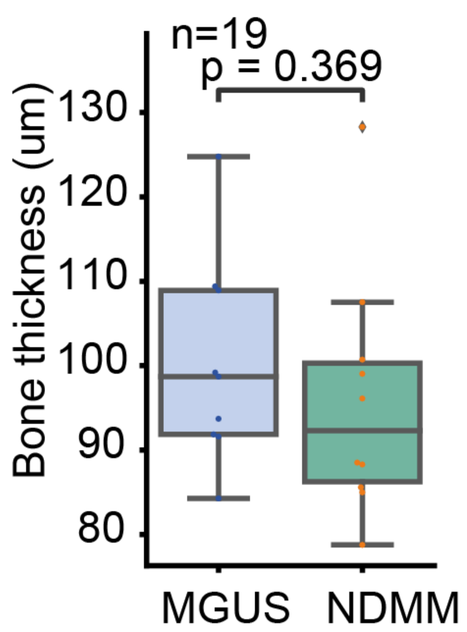



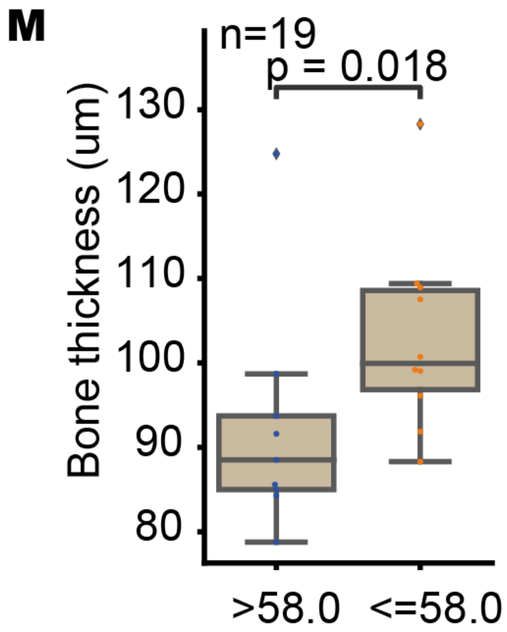



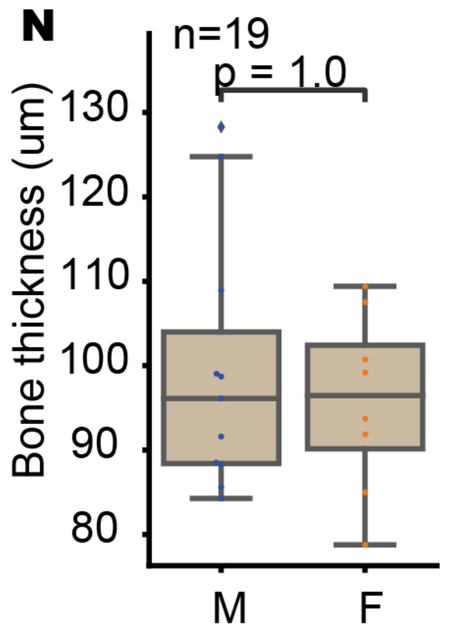



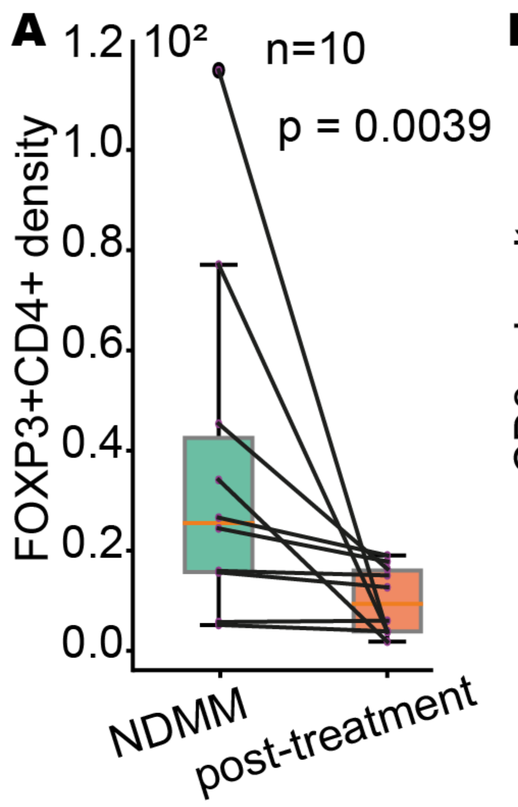



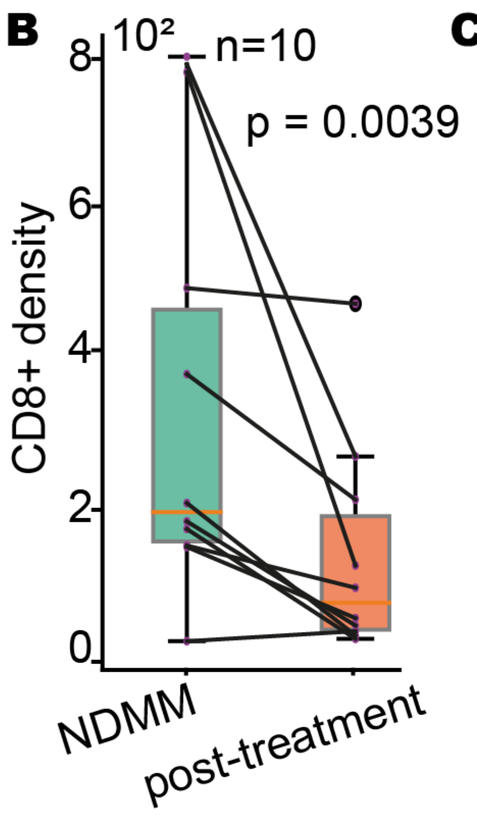



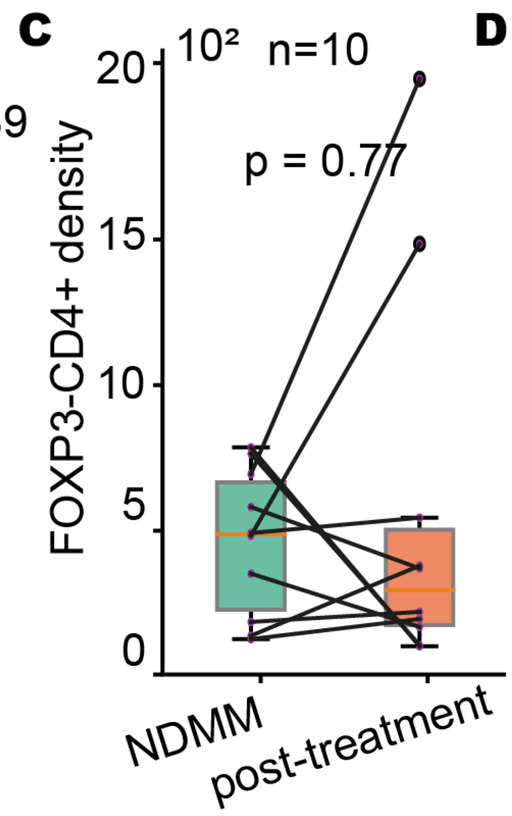



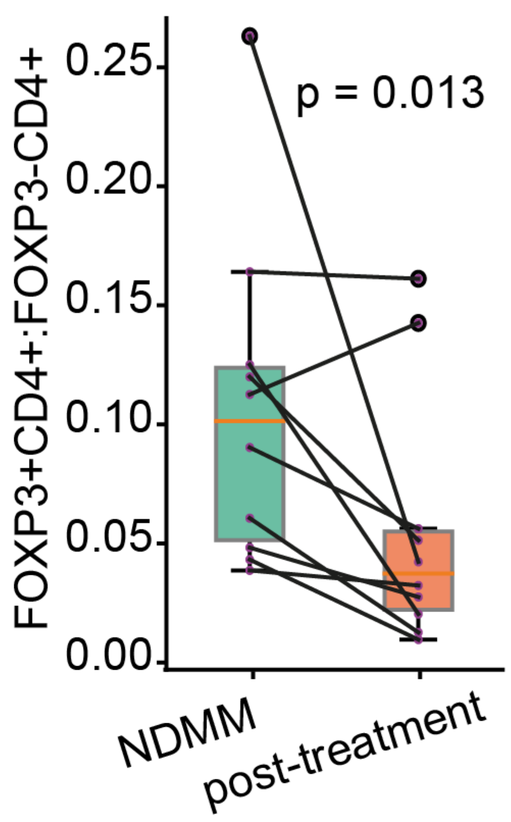



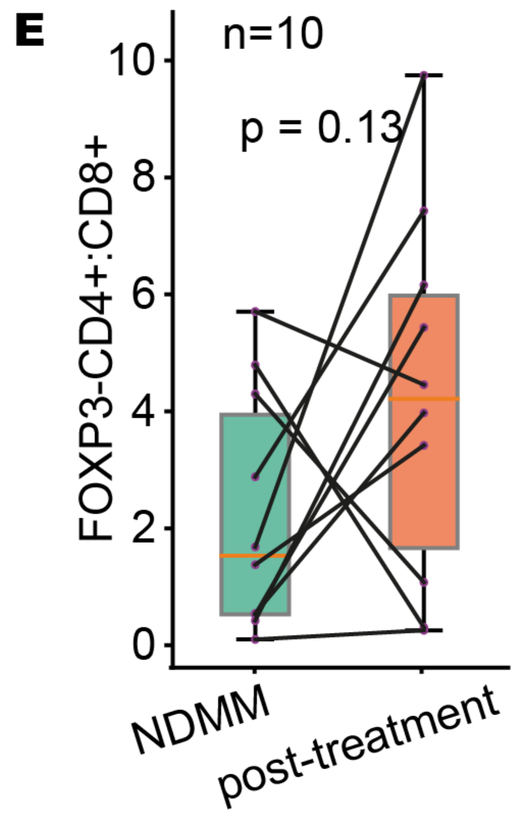
# Figure 5

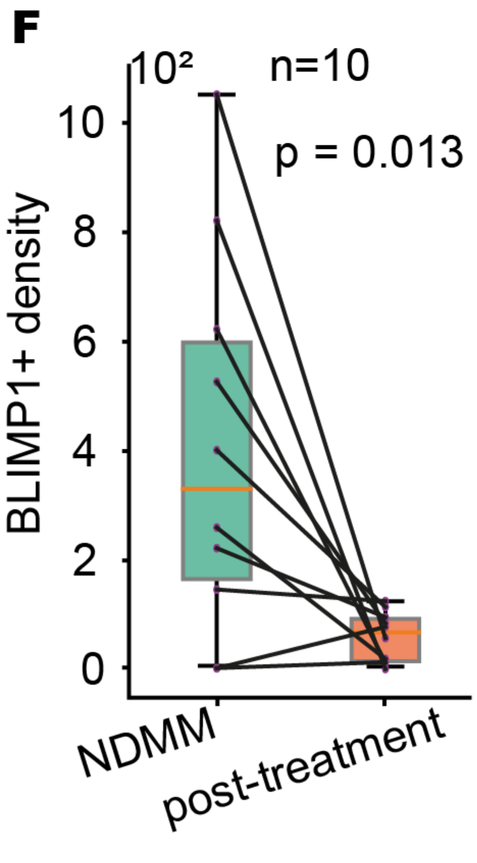

A
10²
n=10
p = 0.0039
FOXP3+CD4+ density
NDMM
post-treatment

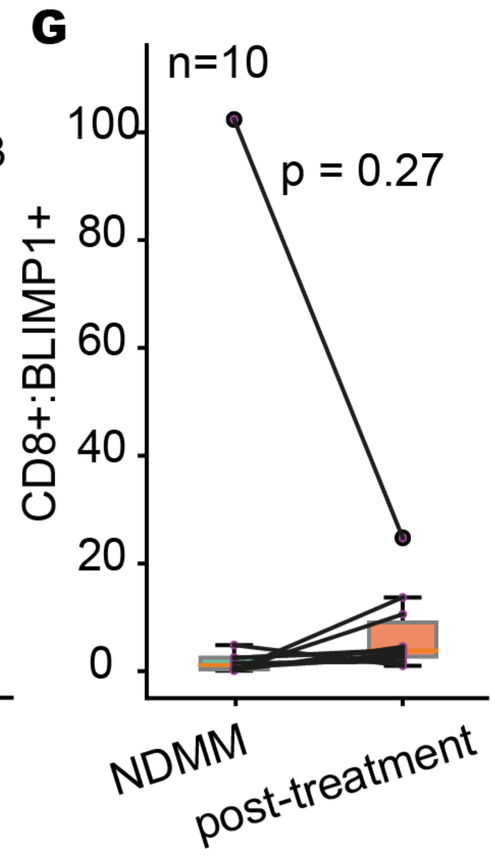

B
10²
n=10
p = 0.0039
CD8+ density
NDMM
post-treatment

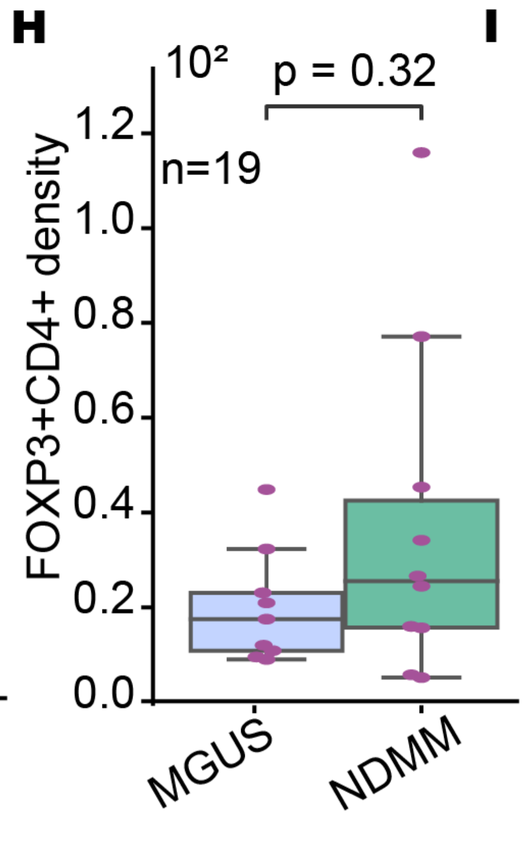

C
10²
n=10
p = 0.77
FOXP3-CD4+ density
NDMM
post-treatment

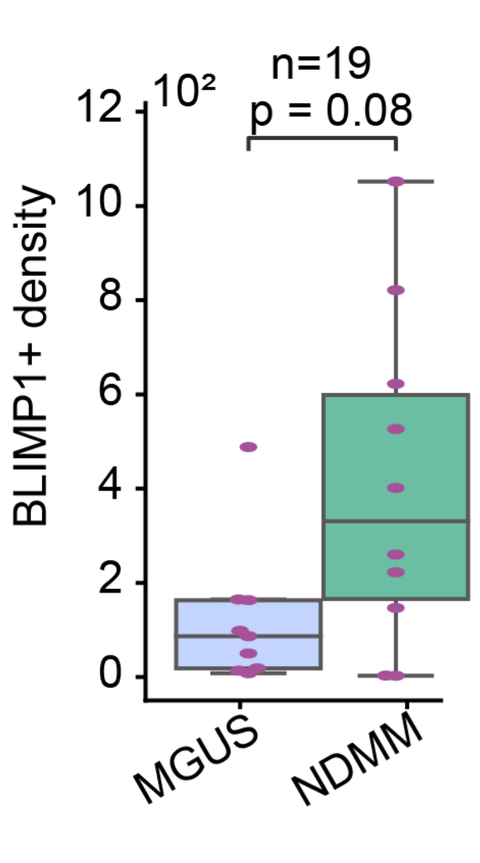

D
p = 0.013
FOXP3+CD4+:FOXP3-CD4+
NDMM
post-treatment

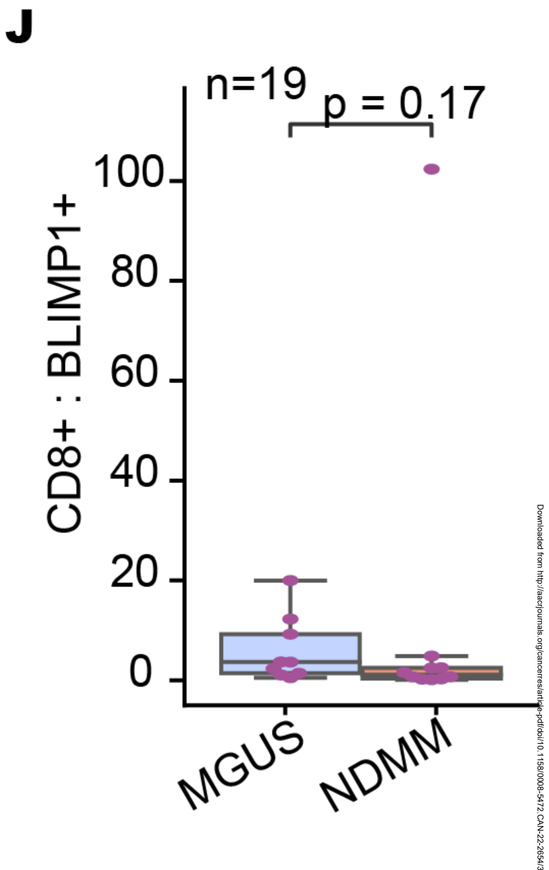

E
n=10
p = 0.13
FOXP3-CD4+:CD8+
NDMM
post-treatment

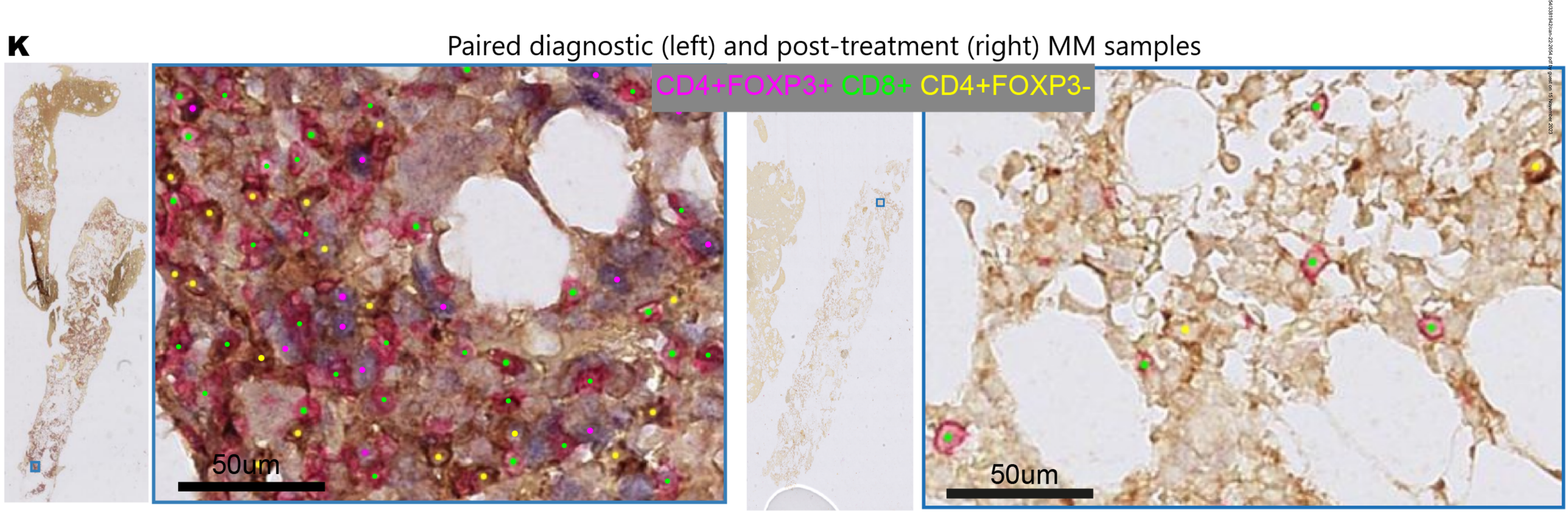

F
10²
n=10
p = 0.013
BLIMP1+ density
NDMM
post-treatment

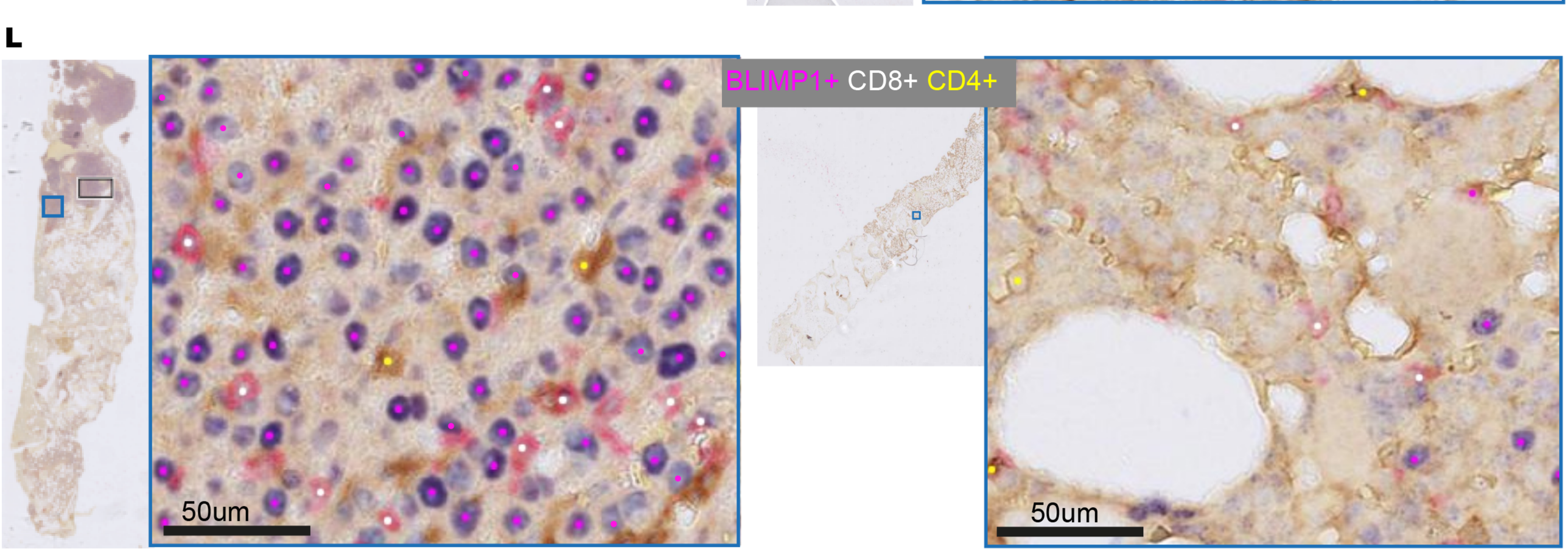

G
n=10
p = 0.27
CD8+:BLIMP1+
NDMM
post-treatment


H
10²
p = 0.32
n=19
FOXP3+CD4+ density
MGUS
NDMM

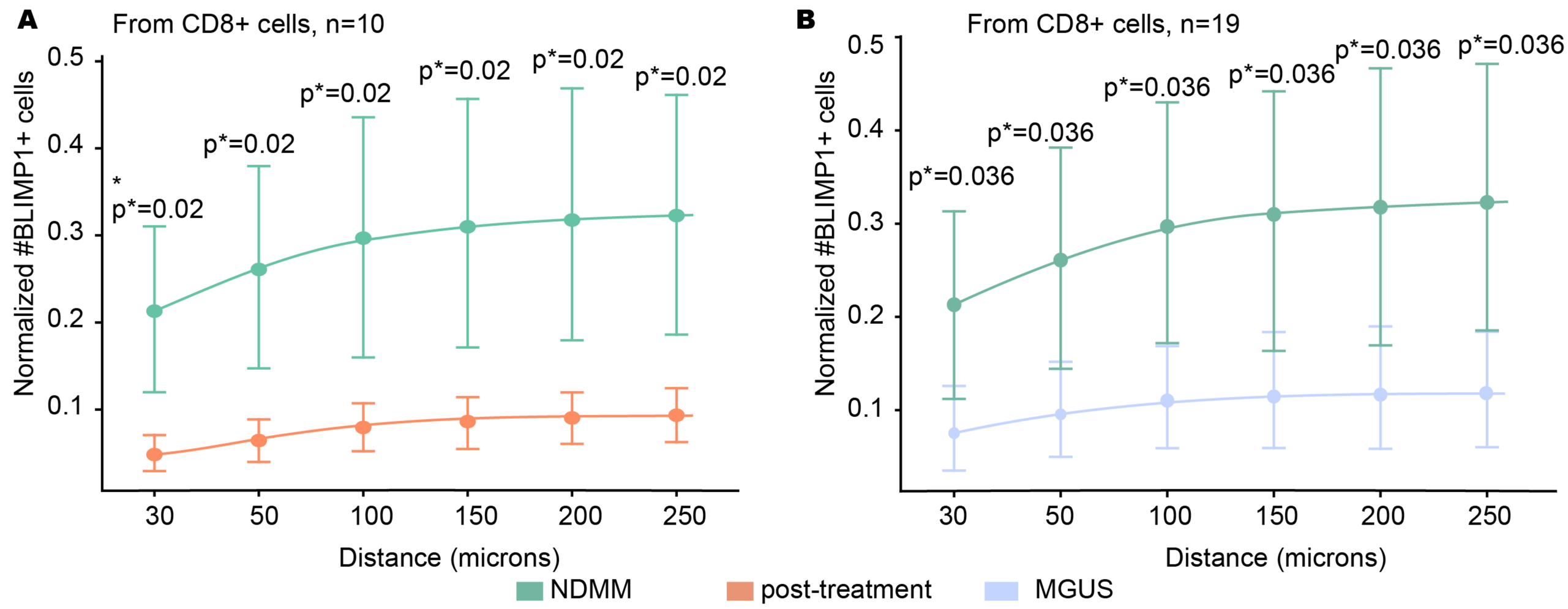

I
10²
n=19
p = 0.08
BLIMP1+ density
MGUS
NDMM

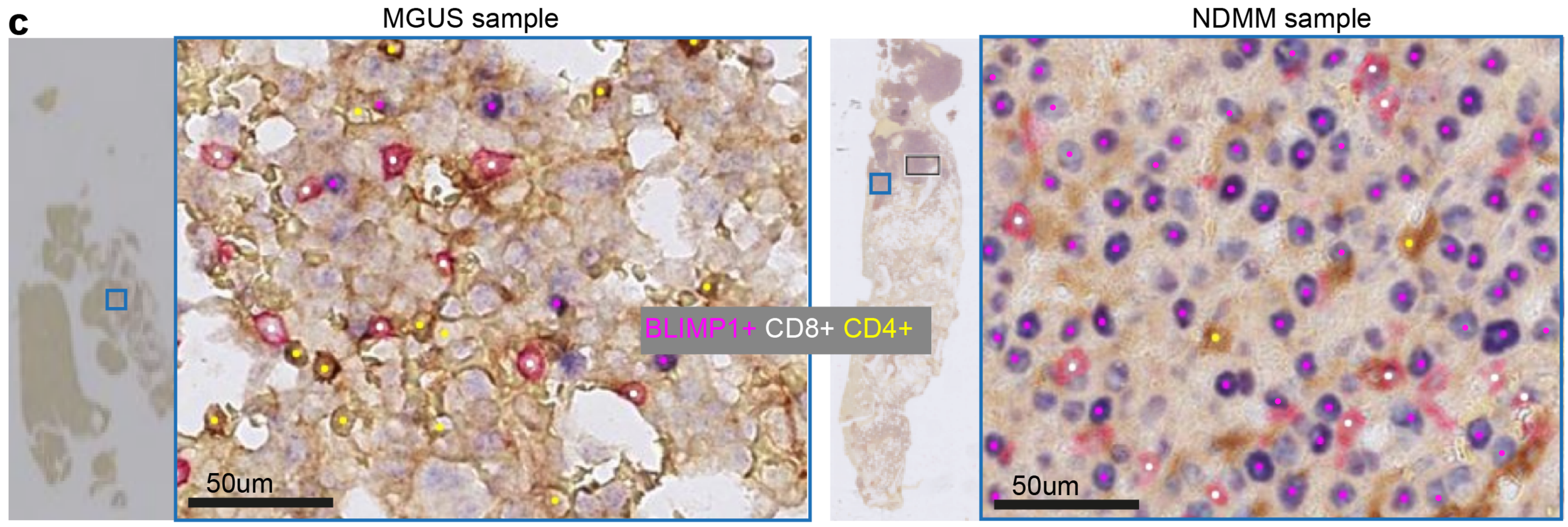

J
n=19
p = 0.17
CD8+ : BLIMP1+
MGUS
NDMM

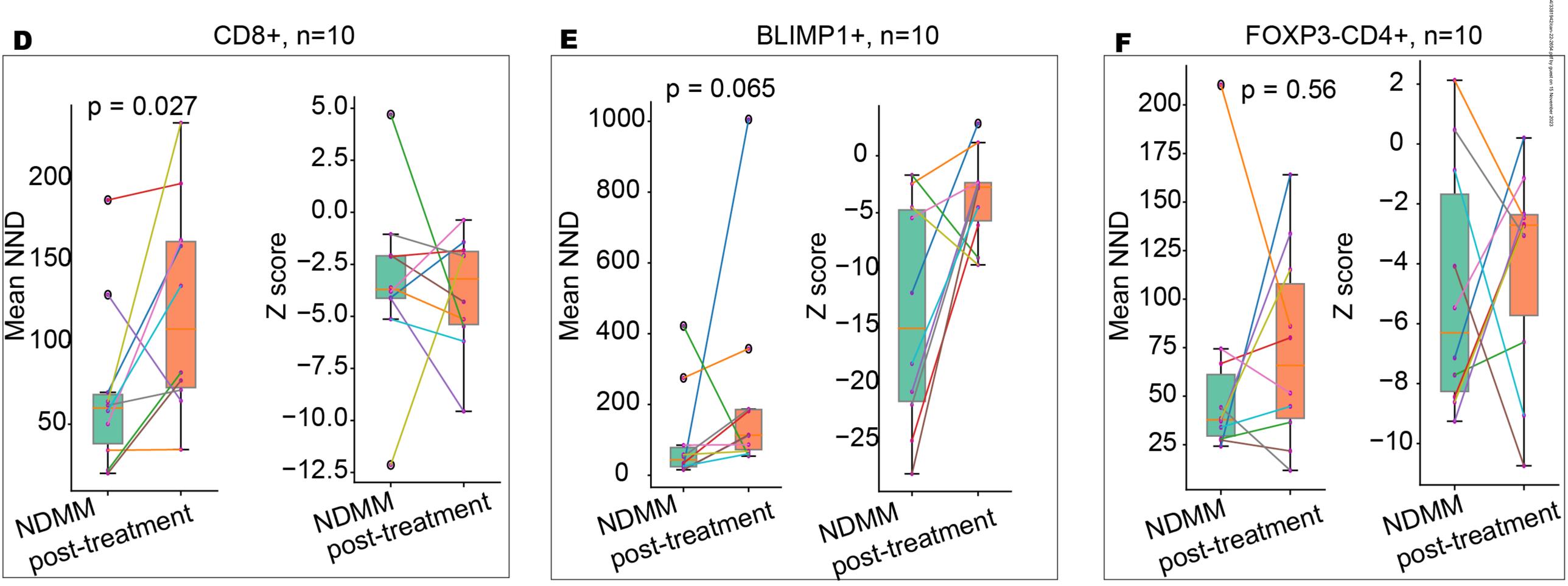

K
Paired diagnostic (left) and post-treatment (right) MM samples
CD4+FOXP3+ CD8+ CD4+FOXP3-
50um
50um

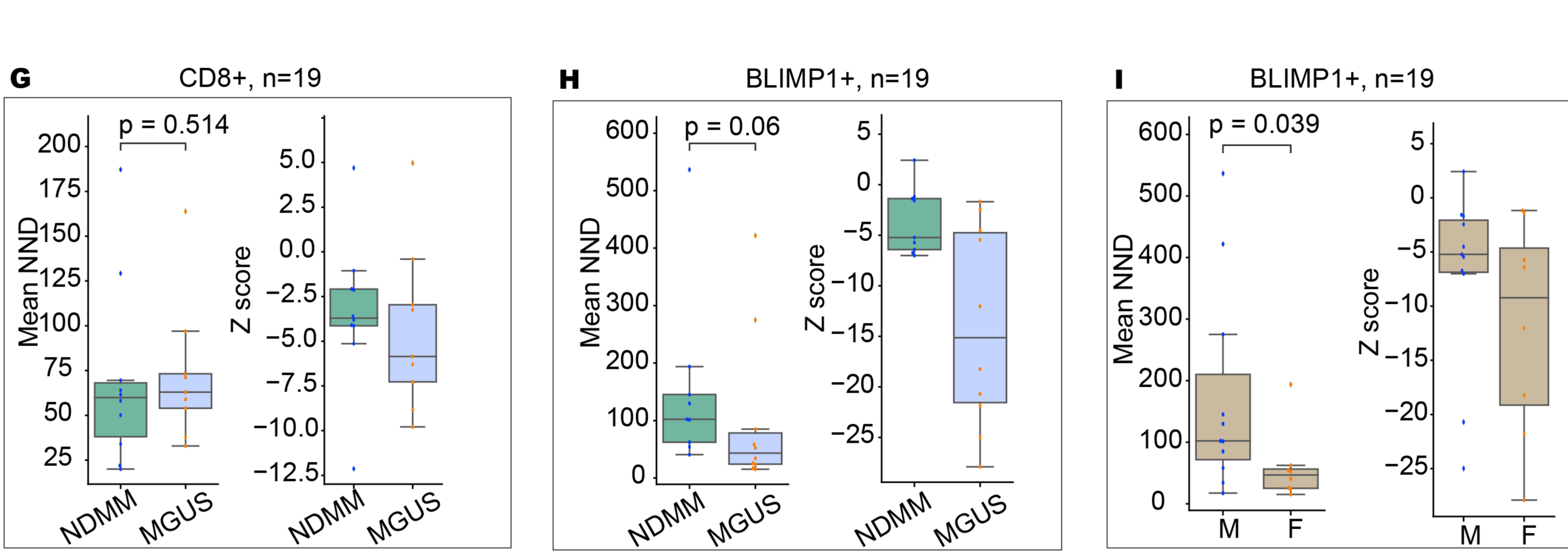

L
BLIMP1+ CD8+ CD4+
50um
50um

# Figure 6


A
From CD8+ cells, n=10
Normalized #BLIMP1+ cells
0.5
0.4
0.3
0.2
0.1
*
p*=0.02
p*=0.02
p*=0.02
p*=0.02
p*=0.02
p*=0.02
30
50
100
150
200
250
Distance (microns)
B
From CD8+ cells, n=19
Normalized #BLIMP1+ cells
p*=0.036
p*=0.036
p*=0.036
p*=0.036
p*=0.036
p*=0.036
Distance (microns)
NDMM
post-treatment
MGUS


c
MGUS sample
NDMM sample
BLIMP1+ CD8+ CD4+
50um
50um


D
CD8+, n=10
p = 0.027
Mean NND
Z score
NDMM
post-treatment
E
BLIMP1+, n=10
p = 0.065
Mean NND
Z score
NDMM
post-treatment
F
FOXP3-CD4+, n=10
p = 0.56
Mean NND
Z score
NDMM
post-treatment


G
CD8+, n=19
p = 0.514
Mean NND
Z score
NDMM
MGUS
H
BLIMP1+, n=19
p = 0.06
Mean NND
Z score
NDMM
MGUS
I
BLIMP1+, n=19
p = 0.039
Mean NND
Z score
M
F