Image Analysis Pipeline for Renal Allograft Evaluation and Fibrosis Quantification

Check for updates

Alton Brad Farris[1], Juan Vizcarra[2], Mohamed Amgad[3], Lee Alex Donald Cooper[3], David Gutman[4] and Julien Hogan[5]

[1]Department of Pathology, Emory University School of Medicine, Atlanta, Georgia, USA; [2]Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, USA; [3]Center for Computational Imaging and Signal Analytics, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA; [4]Department of Neurology, Emory University School of Medicine, Atlanta, Georgia, USA; and [5]Emory Transplant Center, Department of Surgery, Emory University School of Medicine, Atlanta, Georgia, USA

Introduction: Digital pathology improves the standardization and reproducibility of kidney biopsy specimen assessment. We developed a pipeline allowing the analysis of many images without requiring human preprocessing and illustrate its use with a simple algorithm for quantification of interstitial fibrosis on a large dataset of kidney allograft biopsy specimens.

Methods: Masson trichrome–stained images from kidney allograft biopsy specimens were used to train and validate a glomeruli detection algorithm using a VGG19 convolutional neural network and an automatic cortical region of interest (ROI) selection algorithm including cortical regions containing all predicted glomeruli. A positive-pixel count algorithm was used to quantify interstitial fibrosis on the ROIs and the association between automatic fibrosis and pathologist evaluation, estimated glomerular filtration rate (GFR) and allograft survival was assessed.

Results: The glomeruli detection (F1 score of 0.87) and ROIs selection (F1 score 0.83 [SD 0.13]) algorithms displayed high accuracy. The correlation between the automatic fibrosis quantification on manually and automatically selected ROIs was high ($r = 1.00$ [0.99–1.00]). Automatic fibrosis quantification was only moderately correlated with pathologists' assessment and was not significantly associated with eGFR or allograft survival.

Conclusion: This pipeline can automatically and accurately detect glomeruli and select cortical ROIs that can easily be used to develop, validate, and apply image analysis algorithms.

K idney allograft biopsy procedures are currently the cornerstone of the management of transplant recipients, allowing posttransplant monitoring of the graft, diagnosis of complications, such as rejection, and the assessment of allograft prognosis. The complexity of allograft biopsy interpretation is reflected in the regular updates of the Banff classification.[1] This complexity makes the analysis of these biopsy specimens time-consuming for pathologists and partially explains the low interrater agreement among pathologists.[2–4]

The development of digital pathology is an opportunity to improve the standardization and reproducibility of kidney allograft biopsy specimen assessment. During the last Banff meeting, a digital pathology working group was created to standardize practices and improve the scoring of histologic parameters (e.g., interstitial fibrosis and tubular atrophy and inflammation) through algorithm development and classification.[5] In a survey after the meeting, we estimated that >70% of renal pathology departments in the United States have access to whole-slide imaging, supporting the potential for widespread use of morphometric methods on a large number of images either prospectively or retrospectively. There are a growing number of studies reporting the development of algorithms aiming at analyzing whole slide images (WSI) of kidney biopsy specimens and performing various tasks, such as fibrosis quantification, inflammation quantification, or segmentation of various structures.[6–8] However, to

date, most of these studies validated their algorithms on a limited number of selected cases. There is a need to validate these algorithms in large datasets that display the variation in disease presentation and image quality that is often seen in practice. Using such a dataset would allow for a better understanding of the algorithms' true potential in clinical practice and clinical research. Several barriers remain to be addressed to fulfill that goal, such as handling variations in biopsy staining, aging of the slides, biopsy scanning, and removing the need for human annotation. Kidney allograft biopsy evaluation is performed using the Banff classification and focuses on the structures in the cortical area. Therefore, previous studies required manual selection of the cortical area before applying quantitative algorithms, which is problematic when working with hundreds of biopsy specimens. Our aim was to create an analysis pipeline to automatically select cortical ROIs on kidney transplant biopsy specimens allowing the application of image analysis algorithms without the need for manual annotation of cortical regions in WSIs.

Interstitial fibrosis quantification is a common task in the evaluation of kidney biopsy specimens.[1,9] Previous studies demonstrated that morphometric assessment of interstitial fibrosis using special staining, such as collagen III staining or Sirius-red staining, improved reproducibility and association with allograft function.[4,10] Other teams tried to overcome the limitations of color analysis methods by using other imaging techniques, such as Fourier-transform infrared imaging.[11] However, these methods are not currently used in most centers, probably because of the burden of additional staining, slide scanning, and image preprocessing. The development of a tool that is able to perform automatic and reproducible quantification of interstitial fibrosis on stains performed in clinical care on kidney transplant biopsy specimens remains an unmet need.

In this study, we report the development of a pipeline allowing the automatic and fast selection of cortical ROIs based on glomeruli detection. We then incorporate a simple color analysis algorithm to show the use of this pipeline for morphometric quantification of interstitial fibrosis on a large dataset of trichrome-stained kidney allograft biopsy specimens.

METHODS

Study Population

We included all adult patients who had a transplant procedure at Emory University who underwent a kidney transplant biopsy procedure between 2 weeks posttransplantation and 1 year posttransplant between January 1, 2012 and December 31, 2017. Clinical data included recipients', donors', and transplants' characteristics as well as follow-up data, including biopsy results (Banff classification and final diagnosis), immunosuppressive treatments, and GFR assessment, which were collected in the Emory Transplant Data-Mart. For each patient, 3 slides stained with hematoxylin and eosin, 1 slide stained with periodic acid–Schiff and 1 slide stained with Masson trichrome were scanned. An Aperio Scanscope CS (Leica/Aperio Technologies, Inc., Vista, CA) scanner was used for digitizing the whole slide using a 20× objective lens with a numerical aperture of 0.75 coupled with a doubler objective to achieve a whole slide scan at 40× magnification. These images were stored on a secured server. Images were visualized using the Digital Slide Archive and reviewed to remove images containing tissue-processing artifacts including bubbles, section folds, pen markings, and poor staining. Basic image operations including color normalization were performed using HistomicsTK, a Python package for histology images developed by our team.[12]

Glomeruli Detection

We developed a glomerular detection algorithm using a deep learning approach to specifically select cortical ROIs. Our aim was to develop a model that would be trained to recognize glomeruli in Masson trichrome–stained images.

The dataset used to develop the model included 51 trichrome-stained WSIs (36 WSIs as training, 6 WSIs as validation, and 9 WSIs as testing). All glomeruli in this dataset were delineated/annotated using the HistomicsUI front-end of the Digital Slide Archive software[12] (Figure 1). The HistomicsTK Python package was then used to tile the WSI, a process that breaks up the large WSI into smaller images that can be used in deep learning computational models. The WSI was tiled at 10× magnification with tile size of 224 × 224. This resolution was chosen because the average size of glomeruli in our dataset would fit within an image of this size at this magnification.

Every WSI tile containing sufficient glomeruli were labeled as glomerular images. The nonglomerular class was taken from tiles containing tissue but no glomeruli in them. The nonglomerular class was oversampled so the final training dataset would include twice the number of nonglomerular images than the glomerular class. This improves model performance because the nonglomerular class contains considerably more variation than the glomerular images. The training dataset was also color-augmented using color normalization with the Reinhard approach.[13] Four WSIs from the training dataset were selected as the reference images
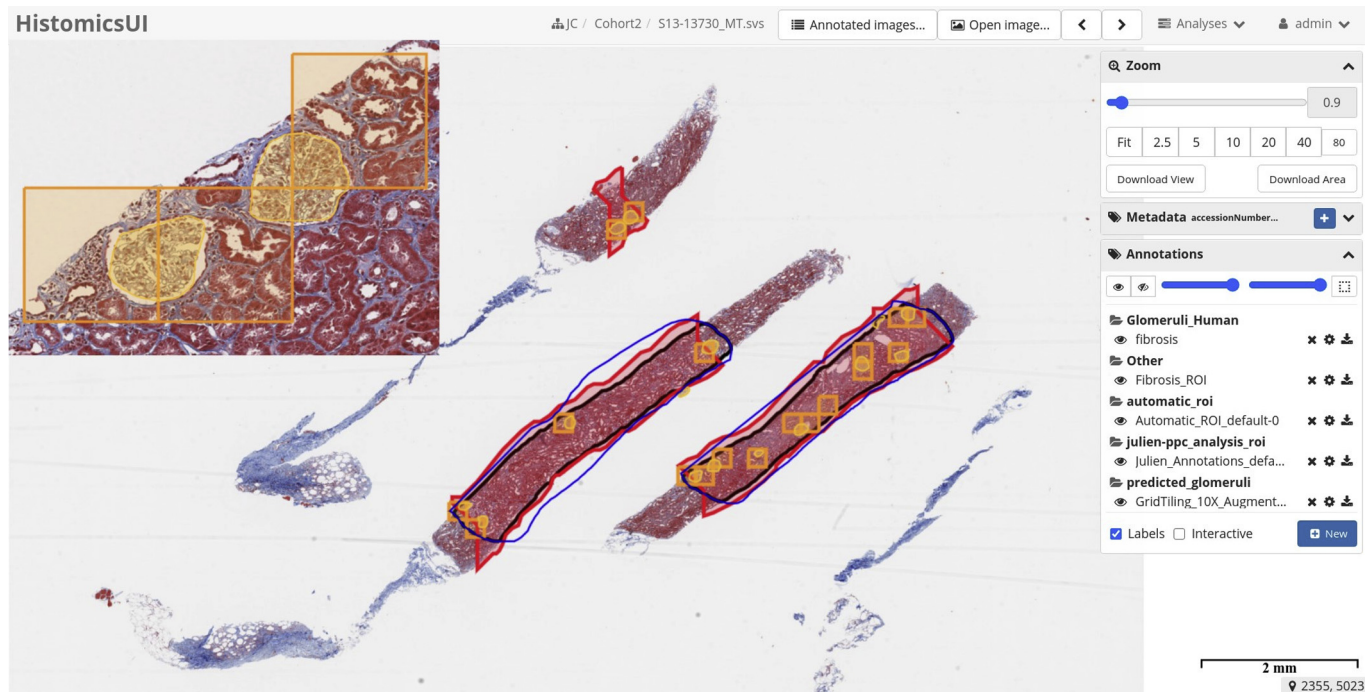
Figure 1. Illustration of the HistomicsUI interface used for viewing whole slide images and to create and view annotations.

for color normalization (Supplementary Figure S1). The training dataset consisted of 3316 glomerular images and 6632 nonglomerular images and the validation dataset consisted of 62 glomerular images and 124 nonglomerular images.

We trained a VGG19 convolutional neural network using keras starting from ImageNet weights using the training and validation dataset, a technique known as transfer learning.[14] The model was trained to classify images into 1 of 2 classes: glomerular or nonglomerular. The validation dataset was used to access training model performance to identify the best set of hyper-parameters to use (for detailed information about model training, see Supplementary Material). The testing dataset was tiled slightly differently with all tiles containing tissue being included (142 glomerular images and 2189 nonglomerular images). The performance of the glomerular detection algorithm was assessed by the F1 score, which is the harmonic mean of the precision (or positive predictive value) and recall (or sensitivity).

Automatic ROI Selection

For each WSI, a low-resolution mask was created for the tissue present in the image, and contours for these tissues were extracted using Python's OpenCV package. The automatic ROI was defined as the region containing all glomeruli predicted in each WSI tissue. The algorithm developed includes the following computer vision steps: 1) find the skeleton of the tissue masks; 2) find all glomeruli within the tissue and find

the skeleton point closest to each glomerulus; 3) convert the problem to a graph and apply Dijkstra's algorithm to find the 2 furthest glomerular points in the skeleton[15]; and 4) thicken the skeleton line back to cover the tissue. These steps are shown in Figure 2.

The performance of this algorithm was evaluated on 51 WSIs where ROIs had been manually delineated and had not been used to train the glomerular detection algorithm. The F1 score, calculated as 2 times the area of overlap divided by the total number of pixels in both ROIs (manual and automatic), was used to assess the accuracy. F1 scores range from 0 to 1, with 1 indicating perfect agreement. We also assessed the correlation between the positive pixel count (PPC) assessment in manual vs. automatic ROIs using the Spearman correlation coefficient.

Fibrosis Quantification Algorithm

Fibrosis quantification was performed using a PPC algorithm developed within HistomicksTK. To validate our PPC algorithm we compared it on a sample of 40 WSIs to a commercial PPC algorithm that we used in previous studies (Leica/Aperio Biosystems, Wetzlar, Germany) and assessed the correlation between the two using the Pearson correlation coefficient.

Association Between Morphometric Fibrosis Quantification and Visual Quantification

Each biopsy specimen was scored for interstitial fibrosis in clinical practice and was assigned a percentage of fibrosis and a ci score defined according to the Banff
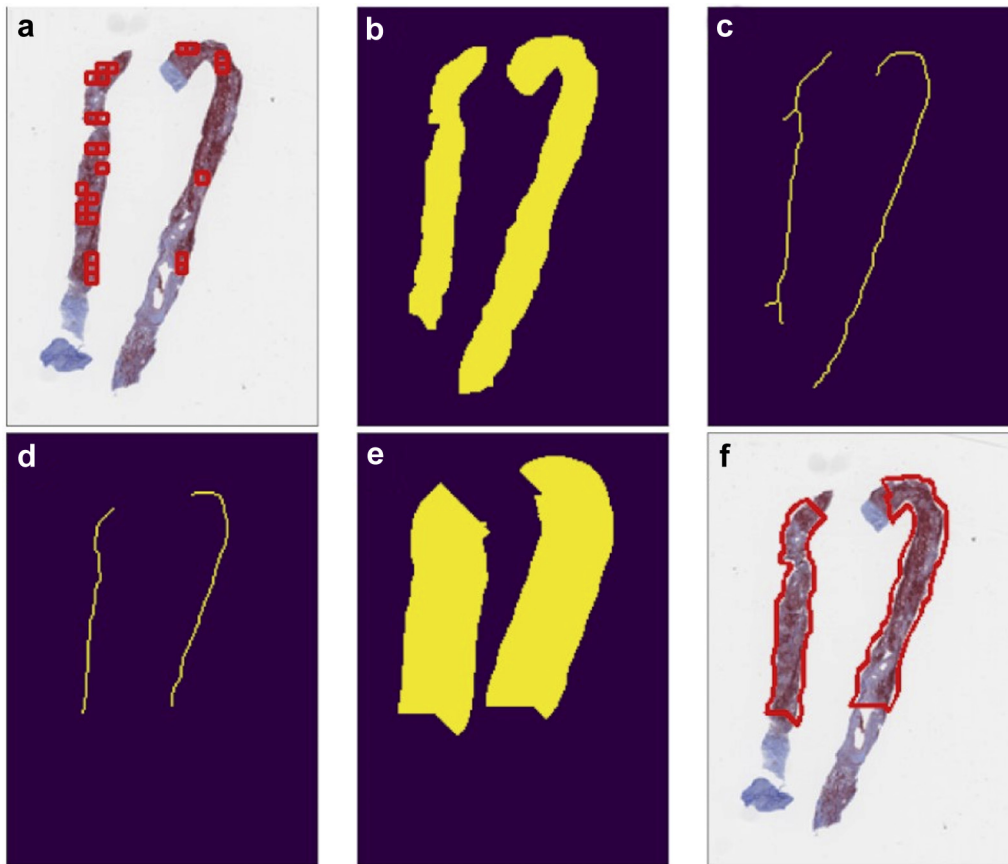
Figure 2. Illustration of the automatic region of interest algorithm.

classification (c_i0, 0%–5%; c_i1, 6%–25%; c_i2, 26%–50%; and c_i3 >50%). To assess intra- and interobserver agreement, a subset of 100 biopsy specimens were reviewed by a single pathologist (22 had previously been analyzed by the same pathologist and 88 had been previously analyzed by 1 of 4 other pathologists) and both the correlation between the continuous fibrosis score (Spearman correlation) and the linearly weighted kappa scores (κ) between the c_i scores were estimated. Similarly, we assessed the agreement between our PPC algorithm result and visual assessment.

Association Between Morphometric Fibrosis Quantification and Graft Function and Outcome

To assess the correlation between graft function and morphometric quantification, we estimated the correlation between the eGFR (estimated using the Chronic Kidney Disease Epidemiology Collaboration formula[16]) at the time of the biopsy procedure and the fibrosis quantification using the Spearman correlation coefficient. The Kaplan–Meier method and log-rank test were used to compare death-censored graft survival by c_i group defined based on morphometric assessment and pathologist scoring. We also used proportional hazard Cox models to assess the association between the percent of fibrosis (continuous evaluation) by both the

pathologists and the morphometric quantifications with death-censored graft survival.

The study was approved by our center's institutional review board (approval IRB00108008). We used *R* software (v. 3.2.1; *R* Core Team, Vienna, Austria) for

Table 1. Patients' characteristics at the time of transplantation and at the time of the first allograft biopsy procedure

Recipients' characteristics	Mean (SD) or n (%)
Age at transplantation, yr	49.6 (13.0)
Age at first biopsy, yr	50.0 (12.9)
Sex	
Male	300 (66.8%)
Female	149 (33.2%)
Race	
African American	260 (57.9%)
White	161 (35.9%)
Other	28 (6.2%)
End-stage kidney disease etiology	
Diabetes	132 (29.3%)
Hypertension	119 (26.5%)
Glomerulonephritis	81 (18.0%)
Other	117 (26.1%)
Donors' characteristics	
Age at donation	39.9 (15.2)
Living donation	142 (31.6%)

R^2: 0.9971 p-value: 5E-64
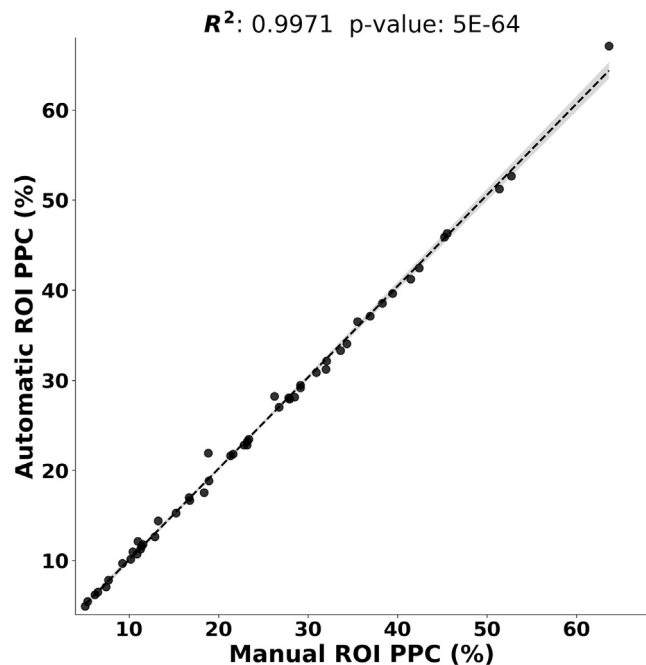
Figure 3. Correlation between automatic fibrosis quantification in manual vs. automatic regions of interest (ROIs) on 51 validation whole slide images. PPC, positive pixel count.

Figure 4. Illustration of the correlation between manual and automatic regions of interest on a validation whole slide image.

all statistical analyses and considered $P < .05$ to be statistically significant; all tests were 2-tailed.

RESULTS

Study Population
A total of 665 trichrome-stained WSIs from 449 individual patients were included in the analysis. Patients' characteristics at the time of transplantation and at the time of the first transplant biopsy procedure can be found in Table 1. Overall, the mean age at kidney transplantation was 49.6 years (SD 13.0 years), 300 were male (66.8%), 260 (57.9%) were African American, and 161 (35.9%) were white. The main causes of end-stage kidney disease were diabetes (131 [29.8%]) and hypertension (119 [27.0%]). One hundred forty-two transplants were from a living donor (31.6%) and the mean donor age was 39.9 years (SD 15.2 years). The mean time between transplantation and first transplant biopsy procedure was 4.1 months (SD 3.1 months).

Glomeruli Detection and Automatic Selection of ROIs
The glomerular detection algorithm was tested on 9 trichrome-stained WSIs including 142 glomerular images and 2189 nonglomerular images. The overall accuracy of the model was good with a sensitivity of 93.0% and a positive predictive value of 81.0% (F1 score 0.87).

The algorithm used to select ROIs was validated on 51 WSIs and showed a high accuracy with an F1 score of 0.83 (SD 0.13). Figure 3 shows the high correlation ($r = 1.00$ [0.99–1.00]) between the PPCs in the manual vs. the automatic ROIs. We also present an image showing both the glomerular detection and ROIs selection on a WSI (Figure 4).

Fibrosis Quantification Algorithm
#### *Correlation Between Visual Fibrosis Quantification by the Pathologists*
One hundred two biopsy specimens were rescored centrally by a single pathologist. Twenty-two had been previously scored by the same pathologist, and intra-observer agreement was high with a correlation 0.85 for the percent of fibrosis and a κ for ci score of 0.73 (0.49–0.97). Interobserver agreement was lower, with a correlation of 0.78 for the percent of fibrosis and a κ for ci score of 0.60 (0.38–0.83).

Correlation Between Automatic Fibrosis Quantification by PPC Algorithms
The PPC algorithm embedded within HistomicsTK was validated against the Aperio PPC algorithm on 42 WSIs. The correlation between the 2 algorithms was high ($r = 0.97$ [0.94–0.98]).

Correlation Between Automatic Fibrosis Quantification by PPC Algorithm and Visual Assessment by Pathologists
The correlation between the fibrosis quantification by the original pathology reading and by the PPC algorithm was moderate, with a correlation coefficient of 0.46 (0.40–0.52). Figure 5 presents the distribution of the fibrosis quantification by the PPC algorithm using ci score, as defined by the pathologists. Overall, 76.5% of the biopsy specimens scored ci1 by the pathologist were given a fibrosis quantification between 6% and 25%, and 40% of those scored ci2 were given a fibrosis quantification between 26% and 50%. The discrepancies were major for the extreme ci score, with only
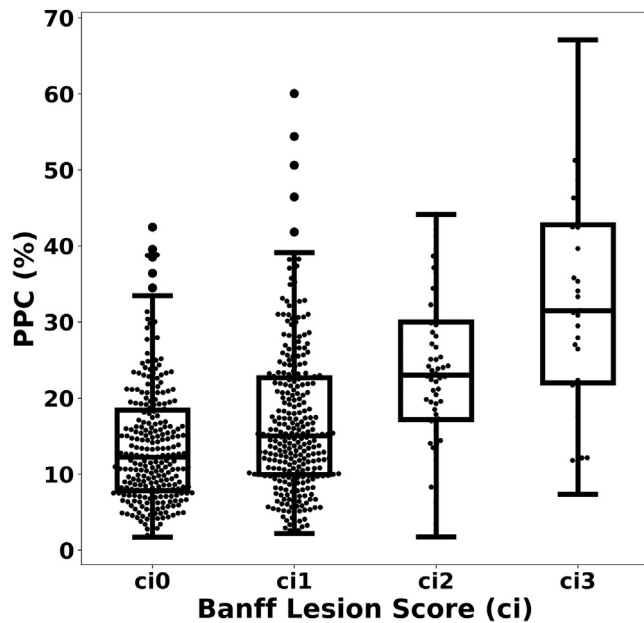
Figure 5. Distribution of the fibrosis quantification by the positive pixel count (PPC) algorithm by ci score as defined by pathologists.

10.3% of ci0 being quantified between 0% and 5% and only 13.3% of ci3 being quantified between 51% and 100%.

No major differences in the correlation between pathologist and automatic quantification were found when restricting the analysis to adequate biopsy specimens per Banff ($r = 0.46$ [0.40–0.53]) or to biopsy specimens without acute rejection ($r = 0.50$ [0.41–0.57]).

Association Between Fibrosis Quantification and Graft Function and Outcome

We found a weak negative correlation between the percent of fibrosis assessed by the pathologists and eGFR at the time of transplant biopsy procedure

($r = -0.14$ [−0.22 to −0.07]) when no significant correlation was found between the automatic quantification of fibrosis and eGFR ($r = -0.07$ [−0.15 to 0.01]; Figure 6).

The median follow-up after the first allograft biopsy procedure was 47 months (interquartile range 29–66 months). We found a significant association between percent of fibrosis assessed by the pathologists and death-censored graft survival (hazard ratio 1.02, $P <$ .001), but the association was not statistically significant between the automatic quantification and the outcome (hazard ratio 1.02, $P = $ .28). Figure 7 shows death-censored graft survival stratified by ci score assessed by pathologists (Figure 7a) and by automatic quantification (Figure 7b). Table 2 presents the 4-year graft survivals by ci score assessed by both visual and automatic fibrosis quantification. No significant differences in graft survival were found between patients classified from ci0 to ci2 by either visual or automatic quantification. Patients classified as ci3 by pathologists experienced poorer outcomes, and the log-rank test reached statistical significance ($P = $ .008). Only 7 patients were classified as ci3 by the automatic quantification—none experienced graft loss, and the log-rank test did not show significant differences in allograft survival by ci score assessed automatically ($P = $ .68).

DISCUSSION

We report the development of a pipeline able to automatically and accurately detect glomeruli and select cortical ROIs on WSIs obtained from Masson trichrome–stained biopsy specimens. We also show that these selected ROIs can then easily be fed into algorithms to perform various tasks, such as fibrosis quantification or structure segmentation.
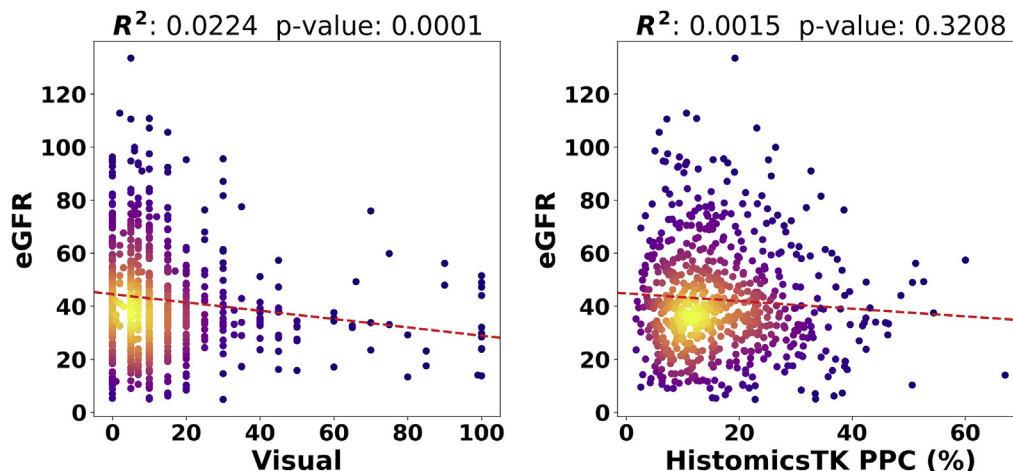
Figure 6. Correlation between visual fibrosis quantification or automatic fibrosis quantification (HTK) and estimated glomerular filtration rate (eGFR) at the time of transplant biopsy procedure. PPC, positive pixel count.
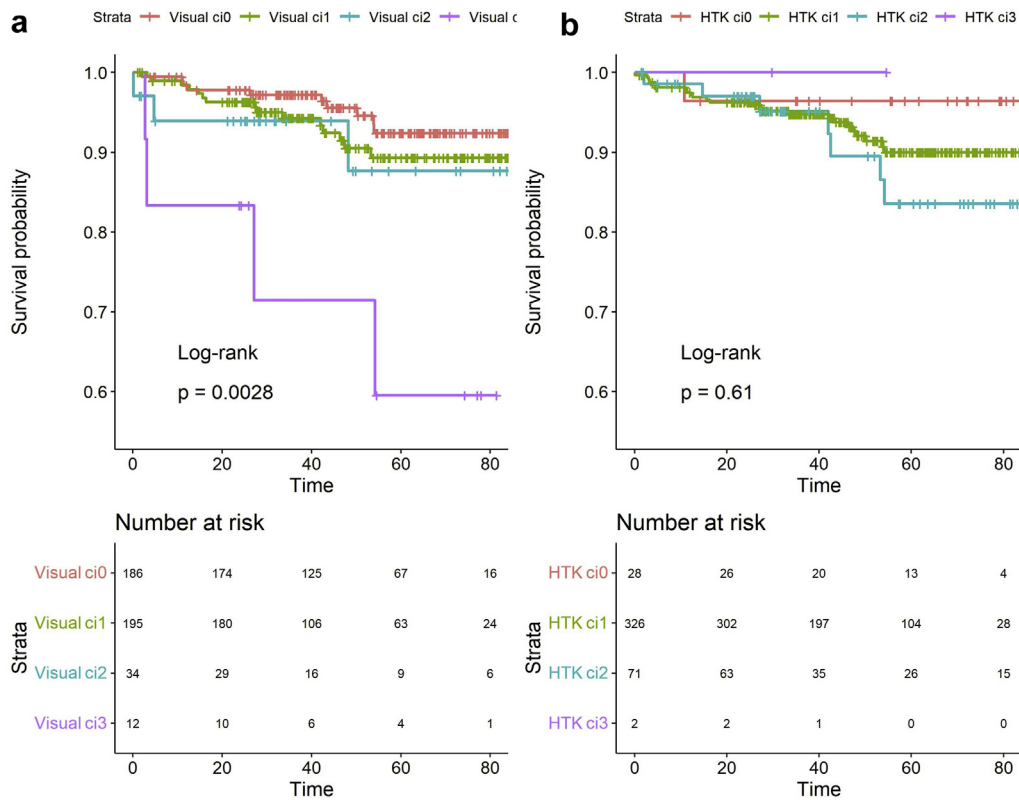
Figure 7. Death-censored graft survival stratified by fibrosis assessed by ci score according to visual pathologist grading (a) and automatic fibrosis quantification (HTK) (b).

The pipeline relies on the automatic detection of glomeruli and the definition of the ROI as the region between the most distant glomeruli. Glomerular detection has consistently been reported as a task that algorithms can be trained to perform with a high level of accuracy. Previously published works have tackled this problem in a multitude of manners, such as using semantic segmentation to delineate the boundaries of glomeruli in periodic acid–Schiff or hematoxylin and eosin–stained mouse and human kidney WSIs.[17] Another group used an image classification approach with overlapping regions to create the glomerular detection algorithm, again on periodic acid–Schiff-stained images.[18] Recently, Bukowy et al.[19] developed a highly accurate region-based convolutional neural network able to detect glomeruli on trichrome-stained WSIs from rats with precision and recall values of 96.9% and 96.8%, respectively. Using a similar method, we were able to accurately detect glomeruli on human kidney transplant biopsy specimens with precision and recall values of 81% and 93%, respectively. The lower precision may be related to the analysis of biopsy specimens instead of nephrectomy specimens. However, this did not significantly impact the accuracy of our ROI detection algorithm as shown by the high concordance between manually defined and automatically selected ROIs.

Several studies have shown that interstitial fibrosis quantification is predictive of renal allograft outcome and may be considered a surrogate marker.[20,21] Indeed, Loupy et al.[22] recently developed a predictive model (iBOX) based on clinical, immunologic, and pathology data that demonstrated high performance in predicting allograft failure at 3, 5, and 7 years posttransplant. Interstitial fibrosis and tubular atrophy were among the pathologic predictors included in the model.[22]

However, the quantification of interstitial fibrosis on biopsy specimens in both native and transplant kidneys is challenging. In our study, we assessed intra- and interobserver agreement for the scoring of interstitial fibrosis and found a relatively good intraobserver agreement with a correlation of 0.85 ($R^2 = 0.72$), consistent with what we previously reported (R^2 0.62–0.90).[10] Interobserver agreement was lower ($r = 0.78$; $R^2 = 0.60$) but within the

Table 2. Four-year death-censored graft survival stratified by ci score estimated by both visual and automatic assessments

ci score	Visual assessment	Automatic assessment
0 (0%–5%)	n = 186, 95% (92%–98%)	n = 28, 96% (90%–100%)
1 (6%–25%)	n = 195, 90% (86%–96%)	n = 326, 92% (89%–96%)
2 (26%–50%)	n = 34, 88% (75%–100%)	n = 71, 89% (80%–99%)
3 (>50%)	n = 12, 71% (48%–100%)	n = 2, no events

highest range compared with previous reports. Indeed, in an international study, the agreement between pathologists for ci score was very low ($\kappa = 0.295$)[3] compared with our study ($\kappa = 0.60$). These results point toward the lack of reliable reference for evaluating new methods to quantify interstitial fibrosis. This probably contributes to the low correlation between visual fibrosis quantification and morphometric assessments. In our study, we found a moderate correlation between pathologist and morphometric quantification ($r = 0.46$ [0.40–0.52]). The correlation was slightly lower than what we found in our previous study with a $R^2 = 0.36$ (compared with $R^2 = 0.21$ in the present study) and may be explained by the inclusion of more diverse diagnoses in the present study, including patients with rejection or other potential complications. Indeed, we found a higher correlation in cases without rejection than in those with acute rejection. It is also interesting to note that the distribution of the percent of fibrosis is much narrower when assessed automatically and that the correlation between visual assessment and morphometric assessment is especially poor in extreme cases with very little or a lot of fibrosis. Although the overall agreement between ci scores from the visual assessment vs. the morphometric assessment was statistically significant but low ($\kappa = 0.08$ [0.04–0.13]), consistent with recent reports using more advanced methods (e.g., Fourier-transform infrared, $\kappa = 0.11$[11]), the discrepancies were major for the extreme ci scores, with only 10.3% of ci0 being quantified between 0% and 5% and only 13.3% of ci3 being quantified between 51% and 100%. This probably reflects the fact that pathologists do not really assess the quantity of fibrosis per se but rather quantify the percent of abnormal cortical tissue. This would explain why, when analyzing a biopsy specimen with abundant fibrosis and tubular atrophy, pathologists can assign 100% fibrosis when the algorithm will only quantify interstitial fibrosis as abnormal and all tubules as nonfibrotic. Conversely, pathologists appear to have a hard time precisely quantifying little amount of fibrosis. It is likely that a threshold effect happens with pathologists quantifying as 0% fibrosis many biopsy specimens with only a little amount of fibrosis, when our algorithm almost always finds some fibrosis. It is also possible that our algorithm could not differentiate some normal structures, such as the glomeruli or the tubular basement membrane, from pathologic fibrosis, systematically overestimating fibrosis especially in cases with very little fibrosis. Our team and others previously reported on the use of special stains, such

as collagen III[10] or Sirius red[10,23] to overcome this limitation, but these stains are not used in routine clinical practice.

Given the lack of a criterion standard for fibrosis quantification, many studies have used the association between fibrosis quantification and allograft function (estimated by serum creatinine or eGFR) or allograft outcome (change in eGFR or graft loss) as a primary outcome to assess the quality of their fibrosis quantification methods. In our study, we did not find a significant association between eGFR at the time of biopsy procedure and fibrosis quantification. Previous studies either found no association or only very weak associations between fibrosis quantification and eGFR.[10,11] This is probably explained by the fact that multiple mechanisms besides fibrosis can impair eGFR. In our cohort including only indication biopsy specimens relatively early after transplantation, it is not surprising that fibrosis, whatever method is used to assess it, only marginally explains kidney function. Servais et al.[7] applied an enhanced color analysis on surveillance biopsy specimens and were able to demonstrate an association between fibrosis quantification and change in fibrosis quantification between 2 surveillance biopsy specimens with change in eGFR, supporting the idea that the association in our study and others is decreased because of the inclusion of allografts with acute complications. Finally, we did find some correlation between ci score assessed by visual assessment and graft survival but not with morphometric assessment of ci score. Grimm et al.[23] reported some correlation between morphometric fibrosis quantification on Sirius red–stained 6-month surveillance biopsy specimens and graft survival, supporting the idea that special staining may provide valuable information to predict graft outcome. In our study, the association between ci score visual assessment and graft survival was driven by an association between ci score 3 and graft loss. This further supports the idea that pathologists tend to assess the total amount of abnormal tissue rather than the precise quantity of fibrosis, potentially explaining the stronger association with graft outcome. In a recent publication, Kolachalama et al.[24] reported the development of a deep neural network algorithm to predict renal survival in patients undergoing both native or allograft biopsy procedures. Their model combines clinical data and trichrome-stained WSI analysis using the commercially available Google Inception algorithm.[24] Although they did not report the performance of a model excluding the image data, they showed that their convolutional neural network model outperformed a model including the same clinical data and a visual scoring of fibrosis. It is important to note that their convolutional neural network model was not trained to quantify

fibrosis but to predict graft survival; therefore, many other predictive features may have been extracted from the WSIs, likely accounting for the better accuracy.

The strength of our study is the development of a pipeline able to accurately select cortical ROIs from WSIs from kidney allograft biopsy specimens, allowing the direct application of image analysis algorithms on large samples of WSIs. The application to fibrosis quantification used to illustrate the use of this pipeline presents several limitations, including the use of early indication biopsy specimens that are less likely to present with severe fibrosis and more likely to present with acute injuries likely to impact the correlation with allograft function. We also acknowledge that the use of specific staining or imaging methods can improve fibrosis quantification and association with allograft function and outcome but are not readily available on large samples of biopsies in most centers, including ours.

In conclusion, we developed a pipeline able to automatically and accurately detect glomeruli and select cortical ROIs on WSIs obtained from Masson trichrome–stained biopsy specimens. We demonstrated that these selected ROIs can easily be used to develop, validate, and apply image analysis algorithms. This pipeline is likely to decrease the need for manual preprocessing of WSIs and to allow the validation of future algorithms on large and unselected batches of WSIs. Our study also supports the need for a more advanced artificial intelligence algorithm to predict kidney allograft outcomes over the development of specialized algorithms aimed at mimicking the Banff classification.

DISCLOSURES

All the authors declared no competing interests.

ACKNOWLEDGMENTS

DATA AVAILABILITY STATEMENT

Python codes of the algorithms are freely available on Github (https://github.com/jvizcar/Fibrosis_Code).

SUPPLEMENTARY MATERIAL

Supplementary File (JPG)

Figure S1. Presentation of the 4 WSIs used as reference for color normalization.

REFERENCES

1. Loupy A, Haas M, Solez K, et al. The Banff 2015 Kidney Meeting Report: current challenges in rejection classification and prospects for adopting molecular pathology. *Am J Transplant*. 2017;17:28–41.

2. Furness PN, Taub N, Convergence of European Renal Transplant Pathology Assessment Procedures (CERTPAP) Project. International variation in the interpretation of renal transplant biopsies: report of the CERTPAP Project. *Kidney Int*. 2001;60:1998–2012.

3. Furness PN, Taub N, Assmann KJM, et al. International variation in histologic grading is large, and persistent feedback does not improve reproducibility. *Am J Surg Pathol*. 2003;27:805–810.

4. Farris AB, Chan S, Climenhaga J, et al. Banff fibrosis study: multicenter visual assessment and computerized analysis of interstitial fibrosis in kidney biopsies. *Am J Transplant*. 2014;14:897–907.

5. Farris AB, Moghe I, Wu S, et al. Banff Digital Pathology Working Group: going digital in transplant pathology. *Am J Transplant*. 2020;20:2392–2399.

6. Moon A, Smith GH, Kong J, Rogers TE, Ellis CL, Farris ABB. Development of CD3 cell quantitation algorithms for renal allograft biopsy rejection assessment utilizing open source image analysis software. *Virchows Arch Int J Pathol*. 2018;472:259–269.

7. Servais A, Meas-Yedid V, Noël LH, et al. Interstitial fibrosis evolution on early sequential screening renal allograft biopsies using quantitative image analysis. *Am J Transplant*. 2011;11:1456–1463.

8. Hermsen M, de Bel T, den Boer M, et al. Deep learning-based histopathologic assessment of kidney tissue. *J Am Soc Nephrol*. 2019;30:1968–1979.

9. Trimarchi H, Barratt J, Cattran DC, et al. Oxford classification of IgA nephropathy 2016: an update from the IgA Nephropathy Classification Working Group. *Kidney Int*. 2017;91:1014–1021.

10. Farris AB, Adams CD, Brousaides N, et al. Morphometric and visual evaluation of fibrosis in renal biopsies. *J Am Soc Nephrol*. 2011;22:176–186.

11. Vuiblet V, Fere M, Gobinet C, Birembaut P, Piot O, Rieu P. Renal graft fibrosis and inflammation quantification by an automated Fourier-transform infrared imaging technique. *J Am Soc Nephrol*. 2016;27:2382–2391.

12. Gutman DA, Khalilia M, Lee S, et al. The Digital Slide Archive: a software platform for management, integration and analysis of histology for cancer research. *Cancer Res*. 2017;77:e75–e78.

13. Reinhard E, Adhikhmin M, Gooch B, Shirley P. Color transfer between images. *IEEE Comput Graph Appl*. 2001;21:34–41.

14. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. *ImageNet: a large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition*. New York, NY: Institute of Electrical and Electronics Engineers; 2009:248–255 pp.

15. Dijkstra EW. A note on two problems in connexion with graphs. *Numerische Mathematik*. 1959;1:269–271.

16. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150:604–612.

17. Lutnick B, Ginley B, Govind D, et al. An integrated iterative annotation technique for easing neural network training in medical image analysis. *Nat Mach Intell*. 2019;1:112–119.

18. Gallego J, Pedraza A, Lopez S, et al. Glomerulus classification and detection based on convolutional neural networks. *J Imaging*. 2018;4:20.

19. Bukowy JD, Dayton A, Cloutier D, et al. Region-based convolutional neural nets for localization of glomeruli in trichrome-stained whole kidney sections. *J Am Soc Nephrol*. 2018;29:2081–2088.

20. Grimm PC, Nickerson P, Gough J, et al. Quantitation of allograft fibrosis and chronic allograft nephropathy. *Pediatr Transplant*. 1999;3:257–270.

21. Cosio FG, Grande JP, Wadei H, Larson TS, Griffin MD, Stegall MD. Predicting subsequent decline in kidney allograft function from early surveillance biopsies. *Am J Transplant*. 2005;5:2464–2472.

22. Loupy A, Aubert O, Orandi BJ, et al. Prediction system for risk of allograft loss in patients receiving kidney transplants: international derivation and validation study. *BMJ*. 2019;366:l4923.

23. Grimm PC, Nickerson P, Gough J, et al. Computerized image analysis of Sirius Red-stained renal allograft biopsies as a surrogate marker to predict long-term allograft function. *J Am Soc Nephrol*. 2003;14:1662–1668.

24. Kolachalama VB, Singh P, Lin CQ, et al. Association of pathological fibrosis with renal survival using deep neural networks. *Kidney Int Rep*. 2018;3:464–475.