

AWARD NUMBER: W81XWH-19-1-0642

TITLE: Quantifying In Situ Adaptive Immunity in Human Tubulointerstitial Lupus Nephritis

PRINCIPAL INVESTIGATOR: Marcus Clark

CONTRACTING ORGANIZATION: The University of Chicago, Chicago, IL

REPORT DATE: December 2022

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE December 2022		2. REPORT TYPE Final		3. DATES COVERED 15Sep2019-14Sep2022	
4. TITLE AND SUBTITLE Quantifying In Situ Adaptive Immunity in Human Tubulointerstitial Lupus Nephritis				5a. CONTRACT NUMBER W81XWH-19-1-0642	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Marcus Clark E-Mail: mclark@medicine.bsd.uchicago.edu				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Chicago 5801 S Ellis Ave Chicago, IL 60637				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Development Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The specific goals of this grant are to use high-dimensional (16 primary antibodies or more) multicolor confocal microscopy to describe the frequency and spatial distributions of the different types of immune cells infiltrating the lupus kidney (Aim 1). We will then use machine learning approaches, including CDM3, to quantify the true complexity of inflammation and identify functional relationships between different T cell populations and the cells that are presenting antigen, and therefore activating, pathogenic T cells (Aim 2). By understanding which cells, functional relationships and mechanisms are associated with progression to renal failure, we will both identify new therapeutic targets and the patients for which such therapies are likely to be efficacious.					
15. SUBJECT TERMS Lupus, lupus nephritis, tubulointerstitial inflammation, adaptive immunity, B cells, human, confocal microscopy, deep machine learning					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Unclassified	18. NUMBER OF PAGES 43	19a. NAME OF RESPONSIBLE PERSON USAMRDC
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code)

TABLE OF CONTENTS

	<u>Page</u>
1. Introduction	4
2. Keywords	4
3. Accomplishments	4
4. Impact	8
5. Changes/Problems	9
6. Products	11
7. Participants & Other Collaborating Organizations	13
8. Special Reporting Requirements	15
9. Appendices	15

1. INTRODUCTION:

The specific goals of this grant are to use high-dimensional (16 primary antibodies or more) multicolor confocal microscopy to describe the frequency and spatial distributions of the different types of immune cells infiltrating the lupus kidney (Aim 1). We will then use machine learning approaches, including CDM3, to quantify the true complexity of inflammation and identify functional relationships between different T cell populations and the cells that are presenting antigen, and therefore activating, pathogenic T cells (Aim 2).

2. KEYWORDS:

Lupus, lupus nephritis, tubulointerstitial inflammation, adaptive immunity, B cells, human, confocal microscopy, deep machine learning

3. ACCOMPLISHMENTS:

What were the major goals of the project?

Specific Aim 1. Determine the frequencies and global distributions of prevalent immune cells in lupus nephritis.	Timeline	Site 1	Date Completed	Percentage of Completion
Major Task 1: Determine the frequency of immune cell populations identified in the AMP.	Months			
Validate primary antibodies	1-3	Dr. Clark	10/2019	100%
Develop multiplex staining panels	2-4	Dr. Clark	11/2019	100%
Stain lupus kidney samples	4-16	Dr. Clark	3/2020	100%
Milestone Achieved: Finish staining of 30 lupus biopsies	16	Dr. Clark	3/2020	100%
Implement supervised and non-supervised segmentation protocols	3-16	Dr. Giger	6/2021	100%
Final machine learning analysis	18-24	Drs. Clark and Giger	12/2021	100%
Milestone Achieved: Finish analysis	24	Drs. Clark and Giger	2/2022	100%

Local IRB/IACUC Approval	done	yes	6/2018	100%
Major Task 2: Determine the spatial distributions of main immune cell populations.				
Apply supervised and non-supervised segmentation protocols for cluster analysis	16-24	Dr. Giger	6/2021	100%
Milestone(s) Achieved: Finish cluster analysis	24	Drs. Clark and Giger	2/2022	100%
Specific Aim 2: Quantify the major adaptive cell networks in lupus TII.				
Major Task 3: Quantify the major adaptive cell networks in lupus TII.				
Staining 30 kidney biopsies and imaging for CDM ³	6-24 mo	Dr. Clark	3/2020	100%
Training DCNN	8-28 mo	Dr. Giger	6/2021	100%
Analysis of DCNN output with TNN	24-36	Drs. Clark and Giger	12/2021	100%
Milestone(s) Achieved: Finish project	36	Drs. Clark and Giger	2/2022	100%

Aim 1: We hypothesize that the relative frequencies and distributions of the main cellular components of immunity (ex., pDCs, T cell subsets, NK cells) in TII will define different mechanistic subsets. We postulate that the cellular mechanisms driving TII will vary between patients for a given degree of TII.

Aim 2: We hypothesize that different adaptive cell networks will characterize different degrees of lupus TII. We postulate that severe TII will be associated with additional and more complex adaptive cell mechanisms.

What was accomplished under these goals?

The major findings to date associated with the project, for Aims 1 and 2, have been published in the Journal of Clinical Investigation:

Abraham R*, Durkee MS, Ai J, Veselits M, Casella G, Asano Y, Chang A, Ko K, Oshinsky C, Peninger E, Giger ML, Clark MR. 2022. Specific *in situ* inflammatory states associate with progression to renal failure in lupus nephritis. *J Clinical Investigation*. 132:e155350. PMID: 35608910. Commentary: Li H, Tsokos MG and Tsokos GC. 2022. Lymphocytes in the neighborhood: good or bad for the kidney? 132:e160657 PMID: 35775489.

Abstract: *Background:* In human lupus nephritis (LN), tubulointerstitial inflammation (TII) on biopsy predicts progression to end-stage renal disease (ESRD). However, only about half of patients with moderate/severe TII develop ESRD. We hypothesized that this heterogeneity in outcome reflects different underlying inflammatory states. Therefore, we interrogated renal biopsies from LN longitudinal and cross-sectional cohorts.

Methods: Data was acquired using conventional and highly multiplexed confocal microscopy. To accurately segment cells across whole biopsies, and to understand their spatial relationships, we developed computational pipelines by training and implementing several deep learning models and other computer vision techniques.

Results: High B cell densities were associated with protection from ESRD. In contrast, CD8, and other CD4-CD8- T cells, were associated with both acute renal failure and progression to ESRD. B cells were often organized into large periglomerular neighborhoods with T follicular helper cells while CD4- T cells formed small neighborhoods in the tubulointerstitium whose frequency predicted progression to ESRD. *Conclusions:* These data reveal that specific *in situ* inflammatory states are associated with refractory and progressive renal disease.

These findings demonstrated that different, clinically meaningful inflammatory states can be defined by quantifying the *in situ* immune state. Performing this study required us to develop, apply and train novel machine learning tools capable of identifying and accurately segmenting complex cell populations within the renal environment.

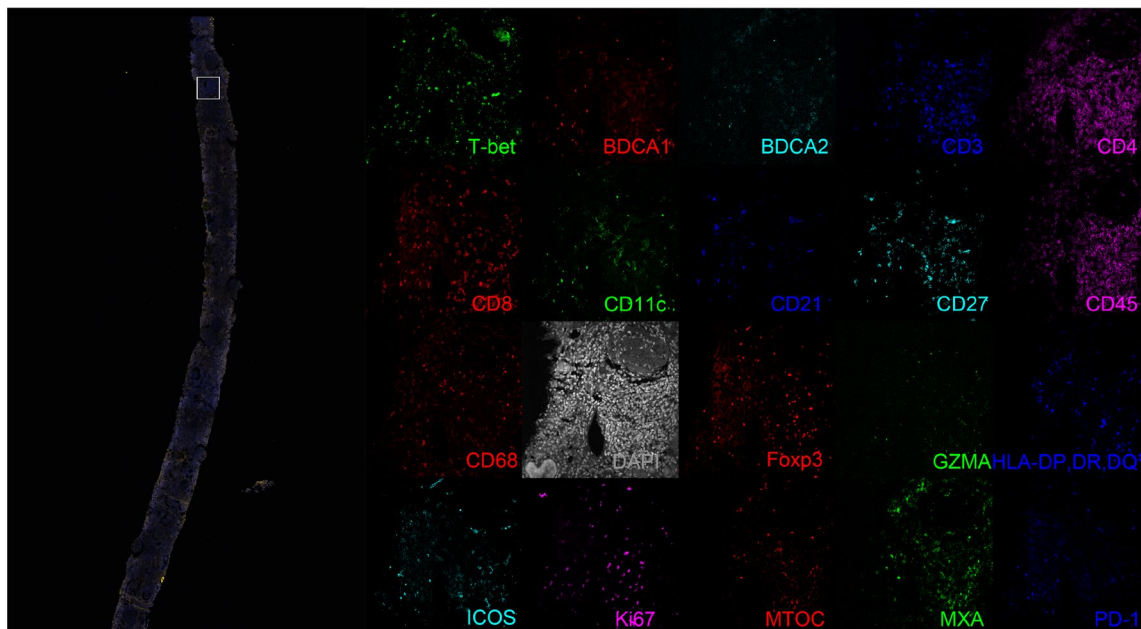


Figure A. 20 color CODEX/Confocal scan of full renal biopsy from lupus nephritis biopsy.

These studies also provide the rationale for a more extensive interrogation of renal biopsies in a way that captures the full complexity of the inflammatory infiltrate and puts it into the histological context of kidney substructures. As described in the previous annual report, we are using CODEX coupled to confocal microscopy to and are currently imaging 40+ colors and scanning whole renal biopsies. Provided in Figure A is a 20 color CODEX example of a whole biopsy scan. The methods to analyze such confocal images will be extensions and adaptations of analytical approaches described in the JCI paper.

What opportunities for training and professional development has the project provided?

In the JCI paper, the first author, 5th author and 6th author were all graduate students. The first author is a MD/PhD student who successfully defended her PhD thesis and returned to medical school. Through the PhD she developed an understanding of human immunology and a deep knowledge of computational biology including advanced machine learning. The 5th author is also a MD/PhD student who is taking aspects of the project forward focusing on developing advanced tools to quantify cognate immunity. The 6th author was a PhD student, graduated and is now in a postdoctoral fellowship focused on translational biology in cancer. Unfortunately, attendance at conferences has been limited during the pandemic.

How were the results disseminated to communities of interest?

These data were presented to the Southern California Rheumatology Group virtually in February 2022 and will be presented at a national conference, Lupus2022 to be held in Tucson in September. Patients attend the Lupus2022 meeting.

What do you plan to do during the next reporting period to accomplish the goals?

In the coming year, we will perform CODEX on 10 renal biopsies from lupus patients and compare this to 10 renal biopsies from renal allograft rejection. This will be done with 43 markers. We will then finalize our computational pipeline including validating cell segmentation approaches using supervised learning (trained on nuclear signature, CD45 or DIC) and then classification algorithm using unsupervised learning. This initial analysis will then be used to finalize the final staining panel, develop hypotheses and then to study a longitudinal cohort of lupus nephritis patients (n>100).

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

Our studies provided the first demonstration that application of artificial intelligence to analyze confocal images of biopsies from lupus nephritis can identify prognostic groups, and potential therapeutic targets, not discernable through other published approaches. These studies suggest that more in-depth studies, that capture the full heterogeneity of inflammation in lupus nephritis, will reveal additional pathogenic processes that can be targeted therapeutically.

What was the impact on other disciplines?

The general approach that we developed in lupus nephritis could be applied to other inflammatory diseases, especially those that affect the kidney. Indeed, we plan to extend our studies to renal biopsies from renal allograft rejection patients.

What was the impact on technology transfer?

Our methods, including novel analytical approaches, were published in the Journal of Clinical Investigation. These methods are available to other investigators to study other disease states.

What was the impact on society beyond science and technology?

Nothing to Report

5. CHANGES/PROBLEMS:

Changes in approach and reasons for change

Nothing to report.

Actual or anticipated problems or delays and actions or plans to resolve them

Nothing to Report

Changes that had a significant impact on expenditures

Funds were spent as proposed and expected. While the overall expected cost of the project was as proposed, we were able to accelerate project development and deliver results in a shorter time than anticipated. The main aims of the grant have been accomplished which required using all DOD funds.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Significant changes in use or care of human subjects

Nothing to Report

Significant changes in use or care of vertebrate animals

Significant changes in use of biohazards and/or select agents

Nothing to report

6. PRODUCTS:

- **Publications, conference papers, and presentations**

Journal publications.

Abraham R*, Durkee MS, Ai J, Veselits M, Casella G, Asano Y, Chang A, Ko K, Oshinsky C, Peninger E, Giger ML, Clark MR. 2022. Specific in situ inflammatory states associate with progression to renal failure in lupus nephritis. J Clinical Investigation. 132:e155350. PMID: 35608910. Commentary: Li H, Tsokos MG and Tsokos GC. 2022. Lymphocytes in the neighborhood: good or bad for the kidney? 132:e160657 PMID: 35775489. Acknowledgement of federal support: yes

Artificial Intelligence and Cellular Segmentation in Tissue Microscopy Images.
Durkee MS, Abraham R, Clark MR, Giger ML. Am J Pathol. 2021 Oct;191(10):1693-1701. doi: 10.1016/j.ajpath.2021.05.022. Epub 2021 Jun 12. PMID: 34129842. Acknowledgement of federal support: yes

See appendix for publications.

Books or other non-periodical, one-time publications.

Nothing to Report

Other publications, conference papers and presentations.

Nothing to Report

- **Website(s) or other Internet site(s)**

Nothing to Report

- **Technologies or techniques**

Bioinformatic techniques, including novel AI approaches, to segment cells in confocal micrographs of human tissue were published in JCI as above.

- **Inventions, patent applications, and/or licenses**

Nothing to Report.

- **Other Products**

Data from the JCI paper was deposited and is available.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Marcus Clark
PI
ORCID ID: 0000-0003-4747-0387
0.30 calendar months
Overall PI

Maryellen Giger
Collaborator
ORCID ID: 0000-0001-5482-9728
0.12 calendar months
Provided guidance for computational portion of grant.

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Alliance for Lupus Research (Clark)
In situ adaptive immunity in Lupus tubulointerstitial inflammation

12/1/20-11/30/23

As noted in the previous report, the DOD grant funded the development of the imaging and computational technologies to analyze renal biopsies. This included application to high dimensional renal images in cross-sectional cohorts (Figures 5-8, JCI paper). The ALR grant supports application of this technology, and the development of additional analytics, to study longitudinal cohorts. Some of this initial analysis is also reported in the JCI paper.

What other organizations were involved as partners?

Nothing to Report

8. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS: QUAD CHARTS:

9. APPENDICES:

Specific in situ inflammatory states associate with progression to renal failure in lupus nephritis

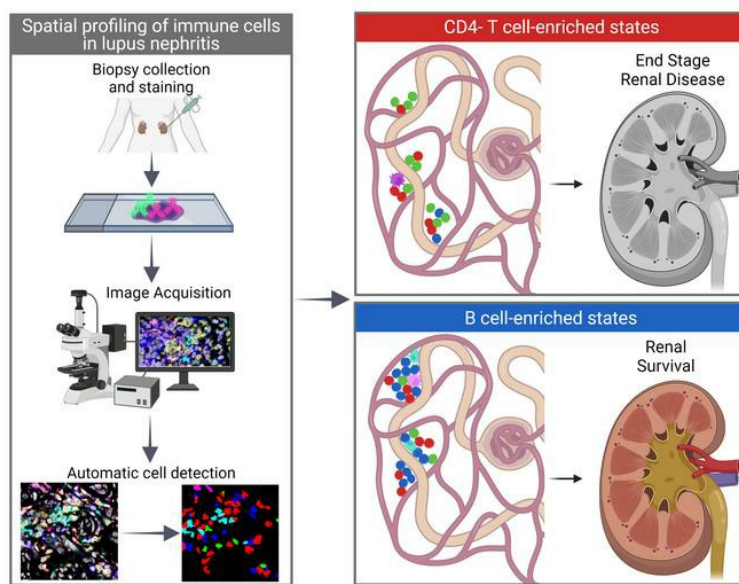
Rebecca Abraham, ... , Maryellen L. Giger, Marcus R. Clark

J Clin Invest. 2022;132(13):e155350. <https://doi.org/10.1172/JCI155350>.

Clinical Medicine

Autoimmunity

Graphical abstract



Find the latest version:

<https://jci.me/155350/pdf>



Specific in situ inflammatory states associate with progression to renal failure in lupus nephritis

Rebecca Abraham,¹ Madeleine S. Durkee,^{1,2} Junting Ai,¹ Margaret Veselits,¹ Gabriel Casella,^{1,2} Yuta Asano,¹ Anthony Chang,³ Kichul Ko,¹ Charles Oshinsky,⁴ Emily Peninger,¹ Maryellen L. Giger,² and Marcus R. Clark¹

¹Section of Rheumatology and Gwen Knapp Center for Lupus and Immunology Research, ²Department of Radiology, and ³Department of Pathology, University of Chicago, Chicago, Illinois, USA. ⁴Division of Rheumatology, University of Washington, Seattle, Washington, USA.

BACKGROUND. In human lupus nephritis (LN), tubulointerstitial inflammation (TII) on biopsy predicts progression to end-stage renal disease (ESRD). However, only about half of patients with moderate-to-severe TII develop ESRD. We hypothesized that this heterogeneity in outcome reflects different underlying inflammatory states. Therefore, we interrogated renal biopsies from LN longitudinal and cross-sectional cohorts.

METHODS. Data were acquired using conventional and highly multiplexed confocal microscopy. To accurately segment cells across whole biopsies, and to understand their spatial relationships, we developed computational pipelines by training and implementing several deep-learning models and other computer vision techniques.

RESULTS. High B cell densities were associated with protection from ESRD. In contrast, high densities of CD8⁺, $\gamma\delta$, and other CD4⁺CD8⁺ T cells were associated with both acute renal failure and progression to ESRD. B cells were often organized into large periglomerular neighborhoods with Tfh cells, while CD4⁺ T cells formed small neighborhoods in the tubulointerstitium, with frequency that predicted progression to ESRD.

CONCLUSION. These data reveal that specific in situ inflammatory states are associated with refractory and progressive renal disease.

FUNDING. This study was funded by the NIH Autoimmunity Centers of Excellence (AI082724), Department of Defense (LRI180083), Alliance for Lupus Research, and NIH awards (S10-OD025081, S10-RR021039, and P30-CA14599).

Introduction

For over 50 years, systemic lupus erythematosus has been thought to result from a break in systemic tolerance and production of pathogenic autoreactive antibodies (1, 2). This canonical model is based on extensive studies of patient blood and spontaneous systemic lupus erythematosus-like animal models (3, 4). In the kidney, the manifestation of systemic autoimmunity is glomerulonephritis (GN). Indeed, lupus nephritis (LN) is usually equated with GN (5). However, tubulointerstitial inflammation (TII)—and not GN—predicts progression to end-stage renal disease (ESRD; refs. 6–9).

Lupus TII is associated with a local immune response very different than the inflammation observed in glomeruli. Indeed, TII is associated with infiltrating B cells, plasma cells, T follicular helper (Tfh) cells, plasmacytoid DCs (pDCs), and myeloid

DCs (mDCs), although these cells are rare in LN glomeruli (10–15). These cell subsets are often organized into lymphoid-like architecture in the tubulointerstitium. This phenomenon is associated with local antigen-driven B cell clonal selection (14, 15), suggesting that adaptive immunity in the tubulointerstitium could play a role in driving renal outcomes. There is, therefore, a compelling need to understand in situ adaptive immunity in human LN.

An initial road map to the lupus kidney was provided by the Accelerating Medicines Partnership-funded (AMP-funded) single-cell RNA sequencing (scRNA-Seq) of cells sorted from LN biopsies (16, 17). While these AMP investigations were informative, there were several limitations. The patient sample was small, and the frequency of each immune cell population was not been related to relevant histological features. A larger deficiency of scRNA-Seq is that all spatial information is lost. We do not know how populations spatially relate to each other in the kidney. This lack of spatial information prevents potential functional relationships from being identified.

Previously, in lupus TII, we used conventional immunofluorescence microscopy coupled to evolving computational and machine-learning approaches to characterize the frequency of specific cell populations and identify cell/cell behaviors indicative of cognate immunity (12, 13, 18). However, a systemic analysis of

► **Related Commentary:** <https://doi.org/10.1172/JCI160657>

Authorship note: RA and MSD contributed equally to this work and are co-first authors.

Conflict of interest: The authors have declared that no conflict of interest exists.

Copyright: © 2022, Abraham et al. This is an open access article published under the terms of the Creative Commons Attribution 4.0 International License.

Submitted: September 27, 2021; **Accepted:** May 19, 2022; **Published:** July 1, 2022.

Reference information: *J Clin Invest.* 2022;132(13):e155350.

<https://doi.org/10.1172/JCI155350>.

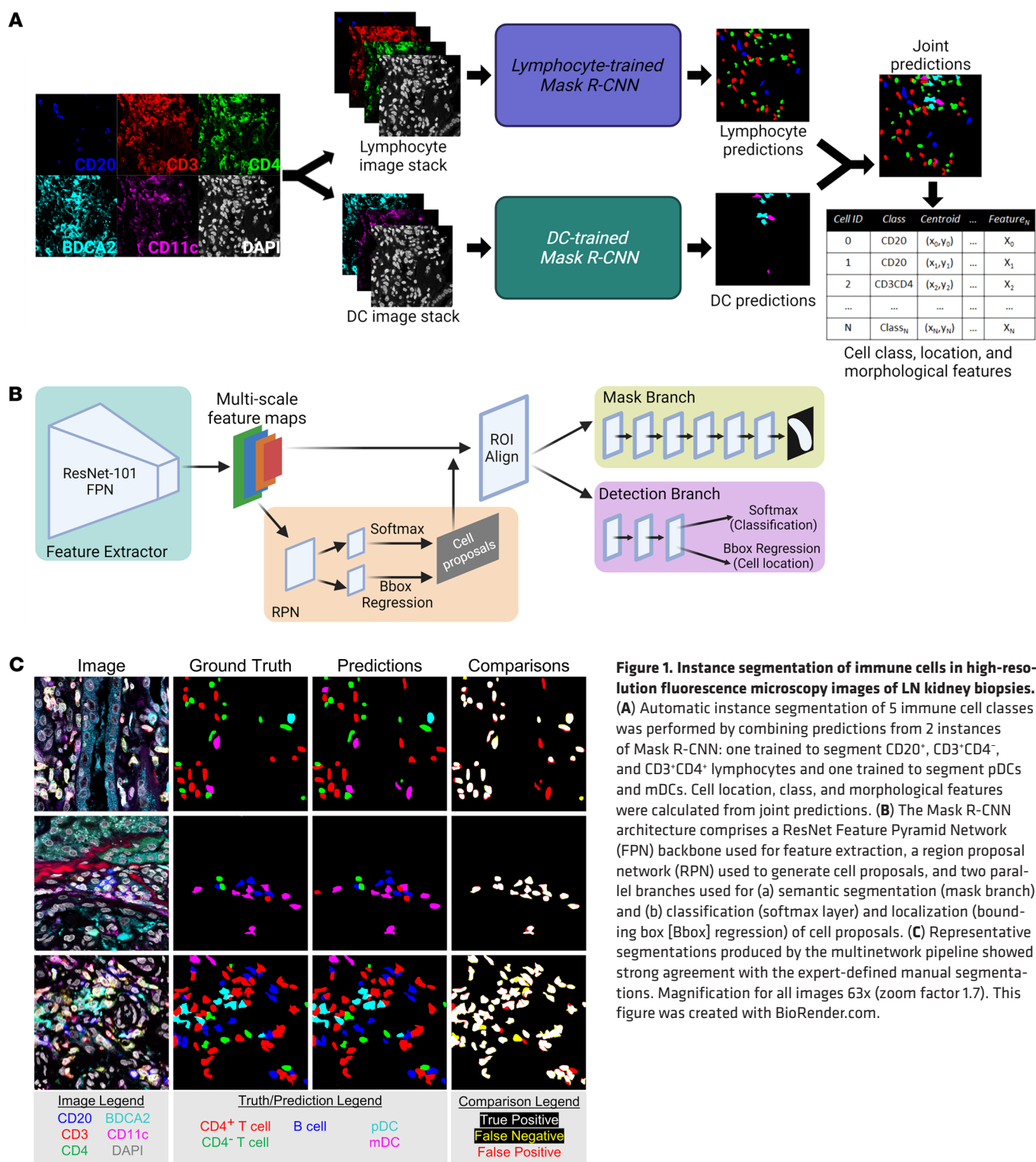


Figure 1. Instance segmentation of immune cells in high-resolution fluorescence microscopy images of LN kidney biopsies. (A) Automatic instance segmentation of 5 immune cell classes was performed by combining predictions from 2 instances of Mask R-CNN: one trained to segment CD20⁺, CD3⁺CD4⁻, and CD3⁺CD4⁺ lymphocytes and one trained to segment pDCs and mDCs. Cell location, class, and morphological features were calculated from joint predictions. (B) The Mask R-CNN architecture comprises a ResNet Feature Pyramid Network (FPN) backbone used for feature extraction, a region proposal network (RPN) used to generate cell proposals, and two parallel branches used for (a) semantic segmentation (mask branch) and (b) classification (softmax layer) and localization (bounding box [Bbox] regression) of cell proposals. (C) Representative segmentations produced by the multinet pipeline showed strong agreement with the expert-defined manual segmentations. Magnification for all images 63x (zoom factor 1.7). This figure was created with BioRender.com.

TII and identification of prognostic features has been historically impeded by the complexities of analyzing immunofluorescence data from chronically inflamed kidneys, including tissue autofluorescence due to scarring, antibody cross-reactivity, and patient heterogeneity. Artificial intelligence algorithms have led to significant progress in automated detection and analysis of cells in confocal images (19–22). These approaches have been used suc-

cessfully in cancer biopsy image analysis (23). However, conventional methods for cell detection and segmentation are not easily generalized to chronically inflamed organs.

Herein, we describe multiple computational pipelines employing deep-learning algorithms that provide high-throughput assessments of cell phenotypes and cellular architectures. These methods have been developed and validated in LN image data sets

consisting of both discrete fields of view and entire biopsy sections images. Integration of these data revealed that CD4⁺ T cell populations, comprising CD8⁺, $\gamma\delta$, and double-negative (CD4⁺CD8⁻ δ ; DN) T cells, often organized into small cellular neighborhoods, are both associated with acute refractory disease and predict progression to renal failure. In contrast, regions of high B cell density were associated with patients who did not progress to renal failure. These and other findings indicate that systemic and in situ autoimmune pathogenic mechanisms are different in LN, and each might require specific targeted therapies.

Results

Accurate segmentation of immune cells in LN kidney biopsies. To probe the relationship between TII and clinical outcome we used a well-characterized cohort of 55 biopsy-proven patients with LN with at least 2 years of follow-up (Supplemental Tables 1 and 2; supplemental material available online with this article; <https://doi.org/10.1172/JCI155350DS1>). Within this cohort, 19 patients progressed to ESRD (ESRD⁺), requiring either dialysis or transplant within the follow-up period, while 36 did not (ESRD⁻). The ESRD⁺ and ESRD⁻ groups did not differ in length of follow-up, duration of disease, or patient age (Supplemental Figure 1, A–C). Additional information about patient treatment can be found in Supplemental Tables 1 and 2. Thirty-eight patients had moderate or severe TII distributed across both outcome groups. Based on previous studies (6), we hypothesized that differences in renal outcome would be related to differences in in situ adaptive immunity, such as frequency and organization of principal cellular effectors. Therefore, we stained each biopsy for 6 markers, CD3, CD4, CD20, CD11c, BDCA2, and DAPI, to characterize 5 classes of immune cells: CD3⁺CD4⁺ T cells, CD3⁺CD4⁻ T cells, CD20⁺ B cells, BDCA2⁺ pDCs, and CD11c⁺ mDCs. Across the 55 biopsies, we captured all regions of interest (ROIs) with detectable CD3⁺ T cells, resulting in 865 ROIs. Image ROIs were 1024 × 1024 pixels, with a pixel size of 0.1058 μ m. These data are referred to as the high-resolution (HR) data set.

LN is often characterized by chronic and intense inflammation in which accurate cell segmentation can be difficult due to the high cell densities and structured background signal (12, 21). Therefore, we trained deep convolutional neural networks (DCNNs) to perform automatic cell detection, classification, and segmentation (collectively known as instance segmentation) on the HR data set. To achieve optimal performance across all cell classes, we split the 5-class cell detection into two tasks: instance segmentation of lymphocytes and instance segmentation of DCs (Figure 1A). For each task, a separate instance of a region-based DCNN architecture, Mask R-CNN, was independently trained (Figure 1B and ref. 24). Each Mask R-CNN was trained on 246 manually segmented images with a validation set of 65 manually segmented images used for hyperparameter tuning. A cell prediction was defined as a true positive prediction if it had an intersection-over-union (IOU) score of greater than 0.25 with a ground truth cell of the same class. Additionally, all cell predictions with a network confidence score of less than 0.3 were rejected. On a test set of 34 images from patients unique to the training and validation data, the lymphocyte detection network and the DC detection network had F_1 scores (Equation

3) of 0.75 and 0.62, respectively, while the overall F_1 score for detection of all 5 cell classes was 0.74, yielding excellent concordance (Figure 1C). The class-specific F_1 scores for the lymphocyte network, DC network, and the combined predictions are reported in Supplemental Table 3. By implementing DCNNs, we achieved rapid and accurate multiclass instance segmentation.

Specific in situ immune cell densities associated with progression to renal failure. Automatic cell segmentations were used to describe and quantify the spatial distribution of all 5 cell classes in the HR data set. A comparison of overall cell densities (total cells/ROI) in ESRD⁻ and ESRD⁺ patients revealed no significant differences (Figure 2A). However, the total cell count per sample was higher in the ESRD⁺ cohort, reflecting larger overall areas of inflammation (Figure 2B). In contrast to overall cell density, there were differences in the cellular constituents of inflammation between the two patient cohorts. Surprisingly, ROIs from ESRD⁻ patients had higher densities of B cells relative to ROIs from ESRD⁺ patients (Figure 2C). In contrast, ROIs from ESRD⁺ patients had increased densities of CD4⁺ T cells (Figure 2D). There were no significant differences in the densities of CD4⁺ T cells, pDCs, or mDCs between patient cohorts (Figure 2, E–G).

Although there were fewer ESRD⁺ patients, on average these patients had more ROIs captured per biopsy. To mitigate any effect from this class imbalance, we performed a bootstrapping analysis. The pools of ESRD⁺ and ESRD⁻ ROIs were iteratively sampled with replacement 1000 times to produce samples of 200 ROIs from each group (ESRD⁺ and ESRD⁻). The distribution of mean cell densities between ESRD⁺ and ESRD⁻ patients revealed distinct, nonoverlapping peaks for both B cells and CD4⁺ T cells (Figure 2, H and I). In contrast, there was substantial overlap in the distribution of sample means between ESRD⁺ and ESRD⁻ patients for CD4⁺ T cells, pDCs, and mDCs (Figure 2, J–L). The 95% confidence intervals of the difference in means between ESRD⁺ and ESRD⁻ patients revealed for both B cells and CD4⁺ T cells did not cross 0 (Supplemental Figure 2, A and B). In contrast, the 95% confidence interval for the difference in means for the remaining cell types did cross 0 (Supplemental Figure 2, C–E). These data indicate that the observed differences in B cell and CD4⁺ T cell densities between ESRD⁺ and ESRD⁻ patients are robust. Furthermore, our results did not significantly change if the 2 patients who received rituximab were removed (Supplemental Table 1 and data not shown). Therefore, we conclude that high B cell densities are associated with a good prognosis, while high densities of CD4⁺ T cells are associated with progression to renal failure.

When we examine these densities on the patient level, we observed that, in patients with high CD4⁺ T cell densities, B cell densities tended to be low (Figure 2M). As indicated by point size, these tended to be ESRD⁺ patients with higher tubulointerstitial (TI) chronicity scores. The converse appeared true, as patients with higher B cell densities tended to have low TI chronicity scores and be ESRD⁻. These data suggest that lupus TII is associated with two or more distinct inflammatory states, each associated with a different prognosis.

Patients who present in renal failure have a skewed in situ inflammatory state. Within the ESRD⁺ group of patients was a small yet distinct cohort of 5 patients that either were in renal failure at

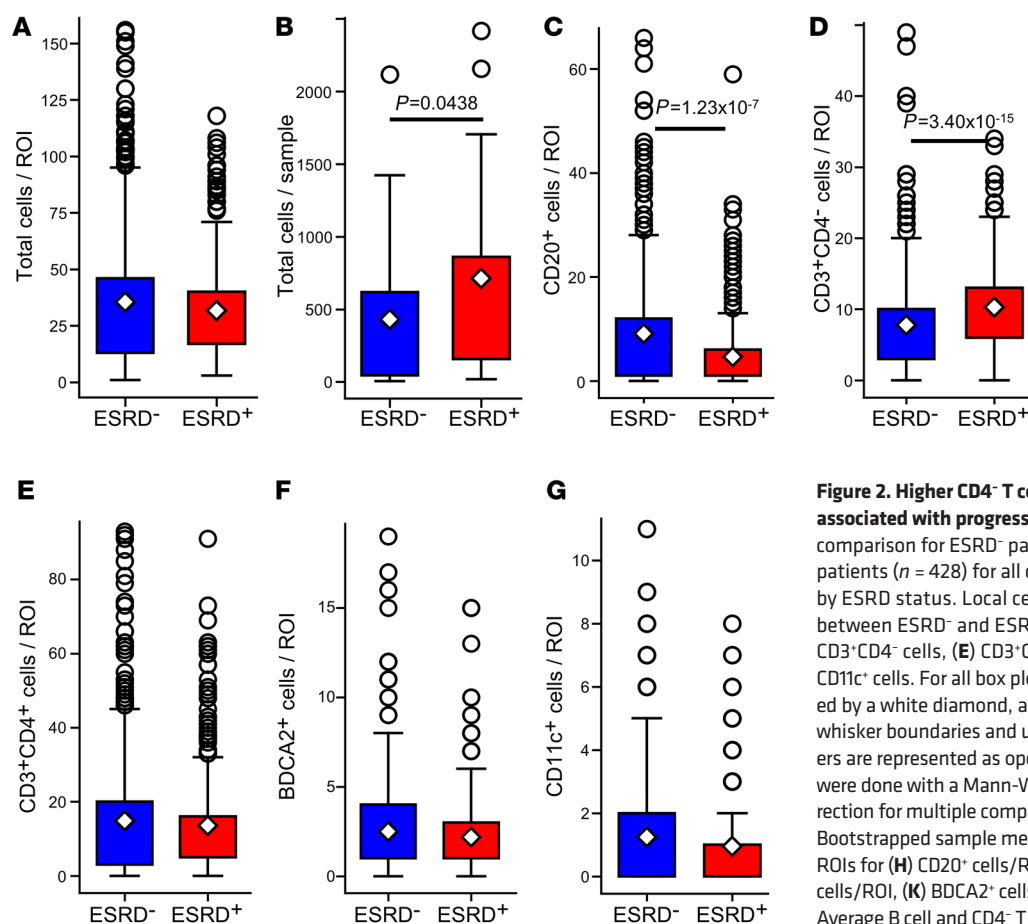
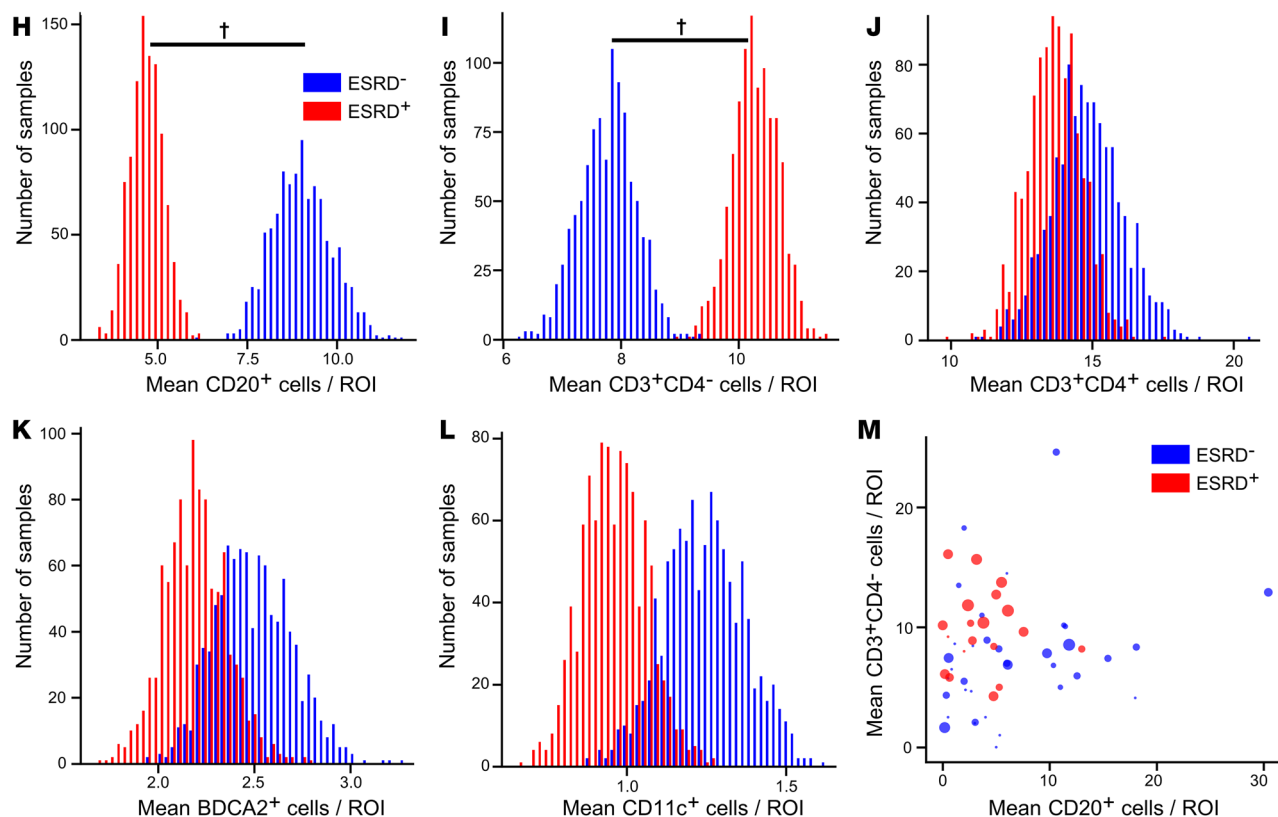


Figure 2. Higher CD4⁺ T cell density and lower B cell density associated with progression to ESRD.

(A) Local cell density comparison for ESRD⁻ patients ($n = 437$ ROIs) and ESRD⁺ patients ($n = 428$) for all cells. (B) Total cells per patient grouped by ESRD status. Local cell density by cell class compared between ESRD⁻ and ESRD⁺ patient for (C) CD20⁺ cells, (D) CD3⁺CD4⁻ cells, (E) CD3⁺CD4⁺ cells, (F) BDCA2⁺ cells, and (G) CD11c⁺ cells. For all box plots, the population mean is represented by a white diamond, and quartile ranges are defined by the whisker boundaries and upper and lower box boundaries. Outliers are represented as open circles. All cell density comparisons were done with a Mann-Whitney U test with a Bonferroni's correction for multiple comparisons, with significant P values noted. Bootstrapped sample means of ESRD⁻ (blue) and ESRD⁺ (red), ROIs for (H) CD20⁺ cells/ROI, (I) CD3⁺CD4⁻ cells/ROI, (J) CD3⁺CD4⁺ cells/ROI, (K) BDCA2⁺ cells/ROI, and (L) CD11c⁺ cells/ROI. (M) Average B cell and CD4⁺ T cell count per ROI for each patient biopsy. Point size is weighted by the TI chronicity score for each patient. *95% confidence interval does not overlap with 0.



the time of biopsy or progressed to renal failure within 2 weeks of biopsy collection. If these patients are treated as their own unique outcome group (ESRD current), differences in the density of specific cell classes become even more apparent (Figure 3). There were progressively fewer B cells/ROI among the ESRD⁻, ESRD⁺, and ESRD current groups, respectively (Figure 3A). The opposite trend was observed for CD4⁺ T cell densities (Figure 3B). In contrast, there were no apparent differences in CD4⁺ T cells or pDCs in the ESRD current patients (Figure 3, C and D). Remarkably, there was a profound depletion of mDCs in the ESRD current cohort (Figure 3E).

A 3-group bootstrapping analysis was performed to assess the effect of class imbalance in patient numbers. ESRD current patients had the lowest mean density of B cells, followed by ESRD⁺ patients, with ESRD⁻ patients having the highest density of B cells (Figure 3F). Confidence intervals for the pairwise differences between bootstrapped samples did not overlap with 0 (Supplemental Figure 3A). An inverse, stepwise relationship was observed for CD4⁺ T cells with progressively higher densities found in the ESRD⁺ and ESRD current patients relative to ESRD⁻ patients (Figure 3G and Supplemental Figure 3B). ESRD current patients were also well separated from the other 2 cohorts with respect to local mDC abundance (Figure 3J and Supplemental Figure 3E). As expected, there were no differences between the 3 groups with respect to CD4⁺ T cells or pDCs (Figure 3, H and I, and Supplemental Figure 3, C and D). These findings indicate that patients with LN that present in renal failure have a skewed inflammatory state with abundant CD4⁺ T cells, relatively few B cells, and a depletion of mDCs.

Specific cellular neighborhoods associated with progressive and refractory renal disease. We next explored the relative in situ spatial relationships between the different immune cell classes. First, for every cell in the data set, we identified the nearest neighbor using centroid-to-centroid distances. All cell classes except mDCs were significantly more likely to have a B cell as their nearest neighbor in ESRD⁻ biopsies (Figure 4A). In contrast, all cell classes were significantly more likely to have a CD4⁺ T cell nearest neighbor in ESRD⁺ biopsies (Figure 4B). Additionally, both B cells and CD4⁺ T cells showed a strong propensity for colocalization with cells of the same type.

Local cellular organization was then probed by grouping cells into spatially discrete neighborhoods. DBSCAN, a density-based clustering algorithm (25), was implemented to define cell neighborhoods using a maximum intercellular centroid-to-centroid distance. Variation in this maximum distance between 50 and 150 pixels resulted in a range of neighborhood sizes varying between those that contained just a few cells (50 pixels) to those that encompassed large areas of inflammation (150 pixels) (Figure 4C and Supplemental Figure 4A). A maximum distance of 100 pixels (~10.6 μ m) was selected, as this distance approximates a cell body and appeared to capture observable groupings of cells across the data set.

Using this 100-pixel cutoff and a minimum neighborhood size of 2, DBSCAN detected 4022 cell neighborhoods. Each neighborhood was characterized by a set of 24 quantitative features, including cell type frequency, cell type proportion, ratios of cell types, cell shape features, and neighborhood area (Supplemental

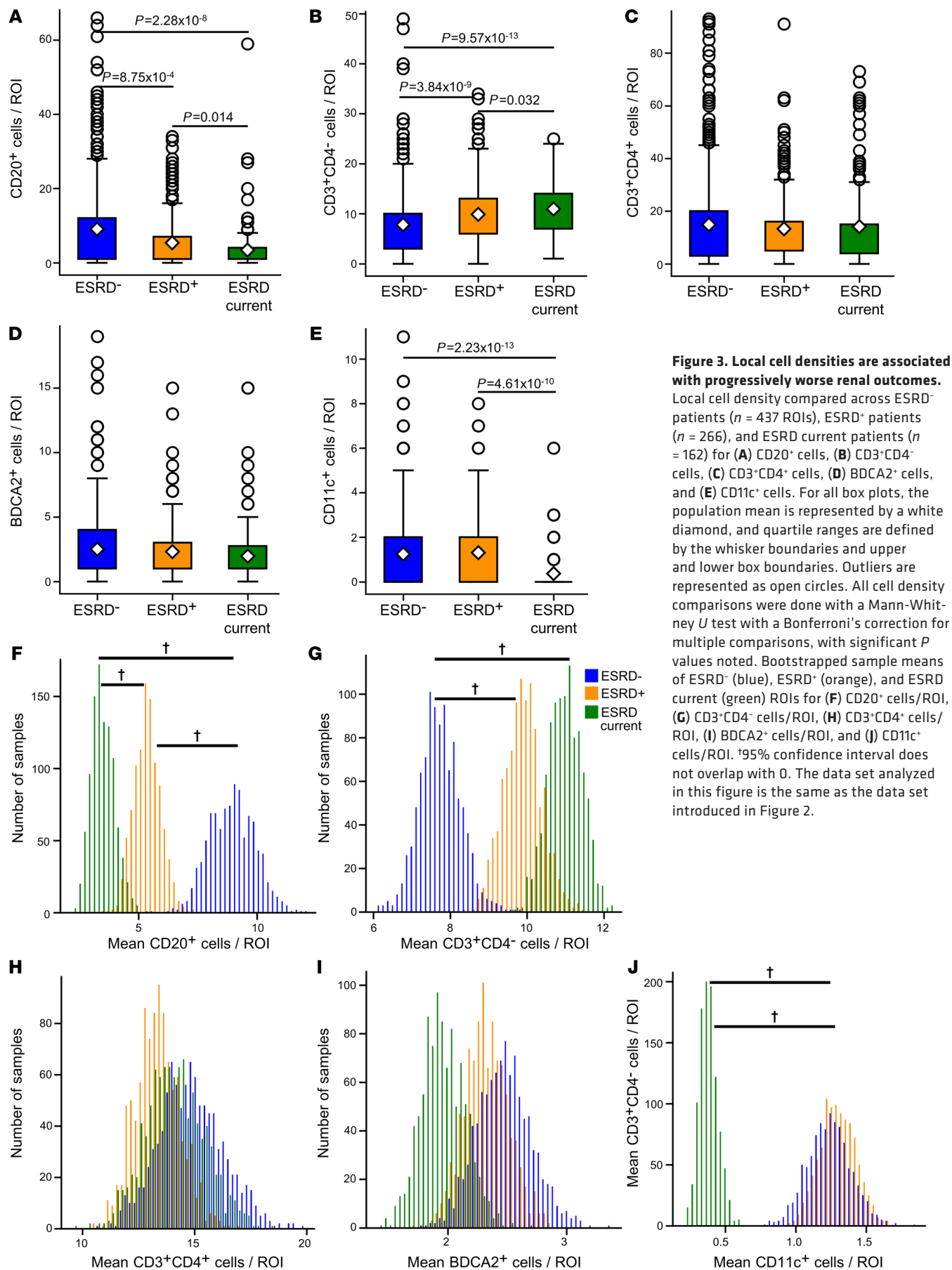
Figure 4B). K-means clustering was then applied to define classes of neighborhoods, with $k = 6$ classes determined ideal by bootstrapping cluster descriptors, including the within-cluster sum of squares (WCSS) and the Δ WCSS (Supplemental Figure 4, C and D). The test score from a leave-one-out t test approach was used to determine which features or combination of features best distinguished the 6 neighborhood groups (Figure 4D). The most distinctive feature(s) for each group was used to describe the cell neighborhoods as follows: (a) B cell-enriched cluster; (b) CD4⁺ T cell-enriched cluster; (c) large lymphocyte-enriched cluster; (d) CD4⁺ T cell-enriched cluster, (e) mDC-enriched cluster; and (f) pDC-enriched cluster (Figure 4E).

Tertiary lymphoid structures (TLSs) have been previously identified in the context of LN (15). Although we cannot explicitly define TLSs in this data set, we hypothesized that some of the large lymphocyte-enriched neighborhoods might approximate TLSs. For example, we noted that within this group 28.6% of the cells were B cells and 48.3% were CD4⁺ T cells. 96.1% of these neighborhoods met the following criteria: (a) contained at least 20 cells, (b) both B cells and CD4⁺ T cells were represented in the neighborhood, and (c) at least 50% of all cells were B cells and/or CD4⁺ T cells. Therefore, the vast majority of large lymphocyte-enriched neighborhoods have features consistent with TLSs.

We then examined how these 6 classes of neighborhoods were distributed between the ESRD⁻ and ESRD⁺ patients. After normalizing by the number of ROIs captured for each patient, ESRD⁻ and ESRD⁺ patients had no difference in their total neighborhood count per ROI (Figure 4F). However, ESRD⁺ patients had a significantly higher prevalence of CD4⁺ T cell-enriched neighborhoods relative to the ESRD⁻ patients (Figure 4G). The per-ROI prevalence of the other classes of neighborhoods did not correlate with renal outcome (Supplemental Figure 5, A–E). We next examined if the CD4⁺ neighborhoods differed among the ESRD⁻, ESRD⁺, and ESRD current patient groups (Figure 4H). ESRD⁺ and ESRD current patients had a statistically higher prevalence of neighborhoods from the CD4⁺ cluster compared with ESRD⁻ patients. These data demonstrate that, on a per patient basis, the prevalence of small CD4⁺ T cell-enriched neighborhoods is strongly associated with progressive renal disease.

Cell detection and segmentation in highly multiplexed, full-biopsy images. To better characterize in situ lymphocyte populations, we performed highly multiplexed (HMP) confocal microscopy on a separate data set of 18 LN biopsies. In this HMP data set, we interrogated a set of 9 markers (CD3, CD4, CD8, ICOS, PD1, FoxP3, CD20, CD138, and DAPI). This HMP panel was obtained using 4-color confocal microscopy and iterative stripping and reprob-ing (26). Additionally, we imaged full biopsy sections rather than capturing isolated ROIs, thereby facilitating a more complete and unbiased spatial analysis.

Full biopsy images for all stains were aligned with the DAPI channel (Figure 5A). Two new instances of Mask R-CNN were trained to perform single-marker and dual-marker instance segmentation (Figure 5B). Briefly, ROIs from the HR data set (pixel size = 0.1058 μ m) were broken up into 512 $\times$ 512 pixel tiles to pretrain each Mask R-CNN. Each network was then fine-tuned using small sets of manually segmented 512 $\times$ 512 pixel tiles from the HMP data set (pixel size = 0.221 μ m). The single-mark-



er Mask R-CNN was used to predict B cells (CD20⁺) and plasma cells (CD138⁺), while the dual-marker Mask R-CNN was used to predict single-positive and double-positive T cells. The 3 main classes of T cells were determined by combining predictions from a CD3/CD4/DAPI image stack with predictions from a CD3/CD8/DAPI image stack at the same location in the tissue: CD4⁺, CD8⁺, and CD4⁺CD8⁺ (DN) (Figure 5C). The dual-marker Mask R-CNN was also used to generate cell predictions on CD3/ICOS/DAPI and CD3/PD1/DAPI images. The resulting single-positive (CD3⁺ICOS⁺ or CD3⁺PD1⁺) and double-positive (CD3⁺ICOS⁺ or CD3⁺PD1⁺) predictions were used to define ICOS and PD1 expression for every putative T cell in the data set. FoxP3 images were binarized by thresholding individual image tiles. T cell predictions with more than 25% overlap with this binary mask were determined to be FoxP3⁺.

CD4⁺ T cells contain CD8, $\gamma\delta$, and other DN T cell populations. T cells comprised over 65% of predicted lymphocytes in the HMP data set (Figure 5D). Plasma cells were the second-most abundant class, comprising approximately 28% of detected lymphocytes. B cells were least prevalent at only approximately 6%. CD4⁺ T cells were the most abundant cell class across all 5 main classes, making up 35% of detected lymphocytes and over 50% of detected T cells (Figure 5E). Surprisingly, CD8⁺ T cells were only slightly more abundant than DN T cells, comprising approximately 17% of detected lymphocytes and approximately 26% of detected T cells.

To further characterize these DN T cells, we interrogated public scRNA-Seq data of immune cells infiltrating the kidneys of patients with LN (16). We identified naive T and CTL clusters in intrarenal immune cells by unsupervised clustering and canonical marker expression (Figure 6A). Within these T cell clusters, 21% were DN, as measured by the unique molecular identifier (UMI) of *CD4*, *CD8A*, and *CD8B* (Supplemental Figure 6A). Several T cell subtypes do not express CD4 nor CD8, including NK T cells and $\gamma\delta$ T cells. Indeed, there was a small increase in *CD3D* in cells assigned to the NK cell class, suggesting a NK T cell phenotype. However, there was not a substantial enrichment for NK T cell markers in the DN subset (Supplemental Figure 6B).

Next, we compared TCR α and δ chain expression (*TRAC* and *TRDC*). Some cells were apparently positive for both *TRAC* and *TRDC*, likely due to sequence homology between these genes (Figure 6B). However, *TRAC*⁺ cells and *TRDC*⁺ cells were both enriched in the DN population (Figure 6C). These results suggest that a portion of the DN T cells observed in LN are $\gamma\delta$ T cells. To further examine this possibility, we stained 8 LN biopsies with antibodies specific for CD3, CD4, CD8, and TCR δ and imaged 281 ROIs (Figure 6D). Per biopsy, 51.4% $\pm$ 21.3% of DN T cells were positive for TCR δ . These findings indicate that a substantial fraction of T cells in LN do not detectably express CD4 or CD8, and approximately half of these DN T cells are $\gamma\delta$ T cells.

In situ distributions of exhausted, regulatory, and helper T cell populations. We then examined the distributions of ICOS, PD1, and FoxP3 in the T cell subsets. Roughly 30% of CD8⁺ T cells in the HMP data set were PD1⁺ (Supplemental Figure 7A), suggesting an exhausted phenotype. Approximately 25% of CD8⁺ T cells were “exhausted” by the definition of PD1⁺ICOS⁺FoxP3⁺ (27). This is consistent with observations from murine lupus models in which exhausted tissue-infiltrating CD8⁺ T cells are relatively common

(28). However, PD1 is only one marker of exhaustion and human lupus renal scRNA-Seq data suggest CD8⁺ T cell exhaustion is infrequent (16).

A surprisingly small percentage (5.41%) of CD4⁺ T cells were FoxP3⁺, while fewer still were also PD1⁺ICOS⁺, suggesting that Tregs comprise only about 2.5% of CD4⁺ T cells (Supplemental Figure 7B). In contrast, even fewer CD8⁺ T cells (1.3%) or DN T cells (0.88%) expressed FoxP3 (Supplemental Figure 7, A and C). Therefore, very few of the tissue-infiltrating CD4⁺ T cells in LN are potentially Tregs.

We additionally identified Tfh cells based on the combination of PD1 and ICOS expression by CD4⁺ T cells (29, 30). 5.05% of the CD4⁺ T cell compartment was PD1⁺ICOS⁺FoxP3⁺. Previous investigations have consistently associated PD1 expression with Tfh-like cells (including T peripheral helper cells) but not necessarily ICOS. Therefore, we applied a less stringent definition of PD1⁺ICOS⁺/FoxP3⁺ to identify this cell subset. This Tfh cell phenotype was associated with roughly 30% of the CD4⁺ T cells (Supplemental Figure 7B). Although PD1⁺ICOS⁺FoxP3⁺ CD4⁺ T cells could be interpreted as exhausted, we chose to use the more expansive Tfh cell definition in our subsequent analysis.

Organization of inflammation across whole biopsies. We next probed potential interacting partners of Tfh cells, Tregs, and exhausted T cells by identifying the class of their nearest neighbors. Most Tregs are closest to other Tregs and other CD4⁺ T cells (Supplemental Figure 7D). In contrast to the expectation that Tfh cells would primarily be in close proximity with CD20⁺ B cells, Tfh cells had other CD4⁺ T cells as their most frequent neighbor, followed by other Tfh cells (Supplemental Figure 7E). Exhausted CD8⁺ T cells were most frequently found near other exhausted CD8⁺ T cells, followed by CD4⁺ T cells and CD8⁺ T cells (Supplemental Figure 7F). Overall, these data demonstrate that across biopsies there is a tendency for the clustering of similar cells together.

Cell neighborhoods in the HMP data set were then defined using DBSCAN with a distance cutoff of 50 pixels, roughly 10 μ m. Most neighborhoods detected in the HMP data set were small (Figure 7A). However, without the constraint of discrete fields of view, we were able to capture larger neighborhoods, with a maximum neighborhood size of 273 cells, relative to the 147 cell maximum in the HR data set. Given the association of CD4⁺ T cell-enriched neighborhoods with ESRD⁺ patients in the HR data, we investigated similar CD4⁺ neighborhoods in the HMP data. We classified CD4⁺ neighborhoods as those that (a) had less than 20 cells and (b) more than or equal to 25% of their cells were either CD8⁺ or DN T cells, as these criteria captured 99.1% of the CD4⁺ neighborhoods observed in the HR data (Figure 7B). A strong majority of the cells in these neighborhoods were CD4⁺ T cells, including 26% DN T cells and 34.2% CD8⁺ T cells (Figure 7C). There was a weak negative correlation ($R = -0.35$) between the number of DN T cells and CD8⁺ T cells in these neighborhoods, suggesting that DN and CD8⁺ T cells are not proportionally represented in a given neighborhood.

Large B and T cell (B-T) neighborhoods were defined by the set of 3 criteria (as described above) that captured most of the large lymphocyte-rich neighborhoods in the HR data. Of nearly 14,000 neighborhoods in the HMP data set, 111 met these criteria (representative clusters in Figure 7D). Within these B-T

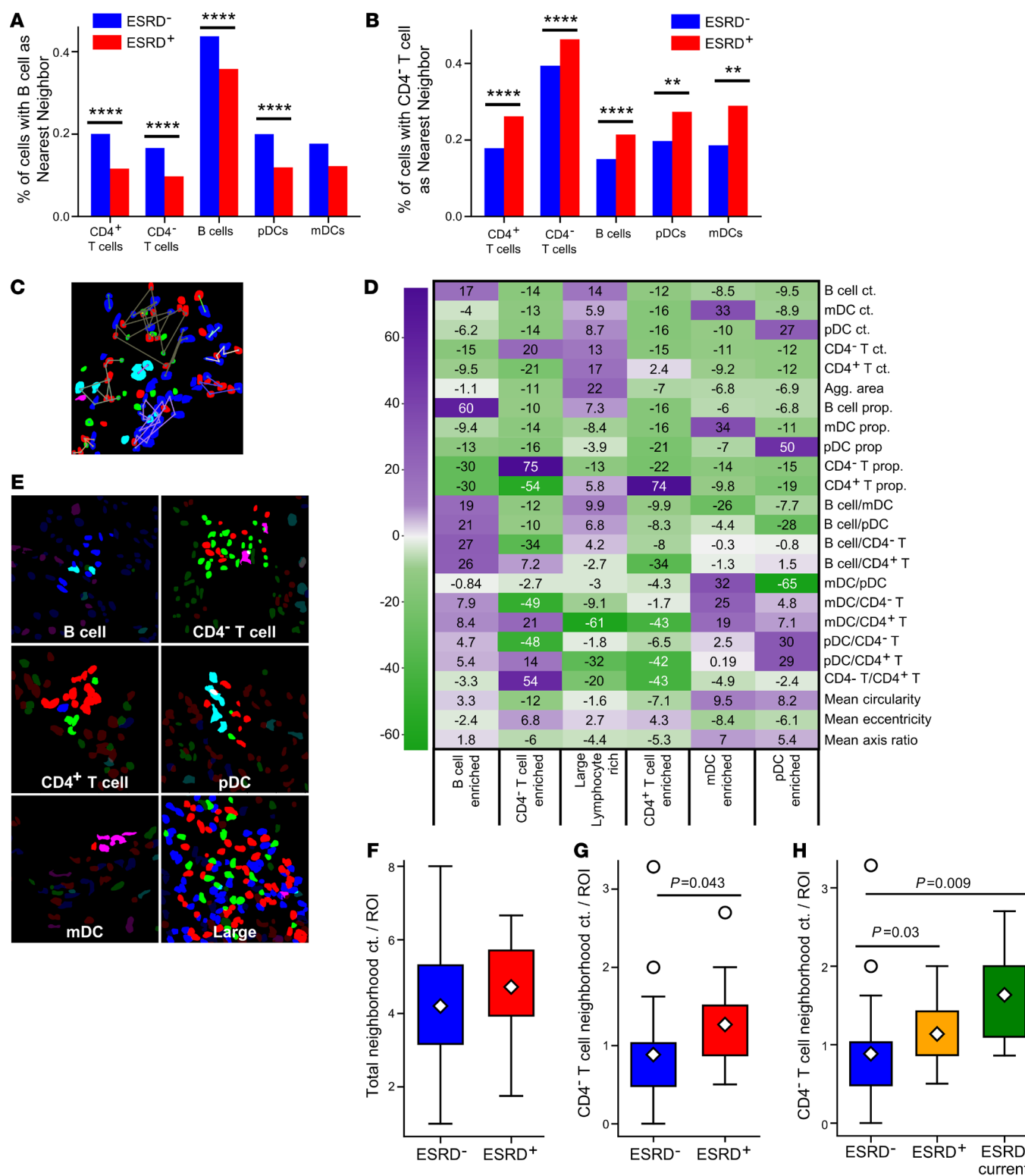


Figure 4. Specific cellular neighborhoods associated with renal failure. Proportions of cells that have (A) CD20⁺ B cells and (B) CD4⁺ T cells as nearest neighbors in ESRD⁻ and ESRD⁺ patients (χ^2 test for independence with Bonferroni's correction for multiple comparisons). (C) Neighborhoods of automatically detected cells were detected by DBSCAN. Automatic cell segmentations and representative neighborhoods (highlighted in E) are shown for images taken at 63x magnification with a zoom factor of 1.7. (D) Heatmap showing test statistics for each feature from leave-one-out *t* tests used to define 6 types of cell neighborhoods, colored by the magnitude of the test statistic. (E) Representative neighborhoods from each defined class. (F and G) The abundance of neighborhoods between the patient cohorts, normalized by the number of ROIs per patient, was compared by Mann-Whitney *U* test with a Bonferroni's correction for (F) all cell neighborhoods and (G) CD4⁺ T cell neighborhoods. A 3-group comparison for CD4⁺ neighborhoods, splitting the ESRD⁺ population into ESRD⁺ and ESRD current patients is shown in H. Significant *P* values after correcting for multiple comparisons are noted. The data set analyzed in this figure is the same as the data set introduced in Figure 2. For all box plots, the population mean is represented by a white diamond, and quartile ranges are defined by the whisker boundaries and upper and lower box boundaries. Outliers are represented as open circles.

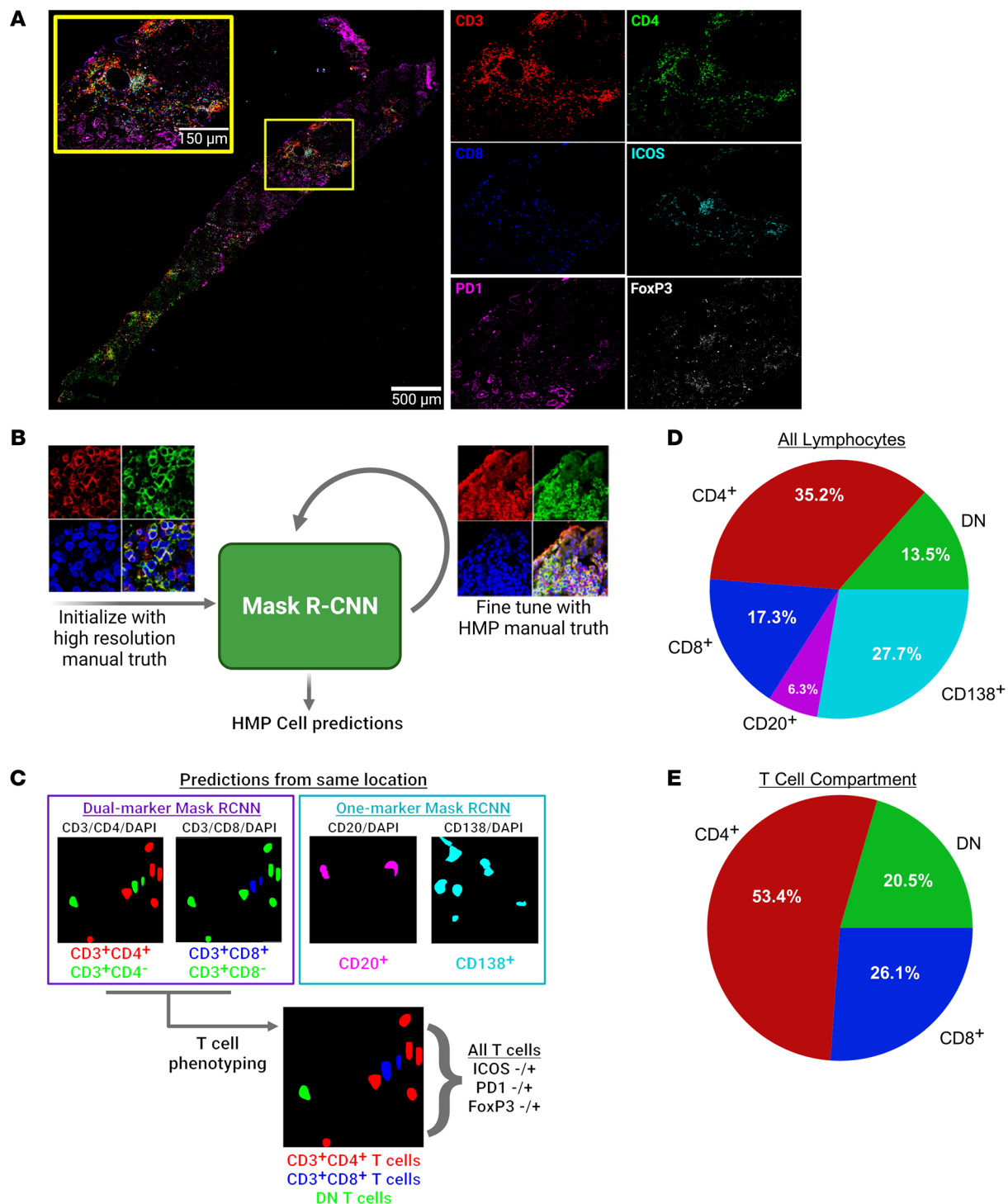


Figure 5. Cell detection, segmentation, and phenotyping in highly multiplexed fluorescence microscopy images. (A) Representative composite of a full biopsy section, shown with merged and with isolated panels of CD3, CD4, CD8, ICOS, PD1, and FoxP3. Scale bar: 150 μ m; 500 μ m (inset). (B) Schematic of procedure for training and fine-tuning a Mask R-CNN for instance segmentation of cells in highly multiplexed microscopy images. High resolution, 63x, zoom factor=1.7 (left); multiplexing image: 63x, zoom factor=1 (right). (C) Dual-marker and single-marker cell predictions are used to establish base lymphocyte classes. All T cell predictions are further described by ICOS, PD1, and FoxP3 expression. (D) Breakdown of frequencies of the 5-base classes in the HMP data set. (E) Frequencies of CD4⁺, DN, and CD8⁺ T cells within the T cell compartment. Images in A–C were created with BioRender.com.

neighborhoods, a vast majority of the lymphocytes were T cells, followed by similar proportions of B cells and plasma cells (Figure 7E). Tfh cells made up 36% of CD4⁺ T cells in B-T neighborhoods (Figure 7F), a significant enrichment compared with

non-B-T neighborhoods ($P = 1.9 \times 10^{-6}$, Mann-Whitney) (Figure 7G). As observed across whole biopsies, within these B-T neighborhoods, homotypic proximity predominated. B cells were located near other B cells, followed by plasma cells and CD4⁺

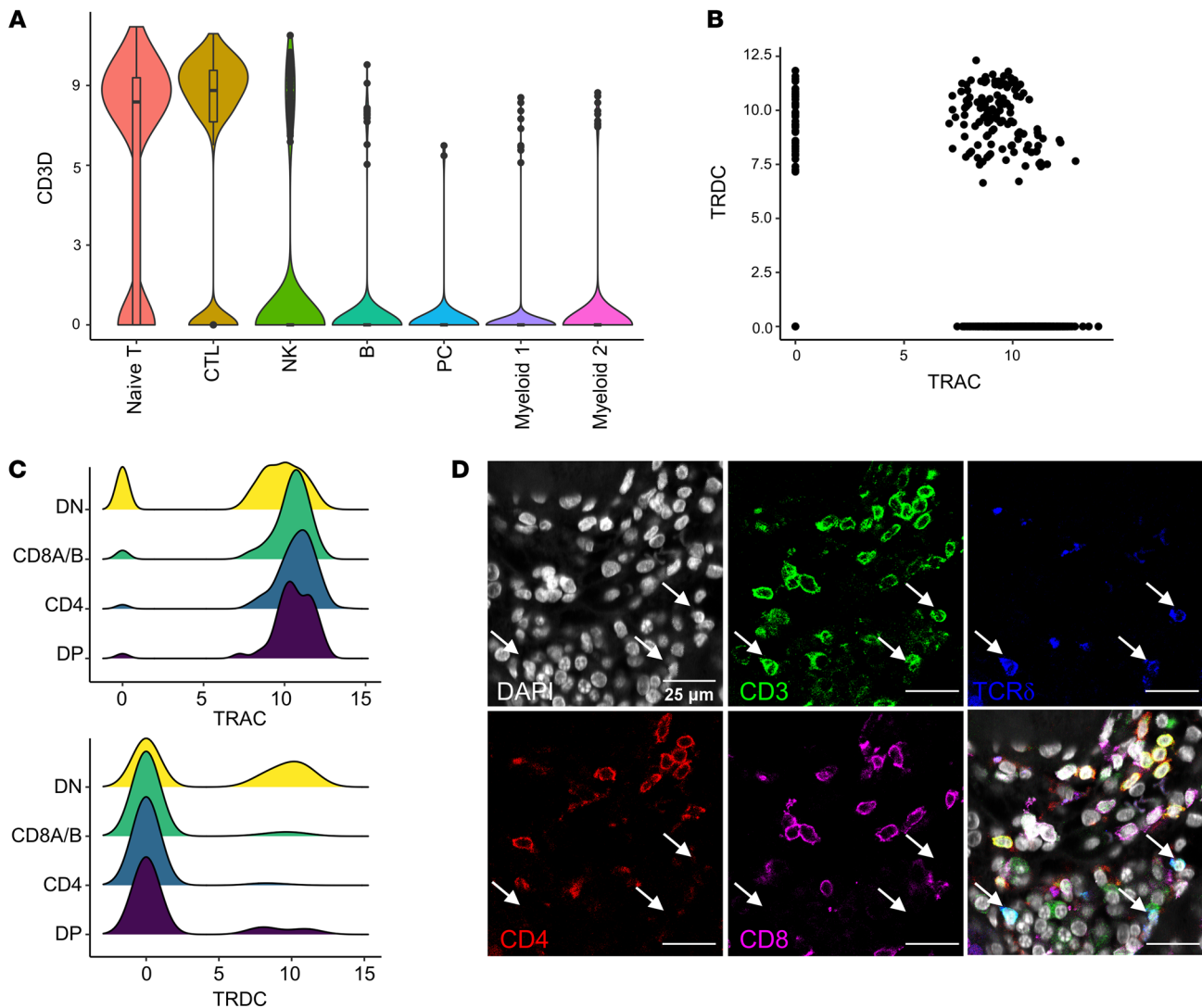


Figure 6. Identifying $\gamma\delta$ T cells in LN. (A) Distribution of *CD3D* in cell clusters identified in scRNA-Seq data from LN kidney samples. (B) Expression of *TRAC* and *TRDC* in T cells identified in scRNA-Seq data. (C) Comparison of *TRAC* and *TRDC* expression in identified double-negative (DN), $CD8^+$, $CD4^+$, and double-positive (DP) T cells. (D) Representative image of DN ($CD4^-CD8^-$) $\gamma\delta$ ($TCR\delta^+$) T cells in LN biopsy, marked by white arrows. Scale bar: 25 μm .

T cells (Figure 7H). Tfh cells in these neighborhoods were most often near other Tfh cells, while overall, Tfh cells were near unspecified $CD4^+$ T cells (Figure 7I and Supplemental Figure 7E). Unassigned $CD4^+$ T cells in B-T neighborhoods were also most likely to be found near other $CD4^+$ T cells, followed by Tfh cells (Figure 7J).

Cellular neighborhoods are differentially distributed relative to renal structures. Whole-biopsy imaging made it possible to characterize the distribution of immune cell populations relative to renal structures. Therefore, we trained a Mask R-CNN to segment the tubular structures in the biopsies, which encompassed proximal and distal tubules, and some vascular structures. Due to a low prevalence, glomeruli were manually segmented (Figure 8A). The relative areas of the tubules, glomeruli, and the resulting TI space were calculated from these structural segmentations (Figure 8B). Glomeruli were indeed much less frequent and comprised the smallest fraction of total tissue. While tubules were abundant, the TI space was the largest structural compartment defined.

The overall proximity of broad lymphocyte classes to tubules and glomeruli was assessed by calculating the minimum distance of a given cell centroid to a pixel defined as either glomerular or tubular. Plasma cells were found to be closer to tubules than all other cell types (Figure 8C and Supplemental Tables 4 and 5). Additionally, B cells and $CD4^+$ T cells were significantly farther from tubules than other cell types, although no difference was found between these 2 populations with respect to tubule proximity. Notably, plasma cells were also farther from glomeruli than all other lymphocytes (Figure 8D and Supplemental Tables 4 and 5). $CD4^+$ T cells were also closer to glomeruli than $CD8^+$ T cells, B cells or plasma cells.

Finally, the locations of the identified $CD4^-$ and large B-T neighborhoods relative to the glomeruli and tubules were assessed. On average, the large B-T neighborhoods were significantly closer to glomeruli than both the $CD4^-$ neighborhoods and all other uncategorized neighborhoods (Figure 8E). In contrast, B-T neighborhoods were significantly farther from tubules than

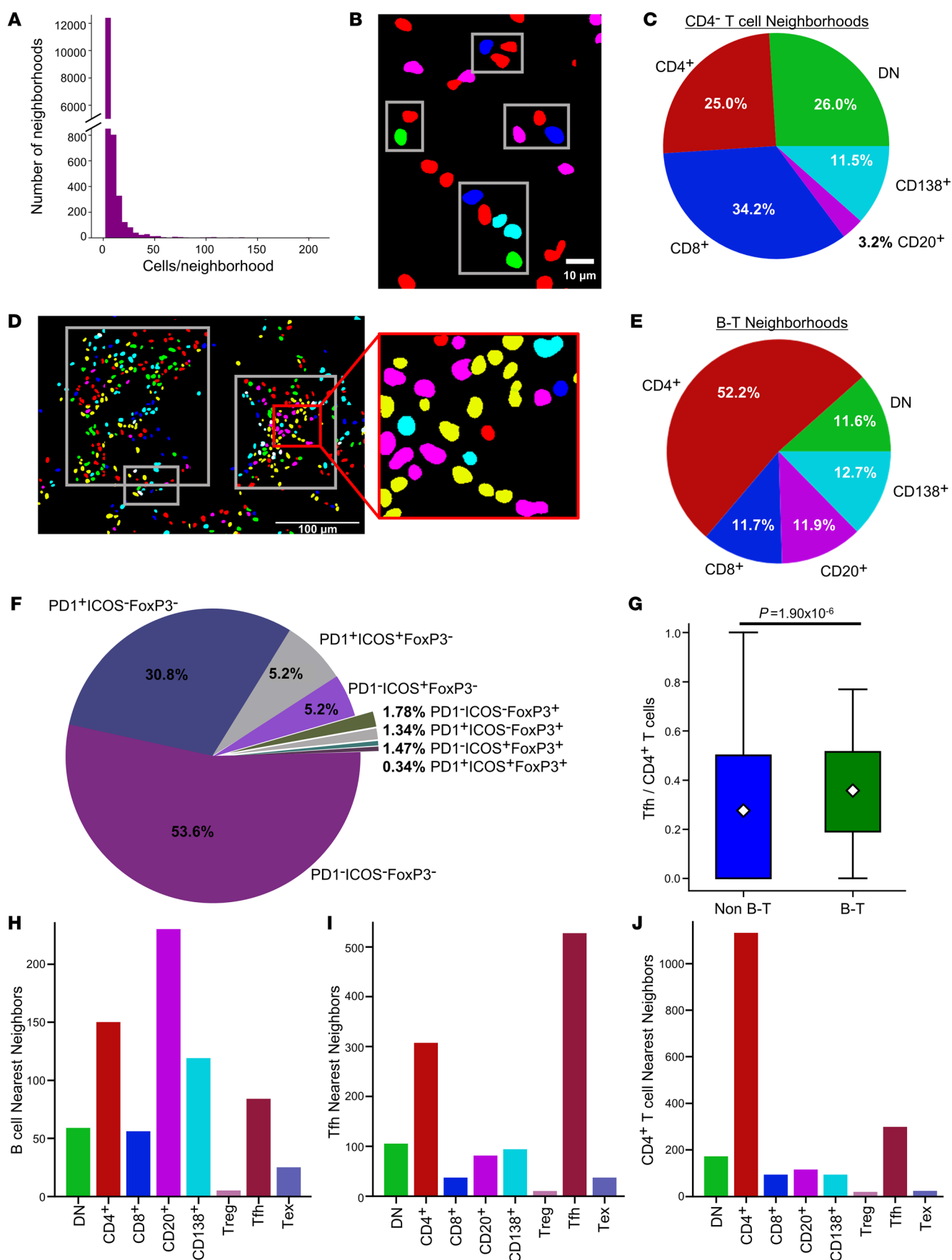


Figure 7. Identification of distinct CD4⁺ and B-T neighborhoods. (A) Distribution of sizes of all cell neighborhoods in the HMP data set. (B) Representative CD4⁺ clusters (red, CD4⁺ T cells; blue, CD8⁺ T cells; green, DN T cells). Scale bar: 10 μ m. (C) Distribution of the 5 main lymphocyte classes in the CD4⁺ T cell neighborhoods. (D) Representative B-T aggregates (outlined by white boxes) (green, DN; red, non-Tfh CD4⁺; yellow, Tfh; blue, CD8⁺; magenta, CD20⁺; cyan, CD138⁺ cells). Scale bar: 100 μ m. (E) Distribution of the 5-base classes of lymphocytes in B-T neighborhoods. (F) Distribution of CD4⁺ T cell phenotypes in B-T neighborhoods. (G) Comparison of proportion of CD4⁺ T cells that are Tfh cells in identified B-T aggregates and non B-T aggregates (Mann-Whitney *U* Test, $P = 1.9 \times 10^{-6}$). The population mean is represented by a white diamond, and quartile ranges are defined by the whisker boundaries and upper and lower box boundaries. Outliers are represented as open circles. The nearest neighbors of (H) CD20⁺ B cells, (I) Tfh cells, and (J) CD4⁺ T cells within B-T aggregates. The data set analyzed in this figure is the same as the data set introduced in Figure 5.

all other cell neighborhoods. CD4⁺ neighborhoods were closer to tubules than B-T neighborhoods but still significantly farther than the other non-B-T neighborhoods (Figure 8F). These data suggest that there is a preferential aggregation of large B-T neighborhoods near glomeruli, while other quantifiable neighborhoods are dispersed among tubules in the TI space. A representative image of this phenomenon can be seen in Figure 8G. These data suggest that cellular neighborhoods, first identified in discrete fields of view, are spatially organized within the renal cortex.

Discussion

Canonically, LN is thought of as arising from a systemic break in B cell tolerance that leads to glomerular antibody deposition and inflammation. This model, supported by large bodies of evidence in both humans and mice (10, 31), has led to clinical trials targeting B cells and Tfh cells (32–38). However, these efforts have yielded either incremental or no improvement over the standard of care. By quantifying cellular organization within confocal microscopy images using deep-learning algorithms, we demonstrated that high regional B cell density is associated with a good prognosis. Rather, it is CD4⁺ T cell populations, including CD8⁺, $\gamma\delta$, and other DN T cells, that are associated with refractory disease and progression to renal failure.

The CD4⁺ population was surprisingly heterogeneous. As expected, CD8-expressing cells were common. However, over 40% of the CD4⁺ cells did not express CD8. Of these, approximately 50% expressed the $\gamma\delta$ TCR. Reexamination of the AMP data confirmed the presence of intrarenal $\gamma\delta$ T cells within the DN T cells. We also observed a substantial population of CD4⁺CD8[−] T cells. These cells appear similar to previously described DN T cells that arise from CD8⁺ self-reactive T cells that have downregulated CD8 expression (39, 40).

It remains to be determined whether a specific CD4⁺ T cell population is associated with progression to renal failure or if these populations share pathogenic roles. Certainly, both CD8⁺ and $\gamma\delta$ T cells can be cytolytic and might provide complementary recognition of different classes of autoantigens (41–44). Intrarenal $\gamma\delta$ T cells have been implicated in chronic renal disease, though the mechanism is unclear (45, 46). The function of DN T cells is not known, but they might retain cytolytic activity, as they are derived from CD8⁺ T cells. Alternatively, they could be a source of inflam-

matory cytokines (39). Resolving heterogeneity in these CD4⁺ T cell populations is necessary to identify the populations most closely linked with ESRD.

In addition to yielding information on cell frequencies in tissue, our analytic pipelines provided precise positions of all cells assayed in the biopsy. This allowed us to define cellular neighborhoods and extract quantitative features, including neighborhood size, shape, and cell constituency. Unsupervised clustering revealed that in individual patients small neighborhoods of CD4⁺ T cells were associated with progression to ESRD. These data suggest that understanding immune cell architectures, even in relatively small patient sample cohorts, can identify prognostically important mechanisms.

Our data identified CD4⁺ T cells, including CD8⁺, $\gamma\delta$, and DN T cells, as potentially important therapeutic targets. This association was particularly striking in patients that presented in renal failure. It is possible that the inflammatory phenotype observed in these patients was not a primary state but arose as a secondary consequence of renal damage and scarring. Indeed, the patients who progressed to ESRD generally had higher chronicity scores than the patients who did not. However, 2 of the ESRD current patients (patients 51 and 55) had high densities of CD4⁺ T cells, high activity indices, and relatively low chronicity scores, suggesting that infiltrating CD4⁺ T cells can precede substantial renal damage. These data suggest that patients exhibit distinct, prognostically meaningful, intrarenal inflammatory trajectories.

Unfortunately, in contrast to the B cell/Tfh cell axis, we have limited therapeutic options that specifically target these T cell populations. One of the few classes available is calcineurin inhibitors. The recent success of the calcineurin inhibitor voclosporin in treating some patients with LN is promising (47). We propose that stratifying patients by the constituency and organization of their renal inflammation might identify those most likely to benefit from the addition of T cell-targeting therapies such as voclosporin.

There are several models that might account for the surprising association between dense regions of B cells and good renal outcome. First, B cells might have a previously unappreciated protective role in the context of renal inflammation (48, 49). Alternatively, it is possible that B cells and subsequent local antibody secretion are somewhat benign or neutral compared with other pathological processes in terms of tissue destruction. The final model is that dense B cell regions are responsible for tissue destruction in a subset of patients, but conventional therapies are effective at inhibiting this process. Indeed, most of our patients were treated with high-dose steroids and induced with cytotoxic therapies, most often mycophenolate. These therapies have been demonstrated to deplete B cells and plasma cells (50, 51).

Large neighborhoods of cells were enriched in both B cells and CD4⁺ T cells, including putative Tfh cells. These neighborhoods have similar features to the T/B aggregates described previously (15). However, our HMP image analysis indicated that these structures are more complex, containing other cell types, including other CD4⁺ T cell and plasma cell populations. These findings cohere with previous studies that suggest an underlying architecture to these large neighborhoods (13). Further work will be needed to understand the rules by which these different neighborhoods organize with respect to each other and the underlying biological processes governing their organization.

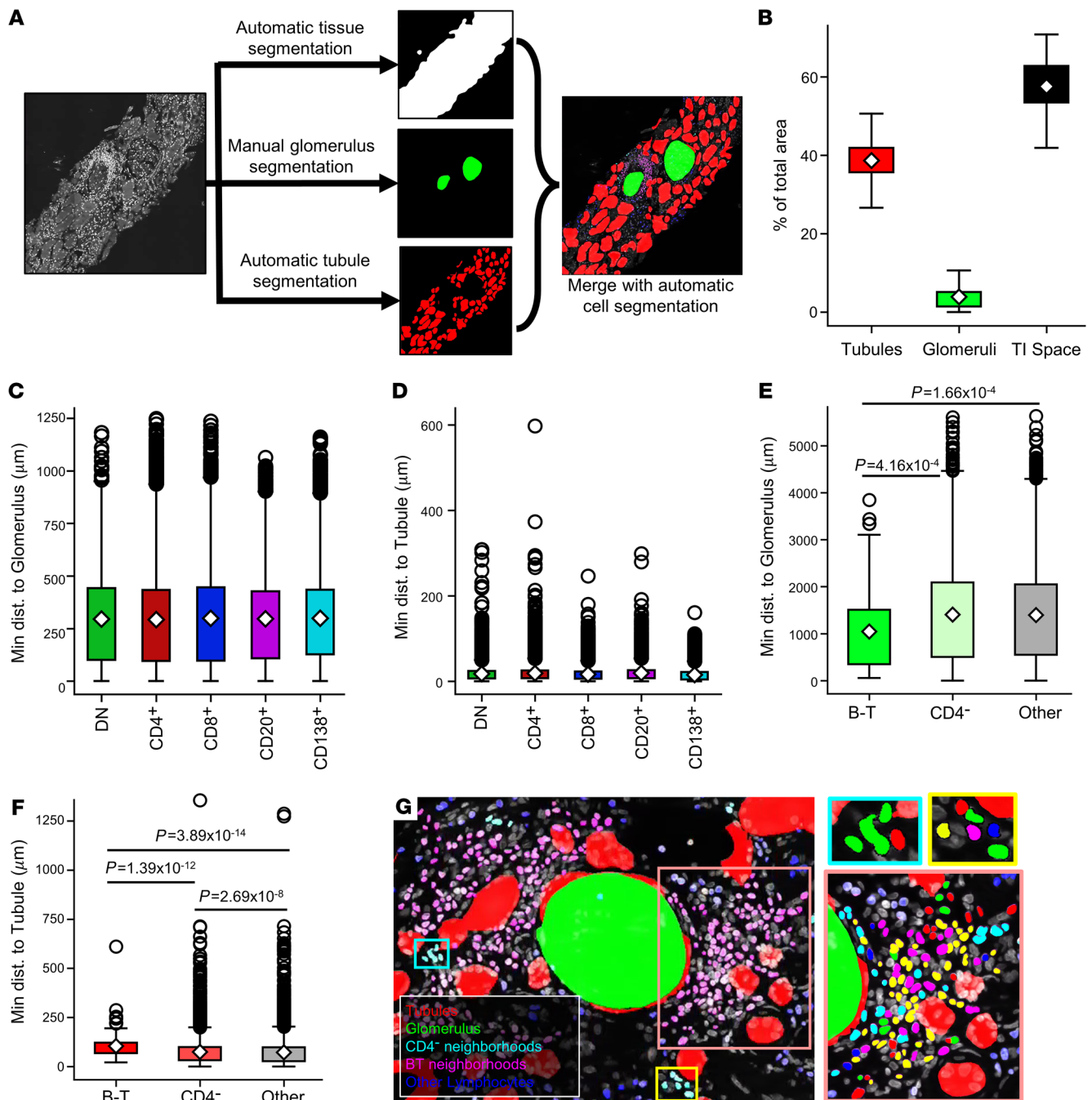


Figure 8. Structural segmentation reveals B-T neighborhood proximity to glomeruli. (A) Structural segmentation of biopsies required 3 steps: (a) automatic segmentation of the tissue in the composite was accomplished through filtering and thresholding the DAPI channel, (b) glomeruli were hand-segmented on the DAPI channel of each biopsy, and (c) a U-Net was trained to segment tubular structures (including both tubules and blood vessels) in 512×512 DAPI tiles. Structural segmentations were merged with the coordinate space of the detected lymphocytes to calculate proximity to kidney structures (magnification 63x, zoom factor 1 in A and G). (B) The TI space was the largest compartment, followed by tubular structures and then glomeruli. (C) Lymphocyte proximity to glomeruli varies slightly across detected classes. (D) Minimum distance of detected lymphocytes to a tubule segmentation also varies across class. Means and P values for all 2-way comparisons in C and D are reported in Supplemental Tables 4 and 5. (E) B-T neighborhoods were significantly closer to glomeruli than all other neighborhood classifications. (F) Conversely, B-T neighborhoods were significantly farther from tubular structures than all other aggregates. CD4⁺ neighborhoods were also significantly farther from tubules than all other non-B-T neighborhoods. The population mean is represented by a white diamond, and quartile ranges are defined by the whisker boundaries and upper and lower box boundaries. Outliers are represented as open circles. (E and F) Significant P values ($P < 0.05$) are noted on plots (Mann-Whitney U test with Bonferroni's correction). (G) Representative B-T and CD4⁺ neighborhoods and cell constituents (for zoom panels at right, green, DN T cells; red, non-Tfh CD4⁺ T cells; blue, CD8⁺ T cells; magenta, CD20⁺ cells; cyan, CD138⁺ cells; yellow, Tfh cells). The data set analyzed in this figure is the same as the data set introduced in Figure 5.

Interestingly, pDC prevalence was not associated with renal outcome in our HR data set. As sources of IFN- α , they have been postulated to play a central role in disease pathogenesis (52–55). However, the outcomes of clinical trials of anti-IFN- α antibodies in lupus have been modest (56, 57). In contrast to pDCs, mDCs were depleted in those patients that presented in renal failure. This was an unexpected finding, as CD1c⁺DC-SIGN⁺ DCs and CD141^{hi}CLEC9A⁺ DCs in the tubulointerstitium of patients with LN have previously been correlated with fibrosis and poor renal function (58). In addition, periglomerular inflammatory mDCs that drive pathology have been recently described in LN (59), though the mDCs in the tubulointerstitium do not appear to share this phenotype. However, tissue-resident DCs have also been appreciated for their role in enforcing peripheral tolerance (60). Therefore, a subset of DCs might have a role in organ tolerance, even in the context of inflammation.

Furthermore, CD11c is a marker that can capture a range of cells in addition to mDCs, including macrophages and even age-associated or DN B cells (61, 62). In our CNN hierarchy, putative cells expressing both CD20 and CD11c would be classified as B cells; such occurrences were rare. Further work, with multiple additional cell markers, will be needed to resolve the complexity of the CD11c populations.

Using deep-learning and other artificial intelligence algorithms, we achieved robust and accurate cell detection across multiple LN image data sets. This enabled a detailed, accurate spatial analysis of in situ adaptive immunity in LN samples. However, we must acknowledge the limitations of this approach — namely that additional work needs to be done to elucidate the pathological mechanisms at play. We have identified a series of striking associations, but we do not yet have insight into *how* some cells are driving progression to renal failure. For example, it is not feasible to directly test whether the CD4⁺ T cell populations identified in this work are poorly responsive to corticosteroids or mycophenolate therapy. Future work should be focused on validating these findings in a separate cohort of patients from a clinical trial of a B cell- or T cell-specific therapy (e.g., rituximab, voclosporin), which would allow the identified cellular states to be related to therapeutic responses.

Several computer vision methods were implemented to establish an analytical pipeline that addressed experimental, biological, and technical limitations. CNNs trained for instance segmentation detected and classified several immune cell classes with high fidelity not only in sparsely populated images, but also in densely packed images. In previous work, we trained and implemented a single Mask R-CNN for 5-class cell segmentation in LN images (21). By splitting this task into 2 separate networks, we were able to mitigate issues with cell class imbalance and variable stain signatures to improve overall detection performance (63). We also trained and implemented CNNs for rapid and robust image and object filtering to optimize immune cell calling in full biopsy sections. This included a network trained to discriminate between image tiles containing positive cell signal and tissue autofluorescence. We also trained a network to segment tubules in order to reject false-positive plasma cell predictions. By combining these CNNs with thresholding and image registration techniques, we automatically mapped several immune cell classes to full-biopsy sections, enabling a robust spatial analysis of in situ autoimmunity.

We implemented artificial intelligence to quantify immune cell populations in human tissue, thereby extracting rich, nonbiased, spatio-cellular data that allowed identification of unexpected pathogenic mechanisms. Even in a relatively small longitudinal cohort, we were able to resolve patient heterogeneity to identify putative pathogenic processes. Remarkably, specific cell densities provided powerful insights into disease pathogenesis. Understanding how these populations were organized into neighborhoods enabled us to associate cellular features of inflammation with prognosis. Further work needs to be done to identify whether these findings apply to other states of renal inflammation, such as mixed renal rejection or ANCA vasculitis. However, preliminary work indicates that the neural networks and approaches used for LN can be extended to other disease states, such as renal allograft rejection and rheumatoid arthritis. Therefore, our data suggest that using machine-learning-assisted spatial analysis to evaluate the complexity, heterogeneity, and organization of in situ inflammation will help to provide a more quantitative understanding of human autoimmunity. Such knowledge is critical for interpreting and applying the wealth of knowledge we have gained from animal models. It is also likely to identify both new therapeutic targets and those patients in which specific strategies are likely to be beneficial.

Methods

Sample staining and image acquisition – HR data set

FFPE kidney biopsies from 55 patients with LN with at least 2 years of clinical follow-up were obtained from the University of Chicago Human Tissue Resource Center (Supplemental Figure 8). FFPE sections were deparaffinized and treated with a citric acid buffer (pH 6.0) for antigen retrieval and blocked with serum prior to antibody staining. Samples were stained with indicated specific antibodies (Supplemental Table 6) and imaged on a Leica SP8 laser scanning confocal microscope at $\times 63$ magnification. Image ROIs were collected in tissue regions with identifiable CD3 signal. Collected images were 1024×1024 pixels $\times$ 6 channels with a $0.1058 \mu\text{m}$ pixel size.

Staining and image acquisition – HMP data set

Samples were stained using a strip and reprobe procedure in which $5 \mu\text{m}$ thick sections of FFPE biopsy sections were iteratively stained according to a procedure outlined by (ref. 26 and Supplemental Figure 8). Sections were deparaffinized and stained with a combination of primary antibodies and secondary antibodies conjugated with Alexa Fluor 488, 546, and 647 fluorophores (Supplemental Table 6). DAPI was included in every iteration of staining. Each round of staining, samples were imaged using a Caliber ID RS-G4 large-format confocal microscope at a magnification of $\times 63$, resulting in a pixel size of 221 nm . After each round of imaging, samples were stripped as described previously (26) and then reprobbed with a new set of primary and secondary antibodies and reimaged until the full marker panel had been imaged.

Staining and image acquisition – $\gamma\delta$ T cells

Eight LN kidney biopsies were stained for CD3, CD4, CD8, TCR δ , and DAPI. Inflamed regions were imaged on the Leica Stellaris 8 confocal microscope, with $\times 40$ magnification and a pixel size of $0.225 \mu\text{m}$. 281 ROIs (35 ± 19 per sample) were obtained and then postprocessed with background subtraction, despeckling, and contrast adjustment using ImageJ (NIH). Cells in these images were quantified by manual count.

Automatic cell detection and segmentation – HR data set

Two instances of a Mask R-CNN architecture (24) were trained to detect and segment cells in this data set. One instance was trained to detect 3 classes of lymphocytes (B cells, CD3⁺CD4⁺ T cells, and CD3⁺CD4⁺ T cells). A second instance was trained to detect 2 classes of DCs (pDCs and mDCs). For this work, ResNet-101 was used as a backbone for the Feature Pyramid Network. Networks were trained with a learning rate of 0.01. Training, validation, and testing data were generated by a single expert. All manually segmented images from a given patient were relegated to the training set (246 images [70%] from 21 unique patients), validation set (62 images [20%] from 4 unique patients), or test set (34 images [10%] from 6 unique patients). Training progress was monitored using Tensorboard, and training was stopped after cell recall stopped improving for all classes. Precision, recall, and F_1 score were used to evaluate network performance (see equations below). Specifically, these metrics were calculated at the object level, not the pixel level. A predicted object was defined as a true positive if it had an IOU score of greater than 0.25, with a ground truth cell of the same class, and if the network confidence in that predicted object was greater than 0.3. Cell predictions that had an IOU of less than 0.25 with a manual segmentation or had a disagreement in class with an overlapping manual segmentation were defined as false-positive predictions. All manual segmentations that were not matched with a true positive prediction were defined as false negatives. Because instance segmentation is first an object detection task, no true negative objects can be defined. All computation for the HR data set was performed using resources at the University of Chicago Research Computing Center. Each instance of Mask R-CNN was trained on a single GPU compute node containing 4 Nvidia GPUs with 12 GB memory per card, 28 Intel E5-2680v4 CPUs at 2.4 GHz, and 64 GB of system memory. A batch size of 4 images was used for training, distributed across the 4 GPU cards. Networks were trained to the point at which the recall for all cell classes stopped increasing.

$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}} \quad (\text{Equation 1})$$

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}} \quad (\text{Equation 2})$$

$$F_1 \text{ score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (\text{Equation 3})$$

Automatic cell detection and segmentation – HMP data set

Image strips (1024 × N pixels) from a Caliber ID microscope were stitched together using cross-correlation of image patches at the strip boundaries. These single-channel composites were then aligned with the DAPI channel from the first round of imaging, again through cross-correlation of image patches. Multichannel composites were then broken into 512 × 512 pixel image tiles. All DAPI tiles were passed through a simple image intensity filter to determine if tissue was present at a given location. All tiles at a given location were filtered out of proceeding analyses if the DAPI filter revealed no tissue at that location. This resulted in over unique 18,000 tile locations, with each location containing 9 unique stains.

Instance segmentation of cells was split into two tasks: instance segmentation of T cells (also referred to as dual-marker detection) and instance segmentation of B cells (also referred to as single-marker detection). Instance segmentation performance was evaluated using precision (Equation 1), recall (Equation 2), and F_1 score (Equation 3).

T cell segmentation. An instance of Mask R-CNN was pretrained to segment single-positive (CD3⁺CD4⁺) and double-positive (CD3⁺CD4⁺) T cells in 512 × 512 pixel image patches from T cell image stacks (CD3/CD4/DAPI) from the HR data set. This pretrained network was fine-tuned with a small set of 211 T cell image stacks from the HMP data set. The HMP fine-tuning image sets contained CD3/CD4/DAPI, CD3/CD8/DAPI, and CD3/ICOS/DAPI image stacks. This fine-tuning set was split into training (169 images, 80%), validation (21 images, 10%), and testing (21 images, 10%) sets, with all images from a given patient confined to a specific set. For fine-tuning, weights were permitted to adjust for all convolutional, max-pooling, and fully connected layers of the pretrained Mask R-CNN. The fine-tuned T cell network was used to make predictions for CD3/CD4/DAPI, CD3/CD8/DAPI, CD3/ICOS/DAPI, and CD3/PD1/DAPI images. The trained dual-marker network had an average F_1 score of 0.85 on all single-positive cell predictions (i.e., CD3⁺CD4⁺ and CD3⁺CD8⁺) and an average F_1 score of 0.92 on all double-positive cell predictions (i.e., CD3⁺CD4⁺ and CD3⁺CD8⁺).

B cell segmentation. An instance of Mask R-CNN was pretrained to segment B cells in 512 × 512 pixel image patches generated from B cell image stacks (CD20/DAPI) from the HR data set. This pretrained network was fine-tuned with a set of 79 B cell image stacks from the HMP data set. This fine-tuning set was split into training (63 images, 80%), validation (8 images, 10%), and testing (8 images, 10%) sets, with all images from a given patient confined to a specific set. For fine-tuning, the weights were permitted to be adjusted for all convolutional, max-pooling, and fully connected layers of the pretrained Mask R-CNN. The fine-tuned B cell network was used to make predictions for B cell (CD20) and plasma cell images (CD138). The trained single-marker network had an average F_1 score of 0.87.

Tubule segmentation

An instance of Mask R-CNN was trained to segment tubular structures, including blood vessels and tubules, in the HMP data set. 300 DAPI tiles (512 × 512 pixels) from 18 patients were manually annotated by a single expert. Manually segmented images were separated into training, validation, and testing sets as follows: 240 images in the training set (80%), 30 images in the validation set (10%), and 30 images in the test set (10%). Data augmentation consisted of random horizontal and vertical flips and rotations. Performance of the tubule segmentation network was assessed at the pixel level, with the trained network yielding an average recall (Equation 2) of 0.74 and an average precision (Equation 1) of 0.79 on the test set of tubule images.

All computation associated with the HMP data set was performed on the MEL computational server in the Radiomics and Machine Learning Facility at the University of Chicago. MEL contains 256 Xeon Gold 6130 CPU cores, 3 TB of DDR4 ECC RAM memory, 24 TB of NVMe SSD storage, and 16 Nvidia Tesla V100 32GB GPU accelerators.

Defining cell neighborhoods through density-based clustering

Cells in both data sets were assigned to clusters using the sklearn (version 0.23.2) implementation of Density-Based Spatial Clustering of Applications with Noise (DBSCAN; ref. 25), using an epsilon of roughly 10 μ m, corresponding to 100 pixels in the HR data set and 50 pixels in the HMP data set and a minimum cluster size of 2. In the HR data set, 24 features of cellular constituency and cell/neighborhood shape were extracted for each cluster, and K-means clustering was then applied to define classes of neighborhoods. The neighborhoods were split into 6 classes, as deter-

mined ideal by bootstrapping cluster descriptors, including the WCSS and the Δ WCSS. The types were characterized using a leave-one-out t test to identify which features of each type of neighborhood distinguished it from the other neighborhoods. In this procedure, the current cluster of reference was treated as the alternative group, all remaining clusters were then binned together as the reference group, and then a t test for was applied for all features used to describe the neighborhoods.

Structural segmentation of kidney biopsies

Full-biopsy segmentation was performed on the DAPI channel for all kidney biopsies. The fully stitched DAPI composite was γ -corrected to increase brightness and then filtered with a Gaussian kernel. The filtered image was normalized and thresholded to create a binary mask of all tissue in the composite. Glomerular segmentation was performed by hand. Due to the relatively low prevalence of glomeruli in this data set, there was not enough glomeruli to train an algorithm to define glomeruli. Tubular structures, including both tubules and blood vessels, were segmented on the DAPI channel of all samples using the Mask R-CNN trained for tubule segmentation (as described above). No markers were included in the staining panel to explicitly identify tubules and blood vessels, and the signatures of these 2 structures are very similar in the DAPI channel. Therefore, the “tubule” network is trained to segment general tubular structures, which include proximal tubules, distal tubules, and blood vessels.

Spatial analyses

All other spatial analyses were performed using Python (3.7.9) and the following packages: pandas (1.2.2) (<https://pandas.pydata.org/>), numpy (1.19.2) (<https://numpy.org/>), sklearn (0.23.2) (<https://scikit-learn.org/stable/>), scipy (1.6.1) (<https://scipy.org/>), and tiff file (2021.1.14) (<https://pypi.org/project/tiff file/2021.1.14/>). Plotting was performed with matplotlib (3.3.2) (<https://matplotlib.org/>) and seaborn (0.11.1) (<https://seaborn.pydata.org/>). The nearest-neighbors calculation was performed by iterating through every cell in the data set and identifying the class of the closest cell by centroid-to-centroid distance.

In the HMP data set, coordinates of the cells in the tiles were adjusted to a composite-level coordinate system by shifting the tile-level coordinates based on the location of the tile in the composite. All subsequent calculations around the distribution of cells in tissue were based on these composite-level locations.

RNA-Seq analysis

scRNA-Seq data for human LN tissue were obtained from the ImmPort repository (accession code SDY997, “SDY997_EXP15176_celseq_matrix_ru10_molecules.tsv” raw data file). Quality control was performed according to Arazi et al. (16), such that cells were removed from the analysis if they expressed fewer than 1000 or more than 5000 genes or if more than 25% of the total UMIs mapped to mitochondrial genes. Gene expression values were normalized to library size (UMI count per million) and scaled by $\log_2$. Clustering implemented in Seurat 3.2.2 and canonical marker expression were used to identify cellular subsets. T cells were analyzed if they were assigned to the “naive T” or “CTL” clusters. T cells were categorized based on *CD4*, *CD8A*, and *CD8B* expression. Cells were categorized as “CD4” when they had detectable expression of *CD4* transcripts but no *CD8A* or *CD8B*. They were instead categorized as “CD8” when they had detectable *CD8A* and/or *CD8B*

expression with no detectable *CD4* transcripts. Cells were categorized as double-positive or DN when they had both/neither *CD4* and/or *CD8A/B*. t-SNE was performed by Rtsne (0.15). Plots were generated by ggplot2 (3.3.2) and ggridges (0.5.2).

Data and materials availability

Code used for the cellular segmentation and spatial analysis can be found at the following repository: https://github.com/durkeems13/LN_image_analysis (commit id: 9281d32).

Statistics

Differences between the groups of patients or groups of cellular neighborhoods were evaluated using the Mann-Whitney U test with the Bonferroni's correction for multiple comparisons. Differences in proportions were evaluated using the χ^2 test for independence with Bonferroni's correction for multiple comparisons. Data are shown as the mean $\pm$ SEM.

The bootstrapping analysis to validate the differences between the groups of patients in the HR data set was performed by sampling the individual pools of ROIs with replacement to produce samples of 200 ROIs for each of the patient cohorts in the 2-group analysis (ESRD⁺ vs. ESRD⁻) and 150 ROIs for each of the patient cohorts in the 3-group analysis (ESRD⁺ vs. ESRD⁻ vs. ESRD current). For each pair of groups, the mean for each random sample was calculated as well as the difference between the 2 group means. This procedure was repeated 1000 times to produce distributions of the difference in means between the groups. The 2 populations were considered to be significantly different from each other if the 95% confidence interval of the differences in means that did not overlap with 0.

Study approval

This study, which used only deidentified human samples, was approved by the University of Chicago Institutional Review Board, with protocol number 15065.

Author contributions

RA, MSD, and MRC conceptualized the study. RA, MSD, JA, MV, GC, YA, MLG, and MRC developed the methodology. RA, MSD, GC, and YA developed software. RA, MSD, AC, KK, EP, and CO provided clinical data. MRC and MLG acquired funding. MRC and MLG supervised the study. RA, MSD, JA, MV, GC, YA, and MRC wrote the original draft of the manuscript, and RA, MSD, JA, MV, GC, YA, MLG, and MRC reviewed, and edited the manuscript. First authorship was split between RA and MSD, with RA named first, because MSD led the development and implementation of the computer vision techniques, and RA led the spatial analyses of cells that comprise the bulk of the biological findings in the results section.

Acknowledgments

Special thanks to Karen Drukker for her input on statistical analyses, Chun-Wai Chan for computational support, and the University of Chicago Research Computing Center for additional computational support and resources. All imaging was performed at the University of Chicago Integrated Light Microscopy Facility. These studies were funded by the NIH Autoimmunity Centers of Excellence (AI082724), Department of Defense (LRI180083), Alliance for Lupus Research, and NIH awards (S10-OD025081, S10-RR021039, and P30-CA14599).

Address correspondence to: Marcus Clark, Section of Rheumatology, MC0930, University of Chicago, 5841 S. Maryland Ave., Chicago, Illinois 60637, USA. Phone: 773.702.0202; mclark@uchicago.edu. Or to: Maryellen Giger or Madeleine Durkee, Department

of Radiology, MC 2026, University of Chicago, 5841 S. Maryland Avenue, Chicago, Illinois 60637, USA. Phone: 773.702.6778; Email: m-giger@uchicago.edu (M. Giger) Phone: 702.506.6249; Email: durkeems@uchicago.edu (M. Durkee).

1. Tan EM, Kunkel HG. Characteristics of a soluble nuclear antigen precipitation with sera of patients with systemic lupus erythematosus. *J Immunol.* 1966;96(3):464–471.
2. Dubois EL, ed. *Lupus Erythematosus*. 2 ed. W.B Saunders; 1974.
3. Mohan C, Putterman C. Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nat Rev Nephrol.* 2015;11(6):329–341.
4. McGaha TL, Madaio MP. Lupus nephritis: animal modeling of a complex disease syndrome pathology. *Drug Discov Today Dis Models.* 2014;11:13–18.
5. Markowitz G, D'Agati VD. Classification of lupus nephritis. *Curr Opin Nephrol Hypertens.* 2009;18(3):220–225.
6. Hsieh C, et al. Predicting outcomes of lupus nephritis with tubulointerstitial inflammation and scarring. *Arthritis Care Res (Hoboken).* 2011;63(6):865–874.
7. Esdaile JM, et al. The clinical and renal biopsy predictors of long-term outcome in lupus nephritis: a study of 87 patients and review of the literature. *Q J Med.* 1989;72(269):779–833.
8. Yu F, et al. Tubulointerstitial lesions of patients with lupus nephritis classified by the 2003 International Society of Nephrology and Renal Pathology Society system. *Kidney Int.* 2010;77(9):820–829.
9. Schwartz M, et al. Pathologic changes in the renal tubule in systemic lupus erythematosus. *Hum Pathol.* 1982;13(6):534–547.
10. Chang A, et al. Cellular aspects of the pathogenesis of lupus nephritis. *Curr Opin Rheumatol.* 2021;33(2):197–204.
11. Kinloch AJ, et al. Anti-vimentin antibodies: a unique antibody class associated with therapy-resistant lupus nephritis. *Lupus.* 2020;29(6):569–577.
12. Liarski V, et al. Quantifying in situ adaptive immune cell cognate interactions in humans. *Nat Immunol.* 2019;20(4):503–513.
13. Liarski V, et al. Quantitative cell distance mapping in human nephritis reveals organization of in situ adaptive immune responses. *Sci Trans Med.* 2014;6(230):230ra46.
14. Kinloch AJ, et al. Vimentin is a dominant target of in situ humoral immunity in human lupus tubulointerstitial nephritis. *Arthritis Rheumatol.* 2014;66(12):3359–3370.
15. Chang A, et al. In situ B cell-mediated immune responses and tubulointerstitial inflammation in human lupus nephritis. *J Immunol.* 2011;186(3):1849–1860.
16. Arazi A, et al. The immune cell landscape in kidneys of patients with lupus nephritis. *Nat Immunol.* 2019;20(7):902–914.
17. Der E, et al. Tubular cell and keratinocyte single-cell transcriptomics applied to lupus nephritis reveal type I IFN and fibrosis relevant pathways. *Nat Immunol.* 2019;20(7):915–927.
18. Ko K, et al. Bcl-2 as a therapeutic target in human tubulointerstitial inflammation. *Arthritis Rheumatol.* 2016;68(11):2740–2751.
19. Amgad M, et al. Joint region and nucleus segmentation for characterization of tumor infiltrating lymphocytes in breast cancer. *Proc SPIE Int Soc Opt Eng.* 2019;10956:109560M.
20. Durkee MS, et al. Artificial intelligence and cellular segmentation in tissue microscopy images. *Am J Pathol.* 2021;191(10):1693–1701.
21. Durkee MS, et al. Quantifying the effects of biopsy fixation and staining panel design on automatic instance segmentation of immune cells in human lupus nephritis. *J Biomedical Optics.* 2021;26(2):022910.
22. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: challenges and opportunities. *Med Image Anal.* 2016;33:170–175.
23. Lewis SM, et al. Spatial omics and multiplexed imaging to explore cancer biology. *Nat Methods.* 2021;18(9):997–1012.
24. He K, et al. Mask R-CNN. Arxiv. Cornell University. <https://doi.org/10.48550/arXiv.1703.06870>. Published March 20 2017. Updated January 24, 2018. Accessed June 21 2022.
25. Schubert E, et al. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Trans Database Syst.* 2017;42(3):1–21.
26. Bolognesi MM, et al. Multiplex staining by sequential immunostaining and antibody removal on routine tissue sections. *J Histochem Cytochem.* 2017;65(8):431–444.
27. McKinney EF, et al. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. *Nature.* 2015;523(7562):612–616.
28. Tilstra JS, et al. Kidney-infiltrating T cells in murine lupus nephritis are metabolically and functionally exhausted. *J Clin Invest.* 2018;128(11):4884–4897.
29. Bocharnikov AV, et al. PD-1hiCXCR5⁺ T peripheral helper cells promote B cell responses in lupus via MAF and IL-21. *JCI Insight.* 2019;4(20):e130062.
30. Rao DA, et al. Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature.* 2017;542(7639):110–114.
31. Liu Z, Davidson A. Taming lupus—a new understanding of pathogenesis is leading to clinical advances. *Nat Med.* 2012;18(6):871–882.
32. Dooley MA, et al. Effect of belimumab treatment on renal outcomes: results from the phase 3 belimumab clinical trials in patients with SLE. *Lupus.* 2013;22(1):63–72.
33. Mysler EF, et al. Efficacy and safety of ocrelizumab in active proliferative lupus nephritis: results from a randomized, double-blind, phase III study. *Arthritis Rheum.* 2013;65(9):2368–2379.
34. Rovin BH, et al. Efficacy and safety of rituximab in patients with active proliferative lupus nephritis: the Lupus Nephritis Assessment with Rituximab study. *Arthritis Rheum.* 2012;64(4):1215–1226.
35. Almaani S, Rovin BH. B-cell therapy in lupus nephritis: an overview. *Nephrol Dial Transplant.* 2019;34(1):22–29.
36. Levy RA, et al. 10 Years of belimumab experience: what have we learnt? *Lupus.* 2021;30(11):1705–1721.
37. Atisha-Fregoso Y, et al. Phase II randomized trial of rituximab plus cyclophosphamide followed by belimumab for the treatment of lupus nephritis. *Arthritis Rheumatol.* 2021;73(1):121–131.
38. Furie R, et al. Two-year, randomized, controlled trial of belimumab in lupus nephritis. *N Engl J Med.* 2020;383(12):1117–1128.
39. Li H, et al. Systemic lupus erythematosus favors the generation of IL-17 producing double negative T cells. *Nat Commun.* 2020;11(1):2859.
40. Crispin JC, et al. Expanded double negative T cells in patients with systemic lupus erythematosus produce IL-17 and infiltrate the kidneys. *J Immunol.* 2008;181(12):8761–8766.
41. Paul S, et al. Role of gamma-delta ($\gamma\delta$) T cells in autoimmunity. *J Leukoc Biol.* 2015;97(2):259–271.
42. Bank I. The role of gamma delta T cells in autoimmune rheumatic diseases. *Cells.* 2020;9(2):462.
43. Kaminski H, et al. Unconventional T cells and kidney disease. *Nat Rev Nephrol.* 2021;17(12):795–813.
44. Deng Q, et al. The emerging epigenetic role of CD8⁺T cells in autoimmune diseases: a systematic review. *Front Immunol.* 2019;10:856.
45. Law BM, et al. Effector $\gamma\delta$ T cells in human renal fibrosis and chronic kidney disease. *Nephrol Dial Transplant.* 2019;34(1):40–48.
46. Yin S, et al. Hyperactivation and in situ recruitment of inflammatory V δ 2 T cells contributes to disease pathogenesis in systemic lupus erythematosus. *Sci Rep.* 2015;5:14432.
47. Rovin BH, et al. Efficacy and safety of voclosporin versus placebo for lupus nephritis (AURORA 1): a double-blind, randomised, multicentre, placebo-controlled, phase 3 trial. *Lancet.* 2021;397(10289):2070–2080.
48. Rosser EC, Mauri C. Regulatory B cells: origin, phenotype, and function. *Immunity.* 2015;42(4):607–612.
49. Zhang B, et al. B cell-derived GABA elicits IL-10(+) macrophages to limit anti-tumour immunity. *Nature.* 2021;599(7885):471–476.
50. Fassbinder T, et al. Differential effects of cyclophosphamide and mycophenolate mofetil on cellular and serological parameters in patients with systemic lupus erythematosus. *Arthritis Res Ther.* 2015;17(1):92.
51. Bancheau R, et al. Personalized immunomonitoring uncovers molecular networks that stratify lupus patients. *Cell.* 2016;165(3):551–565.
52. Blanco P, et al. Induction of dendritic cell differentiation by IFN- α in systemic lupus erythematosus. *Science.* 2001;294(5546):1540–1543.
53. Baechler EC, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci U S A.* 2003;100(5):2610–2615.
54. Niewold TB, et al. Interferon alpha in system-

- ic lupus erythematosus. *J Biomed Biotechnol.* 2010;2010:948364.
55. Furie RA, et al. Type I interferon inhibitor anifrolumab in active systemic lupus erythematosus (TULIP-1): a randomised, controlled, phase 3 trial. *Lancet Rheumatol.* 2019;1(4):208–219.
 56. Khamashta M, et al. Sifalimumab, an anti-interferon- α monoclonal antibody, in moderate to severe systemic lupus erythematosus: a randomised, double-blind, placebo-controlled study. *Ann Rheum Dis.* 2016;75(11):1909–1916.
 57. Morand EF, et al. Trial of anifrolumab in active systemic lupus erythematosus. *N Engl J Med.* 2020;382(3):211–221.
 58. Kassianos AJ, et al. Increased tubulointerstitial recruitment of human CD141(hi) CLEC9A(+) and CD1c(+) myeloid dendritic cell subsets in renal fibrosis and chronic kidney disease. *Am J Physiol Renal Physiol.* 2013;305(10):F1391–F1401.
 59. Parikh SV, et al. A novel inflammatory dendritic cell that is abundant and contiguous to T cells in the kidneys of patients with lupus nephritis. *Front Immunol.* 2021;12:621039.
 60. Kim SJ, Diamond B. Modulation of tolerogenic dendritic cells and autoimmunity. *Semin Cell Dev Biol.* 2015;41:49–58.
 61. Kurts C, et al. Kidney dendritic cells: fundamental biology and functional roles in health and disease. *Nat Rev Nephrol.* 2020;16(7):391–407.
 62. Jenks SA, et al. Distinct effector B cells induced by unregulated toll-like receptor 7 contribute to pathogenic responses in systemic lupus erythematosus. *Immunity.* 2018;49(4):725–739.
 63. Durkee MS, et al. Managing class imbalance and differential staining of immune cell populations in multi-class instance segmentation of multiplexed immunofluorescence images of Lupus Nephritis biopsies. *SPIE Proceedings.* 2021;11603:1160313.



MINI-REVIEW

Artificial Intelligence and Cellular Segmentation in Tissue Microscopy Images



Madeleine S. Durkee,^{*} Rebecca Abraham,[†] Marcus R. Clark,[†] and Maryellen L. Giger^{*}

From the Department of Radiology and the Committee on Medical Physics,^{*} and the Department of Medicine,[†] Section of Rheumatology and Gwen Knapp Center for Lupus and Immunology Research, University of Chicago, Chicago, Illinois

Accepted for publication
May 17, 2021.

Address correspondence to
Madeleine S. Durkee, Ph.D.,
Department of Radiology, Uni-
versity of Chicago, 5841 S.
Maryland Ave, Chicago, IL,
60637; or Maryellen L. Giger,
Ph.D., Department of Radi-
ology, University of Chicago,
5841 S. Maryland Ave.,
MC2026, Chicago, IL
60637. E-mail: durkeems@uchicago.edu or m-giger@uchicago.edu.

With applications in object detection, image feature extraction, image classification, and image segmentation, artificial intelligence is facilitating high-throughput analysis of image data in a variety of biomedical imaging disciplines, ranging from radiology and pathology to cancer biology and immunology. Specifically, a growth in research on deep learning has led to the widespread application of computer-visualization techniques for analyzing and mining data from biomedical images. The availability of open-source software packages and the development of novel, trainable deep neural network architectures has led to increased accuracy in cell detection and segmentation algorithms. By automating cell segmentation, it is now possible to mine quantifiable cellular and spatio-cellular features from microscopy images, providing insight into the organization of cells in various pathologies. This mini-review provides an overview of the current state of the art in deep learning— and artificial intelligence—based methods of segmentation and data mining of cells in microscopy images of tissue. (*Am J Pathol* 2021, 191: 1693–1701; <https://doi.org/10.1016/j.ajpath.2021.05.022>)

Machine learning has led to rapid advancements in the biomedical sciences, particularly image quantitation. Deep learning, a branch of machine learning focused on artificial neural networks, is driving the growth of research in the imaging sciences, radiomics, and computational pathology. In conjunction with more conventional machine-learning methods, deep learning has facilitated the automation of image-analysis tasks such as cell classification and tissue-type identification, thereby reducing the time and effort needed for analyzing microscopy data. This mini-review discusses the recently developed applications of artificial intelligence (AI) to digital pathology and cell-image analysis, focusing on novel machine-learning and deep-learning methods and associated challenges.

Artificial Intelligence in Digital Pathology

Computer visualization and machine learning have been applied in medical imaging for decades. Early computer-visualization techniques in medical imaging were developed

for application in radiography and have since been widely applied in histopathology.^{1,2} Digital pathology produces a wealth of complex data, making automated, high-throughput analysis desirable. Manual annotation of

Supported by National Institute of Allergy and Infectious Diseases (NIH) grants U19 AI082724 (M.R.C.), R01 AR055646 (M.R.C.), R01 AI148705 (M.R.C.), and U01 CA195564 (M.L.G.); and US Department of Defense grant LR180083 (M.R.C.). The content is the responsibility of the authors and does not necessarily represent the official views of the NIH.

Disclosures: Maryellen L. Giger holds stock options in and receives royalties from Hologic, Inc.; holds equity in and is a co-founder of Quantitative Insights, Inc. (now, Qlarity Imaging); holds stock options in QView Medical, Inc.; and receives royalties from General Electric Company, Median Technologies, Riverain Technologies LLC, Mitsubishi, and Toshiba. It is the University of Chicago's conflict of interest policy that investigators disclose publicly actual or potential significant financial interests that would reasonably appear to be directly and significantly affected by the research activities.

This article is part of a mini-review series on the applications of artificial intelligence and deep learning in advancing research and diagnosis in pathology.

cell-level features (eg, surface marker expression) and tissue-level features (eg, tumor boundaries) from microscopy data is a resource-intensive task that usually requires time and the experience of a trained pathologist. The extraction of quantitative information from image data sets for research purposes has historically been limited by the availability and participation of these clinical experts. AI provides a means for high-throughput, standardized, quantitative analysis of pathology images.^{3–5}

The development of early machine-learning applications in medical imaging was facilitated by digitization, which allowed for mathematical operations to be performed on image data.^{1,4} For lesion grading and assessment in radiology and pathology, imaging specialists worked in tandem with clinicians to identify the qualitative features recognizable to the well-trained eye, then mathematically designed filters to enhance such characteristics, leading to computerized lesion detection and segmentation. Once segmented, a lesion can be quantitatively described using human-engineered features such as size, shape, pixel intensity, and local texture.^{4,6} Texture analysis and lesion classification are two of the early building blocks of computational and digital pathology and have been applied in clinical tasks such as diagnosis and grading, and in studies of disease pathogenesis.^{4,6}

Deep learning was widely adopted in the imaging sciences after AlexNet, a deep convolutional neural network (DCNN), achieved a 15% error rate (nearly 10% better than its competitors) on the ImageNet Large Scale Visual Recognition Challenge.⁷ The implications for medical image analysis were immediately apparent. Within a year, multiple studies were published in which DCNNs were applied in medical image analysis, followed by continued growth in the application of deep learning to biomedical image analysis.^{8,9} Since the success of AlexNet, a variety of DCNN architectures have been developed for specialized medical and biological imaging tasks, ranging from lesion detection in magnetic resonance imaging to cell segmentation in microscopy.^{4,10} DCNNs recognize patterns in raw image data without the constraints of human-defined equations. These patterns may not be interpretable by human observers but are determined by the network to be the features most useful in making robust and accurate decisions in image classification or segmentation.

Deep Learning in Multiplexed Microscopy

Deep learning facilitates the rapid analysis of high-content, experimental imaging modalities, including multiplexed microscopy. Multiplexed microscopy refers to the use of image data in which the spatial distribution of multiple biological markers, such as cell surface proteins, has been captured via immunohistochemistry analysis or immunofluorescence. Previously, these type of data were either analyzed manually or used only for the purpose of

presenting representative examples of biological phenomena. DCNNs have been particularly useful for automating the segmentation and classification of regions, structures, and cells in tissue.¹⁰ This automation has facilitated previously intractable quantification of microscopy data.

Simultaneously, multiplexed microscopy has advanced to the point of generating novel computational challenges. Improved microscope designs and staining protocols have greatly increased the number of markers that can be captured in a given sample, allowing for co-localization of upward of 40 markers per individual frame.^{11–13} This level of phenotypic resolution was previously accessible only through tissue-destructive methods such as flow cytometry and RNA sequencing. The destructive processing required for these methods does not conserve the spatial arrangement of cells and other tissue structures. In contrast, highly multiplexed imaging produces rich data sets that include detailed information on phenotype and cell spacing. Novel AI solutions have been implemented in the location and segmentation of these phenotype-rich images.^{11,13}

Cell Detection and Segmentation in Tissue Microscopy Images

A key application of machine learning in tissue microscopy is in the automation of the identification of structures and individual cells within images. Cell segmentation is achieved through iterative training of AI computer-visualization algorithms (Figure 1). These trained models can then be deployed for predicting cell segmentation in new data, allowing for high-throughput mining of quantitative descriptors of tissue pathology. This section describes deep-learning architectures and open-source programs used for performing cell segmentation in microscopy data.

DCNN Architectures for Analyzing Cells in Microscopy Images

The detection and segmentation of cells in *ex vivo* tissue allow for high-throughput quantification of cell features, including cell frequency, cell morphology, cell-specific signal intensity, and spatial distribution of cells. Biomedical image segmentation is often conducted using U-Net or similarly structured architectures, which provide pixel-level classification, referred to as *semantic segmentation*.¹⁴ However, these methods can underperform in dense regions of cells. Object-detection and semantic-segmentation tasks can be combined to perform *instance segmentation*, which can be accomplished through object-based detection networks such as Mask R-CNN or by combining a U-Net with a region-proposal network.^{15,16} Image-translation methods have also shown potential in segmentation of microscopy images.¹⁷ These generative methods can also improve image quality by converting the image to a higher resolution or converting between image modalities.^{18,19}

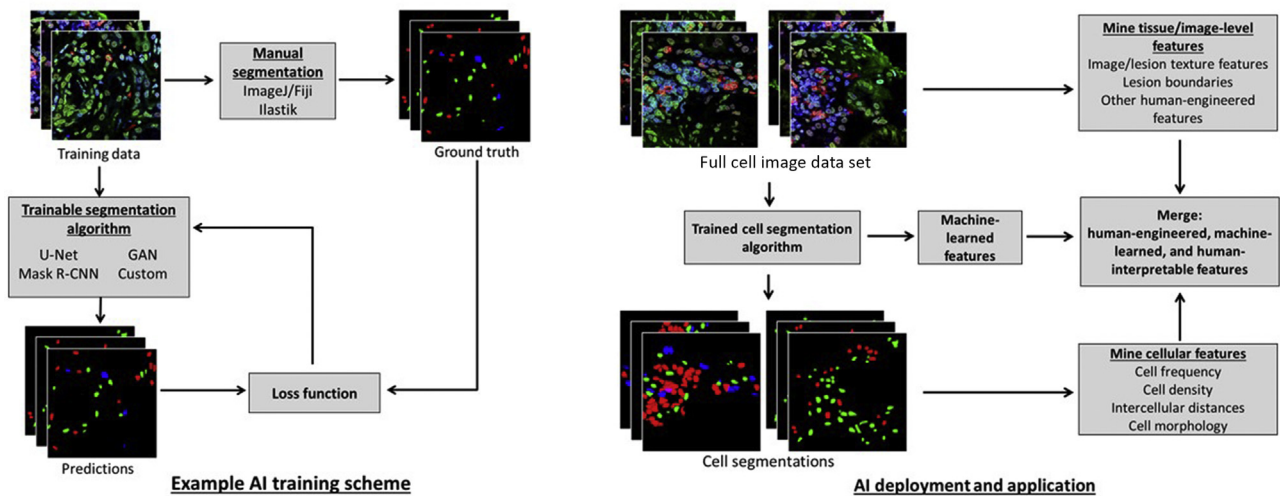


Figure 1 Supervised AI computer-vision algorithms are trained using manual ground truth to achieve accurate cell segmentations on a specific domain of image data (left). A trained algorithm can then be deployed to automatically extract quantitative features from microscopy data to answer clinic- and biology-related questions (right). CNN, convolutional neural network; GAN, generative adversarial networks. Fiji, NIH, Bethesda, MD; <https://imagej.nih.gov/ij>. Ilastik; <http://ilastik.org>. ImageJ, NIH; <http://imagej.nih.gov/ij>.

Semantic Segmentation

Encoder–decoder architectures have dominated biomedical image segmentation since the advent of U-Net and its derivatives. These have been applied to semantic segmentation tasks in medical imaging across scales, with excellent performance on full-organ segmentation in computed tomography and nuclear segmentation in high-resolution microscopy.^{14,20,21} The contracting, or *encoder*, portion of the architecture captures contextual features within the image, while the expanding, or *decoder*, portion of the architecture generates precise localization. In pathology images, semantic segmentation has been used to differentiate between tumor and healthy tissue, or segment pathogenic or prepathogenic areas of tissue. Cell and cell-nucleus segmentation can also be performed through semantic segmentation schemes. Quantitative characteristics acquired from these segmentations, such as nucleus-to-cytoplasm ratio, can be indicative of cancer grade.²² However, semantic segmentation of cells can fail in crowded regions or in images with a low signal-to-noise or signal-to-background ratio.

Instance Segmentation

While semantic segmentation alone is not always sufficient for accurate cell counting in images, it can be combined with object-detection methods to generate object-level segmentation of cells rather than image-level segmentation. This combined task, called *instance segmentation*, generates pixel-level segmentation of individual cells in an image, which allows for the separation of clustered or overlapping cells in an image, resulting in improved cell-frequency data. Region-based CNNs from the Fast R-CNN family perform

multi-object, multi-class segmentation with high accuracy in natural images, and have also shown promise in biomedical image analysis tasks, including cell segmentation in a variety of contexts.^{23,24} Additionally, region-proposal networks have been incorporated into U-Net architectures to perform object-level cell segmentation in biomedical images.^{15,25} NuSeT, a U-Net + region proposal network architecture, has shown high accuracy in nuclear segmentation in a variety of contexts, including segmentation in images from different modalities and different tissues and pathologies.¹⁵

Generative Networks for Segmentation

Generative adversarial networks (GANs) have also been successfully implemented in segmentation in biological images.¹⁷ However, GANs take a different approach to image segmentation. A typical GAN architecture consists of two networks, a generator and a discriminator, which are trained competitively to generate a simulated version of the input image as it would present in a different context, for example translating a daytime landscape image to appear as if it were taken at night. GAN segmentation works by converting an input image into the mask domain, in which each class is represented as a specific pixel value. In addition to image segmentation, image conversion can also mitigate generalizability-related problems in deep learning. de Bel et al¹⁷ have shown the ability of GANs to normalize staining variability across multiple institutions and multiple tissue types. Additionally, generative architectures have been trained to rectify image artefacts, to de-noise low signal-to-noise ratio images, and to generate isotropic resolution in three dimensions.¹⁹ These de-noising and

resolution adjustments have led to improved segmentation of cells in three-dimensional images, videos, and image projections. Keikhosravi et al¹⁸ have trained generative networks to perform cross-modality synthesis on images of hematoxylin and eosin (H&E) staining to produce second harmonic generation images of collagen organization in biopsy samples. A second harmonic generation system is not feasible to operate in the clinic, but by converting standard H&E images, deep-learning methods can provide more detailed information on samples.

Examples of Current Open-Source Software

Several open-source implementations of machine-learning tools have been designed for scientists who may not have the extensive background in programming required for low-level algorithm control. Many of these programs employ pretrained networks for common tasks such as nuclear segmentation. Some enable the training of machine-learning models from scratch. One of the benefits of these open-source pipelines is that they include graphical user interfaces, making them user-friendly and accessible to a broader audience. However, the user-friendly environment comes at the expense of the user losing low-level control of the algorithm, making task-specific optimization via hyperparameter tuning or transfer learning challenging.

ImageJ

ImageJ (NIH, Bethesda, MD; <http://imagej.nih.gov/ij/>; alias Fiji) is an open-source tool that has been used extensively for many image-analysis tasks. Users frequently build and share plug-ins that can be used to perform complex analyses. To date, several plug-ins have been developed to make machine learning–based segmentation more accessible. For example, the Trainable Weka segmentation plug-in allows users to train segmentation algorithms using the ImageJ software graphical user interface.²⁶ Similarly, the laboratory that invented U-Net developed a U-Net plug-in (Unet-segmentation), which uses the Caffe deep-learning framework (University of California, Berkeley, CA) to implement the architecture in ImageJ.²⁰ DeepImageJ can run pretrained TensorFlow models in ImageJ, and is bundled with a set of models for particular tasks. One of the bundled models is a deep CNN that can virtually stain an unlabeled tissue autofluorescence image to produce an approximation of the corresponding H&E, Masson trichrome, or Jones stain image.²⁷

CellProfiler

CellProfiler is a widely used software that has been cited >9000 times since the publication on its original release in 2006.²⁸ It has a point-and-click graphical user interface that allows users with minimal programming experience to string together several image-analysis modules into an

analytical pipeline. CellProfiler comes with >50 modules that allow for standard image-analysis procedures, and further allows users to write their own modules. As of 2018, CellProfiler software version 3.0 featured a pretrained U-Net–based semantic segmentation module (ClassifyPixels-Unet).²⁹ A drawback is that no task-specific training or fine-tuning can be applied, which can limit its utility in tissue microscopy analysis, as tissue background can have a substantial impact on network generalizability. However, the modularity of CellProfiler allows users to integrate trained models from other sources into a large analytical pipeline. For example, Sadanandan et al³⁰ wrote a module that allowed users to run pretrained, Caffe-based models within CellProfiler pipelines.

Ilastik

Ilastik is a deep-learning software package used for performing semantic segmentation tasks by extracting a series of predefined pixel-level features to train random forest classifiers.³¹ It provides a user interface that enables the generation of sparse training labels, and an interactive mode in which users can provide feedback to the network during training. This flexibility has made it a useful tool in the context of highly multiplexed imaging, in which training-set generation is an expensive process.^{32,33} Several recent papers have reported on the combination of CellProfiler with Ilastik for segmenting cells in tissue-microscopy images. For example, a CellProfiler pipeline can be generated to perform preprocessing steps, then Ilastik can be used to produce a probability map for semantic segmentations.^{34,35} This map is then returned to CellProfiler for object detection.

QuPath

QuPath, an open-source platform designed specifically for whole-slide image analysis, can be used for both immunohistochemistry analysis and immunofluorescence imaging.³⁶ Like Ilastik, it employs random forest classifiers that can be trained to segment cell classes based on cell surface marker expression. It is useful for quantifying the distribution of a particular marker across a whole-tissue section.

Downstream Applications of Algorithm Outputs

The output of a given machine-learning or deep-learning algorithm is often not the culmination of a digital pathology analysis. These techniques are widely applied to automate the extraction of quantifiable metrics from cell images to answer biology-related questions. This section addresses how machine-extracted features are used in downstream analyses.

Extracting Human-Interpretable Features from Deep-Learning Algorithms

Beyond automating the laborious parts of image analysis, deep learning provides the advantage of utilizing image features that are not generated or perceived by the user. However, the black-box nature of DCNNs is a common critique of applying deep learning to medical and biological image analysis. DCNNs are excellent for recognizing patterns, but to optimize and apply DCNNs, it needs to be understood whether they are recognizing either patterns specific to the imaging system or protocol, or biological features such as cell density and tissue structure. Minimal insight into how a DCNN arrives at a classification limits the opportunities to strategically improve the algorithm or training data. The ability to explain and interpret how deep-learning algorithms arrive at classification scores has therefore become a topic of active research by AI developers and AI users, respectively. The internal machine-learned features can be extracted to better interpret the rationale of decision making of a trained network. Attention maps allow for visualization of these machine-learned features, highlighting the parts of a given image that were influential in classifying or segmenting the image.³⁷ Faust et al³⁸ have used feature map visualization methods to compare learned features with human-interpretable features of images. Machine-learned features were found to correlate to multiple tissue and cell features, including the presence of epithelial cells, fibrosis, and luminal space.

An understanding of the internal DCNN features and the ways in which they relate to known markers of tissue structure and pathology provides insight into the decision-making capabilities of these AI algorithms. Human-interpretable features can also be generated from DCNN results to better characterize pathologic tissue. U-Nets and their derivatives can be used to segment both cells and tissues within a sample, producing a multiscale map of the cell populations that reside in given regions of tissue. If a network that segments cancerous tissue is combined with a network that segments cells, the resulting segmentations can be mined for human-interpretable features of the spatial distribution of cells, such as lymphocyte density near the cancer boundary and relative density of immune cells and cancer cells within the tumor.³⁹ The method of combining deep automated cell and lesion detection with human-engineered computer-visualization features provides high-throughput computation of interpretable features that previously would have required costly and time-intensive manual annotation of images.

Quantifying Spatial Organization of Cells in Microscopy Images

Automation of annotation facilitates spatial analyses that are more complex than was previously possible. Specifically, determining the interspatial organization of various cell

classes unlocks the capacity to determine the ways in which the cell constituents of tissue interact. This spatial data allows for the generation of novel features that can be used to stratify patients and increase the understanding of cell processes in tissue.

This idea has been applied extensively in the study of tumor immunology. Nearchou et al⁴⁰ examined sequential sections of surgically resected colorectal cancer specimens and developed the *spatial immuno-oncology index*, composed of patient-level spatial features, including mean CD3⁺ density, mean number of lymphocytes in proximity to tumor buds, and ratio of CD68⁺/CD163⁺ macrophages, and used as a predictor of prognosis. Similarly, Lazarus et al⁴¹ examined the abundance and spatial distribution of various subsets of T cells, antigen-presenting cells, and tumor cells in liver metastases of colorectal cancer. The automated detection and segmentation of immune cells facilitated the finding that cytotoxic T lymphocytes were typically positioned further from epithelial cells and antigen-presenting cells that were expressing a programmed cell death checkpoint molecule. Furthermore, they found that high levels of *engagement*, defined as an intercell distance of <15 μm , between cytotoxic T lymphocytes and epithelial cells, helper T cells, and regulatory T cells was predictive of increased survival. Thus, automated image analysis allows for the extraction of features related to the interactions between tumors and the immune system that can be used to predict patient outcomes. Wang et al⁴² used a Mask R-CNN architecture to segment lymphocytes and tumors cells in whole-slide H&E images of hepatocellular carcinoma samples. These segmentations were then used to generate 246 image features that included tumor nuclei morphology, density of lymphocytes and tumors, and the spatial relationships between the two cell types. These features were used to perform unsupervised consensus clustering that identified three distinct subgroups within the patient population.

Predicting prognosis is not the final goal of these analyses. Rather, associating specific intercell interactions with positive or negative patient outcomes is the first step in identifying immunologic phenomena that have potential as therapeutic targets. Understanding the cell interactions taking place in tissue, and the ways in which these interactions correlate with outcome, allows for the identification of interactions that should be promoted and those that should be disrupted.

Challenges in Applying AI to Cell Image Analysis

Use of deep learning for image analysis as a standard in the experiment toolkit necessitates considering the ways in which image data acquisition can be optimized for downstream machine-learning applications (Figure 2). In particular, it is important to understand which features of data

acquisition can influence the performance of segmentation algorithms, including: i) selection bias, ii) sample processing, and iii) imaging system.^{43,44}

Selection Bias

Selection bias in AI for cell imaging has multiple components. This section discusses selection bias as it relates to both tissue sampling and building training, validation, and testing sets of data.

Bias in Tissue Collection and Sampling

Selection bias is always a concern with *ex vivo* imaging, as the full organ or tissue cannot always be excised for *ex vivo* analysis.⁴⁵ In sampling a small region of the tissue, *ex vivo* two-dimensional imaging may or may not contain the most severe region, meaning that at the cellular level, sampling error must be considered for in-depth evaluation of the tissue. While selection bias does affect the capacity of AI to accurately predict patient diagnosis or prognosis, tissue analysis by a pathologist is also subject to selection bias. Further improvements in the specificity of biopsy sample—acquisition methods will be useful in reducing the effects of selection bias on both human and machine decision-making in pathology.

Bias in Training Data

AI methods consist of trainable algorithms that must learn from a predetermined set of samples. If an algorithm is trained using data that is biased toward one population or

class, this bias will be inherent in the trained model when it is applied to new samples.⁴⁶ Certain biases in training data are difficult to address in smaller-scale studies. For example, if a training set is composed of samples from a single institution, or if one institution is more represented than another, the resulting model will likely perform better on samples from this overrepresented institution. The same concept can be applied to patient or sample groups (ie, patient age or sex) present in the training data. This bias can be partially mitigated through training-set normalization or harmonization but should be noted as a limitation of any trained model.

Sample Processing

The fixation method also affects the results of AI analysis of *ex vivo* tissue samples. A significant difference was found between quantitative features of cell morphology in cells detected by deep CNNs in tissue samples that were fresh-frozen compared to formalin-fixed, paraffin-embedded tissue samples.¹⁶ As these cell morphology features can be used as descriptors of disease states, it becomes important to interpret studies on patient diagnosis or disease pathogenesis relative to the way in which the biopsy sample was stored after acquisition. Variability in tissue staining is another concern. Even with the use of the same protocols, findings from H&E, immunohistochemistry, and immunofluorescence staining can vary between technicians, or within samples stained by the same technician on different days. With different institutions and different protocols, stain quality and appearance can vary even further.⁴⁴ These differences in staining can affect the performance of deep-

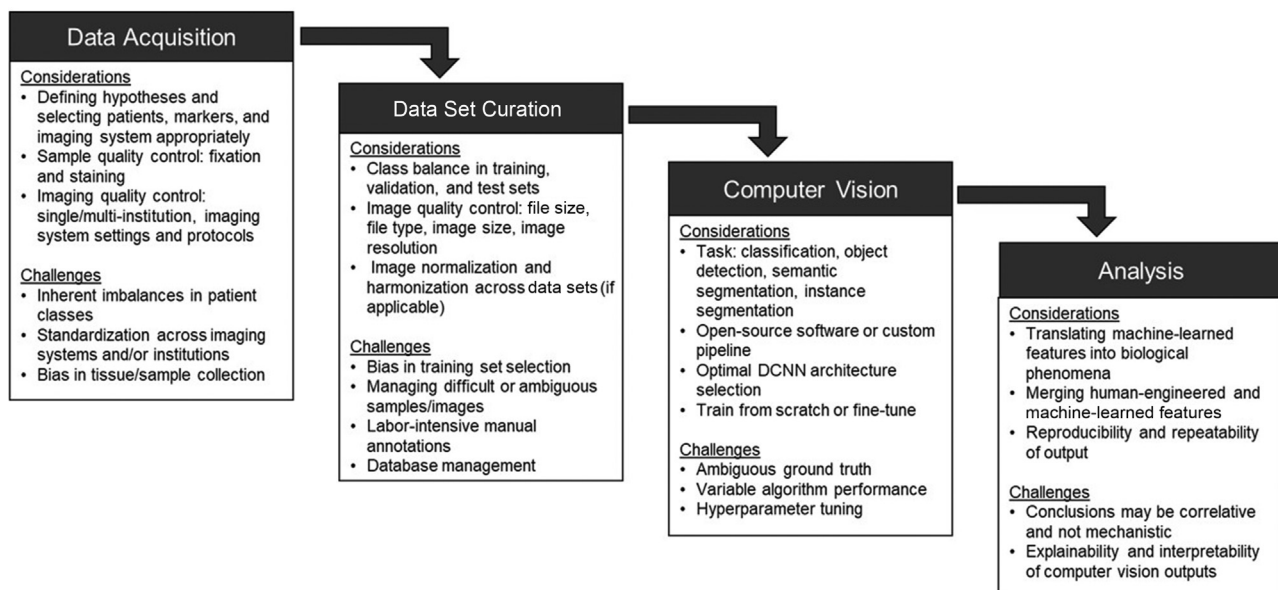


Figure 2 The application of AI to the microscopy of cells requires careful consideration and planning at all stages of the process, including data acquisition, data set curation, selection and deployment of computer-vision methods, and post–deep-learning analysis. If not properly addressed, the challenges associated with each step may limit the scope of the conclusions. DCNN, deep convolutional neural network.

learning algorithms in diagnosis or lesion/cell segmentation, but they also limit the generalizability of these algorithms. While deep-learning algorithms excel in the recognition of patterns in data sets, changes in the hue, relative intensity, or staining specificity can change these patterns, causing AI to learn institution-specific features and thus reducing algorithm generalizability.

Imaging System

Performance of AI in the analysis of microscopy images is also sensitive to the imaging system on which the training data were acquired, and trained algorithms may not generalize well to new microscopes or slide scanners.⁴⁷ Many variables associated with the imaging system can affect performance, including light source power and spectrum, magnification, pixel size, and objective numeric aperture. A large training set representative of the diversity of images used in the algorithm should be built, including images from multiple microscopes or slide scanners, if necessary. Additionally, a small fine-tuning data set can be used to adapt a trained network to a new imaging system or imaging protocol. Generalization between imaging systems is often further complicated by proprietary image file formats offered by various vendors, which can be difficult to convert into a standardized format. There is a need for an industry-wide file standard, analogous to the digital imaging and communications in medicine (DICOM) format used in radiologic imaging modalities.

Image Annotation

Many of the methods discussed in this review are considered supervised learning methods, in which a specialist must annotate a large training set of data to establish ground truth. Generating these training sets is a costly and time-intensive task, but algorithm performance hinges on the acquisition of a well-annotated training set. Fully unsupervised algorithms, such as clustering or dimensionality reduction, are machine-learning tools useful in digital pathology.⁴⁸ Rather than learning the features associated with specific classes, unsupervised methods learn the inherent distribution of the training data. Groups of points that emerge from an unsupervised algorithm will all have similar features, but the groups may or may not have direct relevance to human-interpretable groups. Unsupervised machine-learning algorithms do not require the manual annotation overhead of supervised algorithms. However, they address separate hypotheses, and probe questions such as, How do these data points group together?, compared to supervised questions, such as, What are the similarities between data points in each class?

New developments in AI are pushing toward semi- or self-supervised machine-learning methods. *Contrastive learning*, a form of self-supervision, has recently been

applied to medical imaging tasks to minimize the need for manual annotations. In contrastive learning, a network is trained to learn feature representations of images, ultimately identifying the features that are similar across similar images, without human-defined truth on the class of those images.⁴⁹ AI in digital and computational pathology is currently dominated by data-hungry algorithms, so the application of contrastive learning methods may be useful in mitigating the need for large sets of training data. Additionally, transfer learning can be implemented to leverage trained networks to apply to new image domains, reducing the minimal amount of manual data needed for achieving high accuracy.

Summary

AI in microscopy image analysis is a rapidly evolving discipline that affects both clinical and basic science research. The automation of image classification and cell detection using computer-visualization techniques facilitates high-throughput analysis of information-dense microscopy images, particularly in the case of multiplexed microscopy. These tools are becoming increasingly mainstream, with the availability of open-source software making them accessible to a broader range of users. Machine learning-enabled feature extraction can allow for novel insights into cell environments associated with disease. However, the use of machine-learning methods introduces a set of experiment-related considerations that must be addressed to generate data sets that produce robust and meaningful insights.

References

1. Bansal GJ: Digital radiography. A comparison with modern conventional imaging. *Postgrad Med J* 2006, 82:425–428
2. Giger ML, Chan HP, Boone J: Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM. *Med Phys* 2008, 35:5799–5820
3. Wang H, Shang S, Long L, Hu R, Wu Y, Chen N, Zhang S, Cong F, Lin S: Biological image analysis using deep learning-based methods: literature review. *Digit Med* 2018, 4:157–165
4. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B: Histopathological image analysis: a review. *IEEE Rev Biomed Eng* 2009, 2:147–171
5. Xing F, Yang L: Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. *IEEE Rev Biomed Eng* 2016, 9:234–263
6. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A: Artificial intelligence in digital pathology-new tools for diagnosis and precision oncology. *Nat Rev Clin Oncol* 2019, 16:703–715
7. Simonyan K, Zisserman A: Very deep convolutional networks for large-scale image recognition. *arXiv* 2014. 1409.1556
8. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, Summers RM, Giger ML: Deep learning in medical imaging and radiation therapy. *Med Phys* 2019, 46:e1–e36
9. Tizhoosh HR, Pantanowitz L: Artificial intelligence and digital pathology: challenges and opportunities. *J Pathol Inform* 2018, 9:38
10. Cooper LAD, Carter AB, Farris AB, Wang F, Kong J, Gutman DA, Widener P, Pan TC, Chollet SR, Sharma A, TKurc TM, Brat DJ,

- Saltz JH: Digital pathology: data-intensive frontier in medical imaging. *Proc IEEE Inst Electr Electron Eng* 2012, 100:991–1003
11. Radtke AJ, Kandov E, Lowekamp B, Speranza E, Chu CJ, Gola A, Thakur N, Shih R, Yao L, Yaniv ZR: IBEX: a versatile multiplex optical imaging approach for deep phenotyping and spatial analysis of cells in complex tissues. *Proc Natl Acad Sci U S A* 2020, 117: 33455–33465
 12. Tsujikawa T, Kumar S, Borkar RN, Azimi V, Thibault G, Chang YH, Balter A, Kawashima R, Choe G, Sauer D: Quantitative multiplex immunohistochemistry reveals myeloid-inflamed tumor-immune complexity associated with poor prognosis. *Cell Rep* 2017, 19: 203–217
 13. Goltsev Y, Samusik N, Kennedy-Darling J, Bhate S, Hale M, Vazquez G, Black S, Nolan GP: Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* 2018, 174: 968–981.e15
 14. Ronneberger O, Fischer P, Brox T: U-Net: Convolutional Networks for Biomedical Image Segmentation. New York: Springer. 234-241.
 15. Yang L, Ghosh RP, Franklin JM, Chen S, You C, Narayan RR, Melcher ML, Liphardt JT: NuSeT: a deep learning tool for reliably separating and analyzing crowded cells. *PLoS Comput Biol* 2020, 16: e1008193
 16. Durkee MS, Abraham R, Ai J, Veselits M, Clark MR, Giger ML: Quantifying the effects of biopsy fixation and staining panel design on automatic instance segmentation of immune cells in human lupus nephritis. *J Biomed Opt* 2021, 26:1–17
 17. de Bel T, Bokhorst J-M, van der Laak J, Litjens G: Residual CycleGAN for robust domain transformation of histopathological tissue slides. *Med Image Anal* 2021, 70:102004
 18. Keikhosravi A, Li B, Liu Y, Conklin MW, Loeffler AG, Eliceiri KW: Non-disruptive collagen characterization in clinical histopathology using cross-modality image synthesis. *Commun Biol* 2020, 3:414
 19. Weigert M, Schmidt U, Boothe T, Müller A, Dibrov A, Jain A, Wilhelm B, Schmidt D, Broaddus C, Culley S, Rocha-Martins M, Segovia-Miranda F, Norden C, Henriques R, Zerial M, Solimena M, Rink J, Tomancak P, Royer L, Jug F, Myers EW: Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat Methods* 2018, 15:1090–1097
 20. Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, Böhm A, Deubner J, Jäckel Z, Seiwald K, Dovzhenko A, Tietz O, Dal Bosco C, Walsh S, Saltukoglu D, Tay TL, Prinz M, Palme K, Simons M, Diester I, Brox T, Ronneberger O: U-net: deep learning for cell counting, detection, and morphometry. *Nat Methods* 2019, 16:67–70
 21. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J: UNet++: a nested U-net architecture for medical image segmentation. Edited by Stoyanov D, Taylor Z, Carneiro G, Syeda-Mahmood T, Martel A, Maier-Hein L, Tavares JMRS, Bradley A, Papa JP, Belagiannis V, Nascimento JC, Lu Z, Conjeti S, Moradi M, Greenspan H, Madabhushi A. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer International Publishing, 2018. pp. 3–11
 22. Brennan DJ, Rexhepaj E, O'Brien SL, McSherry E, O'Connor DP, Fagan A, Culhane AC, Higgins DG, Jirstrom K, Millikan RC: Altered cytoplasmic-to-nuclear ratio of survivin is a prognostic indicator in breast cancer. *Clin Cancer Res* 2008, 14:2681–2689
 23. Ren S, He K, Girshick R, Sun J: Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE* 2017, 39: 1137–1149
 24. He K, Gkioxari G, Dollár P, Girshick R: Mask R-CNN. 2017 International Conference on Computer Vision (ICCV 2017). Oct 24-27: Venice, Italy: IEEE Computer Society, 2017. pp. 2961–2969
 25. Lin T-Y, Goyal P, Girshick R, He K, Dollár P: Focal loss for dense object detection. 2017 International Conference on Computer Vision (ICCV 2017). Oct 24-27: Venice, Italy: IEEE Computer Society, 2017. pp. 2980–2988
 26. Rodemerck J, Junker A, Chen B, Pierscianek D, Dammann P, Darkwah Oppong M, Radbruch A, Forsting M, Maderwald S, Quick HH: Pathophysiology of intracranial aneurysms: Cox-2 expression, iron deposition in aneurysm wall, and correlation with magnetic resonance imaging. *Stroke* 2020, 51:2505–2513
 27. Rivenson Y, Wang H, Wei Z, de Haan K, Zhang Y, Wu Y, Günaydin H, Zuckerman JE, Chong T, Sisk AE: Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng* 2019, 3:466–477
 28. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J: CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 2006, 7:1–11
 29. McQuin C, Goodman A, Chernyshev V, Kamensky L, Cimini BA, Karhohs KW, Doan M, Ding L, Rafelski SM, Thirstrup D: CellProfiler 3.0: next-generation image processing for biology. *PLoS Biol* 2018, 16:e2005970
 30. Sadanandan SK, Ranefall P, Le Guyader S, Wählby C: Automated training of deep convolutional neural networks for cell segmentation. *Sci Rep* 2017, 7:1–7
 31. Berg S, Kutra D, Kroeger T, Straehle CN, Kausler BX, Haubold C, Schiegg M, Ales J, Beier T, Rudy M: Ilastik: interactive machine learning for (bio) image analysis. *Nat Methods* 2019, 16:1226–1232
 32. McMahon NP, Jones JA, Kwon S, Chin K, Nederlof MA, Gray JW, Gibbs SL: Oligonucleotide conjugated antibodies permit highly multiplexed immunofluorescence for future use in clinical histopathology. *J Biomed Opt* 2020, 25:1–18
 33. Rashid R, Gaglia G, Chen Y-A, Lin J-R, Du Z, Maliga Z, Schapiro D, Yapp C, Muhlich J, Sokolov A: Highly multiplexed immunofluorescence images and single-cell data of immune markers in tonsil and lung cancer. *Sci Data* 2019, 6:1–10
 34. Wills JW, Robertson J, Summers HD, Minter M, Barnes C, Hewitt RE, Keita ÁV, Söderholm JD, Rees P, Powell JJ: Image-based cell profiling enables quantitative tissue microscopy in gastroenterology. *Cytometry A* 2020, 97:1222–1237
 35. Schürch CM, Bhate SS, Barlow GL, Phillips DJ, Noti L, Zlobec I, Chu P, Black S, Demeter J, McIlwain DR: Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell* 2020, 182:1341–1359
 36. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG: QuPath: open source software for digital pathology image analysis. *Sci Rep* 2017, 7:1–7
 37. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D: Grad-cam: visual explanations from deep networks via gradient-based localization. 2017 International Conference on Computer Vision (ICCV 2017). Oct 24-27: Venice, Italy: IEEE Computer Society, 2017. pp. 618–626
 38. Faust K, Bala S, van Ommeren R, Portante A, Al Qawahmed R, Djuric U, Diamandis P: Intelligent feature engineering and ontological mapping of brain tumour histomorphologies by deep learning. *Nat Mach Intell* 2019, 1:316–321
 39. Diao JA, Chui WF, Wang JK, Mitchell RN, Rao SK, Resnick MB, Lahiri A, Maheshwari C, Glass B, Mountain V, Kerner JK, Montalto MC, Khosla A, Wapinski IN, Beck AH, Taylor-Weiner A, Elliott HL: Dense, high-resolution mapping of cells and tissues from pathology images for the interpretable prediction of molecular phenotypes in cancer. *bioRxiv* 2020. [Preprint] doi:10.1101/2020.08.02.233197
 40. Nearchou IP, Gwyther BM, Georgiakakis ECT, Gavriel CG, Lillard K, Kajiwar Y, Ueno H, Harrison DJ, Caie PD: Spatial immune profiling of the colorectal tumor microenvironment predicts good outcome in stage II patients. *NPJ Digit Med* 2020, 3:1–10
 41. Lazarus J, Maj T, Smith JJ, Lanfranca MP, Rao A, D'Angelica MI, Delrosario L, Girgis A, Schukow C, Shia J: Spatial and phenotypic immune profiling of metastatic colon cancer. *JCI Insight* 2018, 3: e121932

42. Wang H, Jiang Y, Li B, Cui Y, Li D, Li R: Single-cell spatial analysis of tumor and immune microenvironment on whole-slide image reveals hepatocellular carcinoma subtypes. *Cancers* 2020, 12:3562
43. Niethammer M, Borland D, Marron JS, Woosley J, Thomas NE: Appearance normalization of histology slides. *Mach Learn Med Imaging* 2010, 6357:58–66
44. Bancroft JD, Gamble M: *Theory and Practice of Histological Techniques*. New York: Elsevier Health Sciences, 2008
45. Hoppin JA, Tolbert PE, Taylor JA, Schroeder JC, Holly EA: Potential for selection bias with tumor tissue retrieval in molecular epidemiology studies. *Ann Epidemiol* 2002, 12:1–6
46. Jiang H, Nachum O: Identifying and correcting label bias in machine learning. *Proc Mach Learn Res* 2020, 108:702–712
47. Weinstein RS, Graham AR, Richter LC, Barker GP, Krupinski EA, Lopez AM, Erps KA, Bhattacharyya AK, Yagi Y, Gilbertson JR: Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Hum Pathol* 2009, 40:1057–1069
48. Rashidi HH, Tran NK, Betts EV, Howell LP, Green R: Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. *Acad Pathol* 2019, 6:2374289519873088
49. Chen T, Kornblith S, Norouzi M, Hinton G: A simple framework for contrastive learning of visual representations. *Proc Mach Learn Res* 2020, 119:1597–1607