# QuantCell: machine learning based cell annotation from qualitative and quantitative imaging profiles

Wade Boohar[1], Bowen Wang[1], Zachary Thomas[1, 2], Anna Nogalska[1], Rong Lu[1, 2, 3, 4*]

[1]Department of Stem Cell Biology and Regenerative Medicine, Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research, Keck School of Medicine, [2]Department of Biomedical Engineering, [3]Department of Medicine, [4]Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA 90033, USA

* **Correspondence:**

Rong Lu

Keck School of Medicine, University of Southern California

1425 San Pablo Street, BCC 306, Los Angeles, CA 90033

Office: (323) 442 0169

Email: ronglu@usc.edu

**Abstract**

Recent advances in spatial omics enable high-resolution, multiplexed *in situ* imaging of gene and protein expression. A major challenge in analyzing these data is cell annotation, especially in complex tissues with limited molecular markers, overlapping marker expression, or rare cell types such as stem cells. Here, we present QuantCell, a machine learning (ML) based framework that leverages quantitative imaging data to improve cell annotation derived from qualitative data. QuantCell evaluates multiple ML models to identify the best-performing approach and applies a user-defined false discovery rate (FDR) to ensure reliable annotation. Using PhenoCycler imaging of mouse bone marrow, we show that QuantCell increases the number of annotated cells from 33.1% to 90.2% at a 5% FDR, achieving 96.5% accuracy based on the test dataset from conventional annotation. Benchmarking results demonstrate that QuantCell outperforms existing annotation methods in predictive accuracy. It also minimizes performance bias toward abundant cell types, enabling robust detection of rare cell populations. Unlike many existing methods, QuantCell does not require manual cell annotation, which can be labor-intensive for multi-marker cell types and unfeasible for rare cell populations. In summary, QuantCell provides an effective ML solution for spatial omics analysis, with a workflow that is broadly compatible with various imaging platforms and adaptable to project-specific needs.

## Introduction

Recent advances in spatial omics technologies have revolutionized our ability to investigate cellular organization and function within native tissue environments[1–4]. These technologies generate high-resolution, spatially resolved molecular profiles that quantify gene or protein expression at a single-cell or subcellular level[5–9]. A critical step in analyzing such data is to assign cell identities based on spatial molecular profiles, a process known as cell annotation or classification. This step sets a foundation for downstream analyses that infer cellular interactions, spatial organization, and functional states.

Several approaches have been developed to annotate cells based on spatial molecular profiles. Semi-supervised methods are widely used, relying on prior knowledge of marker genes[10–13] or predefined combinations of positive and negative markers for specific cell types[14,15] to guide the annotation of cell clusters generated by unsupervised algorithms. In contrast, unsupervised methods cluster cells without prior knowledge or labels, and infer cell identities directly from the data[16–18]. More recently, supervised approaches have emerged for large-scale datasets, leveraging manual annotations of a subset of cells to train classifiers that subsequently annotate the remaining cells[19,20].

However, existing cell annotation methods face significant challenges when applied to complex tissues containing diverse cell populations with complex molecular markers. Rare cell types, such as stem cells and cancer stem cells, play pivotal roles in physiological and pathological contexts[21–24], but are difficult to manually identify for training existing annotation models. The challenges are further exacerbated in spatial proteomics, where the number of measurable protein markers is constrained by cost and technical limits[25,26]. Yet, understanding protein expression and localization is essential for uncovering fundamental biological mechanisms and disease pathology[27–29]. Recent advances in artificial intelligence have shown great promise in addressing these challenges by leveraging the large-scale data generated by spatial imaging technologies[30–32]. However, the available models and platforms are often not generalizable beyond their original experimental context, as their performance is highly optimized for a specific set of tissues, markers, and technical procedures[33,34].

In this study, we present QuantCell, a machine learning (ML) based framework designed to improve cell annotation using qualitative and quantitative features derived from multiplexed imaging. QuantCell leverages cell segmentation and image processing by QuPath[35,36], which extracts intensity-based summary statistics across imaging channels and cellular compartments. Using predefined sets of positive and negative markers for expected cell types, QuantCell identifies cells with marker expression profiles unambiguously matching one of the marker sets. These cells serve as “ground truth” for training and evaluating multiple ML models, from which QuantCell selects the best-performing model to annotate previously unlabeled cells. To ensure classification confidence, QuantCell uses a false discovery rate (FDR) based thresholding strategy, allowing low-confidence or ambiguous cells to remain unannotated.

To demonstrate the utility of QuantCell, we imaged mouse bone marrow (BM) using the PhenoCycler (formerly CODEX) platform (Akoya Biosciences). PhenoCycler leverages

DNA-barcoded antibodies to enable high-resolution, multiplexed imaging[37]. Mouse BM contains over 20 distinct cell types with a wide range of frequencies[38–42]. Notably, hematopoietic stem and progenitor cells (HSPCs) in the BM are well characterized by complex combinations of cell surface markers[43–45], making BM an ideal system for evaluating the accuracy and robustness of QuantCell.

## Results

### QuantCell concept and design

We demonstrated the performance of QuantCell using data acquired from the PhenoCycler[37] platform. Mouse BM sections were imaged after staining with barcoded antibodies targeting 25 distinct proteins that are established cell type markers (Fig. 1a, Supplementary Fig. 1 and 2). These markers were able to annotate 21 BM cell types (Supplementary Table 1), which collectively comprise the vast majority of cells in the BM. Following the PhenoCycler manufacturer's recommendation, cell segmentation was performed using nuclear staining in QuPath[35], an open-source bioimage analysis software platform (Fig. 1b, Supplementary Fig. 3a). QuantCell performs equally well using nuclei-based and membrane-based cell segmentation (Supplementary Fig. 3). Subsequently, the presence or absence of each protein marker was annotated in QuPath for each segmented cell based on manual qualitative assessment of marker expression for dozens of cells (Fig. 1c).

Cell annotation in QuPath requires manually identifying representative cells for each cell type, which is labor-intensive, especially for rare cell types or those defined by multiple markers. QuantCell leverages qualitative marker expression profiles from QuPath to classify cells based on expected combinations of positive and negative markers for individual cell types (Fig. 1d, Supplementary Table 1). Using this conventional approach based on qualitative marker expression profiles, 33.1% of BM cells were annotated.

To classify the remaining cells, QuantCell leverages quantitative imaging data to improve cell annotation derived from qualitative data. Using conventional annotations as the ground truth, QuantCell trains ML models for cell classification on the quantitative imaging measurements of individual cells (Fig. 1e). This process involves selecting appropriate ML models (Fig. 1f) and fine-tuning their hyperparameters to optimize performance (Fig. 1g). For model selection and training, 80% of the cells annotated by the conventional approach were used, with the remaining 20% reserved as a testing dataset. The testing dataset was used to identify the best-performing model (Fig. 1h), which was then applied to classify cells that were not annotated by the conventional approach (Fig. 1i).

### Selection and optimization of ML models for cell classification

QuantCell systematically compares a variety of classification models based on their macro average precision (AP) score (Fig. 1f). The AP score is derived from the precision–recall curve (Fig. 2a), where precision reflects the accuracy of cell type prediction by the model, and recall represents the fraction of ground-truth cells that are correctly annotated. The AP score, which

corresponds to the area under the precision–recall curve, provides a summary measure of model performance. Precision, recall, and AP are calculated separately for each cell type, and the unweighted average across all cell types defines the macro AP score.

As a demonstration, we compared 10 models: random forest, multilayer perceptron, extra trees, logistic regression, linear SVC, nearest neighbors, decision tree, ridge classifier, LightGBM, and gaussian naive-Bayes (Fig. 2b). The QuantCell framework readily accommodates additional models as appropriate. Based on the AP score comparison, three best-performing ML models were identified (Fig. 2a,b) and further optimized through hyperparameter tuning (Supplementary Table 2). This fine-tuning process is essential to optimize each model's performance, ensuring the selection of the most effective model. During model selection and parameter tuning, principal component analysis (PCA) plots are generated to enable visual assessment. For each cell type, we used (i) all markers in the panel and (ii) only the markers used for its conventional annotation, to produce PCA plots showing conventionally annotated cells, QuantCell-annotated cells, and cells not assigned to that cell type (Supplementary Fig. 4). These PCA plots help guide model selection and parameter tuning by allowing researchers to identify potential domain-shift issues.

While the optimal model selected by QuantCell may vary depending on the dataset, our analysis using PhenoCycler data showed that the random forest model outperformed all other models tested (Fig. 2c) and required substantially less runtime than the second-best model, the multilayer perceptron (Supplementary Fig. 5a). The random forest model operates by aggregating the predictions of multiple decision trees trained on subsets of the data, selecting the most common classification outcome. This ensemble approach reduces overfitting while maintaining strong predictive performance[46]. Random forest models are widely used for classification tasks in many applications due to its robustness, versatility, and efficacy[47,48].

**Impact of the data sources and statistical descriptors on prediction accuracy**

During the ML model selection and optimization, the model was trained on all imaging measurements obtained from QuPath's cell segmentation and marker annotation. This comprehensive quantitative dataset included data captured from multiple cellular compartments, such as membrane, cytoplasm, nucleus, and whole-cell region. For each compartment, multiple statistical descriptors were collected, including the minimum, maximum, standard deviation, median, and mean of fluorescence intensity for each protein marker. Understanding the impact of data quantity, data source, and statistical descriptor on prediction accuracy could help identify data types with the highest and lowest predictive value, including those with detrimental effects that should be excluded to enhance QuantCell performance (Fig. 2d-f). This analysis was conducted using the hyperparameter-optimized random forest model, which demonstrated the highest macro AP score in our evaluation (Fig. 2a-c). Measurements of all markers were included when comparing cellular compartments and statistical descriptors. Cellular compartments were defined using default StarDist[49]-based cell segmentation in QuPath.

Our analyses suggest that prediction precision depends on the amount of training data used. However, we observed minimal difference between using the full training dataset and using only

75% of it, suggesting that our current training dataset provides sufficient predictive power (Fig. 2d). Comparing various data sources and statistical descriptors, our results suggest that the model performed the best when incorporating all cellular compartments and statistical measurements (Fig. 2e). Among the statistical measurements, the mean produced the best prediction accuracy, followed by the median, while the minimum performed markedly worse (Fig. 2e). This is expected as the minimum primarily reflects background intensity. Interestingly, standard deviation outperformed the maximum (Fig. 2e), suggesting that variability in signal may carry more predictive value than peak intensity alone.

Comparing data from different cellular compartments, we found that the intensity measured across the entire cell generated the best prediction accuracy (Fig. 2e). Although most of the protein markers analyzed are localized to the cell surface, our results show that features derived from the cell membrane exhibited the lowest predictive performance (Fig. 2e). This is likely due to the dense cellular architecture of the BM, where closely packed cells can lead to overlapping membrane signals from adjacent cells. As a result, membrane measurements may not accurately represent protein expression levels in individual cells. In addition, our imaging data were acquired using a wide-field microscope rather than a confocal microscope. As a result, the nuclear region likely includes signals originating from the cytoplasm and membrane in the z-dimension.

We also performed a feature exclusion analysis to evaluate whether excluding any type of imaging measurement would improve model performance. Overall, the impact of excluding any single feature, for all 25 protein markers, was negligible (Fig. 2f). Notably, removing features with high individual importance, such as the median or mean intensity of all cellular compartments, or all statistical measures across the whole-cell region (Fig. 2e), resulted in only minor reductions in the macro AP (Fig. 2f). This suggests that the QuantCell prediction remains robust and can compensate using other spatial measurements, even in the absence of the most informative features.

**Evaluation of QuantCell prediction performance**

Using the hyperparameter-optimized random forest model and imaging data from all cellular compartments and statistical measurements, we next evaluated the predictive performance of QuantCell. We began by examining potential misclassifications using the reserved testing dataset annotated by the conventional method (Fig. 3a). While the conventional classification may contain errors, these annotations were used to train QuantCell and therefore served as the ground truth for this analysis.

Our results show that 96.5% of cells classified by QuantCell were consistent with annotations from the conventional method, demonstrating high prediction accuracy across all cell types (Fig. 3a, Supplementary Fig. 6). Precision was also high across all cell types, with a minimum cell type precision rate of 87.5%, underscoring the model's robustness even in unbalanced cell populations (Fig. 3a, Supplementary Table 1, Supplementary Fig. 7). With high precision rates, QuantCell was able to increase the number of annotated cells from 33.1% to 90.2% at a false discovery rate of 5% (Fig. 3b). The cell type frequencies predicted by QuantCell were consistent

with those obtained from the conventional annotation (Fig. 3c, Supplementary Fig. 8a). To independently validate QuantCell's classifications, we performed flow cytometry on unfractionated BM cells from the same mice, focusing on select hematopoietic cell types. The results showed consistent cell type frequencies between QuantCell's annotation and flow cytometry data, further validating QuantCell's classification accuracy (Fig. 3d, Supplementary Fig. 8b).

**Comparison of QuantCell with existing annotation frameworks**

To evaluate QuantCell's performance, we compared it with other cell annotation models developed for proteomics data, which are difficult to annotate due to limited marker availability (Table 1). Astir is a probabilistic model that uses deep recognition neural networks to annotate cell types based on user-defined markers for each cell type[10]. AnnoSpat is a recently proposed semi-supervised approach that first clusters cells, then identifies a subset of confidently annotated cells using predefined positive and negative protein markers[15]. These annotated cells are used to train a neural network, which subsequently annotates the remaining cells. MAPS is another recently introduced model that uses a feedforward neural network trained on a ground truth dataset to perform cell type annotation[19].

We compared each model's performance on the training dataset using balanced accuracy, the unweighted average accuracy across all cell types. This metric was used instead of macro AP because some models do not provide cell type probabilities required for AP calculation. Our analysis shows that QuantCell achieves higher prediction accuracy than existing models (Fig. 3e) and operates 45 times faster than the second most accurate model, MAPS (Supplementary Fig. 5b). QuantCell also demonstrates robust performance across all cell types (Supplementary Fig. 9) and successfully annotates a substantial number of cells that cannot be classified by conventional approaches (Fig. 3b). QuantCell is well suited for analyzing data from experiments where the marker panels may not capture all expected cell types, allowing for flexible and high-confidence annotation even in partially characterized biological systems.

Multilayer perceptron and MAPS are feed-forward neural networks, and both demonstrated relatively strong performance in our evaluation (Fig. 2b, 3e). However, MAPS relies on manually annotated cells for training, which is a major limitation[19] (Table 1). Manual cell annotation is labor-intensive for cell types defined by multiple markers and unfeasible for rare cell populations. In contrast, QuantCell automatically generates training datasets based on molecular marker profiles, eliminating the need for manual cell type annotation.

The other two benchmarked methods, Astir and AnnoSpat, also have limitations in generating training datasets. AnnoSpat's reliance on clustering compromises its performance in identifying rare cell types. While clustering is useful for pattern discovery, this unsupervised approach lacks the sensitivity of supervised learning, particularly in imbalanced datasets where rare cell types can be absorbed into larger, heterogeneous clusters[50]. Furthermore, clustering treats all markers as equally informative. Noise from uninformative markers can compromise the clustering and annotation accuracy. Astir is constrained by its use of only positive markers for annotation (Table 1), assuming that marker expression is higher in the corresponding cell types

231 than in all other cells. This assumption is not valid for markers shared by multiple cell types,
232 which is common among stem/progenitor cells and diseased cells.

233 **Cell type specific prediction performance**

234 A major strength of QuantCell is its strong performance across cell types with widely varying
235 frequencies. We further evaluated QuantCell's performance across individual cell types using
236 the F1 score, the harmonic mean of precision and recall which summarizes predictive
237 performance for each cell type. Our results suggested that the F1 score was significantly
238 influenced by cell frequency and cell circularity (Fig. 4a), likely due to its effect on the sample
239 size and quality for model training. In contrast, features such as cell size and the number of
240 protein markers used in conventional annotation did not show direct correlation with F1 score
241 (Fig. 4a), indicating the robustness and consistency of QuantCell's performance across cell
242 types.

243 To assess the contribution of individual protein markers to cell type classification, we
244 systematically excluded one marker at a time from the training dataset. This analysis revealed
245 the markers that are indispensable for accurately annotating specific cell types (Fig. 4b,
246 Supplementary Fig. 10). These essential markers often, though not always, correspond to those
247 used in conventional annotation (Supplementary Table 1). This discrepancy reflects how the
248 model weighs informative features for each cell type and the redundancy among markers. For
249 example, removing one lineage marker may have little effect on HSPCs because other lineage
250 markers provide similar information. Together, these results validate QuantCell's predictions
251 and highlight its reliability and effectiveness in classifying diverse cell types.

252 **Discussion**

253 In this study, we developed QuantCell, an ML based framework designed to improve cell
254 annotation using qualitative and quantitative imaging data. We validated the performance of
255 QuantCell using multiple complementary approaches. First, we assessed its accuracy on an
256 independently held-out test dataset generated through conventional annotation (Fig. 3a,
257 Supplementary Fig. 7). Second, we confirmed QuantCell predictions using flow cytometry
258 measurements (Fig. 3d, Supplementary Fig. 8b). Third, all marker and cell annotations were
259 independently reviewed by three field experts. Fourth, we evaluated QuantCell on three
260 externally annotated datasets from published studies, demonstrating robust, high-precision
261 performance (Supplementary Fig. 11). Finally, PCA analyses showed that QuantCell-annotated
262 cells tightly colocalize with conventionally annotated cells and are clearly separated from cells
263 not assigned to the corresponding cell type (Supplementary Fig. 4).

264 QuantCell outperforms existing methods in predictive accuracy and offers significant
265 advancements in the following key areas. First, QuantCell optimizes annotation performance
266 across all cell types (Fig. 3a, Supplementary Fig. 7, 9). By avoiding performance bias toward
267 abundant cell types, it provides a more balanced and accurate classification. This is particularly
268 important in biomedical datasets where rare cell types, such as stem cells or diseased cells, are
269 often the primary focus but are poorly detected by existing methods. Second, QuantCell

performs annotation based on a user-defined FDR, enabling stringent control over classification confidence (Fig. 3b). It reduces the risk of misclassification by leaving uncertain cells unannotated, which improves the reliability of biological interpretation. Third, QuantCell does not require manual cell type annotation, which is labor-intensive for cell types defined by multiple markers and unfeasible for rare cell populations. It uses both positive and negative markers to guide training dataset generation, enabling evaluation of each marker's contribution (Fig. 4b, Supplementary Fig. 10). Nonetheless, QuantCell's performance is constrained by the quality of the training data. It may underperform on datasets with sparse or low-dimensional features (Fig. 2e) or for cell types with limited or low-quality training data (Fig. 4a). To improve the quality of the training and testing datasets, one potential approach is to filter out outlier cells annotated by the conventional method. Because the conventional method assigns cell identities based on predefined marker combinations (Supplementary Table 1), cells of the same type are expected to cluster in the corresponding marker space (Supplementary Fig. 4, left panels). By comparing the density of cells assigned to a given type with the densities of other types in this space, it is possible to estimate the probability that each cell truly belongs to its assigned identity. The cell distributions can be visually inspected in PCA plots to determine whether low-probability cells should be removed. Implementing such procedures in the future could further improve model predictions.

Despite this limitation, QuantCell is able to increase the number of annotated cells from 33.1% to 90.2% at a 5% FDR, achieving 96.5% overall accuracy based on the test dataset from conventional annotation (Fig. 3b). Our benchmarking analysis shows that QuantCell outperforms existing platforms in predictive accuracy (Fig. 3e). It has the potential to be compatible with a wide range of spatial omics platforms, including two- and three-dimensional imaging, RNA-based technologies such as seqFISH[51] and MERFISH[52], and sequencing-based approaches like Visium[53], Slide-seq[54], and Stereo-seq[9]. The framework is also adaptable to project-specific priorities. For example, QuantCell is configured by default to optimize performance equally across all cell types, but users can customize weighting schemes to prioritize specific cell populations of interest. Users can also incorporate their preferred supervised ML models in QuantCell for model selection (Fig. 1f). Future studies could further expand QuantCell's capabilities by incorporating transfer learning to enhance performance in data-sparse settings. In addition, incorporating domain adaptation techniques may enable QuantCell to generalize across tissue types, experimental platforms, or imaging modalities with minimal retraining.

In conclusion, QuantCell represents a powerful application of ML in the rapidly advancing field of spatial omics. It improves cell annotation by leveraging large-scale data from high-resolution multiplexed imaging. We demonstrate that QuantCell not only accurately annotates a substantial number of cells left unclassified by conventional methods, but also performs robustly across diverse BM cell types, underscoring its potential to resolve complex cellular phenotypes in heterogeneous tissues.

## Methods

<u>Mice</u>

311 C57BL/6J mice (Jackson Laboratory stock #000664) were maintained in the Research Animal
312 Facility at the University of Southern California (USC) under a 12:12 hour light-dark cycle, with
313 controlled room temperature (22 ± 1 °C) and humidity (55 ± 10%). All animal procedures were
314 approved by the Institutional Animal Care and Use Committee of USC.

315 PhenoCycler antibodies

316 The following primary antibodies were obtained from Akoya Biosciences: RX025-AF488 anti-
317 CD11b (M1/70), RX030-Cy5 anti-CD11c (N418), RX020-ATTO550 anti-CD19 (6D5), RX021-Cy5
318 anti-CD3 (17A2), RX002-ATTO550 anti-CD31 (MEC13.3), RX026-ATTO550 anti-CD4 (RM4-5),
319 RX005-ATTO550 anti-CD44 (IM7), RX007-AF488 anti-CD45 (30-F11), RX010-AF488 anti-
320 CD45R (RA3-6B2), RX027-Cy5 anti-CD71 (RI7217), RX029-ATTO550 anti-CD8a (53-6.7),
321 RX024-Cy5 anti-Ly6g (1A8), RX014-ATTO550 anti-MHCII (M5/114.15), RX003-Cy5 anti-TCRαβ
322 (H57-597), and RX004-AF488 anti-TER119 (TER-119). The following primary antibodies were
323 obtained from ThermoFisher Scientific: BX035-ATTO550 anti-CD105 (209701), BX050-AF647
324 anti-CD115 (AFS98), BX045-Cy5 anti-CD117 (2B8), RX013-AF488 anti-CD34 (RAM34), BX037-
325 ATTO550 anti-CD41 (MWReg30), BX040-AF488 anti-endomucin (polyclonal), BX036-Cy5 anti-
326 F4/80 (CI:A3-1), and BX041-ATTO550 anti-CD127 (A7R34). The following primary antibody was
327 obtained from Leinco Technologies: BX028-AF488 anti-CD32/CD16 (2.4G2). The following
328 primary antibody was obtained from R&D Systems: BX049-ATTO550 anti-LEPr (polyclonal). All
329 primary antibodies obtained from vendors other than Akoya Biosciences were conjugated to
330 Akoya barcodes by Akoya Biosciences.

331 Antibody validation and quality control

332 All antibodies were conjugated and validated by Akoya Biosciences. Successful conjugation was
333 verified by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE), confirming
334 that each conjugated antibody migrated at the expected molecular weight. After verification of
335 conjugation, Akoya staff stained our tissue sections with the antibodies and evaluated antibody
336 performance using a standardized scoring system: –1 (non-specific binding), 0 (no signal/no
337 positive cells), 1 (weak signal), 2 (good staining), and 3 (saturated). Antibody clones and/or
338 dilutions were adjusted based on these evaluations. A blank imaging cycle was used as a negative
339 control. All antibodies included in the final panel received a rating of 2.

340 Bone marrow preparation for flow cytometry analysis

341 Bone marrow cells were isolated from crushed bones of the arms, spine, pelvis, and sternum.
342 Cells were separated from bone debris using a density gradient (Sigma-Aldrich, Cat. #11191).
343 The resulting cell suspension was then divided into two fractions for antibody staining: one for
344 analyzing mature blood cells and the other for identifying hematopoietic stem and progenitor cells
345 (HSPCs).

346 The following antibodies were used to identify mature blood cells: PE anti-TCRαβ (H57-597), PE-
347 Cy5 anti-CD8α (53-6.7), PE-Cy5.5 anti-CD19 (eBio1D3), PE-Cy7 anti-CD45.2 (104), eFluor660

348 anti-B220 (RA3-6B2), APC-eFluor780 anti-CD11b (M1/70), BV510 anti-Gr-1 (RB6-8C5), PE-
349 eFluor610 anti-CD4 (GK1.5), and PerCP-eFluor710 anti-CD115 (AF598).

350 The following antibodies were used to identify HSPCs: PE anti-CD150 (TC15-12F12.2), PE-Cy5
351 anti-CD135 (A2F10), PE-Cy7 anti-Ly-6A/E (D7), AF700 anti-CD45.2 (104), APC-eFluor780 anti-
352 CD117 (2B8), BV510 anti-CD127 (A7R34), eFluor450 Lineage (anti-CD3 (17A2), anti-CD19
353 (eBio1D3), anti-CD11b (M1/70), anti-Gr-1 (RB6-8C5), anti-TER-119 (TER-119), anti-CD4
354 (GK1.5), anti-CD8α (53-6.7), anti-B220 (RA3-6B2)), Biotin anti-CD32/CD16 (93), and BV605
355 Streptavidin (secondary).

356 Bone marrow cells were incubated with the mature blood cell antibody mix for 1 hour, or with the
357 HSPC antibody mix for 1.5 hours, followed by a 20-minute incubation with the secondary antibody.
358 After antibody incubation, samples stained for mature blood cells were resuspended in Dulbecco's
359 Phosphate Buffered Saline (DPBS) (ThermoFisher, Cat. #21600010), supplemented with 2% fetal
360 bovine serum (FBS; Avantor, Cat. #76419-584) and 4′,6-diamidino-2-phenylindole (DAPI).
361 Samples stained for HSPCs were resuspended in DPBS containing 2% FBS and propidium
362 iodide. All samples were analyzed using a FACSAria II (Becton Dickinson) as previously
363 described[55–57]. Flow cytometry data were analyzed using FlowJo software version 10.8.2 (Becton
364 Dickinson).

365 Bone marrow preparation for imaging

366 Bone marrow was carefully flushed from the diaphysis of femurs and tibias using a syringe filled
367 with DPBS with 2% FBS, ensuring gentle handling to obtain the whole, intact bone marrow plugs.
368 The collected bone marrow was immediately fresh-frozen in tissue freezing medium (Electron
369 Microscopy Sciences, Cat. #72592) for downstream applications.

370 Bone marrow imaging

371 Glass coverslips, 22 mm x 22 mm, were pre-treated with poly-L-lysine (Santa Cruz Biotechnology,
372 Cat. #sc-286689) to enhance tissue adherence. The fresh-frozen bone marrow samples were
373 cryosectioned at a thickness of 5 µm and mounted onto the prepared coverslips. Six BM section
374 samples from six mice were imaged and processed, yielding a total of 245,719 cells identified
375 through QuPath segmentation. Cells from all six samples were pooled for downstream analyses.
376 Antibody staining and imaging were performed by Akoya Biosciences personnel using Akoya's
377 established protocols.

378 Cell segmentation

379 Nuclei-based cell segmentation was performed based on DAPI stains in QuPath[35] (version
380 0.3.0) using the StarDist[49] (version 0.3.0) algorithm, following default settings (provided in our
381 GitHub, see Code availability). For nuclei-based segmentation, QuPath automatically infers
382 cytoplasm and membrane compartments by expanding nuclear boundaries using a watershed
383 approach. Membrane-based segmentation was performed with Cellpose-SAM[58] (version 4.0.6)
384 using a composite image of DAPI, F4/80, CD45, Endomucin, Sca1, and CD71. To recover
385 cellular compartments using Cellpose, segmentation was run independently on the composite

386 (whole cell) and the DAPI channel (nucleus). The cytoplasmic compartment was defined by
387 subtracting the nuclear area from the whole-cell area. QuantCell performed comparably well
388 with either membrane-based or nuclei-based cell segmentation (Supplementary Fig. 3). Nuclei-
389 based segmentation was used for all other analyses.

390 Conventional annotation and QuPath settings

391 Qualitative marker annotation was performed in QuPath[35] (version 0.3.0). For each marker,
392 approximately 20-70 positive and 20-70 negative cells from across the tissue section were
393 manually annotated and used to classify all cells in the section. This process involved multiple
394 iterations of model training, manual review, and additional manual refinement of model-
395 generated annotations, until the resulting cell annotations passed human assessment. A cell
396 was considered positive for a marker when its signal clearly exceeded the background signal
397 and exhibited the expected staining pattern for that marker. Signal adjustment and staining
398 patterns were used to minimize the effects of non-specific binding and background signals.
399 Mutually exclusive and co-expressing marker combinations were also used for further
400 verification.

401 To reduce potential cross-contamination from neighboring cells, our annotation strategy
402 incorporated the spatial distribution of marker signal within each cell. During manual binary
403 marker annotation in QuPath, cells were classified as negative when the apparent signal was
404 confined to regions adjacent to neighboring positive cells. In the QuantCell annotation, spatial
405 distribution for each marker was quantified using multiple statistical metrics, including the
406 standard deviation of signal intensity within each cellular compartment.

407 The resulting marker assignments were independently verified by three field experts, including a
408 board-certified pathologist, the director of our imaging core facility, and an experienced tissue
409 biologist. After verification, the protein-expression–by–cell matrix and the binary
410 positive/negative marker annotations were downloaded from QuPath. The binary annotations
411 were then used for the conventional cell annotation. Detailed information on the procedure used
412 to generate the protein-expression-by-cell matrix is available in our GitHub repository (see Code
413 availability).

414 Using the binary positive/negative marker annotations generated by QuPath, QuantCell
415 constructed the training and testing datasets by annotating cells that exactly matched
416 predefined marker combinations (Supplementary Table 1). Cells with ambiguous profiles, such
417 as those matching more than one combination or none at all, were not assigned a cell type and
418 were labeled as “Other”. To further verify cell annotation accuracy, single-cell annotations
419 generated by both the conventional approach and by QuantCell were independently assessed
420 by three experts as mentioned above.

421 Data preprocessing

422 Cells that were successfully annotated by the conventional method were split into an 80/20
423 train-test dataset using train_test_split() from scikit-learn[59], with stratification based on class

distribution. Features were normalized to have a mean of 0 and a standard deviation of 1 using StandardScaler() from scikit-learn. The scaler was fitted to the marker distributions in the training dataset and then applied to both the testing dataset and unannotated cells without refitting, to prevent data leakage. Cell type labels were encoded using LabelEncoder() from scikit-learn after excluding any cells that were originally unannotated.

Model selection

Initial model performance was evaluated across eight different base classifiers from scikit-learn with default parameters, using the training dataset and 10-fold cross-validation. The three best-performing models were selected based on macro average precision (AP) score. These models were further optimized using HalvingGridSearchCV from scikit-learn, with 10-fold cross-validation on the training dataset.

Training size analysis

To evaluate the dependence of model performance on training data availability, we performed a training size analysis using 5-fold cross-validation. Within each fold, the test set was held constant to ensure paired comparisons, while the training set was randomly subsampled at varying proportions (5%, 10%, 25%, 50%, 75%, and 100%). Consequently, the 100% condition represents the full set of remaining 4 folds and is equivalent to the 80/20 train-test split used during preprocessing.

Feature analysis

Performance of the model was evaluated by systematically varying the inclusion or exclusion of specific measures or markers across 10-fold cross-validation on the training set. Each fold was treated as an individual data point. F1 scores were calculated without applying a false discovery rate (FDR) threshold. The model using all features was used as the baseline to assess changes in macro AP and F1 scores

Model thresholding

Thresholding was performed independently for each cell type by examining the predicted cell type probabilities for each cell on the testing dataset and selecting a threshold that ensured the FDR was less than or equal to the specified target. For thresholded analyses in this study, the default FDR target was set to 5%. Thresholding was not applied in any analysis involving the AP score, as AP is calculated by evaluating performance across all possible thresholds.

Benchmarking QuantCell against existing models

For all benchmarking analyses, annotations from the conventional method were used as the ground truth. For Astir, AnnoSpat, and MAPS, mean expression of each marker across the entire cell was used as input, following the pipeline described in the corresponding papers. MAPS was also provided the cell area ($\mu m^2$) as directed.

Astir and AnnoSpat were applied to the entire conventionally annotated dataset. No preprocessing was performed. The models were run with default parameters. The same marker combinations used in QuantCell were applied, except negative markers were excluded for Astir, as they are not supported by this model.

MAPS was trained and tested using the same 80/20 train–test split as QuantCell. Labels were encoded as in the QuantCell workflow prior to training. No preprocessing was applied, and default parameters were used, except for “min_epochs” which was reduced to 5 and “patience” which was reduced to 25 to avoid unnecessary computation.

QuantCell was trained and tested using the same 80/20 train–test split as MAPS. Performance benchmarking was conducted both with thresholding at 5% FDR**.**

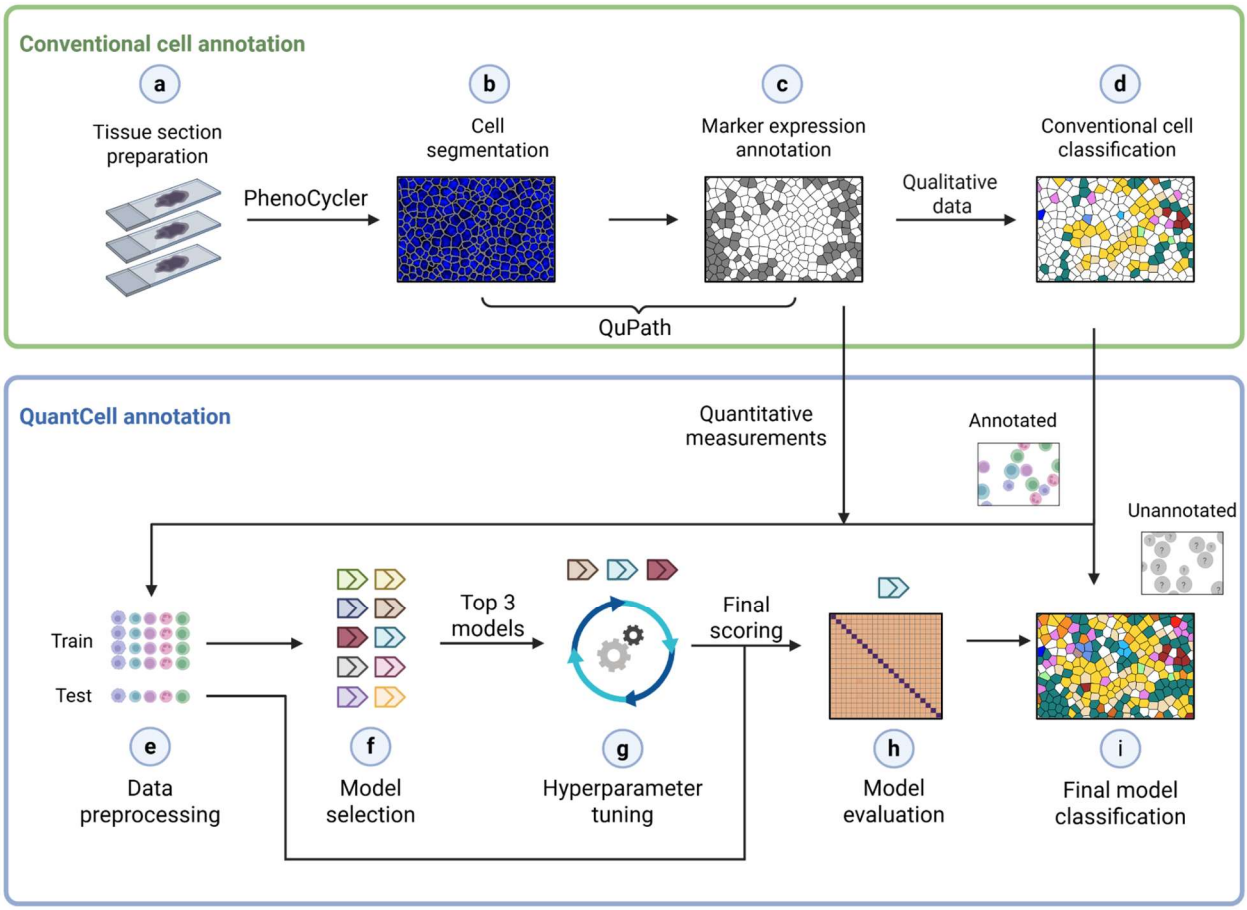
Computational environment

All benchmarking tests and analyses were performed on a dedicated server with Ubuntu 20.04.6 LTS, 2 x Intel(R) Xeon(R) Gold 6136 CPU @ 3.00GHz (24 physical cores, 48 threads total, 49.5 MB L3 cache), 494.7 GB memory. All code is in Python, except for a Bash script used to run AnnoSpat.

Statistics and reproducibility

Statistical tests used in data analysis are detailed in the relevant figure legends and "Methods" sections. No statistical method was used to predetermine sample size. Expression data from markers with poor staining quality or those not included in any marker combinations (Supplementary Table 1) were excluded from analysis. Flow cytometry validation was performed for cell types with gating strategies matching the marker combinations (Supplementary Table 1). Where applicable, 10-fold cross-validation was employed for model selection. A fixed random seed was applied during data splitting to ensure reproducibility across sessions.

483 **Figures**




484

485 **Figure 1. Overview of conventional and QuantCell annotation pipelines.** a) Tissue sections
486 are prepared, stained, and imaged. b) Cell segmentation is performed using QuPath[35], based
487 on 4′,6-diamidino-2-phenylindole (DAPI) nuclei staining. c) Each protein marker is manually
488 annotated as positive or negative in a small subset of cells. These annotations are then
489 propagated to the remaining cells using QuPath. d) Cells are classified based on expected
490 combinations of positive and negative markers for each cell type. e) Data are scaled and
491 encoded. 80% of the cells annotated by the conventional method are used for model selection
492 and training, while the remaining 20% are reserved as a testing dataset. f) Multiple machine
493 learning models are evaluated using 10-fold cross-validation. g) The top three performing
494 models are further optimized through hyperparameter tuning to select the most effective model.
495 h) The final model is evaluated on the reserved 20% testing dataset. i) The final model is
496 applied to classify cells that are unannotated in d), using a 5% false discovery rate threshold.

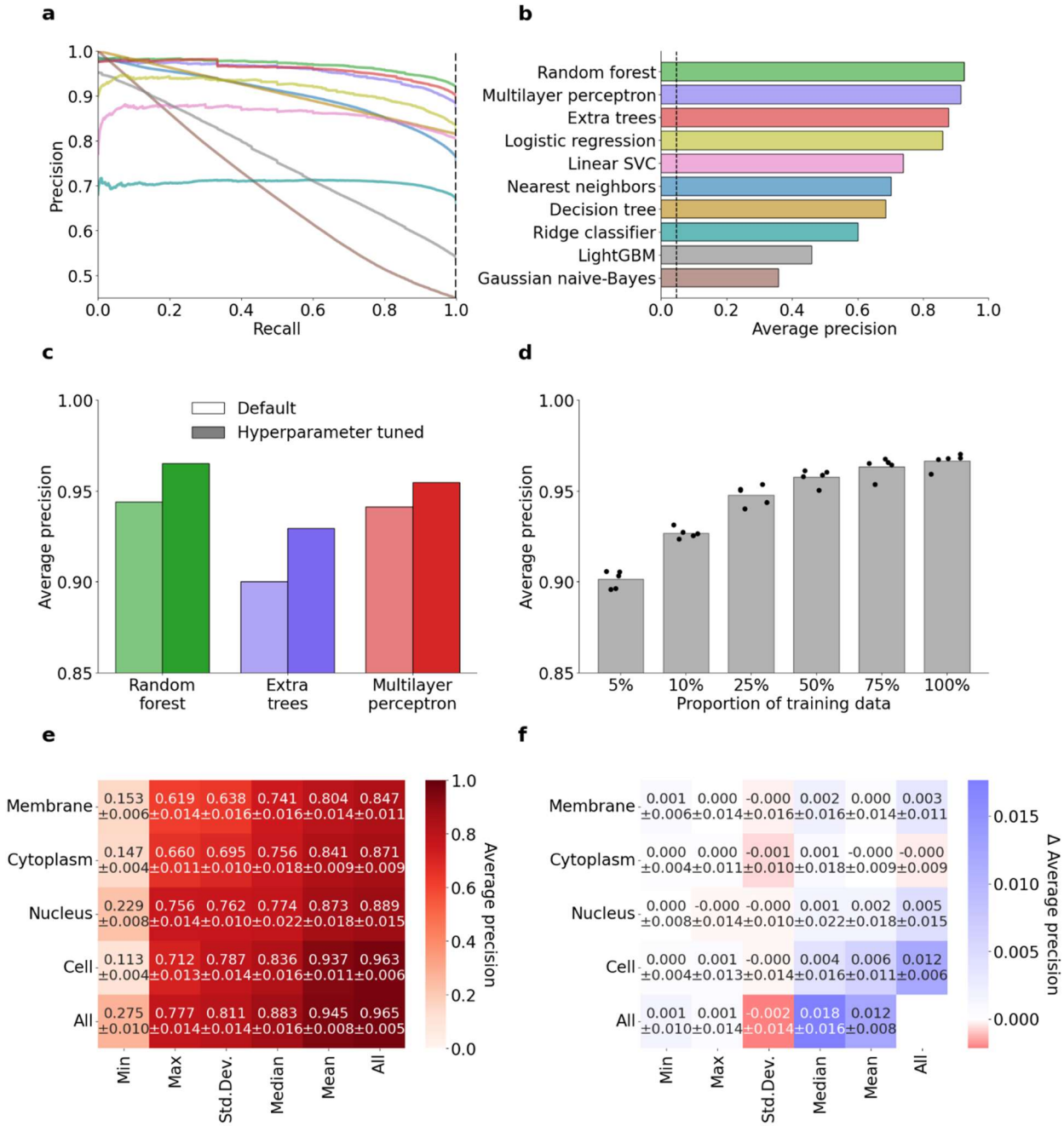



**Figure 2. Selection and optimization of machine learning models for cell classification.** a) Precision-recall curve for base classifiers without hyperparameter tuning. Data represent the unweighted average of all cell types. b) Macro average precision (AP) score for each classifier, equivalent to the area under the precision-recall curve (AUPRC). The dashed line indicates the AUPRC corresponding to random prediction performance. c) Comparison of the default and hyperparameter-optimized performance of the top three models based on AP score. d) Comparison of QuantCell performance using varying proportions of the training dataset. Average scores across 5-fold cross-validation are shown as black dots above each bar. e) Evaluation of feature importance using the best-performing model, hyperparameter-optimized random forest. Shown are the prediction AP scores using training data from the indicated

measurement(s). f) Feature exclusion analysis to assess the contribution of individual features. Shown are changes in macro AP when the model was trained on all measurements except the indicated one(s). e and f) Standard deviations were calculated from 10-fold cross-validation. "Cell" refers to the average intensity across the entire cell.

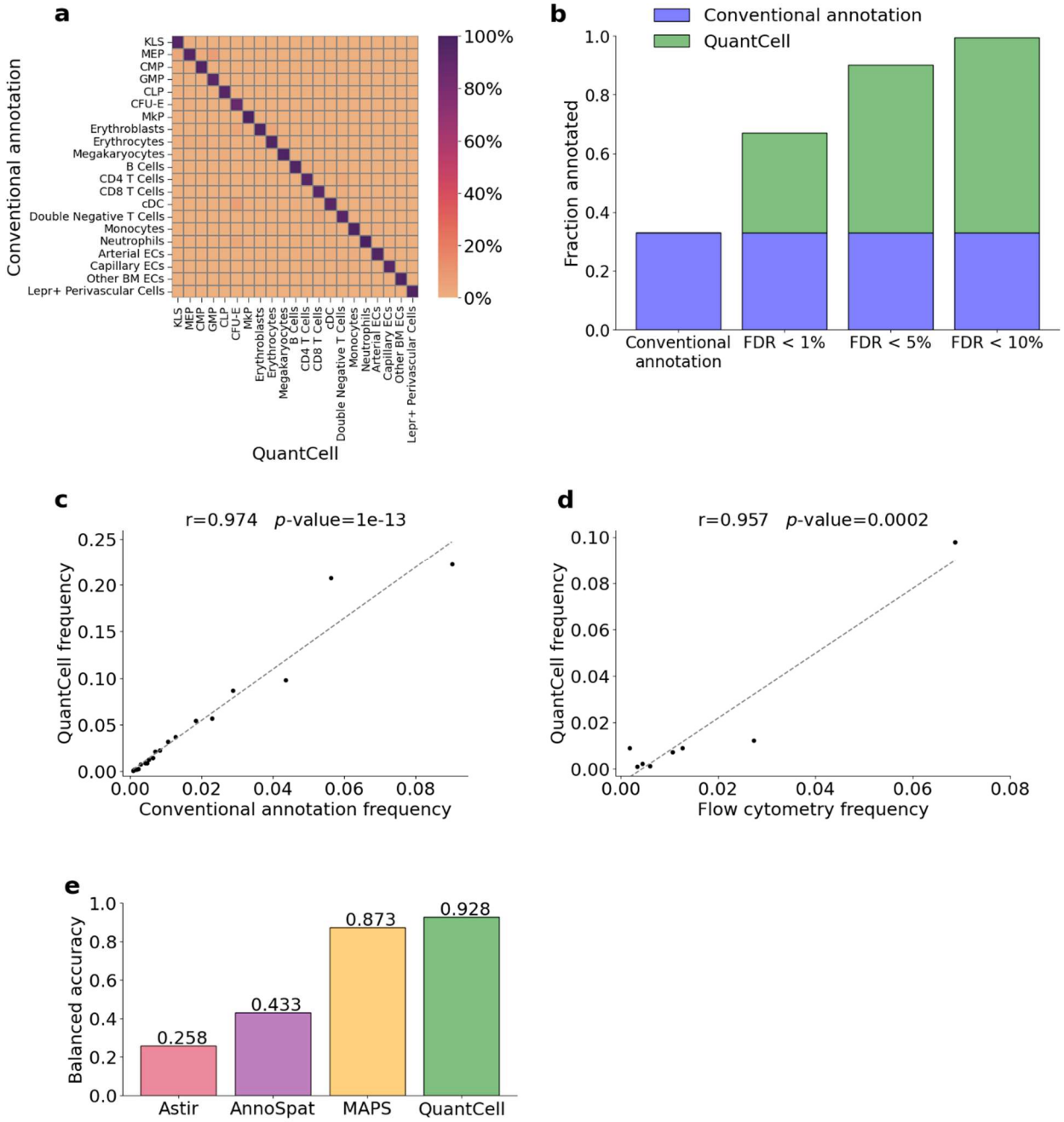



**Figure 3. Evaluation of QuantCell prediction performance.** a) Column-normalized confusion matrix comparing conventional cell annotation with QuantCell annotation. Values on the diagonal are equivalent to precision. b) Fraction of cells classified by the conventional method (blue) and those unclassified by the conventional method but successfully classified by QuantCell (green). QuantCell was configured to ensure that the false discovery rate (FDR) for all cell types are below the specified threshold. c) Comparison of cell type frequencies obtained

519 from conventional annotation and QuantCell annotation, calculated as fractions of all cells,
520 including unannotated ones. d) Comparison of cell type frequencies derived from QuantCell
521 annotations and flow cytometry analyses of the same mice. QuantCell frequencies are
522 calculated among the annotated cell types, which represent most but not all bone marrow cells,
523 whereas flow cytometry frequencies are measured relative to all alive bone marrow cells. e)
524 Benchmarking QuantCell against other spatial proteomics annotation methods. Cell type
525 abbreviations are listed in Supplementary Table 1. c and d) Two-tailed Pearson's correlation
526 coefficient (r) and *p*-value are shown. Each dot represents one cell type. The dashed line
527 represents the best-fit line computed using least squares regression.

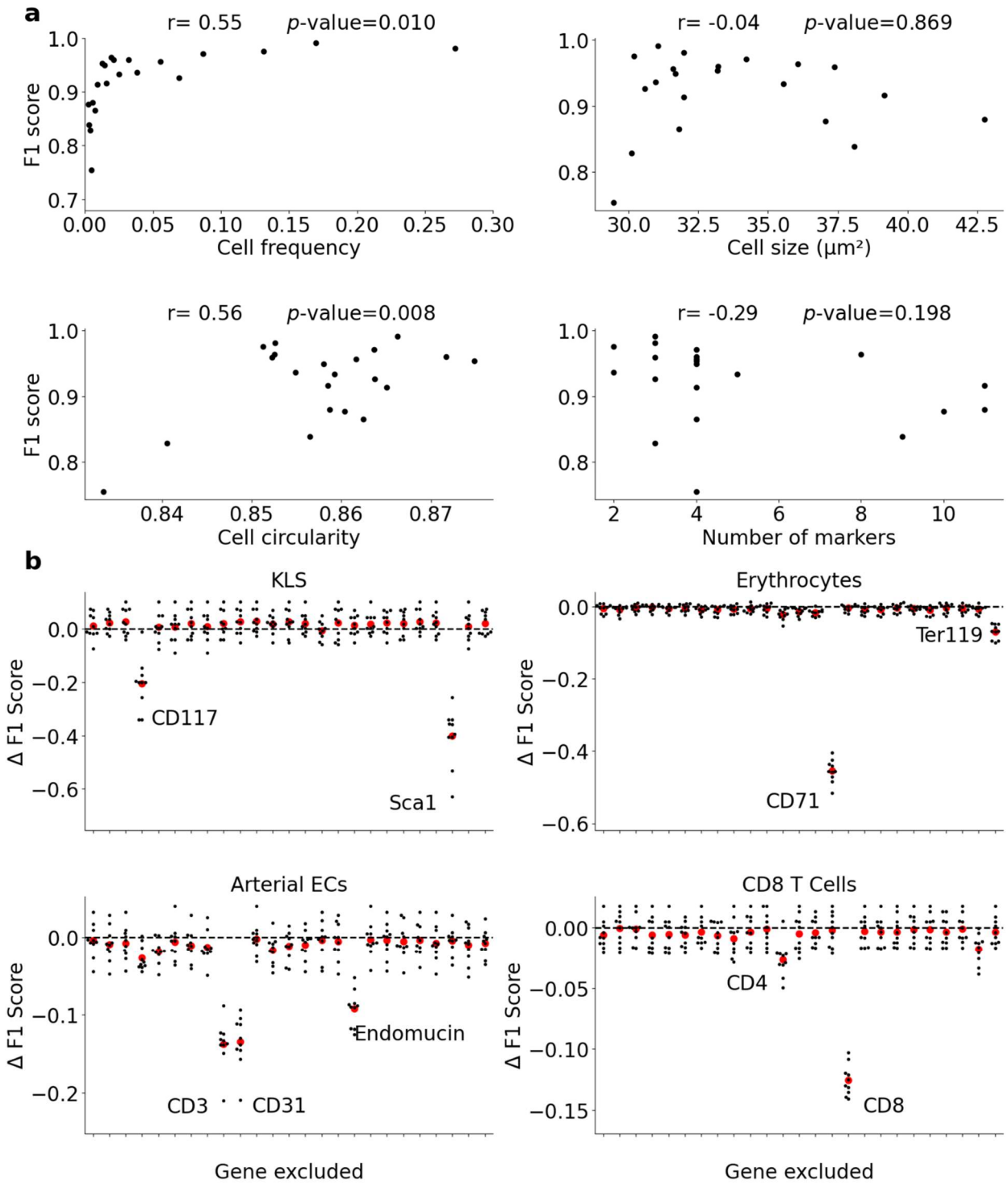



**Figure 4. QuantCell prediction performance across cell types.** a) Comparison of F1 scores without FDR thresholding based on the cell type specific attributes: cell frequency, cell size, cell circularity, and the number of positive and negative markers used in conventional annotation. Each dot represents one of 21 distinct cell types. Two-tailed Pearson's correlation coefficients and corresponding *p*-values are shown. b) Impact of individual protein markers on the prediction performance for representative cell types. Shown are changes in F1 score resulting from the

exclusion of individual protein markers from the training and testing datasets. Each dot represents the change in the F1 score relative to the baseline (no exclusion), calculated using 10-fold cross-validation. Red dots indicate the mean change across all 10 folds. Markers with a mean change exceeding 1.96 standard deviations from the overall distribution (z-score > 1.96) are annotated.

541 **Table 1. Comparison of QuantCell with existing annotation frameworks**

542

| | Method | Input data | Manual annotation required | Marker combos | Uncertainty (allow unlabeled) | Balanced label accuracy |
|---|---|---|---|---|---|---|
| **Astir**[10] | Clustering | Protein-Expression-by-cell matrix & marker combos | No | Positive | Yes | 0.258 |
| **AnnoSpat**[15] | Clustering | Protein-Expression-by-cell matrix & marker combos | No | Positive & negative | Yes | 0.433 |
| **MAPS**[19] | Feed forward neural network | Protein-Expression-by-cell matrix & cell size | Yes | No | No | 0.873 |
| **QuantCell** | Variable classifiers | Protein-Expression-by-cell matrix & marker combos | No | Positive & negative | Yes | 0.928 |

543

544 **References**


545 1. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat. Methods* **18**, 9–14
546 (2021).
547 2. Lundberg, E. & Borner, G. H. H. Spatial proteomics: a powerful discovery tool for cell
548 biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302 (2019).
549 3. Seydel, C. Beyond cell atlases: spatial biology reveals mechanisms behind disease. *Nat.*
550 *Biotechnol.* **43**, 841–844 (2025).
551 4. Cheng, X. *et al.* Application of single-cell and spatial omics in deciphering cellular
552 hallmarks of cancer drug response and resistance. *J Hematol Oncol* **18**, 70- (2025).
553 5. Stadler, C. *et al.* Immunofluorescence and fluorescent-protein tagging show high
554 correlation for protein localization in mammalian cells. *Nat. Methods* **10**, 315–323 (2013).
555 6. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed
556 Imaging. *Cell* **174**, 968-981.e15 (2018).
557 7. Merritt, C. R. *et al.* Multiplex digital spatial profiling of proteins and RNA in fixed tissue.
558 *Nat. Biotechnol.* **38**, 586–599 (2020).
559 8. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA
560 seqFISH. *Nature* **568**, 235–239 (2019).
561 9. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA
562 nanoball-patterned arrays. *Cell* **185**, 1777-1792.e21 (2022).
563 10. Geuenich, M. J. *et al.* Automated assignment of cell identity from single-cell multiplexed
564 imaging and proteomic data. *Cell Syst.* **12**, 1173-1186.e5 (2021).
565 11. Coleman, K., Hu, J., Schroeder, A., Lee, E. B. & Li, M. SpaDecon: cell-type
566 deconvolution in spatial transcriptomics with semi-supervised learning. *Commun. Biol.* **6**, 378–
567 13 (2023).
568 12. Shen, R. *et al.* Spatial-ID: a cell typing method for spatially resolved transcriptomics via
569 transfer learning and spatial embedding. *Nat. Commun.* **13**, 7640–17 (2022).
570 13. Zhou, Y., He, W., Hou, W. & Zhu, Y. Pianno: a probabilistic framework automating
571 semantic annotation for spatial transcriptomics. *Nat. Commun.* **15**, 2848–15 (2024).
572 14. Ianevski, A., Giri, A. K. & Aittokallio, T. Fully-automated and ultra-fast cell-type
573 identification using specific marker combinations from single-cell transcriptomic data. *Nat.*
574 *Commun.* **13**, 1246–1246 (2022).
575 15. Mongia, A. *et al.* AnnoSpat annotates cell types and quantifies cellular arrangements
576 from spatial proteomics. *Nat. Commun.* **15**, 3744–19 (2024).
577 16. Kiselev, V. Y., Andrews, T. S. & Hemberg, M. Challenges in unsupervised clustering of
578 single-cell RNA-seq data. *Nat. Rev. Genet.* **20**, 273–282 (2019).
579 17. Kumagai, Y. BootCellNet, a resampling-based procedure, promotes unsupervised
580 identification of cell populations via robust inference of gene regulatory networks. *PLoS Comput.*
581 *Biol.* **20**, e1012480- (2024).
582 18. Marghi, Y., Gala, R., Baftizadeh, F. & Sümbül, U. Joint inference of discrete cell types
583 and continuous type-specific variability in single-cell datasets with MMIDAS. *Nat. Comput. Sci.*
584 **4**, 706–722 (2024).
585 19. Shaban, M. *et al.* MAPS: pathologist-level cell type annotation from tissue images
586 through machine learning. *Nat. Commun.* **15**, 28–11 (2024).

587 20. Amitay, Y. *et al.* CellSighter: a neural network to classify cells in highly multiplexed
588 images. *Nat. Commun.* **14**, 4302–4302 (2023).
589 21. Nguyen, A., Khoo, W. H., Moran, I., Croucher, P. I. & Phan, T. G. Single Cell RNA
590 Sequencing of Rare Immune Cell Populations. *Front. Immunol.* **9**, 1553–1553 (2018).
591 22. Patel, S. B., Franceski, A. M., Crown, B. L. & Welner, R. S. Leading Edge Techniques in
592 the Quest for Characterizing Rare Hematopoietic Stem Cells. *Curr. Stem Cell Rep.* **10**, 108–125
593 (2024).
594 23. Altshuler, A., Wickström, S. A. & Shalom-Feuerstein, R. Spotlighting adult stem cells:
595 advances, pitfalls, and challenges. *Trends Cell Biol.* **33**, 477–494 (2023).
596 24. Abbaszadegan, M. R. *et al.* Isolation, identification, and characterization of cancer stem
597 cells: A review. *J. Cell. Physiol.* **232**, 2008–2018 (2017).
598 25. Wu, M. *et al.* Spatial proteomics: unveiling the multidimensional landscape of protein
599 localization in human diseases. *Proteome Sci.* **22**, 7–15 (2024).
600 26. Bressan, D., Battistoni, G. & Hannon, G. J. The dawn of spatial omics. *Sci. Am. Assoc.*
601 *Adv. Sci.* **381**, (2023).
602 27. Hung, M.-C. & Link, W. Protein localization in disease and therapy. *J. Cell Sci.* **124**,
603 3381–3392 (2011).
604 28. Blise, K. E., Sivagnanam, S., Banik, G. L., Coussens, L. M. & Goecks, J. Single-cell
605 spatial architectures associated with clinical outcome in head and neck squamous cell
606 carcinoma. *NPJ Precis. Oncol.* **6**, 10–10 (2022).
607 29. Han, J.-D. J. Understanding biological functions through molecular networks. *Cell Res.*
608 **18**, 224–237 (2008).
609 30. Dong, K. & Zhang, S. Deciphering spatial domains from spatially resolved
610 transcriptomics with an adaptive graph attention auto-encoder. *Nat. Commun.* **13**, 1739–12
611 (2022).
612 31. Bae, S. *et al.* CellDART: cell type inference by domain adaptation of single-cell and
613 spatial transcriptomic data. *Nucleic Acids Res.* **50**, e57–e57 (2022).
614 32. Li, B. *et al.* Benchmarking spatial and single-cell transcriptomics integration methods for
615 transcript distribution prediction and cell type deconvolution. *Nat. Methods* **19**, 662–670 (2022).
616 33. Ge, S., Sun, S., Xu, H., Cheng, Q. & Ren, Z. Deep learning in single-cell and spatial
617 transcriptomics data analysis: advances and challenges from a data science perspective. *Brief.*
618 *Bioinform.* **26**, (2025).
619 34. Gohil, V. *et al.* The Importance of Generalizability in Machine Learning for Systems.
620 *IEEE Comput. Archit. Lett.* **23**, 95–98 (2024).
621 35. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis.
622 *Sci. Rep.* **7**, 16878–7 (2017).
623 36. Loughrey, M. B. *et al.* Validation of the systematic scoring of immunohistochemically
624 stained tumour tissue microarrays using QuPath digital image analysis. *Histopathology* **73**, 327–
625 338 (2018).
626 37. Black, S. *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat.*
627 *Protoc.* **16**, 3802–3835 (2021).
628 38. Cesar Nombela-Arrieta & Markus G Manz. Quantification and three-dimensional
629 microanatomical organization of the bone marrow. *Blood Adv.* **1**, 407–416 (2017).

630 39. Baccin, C. *et al.* Combined single-cell and spatial transcriptomics reveal the molecular,
631 cellular and spatial bone marrow niche organization. *Nat. Cell Biol.* **22**, 38–48 (2020).
632 40. Dolgalev, I. & Tikhonova, A. N. Connecting the Dots: Resolving the Bone Marrow Niche
633 Heterogeneity. *Front. Cell Dev. Biol.* **9**, 622519 (2021).
634 41. Ding, L. & Morrison, S. J. Haematopoietic stem cells and early lymphoid progenitors
635 occupy distinct bone marrow niches. *Nature* **495**, 231–235 (2013).
636 42. Bryder, D., Rossi, D. J. & Weissman, I. L. Hematopoietic stem cells: the paradigmatic
637 tissue-specific stem cell. *Am. J. Pathol.* **169**, 338–346 (2006).
638 43. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of
639 mouse hematopoietic stem cells. *Science* **241**, 58–62 (1988).
640 44. Purton, L. E. Adult murine hematopoietic stem cells and progenitors: an update on their
641 identities, functions, and assays. *Exp. Hematol.* **116**, 1–14 (2022).
642 45. Wu, Y.-C., Kissner, M. & Momen-Heravi, F. A comprehensive multiparameter flow
643 cytometry panel for immune profiling and functional studies of frozen tissue, bone marrow, and
644 spleen. *J. Immunol. Methods* **515**, 113444–113444 (2023).
645 46. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
646 47. Chen, X. & Ishwaran, H. Random forests for genomic data analysis. *Genomics* **99**, 323–
647 329 (2012).
648 48. Boulesteix, A.-L., Janitza, S., Kruppa, J. & König, I. R. Overview of random forest
649 methodology and practical guidance with emphasis on computational biology and
650 bioinformatics. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2**, 493–507 (2012).
651 49. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-convex
652 Polygons. *arXiv.org* https://doi.org/10.48550/arxiv.1806.03535 (2018)
653 doi:10.48550/arxiv.1806.03535.
654 50. Weber, L. M. & Robinson, M. D. Comparison of clustering methods for high-dimensional
655 single-cell flow and mass cytometry data. *Cytom. Part J. Int. Soc. Anal. Cytol.* **89**, 1084–1096
656 (2016).
657 51. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA
658 profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
659 52. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved,
660 highly multiplexed RNA profiling in single cells. *Sci. Am. Assoc. Adv. Sci.* **348**, 412–412 (2015).
661 53. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by
662 spatial transcriptomics. *Sci. Am. Assoc. Adv. Sci.* **353**, 78–82 (2016).
663 54. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide
664 expression at high spatial resolution. *Sci. Am. Assoc. Adv. Sci.* **363**, 1463–1467 (2019).
665 55. Jiang, D. *et al.* Quantitative association between gene expression and blood cell
666 production of individual hematopoietic stem cells in mice. *Sci. Adv.* **10**, eadk2132 (2024).
667 56. Nogalska, A. *et al.* Age-associated imbalance in immune cell regeneration varies across
668 individuals and arises from a distinct subset of stem cells. *Cell. Mol. Immunol.* **21**, 1459–1473
669 (2024).
670 57. Nguyen, L. *et al.* Functional compensation between hematopoietic stem cell clones in
671 vivo. *EMBO Rep.* **19**, e45702 (2018).

58. Pachitariu, M., Rariden, M. & Stringer, C. Cellpose-SAM: superhuman generalization for cellular segmentation. 2025.04.28.651001 Preprint at https://doi.org/10.1101/2025.04.28.651001 (2025).
59. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* https://doi.org/10.5555/1953048.2078195 (2011) doi:10.5555/1953048.2078195.
60. Shaban, M. et al. Data for MAPS: Pathologist-level Cell Type Annotation from Tissue Images through Machine Learning. Zenodo https://doi.org/10.5281/zenodo.10067010 (2023).

**Acknowledgements**
We thank Akoya Biosciences for their assistance in establishing the antibody panel and acquiring imaging data, and Drs. G. Fudenberg, A. MacLean as well as the members of the Lu laboratory for their valuable discussions and insights. We also thank Drs S.W. Ruffins and I. Hajjali for assistance with validating marker and cell annotations. This study is supported by R35HL150826, R01DK143671, and R01AG080982. W.B. was supported by USC Provost's Fellowship, QCB Google Alumni Award in Quantitative Biology, and the John Petruska Memorial Award.

**Author contributions**
W.B. and R.L. conceived and designed the project. B.W. designed the antibody marker panel and cell type marker combinations, conducted all wet lab experiments including bone marrow sectioning. B.W. also performed cell segmentation and marker annotation in QuPath. W.B. analyzed the data and wrote all project-related code. B.W., Z.T., A.N., and R.L. assisted with data analysis. W.B. and R.L. wrote the manuscript. All authors reviewed and approved the final manuscript.

**Data availability**
Data supporting the findings of this study are available from the corresponding author upon request.

**Code availability**
QuantCell code is available at https://github.com/ronglulab/QuantCell. All code used for data analysis and figure generation is provided in the "paper" directory of the repository.

**Competing interests**
The authors declare no competing interests.