

Research Article

MarrowQuant 2.0: A Digital Pathology Workflow Assisting Bone Marrow Evaluation in Experimental and Clinical Hematology

Rita Sarkis^{a,b,c}, Olivier Burri^d, Claire Royer-Chardon^e, Frédérica Schyrr^a, Sophie Blum^a, Mariangela Costanza^f, Stephane Cherix^g, Nathalie Piazzon^e, Carmen Barcena^{e,h}, Bettina Bisig^e, Valentina Nardiⁱ, Rossella Sarro^{e,j}, Giovanna Ambrosini^k, Martin Weigert^l, Olivier Spertini^f, Sabine Blum^f, Bart Deplancke^m, Arne Seitz^d, Laurence de Leval^e, Olaia Naveiras^{a,f,*}

^a Laboratory of Regenerative Hematopoiesis, Institute of Bioengineering & ISREC, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; ^b Department of Biomedical Sciences, University of Lausanne (UNIL), Lausanne, Switzerland; ^c Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; ^d Bioluminescence and Optics Core Facility, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; ^e Institute of Pathology, Department of Laboratory Medicine and Pathology, Lausanne University Hospital and Lausanne University, Lausanne, Switzerland; ^f Hematology Service, Departments of Oncology and Laboratory Medicine, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland; ^g Department of Orthopaedics and Traumatology, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland; ^h Department of Pathology, Hospital 12 de Octubre, Madrid, Spain; ⁱ Department of Pathology, Massachusetts General Hospital, Boston, Massachusetts; ^j Institute of Pathology, Ente Ospedaliero Cantonale (EOC), Locarno, Switzerland; ^k Bioinformatics Competence Center (BICC), UNIL/EPFL Lausanne, Switzerland; ^l Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; ^m Laboratory of Systems Biology and Genetics, Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics, Lausanne, Switzerland

ARTICLE INFO

Article history:

Received 22 November 2022

Accepted 15 December 2022

Available online 10 January 2023

Keywords:

adiposity
bone marrow
cellularity
digital pathology
hematopathology
open-source
stroma

ABSTRACT

Bone marrow (BM) cellularity assessment is a crucial step in the evaluation of BM trephine biopsies for hematologic and nonhematologic disorders. Clinical assessment is based on a semiquantitative visual estimation of the hematopoietic and adipocytic components by hematopathologists, which does not provide quantitative information on other stromal compartments. In this study, we developed and validated MarrowQuant 2.0, an efficient, user-friendly digital hematopathology workflow integrated within QuPath software, which serves as BM quantifier for 5 mutually exclusive compartments (bone, hematopoietic, adipocytic, and interstitial/microvasculature areas and other) and derives the cellularity of human BM trephine biopsies. Instance segmentation of individual adipocytes is realized through the adaptation of the machine-learning-based algorithm StarDist. We calculated BM compartments and adipocyte size distributions of hematoxylin and eosin images obtained from 250 bone specimens, from control subjects and patients with acute myeloid leukemia or myelodysplastic syndrome, at diagnosis and follow-up, and measured the agreement of cellularity estimates by MarrowQuant 2.0 against visual scores from 4 hematopathologists. The algorithm was capable of robust BM compartment segmentation with an average mask accuracy of 86%, maximal for bone (99%), hematopoietic (92%), and adipocyte (98%) areas. MarrowQuant 2.0 cellularity score and hematopathologist estimations were highly correlated ($R^2 = 0.92-0.98$, intraclass correlation coefficient [ICC] = 0.98; interobserver ICC = 0.96). BM compartment segmentation quantitatively confirmed the reciprocity of the hematopoietic and adipocytic compartments. MarrowQuant 2.0 performance was additionally tested for cellularity assessment of specimens prospectively collected

These authors contributed equally: Olivier Burri and Claire Royer-Chardon.

These authors co-directed the work: Laurence de Leval and Olaia Naveiras.

* Corresponding author.

E-mail address: olaia.naveiras@unil.ch (O. Naveiras).



from clinical routine diagnosis. After special consideration for the choice of the cellularity equation in specimens with expanded stroma, performance was similar in this setting ($R^2 = 0.86$, $n = 42$). Thus, we conclude that these validation experiments establish MarrowQuant 2.0 as a reliable tool for BM cellularity assessment. We expect this workflow will serve as a clinical research tool to explore novel biomarkers related to BM stromal components and may contribute to further validation of future digitalized diagnostic hematopathology workstreams.

© 2022 THE AUTHORS. Published by Elsevier Inc. on behalf of the United States & Canadian Academy of Pathology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

The use of quantitative digital pathology is growing rapidly because it enables an objective, robust, and automatic assessment of stained slides to inform pathologists on diagnostic and prognostic parameters.^{1–6} However, the implementation of digital pathology comprises many challenges and gaps for translation toward a clinical context, including difficulty integrating clinical workflows within a user-friendly environment. Such integration requires a joint effort from different stakeholders to assimilate the current landscape of multiple separate applications into an integrated image analysis platform and unified workflow compatible with the diagnostic environment.^{1,3–7}

In hematopathology, recent advances in digital and quantitative pathology have renewed interest in developing faster and more quantitative assessments of bone marrow (BM) trephine biopsies, which are a key component in the diagnosis and follow-up of hematologic and nonhematologic disorders.^{8–11} BM biopsy assessment informs on tissue architecture, such as cellularity, necrosis, inflammation, cell lineages, metastatic spread, and stromal modifications,¹² with cellularity being a key factor reflecting hematopoietic function. For instance, low cellularity highlights a central defect in blood production (toxic, constitutional, or idiopathic), whereas high cellularity can suggest a neoplastic transformation.^{8,9,13–18} BM cellularity can be estimated quantitatively (histomorphometry: point-counting) or semiquantitative (visual estimation).^{8,10} It is defined as the relative percentage of the area of the hematopoietic tissue within a BM biopsy specimen and assessed on hematoxylin and eosin (H&E)-stained slides.^{10,13,19–21} It can be either expressed as a percentage of the total marrow space delimited by the endosteal surface^{22,23} or as a percentage of the area occupied by the sum of the hematopoietic and adipocytic compartments,^{8,10,19,24,25} which assumes reciprocity of the hematopoietic and adipocytic compartments within the BM space. Hence, when nonadipocytic elements of the stroma constitute a significant compartment within the BM, the 2 equations might conflict.

Indeed, blood vessels, nerve fibers, and nonadipocytic stromal elements are, in most instances, in minority. Not being the focus of the cellularity assessment, they tend to be neglected when assessing H&E stains. Because these stromal elements are affected in many disorders and different conditions (eg, osteosclerosis, expanded stromal scenarios: fibrosis, edema, or gelatinous transformation),^{8,13,17,18} quantitative efforts that focus exclusively on the adipocytic component of the stroma may limit our understanding of the prognostic parameters associated with stromal remodeling in BM histopathology. For instance, although the assessment of fibrosis is highly standardized,²⁶ very few studies have correlated the architectural characteristics of the BM stroma with clinical outcomes in leukemia.^{27–30} To enable a more thorough assessment of the BM stroma, it is important to remedy the lack of digital pathology tools that

enable the simultaneous tracking of the hematopoietic and stromal BM compartments.

In particular, the historical definition of quantitative BM cellularity assessment comes from the point-counting method, which uses an eyepiece with a graticule. This method is labor intensive, requires large biopsies, and is incompatible with clinical routine. It was first adapted from the method of quantitative morphologic tissue analysis of Chalkley³¹ by Hartsock et al.²⁴ Three different categories were identified when counting with this method: hematopoietic tissue, fat, and other structures. The % cellularity is calculated by dividing the total number of hits on the hematopoietic tissue by the sum of the counted hits on the fat and hematopoietic tissues, expressed in percentages.²⁴ The equivalent visual estimate of cellularity used in clinical routine assessment method is semiquantitative,³² and for highly trained individuals, it correlates with the point-counting method while being faster and simpler.^{19,24,25,33} However, it is still time consuming and semiquantitative and may underestimate cellularity.¹³ Several quantitative digital pathology and deep learning methods have been recently used to quantify BM cellularity and agreed with the gold standard visual semiquantitative estimations.^{34–37} Nevertheless, they were tested only on selected retrospective training and validation data sets focused on the hematopoietic compartment. To further assess the overall heterogeneity of BM compartments, we recently introduced the semiautomated MarrowQuant workflow and extensively validated it on mouse samples.³⁸

In this study, we present MarrowQuant 2.0—an adapted version of MarrowQuant for human BM assessment, implemented as a QuPath script, thus allowing for a user-friendly application.³⁹ We have tested MarrowQuant 2.0 on H&E-stained images from 250 pathologic and nonpathologic BM samples. Our workflow predicts and quantifies 4 major compartments: bone, hematopoietic cells, adipocytes, and the interstitial/microvasculature area (IMV), and groups unassigned pixels into a fifth “other” compartment that includes the expanded stromal compartment (eg, stromal edema and fibrosis). We validated BM cellularity scores in retrospective longitudinal BM trephine biopsies from patients with acute myeloid leukemia (AML) or high-risk myelodysplastic syndrome (MDS) and control orthopedic surgical bone specimens, as well as in prospectively collected samples from clinical routine diagnostics. In particular, we first validated the predictions of the workflow on a training data set, then tested its compatibility with clinical routine samples (test data set), and finally applied the workflow to the most extreme context of BM remodeling (experimental/validation data set), where we could test the assumption of reciprocity implicit in the definition of BM cellularity. Overall, we demonstrate that MarrowQuant 2.0 constitutes a robust, objective, and accurate assessment of BM tissue in a user-friendly platform, open source, and easy to access. Its use may contribute to stromal biomarker discovery, to open up the assessment of specific hematopathologic parameters in research laboratories (with limited access to expert clinicians) and to

homogenize BM cellularity assessments in multisite clinical trials. Potential incorporation in digitalized clinical diagnostic pipelines will face the complex regulatory challenges of pathologist plus machine workstreams.⁴⁰

Materials and Methods

Clinical Samples

This study complied with the Declaration of Helsinki and the local ethical authorities (CER-VD). Three independent sample data sets (H&E-stained images from 250 BM specimens) were collected as detailed further, to form the training, test, and experimental sets from patients treated at the Lausanne University Hospital (CHUV). Figure 1 describes the study design and sample allocation.

Training Set

A retrospective set of 36 anonymous BM trephine biopsies was selected from the Institute of Pathology biobank at CHUV. These biopsies were collected for clinical purposes from patients undergoing treatment for AML or MDS at diagnosis and different times after induction chemotherapy and selected to reflect a range of cellularity or BM remodeling. The samples were processed for

standard pathology diagnosis and analyzed blindly, as previously described.³⁸

Test Set

The test set consisted of 42 prospectively collected trephine biopsies digitalized from all hematopathologic cases received in 2021 over a period of 2 weeks at the Institute of Pathology at CHUV. Diagnostic category, sex, and age are tabulated in Supplementary Table S1.

Experimental/Validation Set

The validation set consisted of 2 cohorts of CHUV specimens: longitudinal follow-up of AML or high-risk MDS trephine biopsies collected for diagnostic purposes and BM specimens collected from the spongy bone in the femoral head or neck of age-matched patients who underwent elective surgical hip replacement. In total, H&E-stained slides from 125 BM specimens were digitalized from 28 patients with AML or high-risk MDS (mean age \pm standard deviation = 55 ± 11 years at diagnosis, sex-balanced) at 4 different time points: diagnosis, the peak of aplasia (days 17-21 postinduction chemotherapy with 7 + 3 cytarabine and idarubicin as in the HOVON/SAKK 132 control arm⁴¹, or C2 with FLAG \pm Idarubicin, or C2 by 5-azatidine for 2 frail patients), hematopoietic recovery after the first cycle of induction chemotherapy (RC1), and hematopoietic recovery after the second cycle of induction

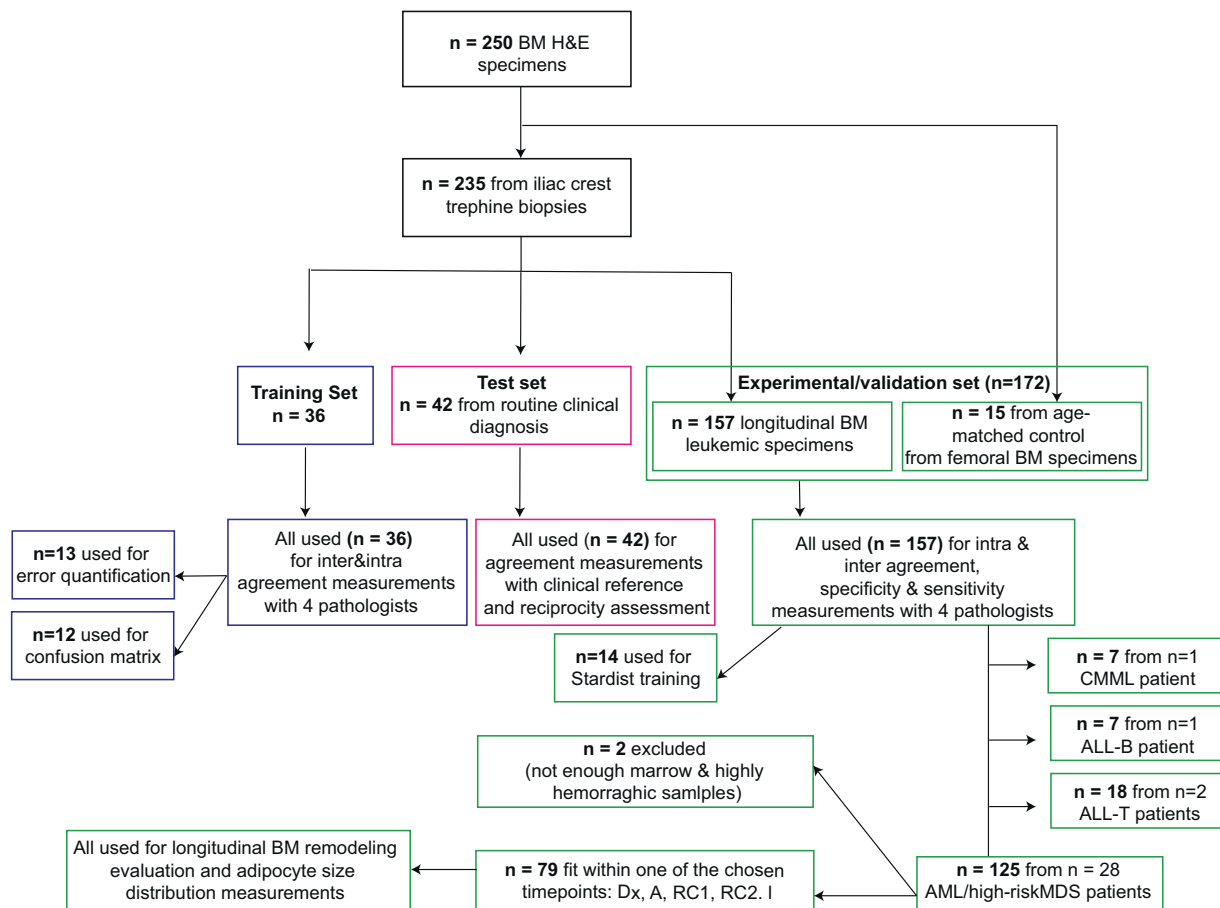


Figure 1.

Study design and image data sets. A, peak of aplasia; ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; BM, bone marrow; CMMML, chronic myelomonocytic leukemia; Dx, diagnosis; H&E, hematoxylin and eosin; MDS, myelodysplastic syndrome; RC1, hematologic recovery after first cycle of induction; RC2, hematologic recovery after second cycle of induction.

chemotherapy (RC2) (Fig. 1). For age-matched control cases, H&E-stained BM samples were obtained from 15 patients who underwent an elective hip replacement surgery ($n = 15$, mean age \pm standard deviation = 57 ± 13 years, sex-balanced). All patients signed a specific consent for the reuse of biological samples and clinical data in the context of our study. Trephine biopsies were collected, routinely processed, and H&E-stained between 2017 and 2021 but scanned synchronously in 2020 and 2022, reflecting a gradient in H&E contrast associated with the sample age.

MarrowQuant 2.0 Workflow

The workflow is summarized in Figure 2 and Supplementary Figure S1 and detailed further. MarrowQuant 2.0 was implemented as a script for the freely available and open-source QuPath software.³⁹ The code is accessible on GitHub (see Code and Data Availability section).

Image Acquisition and Preprocessing

For the training, test, and experimental set specimens, the H&E-stained slides were scanned at the Institute of Pathology at CHUV using a NanoZoomer S60 40 \times objective to generate .ndpi files, with the exception of the H&E-stained slides from the control specimens, which were scanned with an Olympus VS120 slide scanner using a UPLSAPO 20 \times /0.75 objective to generate .vsi files. Both file formats were loaded into a project in QuPath0.3.2 using the BioFormats extension for annotation and MarrowQuant 2.0 for quantification. Once the images were inserted within a QuPath project, the user performed a white color balance selection and 3 annotations (Supplementary Video S1). In brief, the user first annotates the tissue boundaries, which correspond to the regions of interest (ROIs) to be quantified by MarrowQuant 2.0. Then, the user selects both a background reference (white space) and the artifact regions to be excluded from the quantification (ie, imprints of detached bone pieces, retraction artifacts, traces of large blood vessels, or highly hemorrhagic regions). This step took, on average, 2-3 minutes per trained user.

The annotations of the training and experimental sets were performed independently by one expert in hematopathology (L.D.L.) and by 2 individuals without any background in BM histopathology (R.S., S.B.). A first reminder was set to pop as a warning window "If highly heterogenous marrow, seek expert opinion to select ROI." The annotations of the test set were performed by a senior hemopathology resident (C.R.C.) who also participated in the cellularity assessment of the same samples during the diagnostic pipeline at the Institute of Pathology at CHUV.

MarrowQuant 2.0 for Human Tissue

The MarrowQuant workflow, implemented within QuPath software, was first developed on murine bone H&E-stained images and adapted in this study to human BM quantification. In brief, MarrowQuant 2.0 works by segmenting regions based on color and texture compared with those of background.³⁸ It predicts 5 mutually exclusive compartments as output areas for each H&E-stained image in the following order: bone (mint blue mask), hematopoietic cells (dark purple mask), interstitial and microvasculature (pink mask), the later detecting red blood cells and other small eosinophilic structures such as microvasculature, and adipocyte ghosts (yellow mask) (Fig. 2 and Supplementary Fig. S1). Marrow areas not recognized as any of the

abovementioned compartments are categorized as the "other" mask. If the percentage of other compartment is $>25\%$, an error message is displayed to alert the user to manually check the image or seek an expert opinion. This threshold was set based on extensive quantifications of cases with stromal expansion, a component classified as other (Supplementary Fig. S2E and Fig. 3G-H). To calculate the relative contribution of each of the marrow compartments, the "total marrow area (Ma.Ar)," which includes 4 marrow compartments and excludes the bone mask and artifacts, was used as the denominator (Ma.Ar = hematopoietic area + adipocytic area + IMV area + other area). For the percentage of bone compartment, the denominator is defined as the full tissue boundary selected by the user after artifact exclusion.

Bone Marrow Compartment Detection for Cellularity Calculation

For cellularity measurements, 2 separate calculations were performed and embedded within the output of MarrowQuant 2.0 to consider the 2 alternative denominators described in the literature: either the sum of the areas of the hematopoietic and adipocytic compartments used as denominator (Eq. 1: cellularity assessment score or simply cellularity) or the Ma.Ar used as the denominator (Eq. 2: hematopoietic area ratio or simply %hematopoietic area). Equation 1 reflects the clinical, semi-quantitative estimation that constitutes the most prevalent working definition in hematopathology,^{17,24,25,42} which assumes reciprocity of the hematopoietic and adipocytic compartments. Equation 2 reflects the bone morphometry definition of %hematopoietic marrow⁴³ integrating Ma.Ar as the spaces of the skeleton defined by endosteal surfaces.^{22,23,44} Adiposity is defined as the percentage of the Ma.Ar occupied by adipocytes (%Tt.Ad.Ar). Nomenclature and abbreviations for Eqs. 1 and 2 and figures follow recommendations of the International Bone Marrow Adiposity Society²³:

Cellularity assessment score (Cellularity)

$$= \frac{\text{hematopoietic area}}{(\text{hematopoietic} + \text{adipocytic}) \text{ area}} \% = \frac{\text{Hm.Ar}}{\text{Hm.Ar} + \text{Tt.Ad.Ar}} \quad (1)$$

Hematopoietic area ratio (%Hematopoietic area)

$$= \frac{\text{hematopoietic area}}{\text{total marrow area}} \% = \frac{\text{Hm.Ar}}{\text{Ma.Ar}} \quad (2)$$

MarrowQuant Validation: Pathologist Cellularity Assessment

For the training and experimental sets, 4 independent hematopathologists (reviewers 1-4) performed a retrospective visual cellularity assessment based on the digitalized H&E-stained images as follows. Pathologists were blinded to the clinical data associated with the patient samples. Electronic booklets were generated for the pathologists to quickly score each BM specimen, which contained an overview of the whole biopsy (4 \times) along with two 20 \times images (0.22 $\mu\text{m}/\text{pixel}$) for each trephine biopsy ($n = 157 + 36 = 193$ H&E-stained images) (Supplementary Fig. S3; Fig. 2). For intraobserver variability, the same booklet, but with the image order shuffled and relabeled, was sent to the same pathologists to be scored after a washout period of 8 weeks.⁴⁵ For the test set, the senior hematopathology resident (C.R.C.) first visually performed cellularity assessment in the digitalized image. Then, she annotated the image on QuPath and autonomously ran

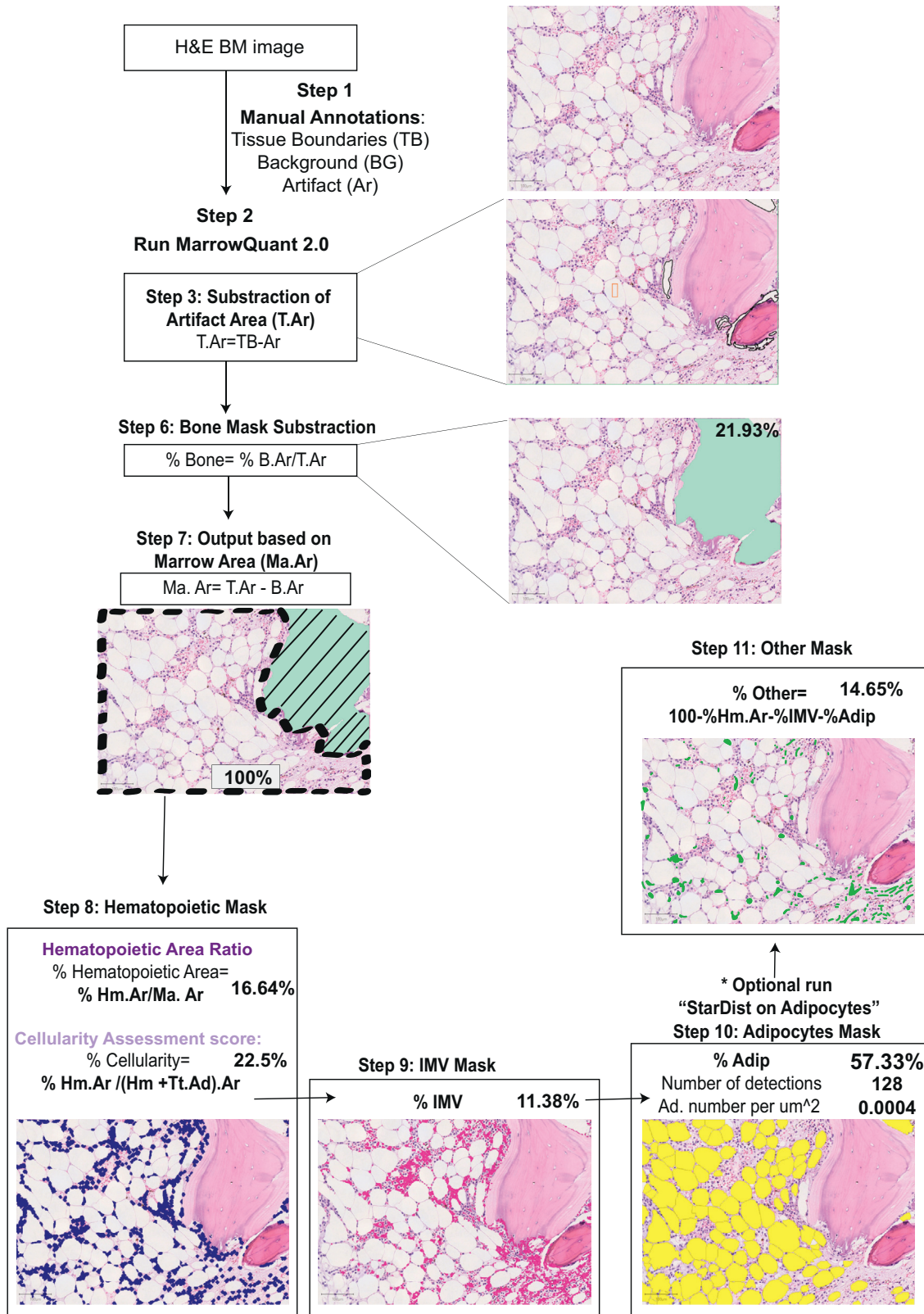


Figure 2.

MarrowQuant 2.0 mask classification and compartment quantification logic. MarrowQuant 2.0 user-defined parameters set at the recommended values for human BM analysis (minimum adipocyte size: 300 μm², maximum adipocyte size: 2,500 μm², minimum circularity: 0.3) (see Supplementary Methods). Percentages are indicated for quantification of the example hematoxylin and eosin image shown. Color code: "bone" mask in mint blue, "hematopoietic cell" mask in dark purple, "IMV" mask in pink, "adipocytes" mask in yellow, and "other" mask (overlayed in green for illustration purposes). If desired, batch run "Stardist for adipocytes" to obtain the adipocyte size distribution and color by size segmentation. B.Ar, bone area; BM, bone marrow; BG, background; Hm.Ar, hematopoietic area; IMV, interstitial and microvasculature; Ma.Ar, marrow area; T.Ar, tissue area; TB, tissue boundary; Tt.Ad.Ar, total adipocyte area.

MarrowQuant 2.0 for quantification. For both validation and test sets, the percent cellularity estimation was extracted from the pathology report (consensus of 2-3 pathologists) for comparison.

From Adipocyte Area Mask to Individual Adipocyte Detection: StarDist for Adipocytes

Segmentation of individual adipocyte ghosts with the MarrowQuant 2.0 workflow alone was suboptimal to derive a precise human adipocyte size distribution. For precise individual adipocyte detection within QuPath, we added an additional step to the workflow after training StarDist to recognize adipocytes. For this purpose, 12 images covering a wide range of cellularity (<5%-95%) were selected from the training and experimental sets. StarDist training was performed on the manually annotated RGB images after extraction from QuPath. Two additional images were used for validation. Default StarDist parameters^{46,47} were used to create the model (32 rays, grid factor of 2×2 , learning rate = 0.0005). Images were augmented by applying flips and rotations, adding Gaussian noise and independent random intensity changes to the R, G, and B channels. The training was performed for 400 epochs on 256×256 -pixel patches with a batch size of 4 (100 steps per epoch). The Jupyter notebook is available on GitHub (see Code and Data Availability section). After threshold optimization built in StarDist, the model was evaluated using the validation images. Then, StarDist for adipocytes was integrated within QuPath as described in [Supplementary Fig. S5](#). An additional code was created to classify the size of each adipocyte and report size distribution for each image: very small ($300\text{--}499 \mu\text{m}^2$), small ($500\text{--}899 \mu\text{m}^2$), medium ($900\text{--}1999 \mu\text{m}^2$), large ($2000\text{--}3499 \mu\text{m}^2$), and very large ($>3500 \mu\text{m}^2$). The performance of StarDist for adipocytes was assessed at different levels during the training process ([Fig. 4G-H](#) and [Supplementary Fig. S4](#)).

Tissues were fixed for 2 hours at room temperature in 4% paraformaldehyde and paraffin embedded. Dewaxed and rehydrated paraffin sections were pretreated in 10 mM Na citrate (pH 6.0) at 95°C for 20 minutes. Blocking was performed with 1% BSA for 60 minutes. Primary antibodies used were as follows: Perilipin-1 (rabbit α -Perilipin-1, clone EPR3753(2), Abcam; dilution 1/200) and CD34 Class II epitope (mouse α -CD34 Class II, clone QBend10, DAKO; dilution 1/50). For immunohistochemistry, chromogenic revelation was performed with ChromoMap DAB kit (Roche Diagnostics), and sections were counterstained with Mayer hematoxylin.

Statistical Analysis

All values quantified by MarrowQuant 2.0 are displayed as mean \pm standard deviation. GraphPad Prism (version 9.3.0; GraphPad La Jolla) was used for Student *t* test or 2-way ANOVA, linear regression (Pearson correlation), and multiple comparison tests were used for [Figures 3-5](#) and [Supplementary Figures S4-S8](#). Statistical definition of outliers followed the Bonferroni outlier test after verification that the data set is normally distributed using D'Agostino-Pearson normality tests. A significance level of .05 was set for all statistical tests. RStudio software (version 1.2.5001; 2013; R Core Team), with the following libraries: caret, irr, pROC, and HandTill2001 was used for agreement analyses (intraclass correlation [ICC], specificity, sensitivity, area under the receiver operating characteristic [ROC] curve [AUC]). To build multiple ROC curves, we used both multiclass macroaveraged AUC function and the AUC metrics computed using the Hand-Till method⁴⁸ ([Table 2](#) and [Supplementary Fig. S6](#)). The

intraobserver variability was calculated with the ICC using an F test. In addition, the interobserver variability was calculated with the ICC using a 2-way random effect model with an absolute agreement. As recommended, ICC values <0.5 were categorized as poor, 0.5-0.75 as moderate, 0.75-0.90 as good, and >0.9 as excellent.⁴⁹ To assess the cellularity scoring of MarrowQuant 2.0, specificity, sensitivity, and AUC tests were performed compared with the clinical references, defined as the score extracted from the pathology report, where sensitivity measures the true positive rate and specificity measures the true negative rate. Cellularity scores were categorized into 3 ranges: low cellularity (scores <25%), as defined for severe aplastic anemia and other BM insufficiency syndromes^{16,50,51}; medium cellularity (25%-50%); and high cellularity (>50%), as applied in several studies for age-adjusted cellularity assessment or myeloid malignancy evaluation.^{24,28,32} Our cellularity measurements (whether from the pathologist scoring or MarrowQuant 2.0 output) are provided as absolute values with no age adjustment.

Code and Data Availability

The codes for MarrowQuant2.0 and StarDist for adipocytes are available on GitHub and Zenodo, along with a detailed tutorial to install the 2 workflows within QuPath 0.3.2: <https://github.com/Naveiras-Lab/MarrowQuant2.0/blob/main/README.md>; <https://zenodo.org/badge/latestdoi/525402383>

Training set H&E-stained images (32 images of 36) are accessible in the Image Data Resource database (Study idr0096, <https://idr.openmicroscopy.org/webclient/?show=project-2104> and <https://idr.openmicroscopy.org/webclient/?show=project-2103> in the experiment B section). Images from the experimental and test data sets are available on request. See [Supplementary Methods](#) for sample processing, histology, immunohistochemistry, code parameter tuning, error quantification, and confusion matrix.

Results

Adaptation and Validation of MarrowQuant 2.0 for Human Bone Marrow Specimens (Training Set)

Adapting our earlier work on MarrowQuant³⁸ for the analysis of human BM specimens ([Fig. 2](#)) required most preset parameters and classification thresholds to be modified and integrated within the code to consider the larger human hematopoietic and adipocyte cell size, larger average cell separation, and higher content of interspersed red blood cells. A training set of H&E-stained images spreading across the full range of BM cellularity was identified ($n = 36$), and an error quantification analysis for BM compartment prediction was performed as detailed in the Error Quantification and Confusion Matrix Sections of the Supplementary Materials and [Supplementary Figures S2](#) and [S3](#). In brief, the minimum adipocyte size was validated through perilipin and CD34 immunohistochemistry in contiguous sections to differentiate small adipocyte ghosts from small vascular structures ([Supplementary Fig. S2A-D](#)). Then, thresholds were adapted to minimize the detection of false adipocytes, adipocyte nuclei, or megakaryocyte misclassification. This includes images from extreme BM remodeling, such that manual correction did not provide significant improvement compared with that of MarrowQuant 2.0 alone ([Supplementary Fig. S3A-C](#)).

For validation, cellularity assessment (Eq. 1) constitutes the only parameter with a clinical reference to which we can compare

the performance of our algorithm. In clinical routine, cellularity is visually estimated in a semiquantitative fashion by an expert hematopathologist (reviewed in a previous study⁸). Cellularity assessment by MarrowQuant 2.0 (Eq. 1) strongly correlated with the mean cellularity assessment of the 4 independent pathologists for the training set (Supplementary Fig. S4A) ($n = 36$, $R^2 = 0.94$). In particular, MarrowQuant 2.0's cellularity values fell within the interobserver variability range, with maximal correlation at medium and high cellularity values. More specifically, the ICC among the 4 reviewer pathologists was 0.96 (95% confidence interval [CI] = 0.94-0.98) and remained unchanged at 0.96 (95% CI: 0.92-0.98) when MarrowQuant 2.0's estimation was incorporated as a fifth reviewer. The mean difference between the scores from MarrowQuant 2.0 and the average of all reviewers was $-0.55\% \pm 7.6\%$. Thus, MarrowQuant 2.0's cellularity assessment performed equivalently to an additional pathologist, indicating noninferiority compared with the clinical reference for the training set.

To compare the overall classification of MarrowQuant 2.0 with that of manual mask classification as measured by the percent true positive pixel classification (visual estimation by the user), we generated a confusion matrix. The average true mask classification rate was 0.86 (Supplementary Fig. S4B). The classification rate of the bone, hematopoietic, and adipocytic classification masks was excellent (0.99, 0.92, and 0.98, respectively). As for the IMV and "other" classification masks, the classification rates were 0.69 and 0.72, respectively. When these 2 classification masks were misclassified, they were confused for one another but not classified as either of the remaining 3 classification masks. The misclassification seemed to arise from expanded stromal or highly hemorrhagic regions within the BM and, thus, did not affect overall cellularity assessment (Eq. 1). In all, for the training set, MarrowQuant 2.0's precision was validated through the classification rate of both average masks across a wide cellularity range and through cellularity assessment compared with that of the clinical reference.

Testing Utility for Routine Clinical Diagnostic Samples and Identifying the Main Source of Outliers (Test Set)

Next, we tested whether MarrowQuant 2.0 could be compatible with the heterogeneity of samples encountered in clinical routine and user-friendly enough to be used in this setting. Over the course of 2 weeks, all hematopathologic cases that were received for routine diagnostic evaluation at the University Hospital were collected and the diagnostic pathology reports generated, whereas all H&E BM-stained BM slides were scanned as a batch to generate digital images. Once digitalized, one of the participating pathologists blindly re-scored all images using the full digital image (pathologist digital score), annotated, and ran MarrowQuant 2.0 for each image.

Initially, we found a strong correlation between the cellularity assessment given by the pathologist on the digital image and MarrowQuant 2.0 ($R^2 = 0.89$, $n = 42$, mean difference $+0.85 \pm 9.9$) but then a correlation of only 0.7 ($n = 42$, mean difference $+1.13 \pm 6.6$) for MarrowQuant 2.0 when compared with the cellularity assessment from the pathology report (Fig. 3A). To understand the discrepancy between the pathology report evaluation and the digital image evaluation when compared with the MarrowQuant 2.0 output (Fig. 3A), we searched for outliers. These belonged to 2 distinct categories: 3 heterogeneous marrow tissues (Fig. 3B-E) and 1 case of myelofibrosis (Fig. 3F-G). The discrepancy in the heterogeneous marrow subgroup was due to the pathologists' selection of specific ROIs during diagnosis (Fig. 3E) as opposed to

the full digital image annotation on MarrowQuant 2.0 scoring (Fig. 3B-D). In fact, during diagnosis, expert hematopathologists excluded purely adipocytic BM areas corresponding to subcortical marrow spaces, which are frequently hypocellular and may be excluded in the assessment of cellularity as per International Council for Standardization in Haematology guidelines⁵² (Fig. 3E). The remaining outlier was a case of myelofibrosis, where MarrowQuant 2.0 overestimated the cellularity because of stromal expansion (MarrowQuant 2.0 score = 60.9%, pathology report = 25%, and pathologist digital image score = 40%) (Fig. 3F-G). Indeed, pathologists confirmed that the cellularity score for this outlier was estimated considering the fibrotic, expanded stromal compartment as part of the denominator, instead of computing exclusively the sum of the adipocytic and hematopoietic areas as denominator. This adapted score corresponds to the alternative cellularity definition (hematopoietic area ratio or %hematopoietic area; Eq. 2), also accepted in hematopathology,^{8,10,19,24,25} which uses the Ma.Ar as a denominator. MarrowQuant 2.0 quantification of this myelofibrosis case using %hematopoietic area (Eq. 2) was 30.1%, falling within the range of intervariability between the pathology report (25%) and the pathologists' estimation of the re-scored digital image (40%) (Fig. 3G), instead of the 60.9% MarrowQuant 2.0 score predicted when using Eq. 1. Thus, we went back to the outlier cases and re-annotated the heterogeneous BM digital images by excluding empty marrow spaces and found coherent scores and a higher correlation of MarrowQuant 2.0 with the pathologist's cellularity assessment ($R^2 = 0.86$ for the pathology report and $R^2 = 0.92$ for the re-scored digital image, $n = 42$) (Fig. 3H).

To diversify our cohort and verify the validity of the hematopoietic mask detection workflow in the context of lymphoid aggregates, we selected a set of H&E-stained images from anonymous patients with lymphoid disorders which associated to with increased cellularity. We then tested the performance of MarrowQuant 2.0 on such cases, and found it to correctly assign lymphoid and lymphoblastoid cells to the hematopoietic mask (Supplementary Fig. S6A-I).

Overall, we could validate the good to excellent performance of MarrowQuant 2.0 compared with the clinical reference, within the user-friendly interface of QuPath, to assess BM cellularity in routine diagnostic samples regardless of the underlying condition assessed. The variety of cases emphasized the importance of expert ROI selection and the challenges associated with the choice of the denominator for BM cellularity assessment in cases of expanded stroma (Eq. 1 vs Eq. 2). MarrowQuant 2.0 systematic reporting of both equations is compatible with future prospective analysis to determine the predictive value of either equation in diagnostic hematopathology.

MarrowQuant 2.0 in an Extreme Bone Marrow Remodeling Context (Experimental Set)

The denominator-dependent discrepancy identified for the cellularity assessment of samples with stromal expansion prompted us to experimentally test the assumption of reciprocity for the hematopoietic and adipocytic marrow compartments, which is implied within the working definition of cellularity (Eq. 1). To investigate the performance of MarrowQuant 2.0 in a full range of cellularities and test the assumption of reciprocity, we focused on patients receiving myeloablative chemotherapy, one of the most extreme cases of BM remodeling and recovery.

We tested MarrowQuant 2.0 on H&E-stained BM images from patients with AML and MDS ($n = 28$) at diagnosis and at 3 time

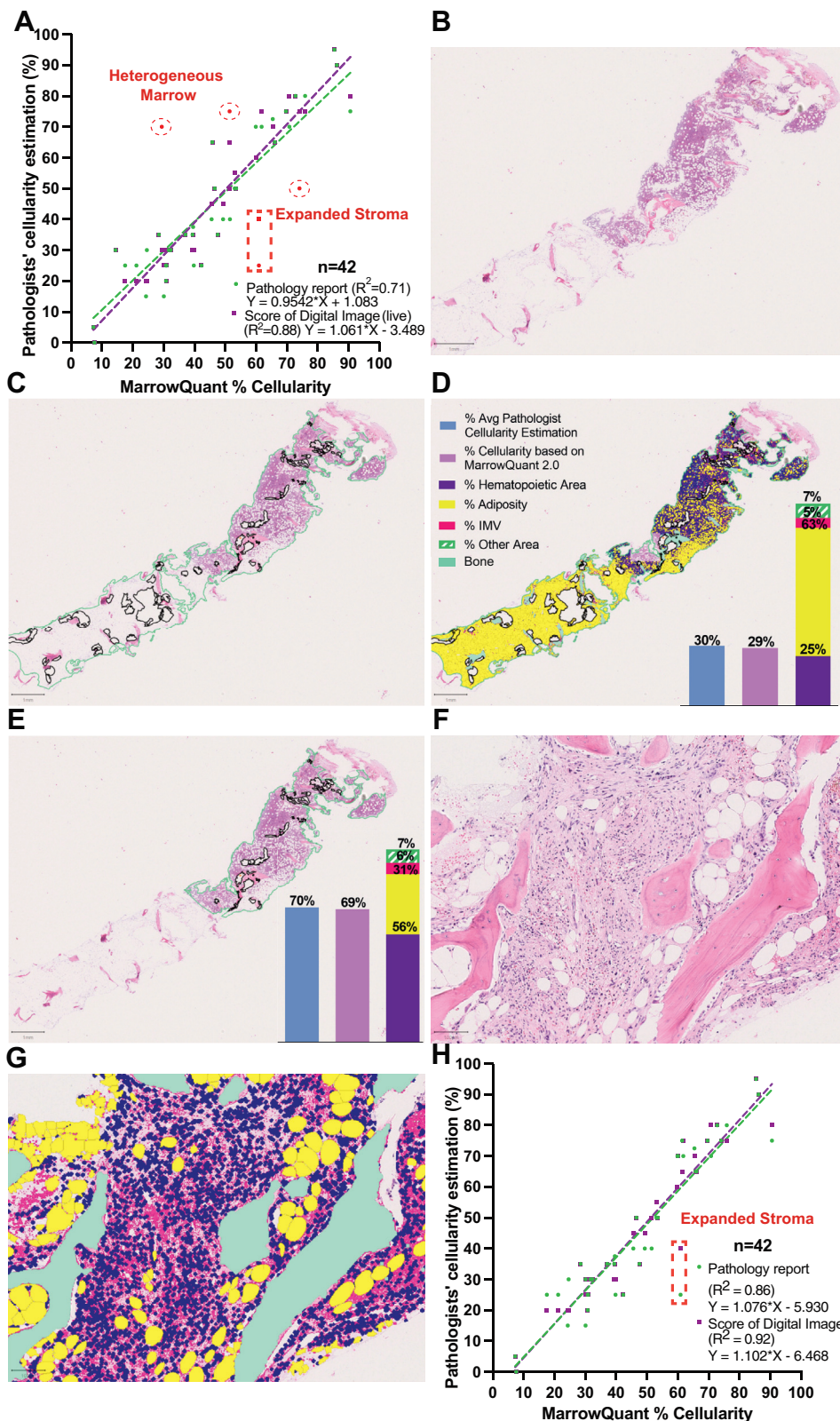


Figure 3.

Use of MarrowQuant 2.0 in a clinical diagnostic setting (test set). Digital hematoxylin and eosin (H&E) slides of 42 successive cases of diagnostic BM trephines biopsies were reviewed by a senior hematopathology resident who double scored cellularity by visual evaluation and using MarrowQuant 2.0. (A) Correlation of %cellularity generated by MarrowQuant with the clinical reference (%cellularity scoring extracted from the pathology report, green dots) or estimated by the reviewing pathologist by visual examination of the digital image (purple dots) before MarrowQuant 2.0 assessment (H&E biopsies: $n = 42$ diagnosis cases [test set], R^2 [pathology report] = 0.7, R^2 [score of digital images (live)] = 0.89). (B-E) Most outlier cases in (A) ($n = 4$ outliers) are explained by heterogeneous marrow ($n = 3/4$), of which an example is presented (B-E): (B) H&E image, quantified by MarrowQuant 2.0 with either of 2 annotations considering either the full marrow full digital tissue (C-D) or the pathologist-selected ROI (E). (B) Before annotating

points after intensive myeloablative chemotherapy, corresponding to the peak of aplasia of the first chemotherapy cycle (A, days 17–21), the hematopoietic recovery after the first cycle of chemotherapy (RC1, days 35–43), and the hematopoietic recovery after the second cycle of chemotherapy (RC2), in addition to an age-matched control group from hip replacement surgical bone specimens ($n = 15$) (Fig. 4A). For all samples, MarrowQuant 2.0 calculated both %cellularity (Eq. 1) and %hematopoietic area (Eq. 2). The later equation is reported in this section to describe the dataset because it does not assume reciprocity and allows for the longitudinal assessment of the whole BM space in stacked plots. As expected, we found significant predominance of the hematopoietic compartment at the diagnosis time point, when samples were infiltrated by the hematopoietic malignancy ($72.4\% \pm 31\%$ for %cellularity and $60.1\% \pm 22.8\%$ for %hematopoietic area) (Fig. 4B–D). Accordingly, the remaining BM compartments measured by MarrowQuant 2.0 were significantly smaller in the diagnosis time point samples than those of the age-matched control group: %adiposity ($18.3\% \pm 23\%$ vs $55.4\% \pm 4\%$), adipocyte density (184 ± 124 adipocytes/ mm^2 vs 425 ± 184 adipocytes/ mm^2), %IMV ($6.6\% \pm 2.5\%$ vs $11.7\% \pm 5.4\%$), and %Other ($18.4\% \pm 5.7\%$ vs $14.3\% \pm 3.4\%$) (Fig. 4B, E, F and Supplementary Fig. S5A). As expected, the reverse scenario was observed at the peak of aplasia time point; both %cellularity ($12.2\% \pm 10.2\%$) and %hematopoietic area ($13.2\% \pm 11.5\%$) (Fig. 4B–D) decreased significantly, whereas %adiposity ($55.8\% \pm 22.6\%$), adipocyte density (504 ± 210 adipocytes/ mm^2), %IMV ($12.2\% \pm 4.9\%$) and %Other ($25.6\% \pm 12.3\%$) compartments increased significantly when compared with those of the diagnostic samples. Surprisingly, no significant differences were detected between the peak of aplasia and the hip replacement control samples, likely reflecting the reported higher adipose content of the femur than the iliac bone in human adults despite their proximity.⁵³ Compared with the aplasia time point, the hematopoietic compartment in the recovery time points significantly increased in RC1 but not in RC2, whereas the %adiposity significantly increased for RC2 but not for RC1 (RC1—%cellularity, $59\% \pm 18.2\%$; %hematopoietic area, $42.7\% \pm 25.7\%$; %adiposity, $35.1\% \pm 17.3\%$; adipocyte density, 346 ± 125 adipocytes/ mm^2 ; RC2—%cellularity, $48\% \pm 20.2\%$; %hematopoietic area, $35.9\% \pm 16.8\%$; %adiposity, $38.9\% \pm 19.5\%$; adipocyte density, 328 ± 127 adipocytes/ mm^2). Overall, the measured average changes in this longitudinal data set were consistent with hematopoietic progenitor depletion at the peak of aplasia and relative hematopoietic progenitor exhaustion on repeated chemotherapy cycles in patients with AML/high-risk MDS. This extensive remodeling scenario enabled us to validate MarrowQuant 2.0 on H&E-stained images with a wide cellularity.

Next, we measured the agreement of MarrowQuant 2.0 with the clinical reference of visual cellularity assessment. We first assessed the intraobserver variability among the 4 independent reviewers after scoring the same set of training and validation images at $t = 0$ and $t = 8$ weeks (washout period), and we observed an excellent agreement among the 4 independent pathologists, with the ICC ranging from 0.95 to 0.99 (Table 1). Then,

we tested the agreement between MarrowQuant 2.0 cellularity assessment and all individual reviewer estimations and found an excellent agreement (ICC = 0.96; 95% CI: 0.93–0.97) (Table 1). Furthermore, compared with all reviewers' average score, the ICC was 0.978 (95% CI: 0.955–0.989). Thus, we conclude that MarrowQuant 2.0 used by an expert performed at least equivalently in cellularity measurements as the visual estimation by a pathologist and that variability between individual pathologist reviewers did not significantly affect our evaluation.

Furthermore, we tested whether MarrowQuant 2.0 performed equally across different cellularity ranges when annotations were performed by an expert (supervised). We performed specificity and sensitivity analyses using the cellularity assessment reported in the pathology report as the ground truth (Table 2). Specificity (true negative rate) ranged between 0.89 and 1.0 and was highest for the 2 extremes: low cellularity and high cellularity. Sensitivity (true positive rate) ranged between 0.81 and 0.99. The lowest sensitivity scores were for low cellularity specimens, which included the cases with stromal expansion (Supplementary Fig. S7). Overall, we could thus validate MarrowQuant 2.0's cellularity assessment as good to excellent when compared with the clinical reference in a scenario of extreme BM remodeling (experimental set) and identified low cellularity cases with stromal expansion as the most problematic for high sensitivity.

Quantification of Adipocyte Size Distribution During Bone Marrow Remodeling

In addition to quantifying the area of different BM compartments using MarrowQuant 2.0, we were interested in quantifying BM remodeling by tracking adipocyte size distribution. One limitation of MarrowQuant is that adipocyte segmentation is based in the Fiji/ImageJ analyze particle function,^{54,55} which calculates the total adipocyte mask but cannot accurately detect individual adipocyte size. To accurately segment individual adipocytes, we developed an adaptation of the StarDist extension of QuPath (<https://github.com/qupath/qupath-extension-stardist>), which segments and measures nuclei size as oval objects fitted within polygons. We applied deep learning to train the model to recognize BM adipocytes instead of nuclei. Our measurements of accuracy and precision validated the use of StarDist for adipocyte size detection (Fig. 4G and Supplementary Fig. S5C, D). Significant differences were detected for the very small ($300\text{--}499 \mu\text{m}^2$) and medium ($900\text{--}1999 \mu\text{m}^2$) adipocyte categories, but not for the small ($500\text{--}899 \mu\text{m}^2$), large ($2000\text{--}3499 \mu\text{m}^2$), or very large ($>3500 \mu\text{m}^2$) categories (Fig. 4H). In particular, we observed a significant increase in the very small adipocytes at diagnosis ($48.5\% \pm 22.2\%$). At RC2, the adipocyte size distribution became comparable with that of the control group (RC2: $22.6\% \pm 9.9\%$; control: $19.5\% \pm 4.1\%$). For the medium adipocytes, a significant decrease was seen at diagnosis ($15.2\% \pm 10.8\%$); then, the percentage was restored at RC2 when compared with the control group (RC2: $30.3\% \pm 6.6\%$; control: $31.8\% \pm 3\%$). In conclusion, we validated

the image, a visual cellularity score of 30% was given for the full digital image and a 70% cellularity score was reported in the consensus pathology report. (D) MarrowQuant 2.0 %cellularity (Eq. 1) was 29.56% when the full trephine was selected as ROI. (E) MarrowQuant 2.0 cellularity (Eq. 1) was 69% when adapting ROI to the pathologist selection. (F–G) BM fibrosis as quantified by MarrowQuant 2.0 (expanded stroma). (F) H&E trephine biopsy of myelofibrosis case (remaining outlier $n = 1/4$ outliers in this test data set) and (G) corresponding MarrowQuant 2.0 BM compartment assignment. (G) Cellularity score reported by the pathology report = 25%, by the senior hematopathology resident on the full digital image = 40%, by MarrowQuant 2.0 (Eq. 1) = 60.9%, %hematopoietic area reported by MarrowQuant 2.0 (Eq. 2) = 30.1%. (H) Correlation of %cellularity generated by MarrowQuant 2.0 (Eq. 1) with the clinical reference %cellularity score estimation as in the pathology report represented in (A) after correction of heterogeneous marrow, outliers (red dots) by adoption of ROI selection to the pathologist-selected ROI. (H&E biopsies $n = 42$ diagnosis cases [test set], adapted R^2 [pathology report] = 0.86, adapted R^2 [score of digital images (live)] = 0.92).

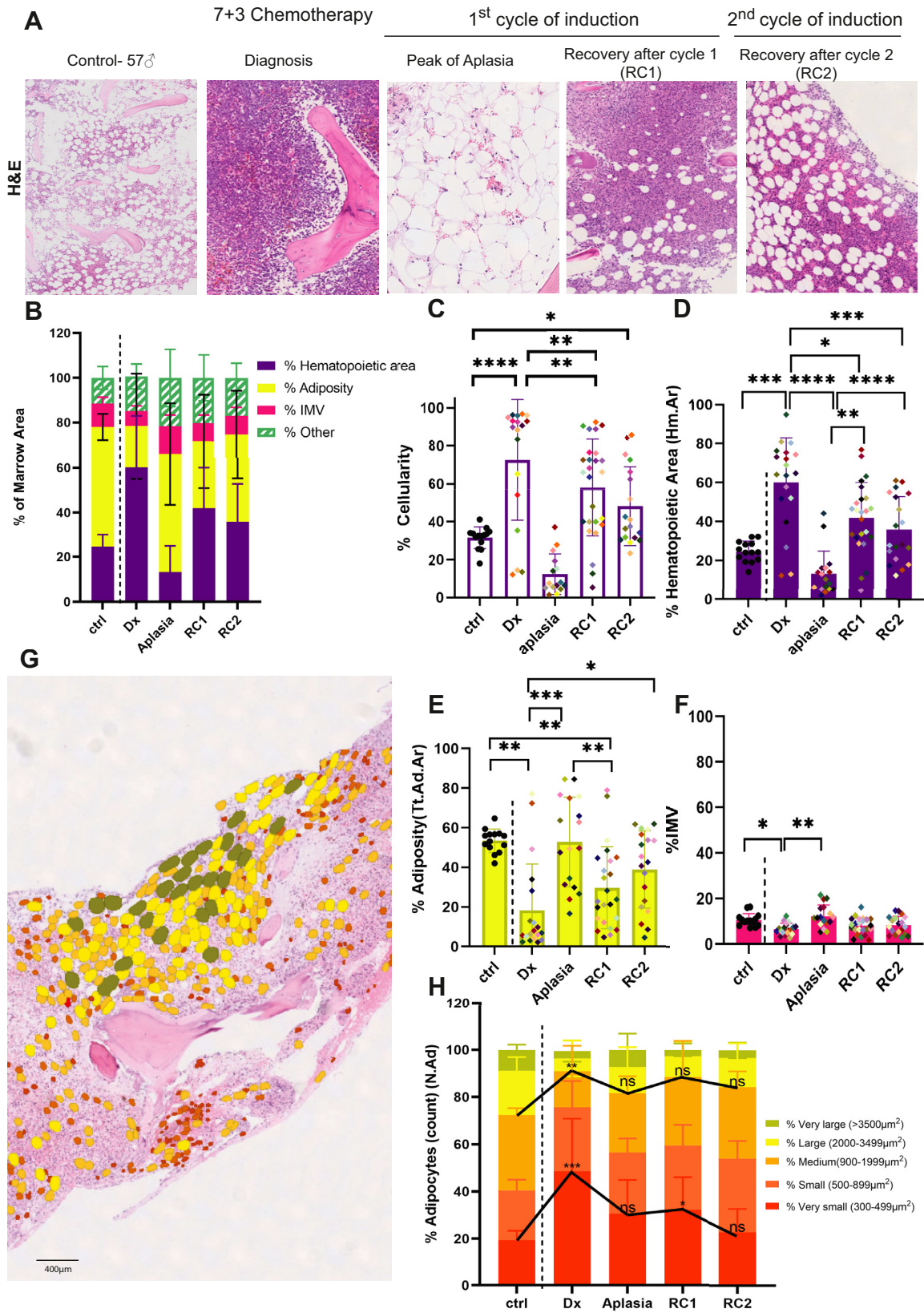


Figure 4.

MarrowQuant 2.0 quantification of hematoxylin and eosin (H&E) images from the experimental set (validation set) (BM remodeling). (A) Representative images of H&E images collected from trephine biopsies of patients with AML/MDS who underwent chemotherapy (55 ± 11 years, sex-balanced) (n = 126 longitudinal specimens from 28 patients) and from bone fragments of age-matched and sex-matched control patients who underwent hip replacement surgery (57 ± 13 years, sex-balanced) (n = 15 images from 15 patients). BM H&E sections from 4 time points were collected (diagnosis [Dx], peak of aplasia, hematological recovery post first induction cycle [RC1], and hematological recovery post

the use of StarDist for the measurement and allocation of BM adipocyte size distribution. We found both the total area and certain size ranges of BM adipocytes to be significantly and longitudinally remodeled in patients with AML/high-risk MDS after intensive chemotherapy.

Reciprocity: Assumption in the Working Definition of Cellularity

The lower sensitivity of MarrowQuant 2.0 for low cellularity samples prompted us to evaluate the limits of reciprocity between the hematopoietic and adipocytic BM compartments. We plotted the adiposity (Tt.Ad.Ar %) versus hematopoietic area (Hm.Ar%) for all images within the experimental set (Fig. 5A and Supplementary Fig. S8A). The hematopoietic and adipocytic areas presented a negative correlation for the 2 data sets ($R^2 = 0.74$, $n = 79$ for experimental set), thus validating a general trend for reciprocity. However, cases with stromal expansion, which could be identified by MarrowQuant 2.0 through a threshold of “other” compartment higher than 25% of the marrow area, clustered outside of the regression line (red dots) and corresponded univocally to the discordant outliers on the %cellularity assessment by the pathologist versus MarrowQuant 2.0 scoring. This is illustrated in Figure 5B-E by the fact that a case of stromal edema post-intensive chemotherapy was scored by the expert pathologists as 5% cellularity (blue column), versus 34% cellularity by MarrowQuant 2.0 (light purple column) when using Eq. 1. The %hematopoietic area (Eq. 2) calculated by MarrowQuant 2.0 was 10% (dark purple column), which was closer to the pathologist’s estimate. The overall concordance between %cellularity estimated by the pathologist (blue column) and %hematopoietic area estimated by MarrowQuant 2.0 (Eq. 2, dark purple column) is illustrated in Fig. 5D for the full experimental data set. All stromal expansion cases are outliers (red dots). The discordance between %cellularity assessed by the pathologist versus by MarrowQuant 2.0 (Eq. 1) was only statistically significant for the stromal expansion subset (Fig. 5E). In conclusion, we found %hematopoietic area (Eq. 2) to reflect more reliably the %cellularity estimated by the pathologist clinical reference assessment in low cellularity BM and stromal expansion cases, as opposed to the working definition (Eq. 1). For high cellularity, Eq. 1 best aligned with the pathologist’s assessment of %cellularity because the maximum values differ for Ma.Ar (Eq. 2 denominator) and Eq. 1. Indeed, %hematopoietic area as measured by Eq. 2 was capped on average at ~80% because the IMV and other compartments occupied, on average, a minimum of 20% of the BM space. Therefore, we tested the use of Eq. 2 for low cellularity samples and Eq. 1 for medium and high cellularity samples in the experimental and test data sets (Fig. 5F-G). Once all cases identified by MarrowQuant 2.0 as <25% cellularity (Eq. 1) were corrected to use Eq. 2 for cellularity assessment, all stromal expansion outlier cases (red dots) correlated with the clinical reference for both experimental (Fig. 5G) and test sets (Fig. 5G). In conclusion, and for best congruency with the clinical reference, we found best results using Eq. 1 for cellularity assessment of

medium to high cellularity samples and Eq. 2 for the assessment of low cellularity samples (Fig. 5H). Thus, the results of Eq. 1 and Eq. 2 are reported in the MarrowQuant 2.0 output, together with a “recommended cellularity value,” which computes Eq. 2 for both low cellularity (<25% hematopoietic area) and expanded stromal cases (> 25% other mask), and Eq. 1 for all other cases.

Robustness of MarrowQuant 2.0

Finally, following the validation of MarrowQuant 2.0 on 2 retrospective data sets and 1 prospectively collected data set, we tested its repeatability when used by experts (Fig. 5I) and its reproducibility when used by nonexperts (Supplementary Fig. S9). Four experts were asked to score the test data set on a fixed ROI predefined by an expert user, which comprised either the totality or a portion of the trephine biopsy judged as contributive for the assessment of BM cellularity. Repeatedly running the MarrowQuant 2.0 workflow on this ROI generated the same cellularity score regardless of how many times we ran the script (ICC = 1.00). However, some variability was detected when comparing the scoring across the 4 experts on the predefined ROI (ICC = 0.90; 95% CI: 0.84<ICC<0.94) (Fig. 5I). MarrowQuant 2.0 cellularity scoring in the predefined ROI correlated with each of the 4 pathologists ($R^2 = 0.81-0.92$) but correlated best with the cellularity estimation derived from the average of the 4 pathologists ($R^2 = 0.92$; ICC = 0.90; 95% CI 0.85<ICC<0.94) (insert in Fig. 5I). Overall, the repeatability of MarrowQuant 2.0 in a predefined ROI was superior to that of the individual pathologists.

To further assess the robustness of the tool, we tested the reproducibility of the results when 2 nonexpert users annotated (ROI selection) and performed MarrowQuant 2.0 analysis for the experimental set when compared with the outcome on annotation by a hematopathology expert. Given the challenges identified for heterogeneous marrow scoring (Fig. 3), nonexpert users were prompted to exclude samples with heterogeneous marrow and seek expert opinion. We found a strong correlation between MarrowQuant 2.0’s cellularity estimation by the 2 nonexpert users and the clinical reference, both when compared with the mean of the 4 independent reviewers and when compared with the score extracted from the pathology report (Supplementary Fig. S9A) ($R^2 = 0.95$). We compared the 2 methods of annotation (expert vs naïve user) and observed an excellent agreement (ICC = 0.98; 95% CI: 0.971-0.98) (Supplementary Table S2). However, we observed a higher correlation with both the pathology report and the 4 expert reviewers when MarrowQuant 2.0 was annotated by an expert compared with nonexpert users ($R^2 = 0.96$ for pathology report score, $R^2 = 0.98$ for average scoring of 4 reviewers) (Supplementary Fig. S9B and Supplementary Table S2). Correlation with the clinical reference was inferior when MarrowQuant 2.0 was annotated by nonexperts and when including the heterogeneous marrow samples ($R^2 = 0.96$) (Supplementary Fig. S9C). Overall, these results suggest that MarrowQuant 2.0 performs very well when annotated by nonexpert users, but it performs

second induction cycle [RC2]). (B-F) MarrowQuant2.0 output (C) %cellularity and (D) %hematopoietic area are highest at diagnosis and lowest at the peak of aplasia. A recovery is observed after the first and second cycles of induction (each patient is assigned 1 color for tracking purposes) ($n = 28$ patients with AML patients; $n = 15$ control patients; $*P < .05$, $**P < .01$, $***P < .0001$ by multiple comparison test). (E) The %adiposity is lowest at diagnosis, reflecting the infiltration by leukemic cells. The highest adiposity is reached at the peak of aplasia ($n = 28$ patients with AML; $n = 15$ control patients; $**P < .01$, by multiple comparison test). (F) The %IMV increases significantly at the peak of aplasia ($n = 28$ patients with AML; $n = 15$ control patients; $*P < .05$, $**P < .01$, $***P < .0001$ by multiple comparison test). (G) Adipocyte size distribution detection with an AI-based model trained on adipocytes, called StarDist for adipocytes, on H&E BM image after the recovery after the first cycle of induction of chemotherapy in an age + gender. (H) %Adipocyte (count) distribution by size category at 5 different time points as quantified by StarDist: very small: 300-499 μm^2 ; small: 500-899 μm^2 ; medium: 900-1999 μm^2 ; large: 2000-3499 μm^2 ; very large: above 3500 μm^2 ($n = 28$ patients with AML/high-risk MDS; $n = 15$ control patients; $*P < .05$, $**P < .01$, by multiple comparison test).

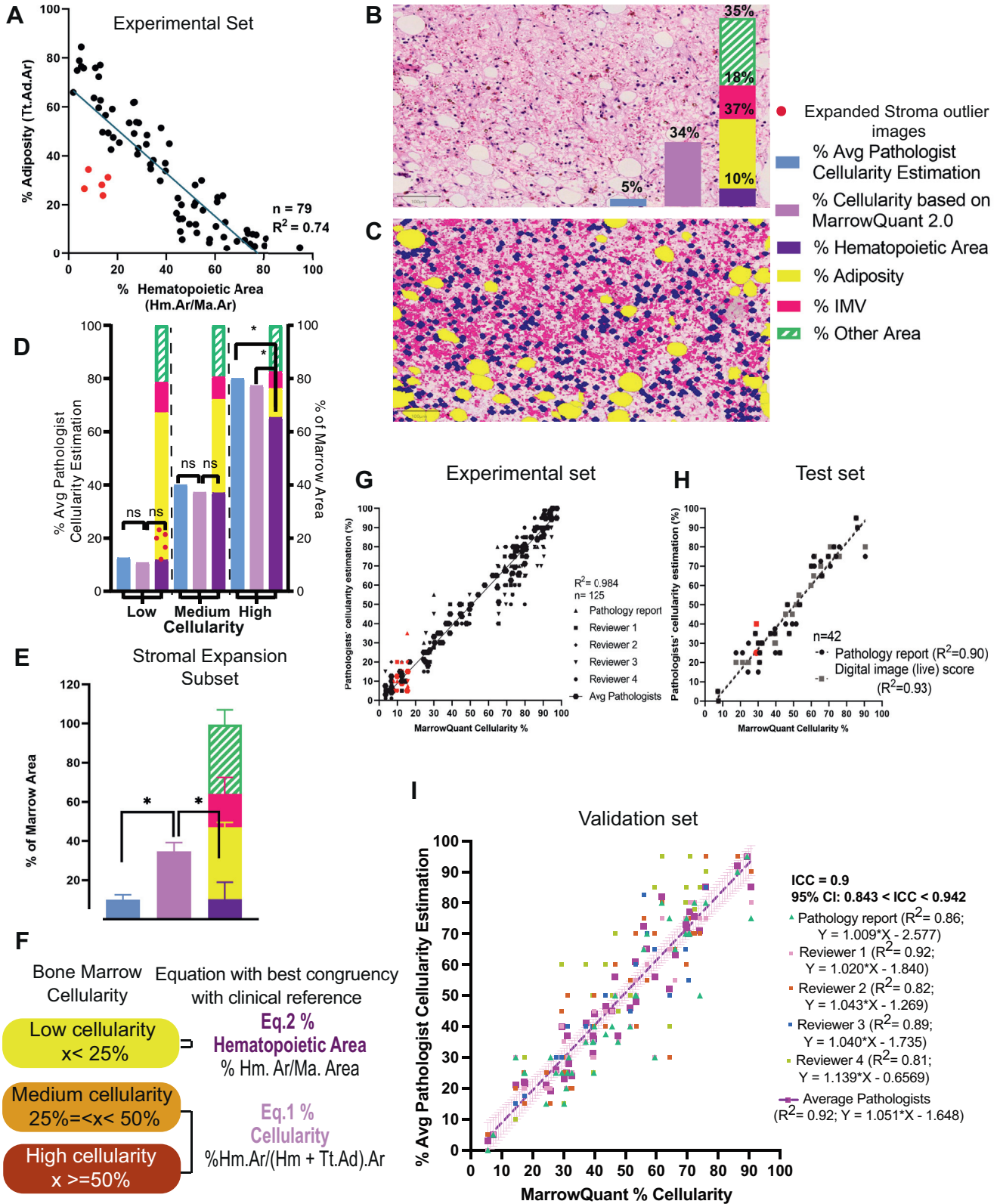


Figure 5. Reciprocity among the hematopoietic and adipocytic compartments within the BM quantified by MarrowQuant 2.0. (A) Reciprocity between the BM %hematopoietic area and % adiposity compartments quantified by MarrowQuant2.0 ($R^2 = 0.74$, $n = 79$, H&E images from experimental set), cases with expanded stroma are highlighted in red. (B) H&E BM biopsy of an expanded stroma case and (C) its corresponding MarrowQuant 2.0 classifications. (D) Comparison of equations: %hematopoietic area (Eq. 2) and %cellularity (Eq. 1) across the 3 ranges of cellularity when compared with the clinical reference ($n = 79$, experimental set). (E) Quantification of the cases with expanded stroma (red dots, $n = 5$, from experimental set). (F) Correlation of %cellularity (Eq. 1) generated by MarrowQuant 2.0 with the clinical reference %cellularity score estimation after correction of outliers (red dots) associated with expanded stroma using %hematopoietic area (Eq. 2) (H&E biopsies $n = 125$ [experimental set], R^2 [average pathologists] = 0.984). (G) Correlation of %

Table 1

Interclass and intraclass correlation coefficient for MarrowQuant 2.0 comparison with pathologist estimation of cellularity

	ICC (95% CI)
Intraobserver variability	
Reviewer 1	0.989 (0.98-0.994)
Reviewer 2	0.967 (0.94-0.981)
Reviewer 3	0.972 (0.95-0.981)
Reviewer 4	0.949 (0.91-0.972)
Interobserver variability	
Among reviewers	0.956 (0.934-0.972)
MQ2.0 expert annotation	
All individual reviewers	0.956 (0.935-0.973)
Average score of reviewers	0.978 (0.955-0.989)

ICC, intraclass correlation coefficient.

Intraobserver (top) and interobserver (bottom) correlation for cellularity estimation among the different reviewers after 1 round of MarrowQuant2.0 operated by expert (n = 172 hematoxylin and eosin-stained images).

better when used by an expert. For non-heterogeneous BM without stromal expansion, the correlation with the clinical reference is still excellent when used by nonexpert users (Supplementary Table S3).

Discussion

In this study, we present MarrowQuant 2.0, a new version of the MarrowQuant workflow³⁸ adapted and validated for BM compartment segmentation of H&E-stained human BM specimens. MarrowQuant 2.0 is implemented within QuPath 0.3.2, and together with a specifically trained StarDist model, it segments the BM images into 5 compartments (bone, hematopoietic, IMV, adipocytic, and other). Together, they quantify both the areas of each BM compartment and the individual BM adipocyte count and size distribution. This semiautomated workflow relies on the user's input to define the ROIs and artifacts. Automatic selection of ROI is not technically limiting for future versions of the workflow. However, it would require consensus in the hematopathology field for computer-logic compatible BM ROI selection/exclusion criteria in defined hematologic pathologies. We have seen that this is one of the most important sources of variability in MarrowQuant 2.0 outputs and have, thus, preferred to give the user control on ROI selection.

We counted on the %cellularity scoring to validate our tool because cellularity assessment is the only BM parameter measured by MarrowQuant that has an established clinical reference.^{10,19,32,56-58} The assessment of cellularity in the diagnostic pathology report or the average cellularity assessment of 4 pathologists (interobserver agreement: ICC = 0.96) was used as the ground truth. We found satisfactory performance when MarrowQuant 2.0 was used by a trained pathologist and once cases of stromal expansion had been excluded ($R^2 = 0.98$; ICC = 0.96-0.98; specificity, 0.89-1.0; sensitivity, 0.81-0.93; AUC = 0.98 for the experimental set, n = 172 BM cases). Cases of stromal expansion could be identified when setting an other mask threshold >25%. Cellularity assessment in these cases was satisfactory (fell within the interobserver variability) when the alternative %hematopoietic

area (Eq. 2) was used to determine cellularity, which uses Ma.Ar as the denominator instead of the sum of the adipocytic and hematopoietic compartments. Overall, MarrowQuant 2.0 outliers are not due to compartment misclassification but to the expansion of the stromal compartment, which is not considered in Eq. 1. Limitations in color contrast inherent to H&E stains prevented us from defining a separate compartment for the expanded stroma, which was thus classified as "other." Nevertheless, MarrowQuant 2.0 displayed a satisfactory average mask classification rate of 0.86 for all BM compartments in the confusion matrix for the training set, with an excellent accuracy for the assignment of bone (99%), hematopoietic (92%), and adipocyte (98%) areas.

Our goal for MarrowQuant 2.0 was to make it accessible for research applications and to test its compatibility with a clinical setting. For the research scale, MarrowQuant 2.0 was able to compartmentalize the BM space and provide cellularity measurements independently of a pathologist or expert hematopathology supervision, (unsupervised/nonexpert annotation performance— $R^2 = 0.95$; ICC = 0.93-0.95; specificity, 0.87-0.99; sensitivity, 0.77-0.93; AUC 0.97 for the experimental set, n = 172 BM cases). However, the performance was even higher when used by an expert pathologist. In a clinical diagnostic setting, the MarrowQuant 2.0 workflow was compatible with integration as an additional quantitative assessment (augmented pathologist) (Fig. 6), thus abiding by the strategy proposed by the Swiss Digital Pathology Consortium.⁵⁹ Overall concordance in the clinical test set was satisfactory for the use of Eq. 1 when excluding cases with stromal expansion ($R^2 = 0.86-0.92$). Based on our data and to best approximate the visual cellularity assessment that constitutes current clinical routine, we recommend the use of Eq. 2 for low cellularity cases (<25%) and for cases of expanded stroma identified by an "other" area superior to 25%. Eq. 1 can be applied for all other cases. The outputs of both Eq. 1 and Eq. 2 are systematically reported by MarrowQuant 2.0., and our recommendation for the choice of equation as discussed above is indicated in the "recommended cellularity value" output column. Future prospective studies in diagnostic hematopathology should more broadly compare this strategy for the choice of equation, and consider the appropriate validation of MarrowQuant 2.0 in a clinical setting abiding by the challenges associated with the implementation of digital pathology tools, both from the regulatory perspective⁴⁰—especially in an open-source environment⁶⁰—and from the perspective of reproducibility on future software updates.⁶¹

The diagnostic or prognostic value of BM cellularity, adiposity, and adipocyte size quantification at the histopathology level has been confirmed in recent studies for MDS, acute lymphoblastic leukemia, AML, and obesity.^{15,28,29,62,63} In the case of extreme BM remodeling upon intensive chemotherapy for patients with AML/high-risk MDS, MarrowQuant 2.0 found a strong reciprocity (negative correlation) between the hematopoietic and adipocytic compartments ($R^2 = 0.68$, n = 42 in test set; $R^2 = 0.74$, n = 79 in experimental/validation set), matching previous quantitative validations in murine BM.^{38,64} In addition, we adapted the deep learning based StarDist extension as an additional workflow within QuPath0.3.2, which can be run in parallel to MarrowQuant 2.0 to provide individual adipocyte segmentation and thus adipocyte size distribution measurements and size-based color coding. This analysis constitutes, to our knowledge, the first

cellularity generated by MarrowQuant 2.0 (Eq. 1) with the clinical reference %cellularity score estimation in the test data set after correction of outliers (red dots) associated with myelofibrosis using %hematopoietic area (Eq. 2) (H&E biopsies n = 42 [test set], R^2 [pathology report] = 0.90, R^2 [digital image] = 0.93). (H) Cellularity measurement as suggested for the digital pathology analysis, classified per cellularity category (low cellularity $x < 25\%$, medium cellularity $25 \leq x < 50\%$, and aged-corrected high cellularity $x \geq 50\%$). Average patient age = 57 ± 9 years. Hm.Ar, hematopoietic area; Ma.Area, marrow area; Ad.Ar, adipocytic area. (I) Correlation of %cellularity generated by MarrowQuant 2.0 with the scoring performed on a preselected ROI by 4 expert reviewers on the validation data set (n = 125, $R^2 = 0.92$ [average pathologists] with ICC = 0.90).

Table 2
Sensitivity, specificity, and area under the ROC curve (AUC) for cellularity estimation in the experimental set

	Sensitivity	Specificity	Macroaveraged AUC	AUC (according to Hand-Till method)
Low cellularity <25%	0.81	0.93	0.98	0.91
Medium Cellularity $25 \leq x < 50\%$	0.88	0.89		
High cellularity $\geq 50\%$	0.99	1		

n = 172 hematoxylin and eosin–stained images when the selection of ROI was performed by an expert hematopathologist; MarrowQuant 2.0 was operated by an expert.

dedicated method to automatically segment adipocyte size distribution at diagnosis and follow-up of patients with AML/high-risk MDS when compared with those of age-matched control patients. We found very small–sized and medium-sized adipocytes to longitudinally remodel after myeloablative chemotherapy, whereas the proportion of other BM adipocyte subsets (large and very large) remains unchanged. Our observations highlight the plasticity of the BM and the specificity of distinct BM adipocyte subsets to undergo size remodeling. This observation fits with previous work by Lu et al,⁶³ who underlined that very small adipocytes at AML diagnosis correlate with poor prognosis, and with the current understanding that at least 2 populations of BM adipocytes coexist (regulated or labile and constitutive or stable subsets) with differential capacity for remodeling.⁶⁵ We conclude from this analysis that the incorporation of the quantitative stromal remodeling parameters provided by MarrowQuant 2.0 and StarDist in experimental hematopathology could be useful to investigate novel diagnostic or prognostic markers in clinical scenarios of intense BM remodeling, including myeloid malignancies, aplastic anemia, BM insufficiency syndromes, and hematopoietic progenitor transplantation or other advanced cellular therapies.

Contrasting the performance of MarrowQuant 2.0 to the results of previous studies and existing algorithms or tools, we found that our results are either comparable or provide an improvement to previous work (reviewed in [Supplementary Table S4](#)). Direct comparison was not possible as these algorithms, unlike MarrowQuant 2.0, are not open source. The closest approach was recently proposed by van Eekelen et al,³⁴ who developed a deep learning approach to automatically detect compartments within the BM compared with the visual estimation of 2 hematopathologists (ICC = 0.78, n = 109). The advantage of our study is that it integrates MarrowQuant 2.0 within the user-friendly platform of QuPath and that we validated our approach in the extreme cases of low and high cellularity (range from below 5%-100% cellularity in our study compared with 20% to 80% cellularity range in the study by van Eekelen et al³⁴). Similarly, Hagiya et al³⁷ used the HALO imaging algorithm to compare automatic cellularity measurements with visual estimation by 3 pathologists (ICC = 0.81, n = 165), and Kim et al³⁶ used nuclear counts to assess the cellularity against visual scoring ($R^2 = 0.816$, n = 325). Our study used not only the visual estimation of independent reviewers for validation but also the scoring extracted from the diagnostic pathology report. When compared with these studies, we observed a

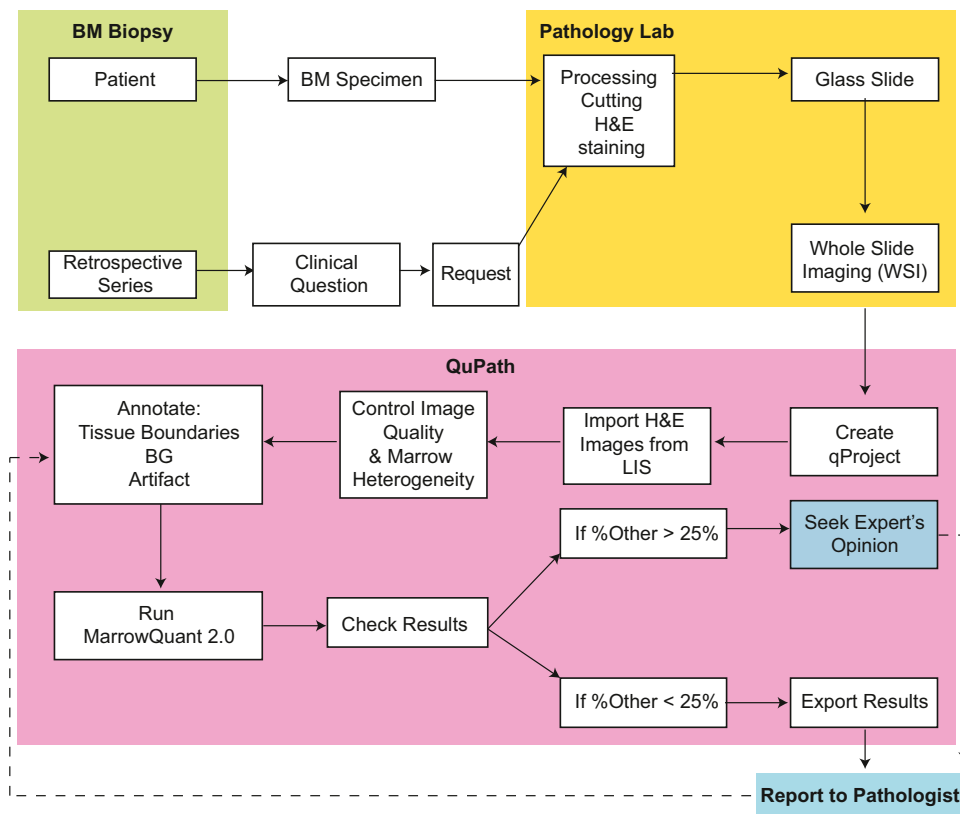


Figure 6. Proposal for MarrowQuant 2.0 potential integration within a digital pathologist plus machine hematopathology workflow. BG, background; LIS, laboratory information system.

higher agreement between the pathology report cellularity scores and MarrowQuant 2.0's cellularity assessment in both nonexpert and expert annotations, suggesting that our approach may provide a more accurate evaluation for the diagnostic pipeline.

Moreover, our algorithm performed similarly to the results reported for the fully automated algorithm developed by Nielsen et al.³⁵ for cellularity assessment, which agreed well with hematopathologists (ICC = 0.80, n = 8). However, their training set was based on 8 BM H&E-stained slides only to fully segment the BM tissue into either red (hematopoietic) or yellow (adipocytic) marrow to generate a heatmap with gradients of cellularity scorings. Compared with our much larger data sets and greater diversity of classification categories, MarrowQuant 2.0 outperforms the previously cited tools in simultaneous segmentation of the full BM space into 5 mutually exclusive BM compartments and overall accuracy of cellularity assessment. Moreover, Brück et al.³⁰ used convolutional neural networks, QuPath, and a multi-regression model to extract and quantify morphologic features from BM biopsies of patients with MDS and MDS/myeloproliferative neoplasms. Their main focus was the correlation with clinical data to build predictive models. Their work relied on the pixel classification power of QuPath to extract the hematopoietic compartment and adipocytes, with the bone plus stroma as a joint entity. Unlike Brück et al.³⁰ the approach used by MarrowQuant 2.0 segments bone and stroma as 2 separate compartments and identifies a separate IMV compartment, thus allowing for comparison of the 2 competing definitions of BM cellularity.

Our study faces some limitations. First, it is a semiautomatic workflow: the user must annotate the ROIs, artifacts, and background. This adds 2-3 minutes of hands-on annotation per image, in addition to the 1-2 min needed to have the full H&E-stained image quantified by MarrowQuant 2.0. However, we strategically kept the semiautomated feature because it gives the user control over the ROI selection, especially in cases where the BM tissue is heterogeneous or significantly modified by stromal expansion. The highly user-friendly and interactive interface offered by QuPath allows for rapid custom annotation of tissues and the exclusion of artifacts that should not be counted in the quantification. Second, it is important to note that we used a 25% low cellularity cutoff for sensitivity and specificity analyses. This value is routinely used as the diagnostic cutoff in the context of BM insufficiency and severe aplastic anemia.^{15,16,50,51} However, others have used a 30%-40% in the context of myeloid malignancies.^{34,43} Finally, MarrowQuant 2.0 classification, based on color and texture thresholding, has the disadvantage of misclassifying parts of the megakaryocyte cytoplasm as IMV or bone. Our training and experimental sets did not contain samples with a high number of megakaryocytes, leading to a nonsignificant misclassification (Supplementary Fig. S3). The algorithm should be used with caution in clinical scenarios where megakaryocyte hyperplasia is expected, including myeloproliferative syndromes. Future versions of the algorithm, which will incorporate deep learning and machine learning,^{34,66} are under development to overcome this problem.

In conclusion, MarrowQuant 2.0 is, to our knowledge, the first robust workflow to simultaneously segment the full BM space in human H&E-stained images into the hematopoietic compartment and 4 stromal compartments (bone, IMV, adipocytic, and other areas). When coupled with StarDist on adipocytes, it automatically provides an adipocyte size classification and distribution for the whole BM biopsy in the user-friendly environment of QuPath. Other applications could be focused on the quantification of bone to total marrow area in the context of osteopenia and osteosclerosis associated to hematologic disorders. We expect our

approach may be useful both to support quantitative BM cellularity and stromal compartment assessment in biomarker discovery for the diagnostic setting and to either provide a novel tool for research laboratories with limited access to expert pathologists or to homogenize BM cellularity assessments in clinical research. Potential integration in future digital clinical diagnosis pipelines should be determined by larger prospective studies and by the challenging regulatory landscape of digital histopathology in open-source digital environments.

Acknowledgments

This work could not have been possible without the expertise and investment of the EPFL Bioimaging and Optics Platform, especially O.B. and A.S. The authors thank the members of the Histology Core Facility, notably Dr Jessica Sordet-Dessimoz; Prof Andrew Janowczyk, Dr Pierre Moulin, Prof Peter Bankhead, Prof Inti Zlobec, Prof Gisou van der Goot, Prof Sandra Rizk, Dr Mohammad Hodroj, Marvy Moujabber, and Wangjie Liu; and the members of the Laboratory of Regenerative Hematopoiesis for critical reading of the manuscript and continuous encouragement to pursue this work.

Author Contributions

O.N., L.D.L., and R.S. conceived ideas, designed experiments, analyzed results, and wrote the manuscript. R.S. performed all experiments and analyses unless otherwise stated. R.S. and O.B. conceptualized, designed, and implemented MarrowQuant 2.0 code. R.S., O.B., and M.W. designed and implemented the "StarDist on adipocytes" version code. R.S., S.Blum, and C.R.C. performed histologic annotations and quantifications. L.D.L., B.B., C.B., R.Sarr, V.N., and C.R.C. evaluated histologic sections for hematopoietic cellularity scoring. R.S. and F.S. contributed to histologic mounting/analysis. MC, SC, OS, and SaB, contributed to patient recruitment and diagnostic reviews. N.P. performed digital image extraction and biobanking management. G.A. performed statistical analysis for ICC calculations. B.D. and A.S. provided significant input on the statistical analysis for tool development and validation. O.N., F.S.C., and L.D.L. initiated the project. All authors edited and reviewed the final manuscript.

Data Availability

The codes for MarrowQuant2.0 and StarDist for adipocytes are available on GitHub (<https://github.com/Naveiras-Lab/MarrowQuant2.0>), along with a detailed tutorial to install the 2 workflows within QuPath 0.3.2. The training set of hematoxylin and eosin images are accessible in the Image Data Resource database (Study idr0096, idr0096-tratwal-marrowquant; <https://idr.openmicroscopy.org/webclient/?show=project2104>, <https://idr.openmicroscopy.org/webclient/?show=project-2102> and <https://idr.openmicroscopy.org/webclient/?show=project-2103>.)

Funding

O.N. and R.S. were supported by the Personalized Health and Related Technologies (PHRT, project #308) and the University of Lausanne unrestricted funds. O.N. was supported by the Swiss National Science Foundation (SNF/SNSF) grants PP00P3/183725

and CRSII5_186271. F.S. was funded by Swiss National Science Foundation (SNF MD/PhD) grant 183986.

Declaration of Competing Interest

The authors state that there are no conflicts of interest to disclose.

Ethics Approval and Consent to Participate

The studies involving human participants were reviewed and approved by Commission cantonale d'éthique de la recherche sur l'être humain du Canton de Vaud (CER-VD). All patients in the experimental data set signed a specific written informed consent form for participation. Images for the training and test sets were anonymous. A specific written informed consent was not required for this tool development study in accordance with the national legislation and the institutional requirements.

Supplementary Material

The online version contains supplementary material available at <https://doi.org/10.1016/j.modpat.2022.100088>

References

- Abels E, Pantanowitz L, Aeffner F, et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the digital pathology association. *J Pathol*. 2019;249:286–294.
- Fraggetta F, L'Imperio V, Ameisen D, et al. Best practice recommendations for the implementation of a digital pathology workflow in the anatomic pathology laboratory by the European Society of Digital and Integrative Pathology (ESDIP). *Diagnostics*. 2021;11(11):2167. <https://doi.org/10.3390/diagnostics11112167>
- Hanna MG, Ardon O, Reuter VE, et al. Integrating digital pathology into clinical practice. *Mod Pathol*. 2022;35(2):152–164. <https://doi.org/10.1038/s41379-021-00929-0>
- Hanna MG, Hanna MH. Current applications and challenges of artificial intelligence in pathology. *Hum Pathol Rep*. 2022;27:300596. <https://doi.org/10.1016/j.hpr.2022.300596>
- Baxi V, Edwards R, Montalto M, Saha S. Digital pathology and artificial intelligence in translational medicine and clinical practice. *Mod Pathol*. 2022;35(1):23–32. <https://doi.org/10.1038/s41379-021-00919-2>
- Steiner DF, Chen PHC, Mermel CH. Closing the translation gap: AI applications in digital pathology. *Biochim Biophys Acta Rev Cancer*. 2021;1875(1), 188452. <https://doi.org/10.1016/j.bbcan.2020.188452>
- Cui M, Zhang DY. Artificial intelligence and computational pathology. *Lab Invest*. 2021;101(4):412–422. <https://doi.org/10.1038/s41374-020-00514-0>
- Bain BJ. Bone marrow trephine biopsy. *J Clin Pathol*. 2001;54(10):737–742. <https://doi.org/10.1136/jcp.54.10.737>
- Kurnick JE, Ward HP, Block MH. Bone marrow sections in the differential diagnosis of polycythemia. *Arch Pathol*. 1972;94(6):489–499.
- Kerndrup G, Pallesen G, Melsen F, Mosekilde L. Histomorphometrical determination of bone marrow cellularity in iliac crest biopsies. *Scand J Haematol*. 2009;24(2):110–114. <https://doi.org/10.1111/j.1600-0609.1980.tb02353.x>
- Arabyarmohammadi S, Leo P, Viswanathan VS, et al. Machine learning to predict risk of relapse using cytologic image markers in patients with acute myeloid leukemia posthematopoietic cell transplantation. *JCO Clin Cancer Inform*. 2022;(6), e2100156. <https://doi.org/10.1200/JCO.21.00156>
- Travlos GS. Histopathology of bone marrow. *Toxicol Pathol*. 2006;34(5): 566–598. <https://doi.org/10.1080/01926230600964706>
- Bain BJ, Clark DM, Wilkins B. *Bone Marrow Pathology*. 5th ed. Wiley-Blackwell; 2019.
- Dugan J, Pollyea DA, Abbott D, Schultheiss T. Bone marrow cellularity and recovery of blood counts after venetoclax and azacitidine in elderly acute myeloid leukemia patients unfit for induction chemotherapy. *Blood*. 2017;130(Suppl 1): 5086. https://doi.org/10.1182/blood.V130.Suppl_1.5086.5086
- Bennett JM, Orazi A. Diagnostic criteria to distinguish hypocellular acute myeloid leukemia from hypocellular myelodysplastic syndromes and aplastic anemia: recommendations for a standardized approach. *Haematologica*. 2009;94(2):264–268. <https://doi.org/10.3324/haematol.13755>
- DeZern AE, Churpek JE. Approach to the diagnosis of aplastic anemia. *Blood Adv*. 2021;5(12):2660–2671. <https://doi.org/10.1182/bloodadvances.202104345>
- Kumar V, Abbas AK, Aster JC, Perkins JA, eds. *Robbins Basic Pathology*. 10th ed. Elsevier; 2018.
- Mohan H. *Textbook of Pathology*. Jaypee Brothers; 2000.
- Ho-Yen DO, Slidders W. Bone marrow cellularity assessed by point-counting. *J Clin Pathol*. 1978;31(8):753–756. <https://doi.org/10.1136/jcp.31.8.753>
- Gruppo R, Lampkin B, Granger S. Bone marrow cellularity determination: comparison of the biopsy, aspirate, and buffy coat. *Blood*. 1977;49(1):29–31. <https://doi.org/10.1182/blood.V49.1.29.29>
- Jamshidi K, Windschitl HE, Swaim WR. A new biopsy needle for bone marrow. *Scand J Haematol*. 2009;8(1):69–71. <https://doi.org/10.1111/j.1600-0609.1971.tb02673.x>
- Dempster DW, Compston JE, Drezner MK, et al. Standardized nomenclature, symbols, and units for bone histomorphometry: a 2012 update of the report of the ASBMR Histomorphometry Nomenclature Committee. *J Bone Miner Res*. 2013;28(1):2–17. <https://doi.org/10.1002/jbmr.1805>
- Bravenboer N, Bredella MA, Chauveau C, et al. Standardised nomenclature, abbreviations, and units for the study of bone marrow adiposity: report of the nomenclature working group of the international bone marrow adiposity society. *Front Endocrinol*. 2020;10:923. <https://doi.org/10.3389/fendo.2019.00923>
- Hartssock RJ, Smith EB, Petty CS. Normal variations with aging of the amount of hematopoietic tissue in bone marrow from the anterior iliac crest: a study made from 177 cases of sudden death examined by necropsy. *Am J Clin Pathol*. 1965;43(4):326–331. <https://doi.org/10.1093/ajcp/43.4.326>
- Al-Adhahd AN, Cavill I. Assessment of cellularity in bone marrow fragments. *J Clin Pathol*. 1983;36(2):176–179. <https://doi.org/10.1136/jcp.36.2.176>
- Thiele J, Kvasnicka HM, Facchetti F, Franco V, van der Walt J, Orazi A. European consensus on grading bone marrow fibrosis and assessment of cellularity. *Haematologica*. 2005;90(8):1128–1132.
- Chen Y, Hoffmeister LM, Zaun Y, et al. Acute myeloid leukemia-induced remodeling of the human bone marrow niche predicts clinical outcome. *Blood Adv*. 2020;4(20):5257–5268. <https://doi.org/10.1182/bloodadvances.2020001808>
- Schemenau J, Baldus S, Anlauf M, et al. Cellularity, characteristics of hematopoietic parameters and prognosis in myelodysplastic syndromes. *Eur J Haematol*. 2015;95(3):181–189. <https://doi.org/10.1111/ejh.12512>
- Kidd PG, Saminathan T, Drachtman RA, Ettinger LJ. Comparison of the cellularity and presence of residual leukemia in bone marrow aspirate and biopsy specimens in pediatric patients with acute lymphoblastic leukemia (ALL) at day 7–14 of chemotherapy. *Med Pediatr Oncol*. 1997;29(6):541–543. [https://doi.org/10.1002/\(sici\)1096-911x\(199712\)29:6<541::aid-mpo4>3.0.co;2-k](https://doi.org/10.1002/(sici)1096-911x(199712)29:6<541::aid-mpo4>3.0.co;2-k)
- Brück OE, Lallukka-Brück SE, Hohtari HR, et al. Machine learning of bone marrow histopathology identifies genetic and clinical determinants in patients with MDS. *Blood Cancer Discov*. 2021;2(3):238–249. <https://doi.org/10.1158/2643-3230.BCD-20-0162>
- Chalkley HW. Method for the quantitative morphologic analysis of tissues. *JNCI J Natl Cancer Inst*. 1943;4(1):47–53. <https://doi.org/10.1093/jnci/4.1.47>
- Tuzuner N, Cox C, Rowe JM, Bennett JM. Bone marrow cellularity in myeloid stem cell disorders: Impact of age correction. *Leuk Res*. 1994;18(8):559–564. [https://doi.org/10.1016/0145-2126\(94\)90036-1](https://doi.org/10.1016/0145-2126(94)90036-1)
- Fong TP, Okafor LA, Schmitz TH, Thomas W, Westerman MP. An evaluation of cellularity in various types of bone marrow specimens. *Am J Clin Pathol*. 1979;72(5):812–816. <https://doi.org/10.1093/ajcp/72.5.812>
- 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 (Phila)*. 2022;54:318–327. <https://doi.org/10.1016/j.pathol.2021.07.011>
- Nielsen FS, Pedersen MJ, Olsen MV, Larsen MS, Røge R, Jørgensen AS. Automatic bone marrow cellularity estimation in H&E stained whole slide images. *Cytometry A*. 2019;95(10):1066–1074. <https://doi.org/10.1002/cyto.a.23885>
- Kim Y, Kim M, Kim Y, Han JH, Han K. Estimation of bone marrow cellularity using digital image nucleated cell counts in patients receiving chemotherapy. *Int J Lab Hematol*. 2014;36(5):548–554. <https://doi.org/10.1111/ijlh.12186>
- Hagiya AS, Etman A, Siddiqi IN, et al. Digital image analysis agrees with visual estimates of adult bone marrow trephine biopsy cellularity. *Int J Lab Hematol*. 2018;40(2):209–214. <https://doi.org/10.1111/ijlh.12768>
- Tratwal J, Bekri D, Boussema C, et al. MarrowQuant across aging and aplasia: a digital pathology workflow for quantification of bone marrow compartments in histological sections. *Front Endocrinol (Lausanne)*. 2020;11:480. <https://doi.org/10.3389/fendo.2020.00480>
- Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1):16878. <https://doi.org/10.1038/s41598-017-17204-5>
- Pell R, Oien K, Robinson M, et al. The use of digital pathology and image analysis in clinical trials. *J Pathol Clin Res*. 2019;5(2):81–90. <https://doi.org/10.1002/cjp2.127>
- Löwenberg B, Pabst T, Maertens J, et al. Addition of lenalidomide to intensive treatment in younger and middle-aged adults with newly diagnosed AML: the HOVON-SAKK-132 trial. *Blood Adv*. 2021;5(4):1110–1121. <https://doi.org/10.1182/bloodadvances.2020003855>

42. Lilleyman J. De Gruchy's Clinical Haematology in Medical Practice. *J Clin Pathol*. 1990;43(4):352-352.
43. Burkhardt R, Kettner G, Böhm W, et al. Changes in trabecular bone, hematopoiesis and bone marrow vessels in aplastic anemia, primary osteoporosis, and old age: a comparative histomorphometric study. *Bone*. 1987;8(3):157-164. [https://doi.org/10.1016/8756-3282\(87\)90015-9](https://doi.org/10.1016/8756-3282(87)90015-9)
44. Beekman KM, Veldhuis-Vlug AG, den Heijer M, et al. The effect of raloxifene on bone marrow adipose tissue and bone turnover in postmenopausal women with osteoporosis. *Bone*. 2019;118:62-68. <https://doi.org/10.1016/j.bone.2017.10.011>
45. Campbell WS, Talmon GA, Foster KW, Baker JJ, Smith LM, Hinrichs SH. Visual memory effects on intraoperator study design. *Am J Clin Pathol*. 2015;143(3):412-418. <https://doi.org/10.1309/AJCPUC3TYMS3QOBM>
46. Weigert M, Schmidt U, Haase R, Sugawara K, Myers G. Star-convex polyhedra for 3D object detection and segmentation in microscopy. In: *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2020:3655-3662. <https://doi.org/10.1109/WACV45572.2020.9093435>
47. Schmidt U, Weigert M, Broaddus C, Myers G. Cell detection with star-convex polygons. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018—21st International Conference*. Spain: Granada; 2018:265-273. https://doi.org/10.1007/978-3-030-00934-2_30. September 16-20, 2018, Proceedings, Part II.
48. Hand DJ, Till RJ. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach Learn*. 2001;45(2):171-186. <https://doi.org/10.1023/A:1010920819831>
49. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
50. Guinan EC. Diagnosis and management of aplastic anemia. *Hematol Am Soc Hematol Educ Program*. 2011;2011:76-81. <https://doi.org/10.1182/asheducation-2011.1.76>
51. Barbui T. The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion. *Blood Cancer J*. 2018;8:15.
52. Lee SH, Erber WN, Porwit A, Tomonaga M, Peterson LC, International Council for Standardization In Hematology. ICSH guidelines for the standardization of bone marrow specimens and reports. *Int J Lab Hematol*. 2008;30(5):349-364. <https://doi.org/10.1111/j.1751-553X.2008.01100.x>
53. Suchacki KJ, Tavares AAS, Mattiucci D, et al. Bone marrow adipose tissue is a unique adipose subtype with distinct roles in glucose homeostasis. *Nat Commun*. 2020;11(1):3097. <https://doi.org/10.1038/s41467-020-16878-2>
54. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9(7):671-675. <https://doi.org/10.1038/nmeth.2089>
55. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676-682. <https://doi.org/10.1038/nmeth.2019>
56. Riley RS, Williams D, Ross M, et al. Bone marrow aspirate and biopsy: a pathologist's perspective. II. interpretation of the bone marrow aspirate and biopsy. *J Clin Lab Anal*. 2009;23(5):259-307. <https://doi.org/10.1002/jcla.20305>
57. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405. <https://doi.org/10.1182/blood-2016-03-643544>
58. Rozman M, Mercader JM, Aguilar JL, Montserrat E, Rozman C. Estimation of bone marrow cellularity by means of vertebral magnetic resonance. *Haematologica*. 1997;82(2):166-170.
59. Janowczyk A, Baumhoer D, Dirnhofer S, et al. Towards a national strategy for digital pathology in Switzerland. *Virchows Arch*. 2022;481:647-652. <https://doi.org/10.1007/s00428-022-03345-0>
60. Harish KB, Price WN, Aphinyanaphongs Y. Open-source clinical machine learning models: critical appraisal of feasibility, advantages, and challenges. *JMIR Form Res*. 2022;6(4):e33970. <https://doi.org/10.2196/33970>
61. Haase R, Schmidt D, Rasband W, et al. *I2K_Poster_Haase_V6_ImageJ_repro.Pdf*. figshare; 2018, 1462758. <https://doi.org/10.6084/M9.FIGSHARE.7409525>. Bytes.
62. Qian S, Tang Y, Tang QQ. Adipose tissue plasticity and the pleiotropic roles of BMP signaling. *J Biol Chem*. 2021;296:100678. <https://doi.org/10.1016/j.jbc.2021.100678>
63. Lu W, Weng W, Zhu Q, et al. Small bone marrow adipocytes predict poor prognosis in acute myeloid leukemia. *Haematologica*. 2018;103(1):e21-e24. <https://doi.org/10.3324/haematol.2017.173492>
64. Tratwal J, Rojas-Sutterlin S, Bataclan C, Blum S, Naveiras O. Bone marrow adiposity and the hematopoietic niche: a historical perspective of reciprocity, heterogeneity, and lineage commitment. *Best Pract Res Clin Endocrinol Metab*. 2021;35(4):101564. <https://doi.org/10.1016/j.beem.2021.101564>
65. Scheller EL, Doucette CR, Learman BS, et al. Region-specific variation in the properties of skeletal adipocytes reveals regulated and constitutive marrow adipose tissues. *Nat Commun*. 2015;6(1):7808. <https://doi.org/10.1038/ncomms8808>
66. Sirinukunwattana K, Aberdeen A, Theissen H, et al. Artificial intelligence-based morphological fingerprinting of megakaryocytes: a new tool for assessing disease in MPN patients. *Blood Adv*. 2020;4(14):3284-3294. <https://doi.org/10.1182/bloodadvances.2020002230>