Computational analysis of the structural progression of human glomeruli in diabetic nephropathy

Brandon Ginley, John Tomaszewski, Kuang-Yu Jen, Agnes Fogo, Sanjay Jain, et al.

Brandon G. Ginley, John E. Tomaszewski, Kuang-Yu Jen, Agnes Fogo, Sanjay Jain, Pinaki Sarder, "Computational analysis of the structural progression of human glomeruli in diabetic nephropathy," Proc. SPIE 10581, Medical Imaging 2018: Digital Pathology, 105810A (6 March 2018); doi: 10.1117/12.2295249

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

Computational analysis of the structural progression of human glomeruli in diabetic nephropathy

Brandon G. Ginley[1], John E. Tomaszewski[1], Kuang-Yu Jen[2], Agnes Fogo[3], Sanjay Jain[4], and Pinaki Sarder[1,*]

[1]Departments of Pathology and Anatomical Sciences,
University at Buffalo – The State University of New York

[2]Department of Pathology,
University of California, San Francisco

[3]Departments of Pathology, Microbiology, and Immunology,
Vanderbilt University School of Medicine

[4]Departments of Medicine – Nephrology,
Washington University School of Medicine

[*]Address all correspondence to: Pinaki Sarder
Tel: 716-829-2265; E-mail: pinakisa@buffalo.edu

ABSTRACT

The glomerulus is the primary compartment of blood filtration in the kidney. It is a sphere of bundled, fenestrated capillaries that selectively allows solute loss. Structural damages to glomerular micro-compartments lead to physiological failures which influence filtration efficacy. The sole way to confirm glomerular structural damage in renal pathology is by examining histopathological or immunofluorescence stained needle biopsies under a light microscope. However, this method is extremely tedious and time consuming, and requires manual scoring on the number and volume of structures. Computational image analysis is the perfect tool to ease this burden. The major obstacle to development of digital histopathological quantification protocols for renal pathology is the extreme heterogeneity present within kidney tissue. Here we present an automated computational pipeline to 1) segment glomerular compartment boundaries and 2) quantify features of compartments, in healthy and diseased renal tissue. The segmentation involves a two stage process, one step for rough segmentation generation and another for refinement. Using a Naïve Bayesian classifier on the resulting feature set, this method was able to distinguish pathological stage IIa from III with 0.89/0.93 sensitivity/specificity and stage IIb from III with 0.7/0.8 sensitivity/specificity, on n = 514 glomeruli taken from n = 13 human biopsies with diagnosed diabetic nephropathy, and n = 5 human renal tissues with no histological abnormalities. Our method will simplify computational partitioning of glomerular micro-compartments and subsequent quantification. We aim for our methods to ease manual labor associated with clinical diagnosis of renal disease.

Key-words: Glomerulus, digital quantification, diabetic nephropathy, automated computation, feature extraction, segmentation, histology, kidney, naïve Bayes

1. INTRODUCTION

The kidney is the organ responsible for filtering solutes out of the blood into urine. The glomerulus (~200 µm in diameter in health) handles the initial filtration of blood to primary filtrate, and is usually the first structure to be damaged in disease[1]. The glomerulus, structurally, is a bundle of fenestrated capillaries encapsulated by a luminous collecting duct. Blood solutes are filtered first through endothelial fenestrations, then selectively filtered by the glomerular basement and visceral epithelial cells (podocytes), after which they are collected in the urinary space. From here, solutes are selectively reabsorbed and everything else is released to the urine. Each human kidney contains approximately 900,000 to 1 million nephrons, each of which is attached to a glomerulus[2]. Damage to any of the basic glomerular compartments leads to instability within the glomerulus that is characteristic of renal disease and failure[3]. Diabetic nephropathy (DN) is a structural glomerular disease, characterized by progressive mesangial space expansion, thickening of basement membranes, and loss of resident cell nuclei. DN can lead to renal failure and death[4]. Medicare reports there are approximately 525K U.S. patients with end stage kidney failure, the care for whom

Medical Imaging 2018: Digital Pathology, edited by John E. Tomaszewski, Metin N. Gurcan,
Proc. of SPIE Vol. 10581, 105810A · © 2018 SPIE · CCC code:
1605-7422/18/$18 · doi: 10.1117/12.2295249

amounts over $24 billion[5, 6]. The center for disease control (CDC) projects one in three Americans will have diabetes by 2050 if current trends continue[7]. Approximately 20-30% of diabetics will develop evidence of nephropathy[8]. The main barrier to improvement of care is the lack of disease markers which can identify disease states early enough such that therapeutic interventions may be initiated to lead to better outcomes[9]. The most important clinical technique to identify glomerular structural damage is imaging of the renal biopsy. While this approach is the gold standard in clinic, it is time consuming, tedious, and semi-quantitative. The highly complex structure of renal tissue creates these issues and has hindered robust computational quantification methods from being developed in the past.

Some examples of clinically relevant glomerular micro-compartmental changes include glomerular volume, podocyte effacement and death, changes in mesangial cellularity and matrix volume, and lumen content[3]. Automatic estimation of these features using computational resources requires development of novel techniques to accurately segment the glomerular boundary, the boundary of luminal compartments, the boundary of mesangium, and the boundary of nuclei. We have developed a glomerular boundary segmentation method in our previous work[10]. In this work, we hand segment the glomerular boundaries, as the focus is not on developing a glomerular boundary segmentation pipeline but rather analyzing differences in healthy and disease DN renal tissue. Further, we have developed a simple technique using classical methods to achieve accurate glomerular compartment segmentations with high accuracy for Periodic Acid-Schiff stained renal tissue. This technique consists of standard color transformations and automated thresholding to develop a preliminary segmentation, followed by a naïve Bayesian classification based correction of the segmentation. Stain deconvolution[11] is used to identify the nuclei, mesangial, and membranous components. *L*a*b** transformation and Otsu's thresholding is used to identify luminal compartments. From the segmented glomerular compartments, we have extracted an array of texture and morphological features that are capable of delineating DN clinical stage. We trained a naïve Bayesian classifier to differentiate DN class based on the extracted features. The classifier achieves optimal performance for determining stage IIa from III with 0.89 sensitivity and 0.93 specificity. The classifier for IIb from III achieves optimal performance scoring 0.7 sensitivity and 0.8 specificity (Fig 3).

We have developed a structural feature quantification pipeline for glomeruli taken from DN diseased human biopsy tissues, stained with periodic acid-Schiff. This technique is particularly simple, and can easily be transferred to other research easily. This technique can easily be extended to other staining types, and possibly to other tissue systems as well. Our ultimate aim is to provide a faster statistical sampling in clinical pathology than the current practice, while ensuring accuracy and precision in diagnosis. Our work aims to motivate the development of digital protocols which quantify glomerular features and motivate the shift of renal pathology to a computational era.

2. RESULTS

2.1. **Glomerular segmentation**

Fig. 1 shows a demonstration of typical segmentations for the glomeruli used in this work. Fig. 1A shows an original glomerulus. The boundary between the glomerulus and surrounding tubules was segmented manually, shown in Fig. 1B. Before the glomerulus is isolated, each image is transformed to both *L*a*b** space[12]. Luminal components are segmented through Otsu's thresholding[12] of the *L** component of the *L*a*b** color space, Fig. 1C. To isolate mesangial components, stain deconvolution for PAS components is used, shown in Fig. 1D. Resulting intensity image is thresholded using local adaptive thresholding. Nuclei in the image are delineated through stain deconvolution, resulting in an intensity image, Fig. 1E, which is segmented using a local adaptive threshold. This segmentation leaves some intra-glomerular pixels unlabeled, shown in binary in Fig. 1G. To classify these pixels, a naïve Bayesian classifier[13] is trained on the already segmented pixels. The Naïve Bayesian classifier is then used to predict the class labels of remaining glomerular pixels. The final segmentation is shown in Fig. 1H.

2.2. **Feature extraction**

We have hand-constructed a computational feature set with the intention of describing the inherent structural progression of DN glomeruli. Some of these features are based in texture analysis, some are morphological, some are distance based, and some have been hand crafted to theoretically accentuate proper glomerular structural conformation. In total, 47 features were extracted from each glomerulus.

2.3. **Singular value decomposition**

A singular value decomposition[14] was performed on the original 47 features extracted from each glomerulus originated from one of the five stages of diabetic nephropathy[15]. We used the first 25 eigenvalues to train a naïve

Bayesian classifier to differentiate DN disease states. The first three principal feature axes of these features are plotted in Fig. 2, stratified by pairwise consecutive disease states.

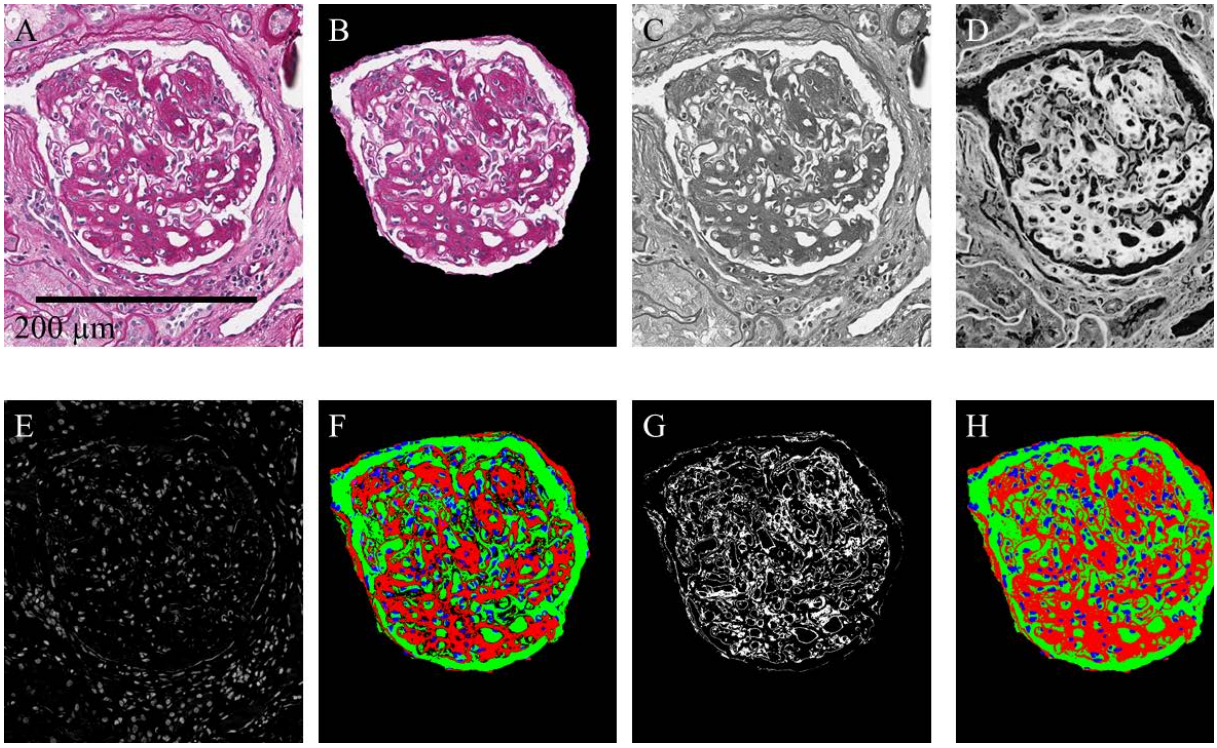

Figure 1. Demonstration of segmentation pipeline. (A) Original image. (B) Image after manual segmentation. (C) *L** component of *L*a*b** space delineating luminal space. (D) Color deconvolution for PAS. (E) Color deconvolution for hematoxylin. (F) Segmentation before correction by the naïve Bayesian classification. (G) Pixels to be classified by the naïve Bayesian classification. (H) Final segmentation.

2.4. **Naïve Bayesian classification**

To estimate the classification performance of the principal features to differentiate two different disease states, we trained two naïve Bayesian classifiers to differentiate between stage IIa and III as well as IIb and III. Structural disease state of each glomerulus was labeled by Dr. Kuang-Yu Jen, a licensed renal pathologist, who attempted to place each glomerulus into the DN stage[15] it most closely resembled (I, IIa, IIb, III, or IV). Because stage IV requires >50% of biopsied glomeruli to be fully sclerosed, any glomerulus that was fully sclerosed was labeled class IV. Out of $n = 514$ total glomeruli, 36 were classified as stage IIa, 94 as stage IIb, and 101 as stage III. Both classifiers treated 50% of the data as holdout and 50% as training data. One Naïve Bayesian model was able to identify stage IIa glomeruli from stage III glomeruli with 0.89/0.93 sensitivity/specificity. The other naïve Bayesian model was able to identify stage IIb glomeruli from stage III glomeruli with 0.7/0.8 sensitivity/specificity. The complete receiver operating curves can be seen in Fig. 3. The classifier for stage IIa from III is particularly efficient, scoring 0.97 area under curve (AUC).

3. METHODS

3.1. **Human data**

Biopsy samples from 13 unique human patients with diabetic nephropathy with chronic kidney disease stages II and III were used. Part of these samples ($n = 7$) were collected from Kidney Translational Research Center at Washington University School of Medicine directed by co-author Dr. Jain, and digital images of the rest came from co-author Dr. Fogo. Thin renal tissue samples with no apparent histological abnormalities from 5 patients were used as control. Three of the cases here came from Dr. Jain, and digital images of the rest from Dr. Fogo. Human data collection procedure followed a protocol approved by the Institutional Review Board at University at Buffalo.

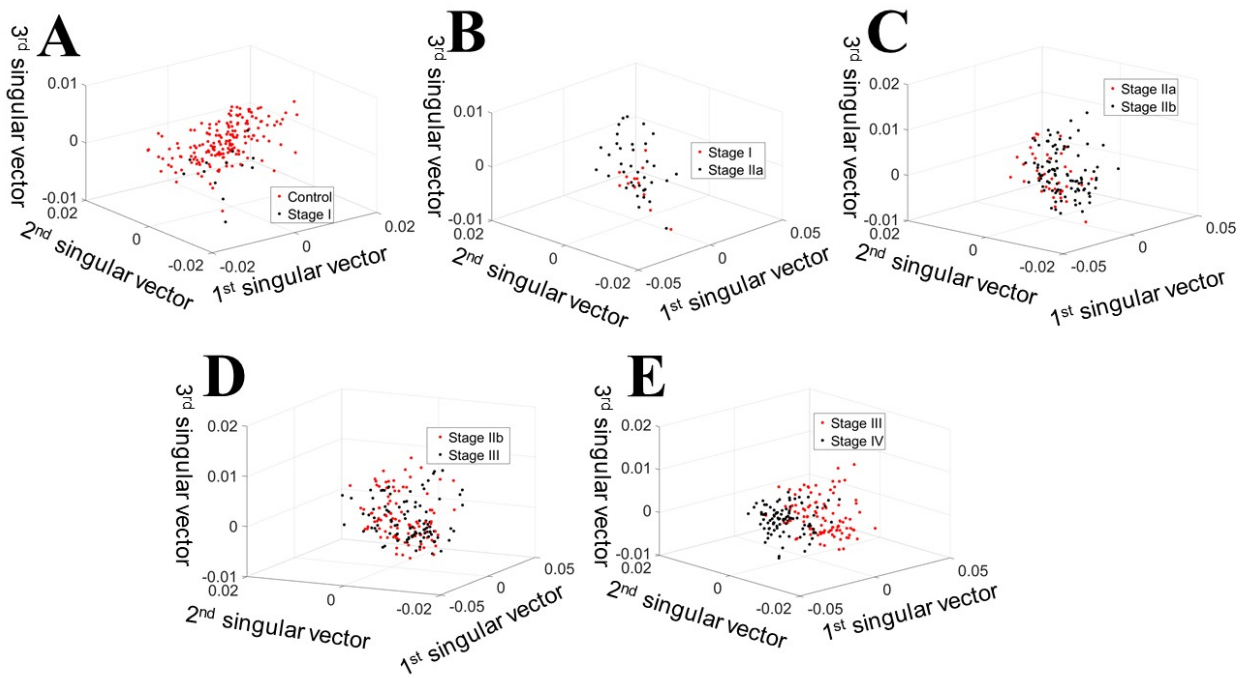

Figure 2. Scatter plot of top three singular vectors of 47 dimensional feature space, stratified by pairwise disease state. (A) Structure of control glomeruli versus stage I. (B) Glomerular structures for stage I versus stage IIa. (C) Stage IIa versus stage IIb. (D) Stage IIb versus III. (E) Stage III versus IV.

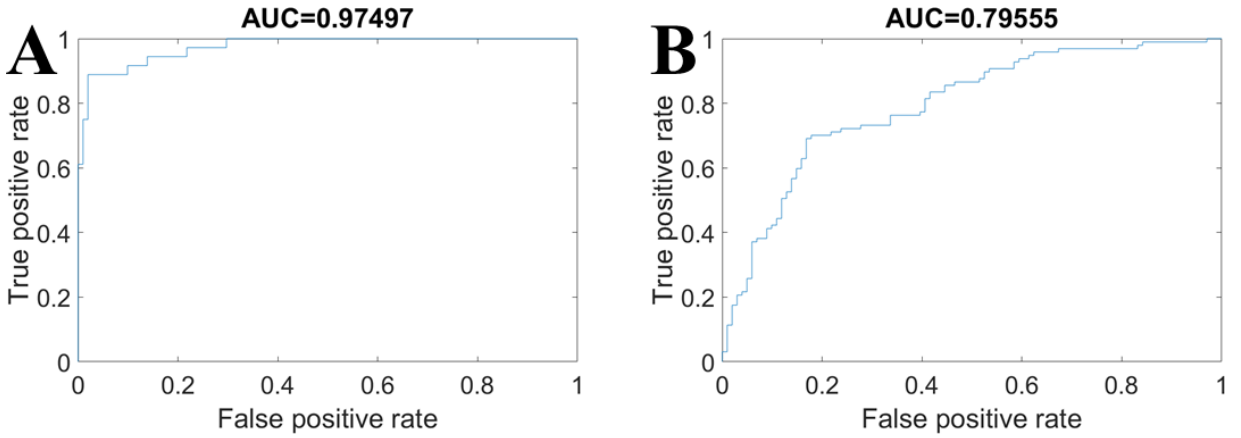

Figure 3. Receiver operating curves for Naïve Bayesian disease classification. (A) Classification of stage IIa from stage III. (B) Classification of stage IIb from III.

3.2. **Imaging and data preparation**

Tissue slices of 2-5 µm were stained using Periodic acid-Schiff, and imaged using a whole-slide imaging scanner (Aperio Versa, Leica, Buffalo Grove, Il). Similar imaging protocol was followed as described in our earlier works[16]. Glomeruli images were manually segmented from each biopsy section for the analysis.

3.3. **Ground-truth annotation and segmentation**

Glomerulus wise ground-truth annotations of DN structural disease state was performed by the co-author Dr. Kuang-Yu Jen. Fig. 4 demonstrates one example glomerulus from each disease classification. Fig. 4A shows a Stage I (healthy/minimal damage) glomerulus, 4B shows stage IIa, 4C shows stage IIb, 4D shows stage III, and 4E shows stage IV. Staging was done following definition available in the literature[15].

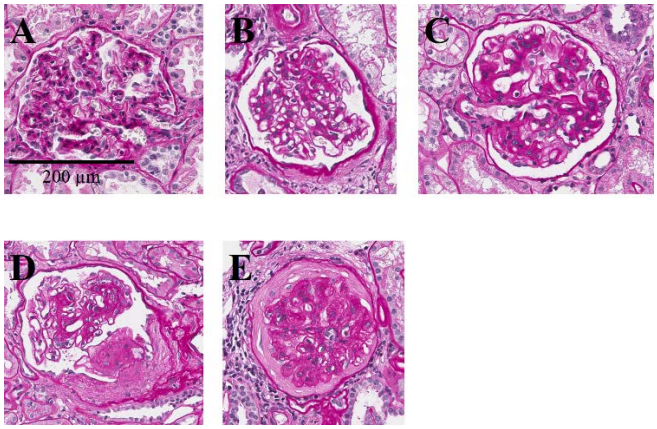

Figure 4. Examples of annotated glomeruli. A. Stage I/healthy. B. Stage IIa. C. Stage IIb. D. Stage III. E. Stage IV.

3.4. **Color space transformation**

Color space transformations performed in this work are standard. Transformation of RGB color space to LAB was performed with using the MATLAB image analysis toolbox. Stain deconvolution was performed using the method developed by Ruifrok *et al.*[11].

3.5. **Thresholding**

Otsu's thresholding is a standard technique used in image analysis to identify foreground pixels from background when there is a distinct bimodal pixel distribution[17]. Otsu's method seeks to minimize the intra-class variance resultant of any one particular threshold. Otsu's thresholding was performed with the MATLAB image analysis toolbox. Otsu's threshold method was used to determine luminal segmentations. Adaptive thresholding[18] was performed for mesangial and nuclear segmentations, and performs thresholding in a local window rather than globally.

3.6. **Feature extraction**

In total, n = 47 features were extracted from each glomerulus, including textural, morphological, and distance descriptions, as well as compartmental containment ratios.

Textural features are calculated on mesangial segmentations and on luminal segmentations. Nuclei were excluded from texture analysis because classification of nuclear texture is a highly complex issue in its own right, dependent on position within the tissue, level of nuclear cut, stain type, transcription status, cell type, etc., and is outside the scope of this manuscript. A linear weighted combination of red, green, and blue channels yields a grayscale image suitable for gray level co-occurrence analysis. Textural entropy, correlation, contrast, and homogeneity were derived from the image gray-level co-occurrence matrix.

Morphological features include the mean area, median area, mode of areas, and convexity for identified compartment objects treated separately (e.g., one single nucleus, isolated piece of mesangium, or capillary lumen).

Distance features are comprised of averaged distances between compartments and other glomerular compartments or landmarks. Glomerular landmarks include the estimated glomerular centroid and the estimated glomerular boundary. The following distance features are extracted for each object: 1) the distance between that object's centroid and similar objects' centroids, 2) the distance to the glomerular boundary, 3) the distance to the glomerular centroid. For each of these categories, average, minimum, and maximum distances were recorded.

Compartmental containment features revolve around analyzing the relationship between the structures of the glomerulus itself. First, each identified lumenal or mesangial compartment is isolated from other compartments. Then, the convex hull of that object is taken. Nuclei, mesangium, or lumen segmentations that fall within this convex boundaries are also extracted and their area recorded. The ratio between the area of the convex object and the area of other compartments enclosed within its boundary become compartmental overlap features. This feature extraction method was applied to all luminal and mesangial sections, and examined the contained amount of luminal, mesangial, or nuclear area contained within. Logically, nuclei contain no other compartments within themselves. For any one nucleus, two features are derived by calculating the fraction of pixels that fall on the nuclear boundary which were segmented as 1) lumen or 2) mesangium. These serve to primarily help the computer understand fractional nuclear composition of each glomerulus.

3.7. **Feature reduction**.

Principal component analysis is ideal to reduce the dimensionality of an original feature space. However, since the feature space was not full rank, principal component analysis is not applicable, and singular value decomposition is used instead[19]. We found that approximately 25 singular vectors must be included to represent 99% of the original feature space. These 25 vectors were used to train the naïve Bayesian classifiers.

3.8. **Classification.**

The 25 singular vectors and pathologist class labels were used to train respective naïve Bayesian classifiers to differentiate any one disease state from another. Hyper parameters of the classification model were iteratively selected using an optimization procedure available from the MATLAB Statistics and Machine learning toolbox. Each classifier was trained using 50% of the available data as holdout.

4. DISCUSSION

We show a near-full extraction pipeline for quantification of human glomeruli in renal histology images. We have also shown that digitally extracted features are able to train supervised machine learners to detect disease states. Though the boundary of the glomerulus was hand segmented in this work, we have demonstrated in the past a working pipeline to computationally identify this boundary[16]. Though the computational methods applied are not new, the physical quantification of human glomeruli in various states of DN structural disease using computational techniques has not been shown in the literature. Our first study suggests that we can precisely and accurately classify disease states by integrating information from multi-dimensional feature space composed from glomerular micro-compartmental structures.

5. CONCLUSION AND FUTURE WORK

This study shows for the first time that automated computational quantification of glomerular images can yield digital features of disease state which can be used for prediction. Automated renal histopathology is the introduction of the answer to the current clinical obstacles in renal histopathology. We expect a higher sample population to yield more robust description of disease states. Future works will aim to investigate higher sample populations, more diverse sample populations, a larger original feature set, and the relationship between structural data and follow-up clinical biometrics such as blood sugar, proteinuria, and estimated glomerular clearance. Future work will also investigate the ability to predict any one particular glomerular disease state from a continuum of disease states, rather than a binary comparison of states.

ACKNOWLEDGEMENT

This project was supported partially by the faculty start-up fund from the Pathology & Anatomical Sciences Department, Jacobs School of Medicine and Biomedical Sciences, University at Buffalo (UB), partially by the UB IMPACT award, and partially by the DiaComp Pilot and Feasibility Program grant #32307-5.

REFERENCES

[1] Pollak, M. R., Quaggin, S. E., Hoenig, M. P. *et al.*, "The glomerulus: the sphere of influence," Clin J Am Soc Nephrol, 9(8), 1461-9 (2014).
[2] Bertram, J. F., Douglas-Denton, R. N., Diouf, B. *et al.*, "Human nephron number: implications for health and disease," Pediatr Nephrol, 26(9), 1529-33 (2011).
[3] D'Amico, G. and Bazzi, C., "Pathophysiology of proteinuria," Kidney Int, 63(3), 809-25 (2003).
[4] Fioretto, P. and Mauer, M., "Histopathology of diabetic nephropathy," Semin Nephrol, 27(2), 195-207 (2007).
[5] Tuttle, K. R., Bakris, G. L., Bilous, R. W. *et al.*, "Diabetic kidney disease: a report from an ADA Consensus Conference," Diabetes Care, 37(10), 2864-83 (2014).
[6] Chronic kidney disease and kidney failure Available from: http://report.nih.gov/nIHfactsheets/ViewFactSheet.aspx?csid=34&key=C
[7] Diabetes Report Card Available from: https://www.cdc.gov/diabetes/library/reports/congress.html

[8] Saran, R., Robinson, B., Abbott, K. C. *et al.*, "US Renal Data System 2016 Annual Data Report: Epidemiology of Kidney Disease in the United States," American Journal of Kidney Diseases, 69(3), (2017).
[9] Kim, S. S., Kim, J. H. and Kim, I. J., "Current Challenges in Diabetic Nephropathy: Early Diagnosis and Ways to Improve Outcomes," Endocrinol Metab (Seoul), 31(2), 245-53 (2016).
[10] Brandon Ginley, J. E. T., Rabi Yacoub, Feng Chen, Pinaki Sarder, "Unsupervised Labeling of Glomerular Boundaries using Gabor Filters and Statistical Testing in Renal Histology," Journal of Medical Imaging, (2017).
[11] Ruifrok, A. C. and Johnston, D. A., "Quantification of histochemical staining by color deconvolution," Analytical and Quantitative Cytology and Histology, 23(4), 291-299 (2001).
[12] Gonzalez, R. C., Woods, R. E. and Eddins, S. L., [Digital Image processing using MATLAB], Gatesmark Pub., S.I.(2009).
[13] Hastie, T., Tibshirani, R., Friedman, J.H., [The Elements of Statistical Learning: Data Mining, Inference, and Prediction], Springer, 768 (2009).
[14] Strang, G., [Introduction to Linear Algebra], Wellesley-Cambridge Press, (2009).
[15] Tervaert, T. W., Mooyaart, A. L., Amann, K. *et al.*, "Pathologic classification of diabetic nephropathy," J Am Soc Nephrol, 21(4), 556-63 (2010).
[16] Ginley, B., Tomaszewski, J. E., Yacoub, R. *et al.*, "Unsupervised labeling of glomerular boundaries using Gabor filters and statistical testing in renal histology," Journal of Medical Imaging, 4(2), 021102: 1-12 (2017).
[17] Gonzalez, R. C., [Digital Image Processing Using MATLAB], (2004).
[18] Bradley, D. and Roth, G., "Adaptive thresholding using the integral image," Journal of Graphics Tools, 12(2), 13-21 (2007).
[19] Golub, G. H. and Kahan, W., "Calculating the singular values and pseudo-inverse of a matrix," Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis, 2(2), 205–224 (1965).