THE UNIVERSITY OF CHICAGO

INTERROGATING SPATIAL FEATURES OF RENAL INFLAMMATION IN HUMAN LUPUS NEPHRITIS USING COMPUTER VISION

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

INTERDISCIPLINARY SCIENTIST TRAINING PROGRAM:

IMMUNOLOGY

BY

REBECCA ABRAHAM

CHICAGO, ILLINOIS

DECEMBER 2021

Table of Contents

List of Figures

List of Tables

Abstract

Lupus Nephritis is a chronic inflammatory kidney disease that arises in the context of systemic lupus erythematosus (SLE). It is characterized by the deposition of immune complexes in the kidney, localized inflammation, and ultimately kidney failure.[1] In the systemic model of lupus pathogenesis, antigen presenting cells activate auto-reactive CD4+ T cells, which in turn provide help to auto-reactive B cells that differentiate into plasma cells that produce auto-antibodies. Many therapeutic approaches (such as B cell depletion) have been attempted to disrupt this pathway. However, despite the rational basis of these approaches, none have demonstrated robust performance in clinical trials.[2,3] This suggests that our understanding of how the adaptive immune system contributes to disease burden is incomplete. This work aims to fill this gap by investigating inflammation that is localized in the kidney, using computer vision to identify cells in tissue, and spatial analysis to define how these cells are organized.

We first related cellular features of inflammation with progression to renal failure using the biopsies of LuN patients for whom we have at least 2 years of clinical follow-up data. Five classes of cells were identified in these biopsies: CD4+ T cells, CD4- T cells, B cells, myeloid dendritic cells, and plasmacytoid dendritic cells. Two striking results emerged from this work—first, we observed that dense cellular neighborhoods of CD4- T cells are associated with progression to renal failure. Second, dense regions of B cells were found in a subset of patients who had preserved renal function.

T cell and B cell phenotypes were further interrogated using a highly-multiplexed dataset from a separate cohort of LuN patients. The richness of markers in this second dataset allowed for investigation into the spatial distribution of more specific cell

phenotypes. We showed that “CD4- T cells” are not exclusively CD8+ but are rather a diverse compartment that include CD3+CD4-CD8- (double negative) T cells, roughly 50% of which might be gamma-delta T cells. In addition, we found that regulatory T cells were relatively rare, while T follicular helper cells were abundant and frequently found in large cellular neighborhoods with B cells. Finally, we found that dense cellular neighborhoods often exist in the context of larger inflammation, while some smaller neighborhoods are isolated within the tissue. Analyzing both image datasets together allowed us to identify features of *in situ* inflammation that associate with patient outcomes in LuN and define metrics that can be used to evaluate inflammatory structures in tissue.

Finally, we developed methods for improving cellular segmentation in inflamed tissue. We investigated the generalizability of cellular segmentation algorithms to other diseases states and demonstrated techniques by which a segmentation algorithm trained for lupus nephritis can be applied to triple negative breast cancer. We also evaluated the effects of sample preparation and staining panel choice on the performance of segmentation algorithms. These insights will aid the development of robust quantitative pipelines for understanding human tissue inflammation.

Chapter 1: Introduction—What is lupus nephritis?

1. Introduction

Lupus nephritis (LuN) is a manifestation of systemic lupus erythematosus (SLE) that is characterized by chronic kidney inflammation and damage.[1] This disease has a complex etiology, with several branches of the immune system playing a role in either driving or mitigating pathology. In order to evaluate the roles of these cells in humans, there are 3 main compartments that we evaluate—blood, urine and kidney. It is relatively easy and non-invasive to collect blood and urine from patients, and then use flow cytometry or transcriptomic analysis to finely detail their cellular constituents. However, these likely provide an incomplete or tangential view of what is happening at the site of inflammation. Evaluating the status of inflammation in the kidney itself is difficult, as renal biopsies have risks associated with them, and are therefore used sparingly in a clinical setting. However, understanding what is happening in the kidney is critical to understanding what is actually causing the kidneys to deteriorate and fail, which might not be obvious from the blood and urine.

Because mechanistic insight is difficult to attain from human studies, typically investigations into human tissue attempt to relate the abundance of specific cells, transcriptional profiles, or cytokines with the type of patient, whether they respond to therapy, pathological metrics (activity indices, chronicity indices) and renal function. The most mechanistic studies we have are clinical trials, which are as close to an experiment as one can do with humans. Here I will describe the clinical picture of lupus nephritis, describe what is known about how various components of the immune system contribute to pathology, and outline the gap in the field that my work is meant to fill.

2. Clinical picture and current standard of care of Lupus Nephritis

It has been estimated that LuN occurs in up to 50% of SLE patients, but it has been observed that there is an increased frequency in patients of African, Asian, and Hispanic descent[4]. Patients with lupus nephritis present with protein and blood in the urine (proteinuria, hematuria), and renal insufficiency[5]. Approximately 10-30% of LuN patients proceed to kidney failure, or end-stage renal disease (ESRD) within 10 years[4,6], resulting in the need for a transplant or dialysis.

The current standard of care for lupus nephritis is an induction period of intensive therapy followed by lower-dose maintenance therapy. Induction usually takes a two-pronged approach— 1) cytotoxic agents such as mycophenolate mofetil (MMF) or cyclophosphamide that target proliferating cells and 2) corticosteroids, which have a global immunosuppressive effect.[6] Maintenance therapy is usually done with MMF and azathioprine, with MMF being the therapy of choice, but azathioprine providing an acceptable alternative in situations like pregnancy where MMF in contraindicated.

Lupus nephritis can follow several courses -- some patients can go into remission for periods of time, interspersed with periodic “flares” of poor kidney function.[5–7] Other patients will have completely normal kidney function and then have sudden, acute kidney failure. Still other patients have lowered kidney function that remains remarkably stable over many years. For this reason, the optimal period of maintenance therapy varies by patient. This disease therefore has a very heterogeneous clinical picture but lacks a correspondingly diverse set of therapies.

3. Kidney biopsy for evaluating/classifying LuN

Kidney biopsies are a powerful tool for diagnosing and evaluating lupus nephritis. The primary method for classifying LuN in the US is the International Society of Nephrology/Renal Pathology Society (ISN/RPS) system, developed in 2003 and revised in 2018.[8] It classifies biopsies into 6 classes with increasing severity: (I) minimal mesangial, (II) mesangial proliferative, (III) focal (<50 involved glomeruli), (IV) diffuse (>50 involved glomeruli), (V) membranous, and (VI) advanced sclerosing. Classes I and II are both considered to be "non-proliferative" while classes III and IV are referred to as "proliferative", the key difference being that proliferative classes have inflammation. This standard scoring system focuses exclusively on glomerular inflammation and damage.

There is an alternative NIH scoring system which provides activity and chronicity indices. In this context, "Activity" is a score between 0 and 24 that captures phenomena such as endocapillary hypercellularity, neutrophils, tubulointerstitial inflammation (TII), cellular crescents, hyaline deposits, and fibrinoid necrosis. Twenty one of 24 points capture glomerular activity while only three points capture TII. In short, it is reflective of ongoing damage events at the time of biopsy. "Chronicity", in contrast, is a 0-12 scale that refers to glomerulosclerosis, fibrous crescents, tubular atrophy, and interstitial fibrosis, all measures of damage that has already occurred at the time of biopsy. It can be broken down into two components reflecting distinctive compartments within the kidney—six points are associated with glomerular damage and six points with tubulointerstitial damage.

Kidney biopsies are also probed for immune complex (conjugates of antibodies with antigen) deposition by staining for immunoglobulin (Ig). "Full house" staining is

described when a biopsy tests positive for IgG, IgA, IgM, C1q, and C3, signifying active antibody deposition and complement activate in the kidneys.[6] These measures are more useful in making a definitive diagnosis rather than staging renal disease activity or damage.

The utility of some of these metrics in predicting renal outcomes has been called into question, as they primarily focus on glomerular pathology and exclude the tubulointerstitium.[9] Tubulointerstitial inflammation (TI) can be scored on a 0-3 scale by immunohistochemical staining for CD45, a marker for cells of a hematopoietic lineage. It has been observed that severe (2-3) TI scores are predictive of renal failure within 5 years of the biopsy. However, within the cohort of patients who have moderate/severe inflammation, roughly half progress to ESRD in five years, while half have relatively stable disease. Similarly, the predictive power of the NIH chronicity scale appears to be driven by the tubulointerstitial component of the score.[10] This variation in patient outcome suggests that glomerular inflammation is not a direct predictor of prognosis, and that the diversity in tubulointerstitial inflammation needs to be investigated further.

4. Mouse models of lupus nephritis

There are several mouse models of lupus nephritis.[11] The MRL^{lpr} model is a spontaneous multigenic model that is driven in part by a loss of function mutation in the *Fas* gene, which mediates apoptosis. It is characterized by multi-system disease, including glomerulonephritis that is characterized by macrophages, T cells, and neutrophils. Disease is independent of FCγRs which is contrary to the observation that mutations in FCγRs are commonly found in some SLE patients.[12] The *gld* line is another spontaneous model that is the result of a mutation in *FasL,* the ligand to *Fas*. These mice

have reduced glomerular pathology relative to the MRL^{lpr} mice.[13] There are several spontaneous lupus models that arise from cross between two strains of mice. BXSB mice result from a cross between C57BL/6 mice and SB/Le mice, and (unlike humans) male mice have greater disease burden. Renal failure in these mice is characterized by proliferative nephritis with prominent neutrophil infiltration, which is not characteristic of humans.[14] F1 mice from a cross between NZB and NZW mice (B/W) are thought to closely recapitulate the human disease. From this cross a stable NZM line has been developed that resembles the parental cross. This multigenic model is one of the most commonly studied mouse models. Intraperitoneal injection of pristane has also been used to investigate auto-antibody production, and these mice have exhibited nephritis, with substantial monocytic infiltrate and antibody deposition. However, this is generally considered to be a suboptimal model due to the other side effects of pristane.

Each of these models recapitulates some, but not all, aspects of lupus nephritis. It has been observed that all of these models are both different from each other and from humans.[15] Furthermore, while each mouse model represents a single genetic state, lupus is genetically very heterogeneous with over 50 susceptibility loci associated with the disease.[16] Therefore, while each mouse model might represent a specific type of lupus, they are not broadly applicable to lupus patients.

Because the work that I will describe has exclusively been focused on studying this disease in humans, I will mostly cover the results of studies that have been done in humans, referring to mouse studies only when they were used as a complement for observations from human studies.

5. Cast of characters—how does adaptive immunity contribute to LuN?

B cells

The dominant model of lupus nephritis pathogenesis is that of auto-antibody mediated damage to the kidney. It is thought that these auto-antibodies are generated as a result of "autovaccination" events, wherein dendritic cells take up self antigens and aberrantly present them to CD4+ T cells, which in turn provide help to auto-reactive B cells, driving their differentiation into antibody-producing plasma cells.[1] This is commonly referred to as a break in immunological "tolerance" to self-proteins. Renal damage occurs when immune complexes (aggregates of bound antibody-antigen pairs) deposit or form de novo in the glomeruli and trigger local complement activation, which in turn engages Fc receptors on leukocytes and tubule cells alike, triggering a cascade of inflammatory responses.[17]

The primary role of B cells in driving lupus nephritis lies in producing auto-antibodies. Lupus nephritis patients often produce antibodies against specific self-derived proteins and molecules. In particular, anti-nuclear antibodies (ANA) are used as biomarkers for diagnosing lupus and, to a lesser degree, monitoring disease activity and severity.[18] Anti-nuclear antibodies are a broad class of antibodies that are specific to macromolecules found in the nucleus, including double stranded DNA (dsDNA) and ribonuclear proteins (RNPs).[19] It has been postulated that anti-nuclear antibodies specifically drive nephritis because extracellular dsDNA binds to the glomerular membrane, triggering antibody deposition and complement fixation at that site.[17,20] In addition to ANA, antibodies against the intermediate filament protein vimentin have been associated with severe tubulointerstitial inflammation (TII).[21]

Auto-reactive plasma cells therefore drive systemic lupus nephritis pathology by generating antibodies against specific “nephritogenic” antigens. Indeed, circulating plasma cells have been found to be elevated in LuN patients, and positively correlated with disease activity.[22,23] There is also evidence of local plasma cell accumulation in the kidney itself.[24] Recently, the presence of a subset of “age associated” CD11c+IgD-CD27- B cells has been identified in both the peripheral blood[25] and kidneys[14] of lupus patients. This subset of cells has been associated with both normal aging and autoimmunity and appear to be hyper-responsive to TLR7 stimulation, which causes them to differentiate into plasma cells.

However, circulating antibodies are not the only role that B cells may play in lupus pathogenesis. There is abundant evidence that B cells that have not yet differentiated into plasma cells infiltrate the kidneys of lupus nephritis patients, leading to the hypothesis that they are having additional tissue-localized effects. For example, they could be playing a role as an antigen presenting cell to kidney-infiltrating T cells. Indeed, in mouse models, the antigen presenting capabilities of B cells is critical for lupus pathogenesis.[26]

Given the abundant evidence that B cells play a critical role in driving LuN pathology, B cell depletion is an obvious therapeutic approach for treating this disease. However, this has not been borne out in clinical trials. Rituximab, an anti-CD20 monoclonal antibody, has consistently failed to show superiority over standard of care in large-scale clinical trials[2], despite the fact that it successfully depletes circulating B cells[27–29] and has been very effective in other autoimmune contexts, such as rheumatoid arthritis.[30] Intriguingly, B cell depletion has also been found to correspond with increases in circulating regulatory T cell markers (CD25, FOXP3, TGFβ) within 1-3 months.[27,29]

Similarly ocrelizumab (another anti-CD20 antibody) did not improve renal responses over standard of care, and even showed increased adverse events in combination with MMF.[31] On the other hand, a third anti-CD20 antibody, obinutuzumab, has shown some promising preliminary results, though the full clinical trial has not been completed.[32] Overall, depletion of this population of B cells has not remarkably improved therapy over MMF/cyclophosphamide. Given that cyclophosphamide and MMF seem to deplete B cells effectively[33], it is possible that anti-CD20 therapy does not confer an additional benefit.

Another approach that has been taken to mitigating the role of B cells in this disease has been inhibition of the cytokine BAFF, also known as BlyS. Anti-BlyS/BAFF antibodies such as belimumab have shown more promise in clinical trials. The BLISS-LN trial tested belimumab on top of standard of care vs standard of care in patients with active LuN, and partial renal response rates increased from 32 to 43%, with a concomitant decrease in anti-dsDNA antibodies in the treatment arm that received the belimumab.[34] Another trial (CALIBRATE) took a multi-pronged approach by testing the efficacy of rituximab, corticosteroid and cyclophosphamide, followed by a course of belimumab in a population of patients with refractory LuN. Though addition of belimumab was safe, it did not significantly improve treatment response rates relative to patients who received just rituximab, corticosteroid and cyclophosphamide. They observed B cell depletion by week 12, and it was better maintained in the belimumab arm; in particular, the people who got belimumab had fewer naïve/transitional B cells and more memory B cells, with a corresponding decrease in IgG. However, ANA+ transitional B cells rebounded, though it was delayed in the belimumab arm.[3] Interestingly, a comparison of the transcriptional profile of responders vs non-responders to therapy via microarray showed that BAFF

expression was associated responders, suggesting that for the patients for whom B cells are driving pathology, standard of care successfully addresses their disease.[35]

The results of all of these trials indicate that simply depleting B cells from the equation does not successfully resolve disease in many patients. Furthermore, some data suggest that current standard of care therapies might be effective in treating B cell-driven pathologies. This implies that other cellular/inflammatory factors might be contributing to the continuing kidney damage, independently of the continuous activity of B cells.

CD4+ T cells

There is also plentiful evidence for CD4+ T cells as a major constituent of LuN in situ inflammation. CD4+ T cells have substantial heterogeneity in their phenotypes and effector functions in terms of their cytokine profiles and effector functions. Therefore there is a variety of roles they might play in the pathogenesis of the disease. Patschan et al observed that circulating CD4+ T cells in lupus nephritis patients were enriched for costimulatory molecules, particularly CD80 and CD134, suggesting that they are actively providing help to other T cells.[36]

The most obvious role of CD4+ T cells in the tissue is that of T follicular helper-like (Tfh) cells. These cells can have the capacity to provide help to B cells and drive their differentiation into plasma cells. Though Tfh are traditionally found in lymphoid tissue, Tfh-like cells, or “peripheral Tfh” have been identified numerous times in the context of autoimmunity, and typically have a PD1+ICOS+/-CXCR5- phenotype.[37–39] Indeed, our laboratory has previously demonstrated that in human lupus nephritis, in situ B cells present antigen to T cells with characteristics of T follicular helper cells including expressing high levels of IL21.[37] Bocharnikov et al performed mass cytometry on PBMCs

from SLE patients and showed that there is an enrichment for a cluster of “T peripheral helper”(Tph) cells that are CD4+CXCR5-PD1+HLADR+. They further identified PD1hi/midCD4+ T cells in LuN kidneys that correlate with in situ B cell abundance and express CXCL13 and IL21, indicating that they have a role as B cell helpers in inflamed LuN kidneys.[39]

There is also some evidence that regulatory T cells (Tregs), identified by the expression of FoxP3 in CD4+ T cells, might be present in the LuN tissue and might help dampen ongoing inflammatory processes through the release of anti-inflammatory molecules like IL-10 and TGFβ. However, there is contradictory evidence about how abundant and active they are in SLE patients, with some studies finding them elevated in the circulation of active SLE patients, and others finding them diminished.[40–43] Tregs have been observed in lupus nephritis renal biopsies by both imaging and single cell RNA seq.[14,41,44] However, it has been proposed that lupus patients might have defective or diminished circulating regulatory T cells.[45] Scherlinger et al found that Tregs isolated from the blood of patients with SLE overexpressed PSGL1, leading to them to bind and aggregate platelets, which serve to reduce their ability to suppress effector T cells.[46] Another study found that Treg and follicular regulatory T cells function (Tregs that exist in germinal centers) was diminished in SLE patients through engagement of the OX40/OX40L axis by antigen presenting cells.[47] It has further been observed that regulatory T cells decrease in patients after immunosuppressive therapy, in conjunction with decreased SLEDAI scores[41]. In summary, the role of regulatory T cells in lupus nephritis is unclear and needs to be further clarified.

The role of other CD4 T helper subsets has also been investigated. Abdirama et al identified circulating and urinary CD4 T cells that were reactive to nuclear antigens in both healthy controls and SLE patients.[48] They found an elevation of IFNγ and IL-10 producing CD4 T cells in patients with active SLE. Tbet+ Th1 cells have been found to be diminished in circulation of LuN patients, with a corresponding increase in urinary Th1 cells relative to both healthy and disease controls.[43] In addition, Th1 cytokines (IL18, IL12, IFNγ) have been observed to the be elevated in LuN patients, and it has been proposed that this contributes to increased type I interferon production by pDCs.[49] Th17 were found to be elevated in non-proliferative LN patients relative to proliferative, and were observed to increase in LN patients post induction therapy.[43]

CD8+ T cells

CD8+ T cells are most often appreciated for their role as cytotoxic effectors, wherein they induce cell death in an antigen-specific manner. The AMP single cell RNA seq data identified at least 3 groups of CD8+ T cells: cytotoxic GZMB+Perforin+PD1- CD8+ T cells, resident memory CD8+ T cells, and GZMK+ T cells. Fava et al[50] paired this dataset with matched urine proteomics and found a chemokine signature in the urine that was associated with lupus patients, and downstream of IFNγ and TNFα signaling. They further found that CD8+ and NK cells produce these chemokines and that the signature observed in urine was correlated with the abundance of CD8+ T cells for 6 out of 30 patients. Accordingly, it has been observed that there is increased abundance of urinary[51] and circulating[52] CD8+ T cells in LuN patients correlated with periods of renal flare. Cyclophosphamide induction has been shown to decrease the frequency of circulating CD8+CD44+CD62L- T cells at 15 weeks post-induction[33]. Several studies have shown

an abundance of CD8+ memory T cells in kidney biopsies of patients[53,54]; showed that they proliferate and express GZMB in addition to a potassium channel Kv1.3; showed somewhat compelling therapeutic efficacy of Kv1.3 blockade in immunodeficient NOD-scid-IL2Rγ^{null} (NSG) mice who had transferred PBMCs from LuN patients.[53] Interestingly, it has been noted that kidney-infiltrating CD8+CD28- T cells are clonally expanded, suggesting an antigen-specific local response.[55]

One point of contention in the literature is whether CD8+ T cells associated with LuN are exhausted. Chronic antigen stimulation by constitutively available self antigens is thought to result in CD8+ T cell exhaustion, and it has been observed that PBMCs from SLE patients were enriched for an exhaustion signature that correlated positive with better patient outcomes.[56,57] Similarly, Buang et al found that circulating CD8+ T cells in SLE patients have aberrant metabolism as a result of interferon exposure.[58] However, the single cell RNA sequencing experiments suggested that CD8+ T cells that reside in the kidneys have low expression of exhaustion markers.[14] Two possible explanations for this discrepancy are 1) the difference between the populations of circulating vs kidney-infiltrating CD8 T cells and 2) differences between measured protein and RNA expression due to regulatory dynamics.

CD4-CD8- (DN) T cells

Though most T cells are canonically either CD4+ or CD8+, several studies have identified a population of CD4-CD8- T cells in the context of lupus nephritis and other autoimmune contexts.[59] Crispín et al identified a population of circulating and kidney-infiltrating TCR$\alpha\beta$ DN T cells in SLE patients that appear to express IL17.[60] Follow up studies from this laboratory suggested that these DN T cells are former CD8+ T cells that

have downregulated CD8 in response to apoptotic debris.[61] Stratigou et al compared the expression of SLAM family receptors on various T cell subsets, including DN T cells, and found that DN T cells identified in LuN patients had alterations in the expression of SLAMF2, 4, and 7 relative to healthy controls, and that increased expression of SLAMF6 on DN T cells prior to treatment was associated with a failure to respond to B cell depletion via rituximab.[62]

One possible subset of CD3+CD4-CD8- T cells in LuN are Natural Killer T (NKT) cells. However, they have not been detected in the kidneys of lupus nephritis patients[14], and their role in mouse models of the disease remains controversial.[63]

Another potential subset of DN T cells are gamma delta T cells, which express gamma and delta T cell receptor chains (rather than the more common alpha and beta chains) and comprise 3-10% of T cells. Yin et al identified a subset of circulating gamma delta T cells in a cohort of 15 lupus patients that were associated with increased disease activity, and further observed these cells in the kidneys of the patients.[64] Similarly, Law et al examined the abundance of gamma delta T cells in several “chronic” kidney diseases, including lupus nephritis, and found that patients with high numbers of gamma delta T cells had more fibrosis and worse kidney function.[65] While there was no mechanistic link provided, these observations suggest that this relatively minor population of T cells might play an outsized role in lupus nephritis.

Therapies targeting T cells

Though T cells have not been targeted with a high level of granularity in lupus nephritis, a few therapies have been geared towards diminishing their activity. Abatacept (CTLA-4Ig) blocks T cell activation (and possibly also B cell activation) by disrupting the

engagement of antigen presenting cells with the co-stimulatory molecule CD28 on T cells. A trial examining the effect of adding abatacept to cyclophosphamide and azathioprine did not show increased efficacy of abatacept over the standard protocol.[66]

However, there has been some clinical success with calcineurin inhibitors (CNI), which also block T cell activation by diminishing some of the early IL2-induced transcriptional alterations.[67] Tacrolimus and cyclosporine, while not first-line treatments for LuN, and occasionally used as add-on therapies for refractory disease.[6] In addition, voclosporin, and newer CNI, recently showed promising efficacy in clinical trials.[68]

These mixed results, along with the heterogeneity of the T cell compartment, suggest that targeting all T cells at once probably has multiple effects. Specifically targeting a particular subset (for example, Tfh), would likely be more beneficial.

Dendritic cells

Dendritic cells (DCs), which serve as “professional” antigen presenting cells, also have a role to play in the pathogenesis of LuN. As previously described, the uptake and presentation of self antigens by dendritic cells is considered to be the inciting incident in the autovaccination process. Two broad categories of DCs, myeloid (mDCs), and plasmacytoid (pDC) have been identified in humans. These two subsets have distinctive surface markers and morphology, and have distinctive roles to play. One model of how these cells might work in tandem to drive pathology suggests that pDCs are driven to maturation by immune complexes and start secreting large quantities of type I interferons. This in turn leads monocytes to mature into mDCs, take up autoantigens, and present them to both CD4+ and CD8+ T cells.[69] It has been observed that circulating mDC and pDC populations are diminished in patients with higher classes of LuN and with higher

disease activity, and that they might be enriched in the renal biopsies of these patients.[49,70–72]

Kassianos et al identified both CD1c+ DC-SIGN+ DCs and CD141hiCLEC9A+ DCs in the tubulointerstitium of LuN patients, and observed that these subsets were positively correlated with fibrosis and poor kidney function. These dendritic cell subsets are notable for their capacity to cross-present antigens to CD8+ T cells. They found that mDCs were a source of TGFβ, a pro-fibrotic cytokine (in addition to being an anti-inflammatory mediator).[73] Another study found that mDCs in early stages of LuN expressed increased co-stimulatory molecules.[71] Additionally, a subset of inflammatory DCs has been identified in the kidneys of LuN patients that are CD11c+FCRγ+CD163+, and found in the periglomerular space near T cells.[74]

Plasmacytoid DCs are typically appreciated for their production of Type I interferon, usually interferon alpha (IFNα). Tucci et al found that patient with Class IV and V LuN had circulating pDCs that highly expressed IL18R, with a corresponding overexpression of IL18 in the patients' glomeruli and an accumulation of pDCs in the periglomerular space.[49] Several studies have found that pDCs release IFNα upon engagement with immune complexes, in particular dsDNA complexed with ANA. One prominent source of this substrate is NETosis, a process in which neutrophils release their DNA to form a net with which to trap microbes. Neutrophil overactivity in SLE patients has been implicated in driving pDC maturation and production of IFNα.[75–77] Somewhat paradoxically, circulating pDCs have been observed to increase in LuN patients at 15 weeks post- cyclophosphamide induction, suggesting that they might be regulated by a cyclophosphamide-sensitive cell subset.[33]

Type I Interferon

The interferon response is a critical feature of lupus nephritis, as evidenced by transcriptomics data from LuN patients.[14,35,50,78] Type I interferons have a broad spectrum of effects, including inducing the maturation of dendritic cells, activating B cells and T cells, and inducing an inflammatory "anti-viral" response in non-immune cells.[69] In particular, they have been shown to drive germinal center formation in autoimmune contexts.[79]

Whole-tissue microarray data shows that interferon response is enriched in patients who respond to therapy during a flare, relative to those who do not.[35] Single cell RNA sequencing data from both immune cells[14] and tubule cells[80] extracted from the kidney biopsies of LuN patients showed a positive relationship between B cell abundance and interferon signaling response. This corresponded with an association between interferon response and higher patient activity indices and poor therapeutic response. This suggests that the characteristic "interferon signature" observed in LuN patients might help drive pathological B cell activation.

However, despite the apparent importance of the interferon response, clinical trials of interferon blockade have been largely unsuccessful. Two anti-IFNα antibodies, anifrolumab and sifalumumab, have been trialed. Sifalumumab failed outright, and anifrolumab succeeded only after the primary endpoints were adjusted downwards.[81–84]

6. Gap in the field

As outlined above, there is a complex array of cells and pathways that contribute to the pathology of lupus nephritis. The standard model of pathogenesis is that mDCs take up auto-antigens and present them to CD4+ T cells, which activate auto-reactive B

cells, which then differentiate into plasma cells and produce auto-antibodies. All of these interactions are enhanced by abundant type I interferons, which are often produced by pDCs. However, attempts to target this mechanism have often fallen short in terms of the successful treatment of lupus nephritis. Therefore, it is necessary to rethink this model and interrogate how these cells are interacting with each other in the context of inflamed tissue.

Flow cytometry and RNA sequencing technology have given us the ability to deeply understand what cells are present in the blood, kidneys, and urine of lupus nephritis patients. Though we have some ideas about the interplay of these cells based on correlating their abundance and prior knowledge of their functions, we largely lack insight into how these cells are interacting with each other in inflamed kidney tissue. Because flow cytometry- and RNA sequencing-based techniques require the disaggregation of the tissue, a critical piece of information is lost: the spatial arrangement of these various cell populations with respect to each other. This spatial information could provide insight into which cell populations are interacting with each other[72], and how these interactions map to divergent patient outcomes.

Immunofluorescence staining coupled with confocal microscopy allows for visualization of cellular distributions in fine detail. However, this type of data is difficult to quantify, as it has historically required a manual count of cells in fields of view, a procedure that is both laborious and time-consuming. Advances in computer vision and machine learning have enabled the automation of this task, enabling high-throughput analysis of large amounts of imaging data, a previously infeasible task. The goal of this work is to use immunofluorescence microscopy to probe the spatial distribution and

phenotype of T cells, B cells, and dendritic cells in the kidneys of lupus nephritis patients, and relate these observations to patient outcome. Our overarching hypothesis is that the difference between patients who progress to renal failure and those that do not lies in how these immune actors are organized in inflamed tissue

Chapter 2: Introduction—Immunofluorescence microscopy and image analysis

1. Introduction

Multiplexed immunofluorescence microscopy allows for the interrogation of the spatial distribution of multiple biological markers in tissue. The biggest barrier to extracting biological insights from this type of data is accurate quantification of cells and structures of interest. Though this is certainly a task that a human can perform, manual annotation of cellular-level features (i.e. surface marker expression) and tissue-level features (i.e. tumor boundaries) is an unfortunate combination of difficult, tedious, and time-consuming. It often requires a highly trained observer, such as a pathologist or microscopist, making it a difficult task to delegate. Therefore, microscopy data has historically been analyzed manually on small datasets, or only used for the purpose of displaying representative examples. Quantifying microscopy data in a high-throughput manner has traditionally not been a very tractable approach.

In recent years the field of digital pathology has been revolutionized by the application of machine learning and computer vision to image analysis tasks. Machine learning is a field of artificial intelligence focused on training computers to perform tasks that they have not been explicitly programmed for. Computer vision is a subfield of machine learning that enables computers to extract information from digital images, in much the same way the human visual system does.[85] Computer vision enables the development of software that can automate image analysis tasks, such as identifying where objects of interest are in an image and classifying those objects. This also opens the door for the development of novel metrics that describe image content, such as intercellular distances. Computer vision provides the means for high-throughput, standardized, quantitative analysis of pathology images.

Even as image analysis techniques have become more sophisticated, multiplexed microscopy technology 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 colocalization of upwards of 40 markers in an individual frame.[86–88]This level of phenotypic resolution was previously only accessible through tissue-destructive methods such as flow cytometry and RNA sequencing. The destructive processing required by these methods does not conserve the spatial arrangement of cells and other tissue structures. In contrast, highly multiplexed imaging produces rich datasets that have both detailed phenotypic and spatial information. Novel analytical solutions are therefore required to locate and segment cells and structures in these phenotype-rich images.[86–88]

In this chapter I will describe immunofluorescence microscopy, the various uses of machine learning for analyzing immunofluorescence data, as well as the spatial analyses that can be performed as a result of automating cellular identification. Much of this chapter is adapted from a mini-review that our group published in the American Journal of Pathology.[89]

2. Introduction to multiplexed immunofluorescence microscopy

Immunofluorescence is a technique whereby cellular markers in tissue are tagged with fluorescent molecules that can be detected with a microscope.[90] This tagging is performed by depositing “primary” antibodies that are specific to the marker of interest on the tissue. The deposition of the antibodies is then detected in one of two ways. The primary antibody might be directly conjugated to a fluorescent molecule (fluorophore) and imaged at the appropriate wavelength. Alternatively, a fluorophore-conjugated

"secondary" antibody that has been raised against the species of animal the primary antibody was raised in might be applied and imaged (**Fig 2.1**). For example, if one is detecting CD8, one might apply an anti-CD8 primary antibody raised in mice, followed by an anti-mouse IgG antibody that has been conjugated with AlexaFluor 488, a fluorescent molecule that is excited at a wavelength of 488 nm, and has a peak emission wavelength of 525 nm. Therefore, the distribution of CD8 in the tissue will be identified by the amount light that is detected within in a range of wavelength values around 525 nm.

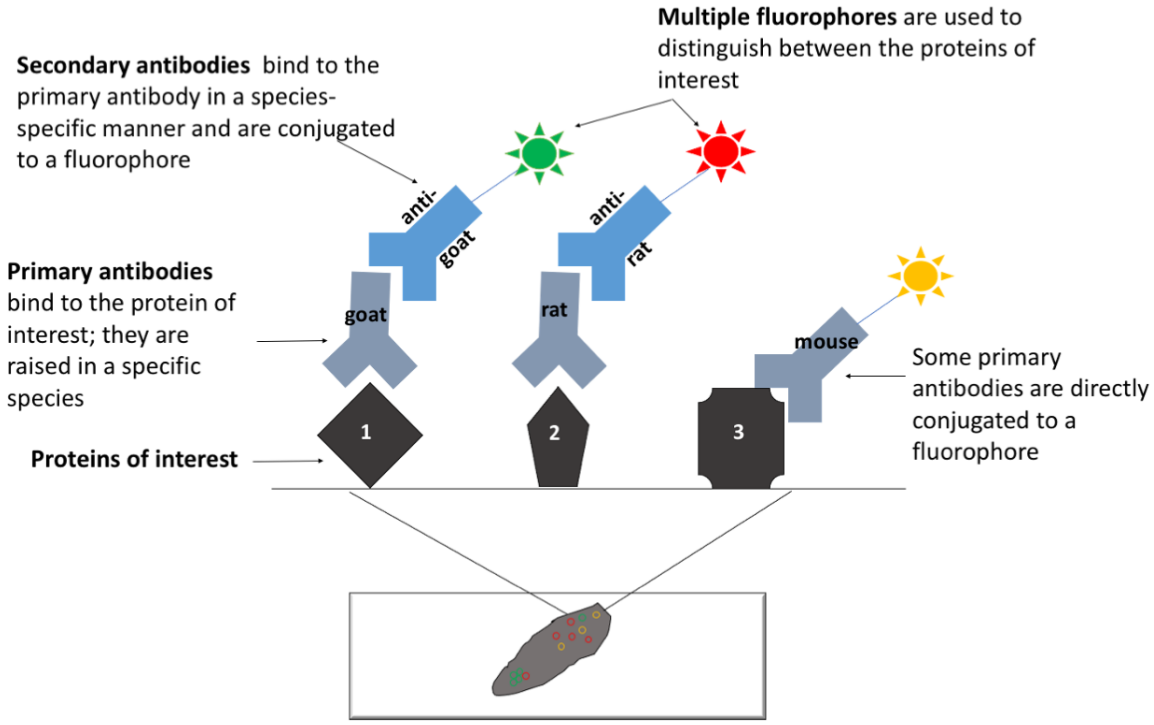

Figure 2.1. Overview of immunofluorescence microscopy. Primary antibodies bind proteins of interest in tissue and are visualized either by direct conjugation to a fluorophore, or through the application of fluorophore-conjugated secondary antibodies.

3. Essential Computer Vision Vocabulary[85]

- *Machine learning*—broad field of computing in which computers (learners) are trained to perform a task based on trial and error, rather than being explicitly programmed; computer vision is machine learning that is focused on image analysis tasks
- *Supervised learning* – type of machine learning in which training is accomplished by providing the learner with annotated examples (or "ground truth"); In the context of image analysis, this means that the algorithm is trained on a set of manually labeled images; all of the machine learning that will be described in this work is supervised learning
- *Artificial Neural Networks (ANN)* – type machine learning algorithm modeled after a human brain; a series of inputs are passed through a series of nodes (or neurons) where an operation is performed that transforms the inputs into a desired output; such that inputs are processed by several nodes in parallel; the operations that happen at the nodes typically include a numerical value or "weight" that drives the transformation of the input, and the optimal values of the weights that produce the correct output are what the network learns during training (**Fig 2.2**). A common type of ANN is a classifier that takes a series of descriptive variables as inputs and outputs a probability that the observation belongs to a particular class.

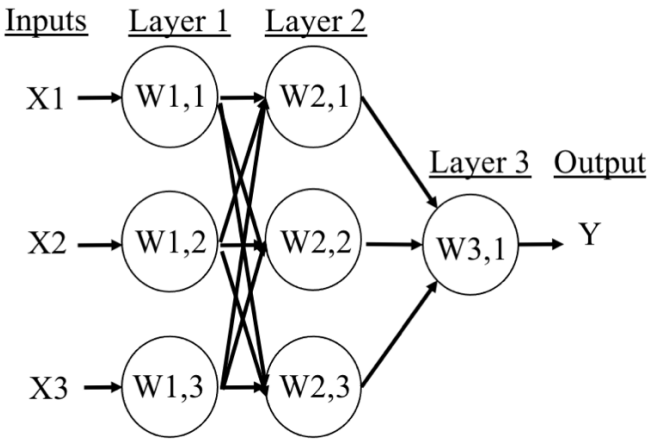

Figure 2.2. Schematic of a basic ANN. Input variables (X) are passed through a series of nodes (represented by the circles), where some function is applied to them that includes a series of weights (W), in order to produce an output (Y)

- *Hyperparameters* – aspects of the network such as the number of nodes, the number of layers, the functions that are used in the nodes, the type of loss function used for training, etc. The collection of a network's hyperparameters is commonly referred to as its *architecture.* These features can be "tuned" to optimize performance on a particular class
- *Training* – In a supervised learning context the weights of a network are randomly initialized. During the process of training the network produces an output using those weights, evaluates how "correct" that output is, and then adjusts the weights based on this evaluation. Repeating this procedure many times over the whole training dataset allows the network to find the optimal values of the weights that produce the correct output most of the time.
- *Deep learning* – style of machine learning that utilizes artificial neural networks (ANN)

- *Deep Convolutional Neural Network (DCNN)* – subclass of ANNs that are used for image analysis; images are passed through a series of filters (convolutions) that are used to extract and describe features of the image; all of the machine learning algorithms described in this work are DCNNs.
- *Segmentation* – task of classifying either pixels (semantic segmentaton) or objects (instance segmentation) in an image
- *Training set* – collection of images used for training a neural network. Typically split into 3 parts:
    - Train: images that the network uses to learn classification
    - Validation: images that the network uses to evaluate network performance during training
    - Test: images that the network has never "seen" before; used to evaluate the performance of the fully trained network
- *Epochs* -- number of times the network iterates through the entire training set during training
- *Metrics* – methods of evaluating network performance. Commonly used metrics include
    - **Precision:** $\frac{True\ Positive}{(True\ Positive+False\ Positive)}$
    - **Recall/Sensitivity:** $\frac{True\ Positive}{(True\ Positive+False\ Negative)}$
    - **Accuracy:** $\frac{True\ Positive+True\ Negative}{(True\ Positive+False\ Negative+False\ Positive+True\ Negative)}$
    - **Intersection over union (IOU):** $\frac{\text{prediction AND ground truth}}{\text{prediction OR ground truth}}$

4. Deep learning in biomedical image analysis

Deep learning caught hold in image science after AlexNet, a deep convolutional neural network (DCNN), achieved a 15% error rate (nearly 10% better than its competitors) on the ImageNet challenge, a competition in which machine learning algorithms were developed to classify a dataset of over 15 million images with 22 thousand classes[91]. The implications for medical image analysis were immediately apparent. Within the year, multiple papers applying DCNNs to pathology and radiology tasks were published, with continued growth in the application of deep learning to biomedical image analysis ever since[92]. A variety of DCNN architectures have been developed for specialized medical and biological imaging tasks since the success of AlexNet, ranging from lesion detection in magnetic resonance imaging to cell segmentation in microscopy.

DCNNs are powerful because they recognize patterns in raw image data without the constraints of human-defined equations. These patterns may not be interpretable by a human observer but are iteratively determined to be the most impactful features in making robust and accurate decisions in image classification or segmentation. Deep learning facilitates rapid analysis of high-content experimental imaging modalities, including multiplexed microscopy. DCNNs have been particularly useful for automating the segmentation and classification of regions, structures, and cells in tissue.

5. DCNN architectures for analyzing cells in microscopy images

DCNNs are typically used for classification tasks in image analysis. This classification can occur at the level of the entire image, individual pixels (semantic segmentation), or individual objects in the image (instance segmentation). There are many types of neural network architectures that are useful for these various classification tasks, and there are

often choices to be made about which architecture to use. Considerations for this include how well it performs that particular type of task, how computationally expensive the method is, and how much training data it requires.

Image Classification

In clinical image analysis, it is often of interest to classify an entire image or region. For example, pathologists might want to classify a biopsy image patch by subtype of cancer[93,94], or a fundoscopic image as "diabetic retinopathy" or "healthy".[95] There are several DCNN architectures that are commonly used in these types of analysis, including AlexNet[91], VGG-16[96], VGG-19[96], ResNet-50[97], and ResNet-100[97]. Though these networks vary in terms of the size and number of layers they have and the hyper-parameters they use, they have a few common features—they all take in images of a particular size, pass them through a series of convolutions in order to extract features from the images, and based on the features assign a probability of that image belonging to one of a pre-defined set of classes. The class that gets the highest probability "wins", and the image is assigned that class. We have used this style of DCNN to classify images in our dataset as "inflamed" or "uninflamed" on the basis of lymphocyte signal.

Semantic segmentation

For images that have complex and varied content, more granular methods of image classification are sometimes required. This next layer of complexity is often achieved with semantic segmentation, or pixel-level classification. Encoder-decoder architectures have dominated biomedical semantic segmentation since the advent of the U-Net DCNN architecture. 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. U-Nets and their derivatives have been applied to semantic segmentation tasks in medical imaging across scales, with excellent performance on full-organ segmentation in CT scans and nuclear segmentation in high resolution microscopy image.[98–100] In pathology images, semantic segmentation has been used to differentiate between tumor and healthy tissue, or segment pathogenic or pre-pathogenic areas of tissue. Cell and cell nucleus segmentation can also be performed through semantic segmentation schemes. Quantitative characteristics acquired from these segmentations, such as nuclear to cytoplasm ratio, can be indicative of cancer grade. However, semantic segmentation of cells can fail in crowded regions or in images with a low signal-to-noise or signal-to-background ratio. In this work, I will describe the tractability of using U-Nets for cellular segmentation in tissue.

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 segmentations of cells rather than image-level segmentation. This combined task, called instance segmentation, generates object-level segmentations of individual cells in an image. This allows for the separation of clustered or overlapping cells, resulting in improved cell frequency data. A relatively simple way to do this is to combine a semantic segmentation network with a region proposal network (RPN), which exclusively identify where an object of interest might be in an image. For example, RPNs have been incorporated into U-Net architectures to perform object-level cell segmentation in biomedical images.[101,102] NuSet, a U-Net + RPN architecture, shows high nuclear segmentation accuracy in a variety of contexts, including segmentation of images from

different modalities and different tissues and pathologies. The “state of the art” in terms of region-based CNNs is the Fast R-CNN family, which can perform multi-object, multi-class segmentation with high accuracy in natural images, and have also shown promise in biomedical image analysis tasks, including cellular segmentation in a variety of contexts.[103,104]

An instance segmentation architecture that has been particularly useful to us is the Mask R-CNN[104], a member of the Fast R-CNN family. This network combines a feature pyramid network with a ResNet-100 backbone with a region proposal network to perform multi-class instance segmentations. In our data, it has been used to great effect in order to identify and segment multiple subsets of immune cells in biopsy tissue.

6. Generalizability of methods

A concept that is of great interest in the machine learning community is that of generalizability. This refers to the ability of a given algorithm to be applied to a task that it was not strictly trained for. The simplest version of this is generalizability to new data of that is not in the training set. One pitfall of training a neural network is over-fitting, wherein the network simply “memorizes” and performs extremely well on the training data but fails when confronted with new data.[105] Network generalizability is routinely evaluated by checking the performance of the network on a independent test set of images it has never encountered.

However, generalizability can also be extended to using a network for a task that is different from its original purpose.[106] If the new task is similar enough, this can be done without changing the network. However, as the task deviates from the original purpose, there is often some manipulation that needs to happen in order to make the algorithm

work. Transfer learning refers to a set of methods by which networks that have been trained on one task can be repurposed.[107] Finetuning is the practice of "freezing" some of the layers of the network, and then either retraining the model or training a new set of layers, using a different set of ground truth for the new task. In both cases, you leverage the training that has already happened in order to use less training data and time for your new task.

The reason why generalizability is a preoccupation of supervised machine learning practitioners is that training algorithms from scratch requires a lot of annotated data and computation time. Developing clever methods of leveraging pre-existing networks for new purposes is therefore an active field of study. In particular, because generating training sets for microscopy data is so time and labor intensive, developing generalizable methods allows us to avoid developing a specialized analytical method for every dataset.

7. Examples of current open-source image analysis software

Several open-source implementations of machine learning tools have been designed for use by scientists who may not have the extensive background in programming required for low-level algorithm control. Many of these programs employ pre-trained DCNNs for common tasks such as nuclear segmentation. Some enable training machine-learning models from scratch. One of the benefits of these open-source pipelines is that they include graphical user interfaces (GUIs), making them user-friendly and accessible to a broader audience. However, this 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/Fiji

ImageJ (also known as 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 produced 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 GUI[108]. Similarly, the lab that invented the U-Net developed a U-Net plug-in ("Unet-segmentation") that uses the caffe deep learning framework to implement the architecture in ImageJ. DeepImageJ can run pre-trained models in ImageJ, and even comes with a set of models for particular tasks bundled with it[109]. 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's Trichrome, or Jones stain image[110].

CellProfiler

CellProfiler is a widely used software that has been cited more than 9000 times since its original publication in 2006[111]. It has a "point and click" GUI that allows users with minimal programming experience to string together several image analysis modules into an analytical pipeline. CellProfiler comes with more than 50 modules that allow for standard image analysis procedures, and further allows users to write their own modules. As of 2018, CellProfiler 3.0 featured a pre-trained U-Net based semantic segmentation module (ClassifyPixels-Unet)[112]. The drawback is that no task-specific training or fine-tuning can be applied, which can limit its utility for tissue microscopy analysis, as tissue background can have 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 al. wrote a module that allowed users to run pre-trained caffe-based models within CellProfiler pipelines[113].

Ilastik

Ilastik is a deep-learning software that performs semantic segmentation tasks by extracting a series of pre-defined pixel-level features to train random forest classifiers.[114] 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[115,116]. Several recent papers have combined CellProfiler with ilastik for segmenting cells in tissue microscopy images. For example, a CellProfiler pipeline can be generated to perform pre-processing steps, then ilastik can be used to produce a "probability map" for semantic segmentations[87,117].These maps are then returned to CellProfiler for object detection.

QuPath

QuPath is an open-source platform that was designed specifically for whole-slide image analysis and can be used for both immunohistochemistry and immunofluorescence imaging[118].Like ilastik, it employs random forest classifiers that can be trained to segment cell classes on the basis of cell surface marker expression. It is useful for quantifying the distribution of a particular marker across a whole tissue section.

Limitations of these tools

While these tools are very useful for making machine learning more widely accessible, they are not ideal for every situation. The inability to fine-tune many of the pre-trained

networks means that if the method doesn't work "out of the box", there is not much recourse.

For example, we evaluated the performance of CellProfiler for segmenting T cells in kidney biopsies, according to a published protocol.[119] We found that the pipeline did not perform well on our data, in the context of tissue that has dense cells and tissue-specific background structures (eg, renal tubules). (**Fig 2.3**). This level of segmentation was not sufficient for the cell frequency data and spatial analysis we were aiming for. Therefore, though it was certainly desirable to use one of these methods to analyze our data, we ultimately used neural network architectures that gave us greater control over the algorithms and the ability to tune them to our purpose.

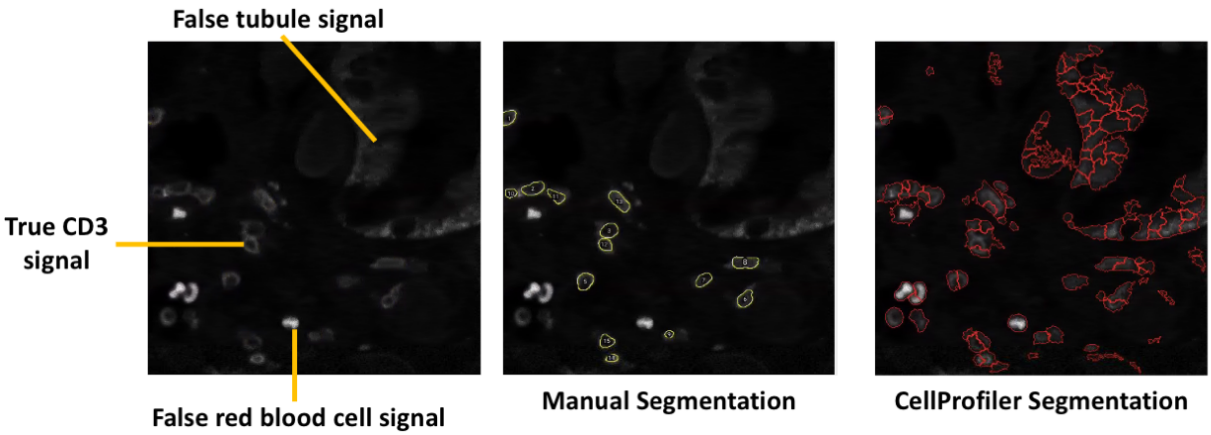

Figure 2.3. Example of CellProfiler performance on kidney biopsy tissue stained for CD3. Representative image contains true CD3 signal in addition to autofluorescent red blood cells, and tubules

8. Quantifying spatial organization of cells in microscopy images

Cellular segmentation is often not the end point of the analysis, but the beginning. These techniques are widely applied to automate the extraction of quantifiable metrics from cellular images to ask biological questions. This section will address how "algorithmically defined" features are used in downstream analyses.

Automatic annotation enables more complex spatial analyses than previously possible. Specifically, interrogating how various cellular classes are spatially organized with respect to each other unlocks the ability to probe how the cellular constituents of tissue interact with each other. This allows for the generation of novel features that can be used to stratify patients, and for increased understanding of the cellular processes that occur in tissue.

This idea has been applied extensively in the study of tumor immunology. Nearchou et al examined sequential sections of surgically resected colorectal cancer specimens and developed a “Spatial Immuno-Oncology Index” (SIOI) that was constructed of patient-level spatial features including average CD3+ density, average number of lymphocytes in proximity to tumor buds and the ratio of CD68+/CD163+ macrophages and used as a predictor of prognosis.[120] Similarly, Lazarus et al examined the abundance and spatial distribution of various subsets of T cells, antigen presenting cells (APCs), and tumor cells in liver metastases of colorectal cancer.[121] The automated detection and segmentation of immune cells facilitated the finding that cytotoxic T lymphocytes were typically positioned further away from epithelial cells and APCs that were expressing the checkpoint molecule PD-L1. Further they found that high levels of “engagement” (defined as intercellular distances under 15 um) between CTLs and epithelial cells, helper T cells, and regulatory T cells was predictive of superior survival. In another study, the Mask R-CNN architecture to segment lymphocytes and tumors cells in whole-slide H&E images of hepatocellular carcinoma samples.[122] 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 3 distinct subgroups within the patient population.

While developing prognostic tools is useful, associating specific intercellular interactions with positive or negative patient outcomes is also the first step in identifying immunological phenomena that have potential as therapeutic targets. Understanding which cellular interactions are taking place in tissue, and how these interactions correlate with outcome allow for the identification of interactions that should be promoted and those that should be disrupted.

9. Gap in the field

In addition to learning about T cell-APC interactions in the specific context of lupus nephritis, the goal of this work is to develop comprehensive pipelines for extracting meaningful biological insight from microscopy data. In order to make full use of the potential of immunofluorescence microscopy, solving the very fundamental problem of how to identify cells in tissue is a key obstacle to overcome. Segmenting cells in culture is a relatively easier task because there is little background autofluorescence, all of the cells are in a single plane of focus, and the density of cells in an area can be controlled. Accordingly, there is a substantial set of tools that have been developed for this type of task that require little additional development for specific experiments. Segmenting cells in tissue is more challenging, as every tissue type has its own morphology, confounding structures, and signal to noise ratio. We have done substantial work to develop methods and best practices for segmenting cells in renal tissue and demonstrate how methods can generalize to other disease contexts.

In addition to cellular segmentation, this work has focused on developing ways of mathematically probing imaging data for spatial relationships. This fundamentally means taking biological hypotheses and translating them into a concrete metrics (such as intracellular distance) that can be extracted from image data and used to compare groups in a statistically robust way. The innovation that was required for this project to succeed was to take the broad hypothesis that “inflammation is organized in distinct ways that can predict patient outcomes” and break it down into discrete, measurable phenomena that are captured in image data.

As I’ve outlined above, this has been done to some degree in the field of cancer, and interesting insights have come even from relatively simple metrics such as lymphocyte-tumor distance. However, there is a lack of similar work in the field of autoimmunity in general, and lupus nephritis in particular. Autoimmunity is just as complex and heterogeneous as cancer and would benefit in a similar way from these types of analysis. Therefore, we have adapted and extended these methods for our purposes, in order to investigate *in situ* inflammation and understand how it is organized.

Computer vision has transformed microscopy image analysis and is a rapidly evolving discipline impacting clinical and basic science research. The automation of image classification and cell detection using computer vision techniques facilitates high-throughput analysis of information-dense microscopy images. Our goal in this work is to develop robust pipelines by which machine learning-enabled feature extraction can generate novel insights into cellular environments associated with disease.

Chapter 3: Cellular segmentation and spatial analysis in the high resolution dataset

The work described in Chapters 3-4 was done in collaboration with several people -- Dr. Junting Ai acquired the data, Dr. Madeleine Durkee led the development of the automated cellular segmentation method, and Gabriel Casella performed some of the spatial analyses

1. Introduction

The first set of questions we wanted to answer were about whether there is a relationship between the inflammatory state of the kidneys of lupus nephritis patients and their renal outcomes. Specifically, we were interested in the distribution of T cells, B cells, and dendritic cells at the time of diagnostic biopsy. In order to answer these questions, a microscopy dataset was acquired for the purpose of interrogating the interactions between T cells and antigen presenting cells in the context of lupus nephritis. This dataset, which I will refer to as the “high resolution” or HR dataset, was collected from formalin fixed, paraffin embedded (FFPE) biopsies from lupus nephritis patients for whom we had at least 2 years of follow up.

In order to extract meaningful information from this dataset we first segmented the following cellular classes in the images: CD3+CD4- T cells, CD3+CD4+ T cells, CD20+ B cells, CD11c+ mDCs, and BDCA2+ pDCs. We then developed a set of spatial analyses with which to probe this data. This work was done as part of an ongoing collaboration with the laboratory of Dr. Maryellen Giger.

2. Data Acquisition

An 865-image dataset of microscopy images was collected from FFPE kidney biopsies of 55 lupus nephritis patients. The biopsies were stained for a panel of 5 cell surface

markers, plus a nuclear marker: CD20 (AlexaFluor 488), CD3 (AlexaFluor 546), BDCA2 (AlexaFluor 594), CD4 (AlexaFluor 647), and CD11c (AlexaFluor 700). These five markers were used to identify five classes of cells: CD3+CD4- T cells, CD3+CD4+ T cells, CD20+ B cells, CD11c+ myeloid dendritic cells, and BDCA2+ plasmacytoid dendritic cells (**Figure 3.1**). The data was collected on the Leica SP8 laser scanning confocal microscope at 63x magnification, with a pixel size of 0.1058 um. The biopsies were imaged at this resolution in order to capture cellular shapes and visualize intercellular interactions. This produced image stacks that were 1024 pixels x 1024 pixels x 6 channels, such that every marker was assigned to a channel.

For this set of patients we had detailed clinical data with at least 2 years of follow-up, and an average of 6 years of follow-up. In particular, we were able to group the patients by whether or not they progressed to end stage renal disease (ESRD), or kidney failure. We split the patients into two groups: 19 ESRD+ and 36 ESRD-. We also did an analysis where we split the cohort into three groups, with the ESRD+ group being stratified by patients who were within 2 weeks of renal failure (ESRD-current, n=5), and those whose kidneys failed later (ESRD+, n=14). This allowed us to relate the cellular content of their biopsies with their clinical data. This clinical data is described in detail in **Chapter 4**.

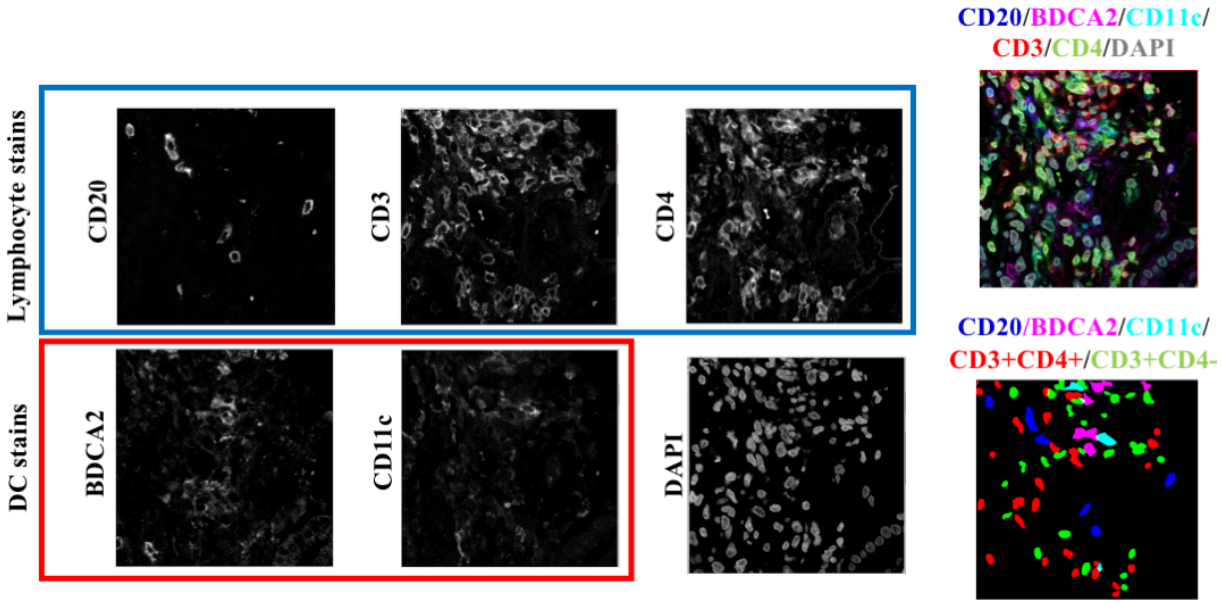

Figure 3.1. Representative images of the staining panel for the high resolution dataset, and example RGB segmentation of the 5 classes of interest

3. Cellular Segmentation

Ground Truth generation

The approach we took to automatically segment cells in this dataset falls under the category of "supervised learning", a machine learning approach in which manual annotations are the gold standard that the network is trained on. It was therefore necessary to manually outline the cells in a training set of 342 images from 31 biopsies. This manual annotation was performed with the ImageJ/Fiji software, using the "freehand selections" tool. Readers were provided with "spectrally unmixed" versions of the images, wherein aberrant and overlapping signal from adjacent wavelengths (aka "spectral bleed-through") were removed from each channel. High quality truth was obtained by having two readers validate the manual annotations.

Mask R-CNN

Mask R-CNN[104] is an instance segmentation algorithm that has three main parts: 1) a feature pyramid network (FPN) with a ResNet-101 backbone for feature extraction and

2) a region proposal network (RPN) for object detection 3) fully connected layers for simultaneous mask generation, bounding box regression, and cell classification.

In the FPN structure, feature maps are generated from the ResNet-101 blocks at multiple scales and then merged. These ‘multiscale’ feature maps are passed through 3x3 convolutional layers in preparation for input into the region proposal network (RPN). At each position of the feature map, anchor boxes of predetermined size and aspect ratio are extracted as patches to be fed into the RPN, which 1) determines whether each patch has a cell in it and 2) defines the bounding box around the cell via bounding box regression. The resulting cell proposals are then re-aligned with the feature map, and then each one progresses in parallel through 1) fully connected layers for classification and further bounding box regression, and 2) mask generation to define cellular boundaries (**Figure 3.2**).[123]

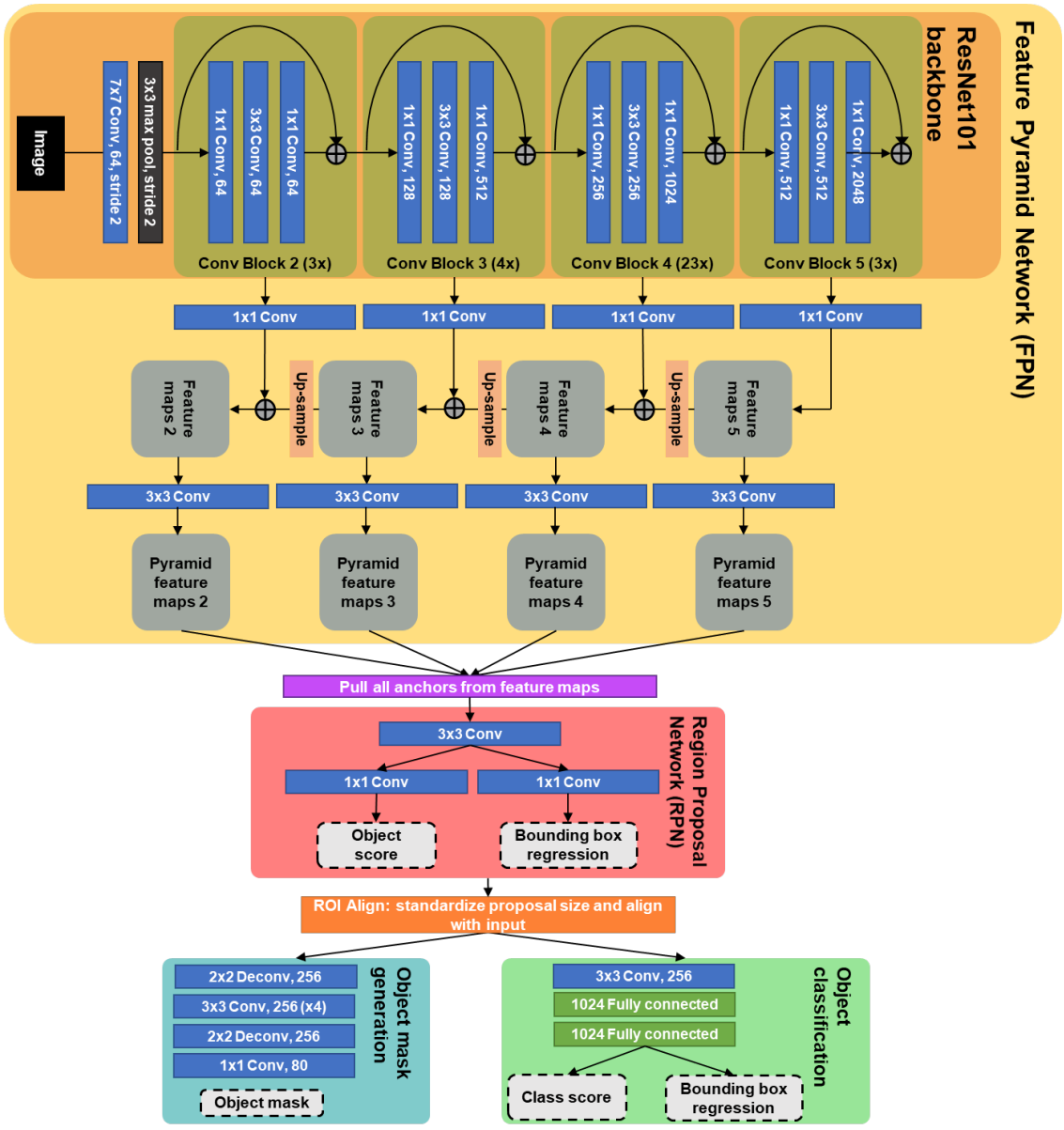

Figure 3.2. Mask R-CNN architecture. Each network trained to segment and classify immune cells is a Mask R-CNN architecture. Object proposals are performed on feature maps from the DCNN, then single objects (cells) are semantically segmented and classified.

Each instance of Mask R-CNN was trained until the mean average recall for all classes stopped improving. A cell was kept for analysis if the network confidence in the prediction was above 0.3. Data augmentation is the practice of altering images in the training set to artificially increase the number of objects that are trained. For this dataset,

random flips and rotations, as well as brightness and gamma augmentation were used for data augmentation.

This work was performed using the Midway2 compute nodes of the University of Chicago Research Computing Center. Each network was trained separately on each dataset, with a batch size of 4 distributed across 4 Nvidia K80 GPUs (12 GB memory each) using the Horovod distributed deep learning framework[124].

Training Strategy

There was substantial class imbalance in the training set, such that approximately 85% of the manual annotations were lymphocytes (CD20+, CD3+CD4-, or CD3+CD4+), and only 15% dendritic cells (BDCA2+ and CD11c+). Due to this imbalance the network performance on lymphocytes was maximized much faster than the performance on dendritic cells. Additionally, the dendritic cell images were qualitatively much noisier than the lymphocyte images and readers had less confidence in the ground truth cell calls. Over-training occurs when the network learns the unique features of the dataset and begins to lose generalizability. In this situation the network was at risk for over-training on lymphocytes before it achieved adequate performance on dendritic cells. Even bolstering the training set with additional DC-rich images did not mitigate this problem.

Therefore the task was split into two—separate instances of Mask R-CNN were trained for lymphocytes and for dendritic cells. The lymphocyte network was trained on stacks of CD20, CD3, CD4, and DAPI, and the DC network was trained on stacks of BDCA2, CD11c and DAPI (**Figure 3.3**). The full 342 image training sets was split into training, validation, and testing sets with an 80/10/10 ratio.

The predictions were then merged in a hierarchical manner, such that if a cell was predicted as two classes, one was chosen. This proved to be a superior method, which uncoupled the performance on lymphocyte classes from performance on DCs. The final network performance is shown in **Table 3.1**.

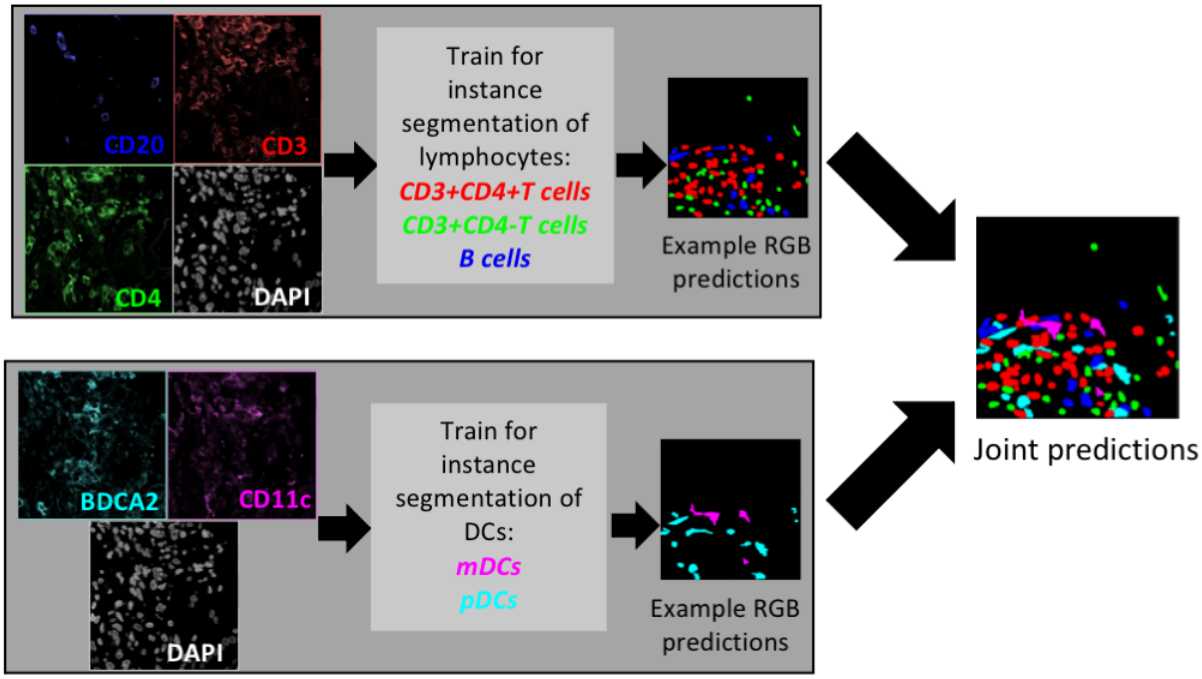

Figure 3.3. Schematic of segmentation. Each ROI is split into two stacks of images (CD20/CD3/CD4/DAPI and BDCA2/CD11c/DAPI) and predicted on by two different neural networks; the outputs are then merged into a final joint prediction.

Table 3.1 Summary of final network performance

	Recall	Precision	Mean IOU
CD3+CD4+ T cell	0.8	0.77	0.73
CD3+CD4- T cell	0.72	0.67	0.72
B cell	0.9	0.87	0.7
pDC	0.74	0.85	0.76
mDC	0.46	0.51	0.64
Average	0.77	0.71	0.71

4. Spatial analysis

All spatial analyses were performed in the programming language Python (3.6.6). The following packages were utilized for analysis: pandas (0.24.1), numpy (1.5.4), sklearn (0.20.2), and scipy (1.2.0). Plotting was performed with matplotlib (3.0.3) and seaborn (0.10.1).

Per ROI analysis

The per ROI analyses were performed by separating the 1024x1024 ROIs into pools based on the patient's renal outcome. Each ROI was treated as an observation and characterized by the number of cells of each class, and these features were compared between the pools using the Mann-Whitney U-test with a Bonferroni correction.

Bootstrapping

In order to mitigate the issue of having different numbers of patients in each renal outcome group, the individual pools of ROIs were sampled with replacement to produce samples of 200 ROIs each for the two class (ESRD+ vs ESRD-) analysis, and 150 ROIs each for the three class (ESRD+ vs ESRD- vs ESRD current) analysis. The average of each feature in the two groups was calculated, as well as the difference between these two averages. This procedure was repeated 1000 times until distributions for the difference in means between the cohorts were produced. A 95% confidence interval of the differences in means that does not overlap with zero suggests that the two populations are significantly different from each other (**Figure 3.4**).

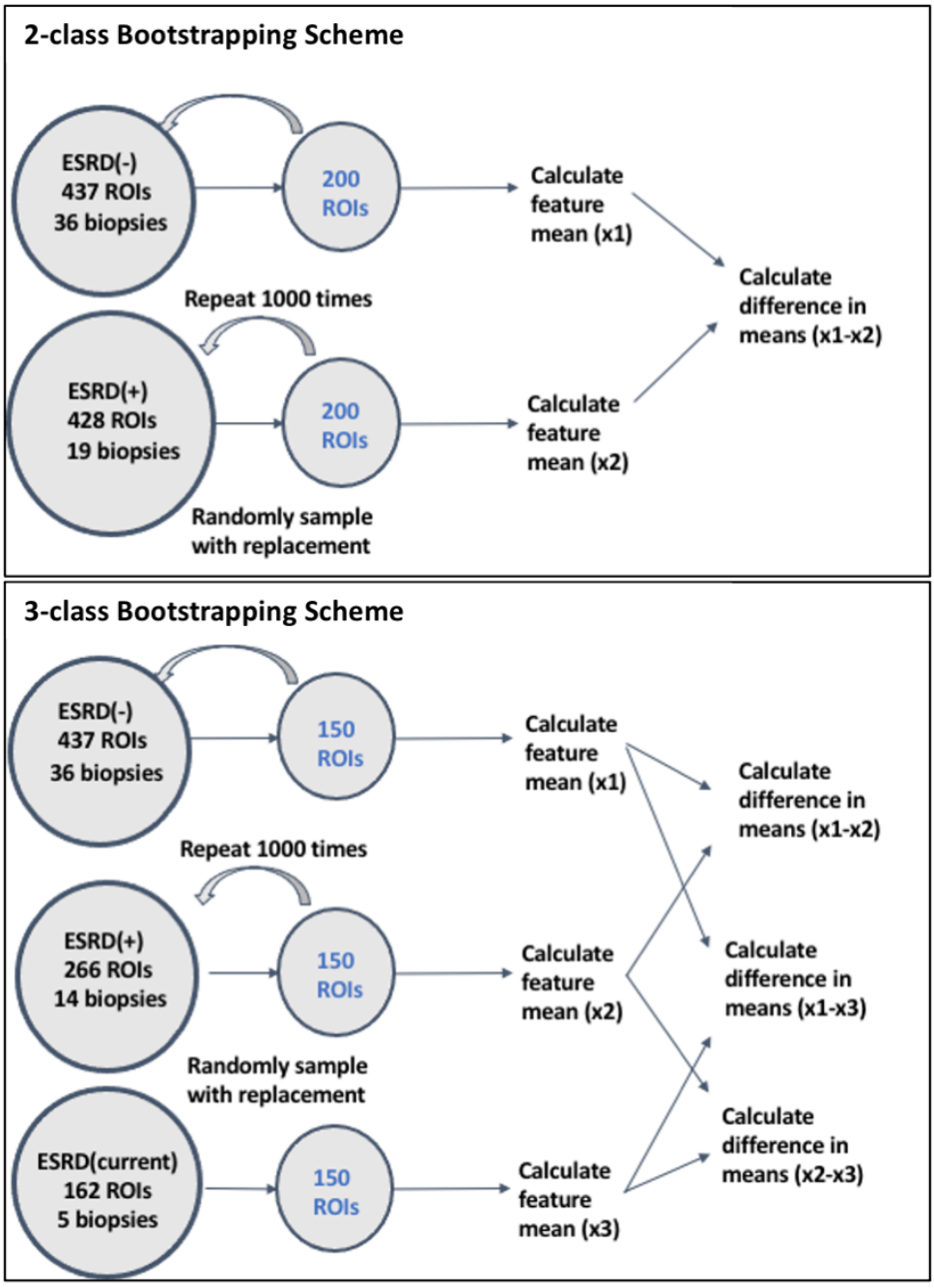

Figure 3.4. Schematic of 2-class and 3-class bootstrapping procedures. Pools of ROIs from each renal outcome group are sampled with replacement, the feature mean for each sample is calculated, as is the difference in the means; the 95% confidence interval for the difference in means is used to determine whether two groups are significantly different.

Nearest Neighbors and Niche Analyses

We wanted to probe the potential interacting partners of each cellular class. We therefore did two related analyses. The first was a nearest neighbors analysis that iterated through every cell in the dataset and identified the class of the segmented cell that was closest to it based on centroid to centroid distance (**Figure 3.5A**). The second was a niche analysis, inspired by the approach taken by Goltsev et al.[125] In this analysis the niche of a each cell was defined as all of the cells that were within 100 pixels of the "center" cell (**Figure 3.5B**). Therefore every cell could be described by how many cells it had near it, and what the classes of those cells were.

Neighborhood analysis

To extend this notion of interacting groups of cells one step further, we then defined "neighborhoods" of cells based on spatial proximity. Cells were assigned to neighborhoods using Density Based Spatial Clustering of Applications with Noise (DBSCAN), a clustering algorithm that assigns points to groups based on whether they are within a minimum size cutoff away from each other, with a defined minimum group size.[126] Any points that cannot be assigned to a group are referred to as "singlets". For this analysis we used a distance cutoff of 100 pixels, which is close to 10 um at this resolution (**Fig 3.5C**).

"Types" of cellular neighborhoods were identified by extracting 24 features that capture cellular constituency as well as cell and cluster shape were extracted for each cluster, and applying K-means clustering with k=6. In order to find the optimal number of k cluster for the K-means clustering algorithm, we adopted a bootstrapped clustering scoring procedure. During each bootstrap iteration, 70% of the normalized neighborhood

data was randomly sampled. Afterwards, the K-means clustering algorithm from k = 2 to k = 24 number of clusters was fitted to the data, the within cluster sum of distances squared, change in within cluster sum of distances squared was calculated for the K clusters. This procedure was repeated a total of N= 250 times. The six clusters were then characterized using a leave-one-out t test such that the current cluster of reference was treated as the alternative group, all the remaining clusters were then binned together as the reference group, and a t-test was performed to identify distinguishing features of that cluster.

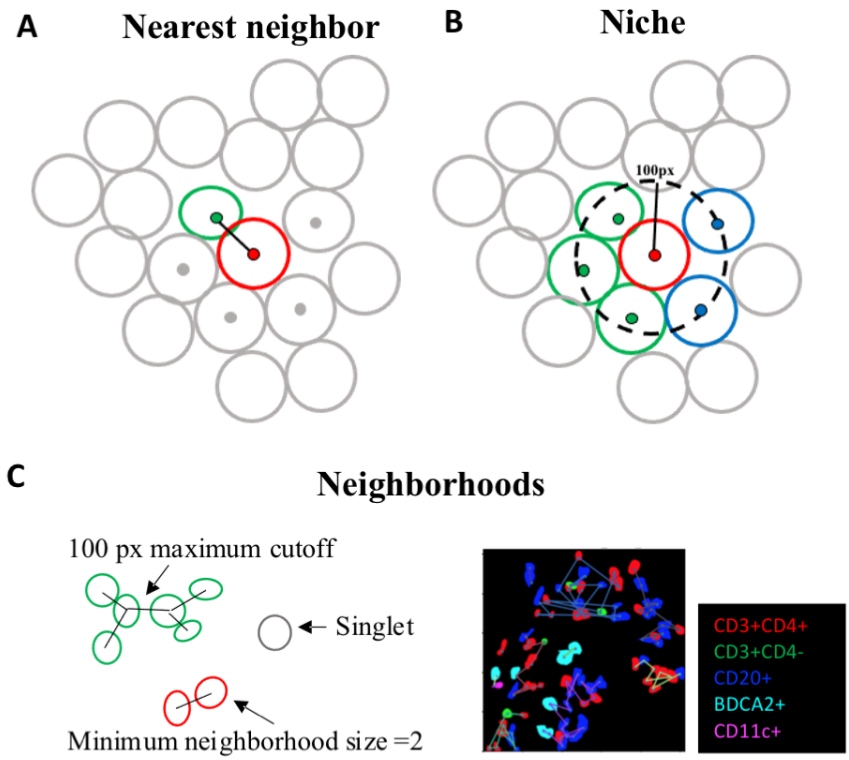

Figure 3.5. Spatial analyses of immune cells. A) Nearest neighbor calculation in which the minimum centroid-centroid distance between each cell and its neighbor is calculated; B) niche calculation in which for every cell, all of the cells that are within 100 pixels are identified; C) neighborhood calculation, in which cells are grouped into distinct neighborhoods using DBSCAN with a minimum cluster size of 2, and a distance cutoff of 100 pixels.

5. Summary

In order to probe the interactions between T cells, B cells, and DCs in lupus nephritis patients, we collected immunofluorescence microscopy data from the diagnostic kidney biopsies of 55 lupus nephritis patients. Accurate segmentation of the cells in the tissue was obtained by training two instances of Mask R-CNN to segment five classes of cells: CD20+ B cells, CD11c+ mDCs, BDCA2+ pDCs, CD3+CD4- T cells, and CD3+CD4+ T cells. Once we had adequate cellular segmentations we performed a series of spatial analyses to quantify the distribution and organization of cells in the tissue.

6. Acknowledgements

I would like to thank and acknowledge the following people who contributed to this part of the project: Dr. Junting Ai from the Clark lab acquired the microscopy data and helped with ground truth generation. Dr. Madeleine Durkee was responsible for training and optimizing the performance of the Mask R-CNN on this dataset. Gabriel Casella performed the neighborhood analysis. Dr. Anthony Chang, Dr. Kichul Ko, Dr. Charles Oshinsky, and Dr. Emily Peninger were responsible for curating this patient cohort. Dr. Karen Drukker provided input on the bootstrapping analyses. The Integrated Light Microscopy Core Facility at the University of Chicago (Dr. Christine Labno, Shirley Bond and Dr. Vytas Bindokas) provided expertise on imaging. The Research Computing Center at the University of Chicago provided the computational resources.

Chapter 4: Identifying inflammatory states that correlate with renal failure in lupus nephritis patients

1. Introduction

The ultimate objective of the high resolution (HR) LuN dataset was to explore the spatial interactions between T cells and antigen presenting cells in the biopsies of our cohort of lupus nephritis patients and relate these features to their clinical outcomes. Though there is a positive relationship between tubulointerstitial inflammation and renal failure[10], inflammation is an imperfect predictor of disease outcome. Indeed, of the patients in our cohort with at least 2 years of follow up who had moderate/severe TII scores of 2-3, roughly 40% did not progress to renal failure. This suggests that the overall level of inflammation is not an accurate predictor of patient outcome in lupus nephritis and motivates the need to probe the nature of inflammation, including which cellular classes are present and how they are interacting with each other. Here, we use spatial proximity as a proxy for interactions, given that immunological phenomena like antigen presentation require that cells are physically near each other. This suggests that the local density of cells that can engage with each other can modulate the overall impact of inflammation on tissue. We hypothesized that understanding such spatial relationships would provide insights into cognate recognition and the organization of the adaptive immune system in *in situ* inflammation.[37,72]

Briefly, this dataset was collected to probe CD3, CD4, CD20, BDCA2 and, CD11c in order to capture CD3+CD4- T cells, CD3+CD4+ T cells, CD20+ B cells, CD11c+ mDCs, and BDA2+ pDCs. Regions of inflammation were selectively imaged based on the presence of a positive CD3 signal. A total of 865 1024x1024 pixel images (0.1058 μm/pixel) from 55 patients were used in the analysis. Because there is variation in the

amount of inflammation present in a given biopsy, the number of images from a given sample varies from 1 to 82. For each patient we have clinical follow up data for at least 2 years, including whether or not they proceeded to end stage renal disease (ESRD) during that time.

2. Clinical description of our patient cohort

The strength of this dataset comes from the detailed clinical information that we have to go with the biopsies. For each biopsy, we have information about the patient's demographics, clinical course, renal function, and whether or not they progressed to kidney failure or end stage renal disease (ESRD), defined as needing dialysis or transplant. In our patient cohort, 19 patients progressed to end stage renal disease (ESRD+), and 36 did not (ESRD-). **Table 4.1** provides a description of these two patient cohorts. These two groups of patients (ESRD+ vs ESRD-) were not significantly different from each other in terms of the time of follow up, duration of disease at the time of biopsy, or patient age at the time of biopsy (**Fig 4.1A-C**). These factors together indicate that differences that are observable between these two groups are not merely a function of a more advanced disease state in the ESRD+ patients.

Table 4.1 Summary of clinical features in high resolution LuN dataset

Feature	Categories	ESRD- (n=36)	ESRD+ (n=19)
Sex	Female	33 (92%)	17 (89%)
	Male	3 (8%)	2 (11%)
Race	African American	29 (81%)	19 (100%)
	Non-African American	7 (19%)	0 (0%)
Induction	Cyclophosphamide	15 (42%)	9 (47%)
	MMF	14 (39%)	5 (26%)
	Unknown	7 (19%)	5 (26%)

Table 4.1 Summary of clinical features in high resolution LuN dataset, continued

Feature	Categories	ESRD- (n=36)	ESRD+ (n=19)
pre-biopsy plaquenil	Yes	18 (50%)	13 (68%)
	No	18 (50%)	6 (32%)
pre-biopsy MMF	Yes	8 (22%)	9 (47%)
	No	27 (75%)	10 (53%)
	Unknown	1 (3%)	0 (0%)
pre-biopsy prednisone >20 mg	Yes	8 (22%)	5 (26%)
	No	28 (78%)	14 (74%)
pre-biopsy azathioprine	Yes	0 (0%)	2 (11%)
	No	35 (97%)	17 (89%)
	Unknown	1 (3%)	0 (0%)
Ace inhibitor/ARB	Yes	28 (78%)	8 (42%)
	No	8 (22%)	11 (58%)
Class 5 glomerulonephritis	Yes	19 (53%)	8 (42%)
	No	17 (47%)	11 (58%)
proliferative glomerulonephritis	Yes	29 (81%)	16 (84%)
	No	7 (19%)	3 (16%)
Moderate-Severe TI (>1) score	Yes	21 (58%)	17 (89%)
	No	15 (42%)	2 (11%)
Moderate-Severe Chronicity (>=4) score	Yes	7 (19%)	14 (74%)
	No	29 (81%)	5 (26%)
Hypertension	Yes	17 (47%)	12 (63%)
	No	19 (53%)	7 (37%)
dsDNA	Yes	30 (83%)	18 (95%)
	No	6 (17%)	1 (5%)
Doubled serum creatinine during follow-up	Yes	9 (25%)	10 (53%)
	No	27 (75%)	9 (47%)

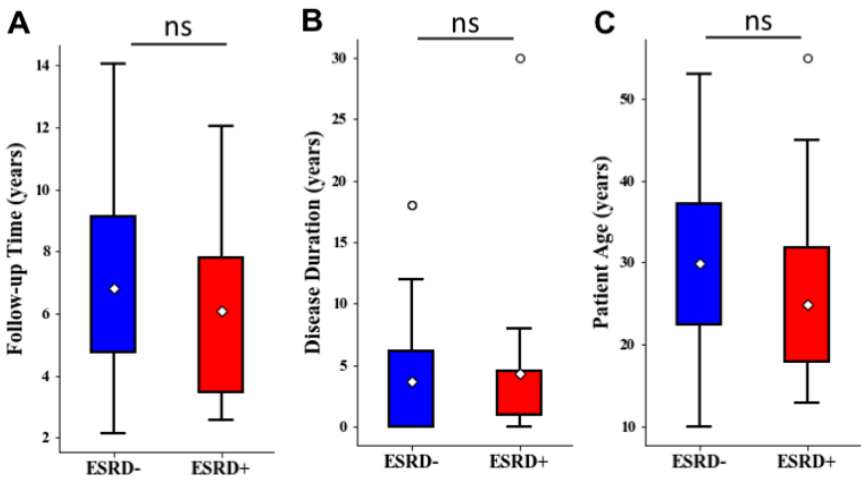

Figure 4.1. Clinical characteristics of patient cohort. Comparison of ESRD+ and ESRD- patients in terms of A) duration of follow up period, B) disease duration and C) patient age at the time of biopsy

3. Comparison of cellular density in ESRD- and ESRD+ patients

To evaluate the relative densities of the five cellular classes in the tissue, we looked at the cellular content of each 1024x1024 pixel region of interest (ROI), representing ~1.2x10^4 um^2 of tissue. Comparison of overall cell densities (total cells/ROI) in ESRD- patients (437 ROIs) and ESRD+ patients (428 ROIs) revealed no significant differences (**Fig 4.2A**). However, the average total cell count per biopsy was higher in the ESRD+ cohort, reflecting larger overall areas of inflammation (**Fig 4.2B**). In contrast to overall cell density, there were apparent differences in the cellular constituents of inflammation between these two patient cohorts. Surprisingly, we observed that ROIs from ESRD- patients were enriched for CD20+ B cells relative to ESRD+ ($p=1.23x10^{-7}$) (**Fig 4.2C**). ROIs from ESRD+ patients showed a significantly higher density of CD4- T cells ($p=3.40x10^{-15}$), suggesting an association between a high density of these cells and renal failure. (**Fig 4.2D**). There were no differences in the number of CD3+CD4+ T cells,

BDCA2+ pDCs, and CD11c+ mDCs per ROI between the two patient cohorts (**Fig 4.2E-G**).

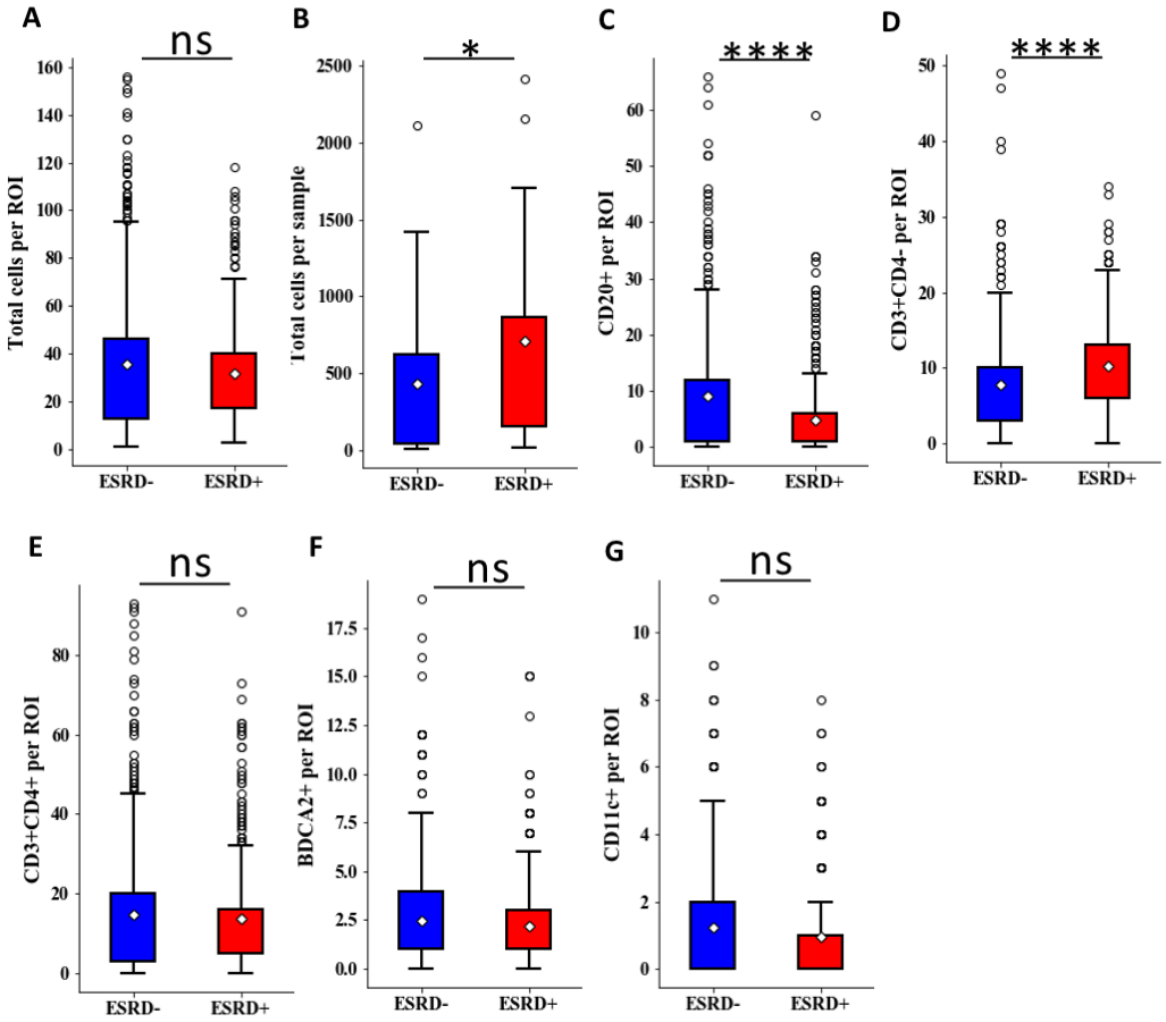

Figure 4.2. Cellular densities in ESRD+ and ESRD-. Difference in the number of cells of each class per ROI between ROIs from patients who progressed to ESRD (ESRD+, n=428 ROIs) and those who do not (ESRD-, n=437 ROIs); A) Total cells/ROI; B) Total cells/sample; C) CD20+ cells/ROI; D) CD3+CD4- cells/ROI; E) CD3+CD4+ cells/ROI; F) BDCA2+ cells/ROI; G) CD11c+ cells/ROI

In order to specifically interrogate regions of dense inflammation, I next considered only ROIs that had at least 20 segmented cells in them. Restricting the analysis in this manner eliminated images from several samples—reducing the total number of patient

samples to 45; 28 ESRD- samples (252 ROIs) and 17 ESRD+ samples (295 ROIs). It also made apparent differences between the inflamed regions in ESRD- and ESRD+ samples. In images from ESRD- patients, there was a substantial increase in the total number of cells ($p=8.55x10^{-5}$) (**Fig 4.3A**). This suggests that, particularly within inflamed regions, there is a tendency for ESRD- ROIs to have more densely packed cells. Accordingly, there were on average more CD20+ cells/ROI ($p=5.94x10^{-16}$), CD3+CD4+ cells/ROI ($p=0.005$), and BDCA2+ cells/ROI ($p=0.043$) in ESRD- ROIs than ESRD+ (**Fig 4.3B, D-F).** Notably, the only cell class that was present at a higher density in ESRD+ were CD3+CD4- T cells ($p=5.56x10^{-5}$), even when all other cell types were diminished (**Fig 4.3C**).

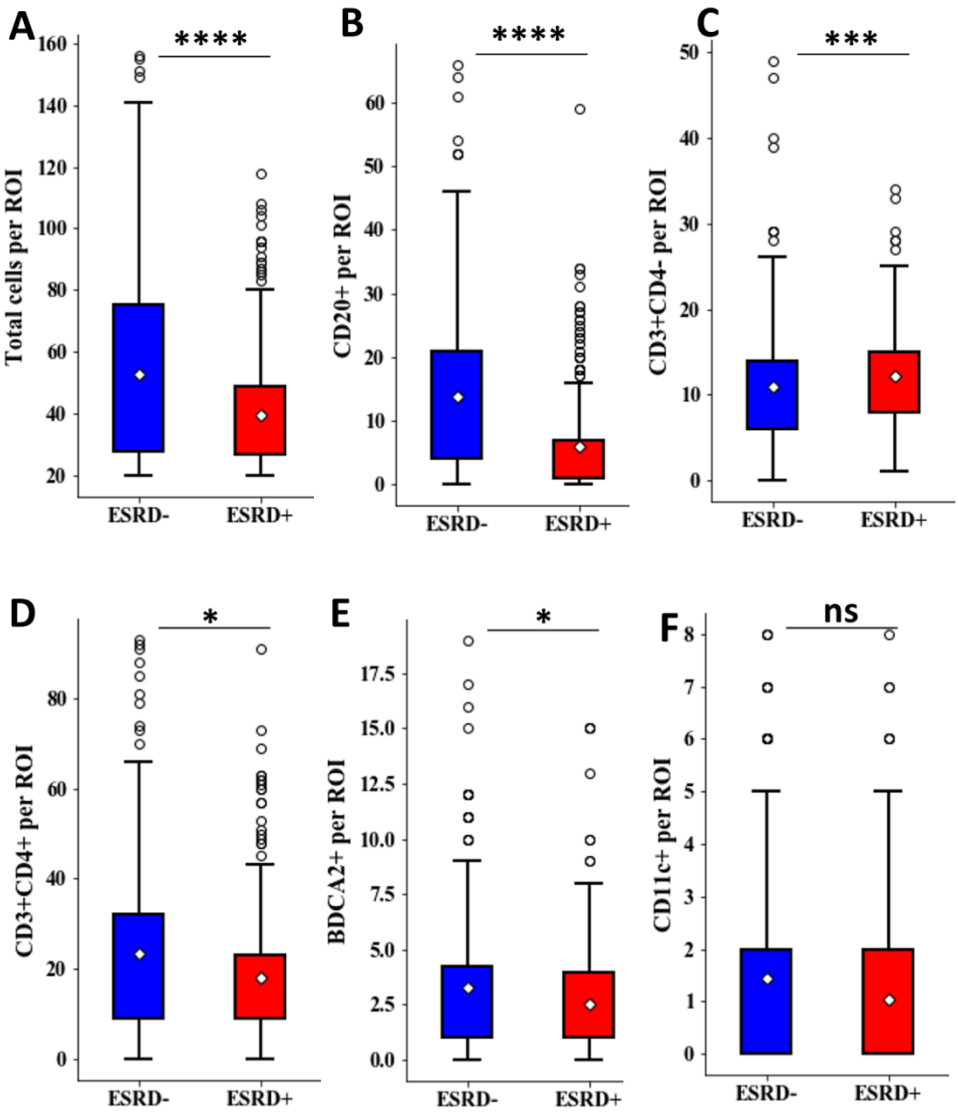

Figure 4.3. Cellular densities in dense ROIs from ESRD+ and ESRD-. Difference in the number of cells of each class per ROI in dense ROIs (n≥20 cells) between ROIs from patients who progressed to ESRD (ESRD+, n=295 ROIs) and those who do not (ESRD-, n=252 ROIs); A) Total cells/ROI; B) CD20+ cells/ROI; C) CD3+CD4- cells/ROI; D) CD3+CD4+ cells/ROI; E) BDCA2+ cells/ROI; F) CD11c+ cells/ROI

Two limitations of this dataset are that there are different numbers of ESRD- and ESRD+ patients, and varying numbers of ROIs per patient. Further, the ESRD+ patients had more ROIs associated with them on average than ESRD- patients. In order to mitigate any effect from these cohort-level and individual-level class imbalances, a bootstrapping analysis was performed, as described in **Chapter 3**, and in **Fig 4.4A**. For every class we

plotted both the distribution of means for each patient cohort, as well as the distribution of the difference in the means calculated for every iteration. We observed that the distributions of the mean CD20+ per ROI and CD3+CD4- T cells per ROI were well separated, and that the confidence intervals for the difference in CD20+ per ROI and CD3+CD4- T cells per ROI both did not overlap with zero (**Fig 4.4B-E**). These striking differences between the ESRD- and ESRD+ pools reinforced our findings from the original analysis. In contrast, the bootstrapping analysis revealed that the observed differences in CD4+ T cell and CD11c+ mDC densities likely reflect underlying class imbalances, while BDCA2 remained the same between the two groups (**Fig 4.4F-K**).

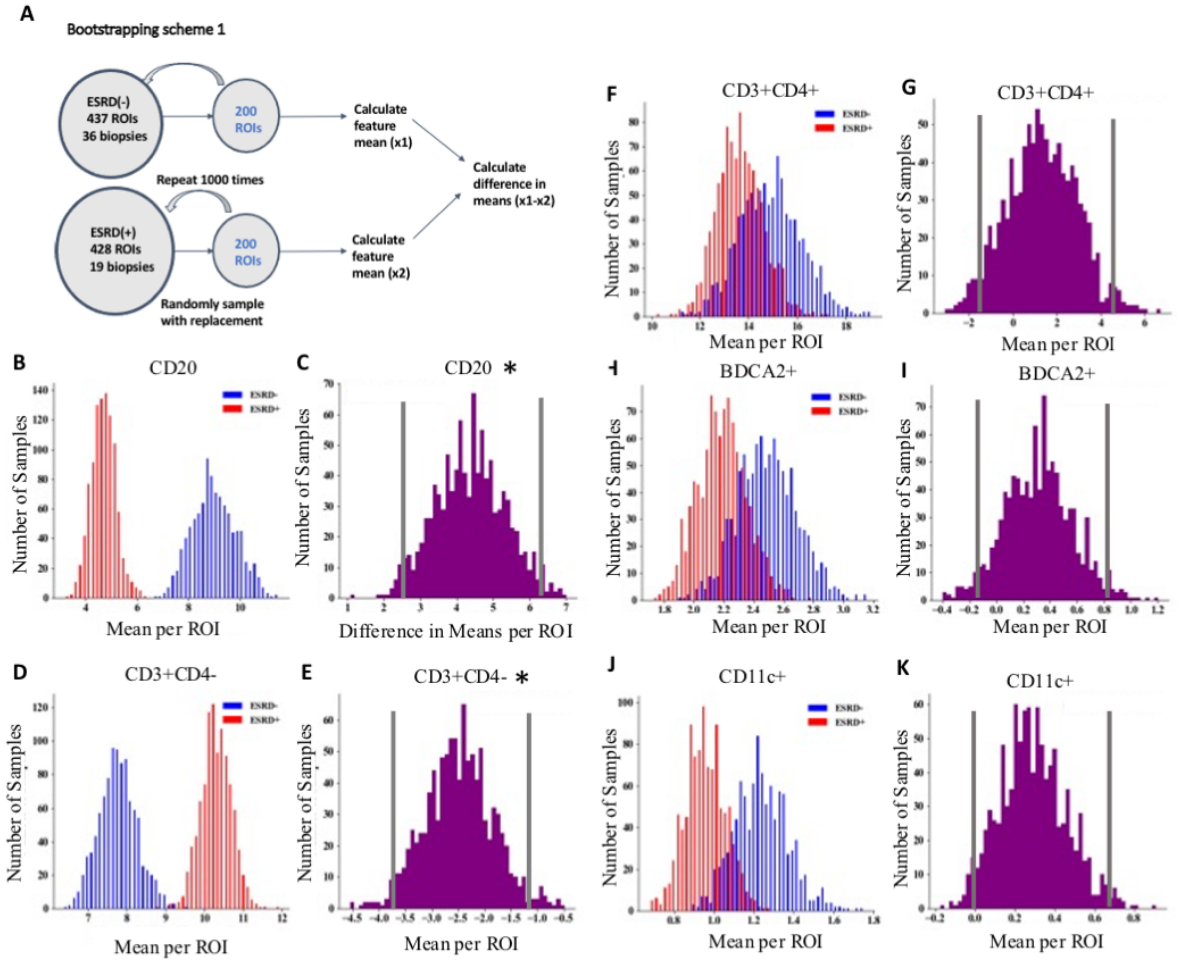

Figure 4.4. Bootstrapping analysis of ESRD+ vs ESRD- ROIs. A) schematic of bootstrapping procedure; B, D, F, H, J—distribution of the mean cells/ROI for 1000 bootstrapped samples; C, E, G, I, K—distribution of the difference in means for every pair of ESRD+ and ESRD- samples (ESRD- — ESRD+), lines indicate the 95% confidence interval, star indicates that the confidence interval does not overlap with zero

Overall, the findings from this analysis suggest that there are differences in how immune infiltrates are organized in lupus nephritis patients who progress to kidney failure and those who do not. There seem to be a subset of patients who have large, dense aggregates that contain many B cells, relatively few CD4- T cells, and remarkably stable kidney function over time. Conversely, there seems to be a subset of patients who have

fewer B cells in a given region and an enrichment for CD4- T cells, and this phenomenon seems to be enriched in patients who have higher tubulointerstitial chronicity and eventually progress to renal failure (**Fig 4.5**).

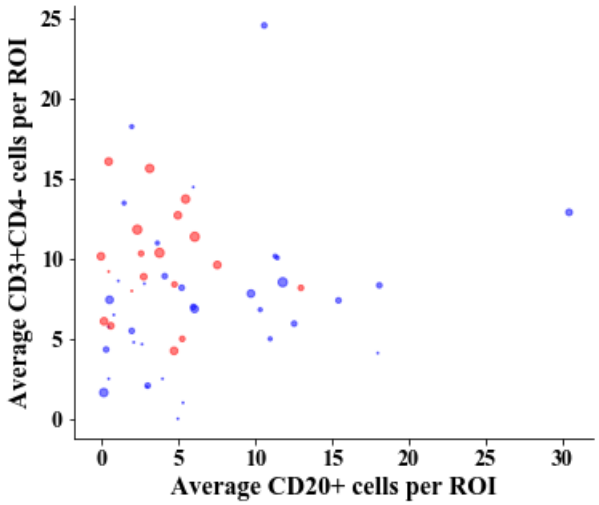

Figure 4.5. Patient level average cell density. Relationship between patient-level average (total number of cells per patient/# ROIs per patient) CD20+ cells per ROI and average CD3+CD4- cells per ROI; each point weighted by tubulointerstitial chronicity; red=ESRD+, blue=ESRD+

4. Comparison of cellular density in ESRD- , ESRD+, and ESRD-current patients

There was a distinct cohort of 5 patients in our dataset who were currently or imminently (within 2 weeks) in renal failure at the time of biopsy. To evaluate whether these patients were distinct from the patients who proceeded to renal failure at a later date, we broke our patient cohort into 3 groups—ESRD-, ESRD+ (defined as people who progressed to renal failure in greater than two weeks), and ESRD-current (defined as the people who were in renal failure within two weeks of the biopsy). When we examined the differences between the three groups of patients, we found that there was no significant difference in the total number of cells in any of the cohorts (**Fig 4.6A**).

ROIs from patients in imminent renal failure (ESRD-current) had many fewer B cells per ROI relative to both the ESRD+, and ESRD- (**Fig 4.6B**). They also had more CD3+CD4- T cells per ROI than the other two groups (**Fig 4.6C**). ESRD-current samples therefore exhibited the most extreme versions of the B cell/CD4- T cell relationships that were found in the analysis of two patient groups. The CD3+CD4+ cells displayed an unusual pattern in which ESRD-current samples showed an intermediate number of CD4+ T cells relative to ESRD- and ESRD+ (**Fig 4.6D**). This is difficult to interpret given the functional heterogeneity of the CD3+CD4+ T cell compartment.[14] Interestingly, there was also an observed decline in mDCs in the ESRD current ROIs relative to the other two groups (**Fig 4.6F**).

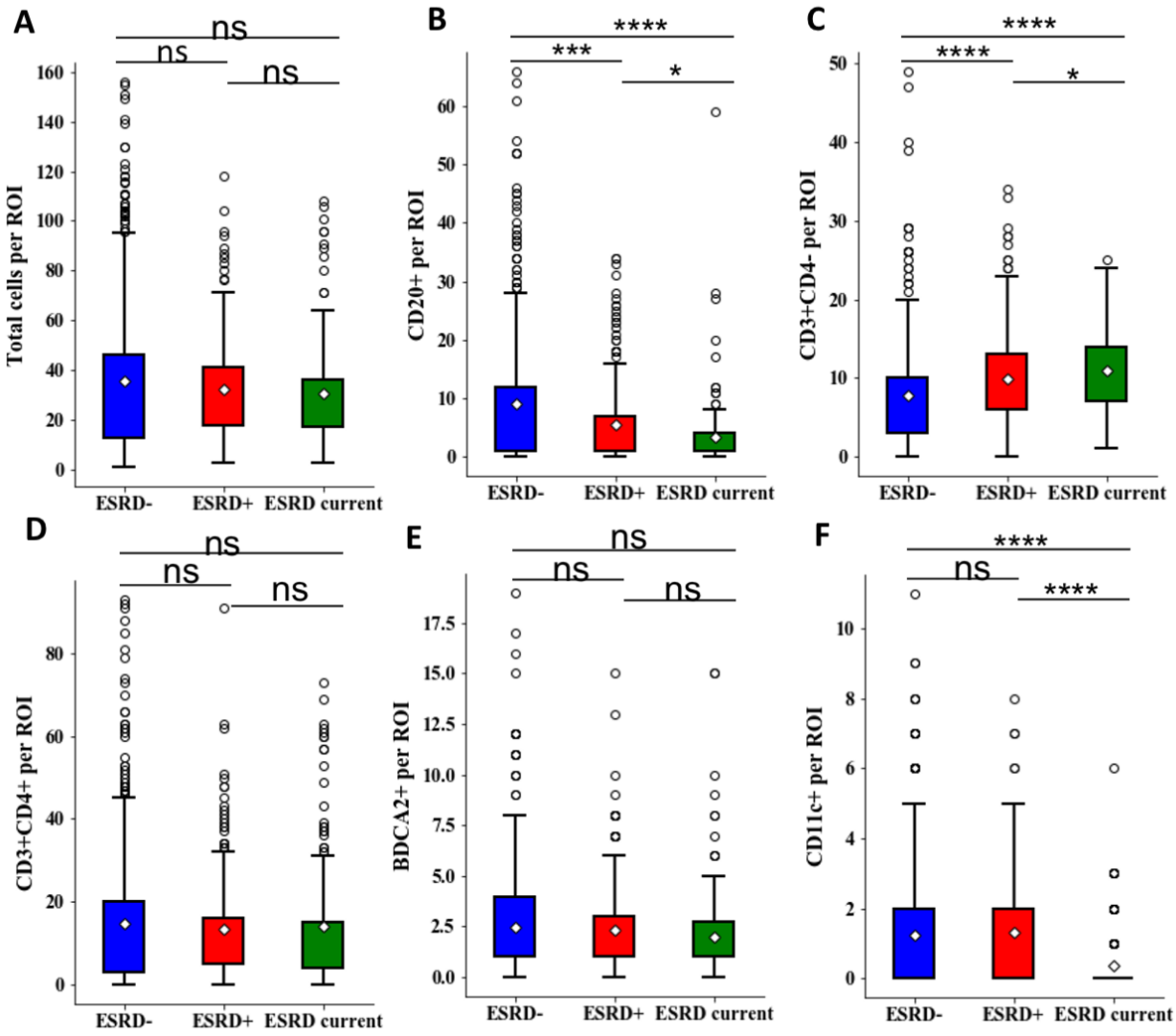

Figure 4.6. Cellular densities in ESRD+, ESRD current, and ESRD-; Difference in the number of cells of each class per ROI between patients who did not progress to ESRD (ESRD-), people who progressed to ESRD in more than 2 weeks (ESRD+) and people who progressed to ESRD within 2 weeks of the biopsy date. A) Total cells/ROI; B) CD20+ cells/ROI; C) CD3+CD4- cells/ROI; D) CD3+CD4+ cells/ROI; E) BDCA2+ cells/ROI; F) CD11c+ cells/ROI

Because we only have 5 of the ESRD "current" patients, we also wanted to ascertain that the differences observed in that cohort were robust. Therefore, we repeated the bootstrapping analysis with all three classes, using a sample size of 150 ROIs. (**Fig 4.7A**). A lower sample size was used for this analysis due to the small number of images from

the ESRD-current patient group (n=162 ROIs). For the number of CD20+ cells per ROI, the pairwise differences between ESRD-, ESRD+ and ESRD-current were all significantly different from zero, suggesting that these were three distinct groups (**Fig 4.7B-C**). For CD3+CD4- T cells the differences between ESRD- vs. ESRD+ and ESRD+ vs. ESRD current were not significantly different from zero, while the difference between ESRD- and ESRD current was significant (**Fig 4.7D-E**). This suggests that the ESRD+ cohort could be in some sort of intermediate phase between the extreme groups with respect to CD3+CD4- T cell density. There were no differences between the groups with respect to CD3+CD4+ T cell and pDC density. It was again observed that the ESRD-current patients were well separated from the other two cohorts with respect mDC abundance, validating the observation that these patients have a particular lack of this cell population (**Fig 4.7J-K**). This is a striking difference that has implications for the dynamics of APC populations during renal failure.

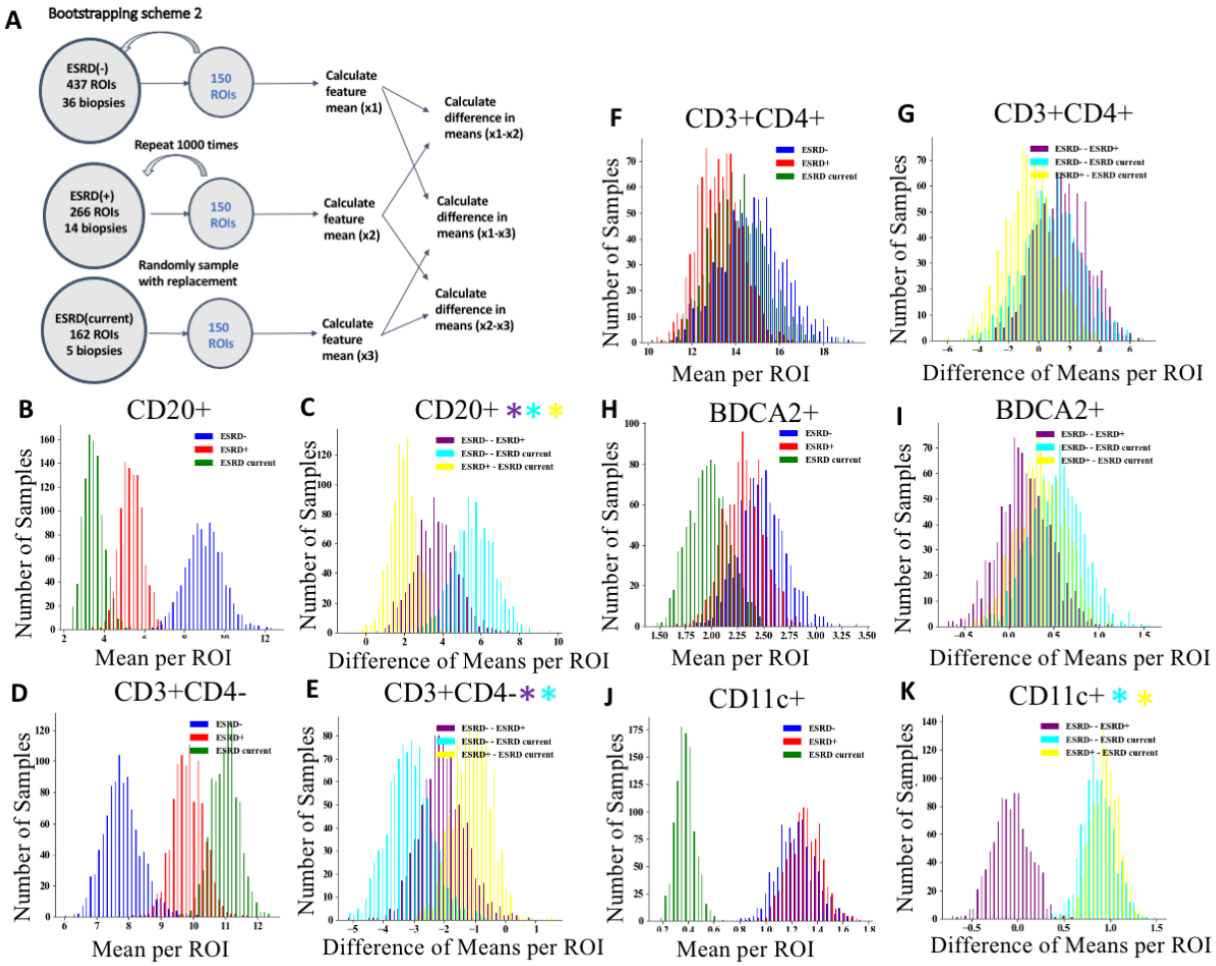

Figure 4.7. Bootstrapping analysis of ESRD+ vs ESRD- vs ESRD current ROIs. A) schematic of bootstrapping procedure; B, D, F, H, J—distribution of the mean cells/ROI for 1000 bootstrapped samples; C, E, G, I, K—distribution of the difference in means for every pair of samples ESRD- — ESRD+ (purple), ESRD+ — ESRD-current (yellow), and ESRD- — ESRD-current (cyan); stars of the corresponding colors indicate that the confidence interval did not overlap with zero

5. Relationship between cellular features and pathologist-defined metrics

End stage renal disease is an important outcome to understand in lupus nephritis, but it was also of interest how cellular features found on biopsy were related to other clinical features available from the biopsy. Clinical decision-making is often driven by these metrics, so we wanted to investigate whether the enrichment for B cells was associated

with other “positive” prognostic indicators, and if enrichment for CD4- T cells is associated with other “negative” prognostic indicators.[8,10]

One clinical metric we considered was chronicity index, a semi-quantitative descriptor assigned by the pathologist that reflects the long-term damage (eg, atrophy, fibrosis) that is observable in the renal tissue. High chronicity at the time of biopsy is associated with progression to renal failure[8] (**Fig 4.8A**), so it was of interest whether any particular cell classes or patterns of inflammation were correlated with chronicity. The ROIs were binned by whether they came from patients who had high tubulointerstitial chronicity (defined as greater than or equal to 4, n=326 ROIs) and those with low chronicity (less than 4, n=539 ROIs). The relationship between tubulointerstitial chronicity and ESRD status in our patient cohort can be found in **Table 4.2**.

This revealed that high TI chronicity ROIs had on average more total cells per ROI, more CD3+CD4+ and CD3+CD4- T cells, fewer CD20+ B cells and no difference in BDCA2+ pDCs and CD11c+ mDCs. The results from low versus high chronicity were very similar to the results from ESRD- and ESRD+, because most of the low chronicity patients were ESRD-, and most of the ESRD+ patients were high chronicity (**Fig 4.8B-G**).

Table 4.2 Relationship between ESRD status and TI chronicity

	Low chronicity <4	**High chronicity ≥ 4**
ESRD-	31 patients	5 patients
ESRD+	8 patients	11 patients
	39 patients	**16 patients**

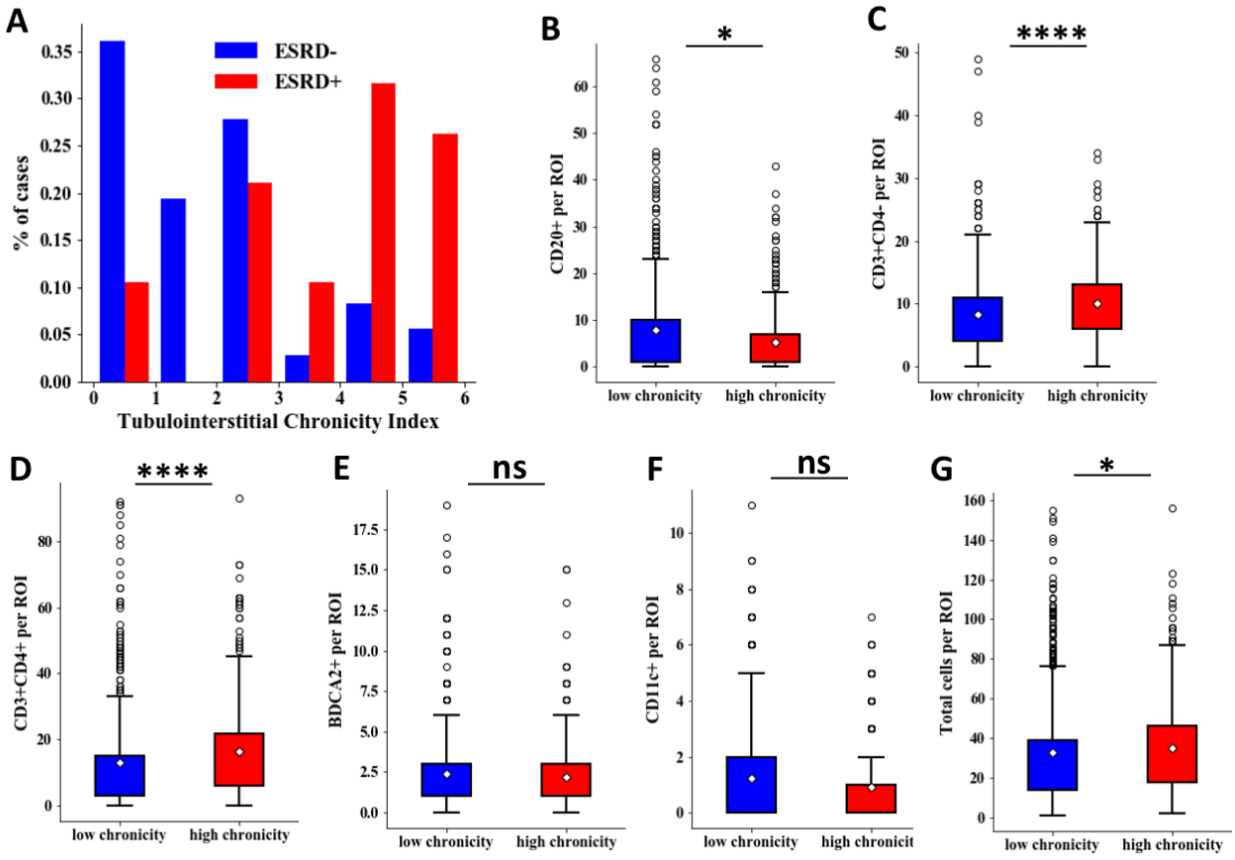

Figure 4.8. Cellular densities in high and low TI chronicity patients; Difference in the number of cells of each class per ROI between patients who had low TI chronicity indices (<4, n=39 patients, n=539 ROIs), and patients with high TI chronicity indices (≥4, n=16 patients, n=326 ROIs). A) distribution of TI chronicity index in patients broken down by ESRD status; B) CD20+ cells/ROI; C) CD3+CD4- cells/ROI; D) CD3+CD4+ cells/ROI; E) BDCA2+ cells/ROI; F) CD11c+ cells/ROI; G) Total cells/ROI

In addition to chronicity is tubulointerstitial inflammation, or TI score, a metric of inflammatory infiltrate.[9] One would expect that total cell counts will track with this metric. However, while very few patients in our cohort with mild TI (0-1) progressed to ESRD, the patients with moderate to severe TI (2-3) did not necessarily progress to kidney failure at

higher rates (**Fig 4.9A**). The relationship between ESRD status and TI inflammation in this cohort is in **Table 4.3**.

When the ROIs were binned by whether they came from a biopsy with high TI score (2-3, n=654) and low (0-1, n=211 ROIs) TI score, it was found that high TI score samples had more CD3+CD4+ T cells per ROI, and fewer CD20+ B cells, and BDCA2+ pDCs per ROI. There was no difference in CD3+CD4- T cells, CD11c+ mDCs, and total cells per ROI (**Fig 4.9B-G**).

Table 4.3 Relationship between ESRD status and TI inflammation

	Low TI 0-1	**High TI 2-3**
ESRD-	15 patients	21 patients
ESRD+	2 patients	17 patients
	17 patients	**38 patients**

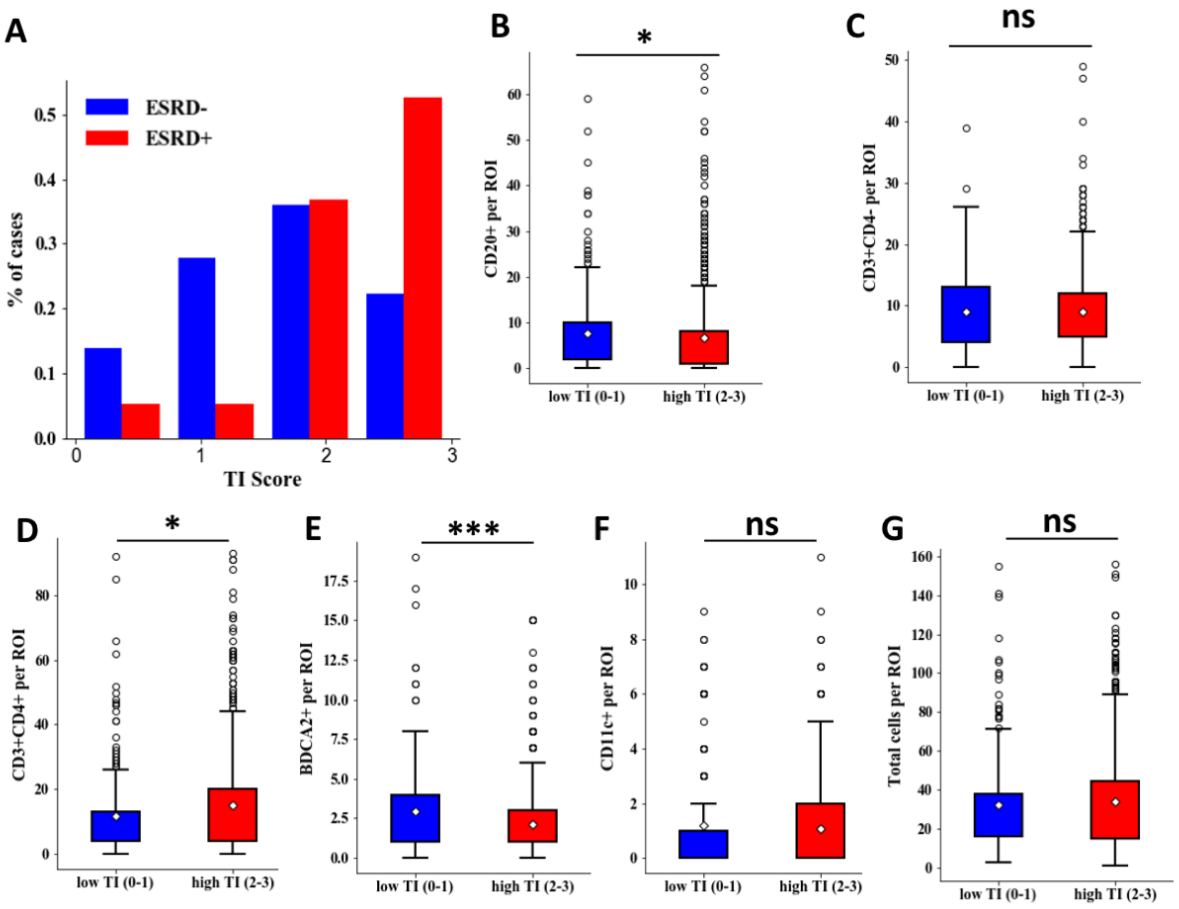

Figure 4.9. Cellular densities in high and low TI inflammation patients. Difference in the number of cells of each class per ROI between patients who had low TI inflammation (<3, n= 17 patients, n= 211 ROIs), and patients with high TI inflammation (≥3, n= 38 patients, n= 654 ROIs). A) distribution of TI score in patients broken down by ESRD status; B) CD20+ cells/ROI; C) CD3+CD4- cells/ROI; D) CD3+CD4+ cells/ROI; E) BDCA2+ cells/ROI; F) CD11c+ cells/ROI; G) Total cells/ROI

The common findings across these different methods of stratifying the patients was that "positive" findings (i.e., lack of progression to ESRD, low TI chronicity, TI inflammation) were associated with an enrichment for CD20+ B cells.

However, the fact that ESRD status, chronicity index, and TI score don't always co-vary is a point of concern for clinical decision-making. Each one of these metrics (TI chronicity, TI inflammation, and cellular features) examines a different aspect of the

tissue. What they have in common is an element of human judgement. In each case, whether it is the pathologist deciding what percentage of the tissue is fibrotic, or a microscopist choosing which areas of the tissue are sufficiently inflamed to image, a person is making a judgement about what aspects of the biopsy to pay attention to. There is variation in this judgement between individual observers, and even within the same observer at different time.[127,128]

6. Relationship between cellular features and renal function at the time of biopsy

In addition to histology metrics, we wanted to know if any of the cellular features tracked with renal function. If so, it would suggest that renal inflammation is perhaps dynamic enough to reflect shorter-term trends in renal function. We chose to look at serum creatinine, as we have both the value that was obtained near the time of biopsy, and an additional measurement taken at least 2.5 years after the original measurement. Therefore, we can calculate the fold change in serum creatinine between these two time points. However, because some of the patients who progressed to kidney failure started with very high serum creatinine and had lower creatinine as a result of treatment, this fold change should be interpreted with care. **Table 4.4** shows the relationship between ESRD status and whether patients doubled their creatinine.

However, a sub-analysis was performed that stratified the ESRD- patients by those who doubled their serum creatinine (n=9 patients, n=105 ROIs) during the time of follow-up and those that did not (n=27 patients, n=332 ROIs) (**Fig 4.10A**). There were

significantly fewer total cells per ROI in the patients who doubled their serum creatinine, but no significant differences in the specific cell classes (**Fig 4.10B-G**).

Table 4.4 Breakdown of patients with respect to ESRD status and whether or not they doubled their serum Cr during the follow up time

	Non-Doubled Serum Cr	**Doubled Serum Cr**
ESRD-	27 patients	9 patients
ESRD+	9 patients	10 patients
	36 patients	**19 patients**

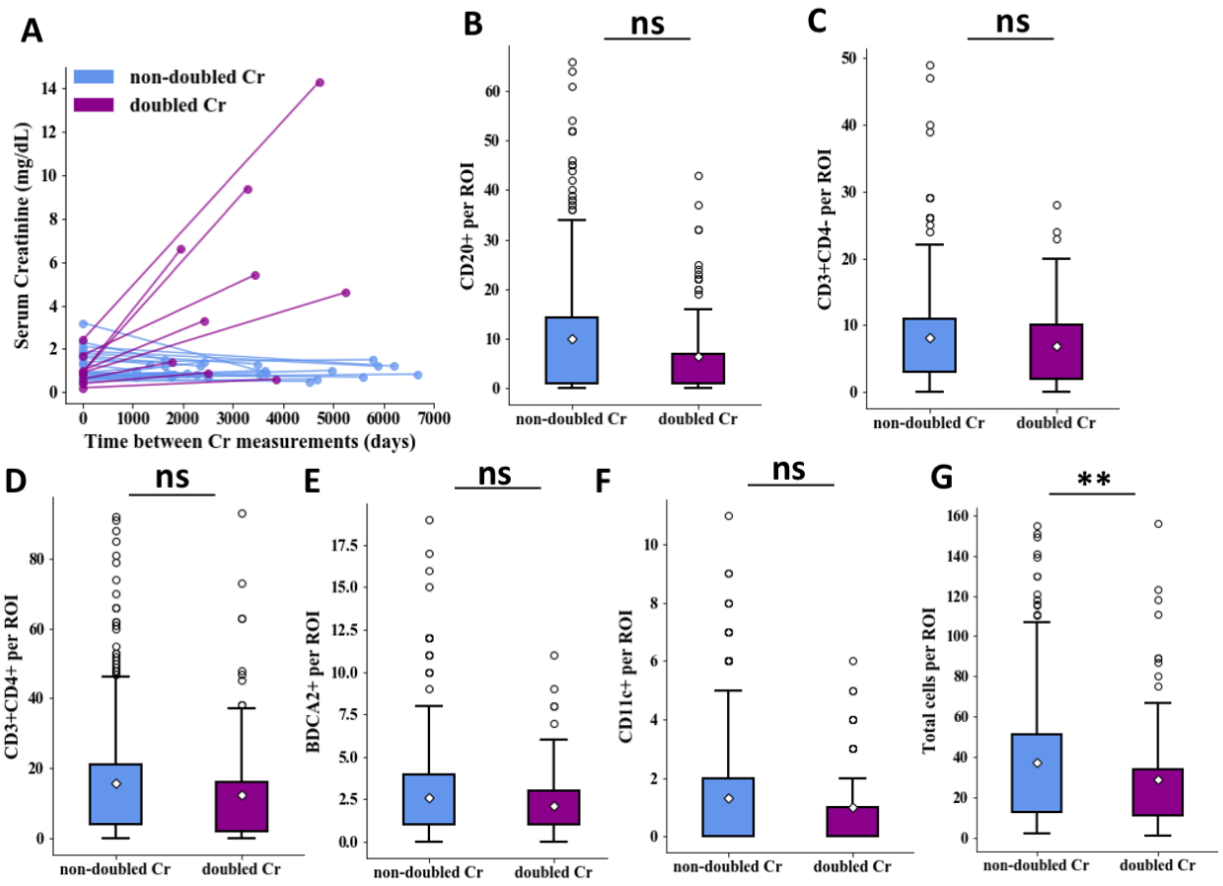

Figure 4.10. Cellular densities in patients with doubled serum Cr and non-doubled serum Cr. Difference in the number of cells of each class per ROI between ESRD- patients who doubled their serum creatinine during the follow-up period (n= 9 patients, n= 105 ROIs), and patients who did not (n= 27 patients, n= 332 ROIs). A) trajectory of serum creatinine in the ESRD- patients; B) CD20+ cells/ROI; C) CD3+CD4- cells/ROI; D) CD3+CD4+ cells/ROI; E) BDCA2+ cells/ROI; F) CD11c+ cells/ROI; G) Total cells/ROI

7. Cellular nearest neighbors and niches

We next probed the close intercellular distances in the tissue. For every cell in the dataset, the class of its "nearest neighbor" was found by identifying the cell with the smallest centroid-centroid distance to the cell in question (**Fig 4.11A**). For both CD20+ B cells and CD3+CD4- T cells there was a strong propensity for co-localization with cells of the same type (i.e. a given B cell's nearest neighbor was another B cell). However, in the context of ESRD- ROIs, there was an increased proportion of all cells that had a B cell as

its nearest neighbor relative to ESRD+ ROIs. For example, 44% of B cells from ESRD- patients compared to only 36% of B cells from ESRD+ patients had another B cell as its nearest neighbor (**Fig 4.11B-C**). In contrast, 39% of CD3+CD4- T cells in ESRD- ROIs had another CD4-T cell as its nearest neighbor, as opposed to 46% of CD3+CD4- T cells in ESRD+ ROIs. In addition, all the other cell types in ESRD+ ROIs displayed an enrichment for CD4- T cells as their nearest neighbors. These observations indicate that in ESRD+ ROIs there is an enrichment of CD4- T cells in cellular aggregates, and in ESRD- ROIs there is an enrichment of B cells in aggregates. The nearest neighbor relationships is a function of both the increased prevalence of these cells in the ROIs and their propensity for clustering with other cells of interest.

We extended this approach by examining the entire "niche" of each cell, defined as all of the cells with centroids within ~10 μm (100 pixels) of the centroid of the cell in question **(Fig 4.11D)**. Differences between the distributions of niche content between cells from ESRD+ and ESRD- biopsies were evaluated using the Kolmogorov-Smirnov test. This analysis allows us to capture both how densely packed cells are and what their most likely interacting partners are. We found that in the context of ESRD-, any given cell is more likely to have more cells within a 10 μm radius ($p=2.1x10^{-189}$) **(Fig 4.11E)**. This calculation validated the observation that cellular niches are denser in ESRD- samples relative to ESRD+ samples. Interestingly, there was an increased frequency of CD20+ B cells in cellular niches from ESRD- samples ($p=8.7x10^{-84}$) and a corresponding increased frequency of CD3+CD4- T cells in ESRD+ niches ($p=6.9x10^{-286}$), in concordance with what was observed in the nearest neighbor analysis (**Fig 4.11F-G**).

For both of these metrics it should be noted that we are only probing cell classes that we imaged. That is to say, if the actual “closest” cell to one of the identified cells is not of one of the 5 classes that we are interested in, the nearest neighbor analysis will not recognize that. This phenomenon will show up in the niche analysis by way of fewer cells being counted in a given cell’s niche. The interpretation of these results is therefore focused on how T cells, B cells and antigen presenting cells are associating with each other, regardless of the local density of other cellular classes.

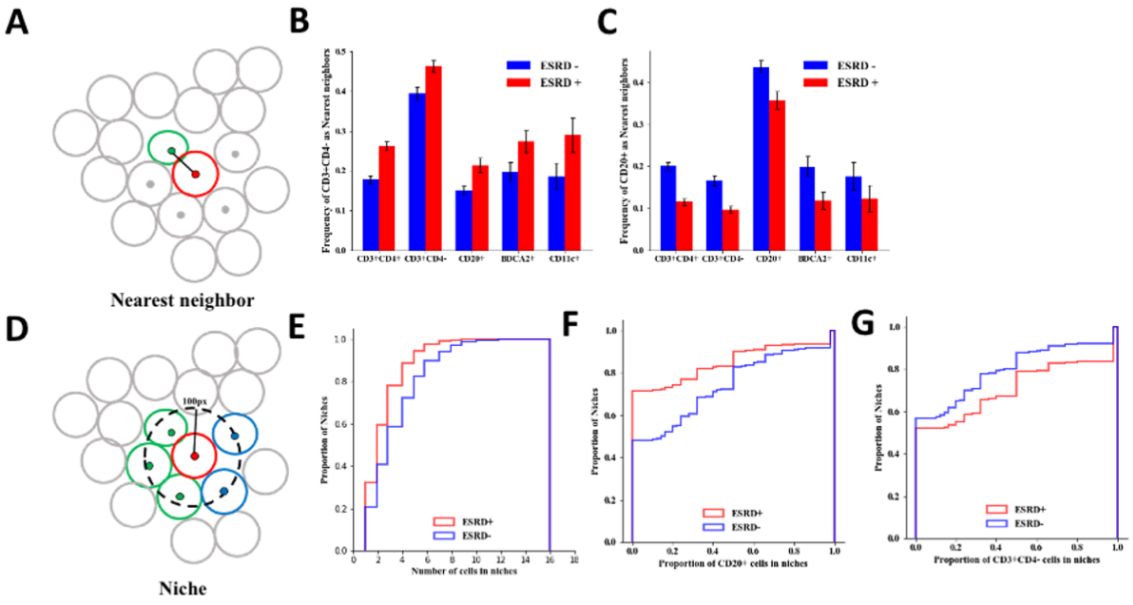

Figure 4.11. Nearest neighbor and niche analyses. A) Nearest neighbor schematic; B) breakdown of what percentage of each cell class has CD20+ B cells and C) CD3+CD4- T cells as their nearest neighbor; D) Niche analysis schematic, radius=100 pixels; E) cumulative distribution functions showing the total niche sizes, F) percentage of CD20+ B cells in niches, and G) CD3+CD4- T cells in niches

8. Cellular neighborhoods

Higher order local cellular organization was then probed by grouping cells into spatially discrete neighborhoods. DBSCAN was implemented to perform this task with a maximum intercellular centroid-centroid distance of 100 pixels (~10.6 μm) and a minimum

neighborhood size of 2, as this best resolved observable regional behavior across the biopsy dataset (**Fig 4.12A**).

We detected 5425 cell neighborhoods in the whole dataset which could be described by a set of 24 quantitative features including cell type frequency, cell type proportion, ratios of cell types, average cell shape features, and the overall area of the neighborhood (**Table 4.5**).

Table 4.5 Features used to define types of neighborhoods

Cell Frequency	CD20+ count CD11c+ count BDCA2+ count	CD3+CD4- count CD3+CD4+ count
Cell Proportion	CD20+ proportion CD11c+ proportion BDCA2+ proportion	CD3+CD4- proportion CD3+CD4+ proportion
Cell Frequency ratios	CD20+/CD11c+ CD20/BDCA2+ CD20+/CD3+CD4- CD20+/CD3+CD4+ CD11c+/BDCA2+	CD11c+/CD3+CD4- CD11c+/CD3+CD4+ BDCA2+/CD3+CD4- BDCA2+/CD3+CD4+ CD3+CD4-/CD3+CD4+
Cell shape	Mean circularity Mean eccentricity Mean major-minor axis ratio	
Aggregate shape	Aggregate area	

K-means clustering was applied to these features to define classes of neighborhoods. A leave-one-out t-test was used to determine which features or combination of features best distinguished the six neighborhood groups (**Fig 4.12B**). The most distinctive feature(s) for each group was used to describe the cell neighborhoods as follows: 1) B

cell enriched, 2) CD4- T cell enriched, 3) Large, lymphocyte-rich, 4) CD4+ T cell enriched, 5) mDC enriched, and 6) pDC enriched (**Fig 4.12C**).

Tertiary lymphoid structures (TLSs) have been previously identified in the context of lupus nephritis.[10] Although we cannot explicitly define TLSs in this dataset, we hypothesized that the large, lymphocyte-rich neighborhoods might approximate TLSs. For example, we noted that within the large, lymphocyte-rich neighborhoods on average 28.6% of the cells were B cells and 48.3% were CD4+ T cells. Furthermore, 96.1% of these neighborhoods met the following criteria: 1) contained at least 20 cells, 2) both B cells and CD4+ T cells were represented in the neighborhood and 3) at least 50% of all cells were B cells and/or CD4+ T cells. Therefore, the vast majority of large, lymphocyte-rich neighborhoods, have features consistent with TLSs.

We then examined how these six neighborhood types were distributed between the ESRD- and ESRD+ patients. After normalizing for the number of ROIs captured for each patient, ESRD+ patients had significantly higher numbers of CD4- T cell enriched neighborhoods relative to the ESRD- patients (p value=0.006) (**Fig 4.12E**). None of the other clusters were different in frequency between ESRD+ and ESRD- patients (**Fig 4.12D,F-I**).

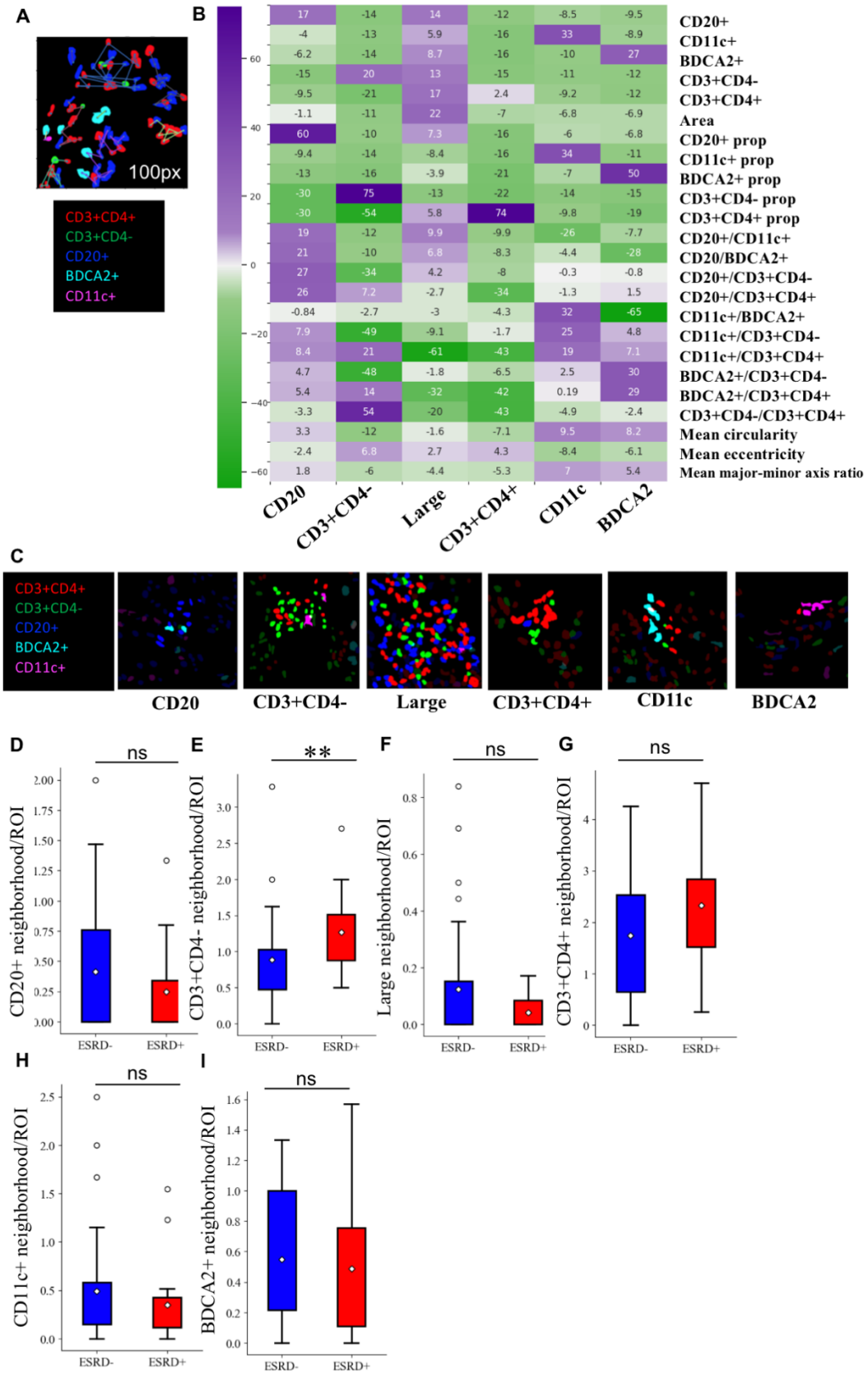

Figure 4.12 Figure caption on next page

Figure 4.12, continued. Spatial analysis of cellular distribution reveals increased CD3+CD4- T cells in cellular neighborhoods. Nearest-neighbors analysis, proportions of cells that have A) representative images of the DBSCAN algorithm performance; B) heatmap showing test statistics associated with each feature for each identified neighborhood type; C) representative examples of each type of aggregate; D-I) comparison of the abundance of each type of neighborhood in each patient cohort, normalized by the number of ROIs per patient; D) CD20+ ; E) CD3+CD4-; F) Large; G) CD3+CD4+; H) CD11c+; I) BDCA2+

9. Discussion

Here we have demonstrated that lupus nephritis patients who proceed to kidney failure have a distinctive inflammatory signature that is characterized by an increased density of CD4+CD4- T cells, and a relatively decreased density of B cells. In particular, we observe that CD3+CD4- T cells are enriched in dense groups of cells that are putatively interacting with each other. In contrast, there are a subset of lupus nephritis patients who have preserved long-term renal function and appear to have dense regions of inflammation that are enriched for B cells, a surprising finding in the context of an antibody-mediated disease. In light of these findings, the clear next step was to obtain more phenotypic details about these cells. In particular, it was critical to more specifically identify the CD3+CD4- T cells, as they appear to be associated with worse patient outcomes. Further work also needed to be done to characterize the distribution of other B cell subsets, in particular plasma cells, in order to resolve the apparent paradox of dense regions of B cells being present in ESRD- patients. These questions will be addressed in the following chapters. Overall, these analyses highlight the utility of examining the spatial distribution of cells in tissue in order to tease apart differences associated with divergent patient outcomes.

Chapter 5: Segmentation and analysis of highly multiplexed datasets

The work described in Chapters 5-6 was done in collaboration with several people -- Dr. Junting Ai and Margaret Veselits acquired the data and assisted with pre-processing, Dr. Madeleine Durkee contributed equally to the development of the automated cellular segmentation method. Dr. Junting Ai and Dr. Yuta Asano performed the follow-up studies on CD4- T cells.

1. Introduction

Though immunofluorescence microscopy is very useful for capturing spatial relationships, it is typically limited in the phenotypic detail that it can provide, due to technical limitations on the number of markers that can be imaged at one time. Developing a robust pipeline for acquiring and analyzing highly multiplexed microscopy (HMP) data was of great interest to our laboratory, as it would allow us to ask much more complex questions about how *in situ* inflammation is organized. For this purpose we collected highly multiplexed datasets in both lupus nephritis and triple negative breast cancer (TNBC). We then used the highly multiplexed LuN data to follow up on the findings from the high resolution dataset.

This part of the project required substantial technical innovation in order to segment cells accurately and perform spatial analysis on these segmentations. Here I will describe these innovations in every step of the process, including data acquisition, pre-processing, segmentation, and spatial analysis (**Fig 5.1**). I will also describe some work that was done to follow up on previous findings about CD4- T cells in the HR dataset. All of this work was done in collaboration with the Giger laboratory, and all data storage, preprocessing,

cellular segmentation, and spatial analysis was performed on MEL, the computation server administered by the Giger laboratory.

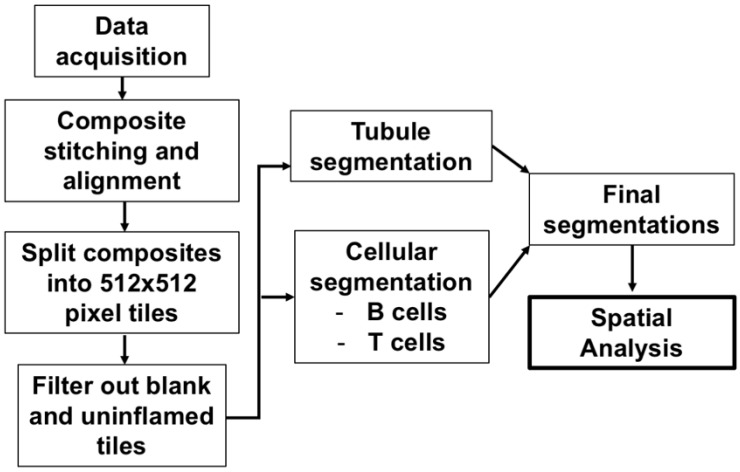

Figure 5.1 Overview of the analytical pipeline for the highly multiplexed

2. Acquisition of Highly multiplexed (HMP) Immunofluorescence Data

Standard immunofluorescence techniques typically place a limit on the number of markers that can be interrogated in a single panel. However, in order to probe the immunological interactions that are taking place in tissue, detailed phenotypic information is desirable. When designing a staining panel, there is often a tradeoff between the number of distinct cellular classes one can probe, and the phenotypic complexity one can achieve within a particular cell class. In order to overcome this tradeoff, our lab utilized a "strip and reprobe" approach to collect data on 20 distinct cellular markers from two different disease states: lupus nephritis, and triple negative breast cancer TNBC.

In both cases, the goal was to collect detailed phenotypic information on T cells, B cells, antigen presenting cells, and the overall tissue context, across the whole biopsy. In order to capture the whole tissue, we used the Caliber-ID RS-G4 large-format confocal microscope, which can efficiently image large tissue sections. The sections were imaged

at a pixel size of 0.221 um and magnification of 63x. In the strip and reprobe procedure, a 2-5 um thick section of a FFPE biopsy was iteratively stained according to a procedure outlined by Bolognesi et al.[129] The sections were deparaffinized and stained with a combination of primary antibodies and secondary antibodies conjugated with AlexaFluor 488, 546, and 647 fluorophores, in addition to DAPI. They were then imaged at the 405, 488, 561, and 640 nm wavelengths. Subsequently the coverslips were removed, and the antibodies were stripped using a combination of sodium dodecyl sulfate (SDS), Tris-HCl, and 2-mercaptoethanol. The samples were re-probed with a new set of primary and secondary antibodies, and then imaged again. Repeating this procedure iteratively, with the DAPI signal serving as a consistent signal between each day, allowed us to capture the large number of markers over the course of several days. The advantage of the large-format microscope was that we could image the whole tissue sample, mitigating the sampling problems inherent in the HR dataset.

Markers were selected based on the specific questions that were of interest in each dataset. They then had to be arranged into sub-panels in such a way that there was only one type of antibody from each species (eg, mouse, rabbit, rat) per sub-panel, and so that weaker antibodies could be used earlier in the procedure, before the aggressive stripping procedure reduced antigen availability.

Though much effort was made to image the same tissue area at the same focal plane every day, it is inevitable that there were some deviations between days that needed to be computationally corrected. This was further complicated by tissue disruption/destruction that occasionally resulted from removing the coverslip multiple times, such that the tissue on the first day was not exactly the same as the tissue on the

final day. The computational approaches used to address these issues will be discussed later in this chapter.

Lupus Nephritis (LuN)

The goal of the highly multiplexed lupus nephritis dataset (n=18 biopsies) was to validate and expand on the findings of the high resolution dataset. Imaging the whole biopsy allowed for the capture of the full range of inflammation, including sparse regions. Some of the biopsies were large enough that they were broken up into multiple imaging regions, for a total of 28 large composite regions. The markers were chosen to expand on the phenotypes captured by CD20, BDCA2, CD11c, CD3, and CD4, as well as provide some contextual information on the tissue (**Table 5.1**).

Table 5.1 Highly multiplexed LuN panel

Cellular Class	Phenotypic markers
T cell	CD3, CD4, CD8, ICOS, PD1, Tbet, FoxP3, Granzyme B, SLAMF7
B cell	CD20, IgG, IgM, CD27, CD138
Dendritic cell	BDCA1, BDCA2, CD11c, CD123, CXCR6, FCRL5, RANTES, MHC-II, Costimulatory molecules (CD40, OX40L, CD80, CD86)
Contextual	MX1

The T cell markers were chosen to capture a range of T cells that are thought to be involved in mediating inflammatory processes in lupus.[14] This includes Tfh cells (CD4+ICOS+/-PD1+),[37–39] exhausted T cells (PD1+),[130] Th1 cells (CD4+Tbet+),[43] regulatory T cells (CD4+FoxP3+),[27,43] cytotoxic CD8 T cells (CD8+GranzymeB+),[50] and cytotoxic CD4 T cells (CD4+SLAMF7+).[131] The B cell markers were chosen to identify naïve B cells (IgM+)[132], class-switched B cells (IgG+),[132] plasma cells (CD138+),[133] and memory B cells (CD27+).[25] The APC markers were chosen to cover myeloid dendritic

cells (BDCA1+, CD11c+),[70,71] plasmacytoid dendritic cells (BDCA2+, CD123+),[71] and "general" APC markers that had been identified in the literature (CXCR6, FCRL5, RANTES, MHC-II, Costimulatory molecules).[14] Finally, MX1 was chosen to show the distribution of interferon response in the tissue.[35]

Though these samples did not have the same level of follow up as the high resolution data, we were able to collect clinical data (renal function, pathology metrics, etc.) about the patients at the time of biopsy, giving us an analogous set of clinical features. (**Table 5.2**).

Table 5.2 Summary of clinical features in highly multiplexed LuN dataset

Feature	Categories	(n=18)
Sex	Female	15 (83%)
	Male	2 (11%)
	Unknown	1 (6%)
Race	African American	11 (61%)
	non African American	3 (17%)
	Unknown	4 (22%)
Induction	Cyclophosphamide	1 (5%)
	MMF	7 (39%)
	Unknown	10 (56%)
pre-biopsy plaquenil	Yes	11 (61%)
	No	3 (17%)
	Unknown	4 (22%)
pre-biopsy MMF	Yes	6 (33%)
	No	8 (45%)
	Unknown	4 (22%)
pre-biopsy prednisone >=20 mg	Yes	4 (22%)
	No	10 (56%)
	Unknown	4 (22%)
pre-biopsy azathioprine	Yes	0 (0%)
	No	14 (88%)
	Unknown	4 (12%)
Ace inhibitor/ARB	Yes	4 (22%)

Table 5.2 Summary of clinical features in highly multiplexed LuN dataset, continued

Feature	Categories	(n=18)
	No	10 (56%)
	Unknown	4 (22%)
Moderate-Severe TI (>=2) score	Yes	17 (95%)
	No	1 (5%)
Moderate-Severe Chronicity (>=4) score	Yes	10 (56%)
	No	6 (33%)
	Unknown	2 (11%)
Hypertension	Yes	9 (50%)
	No	5 (28%)
	Unknown	4 (22%)

Triple-Negative Breast cancer (TNBC)

Another goal of the project was to extend some of the techniques our laboratories had developed for the study of autoimmunity to cancer immunology. It had been previously observed that dense aggregates of immune cells in triple negative breast cancer (lacking overexpression of HER2, progesterone receptor, and estrogen receptor) were associated with positive prognosis.[134] The goal of collecting a highly multiplexed dataset in TNBC (n=18 biopsies) was to ask similar questions about how T cells, B cells, and antigen presenting cells organize and interact in the context of tumor tissue (**Table 5.3**).

Table 5.3 Highly multiplexed TNBC panel

Cellular Class	Phenotypic Markers
T cell	CD3, CD4, CD8, FoxP3, PD1, LAG3, Granzyme-B, SLAMF7, ICOS
APC	CD20, CD138, CD11c, BDCA2, CD68, CD163, MHC-II
Context	MX1, IFN-gamma, pan-cytokeratin, Ki67

These markers were chosen to cover a wide range of T cell phenotypes, including regulatory T cells (CD4+FoxP3+),[135] Tfh cells (CD4+ICOS+/-PD1+),[136] exhausted CD8 T

cells (CD8+PD1+LAG3+),[137] cytotoxic CD4 T cells (CD4+GranzymeB+SLAMF7+), and cytotoxic CD8 T cells (CD8+GranzymeB+).[138] Antigen presenting cells such as B cells (CD20+), plasma cells (CD138+), mDCs (CD11c+), pDCs (BDCA2+), macrophages (CD68+, CD163+)[139], and all other MHCII+ cells were probed in order to determine which APCs were interacting with T cells, and how this relates to regional T cell phenotype. Finally, several "contextual" markers were chosen to examine the local tumor conditions that were associated with dense cellular aggregates, including interferon (IFN-gamma), interferon response (MX1)[140], pan-cytokeratin (tumor cells)[141] and Ki67 (proliferating cells)[142].

Though we are still in the process of analyzing this data, we have completed some interesting preliminary work to show that automatic segmentation procedures developed for lupus nephritis can be leveraged to analyze data from a completely different disease. In addition, the process of collecting this dataset deepened my knowledge of microscopy and gave me useful insights that I was able to apply to the lupus nephritis data.

1. Data preprocessing

Automatic segmentation of cells in the highly multiplexed dataset is analytically challenging. Several challenges arise from the complexity and phenotypic depth of the dataset. First, a substantial amount of pre-processing was required in order to get the data into a form that was tractable for a segmentation algorithm. Briefly, this included 1) stitching the dataset into composites, 2) aligning the resulting composites across rounds of imaging, 3) breaking these aligned composites into tiles, and finally 4) filtering those tiles to identify those that contain immune cells of interest.

Stitching and alignment

The method of data acquisition, full-section imaging on a strip-scanning microscope, made image stitching and alignment necessary to spatially register all stains in the highly multiplexed dataset. The data was acquired by collecting imaging data in strips that were 1024 pixels wide, and the length of the entire imaging region (ranging from 4000 to 70,000 pixels), with an approximate 10% overlap in the imaging region between adjacent strips. To acquire full-section images, it was necessary to stitch the strips together into a single composite for every channel. We applied a sliding-window technique and cross-correlation to identify the area of overlap between consecutive strips.[143]

As part of the stripping and reprobing procedure, sample imaging was done over several days, with the slide being removed from the stage between rounds of staining. Though the same imaging region was selected every day, there was unavoidable shift between days, sometimes wider than the width of a strip. To overcome this technical limitation, all stitched composites across all days of imaging were aligned with the Day 1 DAPI channel. Cross correlation was once again used to calculate the x, y, and rotational shifts necessary to spatially register all stains across all days of imaging.

Due to a combination of the fine and sometimes repetitive details found in these images (particularly DAPI), cross correlation does not work perfectly. We therefore implemented a “semi-automatic method” of alignment. Day 1 DAPI was stitched together using cross-correlation, and then each channel was aligned relative to the Day 1 DAPI composite (**Fig 5.2**). Because there is a difference in signal between DAPI and a different cell surface marker, this cross correlation is less high-fidelity. Therefore, we developed a GUI that allowed the user the to evaluate the quality of alignment and manually adjust the

x, y and rotational shifts until the alignment was sufficient. This semi-automatic stitching and alignment was a laborious process in which all 20 markers were aligned across 28 unique image regions from the 18 biopsies. Once alignment was complete, the composites were broken up into 512x512 pixel tiles, and empty tiles were filtered out based on the intensity of the Day 1 DAPI signal at a given location in the full-section composite. The relative position of each tile within the composite was stored such that segmented cells in a given tile could be mapped back to a position in the composite. Only tiles that contained tissue on the basis of DAPI signal were carried forward.

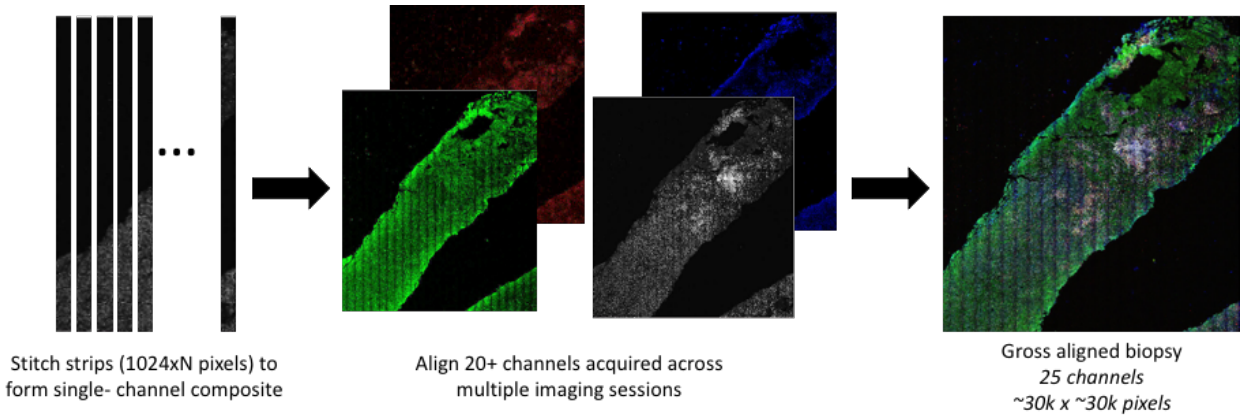

Figure 5.2. Images are stitched into composites and then aligned across days

Tile sorting by lymphocyte signal

At this point in the process, only a fraction of the image tiles contained the immune cell populations that we were interested in. Due to the high potential for false positive segmentations in empty image tiles, an 18-layer "VGG-style"[96] convolutional neural network was trained to classify CD3 and CD20 channel tiles as either "inflamed" or "uninflamed" on the basis of lymphocyte signal (**Figure 5.3A-B**).

This network was trained by splitting the 18 patients from the highly multiplexed dataset into training (n=14), validation (n=2) and test (n=2) sets. 536 CD3 and CD20

image tiles were manually classified as inflamed or uninflamed and placed in the training, validation, or test set based off of this patient split. The CNN was trained with a kernel size of (5,5) and a learning rate of 9e-5, using dropout and data augmentation of random flips and rotations. Training was stopped once the accuracy of the network on the validation set had not improved in 10 epochs. This network was able to distinguish between inflamed and uninflamed regions with high fidelity, producing a receiver operating characteristic (ROC) area under the curve (AUC) of 0.99 +/- 0.008 on the independent test set (**Figure 5.3C**). This curve was used to select the prediction threshold of 0.036, such that tiles with scores greater than this value were considered inflamed. Recall (equal to sensitivity) at this threshold was 0.991. We prioritized recall to reduce the potential that lymphocyte-containing tiles would be missed.

It was determined that the best way to reduce false positives was to segment T cell markers (CD4, CD8, PD1, ICOS) only in tiles that the network identified as inflamed based on the CD3 channel, and to segment B cells (CD20+ and CD138+) only in tiles identified as inflamed based on the CD20 channel. Though there was substantial overlap between these two sets of images, there were some T cell-only tiles, and some B cell-only tiles.

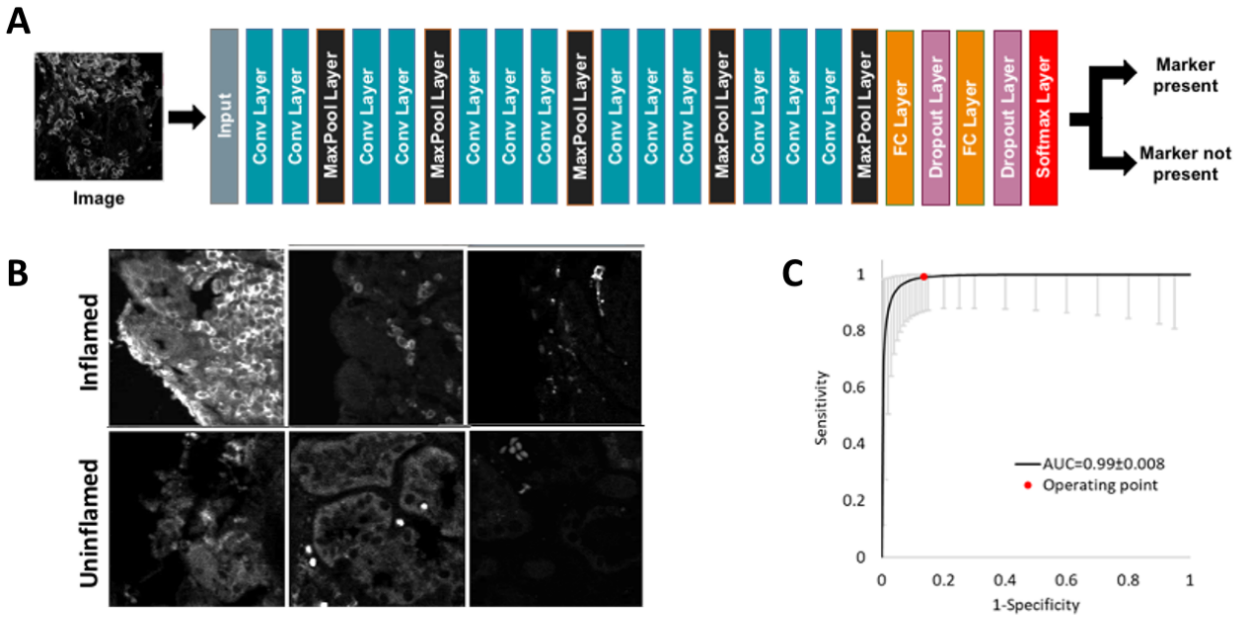

Figure 5.3. Tiles are sorted into "inflamed" and "uninflamed"; A) schematic of the CNN used to identify images that have lymphocytes ("marker present") and those that do not. B) example of training data for "inflamed" and "uninflamed"; C) ROC curve for network performance for the classification task

2. Cellular segmentation

Semi-automated ground truth

A non-trivial bottleneck for segmenting the many different populations of cells in this dataset was the generation of a sufficient amount of high-quality ground truth for each cell class. Previously, ground truth datasets had been generated by manually outlining cells using the free-hand tool in ImageJ, a task that can take on the scale of hundreds of hours for a single training set. With the number of distinct cellular classes that needed to be segmented for the highly multiplexed dataset, it was necessary to speed up this process (**Figure 5.4**). An expert was tasked with identifying the cells of interest in the image stacks using a single point in ImageJ, rather than outlining the whole cell. We then leveraged the Nuclear Segmentation Toolkit (NuSeT)[101], a pre-trained network that had

a combined U-Net/Region Proposal Network architecture, to automate the segmentation of all the nuclei in a tile. NuSeT-segmented nuclei that overlapped with the points marking a cell of interest were kept for the ground truth, while all other nuclei were discarded. The final step was manual correction of some of the nuclear segmentations by an expert, to account for situations in which NuSeT mis-segmented cells in dense regions. The assumption here was that lymphocytes typically have little cytoplasm, and therefore a nuclear segmentation would closely approximate a segmentation based on cell surface markers. Though there was still some manual work involved, this automation reduced the amount of time it took to generate a sufficient amount of ground truth for the highly multiplexed dataset.

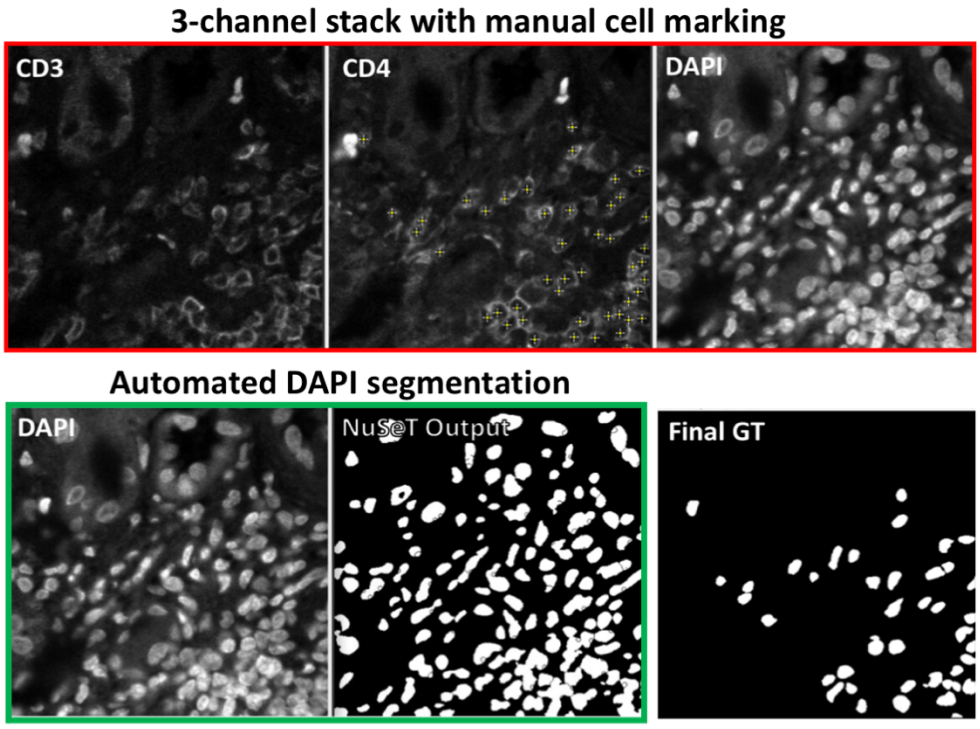

Figure 5.4. Semi-automated ground truth generation; Stacks of images are manually marked by an expert, preliminary DAPI segmentation is performed by NuSeT and the segmented nuclei that overlap with the markers are kept as the final ground truth (GT), after manual correction by the expert

Training—trained on high resolution data, finetuned on highly multiplexed data

Even with our efforts to speed up ground truth generation for this dataset, it was still difficult to produce enough to segment every single cell type with a unique network. This was particularly true for rare cell types, where merely identifying images with positive signal required a substantial amount of time spent looking through raw images for positive signal. It was therefore important to make use of all available ground truth and train networks that could generalize well.

The approach we took was to train two cellular networks—a T cell network and a B cell network. The T cell network was trained to segment 3-channel stacks of CD3/"T cell marker"/DAPI, with the "T cell marker" channel being CD4, CD8, or ICOS. In these images two classes were identified: CD3+T cell marker(+) cells (double positive T cells), and CD3+Tcell marker(-) cells (single positive T cells). The B cell network was trained to segment 2-channel images of either CD20/DAPI or CD138/DAPI, such that only CD20+ or CD138+ cells are identified.

To take advantage of all of the available ground truth, we leveraged the training data from the high resolution dataset to pre-train T and B cell networks. We then fine-tuned these networks with ground truth from the highly multiplexed dataset. During finetuning, the networks were initialized with the weights from the high resolution dataset and trained further for the new task, such that the weights for all convolutional, max-pooling, and fully-connected layers were adjusted.

For the T cell network, stacks of CD3/CD4/DAPI were constructed from the 342-image high resolution training dataset and split into 512x512 tiles (from the original 1024x1024), quadrupling the number of available tiles. An instance of Mask R-CNN[104] was trained on

the high resolution CD3/CD4/DAPI stacks, with a train/validation/test ratio of 75/15/10. This network was then fine-tuned on a set of CD3/CD4/DAPI, CD3/CD8/DAPI, and CD3/ICOS/DAPI stacks (n=211) from the highly multiplexed dataset. For the B cell network this procedure was repeated, but with CD20/DAPI stacks for both initial training and fine-tuning, with n=79 images in the finetuning highly multiplexed set. The performance of this network on a test set of CD138/DAPI stacks was sufficient to negate the need to fine-tune on a training set of CD138 stacks.

Table 5.4 T cell and B cell network performances

T cell detection network			
	Average Recall	Average Precision	Average IOU
CD3+CD4+ T cell	0.93	0.91	0.71
CD3+CD4- T cell	0.85	0.75	0.71
CD3+CD8+ T cell	0.95	0.89	0.72
CD3+CD8- T cell	0.92	0.87	0.72
B cell detection network			
CD20+ B cell	0.93	0.91	0.71
CD138+ B cell	0.85	0.75	0.71

Using the finetuning approach yielded superior network performance compared to training from scratch with limited ground truth. The finetuning procedure allowed the network to learn how to handle lower resolution, different combinations of fluorophores, and new markers. This substantially reduced the amount of highly multiplexed ground truth we needed to generate in order to produce good segmentations; the fine-tuning highly multiplexed datasets each contained less than 250 image tiles each.

Tubule network

One challenge that is particular to segmenting cells in kidney tissue is false signal from renal tubules. Due to their high collagen content, they have a high level of autofluorescence.[144,145] In addition, they have rich expression profiles, including at times CD138.[146] Therefore, the networks consistently produce false positives at the sites of tubules, regardless of efforts to include tubules in the training datasets. In order to eliminate some of the false segmentations derived from autofluorescent structures, another instance of Mask R-CNN[104] was trained to segment "tubules", where tubules are broadly defined as renal tubules and blood vessels. The segmentations were performed on DAPI tiles, and tubule structures were segmented on the basis of nuclear morphology and arrangement into cylindrical structure. The network training set was comprised of 300 manually annotated images from 18 patients, with 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.

The purpose of this network was to produce a binary mask that could be used to subtract cellular segmentations that overlap with tubule segmentations (**Fig 5.5**). Therefore, the network performance was evaluated at the pixel level (semantic segmentation) by comparing the ground truth masks with the predicted masks. This resulted in the following average network performance on the test set:

Table 5.5 Tubule network performance

	Average Recall	Average Precision	Average IOU
Tubule	0.74	0.79	0.62

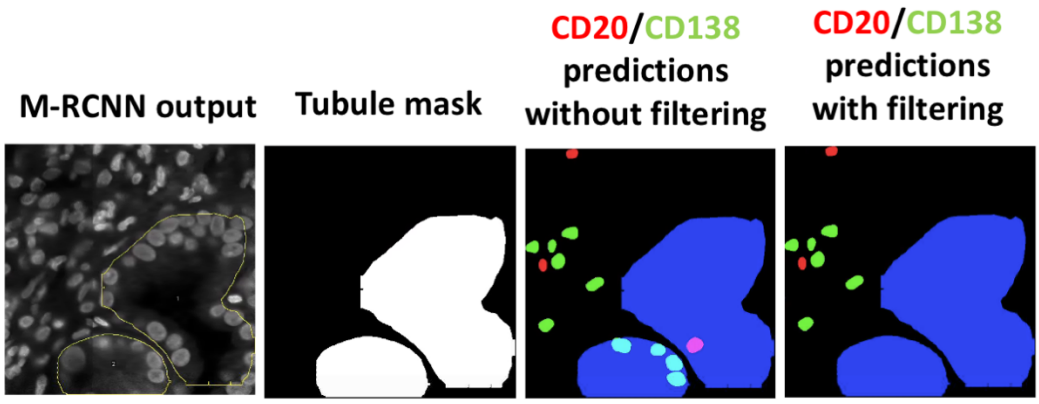

Figure 5.5. Tubule network output and usage in generating final cellular predictions; Mask R-CNN output is turned into a binary mask, and cellular predictions that fall into the mask are eliminated; second two panels—Blue=Tubule, green=CD138, red=CD20; cyan=CD138 overlapping with tubule, magenta=CD20 overlapping with tubule

Post-processing—compiling the results of the various networks

Much like the high resolution dataset, our "final" cellular predictions are the combination of the outputs of multiple networks—the T cell network, B cell network, and tubule networks. Because of the number of classes and networks involved, it was necessary to develop robust methods for combining these outputs and incorporate additional post processing steps.

During the course of generating ground truth for these training sets, we noted a population of CD3+CD4-CD8- (double negative, DN) T cells. Out of 1083 total manually annotated CD3+ T cells, 456 of them were DN, leading us to conclude that they were a substantial population in our dataset, though the non-random sample of images used for ground truth could not be used to accurately estimate their frequency. Therefore, we included DN T cells amongst the cellular classes we were segmenting.

In total we identified 5 "base" classes of cells – CD3+CD4-CD8- (DN), CD3+CD4+, CD3+CD8+, CD20+ and CD138+ (**Fig 5.6A**). The T cell network predictions for the

CD3/CD4/DAPI and CD3/CD8/DAPI tiles were combined to identify the three T cells classes (DN, CD3+CD4+, CD3+CD8+), which the DN class being derived from the overlap of the CD3+CD4- and CD3+CD8- predictions. The identified cells were then further filtered such that predicted cells that overlapped with bright non-specific objects (such as red blood cells) were removed using a noise mask derived from finding signals that are present across all fluorescence channels. In addition, cells that overlapped with the segmented tubules were dropped out, as they were assumed to be spurious bright signal from tubules. For the three T cell subsets (DN, CD4+, CD8+) there was an additional step where each cell was identified as ICOS+/-, PD1+/- and FoxP3 +/- (**Fig 5.6B**). ICOS and PD1 segmentations were derived by using the T cell network described above to predict CD3+PD1+ and CD3+ICOS+ cells, and identifying which of the predicted DN, CD4+ or CD8+ T cells overlapped with each one. Because FoxP3 is a nuclear marker, the T cell network could not perform well on it. Therefore, binary FoxP3 masks were generated by defining positive FoxP3 signal as anything greater than 3.5 standard deviations from the mean signal intensity. FoxP3+ cells were identified as those that had positive FoxP3 signal overlapping with at least 25% of the cell body.

Several subsets of T cells were identified using these three markers: regulatory T cells were defined as CD3+CD4+ PD1-ICOS- FoxP3+, Tfh cells were defined as CD3+CD4+ PD1+ICOS+/-FoxP3- and exhausted CD8 T cells were defined as CD3+CD8+ PD1+ICOS-FoxP3-.

At the end of this procedure, each tile had a set of cells associated with it, with descriptors like class ID and secondary marker expression for each cell.

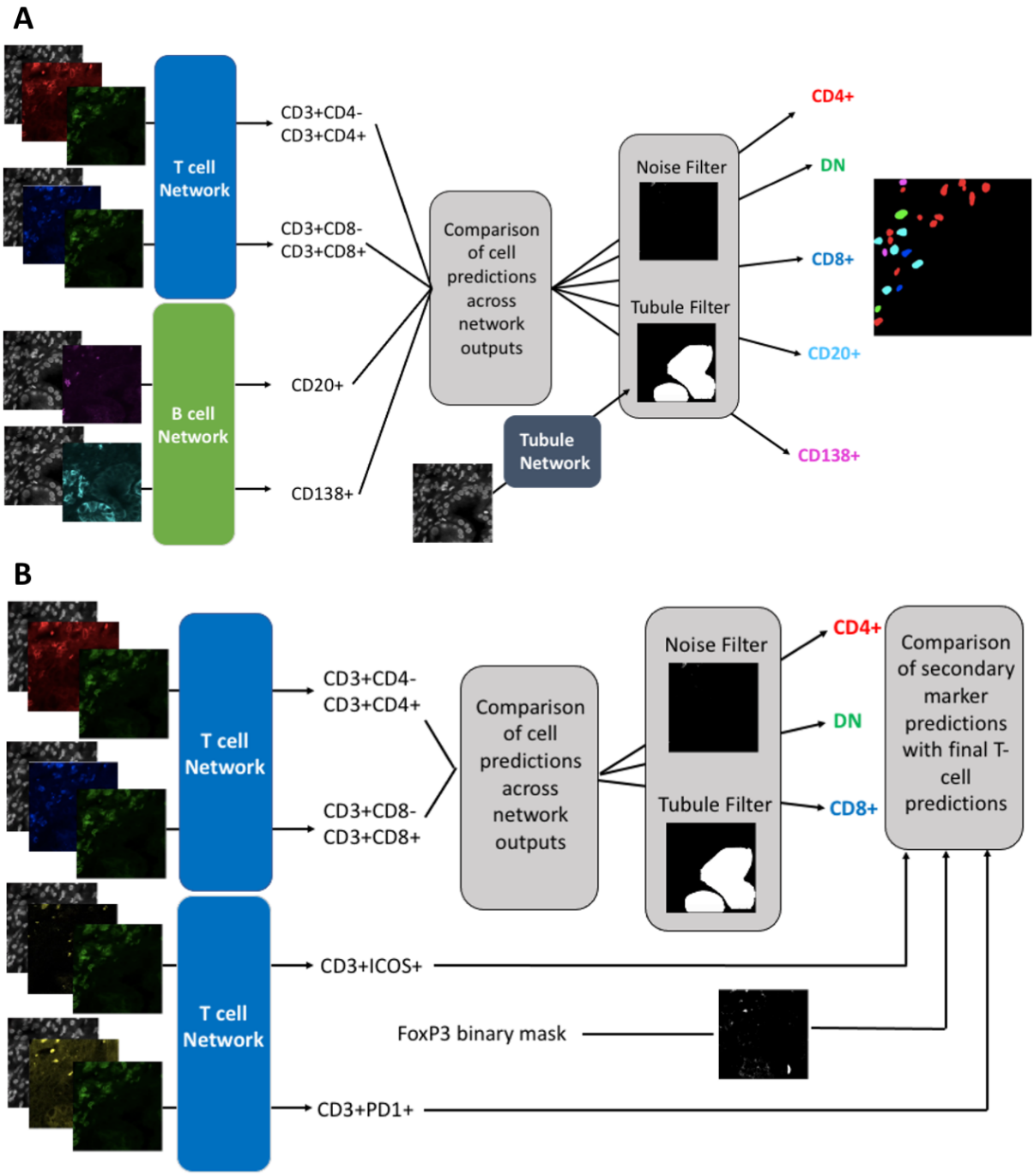

Figure 5.6. HMP Segmentation procedure; A) generation of base classes by consolidating the T cell and B cell network predictions and filtering out predictions that overlap with noise mask and tubules; B) generation of secondary marker classifications by comparing base class predictions with ICOS/PD1 predictions and FoxP3 mask

3. Sample-level processing

MX1 thresholding and analysis

MX1 thresholding was performed as follows: the composite was Z-score normalized, and the resulting tiles were broken into 16x16 pixel sub-regions. If more than 50% of the pixels in the sub-region exceeded a pixel intensity of 40 (out of a maximum of 255), that whole sub-region was categorized as “MX1+”, and if not it was categorized as “MX1-“. This was done to capture continuous MX1 signal and reject small bright spots. The MX1 positivity of sub-regions of tissue was determined by defining a bounding box in the composite that encompassed the area of interest (cells in a cluster), and then calculating what percentage of the bounding box was MX1+.

Area calculation

The total area of tissue in each sample was calculated by analyzing the DAPI composite in the following way: 1) a threshold pixel intensity of 5 was applied to binarize the image; 2) the skimage.morphology (version 0.17.2) functions “area_closing” (area_threshold=5000), “remove_small_holes” (area_threshold=5000), “remove_small_objects” (min_size=100), and “binary dilation” were applied to the image in that order to fill in small areas of tissue that don’t contain nuclei; 3) the number of positive pixels was then calculated and converted to units of um^2 using 0.0488 $um^2/pixel^2$ as the conversion factor. The number of cells per unit area were calculated by dividing the cellular content of the full composite by this calculated area.

4. Spatial Analyses

All spatial analyses were performed in the programming language Python (3.7.9). The following packages were utilized for analysis: pandas (1.2.2), numpy (1.19.2), sklearn

(0.23.2), scipy (1.6.1), and tifffile (2021.1.14). Plotting was performed with matplotlib (3.3.2) and seaborn (0.11.1).

Assigning composite-level coordinates to cells

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.

Nearest Neighbors

The nearest neighbors analyses were performed as described in **Chapter 3**, such that, for every cell in the dataset, the class of its nearest cellular neighbor was identified. The biggest change to this procedure was that the calculation was performed using the composite-level coordinates, such that cells that were in adjacent tiles could be neighbors.

Defining neighborhoods with DBSCAN

As in the HR dataset, Density Based Spatial Clustering of Applications with Noise (DBSCAN)[126] (sklearn 0.23.2) was used to identify cellular neighborhoods with a minimum cluster size ("min_samples")=2, and a distance cutoff of 50 pixels, which converts to approximately 10 um. Once again, doing this calculation on the composite level allowed us to identify neighborhoods that spanned multiple tiles.

Neighborhood minimum distance

In order to evaluate the distances between neighborhoods a bounding box was drawn around each cellular neighborhood, and the coordinates of the center of the box was used

as an approximate neighborhood centroid. The centroid-centroid distance between each neighborhood and the neighborhood closest to it was identified.

5. Follow-up investigation of CD4-CD8- T cells

During the course of our investigation we identified the presence of a population of CD3+CD4-CD8- T cells in our LuN dataset. We followed up on this finding by interrogating a publicly availably single cell RNA sequencing dataset from lupus nephritis patients[14], and by doing additional staining on a small cohort of LuN samples.

RNA sequencing analysis

Single-cell RNA-seq data for human lupus nephritis 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 the original paper[14], such that cells that cells were removed from the analysis if they expressed <1,000 or >5,000 genes, or if more than 25% of the total unique molecular identifiers (UMI) mapped to mitochondrial genes. Gene expression values were normalized to library size (UMI count per million) and scaled by log2. 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 "Naïve 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* or *CD8B* with no *CD4* transcripts. Cells were categorized as DP (double-positive) or DN (double-negative) when they had both/neither *CD4* and/nor

CD8A/B. t-SNE was performed by Rtsne (0.15). Plots were generated by ggplot2 (3.3.2) and ggridges (0.5.2).

Additional staining

Eight lupus nephritis kidney biopsies were stained for CD3, CD4, CD8, TCRδ, and DAPI. Inflamed regions were imaged on the Leica Stellaris 8 confocal microscope, with 40x magnification and a pixel size of 0.225 um. 281 ROIs (35±19 per sample) 1024x1024 ROIs were obtained, and then post processed with background subtraction, despeckling and contrast adjustment using ImageJ. Cells in these images were quantified by manual count.

6. Summary

Highly multiplexed imaging enables detailed phenotyping of infiltrating immune cells in tissue. However, this type of imaging introduces a series of analytical challenges that needed to be addressed in order to accurately segment cells in tissue. For the lupus nephritis dataset, the images were stitched into composites, aligned across days, and broken into tiles. These tiles were then filtered to select for those that had CD3+ or CD20+ lymphocytes. Two instances of Mask R-CNN were used to segment CD4+, CD8+ and DN T cells and CD20+ and CD138+ B cells in the tissue. A third instance of Mask R-CNN was used to identify tubules in the tissue in order to filter out false positives produced by the lymphocyte networks. The T cells were further characterized by the expression of ICOS, PD1, and FoxP3. Once sufficiently high-quality segmentations were obtained, the predictions were stitched back together into composites, allowing for spatial analysis to be performed on the scale of the whole section. It should be noted that we have thus far only segmented a subset of the markers in the lupus nephritis data, and we still need to

apply these methods in the TNBC data. Additional work needs to be done in order to make full use of all of the markers in these dataset. Nonetheless this work has led to additional insights about the cellular constituents of lupus nephritis and provides a framework for analyzing similar datasets.

7. Acknowledgements

I would like to thank and acknowledge the following people who contributed to this part of the project: Margaret Veselits and Dr. Junting Ai collected the highly multiplexed lupus nephritis dataset and assisted with the dataset stitching and alignment. Junting Ai collected the additional dataset on CD4- T cells. Dr. Madeleine Durkee led the development of the cellular segmentation procedure and advised on the spatial analyses. Dr. Yuta Asano performed the RNA sequencing analysis. Dr. Kichul Ko, Dr. Emily Peninger, and Dr. Anthony Chang curated the lupus nephritis dataset, and Dr. Jeffrey Mueller curated the TNBC dataset. The Integrated Light Microscopy Facility at the University of Chicago (Christine Labno, Shirley Bond and Vytas Bindokas) provided expertise on imaging. Chun-Wai Chan administered the MEL computation server and provided troubleshooting help related to its usage.

Chapter 6: Spatial analysis of highly multiplexed dataset—probing B-T interactions

1. Introduction

Given the collective observations from the high resolution (HR) dataset, several questions emerged—what is the phenotype of the CD4- T cells that seem to be associated with poor outcomes? What is the phenotype of CD4+ T cells that are present in large groups of B cells? What is the relative abundance of CD20+ B cells and CD138+ plasma cells? These questions were the driving force behind the analyses we chose to pursue in the highly multiplexed (HMP) dataset.

2. Quantifying lymphocyte abundance in the HMP dataset

Though we had many markers at our disposal, we first segmented the CD20+ B cells, CD138+ B cells, CD4+ and CD8+ T cells for the purpose of this analysis. While we were developing our segmentation procedure, we determined that there was an additional population of CD3+CD4-CD8- (double negative, or DN) T cells, which we also identified throughout our dataset (**Fig 6.1A**).

Of these five classes, CD4+ T cells were the most abundant subset, comprising 35.2% of the segmented lymphocytes. There is substantial between-patient heterogeneity in terms of the density of these cellular classes, which is consistent with previous observations of (**Fig 6.1B**)

As one might expect, there was a high degree of correlation between total CD4+ T cells and CD8+ T cells per total biopsy area within patients (R=0.789) (**Fig 1C**), validating the previous observations that these cell types generally track together. There was also a high degree of correlation between CD20+ and CD138+ B cells (R=0.775), suggesting that an abundance of CD20+ B cells often corresponds with abundant plasma cells (**Fig**

6.1D).

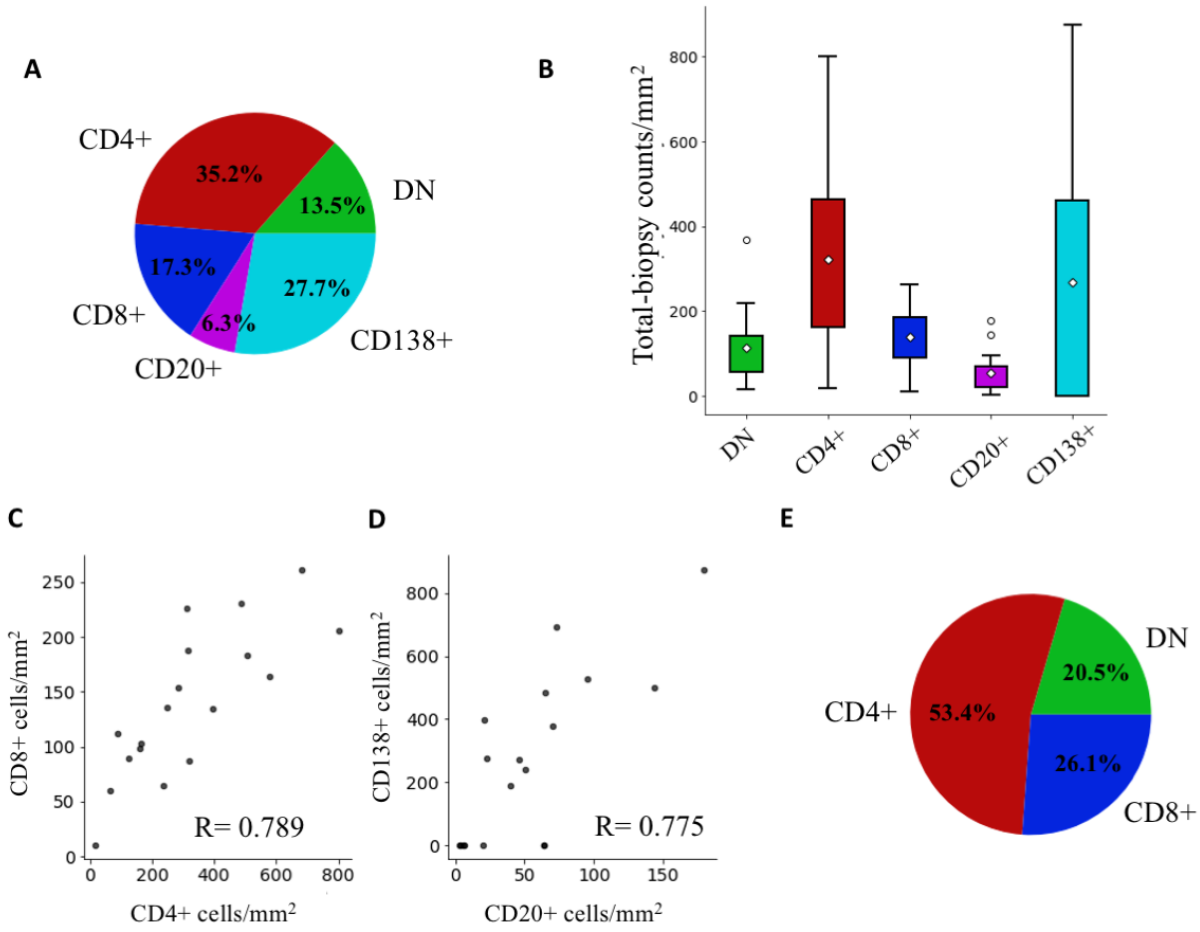

Figure 6.1. Diverse array of lymphocytes are observable in the HMP dataset; A) Distribution of segmented cell classes in entire dataset; B) distribution of the frequency of segmented cell classes per total biopsy area on a per-patient basis; C) Sample-level correlation between density (cells/mm^2) of CD4+ and CD8+ T cells; D) Sample-level correlation between density (cells/mm^2) of CD20+ and CD138+ B cells; E) Distribution of three T cell subsets in the dataset

3. Evaluating phenotypic heterogeneity in the T cell compartment

It was first of critical importance to better understand the T cell compartment. We further phenotyped the T cell subsets by marking each cell as positive/negative for FoxP3,

ICOS, and PD1. These markers were chosen for their ability to identify regulatory T cells (FoxP3), T follicular helper cells (PD1, ICOS), and exhausted T cells (PD1).

"CD4-" T cell compartment

One of the biggest outstanding questions from the HR dataset was whether the entire CD4- T cell compartment is CD8+. In this dataset we have found that 53.4% of the T cells are CD4+, 26.1% are CD8+ and 20.5% are DN (**Fig 6.1E**). It should be noted that DN T cells are difficult to quantify accurately in our system, given that they are defined by the lack of markers and this designation is derived from two different outputs of the T cell network (i.e, CD3+CD4, CD3+CD8-). Nonetheless, their presence was manually confirmed in several biopsies, as described in **Chapter 5**. This indicates that the "CD3+CD4-" T cells that were associated with renal failure in the HR dataset cannot automatically be presumed to be CD8+ T cells.

To further characterize DN T cells, we interrogated public single-cell RNA-seq data of immune cells infiltrating the kidney of lupus nephritis patients[14]. We identified naïve T and CTL clusters in intrarenal immune cells by unsupervised clustering and canonical marker expression (**Fig 6.2A**). Within these T cell clusters, 21% were DN, as measured by the UMI of CD4, CD8A, and CD8B (**Fig 6.2B**). Several T cell subtypes have been reported to be DN, such as NKT cells and gamma delta T cells. Indeed, there was a small increase in CD3d mRNA is cells assigned to NK cells, suggesting a NKT phenotype. We therefore looked for the expression of *NCAM1* (CD56) and *KLRB1* (CD161) in all of the T cell subsets. Neither of these markers were substantially enriched in DN T cells, indicating that DN NKT cells are rare in lupus nephritis (**Fig 6.2C**).

Next, we compared TCR alpha and delta chain expression (*TRAC* and *TRDC*). There were some cells that were apparently positive for both *TRAC* and *TRDC*, likely due to sequence homology between these genes (**Fig 6.2D**). However, *TRAC*- cells and *TRDC*+ cells were both enriched in DN population (**Fig 6.2E**). These results suggest that at least a portion of the DN T cells observed in lupus nephritis are gamma-delta T cells. Given this finding, we confirmed the presence of gamma-delta T cells in a new set of 8 lupus nephritis biopsies. These biopsies were stained with a panel of CD3 CD4, CD8, TCRδ, and 281 ROIs were captured (**Fig 6.2F**). In a given biopsy, 51.4% ± 21.3% of CD3+CD4-CD8- T cells were positive for TCRδ. These findings indicate that within lupus nephritis a substantial fraction of T cells, similar to the prevalence of CD8 T cells, detectably express neither CD4 nor CD8. Within this population, approximately half are gamma-delta T cells.

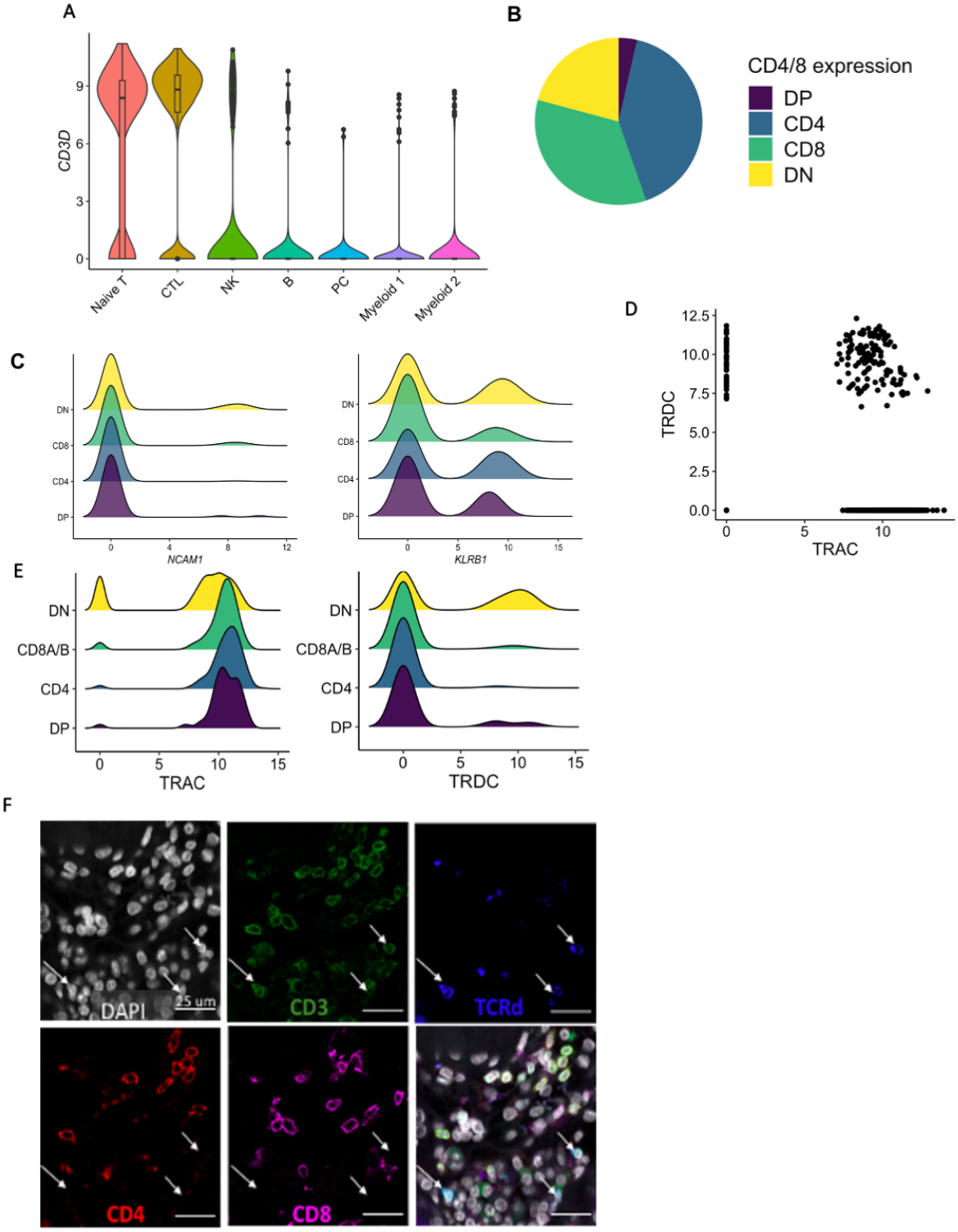

Figure 6.2 Figure caption continued on next page

Figure 6.2, continued. Gamma-delta T cells are a subset of CD4- T cells in lupus nephritis; A) distribution of CD3 in cell clusters identified in single cell RNA sequencing data from LuN kidney samples; B) Distribution of *CD4/8A/8B* expression. DP: double-positive, DN: double-negative; C) Distribution of *NCAM1* and *KLRB1* expression in DP, DN, CD4 and CD8; D) expression of TRAC and TRDC in identified CD3+ cells; E) comparison of TRAC and TRDC expression in identified DN, CD8+, CD4+ and DP T cells; F) representative image of double negative (CD4-CD8-) gamma delta (TCRd+) cells in LuN biopsy, marked by white arrows; representative composite of CD4 negative clusters, red=CD4+ T cells, blue=CD8+ T cells, green=DN T cells

The final question around CD4- T cell phenotype we wanted to address was whether CD8+ T cells in lupus nephritis biopsies were exhausted. Roughly 30% of CD8+ T cells in our dataset are PD1 positive (**Fig 6.3A**). A more stringent definition of exhausted CD8+ T cells (Tex) is CD3+CD8+PD1+ICOS-Foxp3-, and 25% of the total CD8+ T cells were "exhausted" by this definition. This is coherent with observations from murine models in which tissue-infiltrating CD8+ T cells are receiving chronic antigen stimulation that drives them to exhaustion[130], but in contrast to single cell RNA sequencing data that has suggested that CD8+ T cells in the kidneys are infrequently exhausted.[14] This discrepancy could be because of the differential dynamics in protein and RNA expression.

From all of this we conclude that the "CD4-" T cells observed in lupus nephritis are a diverse cohort, that include a subset of CD3+CD4- gamma-delta T cells. Further work will need to be done in order to identify which subset of these cells are associated with progression to renal failure.

CD4+ T cell compartment

We next wanted to understand the diversity of the CD4+ T cell compartment, with a particular focus on regulatory T cells and T follicular helper cells.

We specifically defined regulatory T cells (Tregs) as CD3+CD4+PD1-ICOS-FoxP3+. A surprisingly small percentage (5.41%) of CD4+ T cells were FoxP3 positive, and an even smaller percentage met the more stringent criteria of being PD1-ICOS-, reducing the percentage of identified Tregs to 2.46% (**Fig 6.3B**). In contrast, even fewer CD8+ T cells (1.3%) or DN T cells (0.88%) expressed FoxP3, which is in line with our expectation that the majority of FoxP3+ regulatory T cells are CD4+ (**Fig 6.3A,C**). This finding indicates that very few of the tissue-infiltrating CD4+ T cells in lupus nephritis are Tregs, suggesting that the presence of CD4+ T cells are most often driving pathology rather than inhibiting it.

We additionally identified T follicular helper-like (Tfh) cells based on the expression of PD1 and ICOS. The most stringent definition for these cells using our set of markers would be PD1+ICOS+FoxP3- CD4+ T cells, which made up 5.05% of the CD4+ T cell compartment. If we use a less stringent definition of PD1+ICOS+/-FoxP3-, this percentage goes up to roughly 30% of the CD4+ T cells. It should be noted that one could alternatively interpret PD1+ICOS-FoxP3- CD4+ T cells as exhausted. However, given that previous investigations have consistently associated PD1 expression with Tfh-like cells, we decided to use the less stringent definition to identify this cell subset in our data.[14,38] (**Fig 6.3B**).

Probing regional variation in T cell phenotype

For the three specific T cell subsets that we were interested in (Tex, Treg, Tfh), we probed their potential interacting partners by identifying their nearest neighbors (the cell with the shortest centroid-centroid distance). Most Tregs are in fact closest to other Tregs, and other non-specific CD4+ T cells (**Fig 6.3D**). In contrast to the expectation that Tfh

would primarily be in close proximity with CD20+ B cells, Tfh by far had other CD4+ T cells as their most frequent neighbor, followed by other Tfh (**Fig 6.3E**). Exhausted CD8+ T cells were most frequently found near other exhausted CD8+ T cells, followed by CD4+ T cells and CD8+ T cells (**Fig 6.3F**). These observations lead to the hypothesis that there may be localized niches in which most of the CD8 T cells are exhausted.

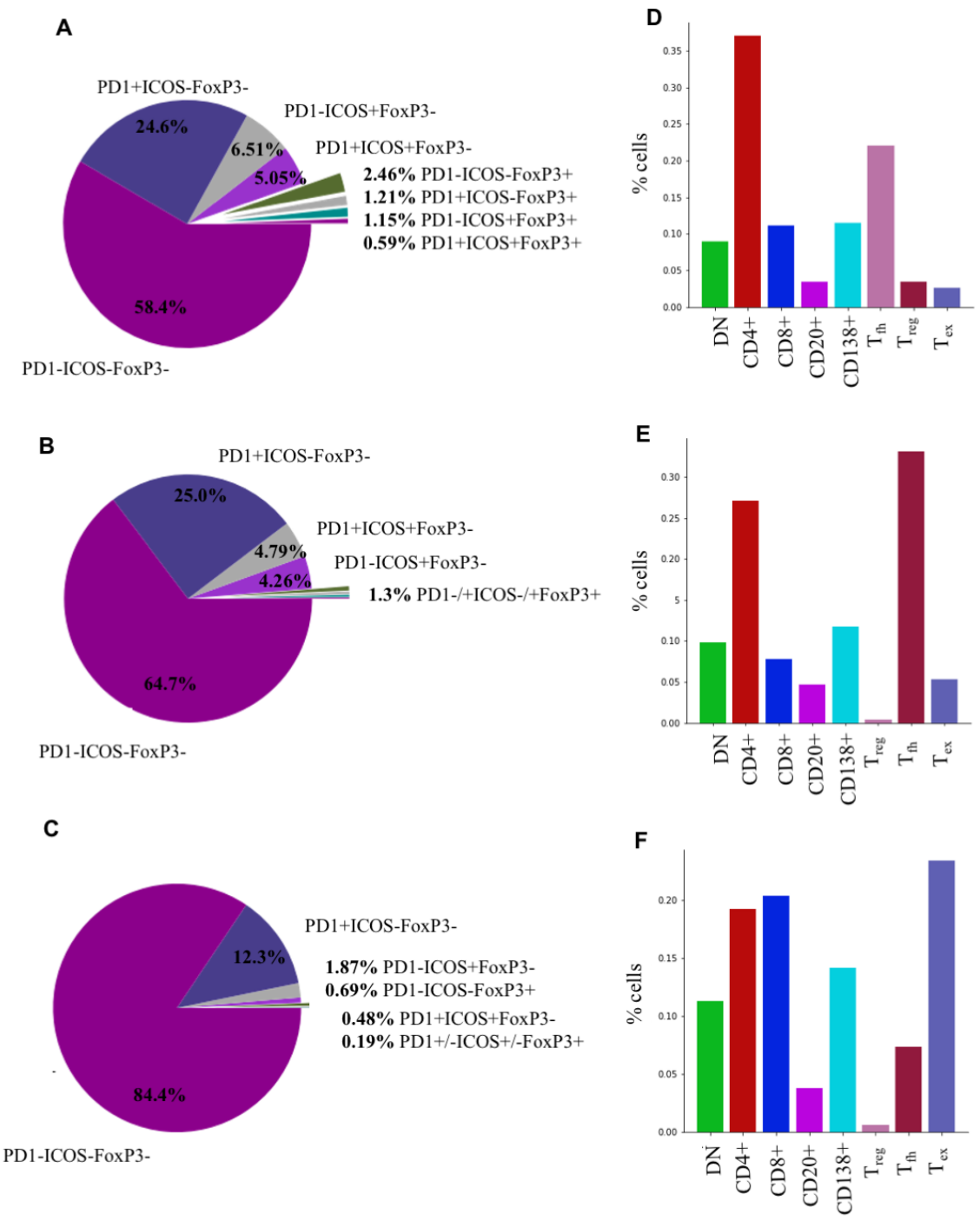

Figure 6.3. Deeper phenotyping of T cells. Distribution of ICOS, PD1, and FoxP3 expression in A) CD4+, B) CD8+ and C) DN T cells; Distribution of nearest neighbors of D) Tregs, E) Tfh, and F) Tex

Notably, 58.4% of CD4+ T cells, 64.7% of CD8+ T cells, and 84.4% of DN T cells (**Fig 6.3A-C**) expressed none of these three secondary markers, which means that there is still substantial phenotypic heterogeneity that was not captured by this panel.

4. Defining inflammatory aggregates in tissue

As in the HR dataset, we wanted to define cellular neighborhoods. We again used Density Based Spatial Clustering of Applications with Noise (DBSCAN)[147] to define these neighborhoods, such that every cell in a tissue sample was either assigned to a cellular cluster with a minimum cluster size of 2 or was defined as a singlet cell. A distance cutoff of 50 pixels was chosen because this is approximately 10 um at the resolution of this dataset, and therefore matches the distance cutoff used in the HR dataset. We identified a total of 13921 neighborhoods across the 18 samples. With this method we were able to capture large aggregates that could span multiple tiles, such that the largest aggregate that was captured had 273 cells in it (**Fig 6.4A**). In contrast, the largest aggregate captured in the high resolution dataset (which was restricted to a to ~$1.2x10^4$ μm^2 of tissue), contained only 147 cells.

We first examined the singlet population, which were 23% of the total segmented cells, to identify any cellular subsets that might not be included in dense cellular aggregates. However, the class breakdown of singlets was not different from the class breakdown of the cells overall (**Fig 6.4B**). This indicates that none of the cell subsets that we were probing were more likely than the rest to be found as singlets and excluded from large cellular clusters.

Next, we looked at “doublets”, or neighborhoods that had exactly 2 cells in them. These constituted 46% of the total clusters, but only accounted for 20% of all of the

segmented cells. **Figure 6.4C** shows the breakdown of cell classes that are present in the doublets, which is again similar to the overall breakdown of classes in the dataset. This again suggests that no particular cell subsets are found in these relatively isolated pairs at a higher rate than the others.

We evaluated all of the possible combinations of class pairs and found that the most common pairs were CD4+/CD4+ (24.7%) and CD138+/CD138+ (17.7%). Only 2% of doublets were CD4+/CD20+, which suggests that most CD4 T cell-B cell interactions are not happening in isolated pairs of cells, but possibly in large neighborhoods (**Fig 6.4D**).

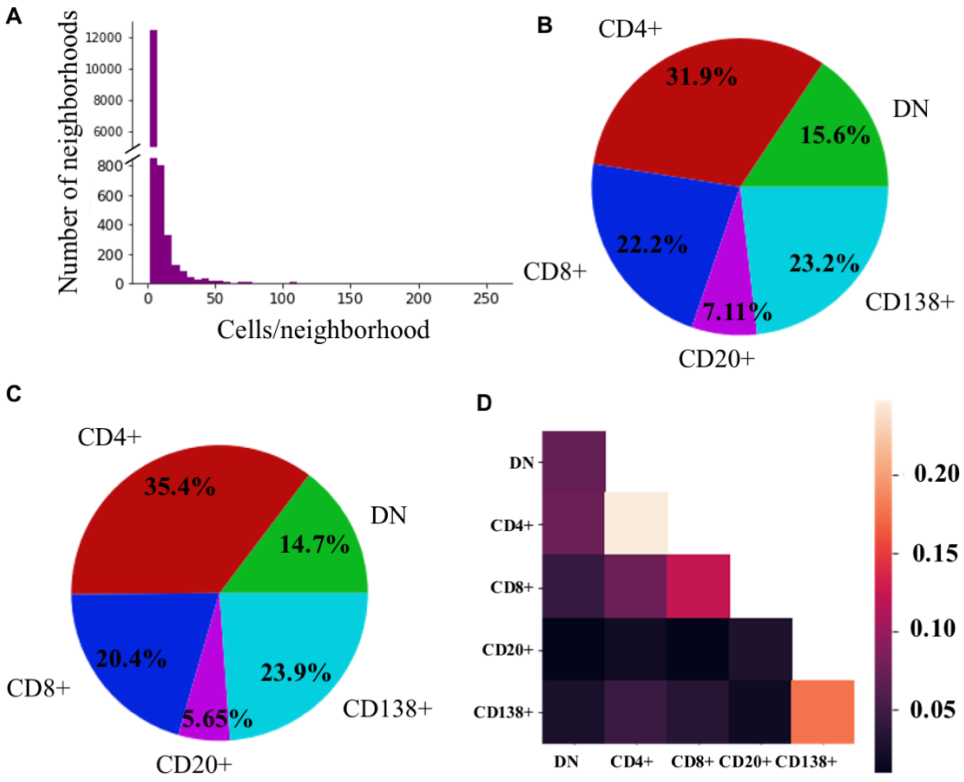

Figure 6.4. Singlets and doublets are a substantial proportion of observed neighborhoods A) distribution of neighborhood size; B) breakdown of five classes amongst the singlets (cells that were not assigned to a neighborhood); C) breakdown of five classes amongst the doublets (cells assigned to neighborhoods with n=2 cells); D) heatmap of pairs of cell classes found in doublets

CD4- neighborhoods

Given the apparent abundance of CD4- T cell enriched neighborhoods in the ESRD+ patients in the HR data, we investigated the CD4- compartment in the HMP data. In accordance with the HR data, we classified CD4- neighborhoods as those that had 1) 25% of their cells as either CD8 T cells or DN T cells; 2) less than 20 cells. This numerical cutoff was chosen because neighborhoods larger than this were classified as "large" in the HR data. **Fig 6.5A** shows a representative image of several of these neighborhoods. A total of 60.2% of the cells in these neighborhoods are CD4-, with 26% DN T cells and 34.2% CD8+ T cells (**Fig 6.5B**). 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 equally represented in a given neighborhood.

B-CD4+ T clusters

We next wanted to identify and characterize the analogous "large, lymphocyte-rich" neighborhoods in this dataset. In the high dimensional dataset, large B-T neighborhoods were defined by the set of criteria that captured most of the "large/mixed identity" neighborhoods in the high resolutions data: 1) a minimum of 20 cells, 2) both B cells and CD4+ T cells were represented in the neighborhood and 3) at least 50% of all cells were B cells or CD4+ T cells. 111 of >13,000 neighborhoods met these fairly stringent criteria (representative clusters in **Fig 6.5C)**. Within these aggregates, 52.2% were CD4+ T cells, 11.9% were CD20+ B cells, 11.7% were CD8+ T cells, 11.6% were DN T cells and 12.7% were CD138+ plasma cells (**Fig 6.5D**). Tfh made up roughly 36% of CD4+ T cells in these B-T neighborhoods (**Fig 6.5E)**. This represents a significantly larger percentage than non-B-T neighborhoods ($p = 1.45 \times 10^{-30}$) **(Fig 6.5F)**. The nearest neighbors of B cells

in these neighborhoods were largely enriched for other B cells, Tfh cells, and non-Tfh CD4+ T cells (**Fig 6.5G**). Tfh cells in these neighborhoods were largely found with other Tfh cells and other unspecified CD4+ T cells (**Fig 6.5H**). Unassigned CD4+ T cells in these aggregates were also most likely to be found near other CD4 T cells, followed by Tfh (**Fig 6.5I**).

To relate this identified type of cluster to the HR dataset, we probed the “large/mixed identity” aggregates and found that 96.1% of these identified aggregates matched the profile of the B-T aggregates (more than 50% cells are CD4+ T cells and B cells, with at least one of each).

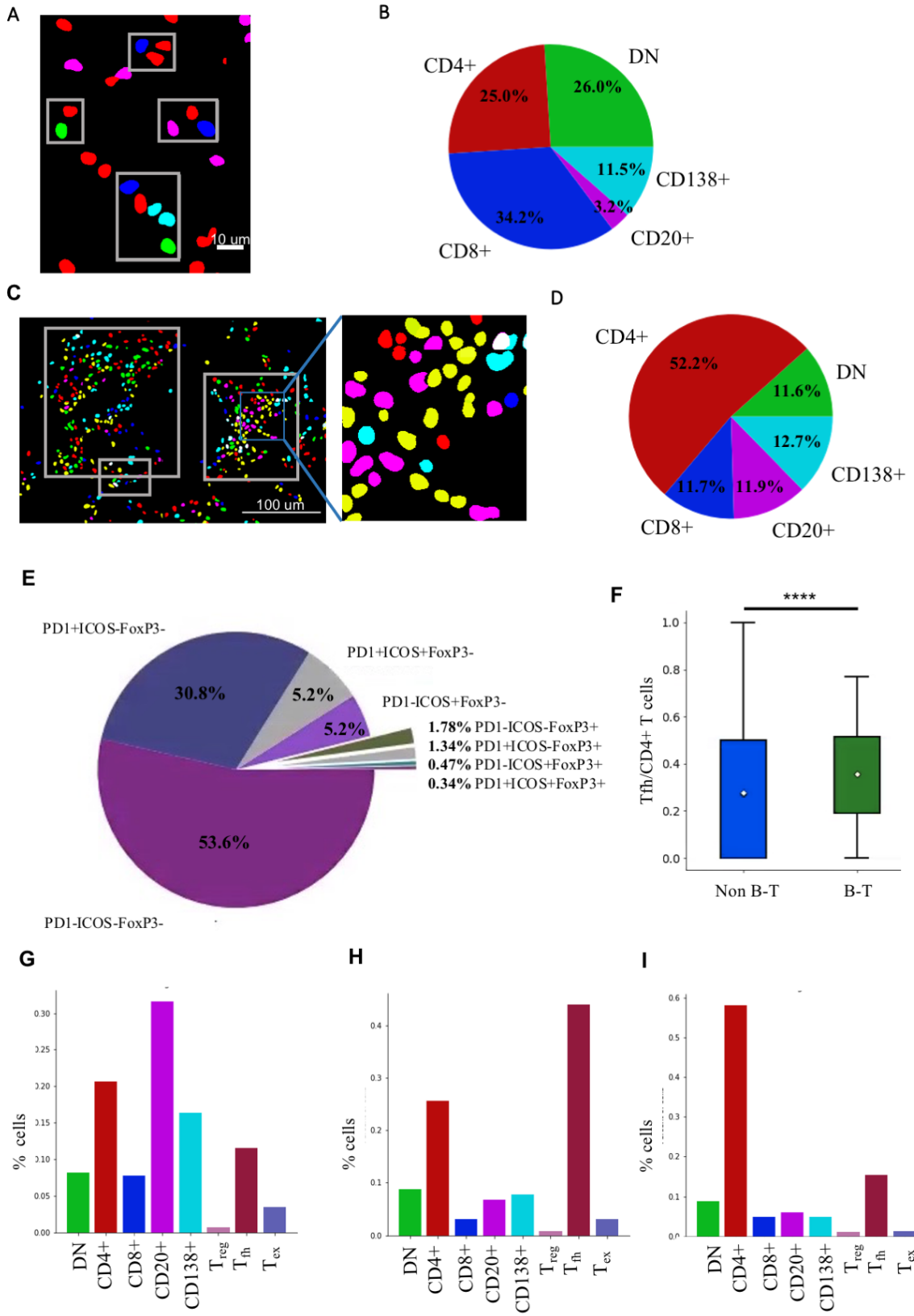

Figure 6.5 Figure caption continued on next page

Figure 6.5, continued. Identification of distinctive CD4- and B-T neighborhoods; A) representative composite of CD4-clusters, red=CD4+ T cells, blue=CD8+ T cells, green=DN T cells; B) breakdown of frequencies of the five base classes in CD4 negative neighborhoods; C) representative composite showing identified B-T aggregates (outlined by white boxes), green=DN, red =non-Tfh CD4+, yellow=Tfh; blue=CD8+, magenta =CD20+,cyan = CD138+; D) breakdown of frequencies of the five base classes in B-T neighborhoods; E) breakdown of CD4 T cell phenotypes in B-T neighborhoods; F) comparison of proportion of CD4+ T cells that are Tfh in identified B-T aggregates vs non B-T aggregates; nearest neighbors of G) CD20+ B cells, H) Tfh, and I) CD4+ T cells within B-T aggregates

Higher level organizations are apparent in inflamed biopsies

Though this neighborhood analysis has provided some useful insights, It should also be noted that the manner in which we defined cellular clusters in both datasets was defined by the approximate size of a cell body and excluded longer-range interactions. We observed that defined clusters often existed within larger apparent aggregates of cells, suggesting that there are multiple scales on which cells might organize (**Fig 6.6A**). In order to get a sense of these higher order architectures (or “cities”), we approximated the minimum distance from each neighborhood to another neighborhood by drawing a bounding box around each neighborhood and calculating the distances between the centers of the neighborhood bounding boxes. The average distance between adjacent neighborhoods was roughly 158.5 pixels, or 35 um. **Fig 6.6B** shows that the distribution of minimum distances to the nearest neighborhood is heavily skewed towards short distances, suggesting that most defined neighborhoods exists within large inflammatory structures. An intriguing relationship emerged when we plotted the relationship between the number of cells in the neighborhoods and the distance to the nearest neighborhood, such that the largest neighborhoods typically were a short distance away from other cellular neighborhoods, while there were some small cellular neighborhoods that were

very far away from the nearest neighborhood (**Fig 6.6C**). This hints at patterns of both dense and sparse inflammation in the tissue. Therefore, though we restricted our analyses to these short-range interactions, exploring these higher-level structures is an important next step.

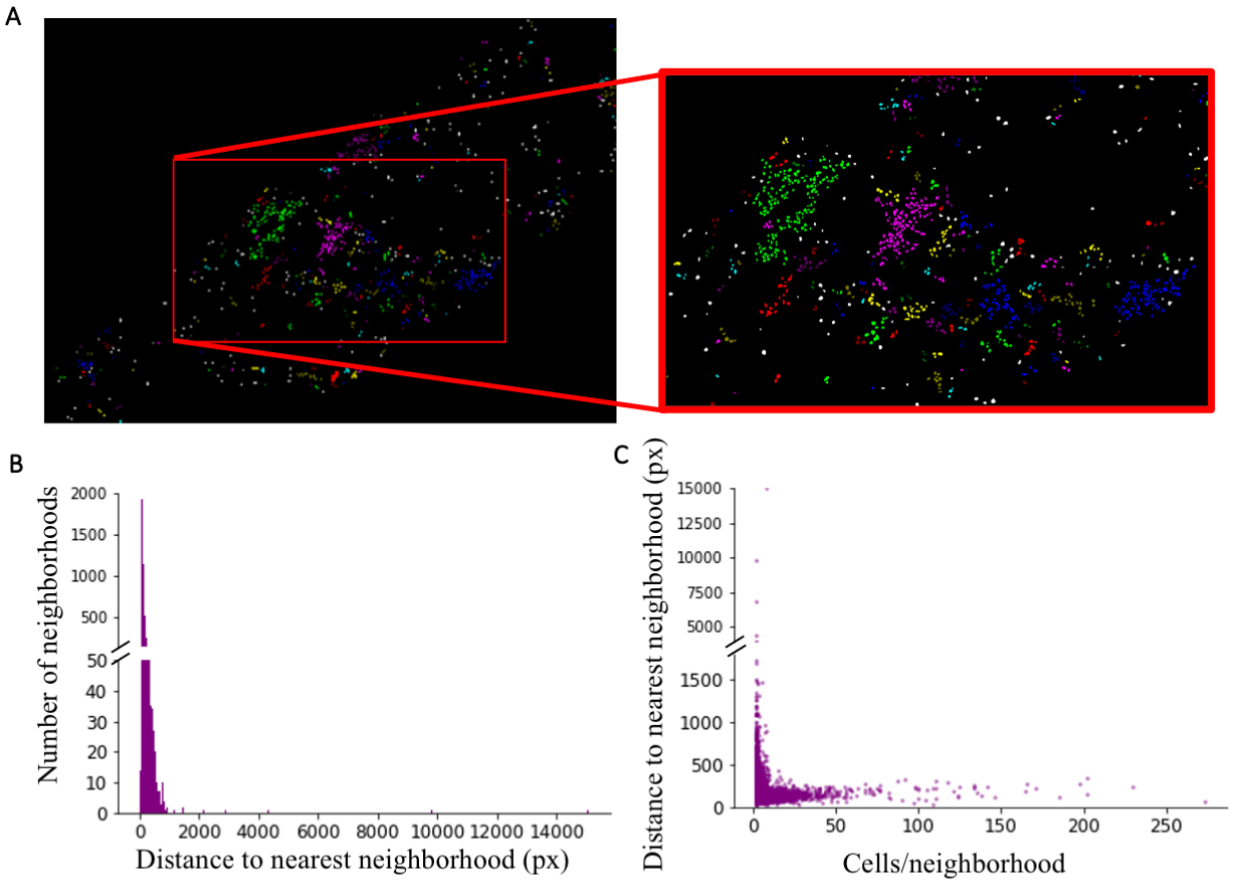

Figure 6.6. Higher orders of neighborhoods can be identified in the tissue. A) Representative image of neighborhood detection; inset: cells that are the same color and spatially close have been assigned to the same neighborhood; distances between neighborhoods were calculated based on the centers of the smallest bounding box that can be drawn around each neighborhood; B) distribution of the minimum distance in pixels of each neighborhood to another neighborhood; C) relationship between the number of cells in a neighborhood and the minimum distance to another neighborhood.

5. Comparing with the high resolution dataset—chronicity as a proxy

Though we don't have long-term outcomes data on the patients from the highly multiplexed dataset, we do have their tubulointerstitial chronicity scores. Tubulointerstitial chronicity roughly correlates with kidney failure, though it is an imperfect proxy. We probed the highly multiplexed dataset by splitting the patients by whether they had high (4 or greater) TI chronicity or low (less than 4). We found that the neighborhoods were larger in high-chronicity patients, both in terms of the average aggregate (5.07 cells in high chronicity vs 3.96 cells in low chronicity, p value = 1.2×10^{-13}), and in terms of the overall distribution of aggregate sizes (p value by Kolmogorov-Smirnov test=1.9×10^{-9}).

In addition to increased phenotypic information, we wanted to use this dataset to validate our findings from the high resolution dataset. Though we do not have long-term outcomes data from these patients, we do have information about their tubulointerstitial chronicity at the time of biopsy. This allows us to compare cellular density high/low chronicity patients in both dataset. The tissue area covered by a 512x512 tile in the highly multiplexed dataset at 0.221 um/pixel is roughly equivalent to the tissue area covered by a 1024x1024 tile in the high resolution dataset at 0.1058 um/pixel, allowing for comparisons of ROIs between the two datasets. If we further restrict the tiles that we examine in the highly multiplexed dataset to those that were positive for T cells, approximating the CD3 entry criterion used during imaging of the high resolution dataset, we can look at roughly equivalent sets of images.

In keeping with the findings from the high resolution dataset, the total cells per ROI were higher in the high chronicity patients than in the low chronicity patients (p value = 6.87×10^{-133}) (**Fig 6.7A**). Similarly, there were higher numbers of both CD4- (encompassed

by both CD8+ and DN) T cells and CD4+ T cells in the high chronicity patients (**Fig 6.7B-C**). However, one finding that was not coherent between the two datasets was that in the highly multiplexed dataset there was also an enrichment for CD20+ B cells per ROI in the high chronicity patients, whereas the equivalent cohort in the high resolution dataset had fewer CD20+ B cells per ROI (**Fig 6.7D**). This discrepancy is possibly due to the low sample number (n=18) in the highly multiplexed dataset, or due to differences in how the data was collected (selective imaging of inflamed regions vs whole-biopsy capture). Further imaging experiments with a sufficiently large number of patients should be aimed towards resolving this apparent contradiction.

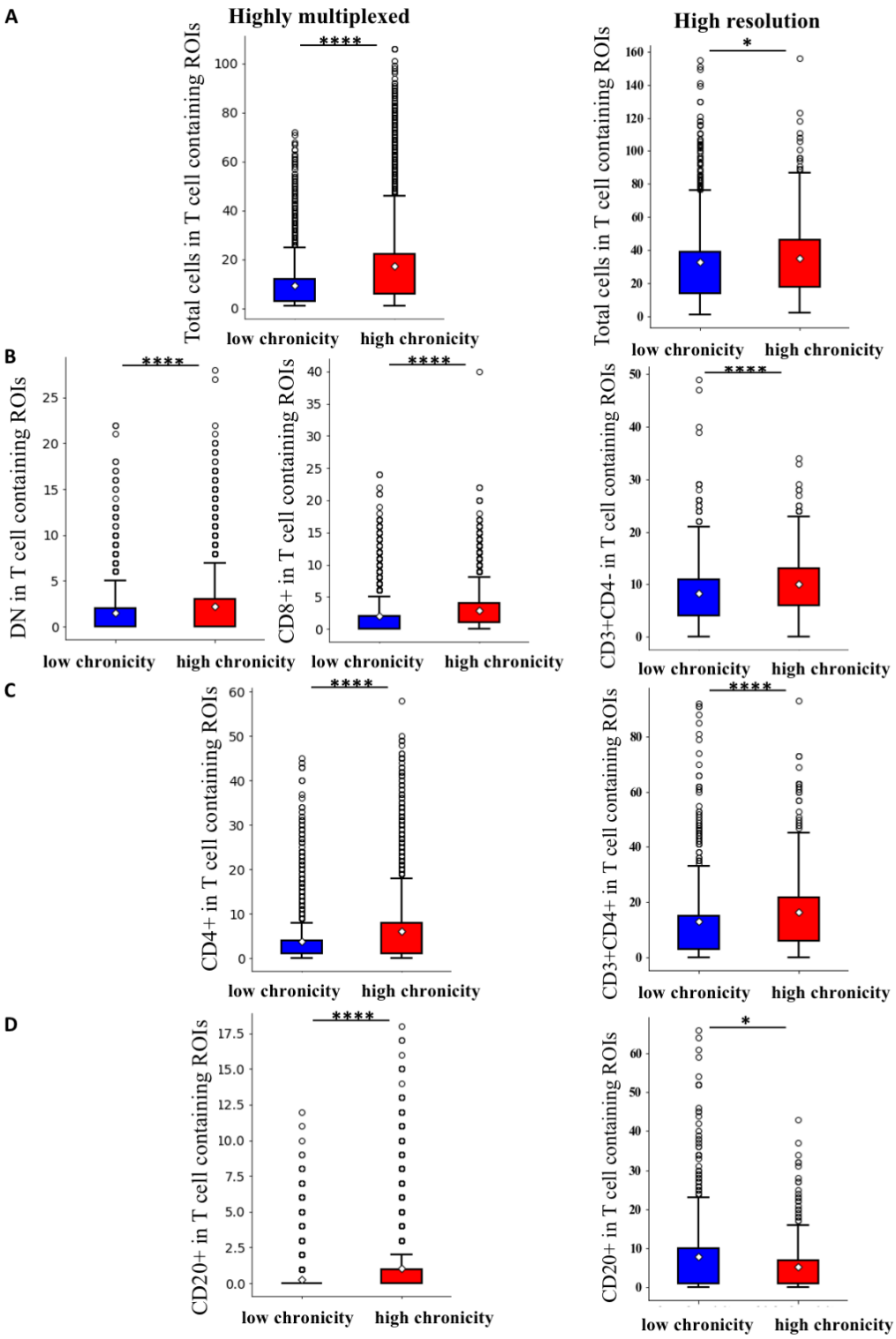

Figure 6.7. Relationship between TI chronicity and inflammation largely consistent between datasets. Comparison of per ROI TI chronicity analysis in HR and HMP datasets; A) total cells; B) DN and CD8+/CD3+CD4-; C) CD4+; D) CD20+

6. Relationship with MX1

One question that was of interest for this dataset was the relationship between interferon signaling and cellular abundance. High interferon signatures in the urine and kidney tissue of patients have been identified in lupus nephritis patients, but studies around the pathogenic significance of this have produced mixed results, with some associating an interferon signature with responders to therapy[35], and others associating it with non-responders[80]. Part of the discrepancy could be that little information exists about the tissue-level regional variation in interferon signaling within patients. We probed this by quantifying the expression of MX1 across the tissue. We found that there was substantial heterogeneity within samples, such that there were regions of high MX1 signaling, and regions of low MX1 (**Fig 6.8A**). This suggests that there is local variation in the intensity of interferon response. It was therefore of interest to identify what features were correlated with increased MX1 expression.

We then wanted to relate the regional abundance of MX1 to the overall presence of lymphocytes, and evaluated to what degree these cells co-localize with MX1 signal. The obvious hypothesis is that MX1 rich regions will have abundant lymphocytes, and vice versa. To our surprise, there was not a strong correlation between local lymphocyte abundance and MX1 signal. While there were certainly observable large aggregates with rich MX1 expression, there were also comparably large aggregates with almost no expression (**Fig 6.8B**). To quantify this, a bounding box was established around each identified cellular cluster, and the percentage of MX1+ pixels within this bounding box was evaluating for each cluster. Of the 356 large ($n \geq 20$) clusters identified in the dataset, only 52 (15%) overlapped with regions that had 50% MX1 expression. In examining the

difference in cellular constituency between MX1 rich (>=50% MX1 expression) and MX1 poor clusters (<50% MX1 expression), we found that there was no difference in overall cellular abundance within clusters, and there was a slight increase in CD4+ T cells in MX1 poor clusters relative to MX1 rich clusters (p=0.015) (**Fig 6.8B-C**). Notably, there was no obvious enrichment for CD8+ T cells in MX1-rich regions, which is surprising given the previous association of these cells with interferon production in the kidneys.[50] Thus, dense aggregates of lymphocytes do not perfectly correlate with regions of high interferon response.

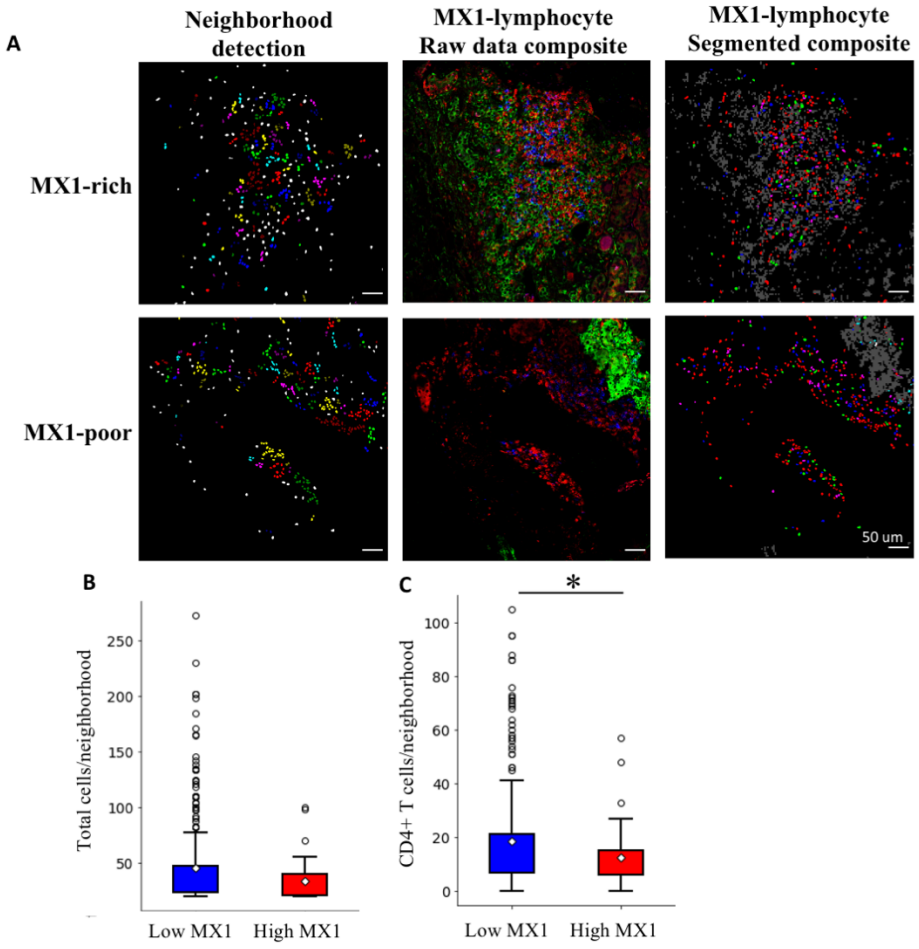

Figure 6.8 Figure caption continued on next page

Fig 6.8, continued. Relationship between MX1 and lymphocyte abundance. A) representative images showing lymphocyte abundance in MX1 rich and MX1 poor regions; Neighborhood detection: segmented cells that are the same color and spatially close have been assigned to the same neighborhood; Raw data composite: Shows distribution of MX1 (green), CD3 (red) and CD20 (blue) marker in microscopy image; Segmented composite: shows distribution of MX1 (gray) relative to segmented cells (green=DN, red=CD4+, blue=CD8+, magenta=CD20+,cyan=CD138+);Distribution of the B) total numbers of cells and C) CD4+ T cells in high MX1 (>=50% of pixels MX1+) and low MX1 neighborhoods (<50% of pixels MX1+)

Discussion

This work has allowed us to further characterize how inflammation is organized in the context of lupus nephritis. While the HR dataset gave us insight into the relationship between lymphocyte distribution and renal failure, the HMP dataset allowed us to delve more deeply into the phenotypes of these cells and examine their spatial features on the scale of a whole biopsy section, rather than a 1024x1024 pixel tile. This allowed us to validate some of the findings from the high resolution dataset around the relationship between high chronicity scores and lymphocyte abundance, though some discrepancies exist between the datasets.

This dataset validated previous observations[60,61,65] of double negative (CD4-CD8-) T cells in the biopsies of LuN patients. Follow up analysis of single cell RNA sequencing datasets and additional imaging revealed that a subset of these cells that are gamma delta T cells. We have also quantified regulatory T cells, T follicular helper cells, and exhausted CD8+ T cells. Interestingly, we have observed that T cells of a similar phenotype were often found to be colocalized. This suggests that specific regions of tissue might be enriched for particular cell populations.

It is difficult to relate these values to the single cell RNA sequencing data[14] that has been previously published as they do not provide frequency data, there seems to be a mismatch between these datasets, such that we are observing higher numbers of exhausted CD8+ T cells and fewer Tregs. This discrepancy could be because, by their own admission, their data is limited to the cells from which they obtained high quality data, and they are therefore not capturing all of the cells in their representative quantities.

We have defined cellular neighborhoods in tissue, and characterized them in terms of their size, constituency, and proximity to each other. This led to the observation that many dense cellular neighborhoods exist in close proximity to other cellular neighborhoods, while smaller neighborhoods are often more sparsely distributed. Amongst these neighborhoods are a subset of large CD20+ B cell, CD4+ T cell rich neighborhoods that are enriched for T follicular helper cells, and a subset of small neighborhoods that are enriched for CD8+ and DN T cells. Finally, we have made the surprising observation that dense neighborhoods of lymphocytes are not always found in regions of high MX1 expression. This suggests that there is variation within these dense neighborhoods such that some are type I interferon-rich, while others are relatively interferon poor.

All of these findings together demonstrate that there is substantial heterogeneity in how lymphocytes are distributed in inflamed tissue, and that highly multiplexed whole-section imaging enables us to probe this heterogeneity. We have developed useful tools for translating visually observable clustering behavior into quantifiable metrics. The next

step is to collect a dataset that has more patients with outcomes data so that we can use these metrics to identify predictors of renal failure.

Chapter 7: Developing generalizable networks—Case studies with the U-net architecture

The work described in this chapter was done in collaboration with several people -- Dr. Junting Ai and Margaret Veselits acquired the lupus data and assisted with pre-processing, and Jordan Fuhrman provided the initial code and advice about U-Nets

1. Introduction

We have used the Mask R-CNN[104] architecture throughout the body of this work, and it has been an effective method of segmenting cells in tissue. However, the Mask R-CNN architecture is very large and computationally expensive, taking on the scale of days to train. As we were developing the methods for the HMP dataset, it became clear that several networks would be necessary to cover all of the classes of interest, and training numerous instances of Mask R-CNN for multiple cellular classes seemed like an inefficient approach. Therefore, we sought out alternative methods of segmentation to accomplish this task with less overhead.

Towards this end, we chose the U-Net architecture, first described by Ronneberger et al.[98,99] Compared to the Mask R-CNN, the U-Net is much less complex, requiring less than an hour to train. The use of U-Nets for similar tasks is well-established. In particular, they have been used for nuclear segmentation of fluorescently labeled cultured cells to great effect[148,149]. In addition, they have been used to segment structures in hematoxylin & eosin-stained and fluorescently labeled slides of paraffin-embedded human tissue.[150–152] One important caveat is that the U-Net architecture is typically used for semantic segmentation tasks, while our need was for robust object-level segmentations. To bridge this gap, we used standard contour-finding algorithms to identify individual objects in the binary U-Net output.

As described previously, training a network for every single marker in our dataset was not feasible. Simply finding and generating ground truth from enough examples for every marker is not tractable, particularly for rare markers. I therefore spent some time developing a workflow in which we used U-Nets to perform semantic segmentation of T cells in human tissue on the basis of multiple markers (nuclear signal and CD3, and another T cell marker), and then defined discrete objects through mask contouring.

I evaluated the extent to which a network trained for a single cell type in one pathology and tissue context can generalize to different cell classes, and how the training set make up can affect generalizability. I also evaluated how well networks trained on one pathology could generalize to perform the same task in a different pathology (TNBC), and to what extent finetuning could be used to improve performance. Though we did not ultimately use the U-Net architecture, this work helped us develop the strategies that we ultimately implemented with the Mask-RCNN. This chapter is an adaptation of work that was submitted to two different conferences—Photonics West 2021, and Medical Imaging 2021.[153,154]

2. Methods

We specifically tested U-Net applicability on T cell markers from the HMP dataset. The data and ground truth was acquired as previously described in **Chapter 5**. The ground truth was converted to binary masks for use with the U-Net architecture. Image stacks were 512 pixels x 512 pixels x 3 channels, with those channels corresponding to CD3, T cell marker, and DAPI (**Figure 7.1A**). For examinations of the efficacy of fine-tuning, the later network layers were frozen, and the early layers were trained on two additional

datasets: 1) CD3/CD8/DAPI images of LuN biopsies, and 2) CD3/CD4/DAPI images of TNBC biopsies (**Figure 7.1B**).

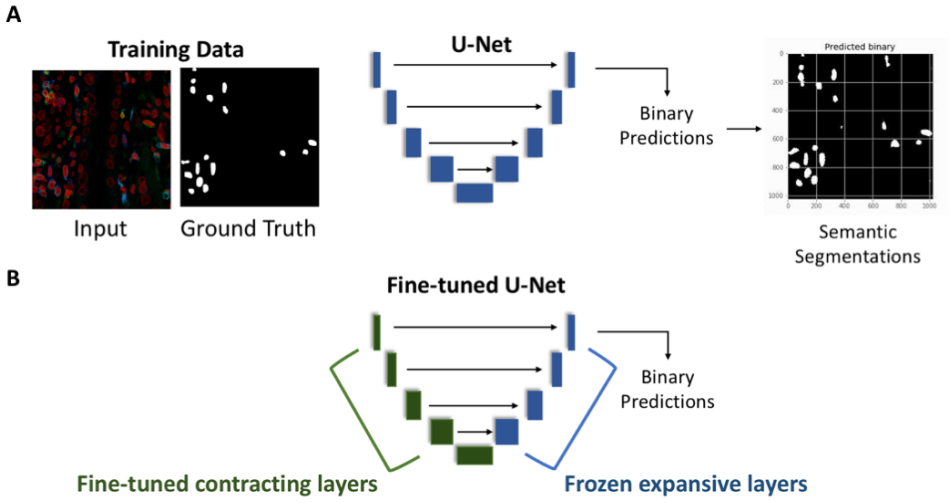

Figure 7.1. Schematic of U-Net training scheme. A) diagram of U-Net with examples of the training inputs and binary outputs; B) schematic of fine-tuning procedure in which contracting layers are re-trained while expansive layers are frozen

Networks were trained with 5-fold cross validation on a single GPU compute node. Object-level network performance was evaluated by predicting on 10-image test sets, performing region proposals on the binary masks (**Figure 7.2**), and then evaluating the object-level precision and recall. The values reported here are the average of the 5-fold networks.

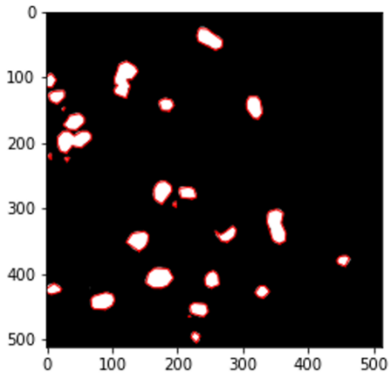

Figure 7.2. Representative region proposals on U-Net output. The skimage.measure function “find_contours” was applied to the binary output of the U-Net to define discrete objects.

3. Generalizing across multiple T cell markers – proto-T cell network

The first question we wanted to answer was how best to segment all of the T cell subsets that we were interested in. Specifically, would varying the content of the training sets make the network more generalizable to other markers, or would biasing towards very clear markers (such as CD4) lead to better overall performance?

To evaluate this, we tested three 90-image training sets (**Table 7.1**)—one that was completely made up of CD3/CD4/DAPI stacks, one that was split evenly between CD4 and CD8, and one that was split between CD4, CD8, and ICOS. Though all of these markers are surface markers for T cells, they do not have exactly the same staining pattern, with ICOS being a little more ambiguous than the other two.

Table 7.1 Description of the three 90-image training sets used to evaluate the effect of training set diversity

Network	Training set classes	# images per class
1	CD4	90
2	CD4/CD8	45
3	CD4/CD8/ICOS	30

Network (1) was trained only on CD3+CD4+ T cells. After several rounds of hyperparameter tuning, the best object-level performance metrics on a test set of 10 CD4 images were a precision of 0.855 and recall of 0.607, when averaged across the five folds of the network. **Figure 7.3** shows an example of the network output on CD3+CD4+ T cells in LuN

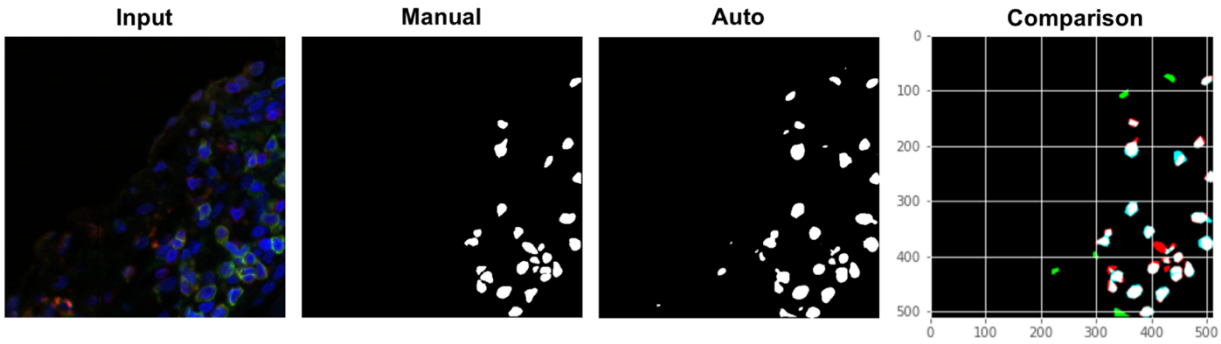

Figure 7.3. Representative CD3CD4 LuN segmentation. Input: 3-channel RGB from test set where R=CD3, G=CD4, B=DAPI; Manual: ground truth binary mask; Auto: binarized predictions with a threshold of 0.3; Comparison: white=true positives, green=false positives, red=false negatives, cyan= false positive near true positive

The precision and recall of this network on CD8 and ICOS stacks were substantially worse. In contrast, precision was improved for CD8 and ICOS by increasing the diversity of images in the dataset; precision for CD4 went down as the training set diversified, but by less than the improvement in the other classes (**Table 7.2**). This suggests that training on more general datasets might produce more generalizable networks that might be worth the tradeoff for performance on individual classes. This idea was ultimately carried forward in what became the Mask-RCNN T cell network.

Table 7.2 Network performance from the three training sets on test sets of CD4, CD8 and ICOS images

Training set	CD4	CD8	ICOS
	Precision		
CD4	0.855	0.689	0.539
CD4/CD8	0.83	0.715	0.587
CD4/CD8/ICOS	0.817	0.755	0.659
	Recall		
CD4	0.607	0.413	0.347
CD4/CD8	0.6	0.515	0.367
CD4/CD8/ICOS	0.573	0.468	0.363

4. Fine-tuning for generalizability

While diverse training sets is one way to improve network generalizability, fine-tuning is another approach one might take to make the most of the availability training data. Rather than one round of training that contains all of the classes that you are interested in, fine-tuning is a procedure by which a network is first trained on one task, and then re-trained to perform a similar task. The reason to use this approach is if the new task is sufficiently different from the original that a training set that contains both types of images would be too ambiguous for the network to be adequately trained.

For this work, I used the network trained only on CD4 (Network 1) as the “base model”. In order to test how well this base model generalized, it was used to segment test sets of CD3+CD8+ T cells from lupus nephritis, and CD3+CD4+ T cells from TNBC. In the case of CD4+ T cells in TNBC (CD3CD4 TNBC), the question was how well the network could generalize to a different tissue context, and a different set of fluorophores, but the same surface markers. The average performance on these test sets is described in **Table 7.3.**

Table 7.3 Network (1) performance on CD4 LuN, CD8 LuN and CD4 TNBC test

Test Set	Precision	Recall
CD3/CD4 LN	0.855	0.607
CD3/CD8 LN	0.689	0.413
CD3/CD4 TNBC	0.704	0.250

As expected, there was a decrease in performance on tasks that were different from the training set. In particular, the sensitivity suffered from a change in disease/tissue context. There are several factors that might account for differences in performance between disease states. One is differential tissue autofluorescence. **Figure 7.4** demonstrates how the same primary antibodies can produce very different signal to noise ratio, depending on the fluorophore used in the secondary antibody, and on the non-specific uptake of the antibody by the tissue.

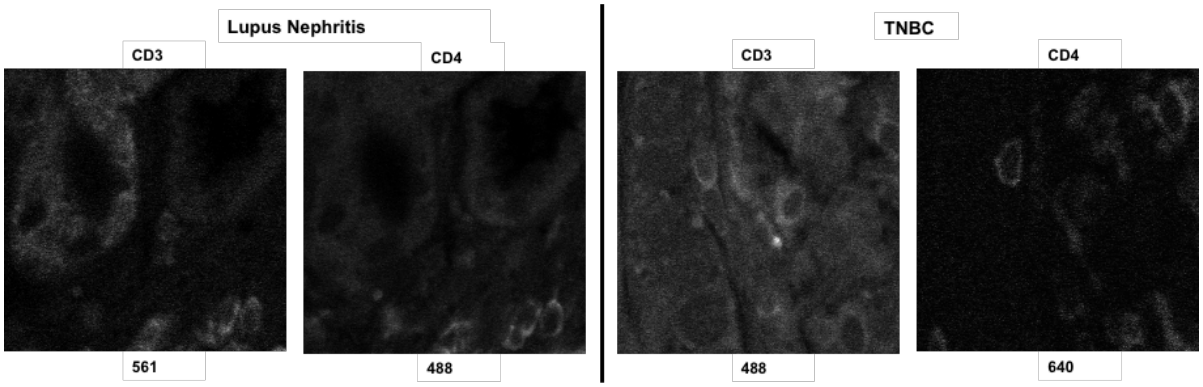

Figure 7.4. Tissue Autofluorescence in LuN vs. TNBC. Representative tissue autofluorescence at the utilized laser lines in the tissue types; the 488 laser line typically produces higher levels of autofluorescence, such that markers imaged on that laser line will have a lower signal-to-noise ratio.

With this as a baseline, I investigated the effect of fine-tuning on network performance. Network (1) was used as a base model, such that the expansive path was frozen, and the earlier contracting layers were re-trained with a set of 50 additional images from each

dataset (**Table 7.4**). This approach was chosen based on the findings of Amiri et al[155], which suggested that this was the ideal way to fine-tune the U-Net architecture. This moderately improved object level recall, while having negligible impact on precision (**Table 7.5**). The number of false positives decreased overall, and the networks were able to better separate adjacent cells. **Figures 7.5 and 7.6** demonstrate the difference in performance before and after fine-tuning.

Table 7.4 Training scheme for U-Net finetuning

Network	Trained on
1	CD3+CD4+ LuN, n=90
4	CD3+CD4+ LuN, n=90 Fine-tuned with CD3+CD8+ LuN, n=50
5	CD3+CD4+ LuN, n=90 Fine-tuned with CD3+CD4+ TNBC, n=50

Table 7.5 Network (4) and (5) performance metrics

Network	Test set	Precision	Recall
(4) Fine-tuned on CD8	CD3/CD8 LuN	0.7	0.545
(5) Fine-tuned on TNBC	CD3/CD4 TNBC	0.721	0.335

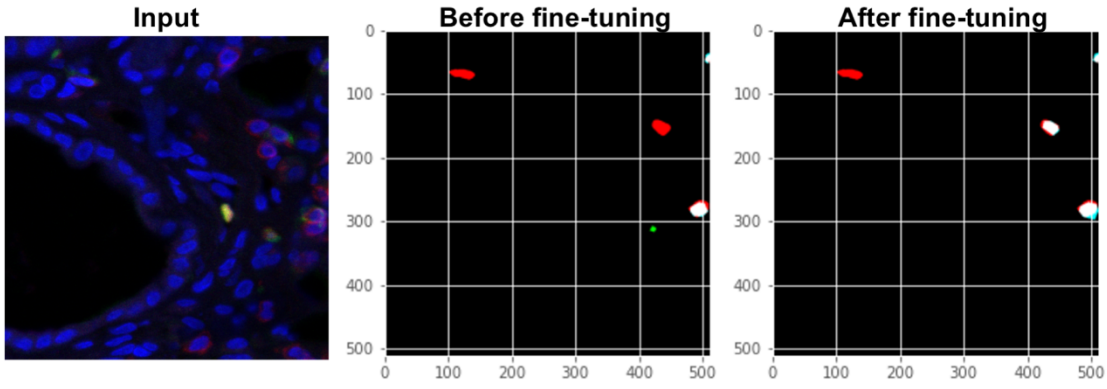

Figure 7.5. Representative CD3CD8 LN segmentation. Input: 3-channel RGB from test set where R=CD3, G=CD8, B=DAPI; Before fine-tuning: performance of Network 1 on CD8 test set image; After fine-tuning: performance of fine-tuned network on the same image; white=true positives, green=false positives, red=false negatives, cyan= false positive near true positive

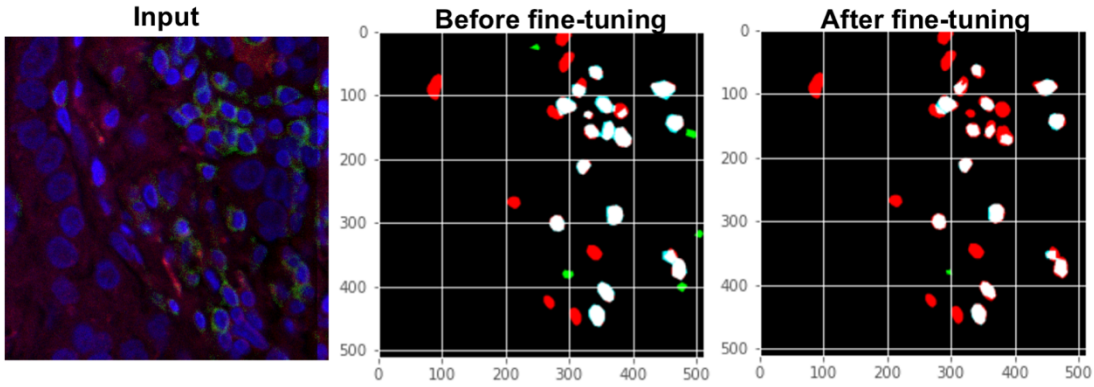

Figure 7.6. Representative CD3CD4 TNBC segmentation. Input: 3-channel RGB from test set where R=CD3, G=CD4, B=DAPI; Before fine-tuning: performance of Network 1 on TNBC test set image; After fine-tuning: performance of fine-tuned network on the same image; white=true positives, green=false positives, red=false negatives, cyan= false positive near true positive

5. Discussion

The first conclusion that we came to from these studies was that though U-Nets can be applied for similar tasks, segmenting specific cell classes on the basis of 3 different fluorescent markers is not an ideal use-case for them. In particular, this approach suffers from poor sensitivity, which is a problem for segmenting and accurately quantifying relatively rare cellular subsets. In examining the way in which the U-Net architecture fails at this task, there seem to be two major issues. First, cells that are touching are identified as a single object, which diminishes recall. Second, there are abundant false positives, which diminishes precision. The takeaway from this was the useful insight that the level of performance that the Mask R-CNN produces justifies how resource intensive it is, as the simpler, less expensive method was genuinely inferior.

However, the time spent investigating U-Nets for this purpose was not wasted. We were able to validate the approach of training a general 3-channel “T cell” classifier, which we ultimately used. Furthermore, we were able to observe that fine-tuning has the

potential to allow us to generalize across pathologies, which would be a powerful tool for future studies. It took a substantial amount of time and effort to develop the methods for segmenting the highly multiplexed lupus nephritis dataset, and we have barely started on the analogous TNBC dataset that was collected. Being able to use the tools developed in lupus as a starting point for the TNBC data opens up the possibility that we will be able to analyze that second dataset with much less additional effort. Developing tools and techniques to robustly and efficiently analyze new datasets, rather than starting from scratch with each experiment, dramatically increases the tractability of quantitative image analysis.

6. Acknowledgements

I would like to thank and acknowledge the following people who contributed to this part of the project: Dr. Junting Ai and Margaret Veselits assisted with data acquisition; Jordan Fuhrman provided me with code and advice for the U-Net analysis, Dr. Madeleine Durkee provided advice and input. Chun-Wai Chan administered the MEL computation server and provided troubleshooting help related to its usage.

Chapter 8: Quantifying the effects of biopsy fixation and staining panel design on automatic instance segmentation of immune cells in human lupus nephritis

The work described in this chapter was done in collaboration with several people -- Dr. Junting Ai acquired the data and Dr. Madeleine Durkee led the development of the automated cellular segmentation method.

1. Introduction

Method generalizability is an important consideration when developing machine learning methods for image analysis. As the preceding chapters highlight, several aspects of data acquisition can impact how easy it is to apply machine learning to a particular task. Because clinical samples are precious and data collection is resource intensive, it is vital to optimize data collection for the chosen analytical method. Specifically, it is important to understand how technical choices regarding sample preparation might influence the quality of data used in automated cell detection algorithms. Here I will describe work that addresses these types of questions, adapted from our paper "Quantifying the effects of biopsy fixation and staining panel design on automatic instance segmentation of immune cells in human lupus nephritis", which was published in the Journal of Biomedical Optics in January 2021.[123]

There are two major considerations that we chose to address. First, does the method of sample preparation influence the detection of cellular morphology? Previous work from our group[72] was performed on fresh-frozen samples. These are relatively expensive to store, and far less widely available than formalin-fixed, paraffin embedded (FFPE) tissue. Extending this technique to FFPE samples would greatly increase the dataset of samples that are available. However, it is well-established that formalin fixation can lead to gross tissue shrinkage[156–158], which could lead to distortions in any findings around cell shape,

size, and distance to other cell types. In this work, we sought to understand whether these deformations cause differences in cell shape and intercellular distances. Given no statistical difference between these two groups, it would be appropriate to group fresh frozen biopsies and FFPE biopsies for analysis of cellular features, which would increase the availability of datasets. For this reason, we evaluated performance of independently-trained Mask R-CNN networks in the task of multi-class instance segmentation of cells in FFPE tissue samples relative to fresh frozen.

Second, we wanted to address the number of stains that are required to identify a cell type. Our previous work utilized two markers to identify each type of dendritic cell. To a human observer, using multiple markers to identify a cellular class results in better discrimination of cell classes. In manual analysis of cells, this approach helps to identify true positives because tissue autofluorescence, stain quality, spectral bleed-through, and non-specific antibody binding can result in ambiguous signal. However, due to the limitation in the number of fluorophores that can be resolved in one imaging session, the choice to use multiple markers per cell type necessarily means that fewer cell types can be investigated in a given panel. This is essentially a tradeoff between robustness and breadth, which has been discussed at length in earlier chapters. Here we investigate whether single markers can be used in a computer vision task to identify and segment cell types with high fidelity, which would allow us expand the set of cell types we examine with a single imaging session.

For these two purposes, we collected 3 datasets from kidney biopsies of lupus nephritis patients: 1) fresh frozen, stained with two markers per antigen presenting cell (APC), 2) FFPE, stained with two markers per APC, and 3) FFPE, stained with one marker

per APC. It should be noted that this third dataset is the HR dataset that was analyzed in depth in **Chapters 2-3**. These findings will allow us to optimize future data collection efforts for the application of computer vision, which will enable rigorous quantification of immune cell subsets in tissue.

2. Methods

Data acquisition

For staining of fresh frozen sections, the sections were removed from -80°C, washed with PBS, blocked with serum and followed by antibody staining. Two distinct antibody panels were utilized to stain the tissue sections; for pDC analysis – CD3 (Alexa Flour 546), CD4 (Alexa Flour 594), BDCA2 (Alexa Flour 488), and CD123 (Alexa Flour 647); mDC analysis – CD3 (Alexa Flour488), BDCA1 (Alexa Flour546), CD4 (Alexa Flour594), and CD11c (Alexa Flour647). DAPI was used with the above to visualize tissue nuclei. Fresh frozen tonsil sections served as controls.

For staining of FFPE sections, the sections were de-paraffinized, treated with citric acid buffer (pH6.0) for antigen retrieval, then blocked and stained with the same process of fresh frozen samples. Double staining on FFPE was done with the same fluorophores as fresh frozen. Single staining was done with CD20 (Alexa Flour 488), CD3 (Alexa Flour546), BDCA2 (Alexa Flour594), CD4 (Alexa Flour647), and CD11c (Alexa Flour700). FFPE tonsil sections served as controls.

Lupus nephritis datasets

Three separate LuN datasets were used in this study to compare two tissue fixation methods (fresh frozen and FFPE) and two staining panels (**Table 8.1**). The first dataset (fresh frozen-DS) was composed of images of fresh frozen LuN biopsies, imaged on a

Leica SP5 laser scanning confocal microscope at 63x magnification. Resulting images were 1024 x 1024 pixels with a 0.1413 μm pixel size (**Table 8.1**). The samples in this dataset were stained with staining panel 1 (**Table 8.2**), using two markers per APC. As a result, a given sample was only stained for two T cell populations and one APC population, either myeloid dendritic cells (mDCs) or plasmacytoid dendritic cells (pDCs), and each image consisted of 3 cell classes: CD3+CD4+ T cells, CD3+CD4- T cells, and one type of dendritic cell (**Fig 8.1**).

Table 8.1 Defining descriptors of the three datasets used to assess DCNN performance on fixation methods and staining panels.

	Fixation method	Staining panel	Microscope	Pixel size (μm)
Fresh Frozen-DS	Fresh Frozen	Panel 1	Leica SP5	0.1337-0.1413
FFPE-DS	FFPE	Panel 1	Leica SP8	0.1058
FFPE-SS	FFPE	Panel 2	Leica SP8	0.1058

Table 8.2 Two staining panels were used to compare DCNN performance on single marker to dual-marker identification of APCs.

	T cell markers	mDC markers	pDC markers	B cell markers	Nuclear marker	Other
Panel 1 (DS)	CD3, CD4	CD11c, BDCA1	BDCA2, CD123	--	DAPI	DIC
Panel 2 (SS)	CD3, CD4	CD11c	BDCA2	CD20	DAPI	--

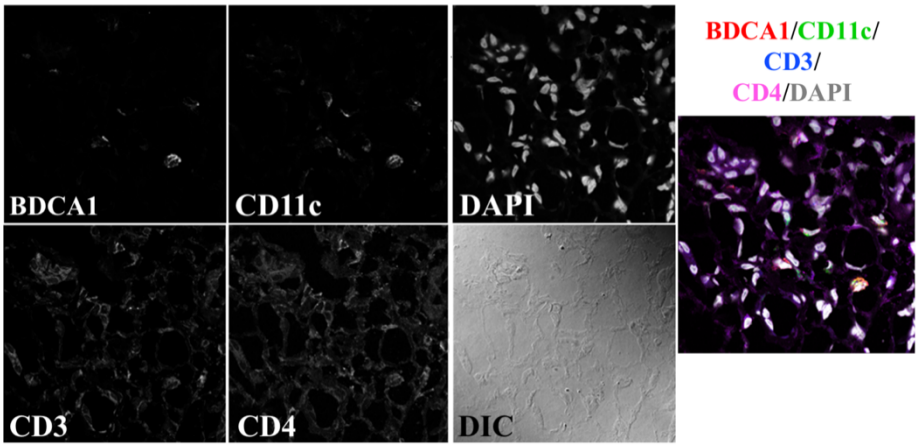

Figure 8.1. Example image from the fresh frozen-DS dataset. A single APC population (mDCs) is probed with two markers. This example shows a mDC image, but pDC images are also in this dataset, with the markers listed in Table 2. T cells are stained for CD3 and CD4, and DAPI is used to identify cell nuclei. All channels are merged in the rightmost panel, with colors corresponding to the above label.

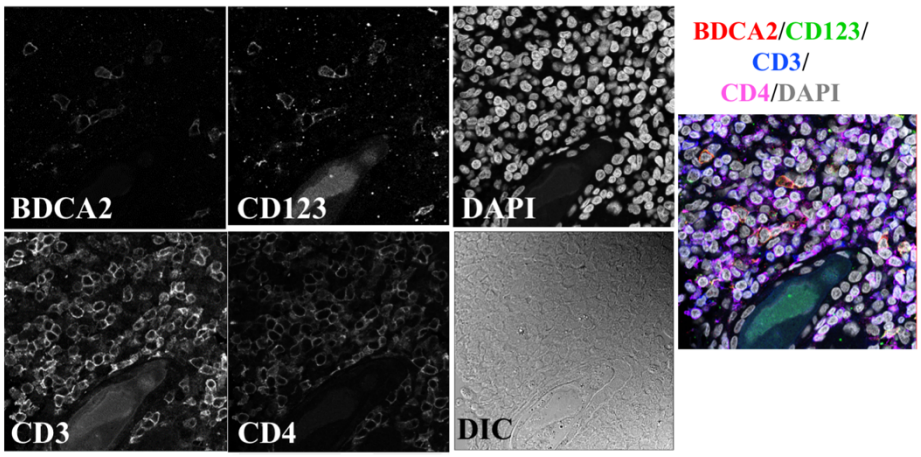

Figure 8.2. Example image from the FFPE-DS dataset. A single APC population (pDCs) is probed with two markers. This example shows a pDC image, but mDC images are also in this dataset, with the markers listed in Table 2. T cells are stained for CD3 and CD4, and DAPI is used to identify cell nuclei. All channels are merged in the rightmost panel, with colors corresponding to the above label.

FFPE-DS and FFPE-SS datasets were composed of images of formalin-fixed paraffin-embedded (FFPE) samples, imaged on a Leica SP8 laser scanning confocal microscope

at 63x magnification. The images remained 1024 x 1024 pixels; however, given the different imaging system, the resulting pixel size for these two datasets is 0.1058 μm. The FFPE-DS dataset was also stained with panel 1, with 3 cell classes per image (**Fig 8.2**). The FFPE-SS dataset was stained with a single marker per APC class. In addition to staining for both pDCs and mDCs in one panel, B cells were also probed in this dataset, resulting in 5 cell classes: 2 T cell populations and 3 APC populations (**Fig 8.3**). The FFPE-SS dataset lacks a differential interference contrast (DIC) channel in order to accommodate an additional cell surface marker while maintaining a constant channel depth. Conservation of channel depth was desirable for this study because keeping this variable consistent preserves the number of trainable parameters in the network. In the DS datasets, the DIC channel was intended to aid in the segmentation of cells, as it mainly contributes cell edge information. Preliminary analysis of the FFPE-SS dataset determined that the pixel-level segmentation was not adversely affected by eliminating this channel. Resulting image stacks were 1024 pixels x 1024 pixels x 6 channels, with each channel associated with a single marker. **Table 8.2** summarizes the key differences in the three datasets.

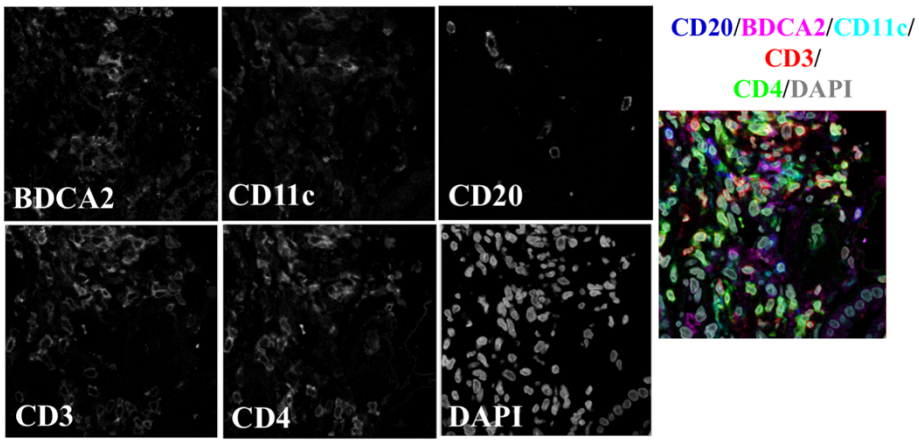

Figure 8.3 Figure caption continued on next page.

Figure 8.3, continued. Example image from the FFPE-SS dataset. Three APC populations (mDCs, pDCs, and B cells) are probed with a single marker each. T cells are stained for CD3 and CD4, and DAPI is used to identify cell nuclei. All channels are merged in the rightmost panel, with colors

Manual segmentation of images for ground truth

For all datasets, a subset of images from each biopsy was selected on which to generate manual truth. All manual segmentations and cell classifications were done in Fiji/ImageJ, as described in **Chapter 3**.

Generation of training sets

Manually segmented images were split into training, validation, and test sets at a 90/5/5 ratio (**Table 8.3**). Validation and test sets were small at the image level, but still contained over 300 cells each, and network performance is measured at the cell level. The FFPE-DS dataset had a smaller ground truth set due to the large number of cells per image in that dataset. The FFPE-SS manual dataset contained more images with a relatively high density of cells. The large number of manually segmented cells in the FFPE-SS manual truth set caused a 90% training set to exceed our GPU memory capabilities (4 nVidia K80 GPUs with 12 GB memory each). The ground truth for this dataset was therefore split into training/validation/testing sets at an 85/7.5/7.5 ratio. Images from a given biopsy were randomly divided up between the training/validation/test sets. This means that, while there were unique sets of images in the training/validation/test sets, images from the same biopsy could be in more than one of these subsets. This was done intentionally for this study to ensure that differences in performance between the separately trained instances of Mask R-CNN were due to the staining panel or fixation method, and not differences between patients in the training and testing sets.

Table 8.3 Training, validation, and test set splits for the manual segmentations in all datasets.

	Total cells in manual set	Total images in manual set	Images in training set	Images in validation set	Images in test set
Fresh Frozen-DS	5166	240	168 [90%]	12 [5%]	12 [5%]
FFPE-DS	7145	160	143 [90%]	8[5%]	8 [5%]
FFPE-SS	10611	342	293 [85%]	26 [7.5%]	26 [7.5%]

Network architecture and training

Three separate instances of a Mask R-CNN architecture[104] were trained to conduct instance segmentation on each of the three datasets, as described in **Chapter 3**. Hyperparameters were tuned to optimize accuracy on multiple class sets. The networks were trained with a learning rate of 0.01 using stochastic gradient descent with momentum. Cells in dense regions were detected with higher accuracy by reducing hyperparameter of the RPN section of the network, anchor stride length. Training was stopped once the mean average recall for all cell classes stopped increasing. Cells that were predicted with at least 0.3 probability were kept for analysis. All of this work was done using the Midway2 compute nodes at the University of Chicago Research Computing Center.

Evaluation of network performance

For this study Mask R-CNN performance was measured by calculating sensitivity, specificity, and Jaccard index, also known as intersection over union (IOU), for a test set. The manual segmentations provided ground truth at the cell level. A cell prediction was

considered a true positive if it had an IOU of at least 0.25 with a manual segmentation of a cell of the same class. Sensitivity and specificity for cell detection and classification were calculated at the cell class level. In addition, IOUs were calculated on a per cell basis and averaged across all cells within a given cell class. Sensitivity, specificity, and IOU were averaged across all cells to provide overall performance metrics for the networks.

Cell shape and distance metrics

After analyzing the detection and segmentation performance of each network on the corresponding test sets, each network was used to predict cell types in larger sets of unlabeled images. Population ratios of each cell type were calculated for each unlabeled dataset and compared to the corresponding ratios in the ground truth dataset. After analyzing the performance of the three trained networks, each network was used to predict cell types in unlabeled images. Cell size, shape and distance features, specifically cell area, cell perimeter, and T cell minimum distance to an DC were calculated for each cell detected by the networks. These shape features were compared across datasets to determine whether the tissue preparation method, stain specificity, or network performance affected cellular features.

3. Network performance on test sets

Mask R-CNN architectures were trained for each of the three separate datasets. Both instances of Mask R-CNN trained on the fresh frozen-DS and FFPE-DS datasets met the stopping criteria at 16k epochs. Performance metrics on test sets for these two datasets are detailed in **Table 8.4.** It is important to note that while DC sensitivity is high, we do not necessarily expect to detect every DC in an unlabeled dataset. These test sets are

relatively small at the image level, and DCs are the least prevalent populations, but there are still over 50 DCs in each test set. Each network trained on a double-stain dataset detected all DCs in the corresponding test set, regardless of sample fixation method.

Table 8.4 A network was trained and tested on each dataset as described in Tables 1-3. Sensitivity, specificity, and Jaccard index (IOU) are shown for the test sets corresponding to the two networks trained on the double-stain datasets.

	CD3+CD4+ T cells	CD3+CD4- T cells	DCs	All (avg)
Sensitivity				
Fresh Frozen-DS	0.77	0.85	1.0	**0.87**
FFPE-DS	0.89	0.84	1.0	**0.91**
Specificity				
Fresh Frozen-DS	0.82	0.84	0.80	**0.82**
FFPE-DS	0.84	0.83	0.96	**0.88**
IOU				
Fresh Frozen-DS	0.79 ± 0.21	0.75 ± 0.24	0.83 ± 0.19	**0.80 ± 0.21**
FFPE-DS	0.77 ± 0.22	0.80 ± 0.19	0.86 ± 0.15	**0.79 ± 0.20**

The instance of Mask R-CNN trained on the FFPE-SS dataset required longer training time (18k epochs), and network sensitivity was poor for mDCs and marginal for pDCs (**Table 8.5**). The poor performance on DCs may be due to the fact that they are more amorphous than lymphocytes like T cells and B cells, which have relatively little cytoplasm and therefore have surface stains that coincide with their nuclei. In contrast, dendritic

cells have long extensions from their cell bodies called dendrites[159], which can reach in and out of the image plane, producing positive signal where there may not be a nucleus to assign it to. Therefore, assigning ground truth to these cells is inherently harder. In the fresh frozen-DS and FFPE-DS datasets, DCs are identified with two markers, while in the FFPE-SS dataset, each DC population is identified with a single marker. With this dataset we tested the hypothesis that using multiple stains to identify DCs bolsters performance, and that using only one marker would impose a cost. The decline of network performance on these cells is likely due to a combination of low signal to noise ratio, variable cell shape, and ambiguous ground truth.

Table 8.5 A network was trained on the FFPE-SS dataset. Sensitivity, specificity, and Jaccard index (IOU) are shown for the FFPE-SS test set.

	CD3+CD4+ T cells	CD3+CD4- T cells	mDCs	pDCs	B cells	All cells (avg)
Sensitivity						
FFPE-SS	0.90	0.85	0.38	0.69	0.75	**0.72**
Specificity						
FFPE-SS	0.86	0.89	0.97	0.95	0.91	**0.92**
IOU						
FFPE-SS	0.81 ± 0.17	0.82 ± 0.18	0.63 ± 0.21	0.74 ± 0.20	0.75 ± 0.21	**0.78 ± 0.19**

The three trained networks described above were used to generate cell predictions on larger unlabeled datasets. **Table 8.6** describes the manual and automatic segmentations for each of the three datasets. Each trained instance of Mask R-CNN was used to generate cell predictions on all images in its corresponding dataset, which included unlabeled versions of all images that had been manually segmented *and* images that

were never manually segmented by an expert. The average number of cells per image is similar between the manual segmentations and automatic predictions for each dataset. Assuming patients in these larger datasets have similar prevalence of each cell type, the manual and automatic segmentation sets should maintain similar ratios of cell types across cell types. Cell types with lower sensitivity values in tables 4-5 are expected to have lower prevalence in the automatic sets compared to the manual counterparts, while cell types with lower specificity are expected to have an increased prevalence in the automatic sets. Absolute numbers and relative amounts of each cell type are listed in **Table 8.6** for both manual segmentations and automatic predictions for all three datasets.

Table 8.6 Cell counts for manual segmentations and automatic predictions in all datasets.

Manual	**Total cells [images]**	**Average cells/ image**	**CD3+CD4 + T cells [%]**	**CD3+CD 4- T cells [%]**	**mDCs [%]**	**pDCs [%]**	**B cells [%]**
Fresh Frozen-DS	5166 [240]	21.5	2688 [52.03]	1161 [22.48]	292 [5.65]	1025 [19.84]	N/A
FFPE-DS	7145 [160]	44.7	4104 [57.44]	2041 [28.57]	483 [6.76]	517 [7.23]	N/A
FFPE-SS	10611 [342]	31.0	3714 [35.00]	2846 [26.82]	768 [7.24]	847 [7.98]	2436 [22.96]
Automatic							
Fresh Frozen-DS	16666 [673]	24.8	8216 [49.30]	4047 [24.28]	2160 [12.96]	2243 [13.46]	N/A
FFPE-DS	16396 [380]	43.1	8351 [50.93]	5340 [32.57]	1186 [7.23]	1519 [9.27]	N/A

Table 8.6 Cell counts for manual segmentations and automatic predictions in all datasets, continued

FFPE-SS	38594 [1332]	29.0	11126 [28.82]	14962 [38.76]	2573 [6.66]	2436 [6.31]	7506 [19.45]

4. Fixation method affects cell shape and network performance

It is widely documented that the processes of fresh freezing and formalin fixation cause different deformations to tissue. Formalin fixation will dehydrate the tissue, causing a contraction[156–158]. **Fig 8.4A-C** shows that this phenomenon is consistent across all cell types. T cells and mDCs show a markedly reduced area in FFPE samples compared to their fresh frozen counterparts (**Fig 8.4A-B**). However, while pDCs are also much smaller in FFPE than fresh frozen samples, the change in area is less than that of T cells and mDCs (**Fig 8.4C**). In FFPE samples, pDCs showed a 31.4% reduction in mean area compared to 54.8% and 55.5% reductions in the mean area of mDCs and T cells, respectively. Similarly, a contraction of cellular perimeter was observed for all classes (**Fig 8.4D-F**). This shrinkage is not only found at the cellular scale but remains consistent at the tissue level. **Fig 8.4G** shows the distribution of minimum distances of T cells to the nearest DC. T cells in FFPE samples show shorter distances to DCs than in fresh frozen samples ($p <<< 0.0001$). The fixation method therefore influences not only measurements of cell size and shape, but of spatial relationships between cells. Both networks exhibited high confidence in the classifications, as measured by the distribution of probabilities assigned by the network, with the FFPE-DS network showing increased

prediction probabilities relative to the fresh frozen-DS network (**Fig 8.4H**).

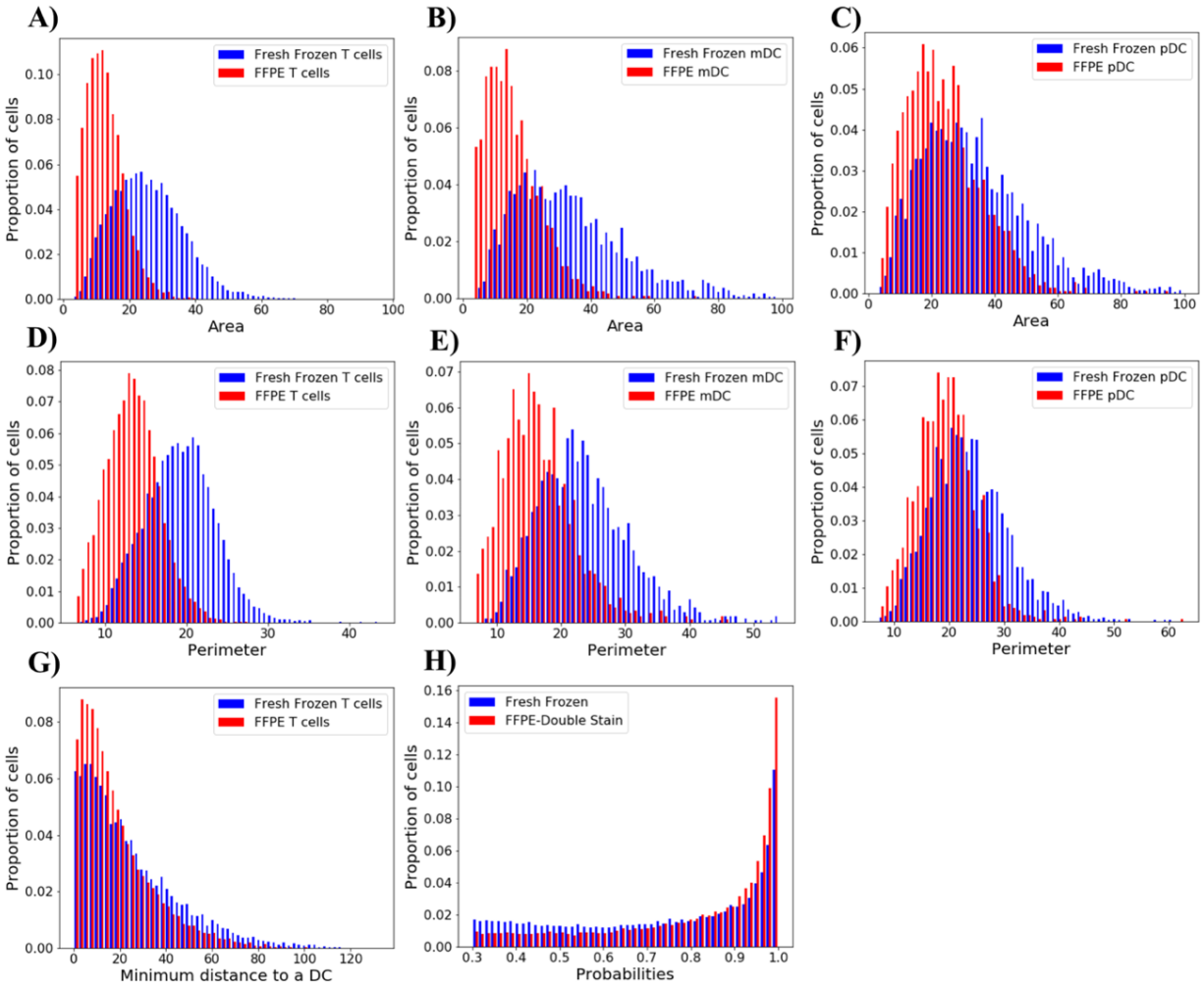

Figure 8.4. Various distributions of area, perimeter, minimum distance to DC, and automatic prediction for the fresh frozen and FFPE fixations. Shape and distance differences exist between cells of the same population when different fixation methods are used. A-C) Areas of T cells (A), mDCs (B), and pDCs (C) are significantly smaller in FFPE samples than fresh frozen samples. D-F) Likewise, perimeters of T cells (D), mDCs (E), and pDCs (F) are significantly smaller in FFPE samples than fresh frozen samples. G) Minimum distances between T cells and the nearest DC are significantly smaller in FFPE than fresh frozen samples. H) Both networks show high confidence in the automatic predictions, although the FFPE probabilities are significantly higher. For all plots, a Kolmogorov-Smirnov test shows a statistical difference between the fresh frozen and FFPE distributions ($p <<< 0.0001$).

5. Effect of staining panel on cellular quantification

Separate staining panels were used on the two FFPE datasets to test the feasibility of using a single marker to identify APC populations. This would allow us to analyze a more diverse set of cells in a given biopsy, overcoming the technical limitations of antibody species and available microscope laser lines. For example, the FFPE-DS dataset can probe a single APC population—either mDCs or pDCs—in a given image, while the FFPE-SS dataset probes three APC populations—B cells, mDCs, and pDCs—in single image. The use of the single-stain system compared to the double-stain system diminished the accuracy of the network for DC populations (**Fig 8.5**). Compared to the network trained on a panel with double-stained DCs, the network trained on the panel with single-stained DCs yielded worse confidence overall in cell detection and classification, as shown by the distribution of probability scores for the DC classes (**Fig 8.5A-C**). This is consistent across all cell types, but particularly noticeable in mDCs (**Fig 8.5C**), which corresponds with the poor sensitivity to mDCs with the network trained on the FFPE-SS dataset (**Table 8.5**). Furthermore, neither mDC nor pDC area remains consistent (**Fig 8.5E-F**), suggesting that the decrease in sensitivity to these cells skews the distribution of cell features.

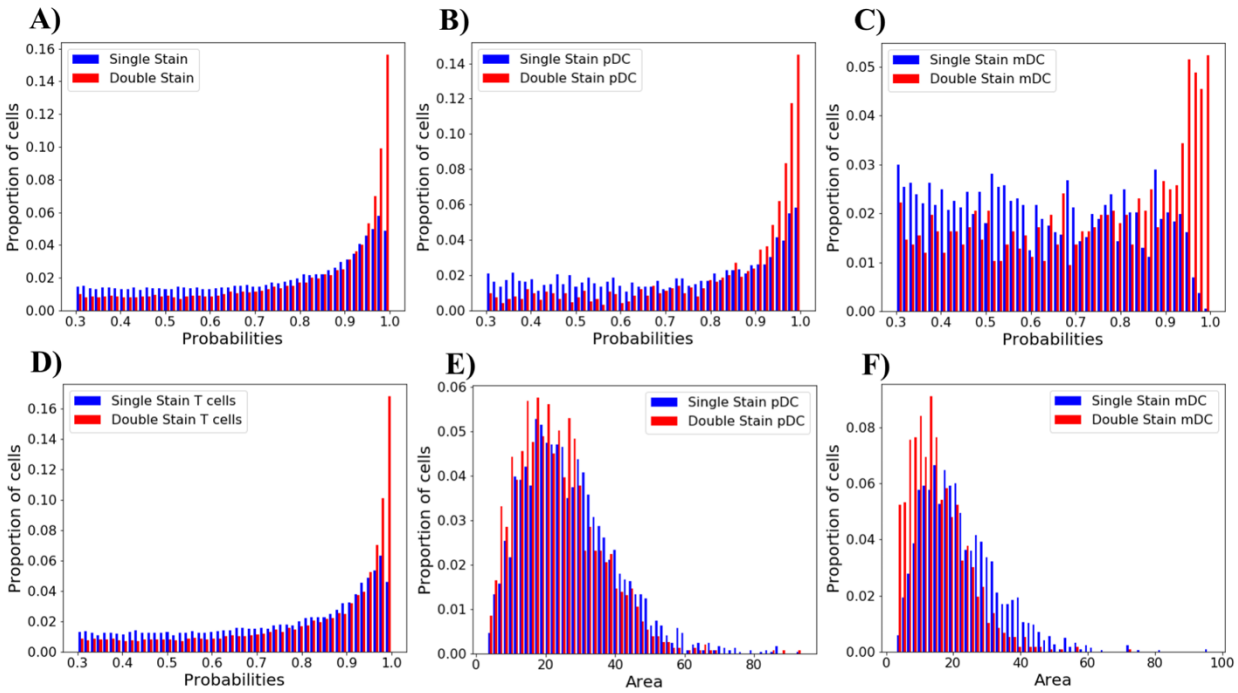

Figure 8.5. Number of stains used to probe a DC population affects the network performance. A) Network confidence in cell classification for all cells is compared between a network trained on a single stain DC panel and a network trained on a double stain DC panel. The network trained on the double stain panel was statistically more confident in its predictions ($p <<< 0.0001$). B) Probabilities of cells classified as pDCs by networks trained on single and double stain DC panels. C) Probabilities of cells classified as mDCs by networks trained on single and double stain DC panels. D) Probabilities of cells classified as either T cell population by networks trained on single and double stain DC panels. A-D have a lower bound of 0.3 because cells below this threshold are automatically rejected by the network. E) Cell area of pDCs detected by networks trained on single stain and double stain panels. F) Cell area of mDCs detected by networks trained on single stain and double stain panels. For all plots in this figure, a Kolmogorov-Smirnov test shows a statistically significant difference between the two distributions ($p <<< 0.0001$).

6. Discussion

The purpose of this work was to investigate how various aspects of data acquisition could influence downstream computer vision applications. In particular, we were interested in how sample preparation and staining panel design influence instance segmentation of cells in tissue.

First we demonstrated that tissue fixation impacts the metrics of cell shape and intracellular distances that can be derived from the network predictions. Tissue expansion and shrinkage in fresh frozen and FFPE tissue, respectively, is well-documented[156–158]. The data presented here quantify these deformations, showing a ~30% decrease in all linear metrics of T cell shape (e.g. equivalent diameter and perimeter), a 52.7% decrease in mean T cell area, and a 24.7% decrease in the minimum distance of a T cell to the nearest DC. These discrepancies in cellular features can have implications for conclusions drawn from data mining images to investigate biological phenomena. Previous from our group has used cellular shape and distance between cell types in fresh frozen LuN biopsies to identify intercellular interactions[72]. Metrics including minimum distance of T cells to a DC and T cell shape features were used to identify which cell populations were more frequently interacting. For findings from fresh frozen data to translate effectively to FFPE LuN biopsies, these differences in cell size, cell shape, and intercellular distances must be taken into consideration.

The second major technical consideration we investigated is the utility of using multiple markers for classifying cells, particularly for difficult classes like dendritic cells. Because a given immunofluorescence experiment is limited to 5-6 markers, there is a real cost associated with using multiple markers per cell type. In panel design there is a tradeoff between robustly identifying a single cell type and interrogating multiple cell types in a single experiment. We evaluated the extent to which using a single stain to identify DC subsets diminished network performance. We observed that the network sensitivity was relatively poor for the single stain dataset, particularly for mDCs and pDCs. DC subsets were particularly impacted by ambiguous staining from single markers,

compounded by relatively low prevalence of these cell types in the dataset. This loss of sensitivity had consequences for calculating cell features downstream, as evidenced by the shift in the observed area distribution for pDCs and mDCs (**Fig 8.5E-F**).

Thus we concluded that using a single marker for detecting difficult or infrequent cell types is not a worthwhile compromise, because the benefit of interrogating multiple cell types is negated by the decrease of algorithm robustness in detecting these infrequent cell types. Using multiple markers bolsters the performance of computer detection of cells, particularly for cell classes of lower prevalence. Though this is the last set of results that I will discuss, this work was done prior to the work described in **Chapters 3-6.** Ultimately, we adopted the approach of training two separate networks for lymphocytes and dendritic cells in order to resolve the poor performance observed in the 5-class instance of Mask R-CNN.

These findings highlight the importance of optimizing image acquisition for computer vision. As using machine learning to quantify multiplexed immunofluorescence imaging becomes increasingly mainstream, it is important to have the analytical method in mind when making choices about tissue preparation, staining panel design and acquisition parameters.

7. Acknowledgements

I would like to thank and acknowledge the following people who contributed to this part of the project: Dr. Junting Ai from the Clark lab acquired the microscopy data and helped with ground truth generation. Margaret Veselits assisted with ground truth generation. Dr. Madeleine Durkee was responsible for training and optimizing the performance of the Mask R-CNN on this dataset, and for conceiving of much of the

analysis. The Integrated Light Microscopy Core Facility at the University of Chicago (Dr. Christine Labno, Shirley Bond and Dr. Vytas Bindokas) provided expertise on imaging. The Research Computing Center at the University of Chicago provided the computational resources.

Chapter 9: Discussion

1. Introduction

In many respects, this work has two parallel "stories" associated with it. The first is a set of findings around lupus nephritis that challenge the standard model of pathogenesis and identify a novel correlate of patient outcomes. The second is a set of technical developments in the rapidly expanding field of quantitative image analysis, which has the potential to overcome previous barriers to understanding human immunity. Here I will discuss the broader implications of both of these stories, including the future directions that arise from them.

2. Insights about lupus nephritis

Organized inflammatory states in lupus nephritis are predictive of outcome

The fundamental question that we were trying to address was whether or not inflammatory states in lupus nephritis are organized. The "null hypothesis" of inflammation is that it is random, and that cells are non-specifically drawn into damaged tissue. However, our observations suggest that this is not the case. The manner in which the components of adaptive immunity (T cells, B cells, antigen presenting cells) were arranged in the kidneys are correlated with patient outcomes, generating the hypothesis that some inflammatory states drive renal failure and some do not. Throughout the course of this work and across two distinct datasets we have interrogated inflammation on multiple scales.

We first performed a relatively simple analysis in which we examined the density of CD4+ T cells, CD4- T cells, B cells, mDCs and pDCs in the tissue. This led to the striking observation that B cells are more densely packed in patients who do not progress to

kidney failure, and that CD4- T cells are present in higher densities in patients who do progress.

We then looked at the next level of complexity and evaluated the nearest neighbors and close cellular niches of each cell. It is true that we cannot make definitive statements about whether these cells are interacting with each other. However, many intercellular interactions require proximity, making a cell's nearest neighbors their most likely contacts. Indeed, when we examined the nearest neighbor relationships in both datasets, several interesting findings emerged.

First, we identified different patterns of cellular organization in ESRD+ and ESRD- patients. Cells were more likely to have a CD4- T cell as their nearest neighbor or in their immediate 10 um niche in the context of ESRD, while they were more likely to have B cells in close proximity in the context of failure to progress to ESRD. Second, we observed in both datasets that there was a strong propensity for cells of the same type to cluster together in tissue. In the highly multiplexed LuN dataset T cells with a shared phenotype were often each other's nearest neighbors, suggesting that there was regional enrichment for T cells of a particular type. This is consistent with a model in which local environment (i.e., cytokine milieu and APC presence) might drive the T cells in a region towards a particular phenotype.

Taking one step up in organizational structure, we then defined neighborhoods of cells. Doing this in the HR dataset had limitations, as we were only able to define groups of cells that were captured in the same 1024x1024 ROI. This prevented us from analyzing larger scales of neighborhoods. Nonetheless, we observed an enrichment for CD4- T cell-dominated neighborhoods in patients who progressed to kidney failure, relative to the

patients who did not. Identifying neighborhoods over the whole biopsy in the HMP dataset enabled us to capture even larger neighborhoods with a higher degree of phenotypic detail.

Finally, in the HMP dataset we were able to evaluate higher levels of organization across the whole biopsy section and observe that cellular neighborhoods (in particular large neighborhoods) tended to group with other cellular neighborhoods, in structures that that one might think of as “cities”. In contrast, some small neighborhoods were found to be isolated from other immune cells, suggesting that there is a tendency for both dense and sparse regions of inflammation rather than an even distribution of cells throughout the tissue.

These findings demonstrate that analyzing the spatial features of inflammation can lead to rich insights about disease and help generate new hypotheses about the determinants of patient outcomes. In particular, we now have reason to re-think some of our assumptions about what aspects of adaptive immunity drive renal failure in lupus nephritis.

Reevaluating the role of B cells in lupus nephritis

When we relate the spatial density of CD20+ B cell rich zones with the renal outcome of patients we find that a higher density of B cells is associated with the patients who did not progress to ESRD during the follow-up period. This is a striking result as our current model of lupus nephritis pathogenesis as an antibody-mediated disease leads to the prediction that dense regions of B cells would be associated with increased antibody production, and therefore worse disease outcomes. There are a couple of possible models that could account for this apparent paradox.

The first model is that currently existing therapies adequately target B cells, such that patients whose pathology is clearly B cell mediated respond well to therapy. This is the simplest explanation, as it does not require us to entirely re-think the role of B cells in driving lupus nephritis. It has been observed that mycophenolate mofetil and cyclophosphamide have the effect of depleting circulating plasma cells in lupus patients.[33] Therapies like rituximab and belimumab also deplete B cells, and successfully treat disease in a subset of lupus nephritis patients.[2,27–29] BAFF expression has been found to be higher in patients who respond to induction therapy, which lends weight to the hypothesis that the patients whose disease is characterized by abundant B cell proliferation are being successfully treated.[35]

However, lupus nephritis is heterogeneous in both its clinical presentation, and in how inflammation occurs in tissue.[6,10] In particular, it has been observed that, though glomerular inflammation often correlates with circulating auto-antibodies, tubulointerstitial inflammation does not.[10] Therefore, though B cells are observed in the tubulointerstitium, it is likely that tubulointerstitial inflammation is driven by mechanisms that are distinct from glomerular inflammation.[9]

Therefore, it is reasonable to postulate that there is a subset of patients who have disease that is driven by additional mediators (such as T cells) in a manner that is at least partially B cell independent. Indeed, our findings around “CD4-“ T cells suggest that there might be some T cell-driven pathology that we are not currently targeting. It is possible that we are simply stratifying between the patients for whom the standard of care is an effective therapeutic approach, and the patients for whom B cell depletion is not sufficient to stop chronic kidney damage.

The best way to evaluate this hypothesis would be to examine repeat kidney biopsies after patients have received treatment. If the patients who respond well to therapy start with abundant B cell-rich regions that dissipate after induction therapy, this would be an argument in favor of the model in which primarily B cell-driven disease is adequately addressed through the standard of care. Currently, repeat kidney biopsies are infrequently used in a clinical setting, though some have argued that this should change, as they are a rich potential data source.[1]

An alternative model is that enrichment for CD20+ B cells indicates a failure to differentiate into plasma cells. It should be noted that CD20+ B cells are not likely to be the subset that would be actively producing antibodies, as B cells express this marker before they develop into plasma cells.[160] It is possible that dense regions of CD20+ B cells accumulate if there is some factor preventing them from maturing into plasma cells, corresponding with a decrease in local antibody production. This bottleneck could be due to a lack of adequate T cell help, or an immunosuppressive environment.[161]

In the highly multiplexed dataset we observed that the density of CD20+ B cells and plasma cells are highly correlated with each other, which makes this model less appealing. However, this was a small cohort of patients for whom we did not have outcomes data. A more focused study evaluating the relationship between plasma cell: CD20+ B cell ratio in a group of patients for whom we have outcomes data would clarify whether an enrichment for CD20+ B cells in patients who have preserved renal function is a reflection of lack of local plasma cell differentiation.

In summary, it is necessary to re-examine the fundamental assumption that B cells are the primary driver of disease in all lupus nephritis patients. The observation that dense

aggregates of B cells are not necessarily associated with kidney failure needs to be confirmed in a validation cohort of patients. Ideally, this would be done in a prospective manner, such that we could biopsy patients at the time of diagnosis, and then track the abundance and distribution of B cells in the kidneys over the follow-up period and relate this to renal function.

"CD4- T cells" are a diverse array of cells that might play a role in driving kidney failure

The second surprising finding from the high resolution data was that CD4- T cells appear to be present at a higher density in patients who do progress to kidney failure, as lupus nephritis is not typically thought of us a T cell-driven disease. However, this premise has been recently called into question by the success of voclosporin, a calcineurin inhibitor that inhibits T cell effector function, in a lupus nephritis trial.[68] Thus, there is some evidence that the role of T cells in driving this disease needs to be scrutinized further.

It was particularly interesting to observe the increased density of this population in the patients who were imminently in kidney failure (ESRD current) at the time of biopsy relative even to other patients who proceeded to failure later on, as it more tightly links the timing of this inflammatory profile with renal failure. Interestingly, one of these five patients presented with acute kidney failure without having received prior therapy, and their kidney biopsy was scored with a tubulointerstitial chronicity index of 2. For this patient the increase in CD4- T cells was not associated with pre-existing fibrosis, nor was fibrosis a pre-requisite for renal failure. Though this cohort was small, they provided compelling impetus to further investigate the link between this cell subset and acute kidney failure in lupus nephritis patients. It would be very interesting to specifically look at the biopsies of a larger cohort of patients who present to clinic in already in kidney failure,

and profile the CD4- T cell compartment in order to more definitively characterize the subset associated with a loss of renal function.

We have already begun the work of characterizing the “CD4-“ T cells in lupus nephritis, starting with the simple question of whether or not they are CD8+ T cells. Indeed, when we investigated this in the highly multiplexed data, we found that there was a distinctive population of CD3+CD4-CD8- (double negative, DN) T cells that made up roughly half of the CD4- T cell compartment. This is coherent with previous observations of DN T cells in the context of lupus nephritis, which have been proposed to be the result of CD8 downregulation on canonical T cells.[60,61] However, this classification is problematic as defining these cells by a negative depends on detecting CD4 and CD8 with high sensitivity, and leaves open the possibility that the population is very heterogeneous. When investigating the possible identities of the CD3+CD4-CD8- T cells, there are two additional possibilities to evaluate—Natural Killer T (NKT) cells, and gamma delta T cells. There is a lack of evidence that NKT cells have a presence in the kidneys of LuN patients or play a significant role in pathogenesis.[14] On the other hand, gamma delta T cells have been observed by both us and other groups in the context of lupus nephritis, though their function is not obvious.[64,65] Given our observations that the CD4- T cell compartment appears to be quite diverse, additional imaging experiments need to be done to determine which of these cellular subsets is associated with renal failure.

The other question that was of interest was the degree to which CD8+ T cells observed in lupus nephritis are exhausted. There is conflicting data on this point, with several studies identifying exhausted CD8+ T cells in the circulation of LuN patients, but with the single cell RNA sequencing data that suggests that very few CD8+ T cells found in the

kidneys of lupus nephritis patients are exhausted.[14,57,58,162] In contrast, our observations in the highly multiplexed data suggest that a high percentage (~30%) of CD8+ T cells express PD1, which suggests that there is a substantial population of exhausted CD8+ T cells. This difference is possibly attributable to differences in the dynamics of transcripts vs proteins. Notably, we also found that these exhausted CD8+ T cells tended to be found near other exhausted CD8 T cells, suggesting that there may be a regional enrichment for an exhaustion phenotype.

CD4+ T cells play several roles in lupus nephritis

CD4+ T cells are amongst the most common lymphocytes found in the biopsies of lupus nephritis patients. When we evaluated whether or not they were differentially distributed between patients who progressed to kidney failure and those that did not, we did not note any significant differences. However, it would not be appropriate to conclude that CD4+ T cells are unimportant in this disease. Rather, this apparent lack of differentiation could be driven by the lack of phenotypic resolution in the CD4+ T cell compartment. CD4+ T cells can be playing a multitude of roles in inflamed tissue, from driving B cells responses, to suppressing inflammation, to creating an “anti-viral”-like Th1 response.[163] Therefore, looking at all CD4+ T cells together is of limited utility.

For this reason we used the HMP data to more finely characterize the CD4+ T cell compartment in inflamed kidney biopsies in terms of PD1, ICOS, and FoxP3 expression. While these markers still do not capture the full range of phenotypic variability within CD4+ T cells, they allowed us to specifically interrogate regulatory T cells and T follicular helper cells.

We found that regulatory CD4+ T cells made up a very small fraction of the CD4+ T cells and were infrequent in all patients. This coheres with other observations that the Treg compartment is deficient in SLE patients.[45–47] This leads to the question of whether Tregs are important for inhibiting progression to renal failure, or if there are other regulatory mechanisms in place.

We also validated the finding that there are dense B-T aggregates that are enriched for Tfh-like CD4+ T cells. However, these B-T aggregates did not typically resemble germinal centers in terms of the frequency with which putative Tfh cells were proximal with B cells. In fact, Tfh cells were frequently found near other Tfh, and other CD4+ T cells that we were not able to characterize fully. Understanding the significance of interactions between CD4+ T cells in lupus nephritis pathogenesis is therefore of great interest for future work.

DC subsets have variable contributions to renal outcome

To our surprise, pDC abundance and distribution turned out to be the same between patients who progressed to renal failure and those who did not. This implies that, though they are present at aberrantly high levels in lupus nephritis kidneys, they are not the primary drivers of tissue damage. They are primarily appreciated for their roles as interferon-producing cells[75–77], but they have also been postulated to present antigen to and activate Tregs[164]. Therefore, they could have complex and contradictory roles in inflamed tissue.

Intriguingly, it was observed that mDCs were particularly diminished in patients who were imminently in renal failure. This is a counter-intuitive result, and contrary to previous observations that mDC abundance in the kidney is correlated with disease activity. [49,70–

[72] This phenomenon needs to be validated in a larger cohort of patients, but the observation leads to the hypothesis that the loss of mDCs might lead to a further degradation of immunological tolerance in the kidney.

Both of these cell subsets still need to be evaluated in the HMP dataset. We will then be able to re-do the spatial analyses and take into consideration their contributions to biopsy-wide organizational structures. In particular, we will be able to evaluate whether they co-localize with T cells of a particular phenotype, and possibly play a role in driving this phenotype.

There is regional variation to the interferon response within kidney tissue

The distribution of MX1, an interferon response gene, had a variety of distributions within the tissue. This suggests that though the interferon response is systemic[15], there is local variation in interferon signaling within inflamed kidney tissue. This naturally led to the question of whether the MX1-rich regions coincided with lymphocyte-rich regions. However, when we examined the relationship between local lymphocyte abundance and interferon signaling we found that not all dense lymphocytic regions were enriched for interferon signaling as expected.

This is one more argument in favor of the notion that not all densely inflamed regions are created equal. The cellular correlates of local interferon signaling therefore still need to be identified. One crucial potential source of type I interferon that is missing from this analysis is plasmacytoid dendritic cells. The relevant markers are in fact included in this dataset, and future work will be done to segment these cells and situate them in the context of the interferon response.

Closing comments on lupus nephritis

Overall, our findings call into question the standard model of lupus pathogenesis wherein dendritic cells take up auto-antigens, present them to CD4+ T cells, which in turn provide help to B cells that then differentiate into plasma cells and start producing large quantities of auto-antibodies.[1] While this is likely the mechanism that occurs in many patients (particularly those who respond to therapy), our findings indicate that there might be other pathways of kidney damage, such as through the activity of CD4- T cells.

These two lupus nephritis datasets emphasize the idea that inflammation in lupus nephritis is diverse in terms of how cells are distributed in the tissue, and that this diversity can be predictive of a patient's trajectory. Understanding the correlates of renal failure is an important tool for identifying new therapeutic targets, as the existing targets have largely failed to live up to their promise. It will also enable us to better interpret and contextualize the data derived from mechanistic studies in animal models.

Another concept that these studies highlight is the importance of the anatomical compartment that is being interrogated. There is sometimes a discordance between observations from the blood and kidneys of lupus nephritis patients, potentially reflecting differences between systemic autoimmunity and *in situ* inflammatory processes. Our argument is that examining inflammation in the kidney is worth the effort, as it more directly probes mechanisms that lead to renal failure.

An important next step that flows from this work would be to further characterize the distribution of B cells and all of the possible CD4- T cells (CD8 T cells, DN T cells) in a sufficiently large cohort of patients with long-term follow-up that can capture the heterogeneity within this patient population. Ideally, tracking these cellular populations

over time through repeat biopsies would be extremely valuable, as it would show the dynamics of tissue inflammation as it relates to disease progression.

This work demonstrates how automatic cellular segmentation opens the door for spatial analyses that can illuminate intercellular relationships in inflamed tissue and place them in a broader context. As spatial analysis in highly multiplexed immunofluorescence imaging data is a relatively nascent field, many of the metrics and techniques used in this analysis had to be developed for our purposes, and can be carried forward into future datasets and questions. The ability to relate the cellular status of inflammation to clinical outcome has the potential to unlock mechanistic insight into why some patients experience kidney failure and why some do not.

3. Insights about the use of computer vision to analyze tissue inflammation

Considerations of data acquisition --- whole tissue vs selective imaging

We have described two datasets that took different approaches to profiling the inflammatory state of biopsy tissue, giving us insight into the benefits and limitations of both. The high resolution data was collected by scanning the tissue for “lymphocyte-rich regions” on the basis of CD3 signal, and selectively imaging those regions. This approach was taken because imaging the whole biopsy on the SP8 microscope would have been prohibitively time-intensive and would have captured a great deal of negative data. However, this approach, as well as the non-standardized size of biopsy samples, meant that the number of images from a given patient was highly variable, ranging from 1-82.

Because the ESRD+ patients were generally more inflamed, this meant that there were many more images per patient in this cohort than in the ESRD- cohort. The selective imaging also meant that this dataset was biased towards the densely inflamed, T-cell rich

regions, at the expense of more sparse regions of inflammation that did not have abundant T cells. It therefore created substantial class imbalance, and we needed to use a variety of statistical methods to overcome these issues.

Despite these limitations, the dataset allowed us to relate cellular features in patient biopsies with the clinical picture of these patients and gain novel insights into the heterogeneous inflammatory states found in lupus nephritis patients. Analyzing the data on the scale of individual cells, neighborhoods, and ROIs allowed us to generate statistically robust results, even from a small group of patients.

On the other end of the spectrum is the highly multiplexed dataset. One key lesson to take away from this study is the value of studying the entire biopsy area and situating local inflammation in the context of the whole sample. The ability to capture large areas of tissue enables the study of larger organizational structures that span multiple image tiles. This was an advantage that our highly multiplexed dataset had over our high resolution data, which specifically interrogated T-cell rich regions, and only at the scale of a 1024x1024 pixel tile. The lack of whole-section data in the high resolution data prevents us from making patient-level observations, as we cannot be confident that the entire range of inflammation for a patient has been captured. While this does not diminish the value of the ROI-level findings, this feature of the dataset limits the hypotheses that can be tested with these images and prevents direct associations with the spatial analysis performed on the highly multiplexed dataset. Capturing whole-section data allows for interrogation of cellular distributions at these multiple scales.

The drawback to this comprehensive approach is that it is resource intensive. The amount of data that is captured with whole-tissue imaging is massive, and very little of it

contains cells of interest. In addition, this approach increases the imaging time, and the computational challenge of extracting useful segmentations from these datasets.

Therefore, for a given experimental question, the acceptable balance of tradeoffs in terms of resource use and data quality should determine which approach is taken.

Considerations of data acquisition—panel design and imaging methods

Another key element of experimental design for imaging experiments is the choice of markers, particularly as it pertains to downstream image analysis applications. In traditional immunofluorescence imaging, one often has to make tradeoffs between capturing several cell types with little phenotypic depth, or characterizing a single cell type with multiple markers. When we investigated the effect of using a single marker to identify cells vs multiple markers, we found that using multiple markers to identify a cell produced higher quality segmentations.

For this reason, stripping and re-probing in order to capture a large number of markers is an appealing approach, because of the phenotypic complexity that it has the potential to unlock while still enabling robust segmentation. However, this methodology introduced additional analytical steps, and the potential for error. Having a procedure in which the slide was removed from the microscope between stains introduced artifacts that proved difficult and sometimes impossible to mitigate computationally. In contrast, several imaging techniques (eg, CODEX)[87] are done in such a way that the slide is fixed on the microscope stage, removing that source of potential error. Another way to address this issue might be through the use of fiducial markers, such as a grid on which the tissue could be placed in order to orient the microscope lens relative to the sample more precisely.

The number of fluorescence channels that can be imaged at once is an important parameter to consider when designing this type of experiment. The microscope we used for the highly multiplexed data acquisition could only reliably image 4 channels, including the DAPI channel. This is on the low end of multiplexed fluorescence microscopy, which can often do on the scale of 6-7 fluorescence channels at a time. This limitation necessitated more rounds of imaging in order to capture the desired number of markers, which increased the amount of alignment that needed to happen, increased the potential for tissue damage between rounds of imaging, and increased the potential for antigen degradation due to repeated stripping. These were all deemed to be appropriate tradeoffs for the ability to image the whole biopsy at high speed, but likely degraded the quality of the data that was produced. Evaluating other microscopes for their potential to image more markers at a time while preserving the quick-scanning functionality would be a useful next step for future experiments.

Considerations of analytical approach --- Roads not taken

What has been described as the “highly multiplexed pipeline” represents a small fraction of what was attempted. Here I will briefly discuss techniques that were tried at various points in the pipeline and eventually discarded.

a. Fully-automated stitching and alignment

Image stitching and alignment (or registration) is a common task in image analysis, and there are consequently many automatic methods that exist for this purpose. In fact, many microscopes come with software that do this automatically, such that many imaging experiments don’t involve a purpose-built program for this task. However, our dataset had the following challenges—large imaging regions and differences in information content

between days. The size of the images was a problem because small regional misalignments that are minor on the scale of a 1024x1024 pixel image become severe when propagated out to the size of the whole image. In addition, because there was often minor tissue damage and shifts in focal plane between days, the DAPI images that were being aligned across days were similar, but at times divergent enough that cross correlation failed when aligning between days. There was substantial variance in the severity of these problems, both between samples and within a single sample across days.

We attempted to perfect a fully automated version of the stitching and alignment code, trying various window sizes for cross correlation, including processing steps like a Sobel filter for edge detection, and even other alignment algorithms like Scale Invariant Feature Transform (SIFT). In all cases the alignment code worked extremely well for some portion of the dataset, and completely failed on other parts of the dataset.

In the end it was concluded that manually adjusting the alignments on a case-by-case basis was more efficient than attempting to find a perfect automatic alignment method that worked for every sample.

b. Generative Adversarial Network for micro-alignment

Because we could only afford misalignment on the scale of the size of a fraction of a cell body (less than 10 pixels), this part of the pipeline needed to be as good as possible, or accurate cellular segmentation would not be possible. Though the semi-automatic method worked quite well, there were still minor stitching artifacts and misalignments that could affect the quality of the segmentation. Therefore, we attempted to use a Generative Adversarial Network (GAN) to correct these small errors. Briefly, this is a type of neural

network that is used to “translate” images from one format to another (eg, black and white to color). In this case, we were using a pix2pix[165] GAN to “translate” the mis-aligned or poorly stitched images into versions that had these artifacts corrected. Visually, the images were much clearer and free of these defects.

However, when we input these images into the segmentation algorithms, the network performance was not very good, producing many false positives. The key issue with the GAN is that it is “generative”, such that it is not constrained to transforming the raw data in a predictable way. In this situation, that lack of constraint meant that it was boosting the signal for some markers spuriously. In the end, this impacted segmentation quality enough to discontinue usage.

c. U-Net architecture

Before adopting the Mask R-CNN architecture for segmentation of this dataset, other network architectures were explored. At the outset of the project, because we were anticipating training several networks, it was appealing to utilize a network that was less computationally expensive. The U-Net neural network architecture is a semantic (pixel-level) segmentation method that has been applied in many biological contexts to segment nuclei. There are many methods of detecting individual objects in semantic segmentations, from watershed algorithms to region proposal networks (RPN). In fact, the U-Net/RPN was used elsewhere in the pipeline to help automate ground truth generation. Both U-Nets with traditional region proposal, and combined U-Net/RPNs were explored as possible methods for cellular segmentation in this dataset.

However, we were not ultimately able to use a U-Net or U-Net/RPN to produce high-quality cellular segmentations in our dataset. Overall, this approach suffers from poor

sensitivity, which is particularly a problem for segmenting and accurately quantifying relatively rare cellular subsets. Mask R-CNN was ultimately used because it allowed us to leverage both knowledge and training data that we already had.

All of these discarded approaches highlight the fact that there is no single obvious way to obtain high-quality segmentations from a dataset. Often, several strategies must be attempted, evaluated, and then discarded before an optimal analytical procedure is developed.

Finetuning for translation of techniques across imaging systems and pathologies

A useful insight that we derived from these studies was how valuable fine-tuning could be for minimizing the time spent developing segmentation algorithms for a given dataset. It is not tractable to spend years developing optimal segmentation algorithms for every single dataset that is collected, especially in situations where a particular cell type is rare or the dataset is small. In segmenting the HMP data, we demonstrated that finetuning can be used to turn a network that has been trained for high resolution into a network that can be applied on lower resolution data. This allowed us to make maximal use of the ground truth we had at our disposal. Similarly, we have observed from our studies with U-Nets that networks trained on one disease state (LuN) can be fine-tuned to segment the same types of cells in a completely different tissue context (TNBC). This is extremely encouraging as it suggests that analyzing the TNBC data will not require that we start from scratch, but rather, that we will be able to leverage the networks and insight that we have already developed to generate high quality segmentations in a shorter timeframe.

Closing thoughts on computer vision

In sum, using computer vision to generate high-quality cellular segmentations in tissue is a complex task, and the difficulty of the task can be heavily influenced by choices made at the stage of experimental design. As highly multiplexed imaging becomes increasingly mainstream, it is important to consider which aspects of experimental design impact downstream analysis.

When designing an experiment that results in several terabytes of data, it is critically important to put forethought into how to extract meaningful information from it. This includes thinking through database management, what kind of pre-processing is required, what kind of neural network architecture would be appropriate for the task, how to design and efficiently generate training sets for this task, and what the standard for "good enough" is. Further, it is vital to be very specific in the questions that are being asked, in order to avoid collecting the wrong data, or collecting it in a suboptimal way.

A reasonable question to ask about this approach is that plenty of people do this task in a less painstaking way[86,87]—is it worthwhile to do it this way? There are various pipelines that have been developed for similar task that seem to accomplish the goal of identifying cells in tissue without the extensive processing. There are two points I would make about this—first is that those pipelines often lack substantial quantification of how their approach is performing.[86,88,125] It is therefore difficult to evaluate whether these methods are "good enough" to extract cellular information. This raises the second point: proper segmentation is not the end goal of this work, though it took the bulk of the time and effort. The end goal is spatial analysis of how cells are arranged in tissue, which will only be as accurate as the cellular segmentations. Taking the time to segment cells

robustly improves the chances of the spatial analysis producing useful insight. Ultimately, developing high-quality methods for quantitative image analysis has the potential to deepen our understanding of human inflammation.

Bibliography

1. Anders H, Rovin B. A pathophysiology-based approach to the diagnosis and treatment of lupus nephritis. *Kidney Int*. 90(3):493-501. doi:10.1016/j.kint.2016.05.017

2. Rovin BH, Furie R, Latinis K, 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. doi:10.1002/art.34359

3. Atisha-Fregoso Y, Malkiel S, Harris KM, 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. doi:10.1002/art.41466

4. Hanly JG, O'Keeffe AG, Su L, et al. The frequency and outcome of lupus nephritis: results from an international inception cohort study. *Rheumatology*. 2016;55(2):252-262. doi:10.1093/rheumatology/kev311

5. Almaani S, Meara A, Rovin BH. Update on Lupus Nephritis. *Clin J Am Soc Nephrol*. 2017;12(5):825 LP - 835. doi:10.2215/CJN.05780616

6. Parikh S V, Almaani S, Brodsky S, Rovin BH. Update on Lupus Nephritis: Core Curriculum 2020. *Am J Kidney Dis*. 2020;76(2):265-281. doi:10.1053/j.ajkd.2019.10.017

7. Sidiropoulos PI, Kritikos HD, Boumpas DT. Lupus nephritis flares. *Lupus*. 2005;14(1):49-52. doi:10.1191/0961203305lu2059oa

8. Umeda R, Ogata S, Hara S, et al. Comparison of the 2018 and 2003 International Society of Nephrology/Renal Pathology Society classification in terms of renal prognosis in patients of lupus nephritis: a retrospective cohort study. *Arthritis Res Ther*. 2020;22(1):260. doi:10.1186/s13075-020-02358-x

9. Clark MR, Trotter K, Chang A. The Pathogenesis and Therapeutic Implications of Tubulointerstitial Inflammation in Human Lupus Nephritis. *Semin Nephrol*. 2015;35(5):455-464. doi:10.1016/j.semnephrol.2015.08.007

10. Hsieh C, Chang A, Brandt D, Guttikonda R, Utset TO, Clark MR. Predicting outcomes of lupus nephritis with tubulointerstitial inflammation and scarring. *Arthritis Care Res (Hoboken)*. 2011;63(6):865-874. doi:10.1002/acr.20441

11. McGaha TL, Madaio MP. Lupus Nephritis: Animal Modeling of a Complex Disease Syndrome Pathology. *Drug Discov Today Dis Models*. 2014;11:13-18. doi:10.1016/j.ddmod.2014.08.002

12. Matsumoto K, Watanabe N, Akikusa B, et al. Fc receptor-independent development of autoimmune glomerulonephritis in lupus-prone MRL/lpr mice.

Arthritis Rheum. 2003;48(2):486-494. doi:10.1002/art.10813

13. Tarzi RM, Sharp PEH, McDaid JP, et al. Mice with defective Fas ligand are protected from crescentic glomerulonephritis. *Kidney Int*. 2012;81(2):170-178. doi:10.1038/ki.2011.319

14. Arazi A, Rao DA, Berthier CC, et al. The immune cell lansdscape in kidneys of patients with lupus nephritis. *Nat Immunol*. 2019;20(July):902-914. doi:10.1038/s41590-019-0398-x

15. Chaussabel D, Quinn C, Shen J, et al. A Modular Analysis Framework for Blood Genomics Studies: Application to Systemic Lupus Erythematosus. *Immunity*. 2008;29(1):150-164. doi:10.1016/j.immuni.2008.05.012

16. Mohan C, Putterman C. Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nat Rev Nephrol*. 2015;11(6):329-341. doi:10.1038/nrneph.2015.33

17. Yung S, Chan TM. Anti-DNA antibodies in the pathogenesis of lupus nephritis--the emerging mechanisms. *Autoimmun Rev*. 2008;7(4):317-321. doi:10.1016/j.autrev.2007.12.001

18. Soliman S, Mohan C. Lupus nephritis biomarkers. *Clin Immunol*. 2017;185:10-20. doi:https://doi.org/10.1016/j.clim.2016.08.001

19. Pisetsky DS. Antinuclear Antibodies in Rheumatic Disease: A Proposal for a Function-Based Classification. *Scand J Immunol*. 2012;76(3):223-228. doi:https://doi.org/10.1111/j.1365-3083.2012.02728.x

20. Nowling TK, Gilkeson GS. Mechanisms of tissue injury in lupus nephritis. *Arthritis Res Ther*. 2011;13(6):250. doi:10.1186/ar3528

21. Kinloch AJ, Chang A, Ko K, et al. Vimentin is a dominant target of in situ humoral immunity in human lupus tubulointerstitial nephritis. *Arthritis Rheumatol*. 2014;66(12):3359-3370. doi:10.1002/art.38888

22. Jacobi AM, Odendahl M, Reiter K, et al. Correlation between circulating CD27high plasma cells and disease activity in patients with systemic lupus erythematosus. *Arthritis Rheum*. 2003;48(5):1332-1342. doi:10.1002/art.10949

23. Wardowska A, Komorniczak M, Skoniecka A, et al. Alterations in peripheral blood B cells in systemic lupus erythematosus patients with renal insufficiency. *Int Immunopharmacol*. 2020;83:106451. doi:10.1016/J.INTIMP.2020.106451

24. Espeli M, Bökers S, Giannico G, et al. Local renal autoantibody production in lupus nephritis. *J Am Soc Nephrol*. 2011;22(2):296-305.

doi:10.1681/ASN.2010050515

25. Jenks SA, Cashman KS, Zumaquero E, 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.e6. doi:10.1016/j.immuni.2018.08.015

26. Chan OT, Hannum LG, Haberman AM, Madaio MP, Shlomchik MJ. A novel mouse with B cells but lacking serum antibody reveals an antibody-independent role for B cells in murine lupus. *J Exp Med*. 1999;189(10):1639-1648. doi:10.1084/jem.189.10.1639

27. Sfikakis PP, Souliotis VL, Fragiadaki KG, Moutsopoulos HM, Boletis JN, Theofilopoulos AN. Increased expression of the FoxP3 functional marker of regulatory T cells following B cell depletion with rituximab in patients with lupus nephritis. *Clin Immunol*. 2007;123(1):66-73. doi:https://doi.org/10.1016/j.clim.2006.12.006

28. Trouvin AP, Jacquot S, Grigioni S, et al. Usefulness of monitoring of B cell depletion in rituximab-treated rheumatoid arthritis patients in order to predict clinical relapse: A prospective observational study. *Clin Exp Immunol*. 2015;180(1):11-18. doi:10.1111/cei.12481

29. Vigna-Perez M, Hernández-Castro B, Paredes-Saharopulos O, et al. Clinical and immunological effects of Rituximab in patients with lupus nephritis refractory to conventional therapy: a pilot study. *Arthritis Res Ther*. 2006;8(3):R83. doi:10.1186/ar1954

30. Cohen MD, Keystone E. Rituximab for Rheumatoid Arthritis. *Rheumatol Ther*. 2015;2(2):99-111. doi:10.1007/s40744-015-0016-9

31. Mysler EF, Spindler AJ, Guzman R, 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. doi:10.1002/art.38037

32. Furie R, Cascino MD, Garg JP, et al. O35 B-cell depletion and response in a randomized, controlled trial of obinutuzumab for proliferative lupus nephritis. *Lupus Sci & Med*. 2020;7(Suppl 1):A27 LP-A28. doi:10.1136/lupus-2020-eurolupus.46

33. Fassbinder T, Saunders U, Mickholz E, 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. doi:10.1186/s13075-015-0603-8

34. Furie R, Rovin BH, Houssiau F, et al. Two-Year, Randomized, Controlled Trial of

Belimumab in Lupus Nephritis. *N Engl J Med*. 2020;383(12):1117-1128. doi:10.1056/nejmoa2001180

35. Parikh S V., Malvar A, Song H, et al. Characterising the immune profile of the kidney biopsy at lupus nephritis flare differentiates early treatment responders from non-responders. *Lupus Sci Med*. 2015;2(1):1-10. doi:10.1136/lupus-2015-000112

36. Patschan S, Dolff S, Kribben A, et al. CD134 expression on CD4+ T cells is associated with nephritis and disease activity in patients with systemic lupus erythematosus. *Clin Exp Immunol*. 2006;145(2):235-242. doi:10.1111/j.1365-2249.2006.03141.x

37. Liarski VM, Kaverina N, Chang A, et al. Cell distance mapping identifies functional T follicular helper cells in inflamed human renal tissue. *Sci Transl Med*. 2014;6(230):1-12. doi:10.1126/scitranslmed.3008146

38. Rao DA, Gurish MF, Marshall JL, et al. Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature*. 2017;542(7639):110-114. doi:10.1038/nature20810

39. Bocharnikov A V, Keegan J, Wacleche VS, et al. PD-1hiCXCR5– T peripheral helper cells promote B cell responses in lupus via MAF and IL-21. *JCI Insight*. 2019;4(20). doi:10.1172/jci.insight.130062

40. Ohl K, Tenbrock K. Regulatory T cells in systemic lupus erythematosus. *Eur J Immunol*. 2015;45(2):344-355. doi:10.1002/eji.201344280

41. Hanaoka H, Nishimoto T, Okazaki Y, Takeuchi T, Kuwana M. A unique thymus-derived regulatory T cell subset associated with systemic lupus erythematosus. *Arthritis Res Ther*. 2020;22(1):88. doi:10.1186/s13075-020-02183-2

42. Shenoy S, Chaurasia S, Edavalath S, et al. Effect of induction therapy on circulating T-helper 17 and T-regulatory cells in active proliferative lupus nephritis. *Int J Rheum Dis*. 2018;21(5):1040-1048. doi:https://doi.org/10.1111/1756-185X.13272

43. Mesquita D, Jr., Kirsztajn GM, et al. CD4+ T helper cells and regulatory T cells in active lupus nephritis: an imbalance towards a predominant Th1 response? *Clin Exp Immunol*. 2018;191(1):50. doi:10.1111/CEI.13050

44. Afeltra A, Gigante A, Margiotta DPE, et al. The involvement of T regulatory lymphocytes in a cohort of lupus nephritis patients: a pilot study. *Intern Emerg Med*. 2015;10(6):677-683. doi:10.1007/s11739-015-1212-x

45. Li W, Deng C, Yang H, Wang G. The Regulatory T Cell in Active Systemic Lupus

Erythematosus Patients: A Systemic Review and Meta-Analysis. *Front Immunol*. 2019;10:159. doi:10.3389/fimmu.2019.00159

46. Scherlinger M, Guillotin V, Douchet I, et al. Selectins impair regulatory T cell function and contribute to systemic lupus erythematosus pathogenesis. *Sci Transl Med*. 2021;13(600):eabi4994. doi:10.1126/scitranslmed.abi4994

47. Jacquemin C, Augusto J-F, Scherlinger M, et al. OX40L/OX40 axis impairs follicular and natural Treg function in human SLE. *JCI Insight*. 2018;3(24). doi:10.1172/JCI.INSIGHT.122167

48. Abdirama D, Tesch S, Grießbach A-S, et al. Nuclear antigen–reactive CD4+ T cells expand in active systemic lupus erythematosus, produce effector cytokines, and invade the kidneys. *Kidney Int*. 2021;99(1):238-246. doi:https://doi.org/10.1016/j.kint.2020.05.051

49. Tucci M, Quatraro C, Lombardi L, Pellegrino C, Dammacco F, Silvestris F. Glomerular accumulation of plasmacytoid dendritic cells in active lupus nephritis: Role of interleukin-18. *Arthritis Rheum*. 2008;58(1):251-262. doi:https://doi.org/10.1002/art.23186

50. Fava A, Buyon J, Mohan C, et al. Integrated urine proteomics and renal single-cell genomics identify an IFN-γ response gradient in lupus nephritis. *JCI Insight*. 2020;5(12). doi:10.1172/jci.insight.138345

51. Dolff S, Abdulahad WH, Arends S, et al. Urinary CD8+ T-cell counts discriminate between active and inactive lupus nephritis. *Arthritis Res Ther*. 2013;15(1):R36. doi:10.1186/ar4189

52. Minning S, Xiaofan Y, Anqi X, et al. Imbalance between CD8+CD28+ and CD8+CD28– T-cell subsets and its clinical significance in patients with systemic lupus erythematosus. *Lupus*. 2019;28(10):1214-1223. doi:10.1177/0961203319867130

53. Khodoun M, Chimote AA, Ilyas FZ, et al. Targeted knockdown of Kv1.3 channels in T lymphocytes corrects the disease manifestations associated with systemic lupus erythematosus. *Sci Adv*. 2020;6(47):eabd1471. doi:10.1126/sciadv.abd1471

54. Zhou M, Guo C, Li X, et al. JAK/STAT signaling controls the fate of CD8+CD103+ tissue-resident memory T cell in lupus nephritis. *J Autoimmun*. 2020;109:102424. doi:https://doi.org/10.1016/j.jaut.2020.102424

55. Winchester R, Wiesendanger M, Zhang H-Z, et al. Immunologic Characteristics of Intrarenal T cells: Trafficking of Expanded CD8 T Cell β-Chain Clonotypes In Progressive Lupus Nephritis. *Arthritis Rheum*. 2012;64(5):1589.

doi:10.1002/ART.33488

56. McKinney EF, Smith KGC. Metabolic exhaustion in infection, cancer and autoimmunity review-article. *Nat Immunol*. 2018;19(3):213-221. doi:10.1038/s41590-018-0045-y

57. McKinney EF, Lee JC, Jayne DRW, Lyons PA, Smith KGC. T-cell exhaustion, co-stimulation and clinical outcome in autoimmunity and infection. *Nature*. 2015;523(7562):612-616. doi:10.1038/nature14468

58. Buang N, Tapeng L, Gray V, et al. Type I interferons affect the metabolic fitness of CD8+ T cells from patients with systemic lupus erythematosus. *Nat Commun*. 2021;12(1):1980. doi:10.1038/s41467-021-22312-y

59. Maccari ME, Fuchs S, Kury P, et al. A distinct CD38+CD45RA+ population of CD4+, CD8+, and double-negative T cells is controlled by FAS. *J Exp Med*. 2020;218(2). doi:10.1084/jem.20192191

60. Crispín JC, Oukka M, Bayliss G, 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. doi:10.4049/jimmunol.181.12.8761

61. Li H, Adamopoulos IE, Moulton VR, et al. Systemic lupus erythematosus favors the generation of IL-17 producing double negative T cells. *Nat Commun*. 2020;11(1):2859. doi:10.1038/s41467-020-16636-4

62. Stratigou V, Doyle AF, Carlucci F, et al. Altered expression of signalling lymphocyte activation molecule receptors in T-cells from lupus nephritis patients-a potential biomarker of disease activity. doi:10.1093/rheumatology/kex078

63. Uchida T, Ito S, Kumagai H, Oda T, Nakashima H, Seki S. Roles of Natural Killer T Cells and Natural Killer Cells in Kidney Injury. *Int J Mol Sci*. 2019;20(10). doi:10.3390/IJMS20102487

64. Yin S, Mao Y, Li X, et al. Hyperactivation and in situ recruitment of inflammatory Vδ2 T cells contributes to disease pathogenesis in systemic lupus erythematosus. *Sci Rep*. 2015;5(1):14432. doi:10.1038/srep14432

65. Law BM-P, Wilkinson R, Wang X, et al. Effector γδ T cells in human renal fibrosis and chronic kidney disease. *Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc*. 2019;34(1):40-48. doi:10.1093/ndt/gfy098

66. Group TAT, Askanase A, Byron M, et al. Treatment of Lupus Nephritis with Abatacept: The Abatacept and Cyclophosphamide Combination Efficacy and Safety Study. *Arthritis Rheumatol (Hoboken, NJ)*. 2014;66(11):3096. doi:10.1002/ART.38790

67. Mok CC. Calcineurin inhibitors in systemic lupus erythematosus. *Best Pract Res Clin Rheumatol*. 2017;31(3):429-438. doi:https://doi.org/10.1016/j.berh.2017.09.010

68. Rovin BH, Teng YKO, Ginzler EM, 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. doi:https://doi.org/10.1016/S0140-6736(21)00578-X

69. Rönnblom L, Pascual V. The innate immune system in SLE: Type I interferons and dendritic cells. *Lupus*. 2008;17(5 SPEC. ISS.):394-399. doi:10.1177/0961203308090020

70. Fiore N, Castellano G, Blasi A, et al. Immature myeloid and plasmacytoid dendritic cells infiltrate renal tubulointerstitium in patients with lupus nephritis. *Mol Immunol*. 2008;45(1):259-265. doi:https://doi.org/10.1016/j.molimm.2007.04.029

71. Wardowska A, Komorniczak M, Bułło-Piontecka B, Dębska-Ślizień MA, Pikuła M. Transcriptomic and Epigenetic Alterations in Dendritic Cells Correspond With Chronic Kidney Disease in Lupus Nephritis. *Front Immunol*. 2019;10:2026. doi:10.3389/fimmu.2019.02026

72. Liarski VM, Sibley A, van Panhuys N, et al. Quantifying in situ adaptive immune cell cognate interactions in humans. *Nat Immunol*. February 2019. doi:10.1038/s41590-019-0315-3

73. Kassianos AJ, Wang X, Sampangi S, Muczynski K, Healy H, Wilkinson R. Increased tubulointerstitial recruitment of human CD141hi CLEC9A+ and CD1c+ myeloid dendritic cell subsets in renal fibrosis and chronic kidney disease. *Am J Physiol Physiol*. 2013;305(10):F1391-F1401. doi:10.1152/ajprenal.00318.2013

74. Parikh S V, Malvar A, Shapiro J, 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. doi:10.3389/fimmu.2021.621039

75. Wang H, Li T, Chen S, Gu Y, Ye S. Neutrophil Extracellular Trap Mitochondrial DNA and Its Autoantibody in Systemic Lupus Erythematosus and a Proof-of-Concept Trial of Metformin. *Arthritis Rheumatol*. 2015;67(12):3190-3200. doi:https://doi.org/10.1002/art.39296

76. Hakkim A, Fürnrohr BG, Amann K, et al. Impairment of neutrophil extracellular trap degradation is associated with lupus nephritis. *Proc Natl Acad Sci*. 2010;107(21):9813 LP - 9818. doi:10.1073/pnas.0909927107

77. Vallin H, Perers A, Alm G V, Rönnblom L. Anti-double-stranded DNA antibodies and immunostimulatory plasmid DNA in combination mimic the endogenous IFN-

alpha inducer in systemic lupus erythematosus. *J Immunol*. 1999;163(11):6306-6313.

78. Baechler EC, Batliwalla FM, Karypis G, 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. doi:10.1073/pnas.0337679100

79. Denton AE, Innocentin S, Carr EJ, et al. Type I interferon induces CXCL13 to support ectopic germinal center formation. *J Exp Med*. 2019;216(3):621-637. doi:10.1084/jem.20181216

80. Der E, Suryawanshi H, Morozov P, 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. doi:10.1038/s41590-019-0386-1

81. Petri M, Wallace DJ, Spindler A, et al. Sifalimumab, a Human Anti–Interferon-α Monoclonal Antibody, in Systemic Lupus Erythematosus: A Phase I Randomized, Controlled, Dose-Escalation Study. *Arthritis Rheum*. 2013;65(4):1011-1021. doi:https://doi.org/10.1002/art.37824

82. Khamashta M, Merrill JT, Werth VP, 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 LP - 1916. doi:10.1136/annrheumdis-2015-208562

83. Furie RA, Morand EF, Bruce IN, 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):e208-e219. doi:10.1016/S2665-9913(19)30076-1

84. Morand EF, Furie R, Tanaka Y, et al. Trial of Anifrolumab in Active Systemic Lupus Erythematosus. *N Engl J Med*. 2019;382(3):211-221. doi:10.1056/NEJMoa1912196

85. Emmert-Streib F, Yang Z, Feng H, Tripathi S, Dehmer M. An Introductory Review of Deep Learning for Prediction Models With Big Data. *Front Artif Intell*. 2020;3:4. doi:10.3389/frai.2020.00004

86. Radtke AJ, Kandov E, Lowekamp B, et al. 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*. 2021;117(52):33455-33465. doi:10.1073/PNAS.2018488117

87. Schürch CM, Bhate SS, Barlow GL, et al. Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell*. 2020;182(5):1341-1359.e19. doi:10.1016/j.cell.2020.07.005

88. Tsujikawa T, Kumar S, Borkar RN, et al. Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. *Cell Rep*. 2017;19(1):203-217. doi:10.1016/j.celrep.2017.03.037

89. Durkee MS, Abraham R, Clark MR, Giger ML. Artificial Intelligence and Cellular Segmentation in Tissue Microscopy Images. *Am J Pathol*. June 2021. doi:10.1016/j.ajpath.2021.05.022

90. Zaqout S, Becker L-L, Kaindl AM. Immunofluorescence Staining of Paraffin Sections Step by Step. *Front Neuroanat*. 2020;14:83. doi:10.3389/fnana.2020.582218

91. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM*. 2017;60(6):84-90. doi:10.1145/3065386

92. Tizhoosh HR, Pantanowitz L. Artificial Intelligence and Digital Pathology: Challenges and Opportunities. *J Pathol Inform*. 2018;9:38. doi:10.4103/jpi.jpi_53_18

93. Wang Y, Guan Q, Lao I, et al. Using deep convolutional neural networks for multi-classification of thyroid tumor by histopathology: a large-scale pilot study. *Ann Transl Med*. 2019;7(18):468. doi:10.21037/atm.2019.08.54

94. Gertych A, Swiderska-Chadaj Z, Ma Z, et al. Convolutional neural networks can accurately distinguish four histologic growth patterns of lung adenocarcinoma in digital slides. *Sci Rep*. 2019;9(1):1483. doi:10.1038/s41598-018-37638-9

95. Shaban M, Ogur Z, Mahmoud A, et al. A convolutional neural network for the screening and staging of diabetic retinopathy. *PLoS One*. 2020;15(6):e0233514. doi:10.1371/JOURNAL.PONE.0233514

96. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv Prepr arXiv14091556*. 2014.

97. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. ; 2016:770-778. doi:10.1109/CVPR.2016.90

98. Ronneberger, Olaf; Fischer, Philipp; Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, Vol 9351.* ; 2015.

99. Falk T, Mai D, Bensch R, et al. U-Net: deep learning for cell counting, detection,

and morphometry. *Nat Methods*. 2019;16(1):67-70. doi:10.1038/s41592-018-0261-2

100. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. *Deep Learn Med Image Anal Multimodal Learn Clin Decis Support 4th Int Work DLMIA 2018, 8th Int Work ML-CDS 2018, held conjunction with MICCAI 2018, Granada, Spain, S.* 2018;11045:3-11. doi:10.1007/978-3-030-00889-5_1

101. Yang L, Ghosh RP, Franklin JM, You C, Liphardt JT. NuSeT: A Deep Learning Tool for Reliably Separating and Analyzing Crowded Cells. *bioRxiv*. January 2019:749754. doi:10.1101/749754

102. Lin T-Y, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. *IEEE Trans Pattern Anal Mach Intell*. 2020;42(2):318-327. doi:10.1109/TPAMI.2018.2858826

103. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(6). doi:10.1109/TPAMI.2016.2577031

104. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. ; 2017:2980-2988. doi:10.1109/ICCV.2017.322

105. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. *J Big Data*. 2019;6(1):60. doi:10.1186/s40537-019-0197-0

106. Jang H-J, Song IH, Lee SH. Generalizability of Deep Learning System for the Pathologic Diagnosis of Various Cancers. *Appl Sci* . 2021;11(2). doi:10.3390/app11020808

107. Deniz E, Şengür A, Kadiroğlu Z, Guo Y, Bajaj V, Budak Ü. Transfer learning based histopathologic image classification for breast cancer detection. *Heal Inf Sci Syst*. 2018;6(1). doi:10.1007/S13755-018-0057-X

108. Rodemerk J, Junker A, Chen B, et al. Pathophysiology of Intracranial Aneurysms: COX-2 Expression, Iron Deposition in Aneurysm Wall, and Correlation with Magnetic Resonance Imaging. *Stroke*. 2020;51(8):2505-2513. doi:10.1161/STROKEAHA.120.030590

109. Gómez-de-Mariscal E, García-López-de-Haro C, Donati L, Unser M, Muñoz-Barrutia A, Sage D. Deepimagej: A user-friendly plugin to run deep learning models in imagej. *bioRxiv*. October 2019:799270. doi:10.1101/799270

110. Rivenson Y, Wang H, Wei Z, et al. Virtual histological staining of unlabelled

tissue-autofluorescence images via deep learning. *Nat Biomed Eng*. 2019;3(6):466-477. doi:10.1038/s41551-019-0362-y

111. Carpenter AE, Jones TR, Lamprecht MR, et al. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biol*. 2006;7(10):R100. doi:10.1186/gb-2006-7-10-r100

112. McQuin C, Goodman A, Chernyshev V, et al. CellProfiler 3.0: Next-generation image processing for biology. Misteli T, ed. *PLOS Biol*. 2018;16(7):e2005970. doi:10.1371/journal.pbio.2005970

113. Sadanandan SK, Ranefall P, Le Guyader S, Wählby C. Automated Training of Deep Convolutional Neural Networks for Cell Segmentation. *Sci Rep*. 2017;7(1):1-7. doi:10.1038/s41598-017-07599-6

114. Berg S, Kutra D, Kroeger T, et al. ilastik: interactive machine learning for (bio)image analysis. *Nat Methods*. 2019;16(12):1226-1232. doi:10.1038/s41592-019-0582-9

115. McMahon NP, Jones JA, Kwon S, et al. Oligonucleotide conjugated antibodies permit highly multiplexed immunofluorescence for future use in clinical histopathology. *J Biomed Opt*. 2020;25(05):1. doi:10.1117/1.JBO.25.5.056004

116. Rashid R, Gaglia G, Chen YA, et al. Highly multiplexed immunofluorescence images and single-cell data of immune markers in tonsil and lung cancer. *Sci Data*. 2019;6(1):1-10. doi:10.1038/s41597-019-0332-y

117. Wills JW, Robertson J, Summers HD, et al. Image-Based Cell Profiling Enables Quantitative Tissue Microscopy in Gastroenterology. *Cytom Part A*. 2020;97(12):1222-1237. doi:10.1002/cyto.a.24042

118. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1):1-7. doi:10.1038/s41598-017-17204-5

119. Diem K, Magaret A, Klock A, Jin L, Zhu J, Corey L. Image analysis for accurately counting CD4+ and CD8+ T cells in human tissue. *J Virol Methods*. 2015;222:117-121. doi:https://doi.org/10.1016/j.jviromet.2015.06.004

120. Nearchou IP, Gwyther BM, Georgiakakis ECT, et al. Spatial immune profiling of the colorectal tumor microenvironment predicts good outcome in stage II patients. *npj Digit Med*. 2020;3(1):1-10. doi:10.1038/s41746-020-0275-x

121. Lazarus J, Maj T, Smith JJ, et al. Spatial and phenotypic immune profiling of metastatic colon cancer. *JCI insight*. 2018;3(22). doi:10.1172/jci.insight.121932

122. 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 (Basel)*. 2020;12(12):1-15. doi:10.3390/cancers12123562

123. 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(2). doi:10.1117/1.JBO.26.2.022910

124. Sergeev A, Del Balso M. Horovod: fast and easy distributed deep learning in TensorFlow. *arXiv Prepr arXiv180205799*. 2018.

125. Goltsev Y, Samusik N, Kennedy-Darling J, et al. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell*. 2018;174(4):968-981.e15. doi:10.1016/j.cell.2018.07.010

126. Schubert E, Sander J, Ester M, Kriegel HP, Xu X. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Trans Database Syst*. 2017;42(3). doi:10.1145/3068335

127. Pavlides M, Birks J, Fryer E, et al. Interobserver Variability in Histologic Evaluation of Liver Fibrosis Using Categorical and Quantitative Scores. *Am J Clin Pathol*. 2017;147(4):364-369. doi:10.1093/ajcp/aqx011

128. Allard FD, Goldsmith JD, Ayata G, et al. Intraobserver and Interobserver Variability in the Assessment of Dysplasia in Ampullary Mucosal Biopsies. *Am J Surg Pathol*. 2018;42(8). https://journals.lww.com/ajsp/Fulltext/2018/08000/Intraobserver_and_Interobserver_Variability_in_the.13.aspx.

129. Bolognesi MM, Manzoni M, Scalia CR, et al. Multiplex Staining by Sequential Immunostaining and Antibody Removal on Routine Tissue Sections. *J Histochem Cytochem*. 2017;65(8):431-444. doi:10.1369/0022155417719419

130. Tilstra JS, Avery L, Menk A V, et al. Kidney-infiltrating T cells in murine lupus nephritis are metabolically and functionally exhausted. *J Clin Invest*. 2018;128(11):4884-4897. doi:10.1172/JCI120859

131. Della-Torre E, Bozzalla-Cassione E, Sciorati C, et al. A CD8α− Subset of CD4+SLAMF7+ Cytotoxic T Cells Is Expanded in Patients With IgG4-Related Disease and Decreases Following Glucocorticoid Treatment. *Arthritis \& Rheumatol*. 2018;70(7):1133-1143. doi:https://doi.org/10.1002/art.40469

132. Rawlings DJ, Metzler G, Wray-Dutra M, Jackson SW. Altered B cell signalling in autoimmunity. *Nat Rev Immunol*. 2017;17(7):421-436. doi:10.1038/nri.2017.24

133. Minowa K, Amano H, Nakano S, et al. Elevated serum level of circulating syndecan-1 (CD138) in active systemic lupus erythematosus. *Autoimmunity*. 2011;44(5):357-362. doi:10.3109/08916934.2010.545846

134. Keren L, Bosse M, Marquez D, et al. A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell*. 2018;174(6):1373-1387.e19. doi:10.1016/j.cell.2018.08.039

135. Wang L, Simons DL, Lu X, et al. Connecting blood and intratumoral Treg cell activity in predicting future relapse in breast cancer. *Nat Immunol*. 2019;20(9):1220-1230. doi:10.1038/s41590-019-0429-7

136. Gu-Trantien C, Migliori E, Buisseret L, et al. CXCL13-producing T FH cells link immune suppression and adaptive memory in human breast cancer. *JCI Insight*. 2017;2(11):1-17. doi:10.1172/jci.insight.91487

137. Ahrends T, Spanjaard A, Pilzecker B, et al. CD4+T Cell Help Confers a Cytotoxic T Cell Effector Program Including Coinhibitory Receptor Downregulation and Increased Tissue Invasiveness. *Immunity*. 2017;47(5):848-861.e5. doi:10.1016/j.immuni.2017.10.009

138. Kawaji H, Kubo M, Motoyama Y, et al. Functional analysis of tumour infiltrating lymphocytes in triple negative breast cancer focusing on granzyme B. *Ann Oncol*. 2019;30:v5. doi:10.1093/annonc/mdz238.015

139. Azizi E, Carr AJ, Plitas G, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell*. 2018;174(5):1293-1308.e36. doi:10.1016/j.cell.2018.05.060

140. Kim Y-A, Lee HJ, Heo S-H, et al. MxA expression is associated with tumor-infiltrating lymphocytes and is a prognostic factor in triple-negative breast cancer. *Breast Cancer Res Treat*. 2016;156(3):597-606. doi:10.1007/s10549-016-3786-z

141. Rajković N, Li X, Plataniotis KN, Kanjer K, Radulovic M, Milošević NT. The Pan-Cytokeratin Staining Intensity and Fractal Computational Analysis of Breast Tumor Malignant Growth Patterns Prognosticate the Occurrence of Distant Metastasis. *Front Oncol*. 2018;8:348. doi:10.3389/fonc.2018.00348

142. Inwald EC, Klinkhammer-Schalke M, Hofstädter F, et al. Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. *Breast Cancer Res Treat*. 2013;139(2):539-552. doi:10.1007/s10549-013-2560-8

143. Rankov V, Locke RJ, Edens RJ, Barber PR, Vojnovic B. An Algorithm for image stitching and blending. In: *Proc.SPIE*. Vol 5701. ; 2005. https://doi.org/10.1117/12.590536.

144. Alexakis C, Maxwell P, Bou-Gharios G. Organ-Specific Collagen Expression: Implications for Renal Disease. *Nephron Exp Nephrol*. 2006;102(3-4):e71-e75. doi:10.1159/000089684

145. Croce AC, Bottiroli G. Autofluorescence spectroscopy and imaging: a tool for biomedical research and diagnosis. *Eur J Histochem*. 2014;58(4):2461. doi:10.4081/ejh.2014.2461

146. Celie JWAM, Katta KK, Adepu S, et al. Tubular epithelial syndecan-1 maintains renal function in murine ischemia/reperfusion and human transplantation. *Kidney Int*. 2012;81(7):651-661. doi:10.1038/ki.2011.425

147. Salton do Prado K. How DBSCAN works and why should we use it? https://towardsdatascience.com/how-dbscan-works-and-why-should-i-use-it-443b4a191c80. Published 2017. Accessed January 12, 2019.

148. Kromp F, Fischer L, Bozsaky E, et al. Deep Learning architectures for generalized immunofluorescence based nuclear image segmentation. *CoRR*. 2019;abs/1907.1. http://arxiv.org/abs/1907.12975.

149. Long F. Microscopy cell nuclei segmentation with enhanced U-Net. *BMC Bioinformatics*. 2020;21(1):8. doi:10.1186/s12859-019-3332-1

150. Pan X, Li L, Yang D, He Y, Liu Z, Yang H. An Accurate Nuclei Segmentation Algorithm in Pathological Image Based on Deep Semantic Network. *IEEE Access*. 2019;7:110674-110686. doi:10.1109/ACCESS.2019.2934486

151. Bulten W, Bándi P, Hoven J, et al. Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard. *Sci Rep*. 2019;9(1):864. doi:10.1038/s41598-018-37257-4

152. Isaka S, Kawanaka H, Aronow BJ, Prasath VBS. Multi-Class Segmentation of Lung Immunofluorescence Confocal Images Using Deep Learning. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. ; 2019:2362-2368. doi:10.1109/BIBM47256.2019.8983146

153. Abraham R, Durkee MS, Veselits M, et al. Application and generalizability of U-Net segmentation of immune cells in inflamed tissue. In: *Proc.SPIE*. Vol 11603. ; 2021. https://doi.org/10.1117/12.2581063.

154. Abraham R, Durkee MS, Veselits M, et al. Application of U-Nets for the automatic segmentation of immune cells in high-dimensional microscopy images of kidney biopsies. In: *Proc.SPIE*. Vol 11647. ; 2021. https://doi.org/10.1117/12.2577788.

155. Amiri M, Brooks R, Rivaz H. Fine-Tuning U-Net for Ultrasound Image Segmentation: Different Layers, Different Outcomes. *IEEE Trans Ultrason Ferroelectr Freq Control*. 2020;67(12):2510-2518.

doi:10.1109/TUFFC.2020.3015081

156. Chen C-H, Hsu M-Y, Jiang R-S, Wu S-H, Chen F-J, Liu S-A. Shrinkage of head and neck cancer specimens after formalin fixation. *J Chin Med Assoc*. 2012;75(3):109-113. doi:10.1016/j.jcma.2012.02.006

157. Tran T, Sundaram CP, Bahler CD, et al. Correcting the Shrinkage Effects of Formalin Fixation and Tissue Processing for Renal Tumors: toward Standardization of Pathological Reporting of Tumor Size. *J Cancer*. 2015;6(8):759-766. doi:10.7150/jca.12094

158. Eltoum I, Fredenburgh J, Myers RB, Grizzle WE. Introduction to the Theory and Practice of Fixation of Tissues. *http://dx.doi.org/101179/his2001243173*. 2013;24(3):173-190. doi:10.1179/HIS.2001.24.3.173

159. Verdijk P, van Veelen PA, de Ru AH, et al. Morphological changes during dendritic cell maturation correlate with cofilin activation and translocation to the cell membrane. *Eur J Immunol*. 2004;34(1):156-164. doi:10.1002/eji.200324241

160. Kläsener K, Jellusova J, Andrieux G, et al. CD20 as a gatekeeper of the resting state of human B cells. *Proc Natl Acad Sci*. 2021;118(7):e2021342118. doi:10.1073/pnas.2021342118

161. Olatunde AC, Hale JS, Lamb TJ. Cytokine-skewed Tfh cells: functional consequences for B cell help. *Trends Immunol*. 2021;42(6):536-550. doi:10.1016/j.it.2021.04.006

162. McKinney EF, Smith KGC. Metabolic exhaustion in infection, cancer and autoimmunity review-article. *Nat Immunol*. 2018;19(3):213-221. doi:10.1038/s41590-018-0045-y

163. Rose T, Dörner T. Drivers of the immunopathogenesis in systemic lupus erythematosus. *Best Pract Res Clin Rheumatol*. 2017;31(3):321-333. doi:10.1016/j.berh.2017.09.007

164. Ito T, Yang M, Wang Y-H, et al. Plasmacytoid dendritic cells prime IL-10–producing T regulatory cells by inducible costimulator ligand . *J Exp Med*. 2007;204(1):105-115. doi:10.1084/jem.20061660

165. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. ; 2017:1125-1134.