Automated renal histopathology: digital extraction and quantification of renal pathology

Pinaki Sarder, Brandon Ginley, John Tomaszewski

Pinaki Sarder, Brandon Ginley, John E. Tomaszewski, "Automated renal histopathology: digital extraction and quantification of renal pathology," Proc. SPIE 9791, Medical Imaging 2016: Digital Pathology, 97910F (23 March 2016); doi: 10.1117/12.2217329

SPIE.

Event: SPIE Medical Imaging, 2016, San Diego, California, United States

Automated renal histopathology: Digital extraction and quantification of renal pathology

Pinaki Sarder[+,]*, Brandon Ginley[+], John E. Tomaszewski

Pathology and Anatomical Sciences, University at Buffalo – the State University of New York, Buffalo, USA.

[+]These authors contributed equally.

* indicates corresponding author. Email: pinakisa@buffalo.edu

ABSTRACT

The branch of pathology concerned with excess blood serum proteins being excreted in the urine pays particular attention to the glomerulus, a small intertwined bunch of capillaries located at the beginning of the nephron. Normal glomeruli allow moderate amount of blood proteins to be filtered; proteinuric glomeruli allow large amount of blood proteins to be filtered. Diagnosis of proteinuric diseases requires time intensive manual examination of the structural compartments of the glomerulus from renal biopsies. Pathological examination includes cellularity of individual compartments, Bowman's and luminal space segmentation, cellular morphology, glomerular volume, capillary morphology, and more. Long examination times may lead to increased diagnosis time and/or lead to reduced precision of the diagnostic process. Automatic quantification holds strong potential to reduce renal diagnostic time. We have developed a computational pipeline capable of automatically segmenting relevant features from renal biopsies. Our method first segments glomerular compartments from renal biopsies by isolating regions with high nuclear density. Gabor texture segmentation is used to accurately define glomerular boundaries. Bowman's and luminal spaces are segmented using morphological operators. Nuclei structures are segmented using color deconvolution, morphological processing, and bottleneck detection. Average computation time of feature extraction for a typical biopsy, comprising of ~12 glomeruli, is ~69 s using an Intel(R) Core(TM) i7-4790 CPU, and is ~65X faster than manual processing. Using images from rat renal tissue samples, automatic glomerular structural feature estimation was reproducibly demonstrated for 15 biopsy images, which contained 148 individual glomeruli images. The proposed method holds immense potential to enhance information available while making clinical diagnoses.

Keywords: Proteinuria, glomerulus, computational pathology, Gabor analysis, principal component analysis, morphological processing,

1. INTRODUCTION

The economic burden resultant of renal disease, and the treatment thereof, is massive. According to the National Institute of Health, the Medicare program spends approximately $24 billion per year to care for over 525,000 patients with end-stage renal failure.[1] Renal diseases which cause abnormal quantities of protein to be filtered from the blood and excreted in the urine are deemed proteinuric. Proteinuria is a subtle indicator of a much larger underlying issue, structural damage within the nephron, the most basic functional unit of the kidney. Specifically, it has been shown that the structural damage which causes proteinuria begins in the glomerulus, a dense region of capillaries at the beginning of the nephron which performs the initial protein filtration from blood plasma.[2-8] A broad schematic of the structural hierarchy in renal anatomy is shown in Fig. 1, and cell types which inhabit the glomerulus are illustrated. In histology, renal pathological markers used to diagnose disease are highly diverse and heterogeneous; accurate diagnosis of the type of renal disease causing proteinuria relies on very time intensive manual qualification/quantification by a trained specialist. This time is further prolonged due to concerns over intra-rater and inter-rater reliability, which magnifies time to diagnosis. A large time to diagnosis leads to a degradation in the efficacy of potential treatments, and reduces the chances of diseased kidney tissue to make a functional recovery.

Computational, histological quantification, on the other hand, holds the potential to massively decrease the required time and manual effort required to accurately diagnose renal disease. Quantifying renal tissue via computational histopathology will allow clinicians to make a stronger, more informed, more rapid decision on the underlying disease and treatment path for patients exhibiting proteinuria. Such a tool would hold massive clinical potential and improve the quality of renal healthcare. It has been shown that the earlier a renal disease is diagnosed and treated, the higher the chances for

Medical Imaging 2016: Digital Pathology, edited by Metin N. Gurcan, Anant Madabhushi, Proc. of SPIE Vol. 9791, 97910F · © 2016 SPIE
CCC code: 1605-7422/16/$18 · doi: 10.1117/12.2217329

renal tissue to make a functional recovery.[9, 10] In turn, computational renal pathology has the potential to reduce the economic burden of end-stage renal healthcare in the United States. We have developed an automated method capable of estimating pathologically relevant glomerular features from histologically stained renal tissues of a normal healthy rat model. The field-of-view examined by the automated method is designed to mimic that of a renal needle biopsy commonly performed by pathologists. Our method uses tools from image and signal processing[11] to robustly solve a problem for which there exists no automated solution to date.

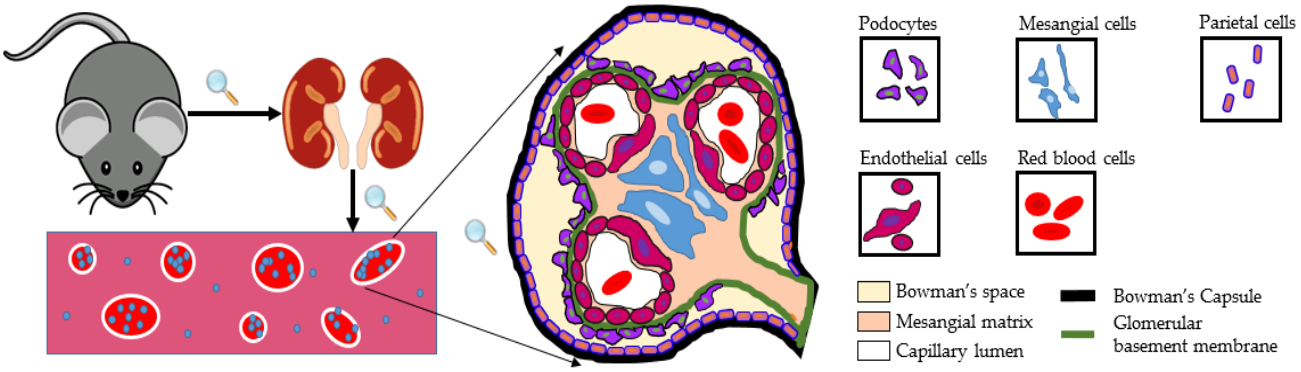


Figure 1. Schematic overview of the data extraction pipeline used to estimate pathologically relevant renal features from renal rat models. Rats are sacrificed, and kidneys are harvested. Slices of 3 μm thickness are cut from kidney, stained, fixed, and mounted to a microscope slide. Examination of the microscopic structure of a glomerulus shows a diverse anatomy, including podocytes, mesangial cells, parietal epithelium, endothelial cells, red blood cells, mesangial matrix, glomerular basement membrane, capillary loops, Bowman Space, and luminal space. Basic glomerular renal anatomy has been illustrated.[12]

Our automated method can be defined by the following extraction pipeline: Glomerulus approximation from needle biopsy mimicking field-of-view, definition of individual glomerular compartments, followed by segmentation and quantification of intra-glomerular compartments. Compartments which can be segmented by our method include cell nuclei and Bowman's and luminal spaces. Our results of this segmentation show high accuracy while also significantly reducing examination time. In regards to biopsy glomerular compartment approximation, average accuracy of estimating glomeruli locations is 88%, this computation performs 60X faster than manual examination, and is scored on n = 15 biopsy images. Definition of intra-glomerular space performs with an average 86% accuracy, the computation is nearly 20X faster than manual, and is scored on n = 148 glomerulus images. Nuclei segmentation performs with an accuracy of 92%, the computation is over 80X faster than human driven examination, and is analyzed for n = 50 glomerulus images. The percentage accuracy above is determined against manual segmentation results. The overall pipeline requires ~69 s of computational time, on average, in an Intel® Core™ i7-4790 CPU @ 3.60GHz desktop computer.

The following sections will explicate further on sample preparation and imaging, the exact steps of our computational method, major results, discussion on the importance of these results, and concluding statements.

2. RESULTS

In this section, we first present our proposed method to segment individual glomeruli from needle biopsies. We then present how to segment the boundaries and important glomerular micro-compartments, such as Bowman's and luminal spaces and resident cell nuclei, from individual glomeruli. Except for the Bowman's space and luminal space segmentation, the method was evaluated in 15 renal needle biopsies of a normal healthy rat model, including 148 individual glomeruli, stained by periodic acid–Schiff (PAS). Bowman's and luminal space segmentation was evaluated on 300 individual glomeruli images from normal healthy murine renal tissue, stained by hematoxylin and eosin (H&E).

2.1. Estimating glomeruli locations from needle biopsies

In order to mimic the field-of-view considered by pathologists in needle biopsies, images with ~10-15 glomeruli are cropped from whole slide image of kidney, similar to the image shown in Fig. 2A. The biopsies are stained using PAS. Typical biopsy sized images, uncompressed for maximum resolution, are approximately 7000 X 4000 pixels, spanning 1.75 X 1 mm^2.

Glomeruli regions have densely packed, high intensity nuclei when examined in grayscale, as opposed to the tubular regions, whose nuclei are sparser. This observation is shown in Fig. 2B, after transforming the original image to

normalized grayscale. This operation is parameter free. Based on our observation, thresholding for nuclei and Gaussian blurring of the resulting image create an approximate glomerular heat map, from which glomeruli locations can be estimated; see Fig. 2C. The nuclei thresholding step requires one parameter which is not constant, and can take on values of 0.5 or 0.6. Small noisy regions which are not nuclei are eliminated from the image by thresholding via object area, using an area threshold parameter = 100 pixels. Smoothing of nuclei is accomplished using a Gaussian filter, with standard deviation = 8 pixels. These area threshold and Gaussian filter parameters are constant for all samples. Note that color deconvolution method[13] can also be used to obtain an intensity image similar to Fig. 2B, with densely packed nuclei. Such image can also lead to an image similar to what is shown in Fig. 2C, from which glomeruli locations can be estimated. If this color deconvolution based route is followed, nuclei thresholding parameter, as described above, can be made constant. This latter procedure however requires longer computation time. Upon obtaining Fig. 2C, with high intensity regions corresponding to glomeruli locations, a second intensity threshold parameter, which takes on values between 0.09 and 0.12, is used to locate the high intensity regions corresponding to glomeruli locations. The resulting image is further processed using morphological operators[11] to increase the accuracy of segmenting glomeruli locations.

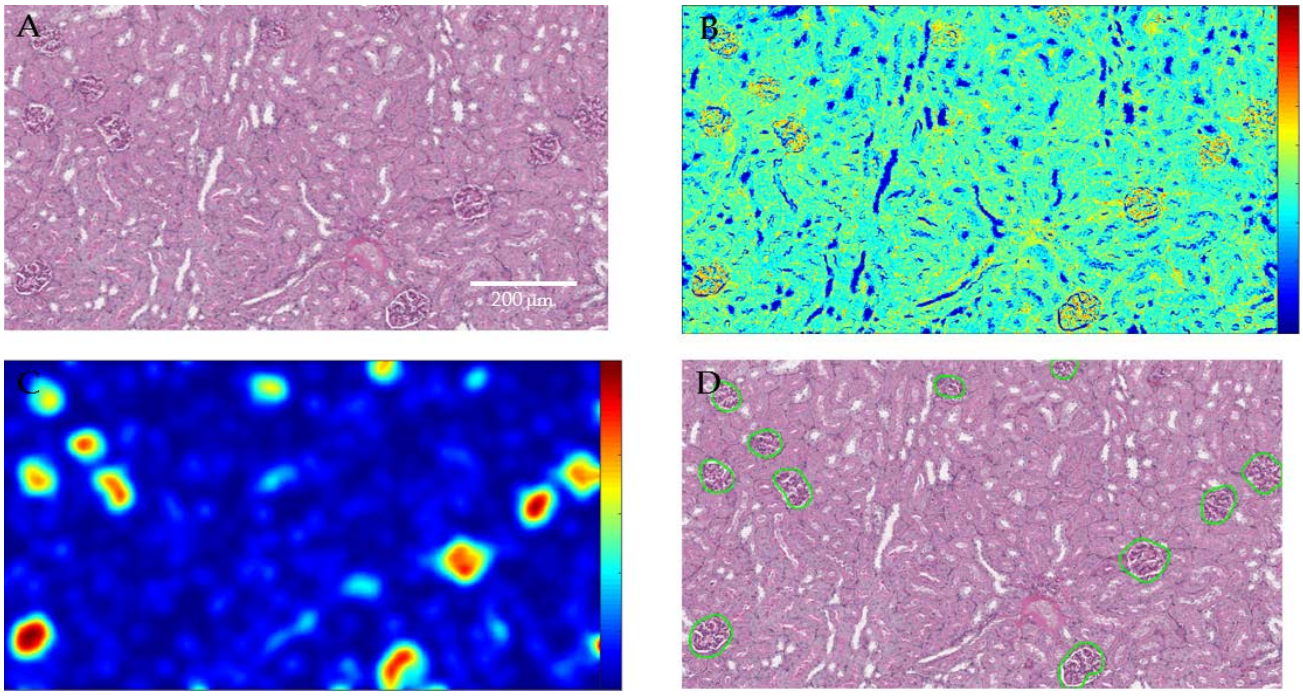


Figure 2. Image processing pipeline demonstrating the proposed method to estimate locations of glomeruli within a field-of-view mimicking that of a needle biopsy performed in clinic. (A) Original starting image (6844 X 3751 pixels) mimicking needle biopsy. (B) Grayscale intensity distribution of the pixels comprising the true color (RGB) image. Nuclei (high red intensity) show high clustering within glomeruli regions as compared to tubular regions. (C) Grayscale intensity image is blurred using a Gaussian filter. Regions of high intensities correspond to approximate glomeruli locations. Morphological processing is applied to the filtered image, to approximate the glomeruli boundaries shown in (D). At each approximate location, 250 pixels are cropped outwards in all directions, based on prior knowledge of average glomerulus size.

Overall, the glomeruli location estimation step contains two intensity thresholding parameters which need tuning. This step was tested for n = 15 samples. On average, this step of the computation requires 4 s to extract most of the glomeruli from needle biopsies. Use of color deconvolution method,[13] by replacing the grayscale transformation, as well as by eliminating the first nuclei threshold parameter, will increase the total average time of computation presented herein by 10 s, with similar performance in estimation.

To complete the extraction of individual glomeruli, a rough bounding box around each identified glomerulus is employed. A bounding box is the smallest rectangle of image which can completely contain a certain region of interest. In our method, this is accomplished by isolating each approximate glomerulus region, and extending the boundaries of that region 250 pixels in all directions. This method captures most of the glomeruli, aside from those samples which have areas that are several deviations above the average glomerulus area. Individual cropped glomeruli images are saved and analyzed further as discussed below.

2.2. Estimating glomerular structural compartments from color knowledge

The glomerular structural compartments discussed herein are embedded in true color (RGB) information. The red, green, and blue components of a single glomerulus are shown in Fig. 3B-D, as well as the true color image for comparison, in Fig. 3A. High intensity pixels in the red component depict nuclei locations to a greater degree than other structures. Similarly, high intensity pixels in the green component depict Bowman's and luminal compartments, and high intensity pixels in the blue component depict extracellular matrix. These structures, when examined in the original true color image, are significantly more difficult to isolate computationally. However, individual color components are still contaminated with high levels of spatially heterogeneous noise. Segmenting structures from these true color components thus requires high levels of unstable parameterization.

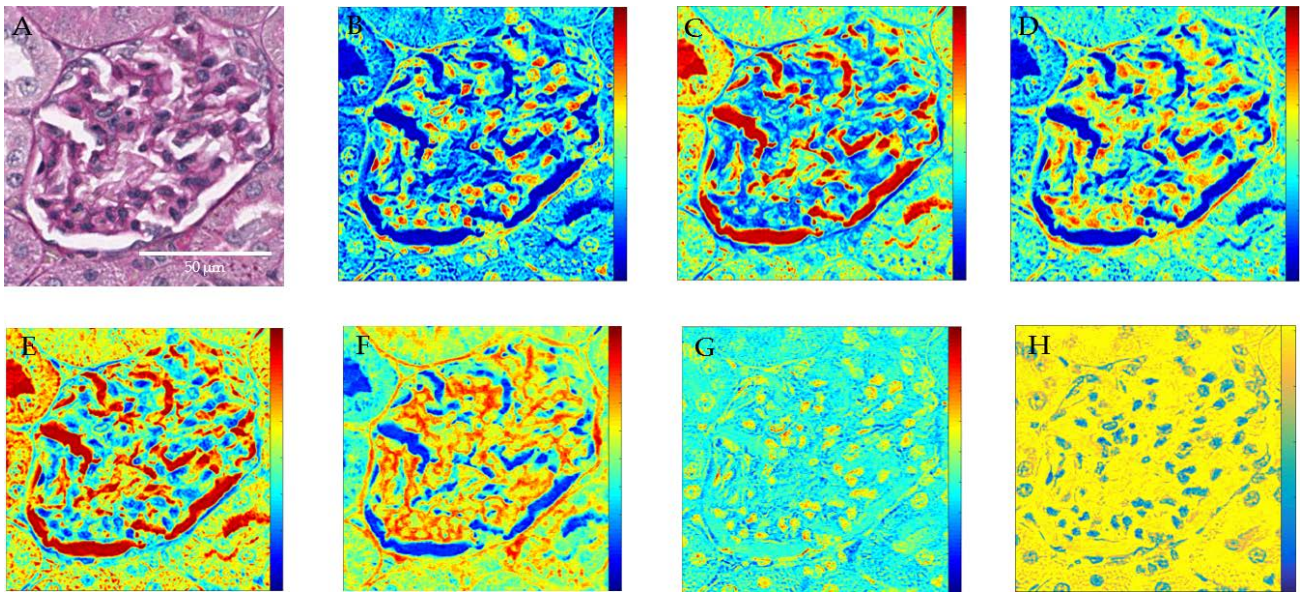


Figure 3. Principal component analysis (PCA) and color deconvolution delineate important glomerulus structures, which are not readily discernable by the human eye from individual true color images. (A) Individual glomerulus image in true color. (B-D) Red, green, and blue components of the true color information, respectively. (E-G) High intensity regions in the respective principal components increase the contrast of compartments such as the Bowman's and luminal spaces (E), extracellular matrix (F), and nuclei locations (G). (H) Low intensity regions of one of the color deconvolved output images approximate nuclei locations while simultaneously eliminating nuclei-irrelevant information.

Principal component analysis (PCA)[14] and spectral color deconvolution methods[13] offer parameter-free alternatives to further isolate structural components. The three principal components of the original true color data are shown in Fig. 3E-G. High intensity pixels of the three principal components allow to better discern the Bowman's and luminal space (3E), extracellular matrix (3F), and nuclei information (3G) contrasts, respectively, while reducing a large portion of background noise.

Spectral color deconvolution also offers several advantages over the native color components. Spectral deconvolution uses knowledge of the staining practice used to separate components by projecting the image in the color space defined by the individual stains. An example of one of the components returned by spectral deconvolution is shown in Fig 3H. Low intensity pixels in this image delineate nuclear contrast. We use this color deconvolved image for segmenting the nuclei contrast as described in the next section.

The result described here holds reproducibly similar for any PAS stained single glomerulus image.

2.3. Segmenting individual glomerulus boundary

Examining Fig 4A, individual glomerular compartments cannot be automatically quantified unless an exact glomerular boundary is first defined. Gabor filtering textural analysis[15] is a very convenient solution to this problem, as the intra-glomerular space is more textural than extra-glomerular regions. Gabor filtering involves filtering textural information from an image using a series of filters with varying orientations and frequencies, modulated by a two-dimensional Gaussian

envelope. The orientation and frequency parameters are constant. We were able to segment n = 144 out of n = 148 total glomeruli samples with high accuracy, as compared to manual annotation. Result is presented in Fig. 4B-C. Four samples which we were not able to segment well using Gabor analysis depict staining variation and are noisy. Histological staining can often be imperfect, and some tissues have the possibility to become stained darker than the others. If an individual glomerulus image contains both lighter stain variation and much darker stain variation, Gabor analysis will segment the glomerulus along one side of the line describing the staining difference.

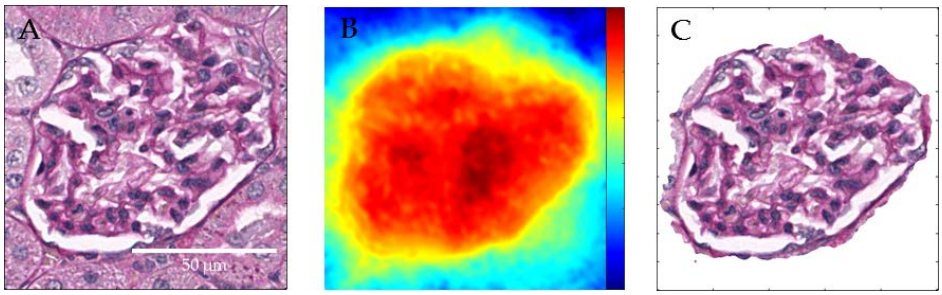


Figure 4. Automatic segmentation of glomerulus boundary from histopathological images. Example is shown using (A) a periodic acid-Schiff (PAS) stained image. Extent of the PAS stained glomerulus can be automatically segmented using Gabor texture analysis. (B) Intensity image of a component of the Gabor texture. (C) Segmented glomerulus.

2.4. Segmenting Bowman's and luminal spaces

We next discuss segmentation of Bowman's and luminal spaces from single glomerulus images. These structural features are important to differentiate disease states from normal healthy states during pathological examination. We found that these structures are more prominent in H&E stained glomeruli than in PAS stained glomeruli. We present here a method to segment these structures using morphological processing. Because of the unavailability of the H&E stained tissue of the rat model, which was used for the other glomerular structural feature estimation presented in this manuscript, we used H&E stained renal tissue slice of a normal healthy mouse model to present our proof-of-concept. The Bowman's and luminal spaces appear to be the white area inside the glomerulus, as can be seen in Fig. 5A. Here Bowman's space is the white area that surrounds the glomerulus, and the rest of the small white spaces are luminal spaces. These regions are extracted after intensity thresholding of the normalized green component of the true color H&E image, where Bowman's and luminal spaces exhibit high intensity. This thresholding parameter can be determined automatically using Otsu's method for grayscale thresholding,[16] or can be decided manually. Manual threshold decision can improve the accuracy of the algorithm in some samples, but also increases the number of tunable parameters for the overall segmentation pipeline. Note that even in the case of manual threshold decision, the threshold parameter requires little tuning and does not vary greatly between samples. Intensity threshold value generally lies at 0.7 or 0.8 for the substantial number of H&E stained glomeruli tested herein. After intensity segmentation, there also exists another parameter, an area threshold, which was used to distinguish luminal areas from the Bowman areas. This parameter is highly constant in samples with strongly defined Bowman's space (Fig. 5A), at 2500 pixels, but is more variable for samples with poorly defined Bowman space. For Bowman's space segmentation, the glomerular boundary can be first defined via Gabor filtering, or by parametric ellipse fitting of Bowman's space segments. Once the glomerular boundary is defined, any white areas which are not attached to the outer Bowman's space can be segmented as luminal spaces (Fig. 5C). This method was tested for n > 300 individual H&E images, with typical computational speed ~0.01 s per glomerulus.

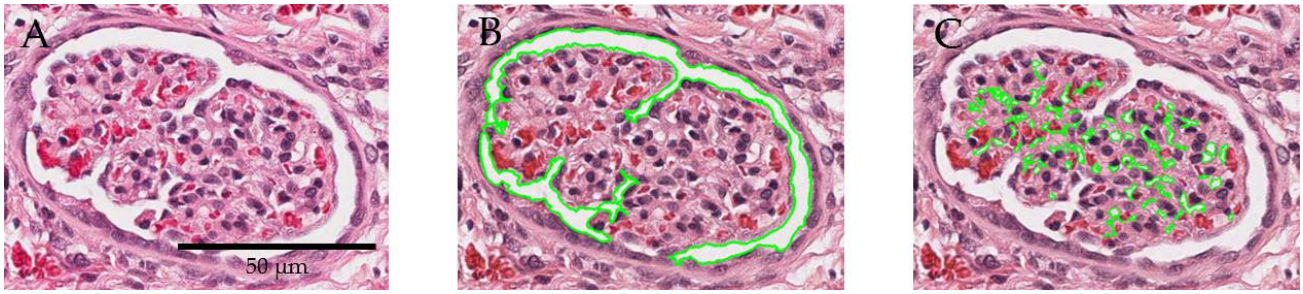


Figure 5. Automatic segmentation of Bowman's and luminal spaces using hematoxylin and eosin (H&E) stained glomeruli. Note that only this result in this manuscript has been produced using murine renal tissue. (A) Original glomerulus image of an H&E stained image. (B) Bowman's space and (C) luminal space segmented using morphological processing.

2.5. Segmenting glomerular nuclei

Perhaps the most important pathological feature examined by our method is the nuclei count. Several research groups have explored this area, yet there exists no universal method for nuclear segmentation from histopatological images.[17] For the PAS stained renal tissue samples from rat, our method segments glomerular nuclei via color deconvolution,[13] morphological processing,[11] and bottleneck detection.[18] Our nuclei segmentation method involves three major steps, which include five parameters, and one optional step, which includes two parameters. We tested the segmentation using 50 individual glomeruli images obtained using the single glomeruli segmentation method presented in Section 2.1 for the PAS stained rat tissue sample. The five parameters from first three major steps were found to be constant for these samples. Fig. 6A shows a single glomerulus with multiple nuclei and Fig. 6B shows the segmentation result.

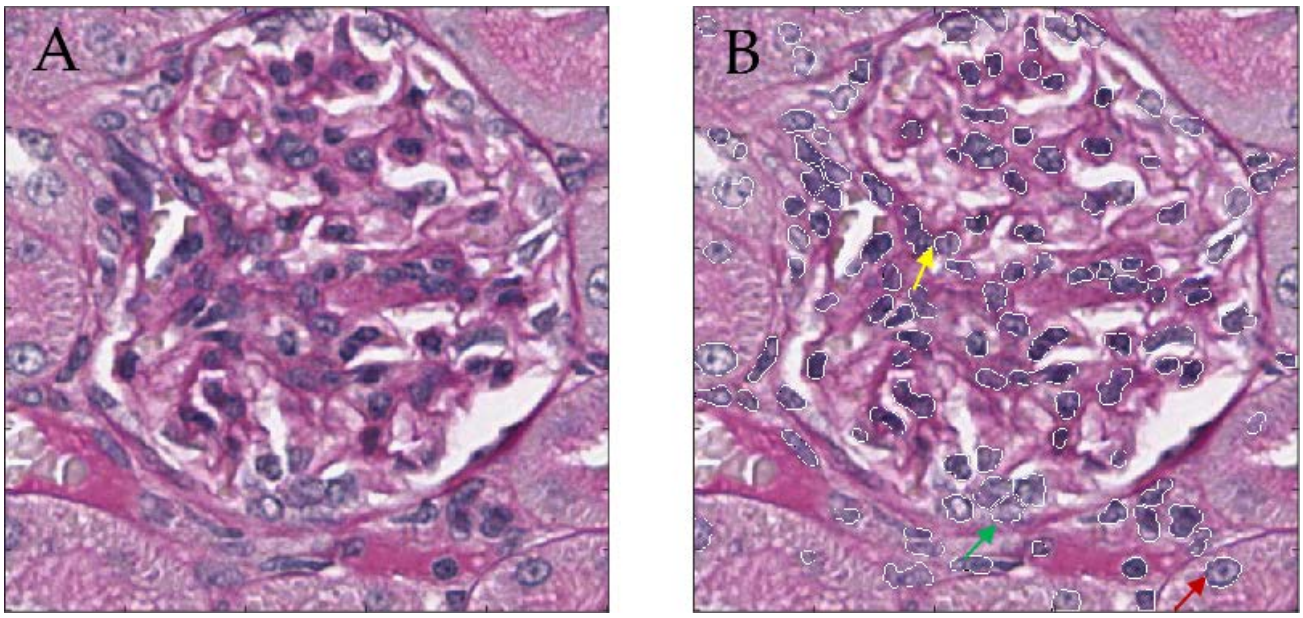


Figure 6. Automatic segmentation of nuclei from an individual glomerulus. (A) Original image of a periodic acid-Schiff stained glomerulus. (B) Segmented nuclei. The single nuclei that have area bounded by a minimal and a maximal threshold are clustered using morphological processing. The nuclei which are either clustered together or have area bigger than the maximal threshold described above, are typically similar to the ones shown using red, green, and yellow arrows. The nucleus shown using the red arrow is a single nucleus with area larger than the maximal threshold described above. The clustered nuclei shown using yellow arrow has a bottleneck that separates at least one single nucleus. The green arrow depicts three nuclei joined together, which have visible bottlenecks, in addition with small white regions visible in each single nucleus.

The first step of the nuclei segmentation method employs the color deconvolution method to obtain nuclei contrast enhanced image similar to the one shown in Fig. 3H. The image is normalized in 0-1 scale, and inverted to obtain an image with nuclei locations having higher intensities. An intensity threshold of 0.15 selects the tentative nuclei spots. Regions with area above 100 pixels are selected. A morphological opening using a disk of 5 pixel radius is used to smooth nuclei boundaries, separate single nuclei, and eliminate small noisy nuclei. Single nuclei are selected from the resulting image as the ones that have area smaller than 550 pixels.

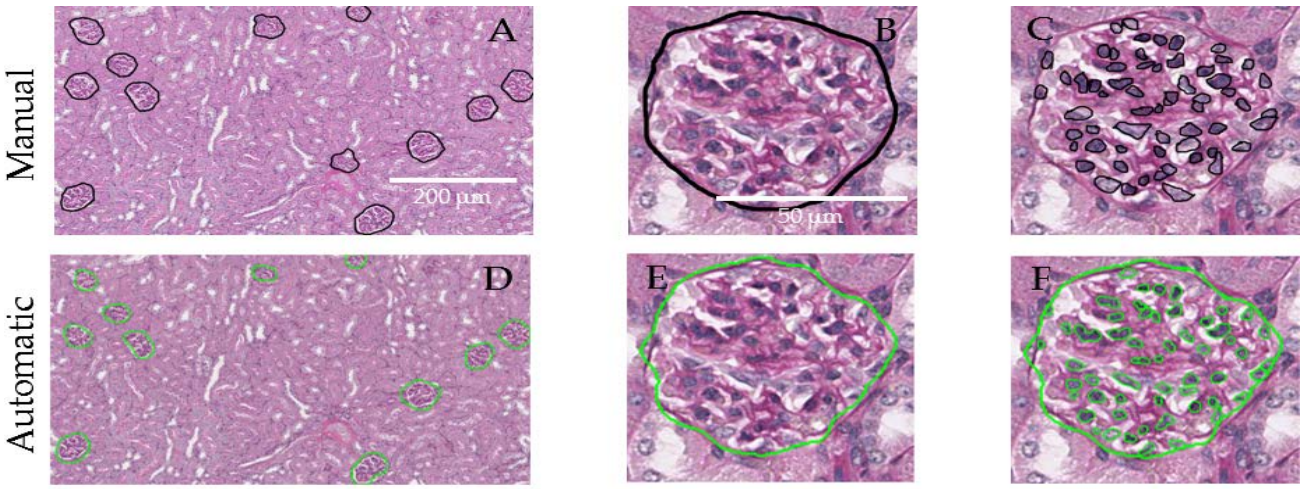
From the second step onwards, we use various techniques to segment the clumped nuclei from the nuclei regions that are not declared as single nuclei in the first step. We first check if the area of a nuclei segment in this step is at least covering a fraction of 0.8 of the convex area generated from this segment. One such example is shown using red arrow in Fig. 6B. The other cluttered nuclei are examined to see if a single nucleus is attached with the rest of the segment via a single pixel. In such a case, the single nucleus is segmented by disconnecting the single pixel. Each of the resulting segments is declared to be a single nucleus if its area lies between 100 to 550 pixels or if its area covers at least a fraction of 0.8 of the convex area formed by it.

In the next step, we use bottleneck detection method[18] to separate single nuclei from the rest of the nuclei segments. This method separates nuclei clumps where two convex shapes are attached in a bottleneck. One example is shown using a yellow arrow in Fig. 6B. Estimating and breaking the structures along the two points that best define the bottleneck isolation of single nuclei from many clumped nuclei. Like the previous step, upon segmenting the clumped

nuclei here, we declare each of the resulting segments to be a single nucleus if its area is measured to be in between 100 to 550 pixels or if its area constitutes at least 0.8 fraction of the convex area formed by this segment. This step is parameter free.

The rest of the segments typically show three or more nuclei clumped together, with each nucleus touching more than one nucleus; see the region pointed by the green arrow in Fig. 6B. Separating these nuclei is not possible by any of the methods described above. In order to separate them, we present here an optional step, performance evaluation of which is beyond the scope of this proceedings manuscript. We observed that the clumped nuclei discussed herein depict small clustered white spots in their nuclei. These white spots are segmented, and smoothened by a Gaussian filter to obtain one intensity peak per nucleus. These peaks are used as starting centroids to segment the single nuclei from the whole nuclei clump using k-means clustering.[11] The Gaussian smoothing step involves two parameters. The first one is an area threshold defining the minimal area of the nuclei clumps that can go through this optional step. The second one is the standard deviation of the Gaussian filter described in this step. Both of these parameters were found to be constant for the samples analyzed for nuclei segmentation.

2.6. Performance analysis



Segmentation method	Speed (s)		Accuracy (%)	
	Manual	Automatic	Manual	Automatic
Glomerular location estimation (A, D)	298 ± 83	4.99 ± 1.3	100	87.8 ± 9.4
Glomerular boundary segmentation (B, E)	18.1 ± 6.7	1.08 ± .24	100	86.0 ± 11
Nuclei segmentation per glomerrulus (C, F)	332 ± 87	4.27 ± .89	100	92.3 ± 13

Figure 7. Comparison of computational method to manual segmentation. Samples of automatic and manual segmentation are shown for each step in the glomerular estimation process. Table shows comparison of speed and accuracy of both methods, with manual segmentation result is used to be the ground truth for accuracy metrics. Quantification is represented as mean value ± standard deviation. Glomerular location estimation, glomerular boundary segmentation, and nuclei segmentation are quantified for n = 15 renal biopsies, n = 148 single glomerulus segmented from these biopsies, and n = 50 glomeruli taken from these single glomeruli, respectively. Though the method is not 100% accurate, computational estimation can be performed 20-100X faster than the manual estimation.

Our overall segmentation, on average, performs quantification of select renal features, as discussed in Fig. 7, from renal biopsies 65X faster than manual segmentation. Automatic method took an average of ~69 s to completely segment and quantify select glomerular compartments. A summary of computation times is available in Fig. 7. The shortest segmentation is glomerular boundary estimation. This is a result of the fact that there is only one object to segment, and image sizes are much smaller than biopsy images. The second shortest step is glomerular location estimation, which requires a large amount of time due to the massive image sizes and many structures to segment. Nuclei segmentation is

the longest step, resultant of a high number of structures which need to be segmented, in addition to fuzzy definition of nuclear boundaries.

To compute accuracy of our method, n = 15 renal biopsies, n = 148 single glomerulus segmented from these biopsies, and n = 50 glomeruli taken from these single glomeruli were manually annotated, and taken as ground truth (Fig. 7A-C). Examples of automatically annotated masks can be seen in Fig. 7D-F. Individual glomeruli images were scored as a function of overlap between computationally obtained segments and manual masks. For each sample, manual masks were overlaid with automatically obtained segments, and the area of the region of overlap was computed. Percentage of this overlap with respect to manual mask area was computed to define accuracy. Glomerular location estimation and nuclei segmentation results were scored on a hit-or-miss basis compared to manual. Accuracy, for these images, is defined as number of automatically labeled structures divided by total number ground truth structures. Overlap of manual mask and automatically segmented regions was not considered here, due to the tendency of a manual hand to overshoot true structure boundaries. Computer mouse devices are not precise enough to exactly segment structures on a pixel by pixel basis.

Note that we discard the performance accuracy of Bowman's space and luminal space segmentation here for consistency. This is because, for segmenting these structures, we only present in this paper our proof-of-concept using murine renal tissue images, and did not use the rat renal tissue slides as used for the rest of the computational analysis.

3. METHODS

3.1. Renal needle biopsies: Tissue slicing, slide preparation, and digital imaging

Intact kidney tissues from normal healthy untreated rat and mouse were collected from Dr. Tracey Ignatowski (Pathology & Anatomical Sciences, University at Buffalo) and Dr. Margarita Dubokovich (Pharmacology & Toxicology, University at Buffalo), respectively, sacrificed under their protocol. Tissue slices with 2 μm thickness were cut from paraffin embedded renal tissues, stained with PAS or H&E, and imaged using a whole slide scanner (Aperio) at 40X magnification.

3.2. Computational analysis

We briefly review the concepts of color deconvolution, Gabor filtering, and bottleneck detection below.

3.2.1. Color deconvolution

According to Lambert-Beer's law, the detected intensity of light transmitted through the histopathology specimen, with an amount c of stain and absorption factor a is described by,[13]

$$I = I_0 \exp(-ac), \tag{1}$$

where I_0 is the intensity of light entering the specimen, and I the intensity of light detected after passing the specimen. Optical density is defined as,

$$o = -\log_{10} \frac{I}{I_0} = ac. \tag{2}$$

For three stains, $\boldsymbol{M}$ is a normalized matrix of size 3×3 of the optical density matrix formed by rows corresponding to the stains and columns corresponding to the measured optical density in the red, green, and blue channels, and is defined as,

$$m_{ij} = \frac{o_{ij}}{\sqrt{\sum_{1}^{3} o_{ik}^2}}. \tag{3}$$

If the measured intensities of the true color image for a pixel p in the red, green, and blue channels are I_{R}, I_{G}, and I_{B}, and thus, $\boldsymbol{y} = \begin{bmatrix} I_{\mathrm{R}} & I_{\mathrm{G}} & I_{\mathrm{B}} \end{bmatrix}$, we have,

$$\boldsymbol{y} = \boldsymbol{c}\boldsymbol{M}, \tag{4}$$

where $\boldsymbol{c} = \begin{bmatrix} c_1 & c_2 & c_3 \end{bmatrix}$ is the vector defining the amount of three stains, and is given by,

$$\boldsymbol{c} = \boldsymbol{y}\boldsymbol{M}^{-1}, \tag{5}$$

which describes the orthogonal representations of the stains forming the pixel p in the image.

3.2.2. Gabor filter bank based texture segmentation

We review briefly the unsupervised texture segmentation method using Gabor filters, which was originally proposed by Jain *et al.*[15] A 2D Gabor function consists of a sinusoidal plane wave of a given frequency and orientation, modulated by a 2D Gaussian envelope. Such a filter is given by,

$$h(x,y) = \exp\left\{-\frac{1}{2}\left[\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right]\right\}\cos\left(2\pi u_0 x+\phi\right), \tag{6}$$

where u_0 and ϕ are the frequency and phase of the sinusoidal plane wave along the x axis with 0^o orientation, and σ_x and σ_y are the space constants of the Gaussian envelope along the x and y axes, respectively. A Gabor filter with arbitrary orientation, θ_0, can be obtained via a rigid rotation of the xy plane. In frequency domain, with $\phi = 0$, the Gabor function is given by,

$$H(u,v) = A\left(\exp\left\{-\frac{1}{2}\left[\frac{(u-u_0)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right]\right\}+\exp\left\{-\frac{1}{2}\left[\frac{(u+u_0)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right]\right\}\right), \tag{7}$$

where $\sigma_u = 1/2\pi\sigma_x$, $\sigma_v = 1/2\pi\sigma_y$, and $A = 2\pi\sigma_x\sigma_y$. To resolve finer textures, it is desirable to have smaller bandwidth in the spatial-frequency domain, while accurate localization of texture boundaries requires filters that are localized in the spatial domain. However, the width of the filter in the spatial domain and bandwidth in the spatial-frequency domain are inversely related.

Texture segmentation was performed using an array of filters generated using (7), following the algorithm developed by MathWorks.[19] Orientations were chosen to be in $\left[0,150^o\right]$ in steps of 30^o. Wavelengths were chosen to be in increasing powers of two starting from $4/\sqrt{2}$ up to the hypotenuse length of the input image. The input image is filtered through the filter array formed by such set-up. The output image is Gaussian-smoothed, normalized, and clustered via k-means to obtain the foreground texture, which is the glomerulus in our work.

3.2.3. Clump splitting via bottleneck detection

We perform bottleneck segmentation for nuclei segmentation discussed in Section 2.5 using the method discussed by Wang *et al.*[18] Namely, for two points A and B on the boundary contour of an object, we optimize the cost function between points A and B for splitting as:

$$E_s(A,B) = \frac{d(A,B)}{\min\{l(A,B), l(B,A)\}}, \tag{8}$$

where $d(A,B)$ denotes the distance between points A and B and $l(A,B)$ denotes the clockwise geodesic length from point A to B on the boundary of the clump. We assume that the line joining A and B is inside the clump. A^* and B^* are the points where $E_s(A,B)$ is minimized. Upon the optimization, the method provides a pair of points for splitting clumped nuclei, with a goal to segment single individual nuclei.

4. DISCUSSION

One of the biggest challenges of developing any computational method to analyze renal pathology is the diversity of disease state. While a healthy glomerulus may look one way, diseased glomeruli can look similar, or drastically different, or even be non-existent. In this regard, the method proposed here may not work for different disease models. For example, some renal diseases may cause vast, widespread proliferation of nuclei,[20] which would effectively call for an extension of the glomerular location estimation presented in this article. Moreover, in other diseases, such as acute kidney injury[21],

damages are seen in the structure of tubules rather than glomeruli. The diversity of renal pathological markers makes it incredibly difficult to encompass all encounterable situations within one model.

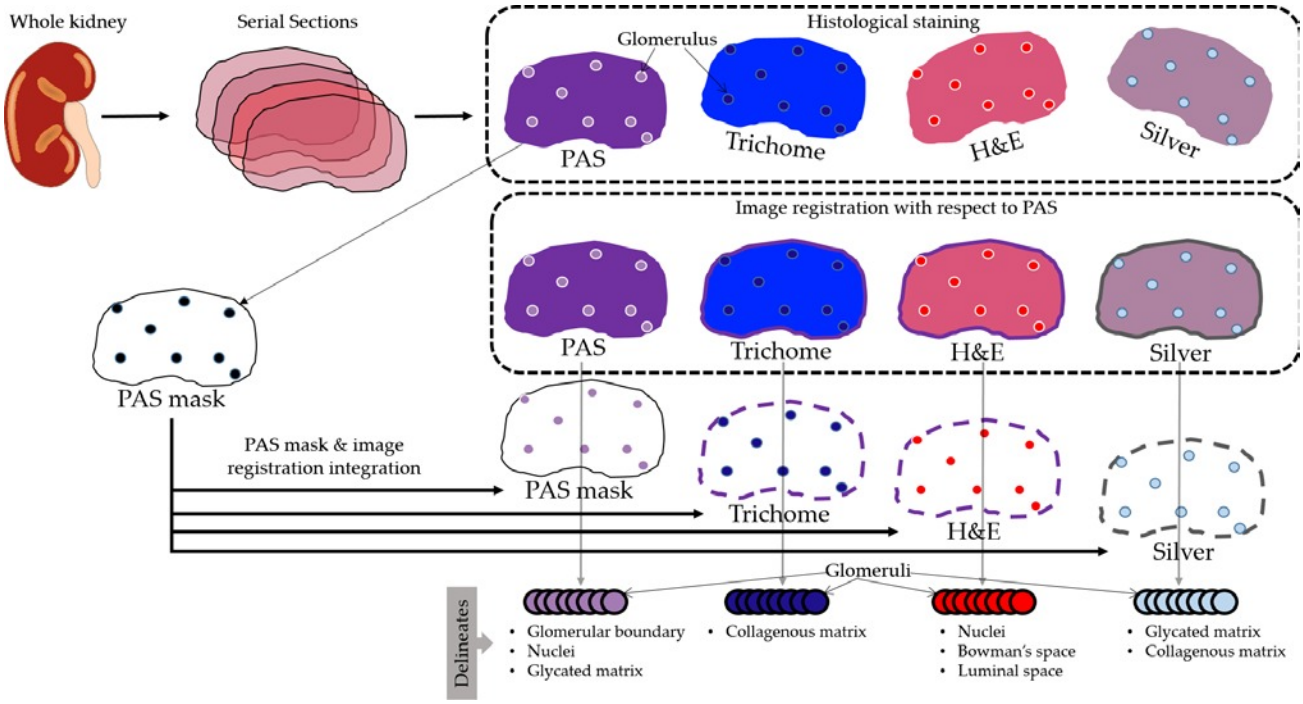


Figure 8. A sketch of the multimodal glomerulus segmentation approach. Serial sections are cut and stained with different histopathological staining to delineate diverse glomerular features. First, serial slice will be stained using periodic acid-Schiff (PAS), and used to approximate glomerular locations. Second, third, and fourth slices, which could be stained using trichome, hematoxylin and eosin (H&E), and silver, will be registered with PAS image. The PAS extracted glomerular mask will be applied to the registered images to extract single glomeruli. Features describing different biological processes will be estimated from single glomeruli using computational image analysis.

No matter how reproducible the method, it is still difficult to generalize conclusions on the biological accuracy of computational structural estimation. This is inherent within the problem of renal pathology; even trained pathologists may analyze the same image twice and return different numbers for quantification of glomerular compartments; due mostly to diversity and fuzzy glomerular compartment boundaries. In the current work, we found that the most difficult structure to segment is the luminal space. Luminal space is epithelial tissue, which shows white in staining like H&E and PAS, and is not attached to the Bowman's space. However, because histological staining shows only a 2D representation of a 3D structure, defining luminal content is tricky. Or, for example, referencing Fig. 3F, the extracellular matrix may be segmentable, but the true question becomes: What is the proper way to describe this matrix? Should intensity distribution be considered, or area, or average color value? Computational methods can often be derailed by the highly heterogeneous environment where there exists vast variation between compartments. Regions which have abnormal cellularity may appear as false glomeruli in estimating their locations. Staining variations and textural abnormalities may improperly define glomerular boundary, or hypo-eosinophilic nuclei may be highly integrated into the background, and produce minimal contrast in the resulting image. For each level of noise present within a renal tissue image, the number of parameters required to distinguish structures from noise will increase drastically.

The advantage of pattern recognition with machine learning, however, is that the automated methods for structural estimation may not need to be perfect to classify different diseases with high accuracy. It is possible to generate automatically certain pathological features with high reproducibility and no parameterization, even when they may not segment a desired structure perfectly. Features extracted this way can readily be employed in analyzing an extremely large amount of data in a feasible time window with no manual parameter tuning. "Large data" refers to whole slide images, which could contain up to 10^{12} pixels at 40X resolution. We expect that mining large amounts of pathological data in different diseases, resolutions, or time courses, may lead to the discovery of hidden pathological trends which cannot be discerned by human eye, and only observable computationally. We have found, through this current work, that the most

reproducible methods for H&E and PAS stained images, are glomerular structural analysis using Gabor filtering as well as color deconvolution and PCA. Color deconvolution and PCA have been demonstrated reproducibly in a wide variety of automated clinical practices, from cancerous nuclei segmentation to genetic variation analysis of populations.[22, 23] These methods perform with high accuracy and have zero tunable parameterization. Our nuclei segmentation method could also analyze the average glomerular cellularity for the entire kidney slice, with no manual tuning required. We envision using these robust methods to extract important features from large numbers of glomeruli from whole slide images.

No one single method will be able to differentiate the plethora of renal diseases without the integration of many independent components. However, different types of stains may delineate features unique to certain disease types. Due to this fact, we propose to use a multimodal method of information collection, summarized in Fig. 8. Renal slices would be cut serially, but each subsequent slice would be stained with different histopathological markers. Glomerular segmentation methods will be developed for each stain type. In this manner, one could garner all possible renal pathological markers of a diseased tissue through several stains, and get a fully accurate profile of the renal tissue. Because glomeruli may not be easily segmentable in some types of stains, we propose an image registration based of glomeruli extraction pipeline. Glomeruli would be segmented from PAS stained slices first. Approximate glomeruli mask constructed from the PAS image would be registered to slices from different stains to segment glomeruli in other images. This method will require development of a new type of image registration method. Traditional image registration method matches an approximate image with its exact boundaries.[24] However, the difficulty in doing the same for serial renal slices is that there exists small variations in structures from one slice to another. As slices are cut from interior to posterior, renal structures tend to move, ever so slightly. Traditional image registration attempts to match a moving image, or noise contaminated image, to a fixed image. However, because all structures within the renal anatomy do not shift in the same direction, traditional image registration is not effective.

5. CONCLUSION

We have demonstrated an automated and quantitative method to computationally analyze renal histopathology images. Our computational method shows that diverse renal heterogeneous tissues can be broken down into separate compartments quantitatively, and analyzed in a way similar to that of a trained pathologist. We have found that parameter-free methods, such as principal component analysis, color deconvolution method, and Gabor texture analysis, as well as our nuclei segmentation method have potential to delineate important pathological glomerular features reproducibly from large renal pathology data. Our segmentation process and accuracy quantification will be extended in a future work to include more glomerular compartments. Other aims include integrating multi-modal large image data obtained by imaging renal tissues by staining those using diverse histopathological markers. We envision that our computational tool will make it possible to accurately classify different renal disease types using pattern recognition and machine learning.

ACKNOWLEDGEMENT

This project was supported by the faculty start-up funds from the Pathology & Anatomical Sciences Department, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo. We thank Dr. Tracey Ignatowski (Pathology & Anatomical Sciences, University at Buffalo) and Dr. Margarita Dubokovich (Pharmacology & Toxicology, University at Buffalo), for providing us the rat and mouse renal tissues for this work. We thank the histology core facility (Pathology & Anatomical Sciences, University at Buffalo) for performing histopathological staining of tissues. We thank master's students Jiaqiao Han, Darshana Goving, and Sixing Zhou (Biomedical Engineering, University at Buffalo) for conducting the manual segmentation for analyzing the performance of our computational method discussed in Section 2.6.

REFERENCES

[1] Chronic kidney disease and kidney failure Available from: http://report.nih.gov/nIHfactsheets/ViewFactSheet.aspx?csid=34&key=C

[2] Davidson, A. and Diamond, B., "Autoimmune Diseases," New England Journal of Medicine, 345(5), 340-350 (2001).

[3] Lichtnekert, J., Kulkarni, O. P., Mulay, S. R. *et al.*, "Anti-GBM glomerulonephritis involves IL-1 but is independent of NLRP3/ASC inflammasome-mediated activation of caspase-1," PLoS One, 6(10), e26778 (2011).

[4] Kopp, J. B., Nelson, G. W., Sampath, K. *et al.*, "APOL1 genetic variants in focal segmental glomerulosclerosis and HIV-associated nephropathy," J Am Soc Nephrol, 22(11), 2129-37 (2011).

[5] Sethi, S., Glassock, R. J. and Fervenza, F. C., "Focal segmental glomerulosclerosis: towards a better understanding for the practicing nephrologist," Nephrol Dial Transplant, 30(3), 375-84 (2015).

[6] Fernandez-Fernandez, B., Ortiz, A., Gomez-Guerrero, C. *et al.*, "Therapeutic approaches to diabetic nephropathy--beyond the RAS," Nat Rev Nephrol, 10(6), 325-46 (2014).
[7] Lai, W. L., Yeh, T. H., Chen, P. M. *et al.*, "Membranous nephropathy: a review on the pathogenesis, diagnosis, and treatment," J Formos Med Assoc, 114(2), 102-11 (2015).
[8] Saha, T. C. and Singh, H., "Minimal change disease: a review," South Med J, 99(11), 1264-70 (2006).
[9] Glomerular disease primer Available from: http://www.niddk.nih.gov/research-funding/at-niddk/labs-branches/kidney-disease-branch/kidney-diseases-section/glomerular-disease-primer/Pages/glomerular-disease-primer.aspx
[10] Kidney disease: Early detection and treatment Available from: https://www.nlm.nih.gov/medlineplus/magazine/issues/winter08/articles/winter08pg9-10.html
[11] Gonzalez, R. C. and Woods, R. E., [Digital Image Processing], Prentice Hall, 1-976 (2007).
[12] Young, B., O'Dowd, G. and Woodford, P., [Wheater's Functional Histology: A Text and Colour Atlas], Churchill Livingstone, 464 (2013).
[13] Ruifrok, A. C. and JOHnston, D. A., "Quantification of histochemical staining of color deconvolution," Anal. Quant. Cytol. Histol., 23(4), 291-299 (2001).
[14] Hogg, R. V., Craig, A. and McKean, J. W., [Introduction to Mathematical Statistics], Prentice Hall, Upper Saddle River, New Jersey, 1-692 (2004).
[15] Jain, A. K. and Farrokhnia, F., "Unsupervised Texture Segmentation Using Gabor Filters," Pattern Recognition, 24(12), 1167-1186 (1991).
[16] Muthu Rama Krishnan, M., Choudhary, A., Chakraborty, C. *et al.*, "Texture based segmentation of epithelial layer from oral histological images," Micron, 42(6), 632-41 (2011).
[17] Irshad, H., Veillard, A., Roux, L. *et al.*, "Methods for nuclei detection, segmentation, and classification in digital histopathology: a review-current status and future potential," IEEE Rev Biomed Eng, 7, 97-114 (2014).
[18] Hui, W., Hong, Z. and Ray, N., "Clump splitting via bottleneck detection," 18th IEEE International Conference on Image Processing (ICIP), 61-64 (2011).
[19] Texture segmentation using Gabor filters. Available from: http://www.mathworks.com/help/images/texture-segmentation-using-gabor-filters.html
[20] Kanwar, Y. S., Sun, L., Xie, P. *et al.*, "A glimpse of various pathogenetic mechanisms of diabetic nephropathy," Annu Rev Pathol, 6, 395-423 (2011).
[21] Basile, D. P., Anderson, M. D. and Sutton, T. A., "Pathophysiology of acute kidney injury," Compr Physiol, 2(2), 1303-53 (2012).
[22] Di Cataldo, S., Ficarra, E., Acquaviva, A. *et al.*, "Achieving the way for automated segmentation of nuclei in cancer tissue images through morphology-based approach: a quantitative evaluation," Comput Med Imaging Graph, 34(6), 453-61 (2010).
[23] Novembre, J. and Stephens, M., "Interpreting principal component analyses of spatial population genetic variation," Nat Genet, 40(5), 646-649 (2008).
[24] Oliveira, F. P. and Tavares, J. M., "Medical image registration: a review," Comput Methods Biomech Biomed Engin, 17(2), 73-93 (2014).