Work Flows for Cellular Epidemiology, From Conception to Translation

Anton D. Nathanson[1]†, Lucy Ngo[1]†, Tomasz Garbowski[2], Abhilash Srikantha[3], Christian Wojek[3], Dirk Zeidler[2], Melissa L. Knothe Tate[1]*

submission as Tools and Resources *type article*

[1]MechBio Team, Graduate School of Biomedical Engineering, University of New South Wales, Sydney, Australia.

[2]Carl Zeiss Microscopy GmbH, Oberkochen, Germany

[3]Corporate Research and Technology, Carl Zeiss AG, Oberkochen, Germany

[4]Cleveland Clinic, Department of Orthopaedic Surgery, Cleveland, Ohio, USA

†Co-first authorship

Abstract: Changes in cell connectivity and morphology, observed and measured using microscopy, implicate a cellular basis of degenerative disease in tissues as diverse as bone, kidney and brain. To date, limitations inherent to sampling (biopsy sites) and/or microscopy (trade-offs between regions of interest and image resolution) have prevented early identification of cellular changes in specimen sizes of diagnostic relevance for human anatomy and physiology. This manuscript describes work flows for human tissue-based cell epidemiology studies. Using recently published sample preparation methods, developed and validated to maximize imaging quality, the largest-to-date scanning electron microscopy map was created showing cellular connections in the femoral neck of a human hip. The map, from a patient undergoing hip replacement, comprises an 11 TB dataset including over 7 million electron microscopy images. This map served as a test case to implement machine learning algorithms for automated detection of cells and identification of their health state. The test case showed a significant link between cell connectivity and health state in osteocytes of the human femur. Combining new, rapid throughput electron microscopy methods with machine learning approaches provides a basis for assessment of cell population health at nanoscopic resolution and in mesoscopic tissue and organ samples. This sets a path for next generation cellular epidemiology, tracking outbreaks of disease in populations of cells that inhabit tissues and organs within individuals.

*Corresponding author

Melissa L.Knothe Tate, Ph.D.

Paul Trainor Chair of Biomedical Engineering

University of New South Wales

UNSW Sydney, NSW

Australia

m.knothetate@unsw.edu.au

Introduction

In the complex biosystem of the human body, the prediction of emergent behaviors and their underlying mechanisms is key to early diagnosis, intervention and/or prevention of disease. Yet such predictions have presented a scientifically and technically untenable research challenge. This is due to the intrinsic hurdles of imaging cell populations across tissues and organs and analyzing data from billions of interconnected cells. The combination of geospatial approaches, using multibeam scanning electron microscopy (multiSEM), with the Google Maps JavaScript API, has enabled cellular studies of aging human tissues, from bone to brain (*1–4*). In contrast to single beam electron microscopy where nano-micrometer scale regions of interest (ROI) are imaged in millimeter-sized samples, multiSEM uses parallel beams (up to 91 in current commercial systems) (*4*), enabling rapid throughput, nanoscale resolution imaging of millimeter scale ROIs in milli- to centimeter sized, tissue and organ samples (*19,21-24*). Development of workflows integrating such geospatial approaches with advances in machine learning algorithms are expected to enable high-throughput, rapid image feature detection and classification, cornerstones of early diagnostics.

Degenerative processes associated with aging and disease are observable in our cells, the living inhabitants of our tissues and organs. Cell viability and connectivity play a critical role in human health. Their loss is associated with loss of function in degenerating cellular networks. Until recently, an understanding of multiscale structure-function relationships, within, between and across tissues has been limited by imaging capabilities (*2*). Imaging fields of view have, until recently, been limited to either local (cellular) or global (tissue, organ) perspectives, preventing early detection of disease emergence. Early detection greatly increases chances of successful treatment of cancer, dementia, osteoporosis and chronic kidney disease (*2*).

This manuscript describes workflows for human tissue- and organ-based, cell epidemiology studies. We integrate a recently developed sample preparation protocol (*25*) with previously described image acquisition, stitching and mapping methods *(3, 14–16, 17)*, to acquire the largest to date multiSEM map of a human hip, an 11 TB dataset comprising over seven million electron microscopy images (Fig. 1). We then implemented machine learning algorithms, which were trained using geographical data from manual pinning of relevant landmarks previous maps of human tissues and their cellular inhabitants (Fig. 3-8), to determine feasibility of the method for rapid throughput imaging and diagnostics.

Methods

Sample Preparation Workflow Traditional electron microscopy necessitates specimen size be limited to millimeters on edge and progression to much larger specimens, up to 10 cm in diameter, challenges basic physical laws, *e.g.* maximal pathlengths of penetration for chemical fixatives and embedding media, as well as preparation of sample surfaces for imaging. Standard EM methods serve well for mouse sized samples and smaller (*19–21*), but new protocols needed to be developed to prepare macroscopic samples from human and large mammals for imaging with multiSEM (*25*).

Hence we sought to improve the preparation of embedded samples to facilitate imaging of human and large mammalian tissues at nanometer scale resolution. Study samples were obtained

from the human femoral neck (Cleveland Clinic Institutional Review Board), which is normally discarded after hip replacement surgery. Sheep (*Ovis aries*) femur and Guinea pig (*Cavia porcellus*) knee samples were also used to test multimodal imaging modalities.

The workflow applies to any human and large mammalian tissues available for collection and applicable for the specific study design. For the purposes of the current study, we were interested in mapping the connectivity of osteocytes in the human hip. 3 - 5 mm coronal and transverse sections of femoral neck (up to 10 cm in diameter) were acquired from patients (Cleveland Clinic IRB approved) undergoing hip replacement; these tissues are discarded in the normal course of arthroplasty. The sections were prepared using methods previously developed to image at nanoscopic resolution the osteocyte and their interconnecting processes using atomic force microscopy studies (*22,23,25*)**.**

The tissues (undecalcified) were fixed in 2.5% glutaraldehyde, 4% formaldehyde, and 0.2M cacodylate buffer at 4°C and then processed for bulk embedding in poly(methyl methacrylate) (PMMA). Curing in a vacuum in a cold environment facilitated gradual polymerization to minimize artifacts. Once the embedding medium polymerized, specimen surfaces to be imaged were milled using a CNC-mill for mirror-like planarity. Sample surfaces were then prepared for etching to reveal the cellular network, after which carbon coating and imaging took place.

Stepwise etching was carried out using 0.02M HCl for 90s and/or 10% NaOCl, to image the respective organic or inorganic phase of the extracellular matrix (*22,23,25*)**.** By etching away the matrix around the cells, layer by layer and capturing each step with multiSEM and other imaging modalities, it is possible to render the three dimensional network for later multi-modal, correlative microscopy analysis. Prior to multiSEM imaging, specimens were carbon sputtered.

MultiSEM imaging Samples was imaged with a Zeiss MultiSEM 505 imaging system. This microscope uses 61 parallel electron beams arranged hexagonally to minimize electron-optical aberrations. Imaging typically used a landing energy in the range of 1–3 keV, 100 ns of dwell time per beam, resulting in a resolution of 3.5nm.

For the newest dataset (Fig. 3), a surface area spanning 18.1 cm^2 was imaged, resulting in 120,262 hexagonally shaped multi-beam fields of view (mFOV), made of more than 7.335 million high-resolution image tiles and a total of 10.7 billion megapixels (Table 1). Each mFOV was composed of 61 rectangular, single-beam image tiles arranged in a flat, hexagonal pattern (Fig 1), with a frame size of 1288 x 1120 pixels for each tile. The stage used for this study operated with a precision of 2 μm. Image files were stored as bitmap files, accumulating circa 10.98 TB of storage space. Pixel coordinates for individual single-beam images were available from the microscope metadata, providing a first approximation for relative positioning.

Machine Learning Workflow
Overview
Recent work from our lab (*17*) has investigated the use of the navigable bone maps, and has sought solutions for the technical challenges inherent to handling and analyzing massive imaging data sets for biomedical and diagnostic applications. A previous proof-of-principle study to

3

elucidate the relationship between bone tissue and cell health in the human femoral neck, the Google JavaScript API was used to enable labeling of landmarks on a navigable map. A blinded observer marked blood vessel edges, viable and pyknotic osteocytes. This coordinate-based mapping enabled testing of specific hypotheses related to bone health in terms of osteocyte viability, transport path distances, and network relationships. While adequate for feasibility testing, this manual approach was not conducive to the pipeline scalability necessary for the magnitude of current datasets (Table 1).

Here we aimed to combine high throughput multiSEM with advanced convolutional machine learning algorithms to increase throughput for cellular and feature detection in such large samples, while reducing observer bias and sampling errors (*26,27*). We implemented a machine learning algorithm for osteocyte detection based on the YOLO convolutional neural network, originally described by Redmon *et al.* (*28*) (Figure 4). Using the osteocyte coordinates extracted from this algorithm, a novel script was created to compare the location of each osteocyte with every other osteocyte and extract the minimum distance to its nearest neighbor. This novel framework enabled high-throughput, large-scale analysis of cellular network relationships in human samples, consisting of over seven million multiSEM images and 11TB of data. We compared and tested basic network hypotheses using automated, machine learning-based and observer-marked (pinned) osteocyte detection in EM maps of human hips to test feasibility of the automated workflow.

Data Description

The machine learning workflow feasibility study was based on two mSEM datasets, including images and coordinates for generating the respective stitched images (Table 1). The smaller dataset ("Previous Dataset", Table 1), which included manually entered (pins dropped, see above) expert annotation indicating health status as "viable" and "pyknotic" (cells in necrotic or apoptotic state), was downsampled by a factor of 8 to obtain the stitched image of size 44288 x 39936 pixels or 560MB. These annotations alone were insufficient for the object detection task, and were thus extended further to bounding boxes by a non-expert (equivalent to a "blinded observer", Fig. 4). The resulting dataset, compiled for object detection training, contains samples of 630 viable and 49 pyknotic cells.

The "Current Dataset" (Table 1) was much larger, *i.e.* on the order of 10TB, and included the coordinate information used to obtain the stitched image. Similar to the smaller, "Previous Dataset", the "Current Dataset" was downsampled by a factor of 8. In contrast to the previous, smaller dataset, this dataset included no annotations for cell health states.

To test the feasibility and utility of using machine learning to automatically detect viable and pyknotic cells, the model was trained on the "Previous Dataset" and tested on the "Current Dataset".

Training the Algorithm

The following quantitative evaluation was based on the annotated ("Previous") dataset, using a 60/40 train/validation split. Due to the limited training data and high class imbalance, we performed data augmentation to obtain 10,000 examples for each class. Augmented samples were obtained by randomly selecting the following operations: rotation [0, 360•], image scaling [0.8x to 1.2x], gamma correction factor [0.95 to 1.05], scaling illumination [0.8 to 1.2] and

translation [-80 to 80] (Figure 6). Training examples were obtained by cropping to 500x500 pixels, centered at the transformed images. Further, single channel mSEM images are converted to 3-channel gray scale images by replicating data.

Object Detection Model

We chose the YOLO object detection model due to its performance [YOLO9000] and reasonable runtimes. As the proposed approach was flexible with respect to model choice, quantitative analysis presented in our experiments serves as a baseline for comparing with other techniques. The 2-class detection model is obtained by finetuning all weights of a model pretrained on ImageNet [ImageNet]. For finetuning, we use momentum 0.9, learning rate 0.0001, batch size 64, decay 0.0005, maximum batch size 45000 and step-policy at 100, 25000 and 35000 iterations to scale down the learning rate by 10, 0.1 and 0.1 respectively. The anchors are left unchanged. Early stopping is used with a loss threshold of 1^{e-8}. The model was trained on GTX1080 in under 36 hours. During testing, detections for a stitched image are obtained by accumulating those obtained by testing non-overlapping 500x500 crops. Each detection encodes the bounding box [x, y, width, height], a confidence score [0 to 100] and a class id.

Model Evaluation

We used the AUC metric [PascalVOC, microsoftCOCO] in order to evaluate the object detection performance. Under this metric, a detected bounding box is said to be successful if its intersection over union (IoU) ratio with respect to the groundtruth bounding box is greater than a threshold, *e.g.* 0.50. Precision-recall curves for a random 60/40 train/validation split are indicative of performance (Figure 7). Results were observed to be consistent across four different splits. The AUC0.30 for living and deceased were 0.76 and 0.79 respectively (Figure 7). This is a relaxed measure as the overlap for detected and groundtruth bounding boxes is not stringent. For a more stringent criterion, *i.e.* AUC0.50, the performance drops slightly to 0.727 and 0.782. At a highly stringent AUC0.75, the performance drops to 0.60 and 0.67. As the cells are localized "roughly" (simple identification, not at a particular location of the cell), AUC0.50 provides an optimal trade-off between object detection performance and its localization capability (Figure 7).

A qualitative analysis of detected samples at a high detection confidence of 75% shows detection and a rare missed detection (Figure 3D). Based on this data, the false positives are either structurally similar to true positive detections or illustrate noisy annotations. This indicates the need for more training data or iterative model learning.

With regard to runtimes, testing for the first ("Previous Dataset") took 2 hours on a GTX1080 system. Under similar conditions, the current data set ("Current Dataset") took *circa* 100 hours to be tested.

Integrated Approaches to Test Network Based Hypotheses

Our initial focus investigated network relationships in tissue samples based on the path distances between all osteocytes detected using the YOLO algorithm (Figures 3, 4). The neural network identified a total of 206,180 osteocytes in 100 hours on GPU (GeForce GTX 1080) compared to the manual pinning method of our previously published work (*17*) that identified 708 osteocytes manually over several weeks.

Previous work, using manual pinning methods with the Google Maps JavaScript API, tested two network hypotheses, showing no significant relationship in path distance between viable osteocytes and the nearest blood supply albeit significant relationships between osteocyte viability and distance to nearest the viable osteocyte (*17*). The current investigation built further on this work to include the relationship from each osteocyte (whether viable or pyknotic) to all others per sample (Figure 8). A new script was written to measure the distance from every identified osteocyte to others in its near surrounding. From this, the identity of the osteocyte with the closest absolute distance was extracted.

ML algorithm training

Automated object detection algorithms such as the YOLO neural network (*28*) facilitate rapid throughput diagnostic assessment of imaging datasets, also mitigating the effects of observer bias. Initially, YOLO was trained for automated osteocyte detection, using 629 annotated cells, which were further augmented to 10^6 examples through variation by rotation, scale and contrast. Unseen images were then processed with YOLO and automatically detected objects were identified by bounding boxes. Latest testing of the YOLO algorithm has proved successful in detecting osteocytes in a complete cross-section of the human femoral neck in less than 100 hours, and early hypotheses have begun testing. Collection of more (*circa* 1000) false- and missed-detections (*cf.* Figure 3), to obtain a more representative training dataset, provides a straight-forward approach to improve detector performance and is currently underway.

Data analytics were compared for both pipelines and showed similar results. The modified YOLO neural network was readily trainable in context of previous as well as new image features in mesoscopic tissue maps created using multiSEM, and Google Maps API. It will be key for large-scale epidemiology studies that all relevant cell types can be automatically identified, from osteocytes to osteoclasts to blood vessel edges. With procedures now in place for the timely re-training, existing training sets can also be refined to increase the performance of the algorithm, currently performing at better than 92% accuracy for osteocyte detection and classification.

Results

Connectivity of cells in human hip tissue samples implicate network transport patency in osteocyte survival throughout life, and in health as well as disease (*e.g.* osteoporosis, osteoarthritis). Our previous study indicated significant relationships in path distances between
5   viable cells and no significant relationships in path distances between viable cells and the distance to the nearest blood supply. Here, by training the algorithm to identify detected osteocytes as viable (three or more cell processes) or pyknotic (defined as in a state of necrosis or apoptosis, less than three cell processes observed) and classifying cell-to-cell connections as either between two viable osteocytes (viable to viable), between a viable and a pyknotic
10   osteocytes (pyknotic to viable) of between pyknotic osteocytes (pyknotic to pyknotic), we could test further hypotheses that viable osteocytes exhibit shorter path distances to other viable osteocytes compared to path distances to pyknotic osteocytes. (Figure 8)

Significant differences were observed in the mean distances across the different connection
15   types. In both cases, healthy-pyknotic connection distances were significantly larger compared to viable-viable cases. Distances between pyknotic cells were significantly ($p < 0.0001$) lower than those between viable cells. This result was not expected, suggesting either that osteocytes die in clusters and/or networks, unrelated to their distance to the nearest healthy osteocyte or blood supply (as initially thought in previous experiments) or that dying cells send out death signals
20   that impact the local population, not unlike a disease outbreak in an epidemiological context. Future work will implement network analysis tools, such as clustering and central node analysis.

Discussion

25   The multiSEM platform allows for scanning electron microscopy of specimens up to 10 cm in diameter at nanometer length scale. The method may be expanded in the future using methods such as those of Denk *et al.*, who integrate in-chamber milling to achieve serial block-face images that can be constructed in 3D to achieve volumetric data sets in nanoscopic detail (*19*, *20*). Alternatively, Lichtman *et al.* use an automated tape capture system to image ultrathin serial
30   brain slices using SEM, again reconstructing volumes in 3D to create digital tissue volume sets (*21*, *24*). In the current method, a novel etching protocol was adapted from atomic force microscopy protocols (*22*, *23*) to enable serial imaging of nanoscopically etched macroscopic samples, which can be reconstructed digitally in three and higher dimensions when etching organic and inorganic phases of tissues, with multiSEM alone as well in conjunction with other
35   modes of microscopy, for instance multimodal, correlative microscopy. This rapidly developed advancement using multiSEM will further expand the breadth of investigations, enabling correlative studies of structure and function within, spanning across and between length scales, as well as imaging modalities.

40

Current clinical uses of high throughput electron microscopy include the diagnosis of malignant schwannomas (*5*) and other kidney disorders. The pipeline from biopsy to diagnosis is geographically circuitous and occurs in anything but real time. Namely, biopsies are first express-shipped daily from clinics across the world to one of the few major medical clinic core
45   laboratories that offer electron microscopy services. Then EM images are acquired and shared with the pathology department for diagnosis and data entry for later reading of results by the

doctors who originally sent the biopsy. In addition to the obvious limitations of this approach, with regard to lack of rapid diagnostics and loss of time between biopsy and treatment, regions of interest are limited, and diagnostics can be fraught with sampling error, depending on biopsy sampling sites and numbers. These issues provide the impetus to advance throughput and automation of advanced EM technologies for clinical diagnostic applications such early detection of kidney disease (*6*, *7*) as well as early detection of degenerative conditions affecting other physiological systems.

While EM resolution imaging is not yet feasible in live patients *in situ*, advancement of the method is expected to enable a new era of cellular epidemiology. Epidemiology denotes the study of the incidence and spatiotemporal distribution of health events or states related to health. Tracking of such events, *e.g.* tracking epidemic outbreaks worldwide, enables early detection and prevention, thus promoting health of the world's inhabitants. Analogously, cellular epidemiology tracks outbreaks of health-related states or events within and between individuals, enabling early detection and therapy and paving the way for the development of next generation theranostics, *i.e.* combined diagnostics and therapies.

The sample preparation and machine learning algorithms enable rapid, automated identification of image features, including osteocytes, determination of their health status, and first analyses of pathlengths and relationships between healthy and sick cells in osteocyte networks. The resulting dataset analysis demonstrates, for the first time to our knowledge, a highly significant correlation between cell health and proximity to a viable network of osteocytes in aged human femoral neck tissue (Figure 8). Thus, combining multiSEM with new machine learning techniques enables large-scale network health assessment in mesoscopic tissue and, ultimately, organ-sized samples.

The current study demonstrates a clear link between osteocyte connectivity and cell health status in osteoarthritic human femoral necks, validating previous work in smaller volumes of healthy, and diseased tissue imaged using confocal microscopy *(12)*. Further studies will elucidate the exact nature of these networks, the information and signals and cargo they transfer, as well as how they can be harnessed to promote tissue health throughout life. Data mining of the vast datasets acquired with multiSEM and rendered with Google Maps API will enable epidemiological investigations of cellular inhabitants within and between individual patient's tissues and organs, as well as cohorts of such patients. These datasets and resulting epidemiological insights are expected to spawn next generation theranostic as well as diagnostic device design.

References and Notes:

1. M. L. Knothe Tate. Microscopy and the Modern Magellans - From Bee Brains to Human Hips, In: *A new age in scanning electron microscopy: Applications in the life sciences* (*Science*/AAAS, 2017), pp. 19–23.
2. M. L. K. Tate, T. Fath. The Only Constant Is Change: Next Generation Materials and Medical Device Design for Physical and Mental Health. *Adv. Healthc. Mater.* **5**, 1840–1843 (2016).

3. A. F. Pereira, D. J. Hageman, T. Garbowski, C. Riedesel, U. Knothe, D. Zeidler, M. L. Knothe Tate. Creating High-Resolution Multiscale Maps of Human Tissue Using MultiSEM. *PLoS Comput. Biol.* **12**, 1–18 (2016).

4. A. L. Eberle, S. Mikula, R. Schalek, J. W. Lichtman, M. L. Knothe Tate, D. Zeidler. High-resolution, high-throughput imaging with a multibeam scanning electron microscope. *J. Microsc.* **259**, 114–120 (2015).

5. J. B. Taxy, H. Battifora, Y. Trujillo, H. D. Dorfman. Electron microscopy in the diagnosis of malignant schwannoma. *Cancer*. **48**, 1381–1391 (1981).

6. I. Shore, J. Moss. In *Current Diagnostic Pathology* (2002), vol. 8, pp. 207–215.

7. G. A. Mokhtar, S. M. Jallalah. Role of Electron microscopy in evaluation of native kidney biopsy a retrospective study of 273 cases. *Iran. J. Kidney Dis.* **5**, 314–319 (2011).

8. N. Kasthuri, K. J. Hayworth, D. R. Berger, R. L. Schalek, J. A., Concello, S. Knowles-Barley, D. Lee, A. Vázquez-Reina, V. Kaynig, T. R. Jones, M. Roberts, J. L. Morgan, J. C., Tapia, H. S. Seung, W. G. Roncal, J. T. Vogelstein, R. Burns, D.L. Sussman, C. E. Priebe, H. Pfister, J. W. Lichtman. Saturated Reconstruction of a Volume of Neocortex. *Cell*. **162**, 648–661 (2015).

9. J. T. Neumann, T. Garbowski, W. Högele, T. Korb, S. Halder, P. Leray, R. Garreis, M. le Maire, D. Zeidler. High-throughput multi-beam SEM: Quantitative analysis of imaging capabilities at IMEC-N10 logic node. Proceedings Volume 10145, Metrology, Inspection and Process Control for Microlithography XXXI 101451S, http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2257980), p. 101451S (2017).

10. A. L. Keller, D. Zeidler, T. Kemen. High throughput data acquisition with a multiSEM, 92360B (2014).

11. O. Sporns, G. Tononi, R. Kötter. The human connectome: A structural description of the human brain. *PLoS Comput. Biol.* **1** (2005), pp. 0245–0251.

12. M. L. Knothe Tate, M. Detamore, J. R. Capadona, A. Woolley, U. Knothe. Engineering and commercialization of human-device interfaces, from bone to brain. *Biomaterials*. **95**, 35–46 (2016).

13. M. Knothe Tate, T. Falls, S. Mishra, R. Atit. In *New Perspectives in Mathematical Biology*, S. Sivaloganathan, Ed. (American Mathematical Society, Providence, Rhode Island, 2010; http://www.ams.org/fic/057), pp. 113–134.

14. L. W. Swanson, J. W. Lichtman. From Cajal to Connectome and Beyond. *Annu. Rev. Neurosci.* **39**, 197–216 (2016).

15. D. Hageman, A. F. Pereira, D. Zeidler, U. Knothe, L. Gardner, M. L. Knothe Tate. Cellular Epidemiology of Human Disease using Biogeographic and Google Maps Approaches - Towards Definition of Cell Network Indices for Rapid Diagnostics. *Ann. Biomed. Eng.* (2016).

16. H. Chang, D. Docheva, U. R. Knothe, M. L. Knothe Tate. Arthritic Periosteal Tissue From Joint Replacement Surgery: A Novel, Autologous Source of Stem Cells. *Stem Cells Transl. Med.* **3**, 308–317 (2014).

17. M. L. Knothe Tate, D. Zeidler, A. Pereira, D. Hageman, T. Garbowski, S. Mishra, L. Gardner, U. R. Knothe. Organ-to-Cell-Scale Health Assessment Using Geographical Information System Approaches with Multibeam Scanning Electron Microscopy. *Adv. Healthc. Mater.* **5**, 1581–1587 (2016).

18. J. Kornfeld, W. Denk. Progress and remaining challenges in high-throughput volume

electron microscopy. *Curr. Opin. Neurobiol.* **50** (2018), pp. 261–267.

19. W. Denk, H. Horstmann. Serial Block-Face Scanning Electron Microscopy to Reconstruct Three-Dimensional Tissue Nanostructure. *PLoS Biol.* **2**, e329 (2004).

20. B. Titze, W. Denk. Automated in-chamber specimen coating for serial block-face electron microscopy. *J. Microsc.* **250**, 101–110 (2013).

21. K. J. Hayworth, J. L. Morgan, R. Schalek, D. R. Berger, D. G. C. Hildebrand, J. Lichtman. Imaging ATUM ultrathin section libraries with WaferMapper: a multi-scale approach to EM reconstruction of neural circuits. *Front. Neural Circuits.* **8**, 1–18 (2014).

22. G. C., H. F. Knapp, A. Stemmer, P. Niederer, M. L. Knothe Tate. Investigation of the morphology of the lacunocanalicular system of cortical bone using atomic force microscopy. *Ann. Biomed. Eng.* **29**, 1074–1081 (2001).

23. H. F. Knapp, G. C. Reilly, A. Stemmer, P. Niederer, M. L. Knothe Tate. Development of preparation methods for and insights obtained from atomic force microscopy of fluid spaces in cortical bone. *Scanning.* **24**, 25–33.

24. K. J. Hayworth, C. S. Xu, Z. Lu, G. W. Knott, R. D. Fetter, J. C. Tapia, J. W. Lichtman, H. F. Hess. Ultrastructurally smooth thick partitioning and volume stitching for large-scale connectomics. *Nat. Methods.* **12**, 319–322 (2015).

25. L. Ngo, A. Nathanson, T. Garbowski, U. Knothe, D. Zeidler, M.L.K. Tate (2019) Sample preparation protocol enabling nano-to-mesoscopic mapping of cellular connectomes and their habitats in human tissues and organs, MS ID#: BIORXIV/2019/533448.

26. V. Gulshan, L. Peng, M. Caram, M. C. Stumpe, D. Wu A. Narayanaswarmy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelso, J. L. Mega, D. R. Webster. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA - J. Am. Med. Assoc.* **316**, 2402–2410 (2016).

27. A. Esteva, B. Kuprel, R. A. Novoa, S. M. Swetter, H. M. Blau, S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* **542**, 115–118 (2017).

28. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection (2015), doi:10.1109/CVPR.2016.91.

Author contributions: The work was conceived over a seven year collaboration by M.L.K.T. and D.Z. with recent input from A.D.N and L.N. who are currently leading the project. The specimens were collected and prepared for imaging by M.L.K.T., A.D.N., and L.N. with assistance from the Cleveland Clinic and Zeiss teams. Image acquisition was carried out by the Zeiss Team, led by D.Z. and T.G. in the Demonstration Labs at Zeiss Microscopy GmbH. Sample preparation protocols were developed and described by M.L.K.T., L.N, and A.N. The machine learning algorithm was developed and tested by A.S. and C.W. with D.Z., M.L.K.T. and A.D.N. Data analysis was carried out by A.D.N. with input from L.N., A.S., C.W., M.L.K.T., and D.Z. The manuscript was written by A.D.N., L.N., A.S., and M.L.K.T. and was revised and approved of by all co-authors.

Competing interests: Zeiss provided in kind support for this project, which is of a fundamental and translational nature.

Data availability: Maps are provided for use by the scientific community and lay public at mechbio.org. Datasets may be shared with the scientific community free of charge via a materials transfer agreement and with industry collaborators through a licensing agreement.

Supplementary Materials

Materials and Methods

1 Specimen Acquisition

One challenge for the successful implementation of human sample studies is finding a surgeon who will partner in the research. The surgeon not only serves as the clinical investigator on the ethics protocol, but also his/her team facilitates the identification and screening of potential study participants, attainment of informed consent/enrolment of the patient, and, perhaps most importantly, serves as the first step in procuring and preparing samples for multiSEM imaging. Similarly, identification of personnel in the pathology and/or tissue procurement department of the clinic is essential to facilitate rapid sectioning and immediate fixation of the specimen.

Human femur samples were collected at the Orthopaedic and Rheumatologic Institute of Cleveland Clinic, in accordance with Institutional Review Board (IRB) protocol #12-335 (*16*). This involved collection of tissues ordinarily discarded during hip arthroplasty (age and gender not disclosed). The tissues were prepared according to techniques adapted from previous atomic force microscopy studies (*22*).

Power analyses conducted for previous studies determined a sample size of 5 patients per cohort was necessary. *Post hoc* analyses on these data indicated the robustness of study outcomes; from a translational perspective, even if more than 700 patients were tested to obtain statistical significance, differences (in periosteum derived stem cell regenerative capacity attributable to age and/or disease state) would remain small, due to the small to medium effect size (of age and disease), even if significantly different (*16*).

2 Sample Fixation and Post-Fixation

Biological samples require fixation prior to embedding and implemented were protocols typical for transmission electron microscopy. Immediately after plane sectioning in pathology, specimens were placed in glutaraldehyde, formaldehyde in sodium cacodylate buffer. Thereafter, the sample was washed with buffer. Fixation time is largely dependent on the specimen size and longer times are required to ensure adequate penetration by diffusion. Following fixation, the samples are washed in buffer before sequential dehydration with ethanol (30%, 50%, 75%, 95%).

Biological samples are typically non-conductive and require an increase in secondary electron emission and consequently contrast to discern biologically relevant details. Typically, electron microscopy employs heavy metal staining, uranyl acetate (UA) and lead citrate, combined with osmium tetroxide (OsO_4) in the fixation process to increase contrast. These contrast agents are commonly selected for their ability to highlight subcellular detail. For identification of cells after etching steps (see below), it was also determined that use of UA and OsO_4 was neither necessary nor desirable since the stains causes light and UV sensitivity in the specimen and puts an exogenous colour on the specimen surface, impairing multimodal imaging, *e.g.* using confocal microscopy.

In this protocol we were able to eliminate the use of heavy metals in the fixation and staining process as carbon coating alone provided sufficient contrast and sample stability. This exclusion of additional pre-embedding steps allowed the process to be greatly streamlined, which is especially important with large samples where considerable time may be required for complete infiltration.

3 Resin Embedding

Epoxy resins are commonly used as the embedding media for morphological samples. To find the optimal embedding protocol, initial attempts to use EPON® (epoxy based) embedding medium proved unfruitful as it is traditionally used for samples 1-2 mm on edge. EPON® embedded samples showed topographic differences on the surface of up to 100µm. Incomplete or inconsistent infiltration of EPON® throughout the specimen was also observed and likely resulted from the larger-than-typical sample sizes. It is possible that the viscosity of EPON® was too high, too few dilution steps were undertaken, and/or the immersion was too short, resulting in insufficient EPON® penetration into the tissue pores. Additionally, there may have been ethanol and propylene oxide remaining in the sample, which would have diluted the EPON® and interfered with polymerization.

For these reasons as well as an abundance of experience with polymethyl methacrylate (PMMA) embedding as well as associated *post hoc* etching protocols used on surfaces of PMMA embedded samples, we opted to adapt PMMA embedding protocols to enable longer penetration times and deep embedding of large samples.

We opted to increase the time for dehydration, and to decrease typical viscosity PMMA embedding medium, to ensure complete infiltration of tissue pores and ensure the sample surface would exhibit minimal perturbations. As such tissues were processed for bulk embedding in PMMA to promote gradual polymerization within a vacuum environment. The results were superior to EPON® embedding and were adopted as the protocol standard.

4 Sample Size and Planarity

The rapid throughput of multiSEM enables practical specimen size is far greater than traditional EM, reaching up to 100mm×100mm×30mm. Samples require a surface flatness of ≤ 500 nm/100 µm (peak-to-peak) parallel to the surface of the sample holder, to prevent edge effects caused by increased electron emission at edges and peaks within the specimen.

Upon polymerization of the embedding medium, the specimens were initially polished to achieve mirror-like planarity. Polishing with an automated polishing head resulted in rapid and excessive removal the sample face. Whereas, polishing by hand on an automated polishing wheel with different grit sanding and polishing papers resulted in scratches and opening of macroscopic cracks on surfaces, likely via release of residual stresses. These cracks caused charging artifacts and charging at crack interfaces during imaging.

Precision CNC-milling and hand polishing were used and showed superior results to automated polishing methods. For CNC milling, specimens were placed in the lathe and a carbide insert without a honed edge was used to face the plastic. Then acrylic polish (NOVUS #2) on a soft cloth with some water was used to polish the surface. The CNC-milled specimens did not exhibit any surface cracks and achieved the required planarity.

13

5 Surface Preparation and Serial Imaging

Finally, to reveal the underlying cellular networks as well as molecular details of the collagen and apatite composite matrix, we tested controlled chemical etching, a technique previously developed and used for AFM (*22*, *23*). Selective etching, between imaging steps, involved sectioning off regions of interest by covering the surrounding areas (with electrical tape or nail polish) and creating a resist. Controlled removal of the extracellular matrix involved 0.02M Hydrochloric acid (HCl) and 10% sodium hypochlorite (NaClO), removing organic and inorganic material, respectively. This enabled imaging of   the extracellular matrix and residing cells from correlating tissues of the hip joint complex. Following the initial chemical etch, the sample required carbon coating, to enhance electron conductivity, and was placed under vacuum before imaging using SEM.

Subsequent etching steps allow the visualization of serial planes of the sample and the collection of 3D data. This requires the removal of the carbon coat by immersing the sample in methanol in an ultrasonic cleaner. Once the coating has been removed the sample can be etched again. This can be repeated until the desired depth is reached. The resultant image set may be of considerable size (in the range of tens of terabytes) and a suitable data handling pipelines are required for effective high throughput data processing.

Tables, Figures and Legends

Table 1. Dataset Information.

	Current Dataset	Previous Dataset (*3*)
Approximate Scanned Area (mm^2)	1,810	5.69
Number of mFOVs	120,262	897
Total number of images	7,335,982	54,717
Number of pixels (megapixels)	$1.07{\times}10^{10}$	75,276
Megabytes/ mFOV	84.8	88.2
Total terabytes	10.98	0.08
website for map of dataset	http://www.mechbio.org/sites/mechbio/files/ maps7/index.html username: mechbio password: #google-maps	http://www.mechbio.org/sites/mechbio/files/ maps5/index.html

5

10

15

Figure 1 (next page). Workflow feasibility for a new era of cellular epidemiology - case study on the largest mSEM map to date. A region of the human femoral head was imaged with the 61 beam multiSEM microscope (see Table 1 for details of data set). The femoral head was removed, in the normal course of hip replacement, from the hip of an aged patient with severe osteoarthritis; erosion of the joint surface is evident on the upper left quadrant of the top image. Seamlessly zooming in closer into a region of interest (dashed box) reveals cellular connections, which branch further at the subcell length scale. The complete map can be accessed and explored via http://www.mechbio.org/sites/mechbio/files/maps7/index.html [prior to publication, input of username and password are required to access the dataset]. username: mechbio password: #google-maps
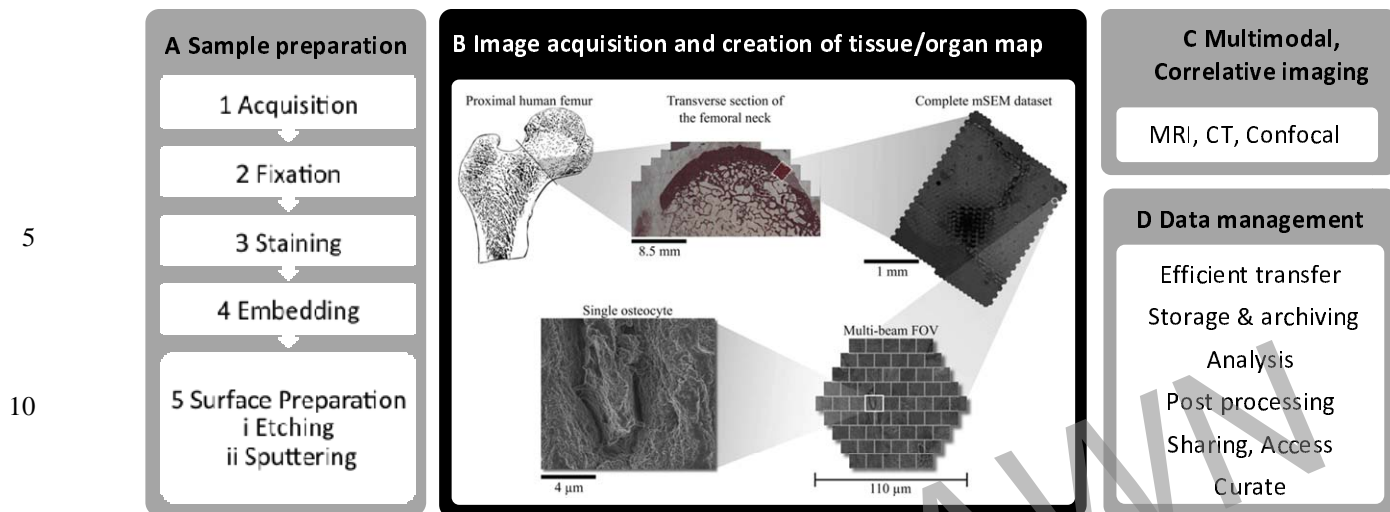
15

5

10
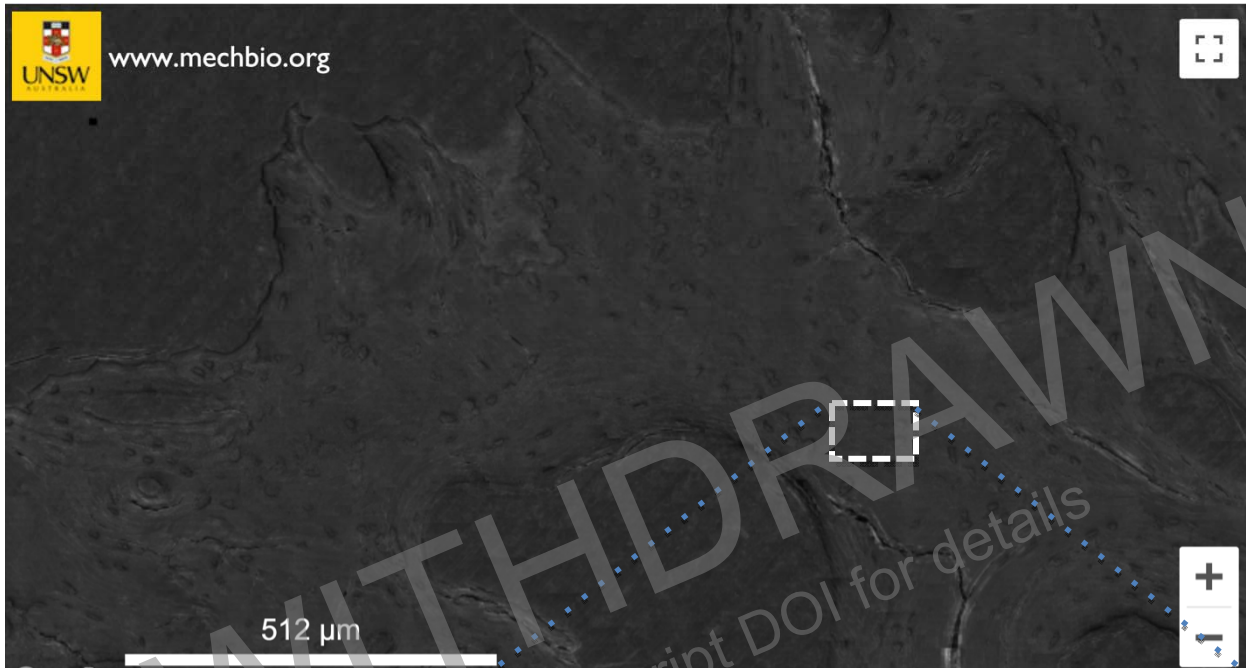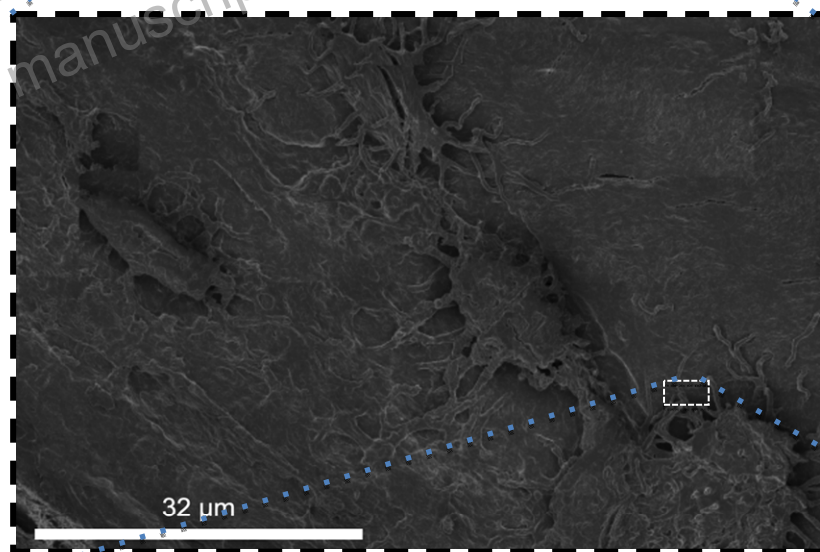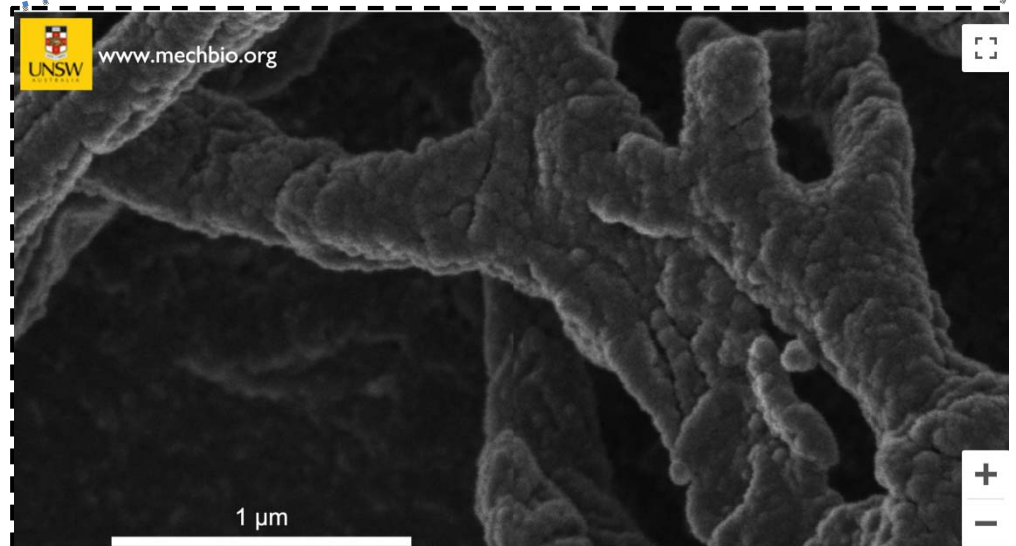
Figure 2. Workflow for acquisition and epidemiological analysis of nanoscale cell population data using mesoscopic samples.
A Sample preparation work flow is a critical first step in insuring data quality and impact.
B Image acquisition and creation of organ maps, described in previous publications *(3, 4, 15–17), and used with permission (4)*.
C Multimodal imaging and **D** data management workflows enable epidemiological study of cellular populations within individuals.
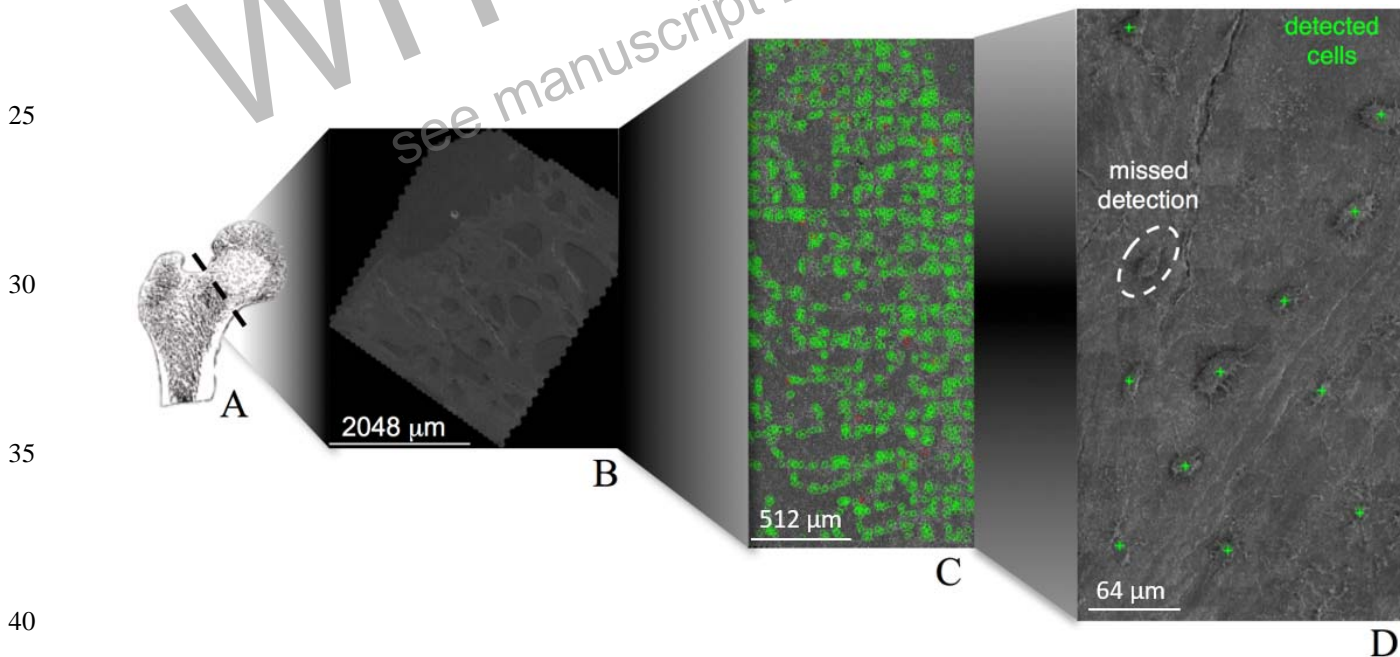
Figure 3. Pipeline utilizing a trained neural network for automatic osteocyte detection in human femoral neck tissue from patients undergoing hip arthroplasty.
A,B This largest-to-date imaged multiSEM dataset comprises 11 TB of images acquired with the 61-beam Zeiss MultiSEM 505, allowing for automatic detection of over 200,000 osteocytes (**C,D**).
The complete map can be accessed via
http://www.mechbio.org/sites/mechbio/files/maps7/index.html
[prior to publication, input of username and password are required to access the dataset].
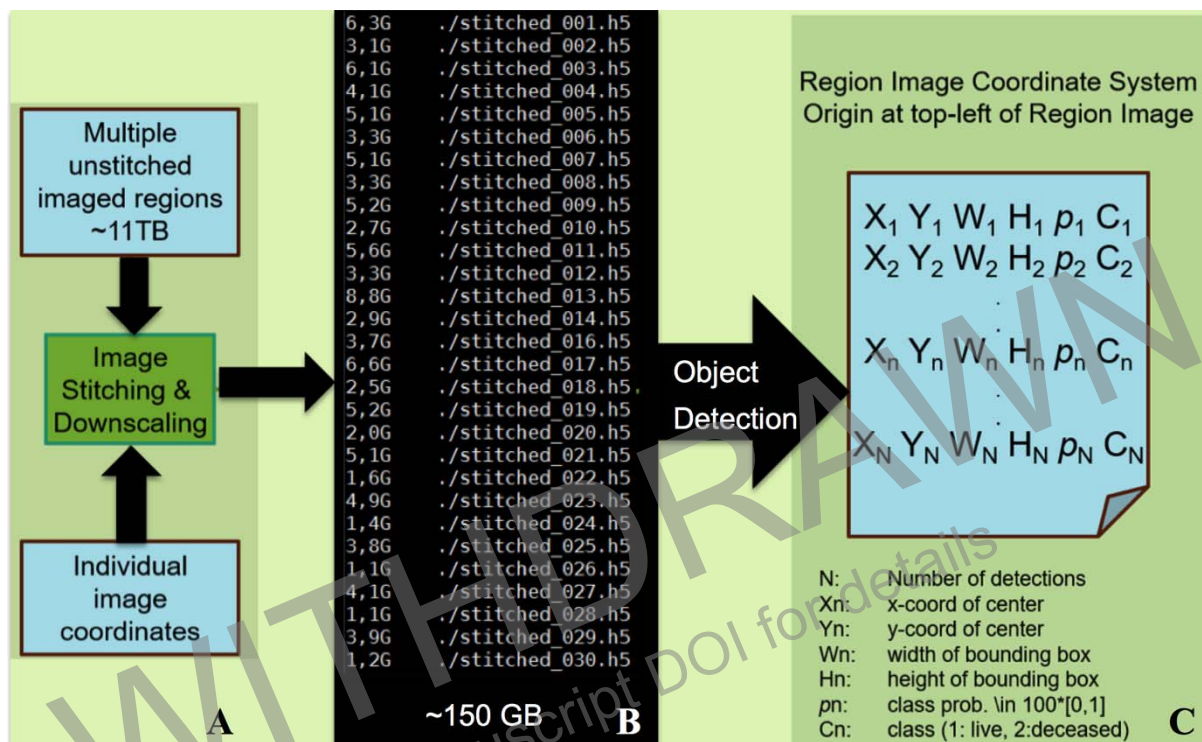username: mechbio
password: #google-maps

Figure 4. YOLO (You Only Look Once) machine learning system description.
The multiSEM output is pre-processed **A** by stitching individual images into region-wide panoramas by virtue of the recorded image coordinates. In the interest of computational efficiency, the resulting image is down-scaled such that individual cells occupy *circa* 200x200 pixels. Here, the 11TB multiSEM output is stitched into 30 image regions, amounting to 150GB of data after downscaling. The pretrained object detector **B** is applied. The output is a file listing the location of each detected cell as bounding box (X,Y,W,H), class C (viable/pyknotic) and associated confidence p (**C**) . The object detector is pretrained using 600 living and 50 pyknotic examples *(17)* and takes 12hrs to train on a single GTX1080. The testing phase on the 150GB dataset lasts 100hrs (**B**).
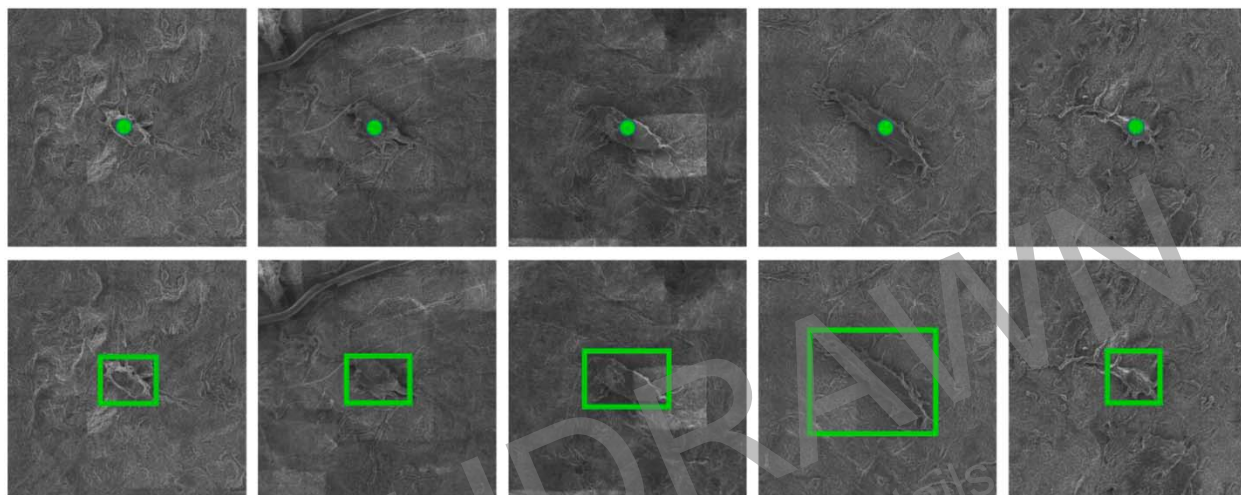
Figure 5. Manual annotations (dropped pins) and extended bounding box notations for object detection. The median width and height of the bounding boxes are 129 and 126 pixels respectively.
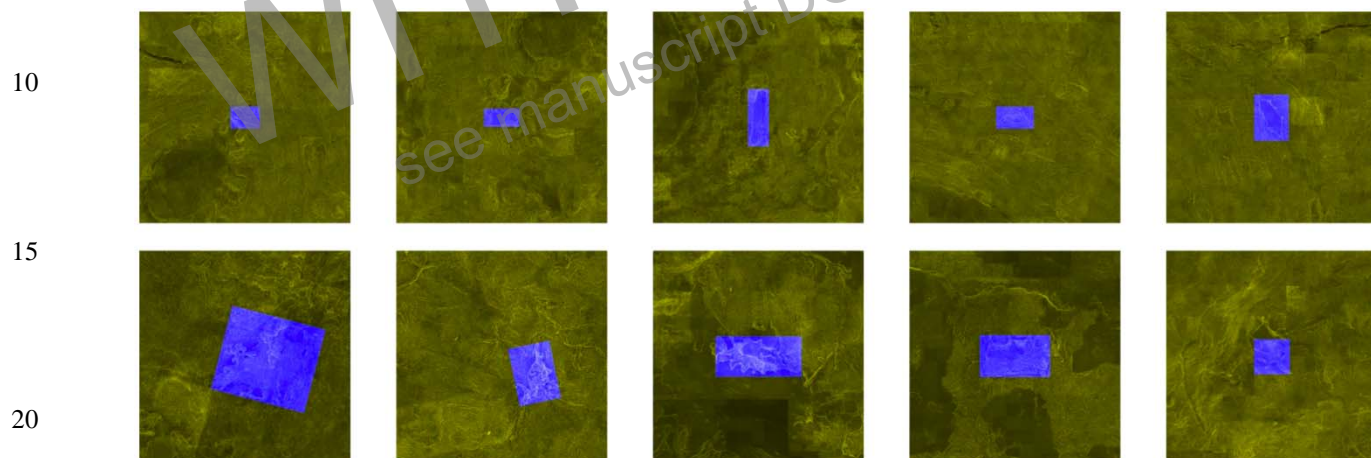
Figure 6. Annotated and augmented examples. To aid visualization, cell regions of interest are overlaid in blue with the background in yellow. Top: viable cells, annotated manually by pinning, were bounded with a tight bounding box. Bottom: image augmentation was performed through scale, rotation, illumination and translation transformations on annotated samples. Note the resulting transformed foreground regions.
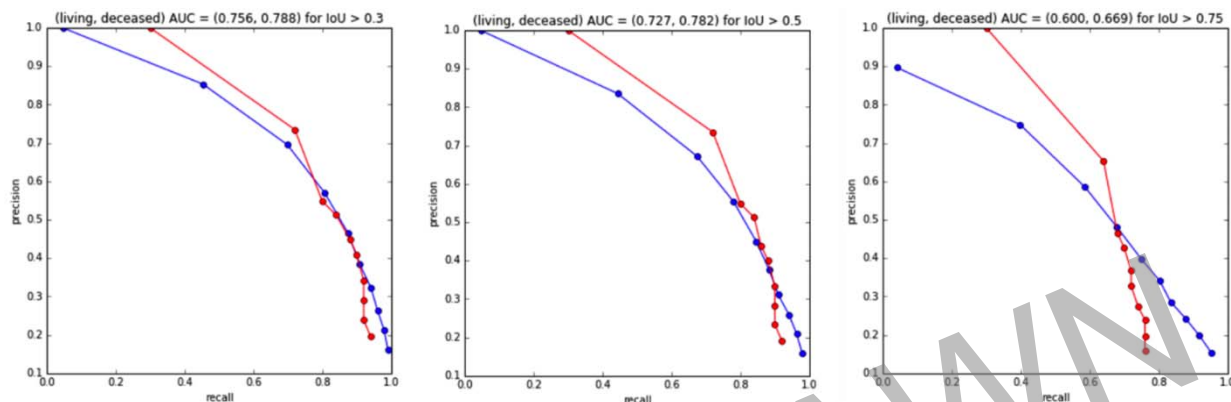
Fig. 7. Precision-recall curves for increasing the IoU threshold. Object detection performance is depicted for viable (blue) and pyknotic (red) cells for increasingly stringent overlap criteria. A practical tradeoff is IoU>0.5 which preserves detection performance of 0.73 and 0.78 while accurately localizing cells.
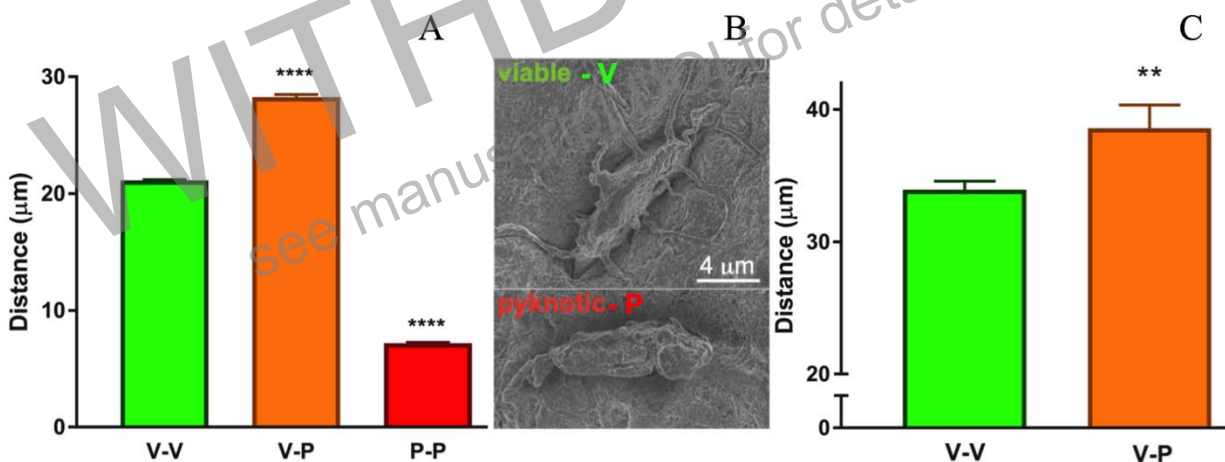

Figure 8. **Machine learning algorithms enable automatic detection of cells and their health states.** This scales up to enable rapid throughput detection of cells and testing of significant relationships among cells, i.e. cellular epidemiological studies of mesoscopic tissue and organ samples.
(**A,B**) Mean distance to the nearest osteocyte is depicted for viable - viable (V - V) osteocyte connections (n=158,905), viable-to-pyknotic connections (V - P, n = 12,986) and pyknotic-to-pyknotic (P - P) connections (n=34,288). Cells were identified and categorised using the trained YOLO neural network. Kruskal-Wallis with Dunn's multiple comparisons test was performed. (**B,C**) The same analysis was performed on a previous, manually pinned data set, comparing healthy-to-healthy connections (n=619) and healthy-to-pyknotic connections (n=89). Mann-Whitney t-test was performed. **p<0.005, ****p<0.0001 compared to healthy-healthy connections. Data are expressed as mean ± SEM.