QuantCell: machine learning based cell annotation integrating 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 imaging of RNA and protein expression, but cell annotation remains challenging, particularly in complex tissues with numerous markers or rare cell types. Here, we present QuantCell, a machine learning framework that leverages quantitative imaging data to improve annotation derived from qualitative profiles. QuantCell evaluates multiple models and applies a user-defined false discovery rate to ensure high-confidence annotation. Using PhenoCycler imaging of mouse bone marrow, QuantCell increased annotated cells from 33.1% to 90.2% at 5% FDR, achieving

96.5% accuracy. QuantCell supports diverse imaging platforms and robustly detects rare cell populations.

Keywords: spatial omics, spatial proteomics, cell annotation, machine learning, multiplexed imaging, quantitative imaging, single-cell analysis

Background

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. We further evaluate QuantCell using ground truth data from published studies and compare its performance with existing machine learning models.

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).

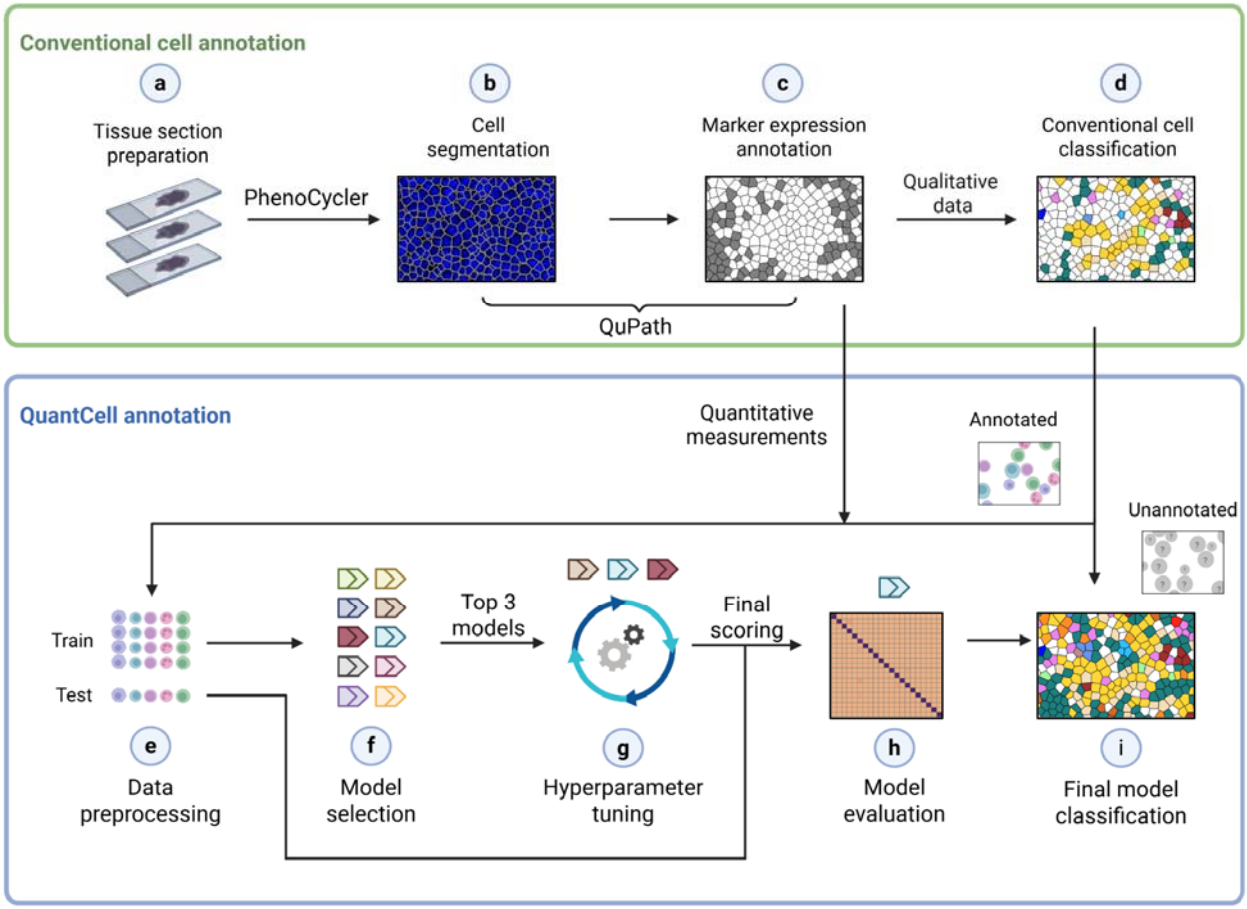

Figure 1. Overview of conventional and QuantCell annotation pipelines. a) Tissue sections are prepared, stained, and imaged. b) Cell segmentation is performed using QuPath [35], based on 4′,6-diamidino-2-phenylindole (DAPI) nuclei staining. c) Each protein marker is manually annotated as positive or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

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.

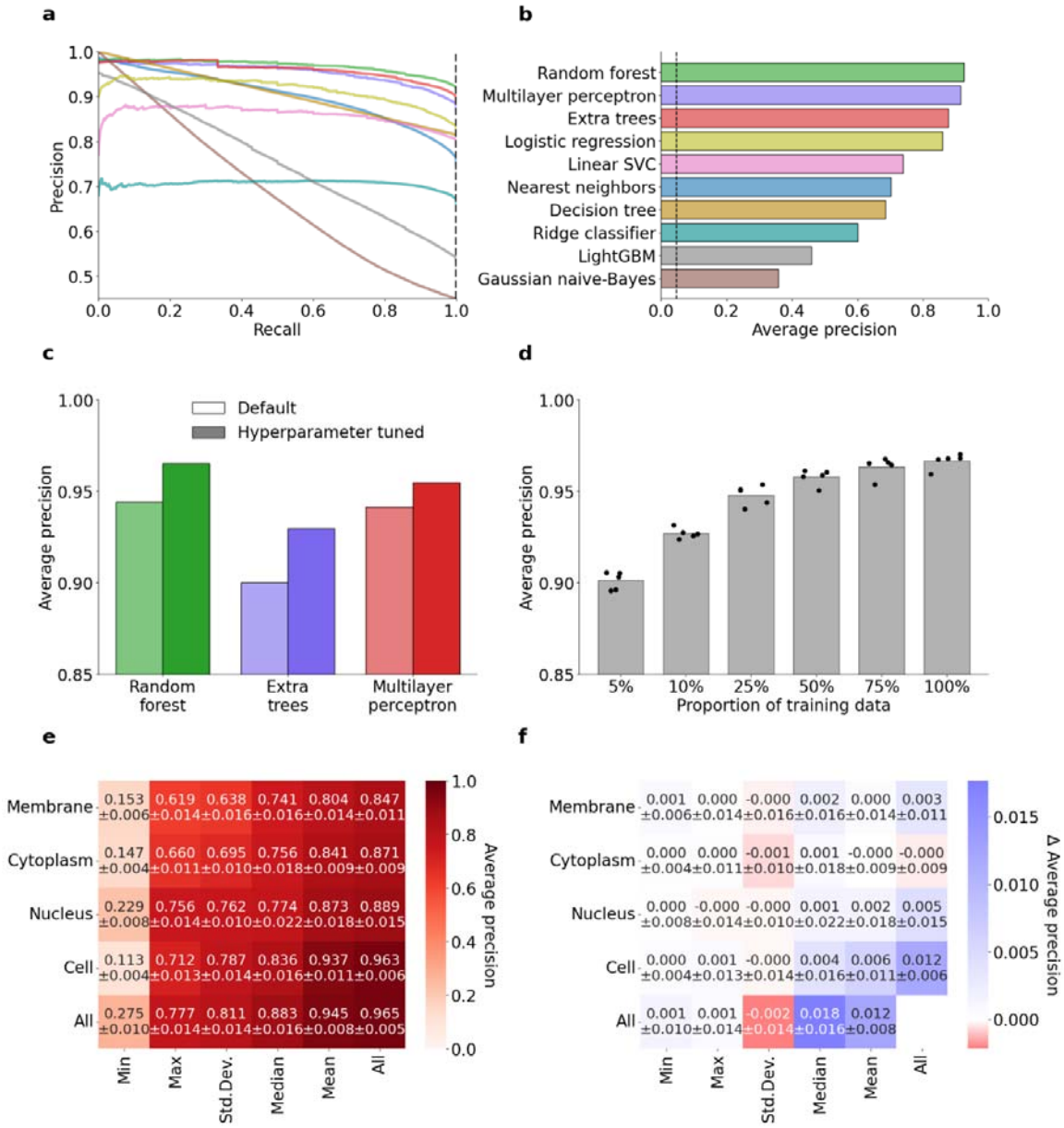

Figure 2. Selection and optimization of machine learning models for cell classification. a) Precision-recall curve for base classifiers without hyperparameter tuning. Data represent the unweighted average of all cell types. b) Macro average precision (AP) score for each classifier, equivalent to the 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.

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 using the corresponding conventional annotation markers (Supplementary Fig. 4). These PCA plots facilitate model selection and parameter tuning by enabling visualization of the overlap between conventionally annotated cells and QuantCell-annotated cells, as well as their separation from cells not assigned to that cell type.

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. 5). 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-based [49] 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.

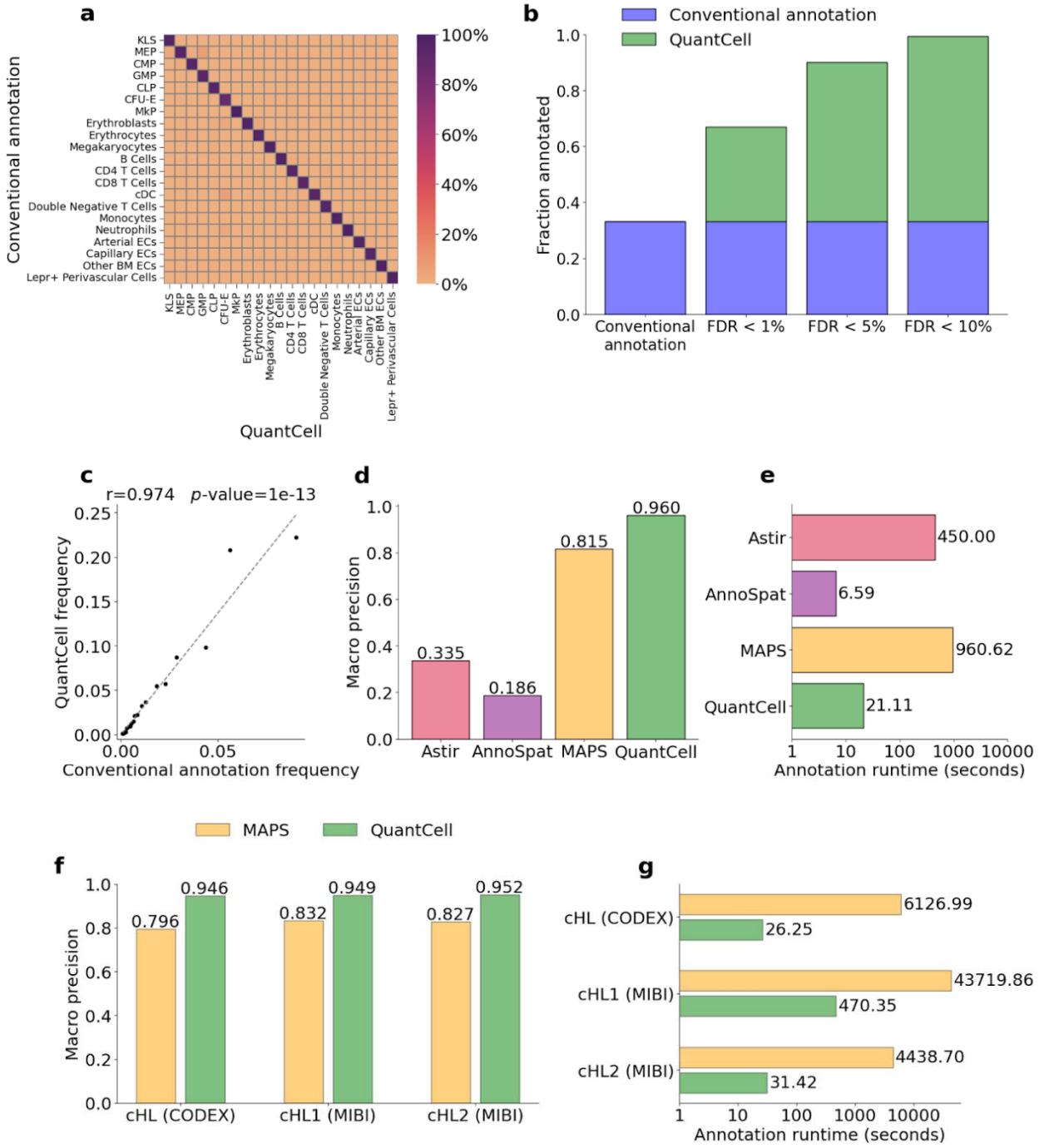

Figure 3. Evaluation of QuantCell prediction performance. a) Column-normalized confusion matrix comparing conventional cell annotation with QuantCell annotation. Values on the diagonal are equivalent to precision. b) Fraction of cells classified by the conventional method (blue) and those unclassified 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 from conventional annotation and QuantCell annotation, calculated as fractions of all cells, including unannotated ones. Two-tailed Pearson's correlation coefficient (r) and *p*-value are shown. Each dot represents one cell type. The dashed line represents the best-fit line computed

using least squares regression. d) Benchmarking QuantCell against other spatial proteomics annotation methods. e) Runtime required to train models and annotate unknown cells for QuantCell and the spatial proteomics annotation methods compared in panel d. f) Benchmarking QuantCell against MAPS using three independent datasets with previously established ground truth annotations [50]. g) Runtime required to train models and annotate testing dataset for QuantCell and MAPS, corresponding to panel f.

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 underscoring the model's robustness even in unbalanced cell populations (Fig. 3a, 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. 8).

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].

Table 1. Comparison of QuantCell with existing annotation frameworks

	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

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 precision than existing models (Fig. 3d, Supplementary Fig. 9) and operates 45 times faster than the second most accurate model, MAPS (Fig. 3e). Using published dataset and ground truth annotation [50], we further demonstrate that QuantCell achieves higher precision and massive decreases in runtime (Fig. 3f-g). We used macro precision, the unweighted average precision across all cell types, as the primary metric for comparison, because it reduces bias toward abundant cell types, ensuring that rare populations contribute equally to the evaluation. Precision was chosen given the importance of limiting false positive classifications. QuantCell also demonstrates robust performance across all cell types (Fig. 3a, 3c) 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, 3d). 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 [51]. 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 than in all other cells. This assumption is not valid for markers shared by multiple cell types, which is common among stem/progenitor cells and diseased cells.

Cell type specific prediction performance

A major strength of QuantCell is its strong performance across cell types with widely varying frequencies. We further evaluated QuantCell's performance across individual cell types using the F1 score, the harmonic mean of precision and recall which summarizes predictive

performance for each cell type. Our results suggested that the F1 score was significantly influenced by cell frequency and cell circularity (Fig. 4a), likely due to its effect on the sample size and quality for model training. In contrast, features such as cell size and the number of protein markers used in conventional annotation did not show direct correlation with F1 score (Fig. 4a), indicating the robustness and consistency of QuantCell's performance across cell types.

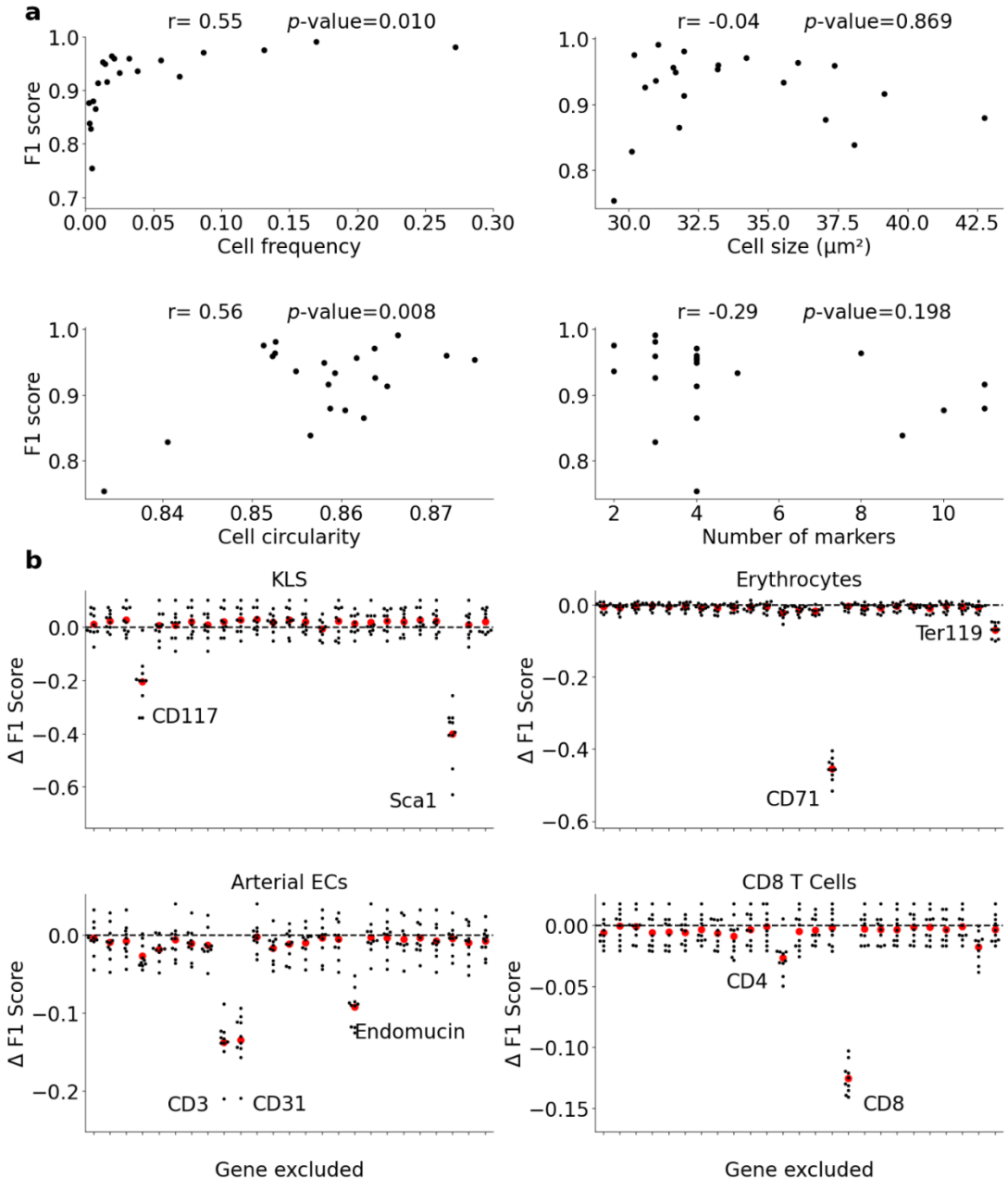

Figure 4. QuantCell prediction performance across cell types. a) Comparison of F1 scores without FDR thresholding based on the cell type specific attributes: cell frequency, cell size, cell circularity, and the number of positive and negative markers used in conventional annotation. Each dot represents 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.

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

Discussion

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

QuantCell outperforms existing methods in predictive accuracy and offers significant advancements in the following key areas. First, QuantCell optimizes annotation performance across all cell types (Fig. 3a, Supplementary Fig. 7). By avoiding performance bias toward abundant cell types (Supplementary Fig. 9), it provides a more balanced and accurate classification. This is particularly important in biomedical datasets where rare cell types, such as stem cells or diseased cells, are 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). 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. 3d-f). 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 [52] and MERFISH [53], and sequencing-based approaches like Visium [54], Slide-seq [55], 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.

Conclusion

In summary, QuantCell provides a robust and versatile machine learning framework for cell annotation in spatial omics datasets. By integrating qualitative marker definitions with quantitative imaging features and applying a user-defined false discovery rate to control classification confidence, QuantCell substantially increases the number of annotated cells while maintaining high accuracy. It outperforms existing annotation methods in predictive accuracy, minimizes performance bias toward abundant cell types, and enables 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 impractical for rare cell populations. QuantCell is compatible with a wide range of spatial omics platforms and can be tailored to project-specific priorities. Its flexibility and precision enable accurate characterization of complex cellular phenotypes, including those with highly imbalanced cell populations.

Methods

Mice

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

PhenoCycler antibodies

The following primary antibodies were obtained from Akoya Biosciences: RX025-AF488 anti-CD11b (M1/70), RX030-Cy5 anti-CD11c (N418), RX020-ATTO550 anti-CD19 (6D5), RX021-Cy5 anti-CD3 (17A2), RX002-ATTO550 anti-CD31 (MEC13.3), RX026-ATTO550 anti-CD4 (RM4-5), RX005-ATTO550 anti-CD44 (IM7), RX007-AF488 anti-CD45 (30-F11), RX010-AF488

anti-CD45R (RA3-6B2), RX027-Cy5 anti-CD71 (RI7217), RX029-ATTO550 anti-CD8a (53-6.7), RX024-Cy5 anti-Ly6g (1A8), RX014-ATTO550 anti-MHCII (M5/114.15), RX003-Cy5 anti-TCR□β (H57-597), and RX004-AF488 anti-TER119 (TER-119). The following primary antibodies were obtained from ThermoFisher Scientific: BX035-ATTO550 anti-CD105 (209701), BX050-AF647 anti-CD115 (AFS98), BX045-Cy5 anti-CD117 (2B8), RX013-AF488 anti-CD34 (RAM34), BX037-ATTO550 anti-CD41 (MWReg30), BX040-AF488 anti-endomucin (polyclonal), BX036-Cy5 anti-F4/80 (CI:A3-1), and BX041-ATTO550 anti-CD127 (A7R34). The following primary antibody was obtained from Leinco Technologies: BX028-AF488 anti-CD32/CD16 (2.4G2). The following primary antibody was obtained from R&D Systems: BX049-ATTO550 anti-LEPr (polyclonal). All primary antibodies obtained from vendors other than Akoya Biosciences were conjugated to Akoya barcodes by Akoya Biosciences.

Antibody validation and quality control

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

Bone marrow preparation for imaging

Bone marrow was carefully flushed from the diaphysis of femurs and tibias using a syringe filled with DPBS with 2% FBS, ensuring gentle handling to obtain the whole, intact bone marrow

plugs. The collected bone marrow was immediately fresh-frozen in tissue freezing medium (Electron Microscopy Sciences, Cat. #72592) for downstream applications.

Bone marrow imaging

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

Cell segmentation

Nuclei-based cell segmentation was performed based on DAPI stains in QuPath [35] (version 0.3.0) using the StarDist [49] (version 0.3.0) algorithm, following default settings (provided in our GitHub, see Code availability). For nuclei-based segmentation, QuPath automatically infers cytoplasm and membrane compartments by expanding nuclear boundaries using a watershed approach. Membrane-based segmentation was performed with Cellpose-SAM [56] (version 4.0.6) using a composite image of DAPI, F4/80, CD45, Endomucin, Sca1, and CD71. To recover cellular compartments using Cellpose, segmentation was run independently on the composite (whole cell) and the DAPI channel (nucleus). The cytoplasmic compartment was defined by subtracting the nuclear area from the whole-cell area. QuantCell performed comparably well with either membrane-based or nuclei-based cell segmentation (Supplementary Fig. 3). Nuclei-based segmentation was used for all other analyses.

Conventional annotation and QuPath settings

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

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

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

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

Data preprocessing

Cells that were successfully annotated by the conventional method were split into an 80/20 train-test dataset using train_test_split() from scikit-learn [57], 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 (μm^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**.**

Benchmarking on external datasets

We benchmarked QuantCell against MAPS using three classical Hodgkin lymphoma (cHL) datasets from the original MAPS study [19,50]: two generated using multiplexed ion beam imaging (MIBI) and one using CODEX (now referred to as PhenoCycler). In these datasets, ground truth annotations were established through cluster-based annotation followed by manual curation. Because marker-combination–based annotation strategies were not available, Astir and AnnoSpat could not be evaluated on these datasets.

QuantCell was assessed directly against the provided ground truth annotations. To ensure a fair comparison, QuantCell was configured using the same 80/20 train–test split as MAPS, and performance was evaluated at a 5% false discovery rate (FDR) threshold. MAPS was re-run in the same computational environment rather than using the performance metrics reported in the original publication (see Computational environment). Reported runtime includes both model training and prediction.

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. 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.

Declarations

Ethics approval and consent to participate

All animal procedures were approved by the Institutional Animal Care and Use Committee of USC.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this study are available from the corresponding author upon request. 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 that they have no competing interests.

Funding

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.

Authors' 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.

Acknowledgments

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.

References

1. Marx V. Method of the Year: spatially resolved transcriptomics. Nat Methods. 2021;18:9–14. https://doi.org/10.1038/s41592-020-01033-y

2. Lundberg E, Borner GHH. Spatial proteomics: a powerful discovery tool for cell biology. Nat Rev Mol Cell Biol. Nature Publishing Group; 2019;20:285–302. https://doi.org/10.1038/s41580-018-0094-y

3. Seydel C. Beyond cell atlases: spatial biology reveals mechanisms behind disease. Nat Biotechnol. Nature Publishing Group; 2025;43:841–4. https://doi.org/10.1038/s41587-025-02699-5

4. Cheng X, Peng T, Chu T, Yang Y, Liu J, Gao Q, et al. Application of single-cell and spatial omics in deciphering cellular hallmarks of cancer drug response and resistance. J Hematol OncolJ Hematol Oncol. England: BioMed Central; 2025;18:70-. https://doi.org/10.1186/s13045-025-01722-1

5. Stadler C, Rexhepaj E, Singan VR, Murphy RF, Pepperkok R, Uhlén M, et al. Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. Nat Methods. Nature Publishing Group; 2013;10:315–23. https://doi.org/10.1038/nmeth.2377

6. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, et al. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. Cell. 2018;174:968-981.e15. https://doi.org/10.1016/j.cell.2018.07.010

7. Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G, et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. Nat Biotechnol. New York: Nature Publishing Group US; 2020;38:586–99. https://doi.org/10.1038/s41587-020-0472-9

8. Eng C-HL, Lawson M, Zhu Q, Dries R, Koulena N, Takei Y, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. Nature. 2019;568:235–9. https://doi.org/10.1038/s41586-019-1049-y

9. Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. Cell. United States: Elsevier Inc; 2022;185:1777-1792.e21. https://doi.org/10.1016/j.cell.2022.04.003

10. Geuenich MJ, Hou J, Lee S, Ayub S, Jackson HW, Campbell KR. Automated assignment of cell identity from single-cell multiplexed imaging and proteomic data. Cell Syst. United States: Elsevier Inc; 2021;12:1173-1186.e5. https://doi.org/10.1016/j.cels.2021.08.012

11. Coleman K, Hu J, Schroeder A, Lee EB, Li M. SpaDecon: cell-type deconvolution in spatial transcriptomics with semi-supervised learning. Commun Biol. London: Nature Publishing Group UK; 2023;6:378–13. https://doi.org/10.1038/s42003-023-04761-x

12. Shen R, Liu L, Wu Z, Zhang Y, Yuan Z, Guo J, et al. Spatial-ID: a cell typing method for spatially resolved transcriptomics via transfer learning and spatial embedding. Nat Commun. London: Nature Publishing Group UK; 2022;13:7640–17. https://doi.org/10.1038/s41467-022-35288-0

13. Zhou Y, He W, Hou W, Zhu Y. Pianno: a probabilistic framework automating semantic annotation for spatial transcriptomics. Nat Commun. London: Nature Publishing Group UK; 2024;15:2848–15. https://doi.org/10.1038/s41467-024-47152-4

14. Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nat Commun. London: Nature Publishing Group UK; 2022;13:1246–1246. https://doi.org/10.1038/s41467-022-28803-w
15. Mongia A, Zohora FT, Burget NG, Zhou Y, Saunders DC, Wang YJ, et al. AnnoSpat annotates cell types and quantifies cellular arrangements from spatial proteomics. Nat Commun. London: Nature Publishing Group UK; 2024;15:3744–19. https://doi.org/10.1038/s41467-024-47334-0
16. Kiselev VY, Andrews TS, Hemberg M. Challenges in unsupervised clustering of single-cell RNA-seq data. Nat Rev Genet. London: Nature Publishing Group UK; 2019;20:273–82. https://doi.org/10.1038/s41576-018-0088-9
17. Kumagai Y. BootCellNet, a resampling-based procedure, promotes unsupervised identification of cell populations via robust inference of gene regulatory networks. PLoS Comput Biol. United States: Public Library of Science; 2024;20:e1012480-. https://doi.org/10.1371/journal.pcbi.1012480
18. Marghi Y, Gala R, Baftizadeh F, Sümbül U. Joint inference of discrete cell types and continuous type-specific variability in single-cell datasets with MMIDAS. Nat Comput Sci. New York: Nature Publishing Group US; 2024;4:706–22. https://doi.org/10.1038/s43588-024-00683-8
19. Shaban M, Bai Y, Qiu H, Mao S, Yeung J, Yeo YY, et al. MAPS: pathologist-level cell type annotation from tissue images through machine learning. Nat Commun. London: Nature Publishing Group UK; 2024;15:28–11. https://doi.org/10.1038/s41467-023-44188-w
20. Amitay Y, Bussi Y, Feinstein B, Bagon S, Milo I, Keren L. CellSighter: a neural network to classify cells in highly multiplexed images. Nat Commun. London: Nature Publishing Group UK; 2023;14:4302–4302. https://doi.org/10.1038/s41467-023-40066-7
21. Nguyen A, Khoo WH, Moran I, Croucher PI, Phan TG. Single Cell RNA Sequencing of Rare Immune Cell Populations. Front Immunol. Switzerland: Frontiers Media S.A; 2018;9:1553–1553. https://doi.org/10.3389/fimmu.2018.01553
22. Patel SB, Franceski AM, Crown BL, Welner RS. Leading Edge Techniques in the Quest for Characterizing Rare Hematopoietic Stem Cells. Curr Stem Cell Rep. Cham: Springer International Publishing; 2024;10:108–25. https://doi.org/10.1007/s40778-024-00240-z
23. Altshuler A, Wickström SA, Shalom-Feuerstein R. Spotlighting adult stem cells: advances, pitfalls, and challenges. Trends Cell Biol. England: Elsevier Ltd; 2023;33:477–94. https://doi.org/10.1016/j.tcb.2022.09.007
24. Abbaszadegan MR, Bagheri V, Razavi MS, Momtazi AA, Sahebkar A, Gholamin M. Isolation, identification, and characterization of cancer stem cells: A review. J Cell Physiol. United States: Wiley Subscription Services, Inc; 2017;232:2008–18. https://doi.org/10.1002/jcp.25759
25. Wu M, Tao H, Xu T, Zheng X, Wen C, Wang G, et al. Spatial proteomics: unveiling the multidimensional landscape of protein localization in human diseases. Proteome Sci. England: BioMed Central Ltd; 2024;22:7–15. https://doi.org/10.1186/s12953-024-00231-2
26. Bressan D, Battistoni G, Hannon GJ. The dawn of spatial omics. Sci Am Assoc Adv Sci. United States: The American Association for the Advancement of Science; 2023;381. https://doi.org/10.1126/science.abq4964
27. Hung M-C, Link W. Protein localization in disease and therapy. J Cell Sci. England; 2011;124:3381–92. https://doi.org/10.1242/jcs.089110

28. Blise KE, Sivagnanam S, Banik GL, Coussens LM, Goecks J. Single-cell spatial architectures associated with clinical outcome in head and neck squamous cell carcinoma. NPJ Precis Oncol. London: Nature Publishing Group UK; 2022;6:10–10. https://doi.org/10.1038/s41698-022-00253-z

29. Han J-DJ. Understanding biological functions through molecular networks. Cell Res. London: Nature Publishing Group UK; 2008;18:224–37. https://doi.org/10.1038/cr.2008.16

30. Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nat Commun. London: Nature Publishing Group UK; 2022;13:1739–12. https://doi.org/10.1038/s41467-022-29439-6

31. Bae S, Na KJ, Koh J, Lee DS, Choi H, Kim YT. CellDART: cell type inference by domain adaptation of single-cell and spatial transcriptomic data. Nucleic Acids Res. England: Oxford University Press; 2022;50:e57–e57. https://doi.org/10.1093/nar/gkac084

32. Li B, Zhang W, Guo C, Xu H, Li L, Fang M, et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. Nat Methods. New York: Nature Publishing Group US; 2022;19:662–70. https://doi.org/10.1038/s41592-022-01480-9

33. Ge S, Sun S, Xu H, Cheng Q, Ren Z. Deep learning in single-cell and spatial transcriptomics data analysis: advances and challenges from a data science perspective. Brief Bioinform. England; 2025;26. https://doi.org/10.1093/bib/bbaf136

34. Gohil V, Dev S, Upasani G, Lo D, Ranganathan P, Delimitrou C. The Importance of Generalizability in Machine Learning for Systems. IEEE Comput Archit Lett. New York: IEEE; 2024;23:95–8. https://doi.org/10.1109/LCA.2024.3384449

35. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open source software for digital pathology image analysis. Sci Rep. London: Nature Publishing Group UK; 2017;7:16878–7. https://doi.org/10.1038/s41598-017-17204-5

36. Loughrey MB, Bankhead P, Coleman HG, Hagan RS, Craig S, McCorry AMB, et al. Validation of the systematic scoring of immunohistochemically stained tumour tissue microarrays using QuPath digital image analysis. Histopathology. England: Wiley Subscription Services, Inc; 2018;73:327–38. https://doi.org/10.1111/his.13516

37. Black S, Phillips D, Hickey JW, Kennedy-Darling J, Venkataraaman VG, Samusik N, et al. CODEX multiplexed tissue imaging with DNA-conjugated antibodies. Nat Protoc. 2021;16:3802–35. https://doi.org/10.1038/s41596-021-00556-8

38. Cesar Nombela-Arrieta, Markus G Manz. Quantification and three-dimensional microanatomical organization of the bone marrow. Blood Adv. Elsevier; 2017;1:407–16.

39. Baccin C, Al-Sabah J, Velten L, Helbling PM, Grünschläger F, Hernández-Malmierca P, et al. Combined single-cell and spatial transcriptomics reveal the molecular, cellular and spatial bone marrow niche organization. Nat Cell Biol. 2020;22:38–48. https://doi.org/10.1038/s41556-019-0439-6

40. Dolgalev I, Tikhonova AN. Connecting the Dots: Resolving the Bone Marrow Niche Heterogeneity. Front Cell Dev Biol. Switzerland; 2021;9:622519. https://doi.org/10.3389/fcell.2021.622519

41. Ding L, Morrison SJ. Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. Nature. 2013;495:231–5. https://doi.org/10.1038/nature11885

42. Bryder D, Rossi DJ, Weissman IL. Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. Am J Pathol. 2006;169:338–46. https://doi.org/10.2353/ajpath.2006.060312

43. Spangrude GJ, Heimfeld S, Weissman IL. Purification and characterization of mouse hematopoietic stem cells. Science. 1988;241:58–62.

44. Purton LE. Adult murine hematopoietic stem cells and progenitors: an update on their identities, functions, and assays. Exp Hematol. 2022;116:1–14. https://doi.org/10.1016/j.exphem.2022.10.005

45. Wu Y-C, Kissner M, Momen-Heravi F. A comprehensive multiparameter flow cytometry panel for immune profiling and functional studies of frozen tissue, bone marrow, and spleen. J Immunol Methods. Netherlands: Elsevier B.V; 2023;515:113444–113444. https://doi.org/10.1016/j.jim.2023.113444

46. Breiman L. Random Forests. Mach Learn. Dordrecht: Springer Nature B.V; 2001;45:5–32. https://doi.org/10.1023/A:1010933404324

47. Chen X, Ishwaran H. Random forests for genomic data analysis. Genomics. 2012;99:323–9. https://doi.org/10.1016/j.ygeno.2012.04.003

48. Boulesteix A-L, Janitza S, Kruppa J, König IR. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdiscip Rev Data Min Knowl Discov. Hoboken, USA: John Wiley & Sons, Inc; 2012;2:493–507. https://doi.org/10.1002/widm.1072

49. Schmidt U, Weigert M, Broaddus C, Myers G. Cell Detection with Star-Convex Polygons. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. Med Image Comput Comput Assist Interv – MICCAI 2018. Cham: Springer International Publishing; 2018. p. 265–73. https://doi.org/10.1007/978-3-030-00934-2_30

50. Shaban M, Bai Y, Qiu H, Mao S, Yeung J, Yeo YY, et al. Data for MAPS: Pathologist-level Cell Type Annotation from Tissue Images through Machine Learning. Zenodo; 2023. https://doi.org/10.5281/zenodo.10067010

51. Weber LM, Robinson MD. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. Cytom Part J Int Soc Anal Cytol. 2016;89:1084–96. https://doi.org/10.1002/cyto.a.23030

52. Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L. Single-cell in situ RNA profiling by sequential hybridization. Nat Methods. New York: Nature Publishing Group US; 2014;11:360–1. https://doi.org/10.1038/nmeth.2892

53. Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. Spatially resolved, highly multiplexed RNA profiling in single cells. Sci Am Assoc Adv Sci. Washington: American Association for the Advancement of Science; 2015;348:412–412. https://doi.org/10.1126/science.aaa6090

54. Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Sci Am Assoc Adv Sci. United States: American Association for the Advancement of Science; 2016;353:78–82. https://doi.org/10.1126/science.aaf2403

55. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. Sci Am Assoc Adv Sci. United States: American Association for the Advancement of Science; 2019;363:1463–7. https://doi.org/10.1126/science.aaw1219

56. Pachitariu M, Rariden M, Stringer C. Cellpose-SAM: superhuman generalization for cellular segmentation. bioRxiv; 2025. p. 2025.04.28.651001. https://doi.org/10.1101/2025.04.28.651001
57. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. Microtome Publishing; 2011; https://doi.org/10.5555/1953048.2078195