PROCEEDINGS OF SPIE

Examining structural changes in diabetic nephropathy using inter-nuclear distances in glomeruli: a comparison of variously automated methods

Olivier Simon, Rabi Yacoub, Sanjay Jain, John Tomaszewski, Pinaki Sarder

SPIE.

Examining Structural Changes in Diabetic Nephropathy using Inter-Nuclear Distances in Glomeruli: A Comparison of Variously Automated Methods

Olivier Simon[1], Rabi Yacoub[2], Sanjay Jain[3], John E. Tomaszewski[1], and Pinaki Sarder[1,*]

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

Departments of [3]Medicine – Nephrology,
Washington University School of Medicine

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

ABSTRACT

In diabetic nephropathy (DN), hyperglycemia drives a progressive thickening of and damage to the glomerular filtration surfaces, as well as mesangial expansion and a constriction of capillary lumens. This leads at first to high blood pressure, increased glomerular filtration and micro-proteinuria, and later (if untreated) to severe proteinuria and end-stage renal disease (ESRD). Though, it is well known that DN is accompanied by marked histopathological changes, the assessment of these structural changes is to a degree subjective and hence varies between pathologists. In this work, we make a first study of glomerular changes in DN from a graph-theoretical and distance-based standpoint, using minimal spanning trees (MSTs) and distance matrices to generate statistical distributions that can potentially provide a "fingerprint" of DN. We apply these tools to detect notable differences between normal and DN glomeruli in both human disease and in a streptozotocin-induced (STZ) mouse model. We also introduce an automated pipeline for rapidly generating MSTs and evaluating their properties with respect to DN, and make a first pass at three-dimensional MST structures. We envision these approaches may provide a better understanding not only of the processes underway in DN progression, but of key differences between actual human disease and current experimental models.

Key-words: Diabetic nephropathy, Whole-slide image analysis, Graph, Minimum-spanning tree, Kolmogorov-Smirnov test

INTRODUCTION

Diabetic-related kidney disease, and particularly diabetic nephropathy (DN), is a large and rapidly growing public health problem, with DN accounting for 42% of all patients with end-stage renal disease (ESRD)[1]. Diabetic nephropathy develops in nearly 40% of patients with diabetes and is the leading cause of chronic kidney disease (CKD) in patients with ESRD[2]. There remains no cure, though methods such as control of hyperglycemia and hypertension can help manage the risks of the disease, and new drugs such as SGRT2 inhibitors show promise as well[3]. Further, given that diabetes afflicts nearly 10% of the US population and pre-diabetes is predicted to affect more than a third of the US population at some point in their lifetimes[2, 4], the urgency of developing of better understanding and treatment of DN is clear.

Although a variety of types of changes are known to occur in the kidney over the course of DN, including in metabolism, oxidative stress and gene expression[2, 5], morphological and structural changes in the glomeruli, the key blood-filtering units of the kidney, remain central to appraisal of the disease[6, 7]. These include basement membrane thickening, mesangial expansion and the appearance of dense "Kimmelstein-Wilson nodules"[1]. Therefore, much may be gained by characterizing these changes more objectively.

The tools of graph theory are highly suited to this task. In particular, minimal spanning trees (MSTs), a graph-theoretical construct whereby the points of a network are joined using the shortest total connection distance, have already proven useful for characterization of subtle structural changes in diverse fields[8-10]. They have also proven useful in a surprising range of disease classification tasks in histopathology, such as in cancers of the bladder, uvea, breast, and prostate[1, 11-14]. In particular, MSTs have the advantage of being independent of anisotropic image characteristics[13]. Therefore, MSTs and distance matrices of nuclear positions within glomeruli in DN versus healthy glomeruli are promising candidates to help discover structural hallmarks of DN progression and to aid diagnosis.

In the following work, we pursue the question of whether inter-nuclear distances, particularly in the form of the MST drawn between nuclear centers, have potential as a means of detecting DN progression. We pursue a number of approaches, including manual selection of nuclei and increasingly automated pipelines to speed calculation of the MST. We find that the MST is a subtle but reasonably robust indicator of structural changes in the glomerulus (and possibly the surrounding tissues) in progression from healthy kidney to diabetic nephropathy. We also demonstrate the construction of three-dimensional MSTs of glomerular sections, which hold promise for illuminating aspects of glomerular change that may be overlooked by two-dimensional MSTs.

METHODS AND RESULTS

A standard streptozocin (STZ) treated mouse model[15] was used. C57BL/6J background mice (7 weeks old) were injected with STZ. They develop a mild form of diabetes mellitus type I, and after 25 weeks, mild-moderate DN. All animal studies were performed in accordance with protocols approved by the University at Buffalo, Animal Studies Committee.

Biopsy samples from human diabetic nephropathy patients with chronic kidney disease (CKD) stage II and stage III were collected from Kidney Translational Research Center at Washington University School of Medicine directed by co-author Dr. Jain. The glomerular structural changes in these biopsies suggest DN related changes spanning different DN stages as discussed in Tervaert *et al.*[6]. As control, renal tissue samples of non-diabetic patients with renal cell carcinoma were considered. We used tissue sections with no apparent histological structural changes as verified by Dr. Jain and Dr. Tomaszewski. Human data collection procedure followed a protocol approved by the Institutional Review Board at University at Buffalo.

Tissue slices of 2 µm (for murine data) and 2-5 µm (for human data) were stained using Periodic acid-Schiff, and imaged using a whole-slide imaging scanner (Aperio Versa, Leica, Buffalo Grove, Il). We followed similar imaging protocol as described in our earlier works [16].

From a dataset of images of glomeruli hand-cropped from whole-slide histopathology images, twenty glomeruli were selected from each of the following four classes: control mouse; mouse treated with streptozotocin (STZ), a widely used model of DN in type I diabetes[7]; human control; and human kidney with diabetic nephropathy, stages 2 and 3[6]. These were taken in roughly equal proportions from 7 control mice, 5 STZ mice, 3 human controls and 6 human subjects with diabetic nephropathy. Glomeruli for study were chosen as randomly as possible while still ensuring that they were intact and not duplicates. Next, locations of nuclei were identified by hand and stored as .csv files for each image. From these lists of nuclear locations, distance matrices and MSTs were generated in MATLAB. Probability-normalized histograms of the resulting distance distributions were created after aggregating the distance, and also MST weights for all glomeruli within each of the four classes. Two-sample Kolomogorov-Smirnov (K-S) comparison tests were carried out on the distributions using MATLAB's *kstest2* function.

Locations of nuclei were pinpointed by hand in a total of 80 glomeruli, and then distance matrices and MSTs (Fig. 1) were constructed for each glomerulus. In both human and mouse, the number of nuclei per glomerulus was substantially higher in the DN/STZ cases (Table 1); in human this effect attained statistical significance ($p = 0.0111$, two-sample t-test) while in mouse the significance was lower ($p = 0.1425$). Histograms of the aggregated MST edge lengths showed a noticeable leftward shift in the distribution for human glomeruli going from normal to DN, and a smaller shift in the same direction in mouse (Fig. 2a-b). On the other hand, the distribution of all inter-nuclear distances showed a distinct rightward shift going from normal to DN, in both human and mouse (Fig. 3a-b). The Kolmogorov-Smirnov test indicated that all but the mouse MST shift had extremely high significance (Table 2).
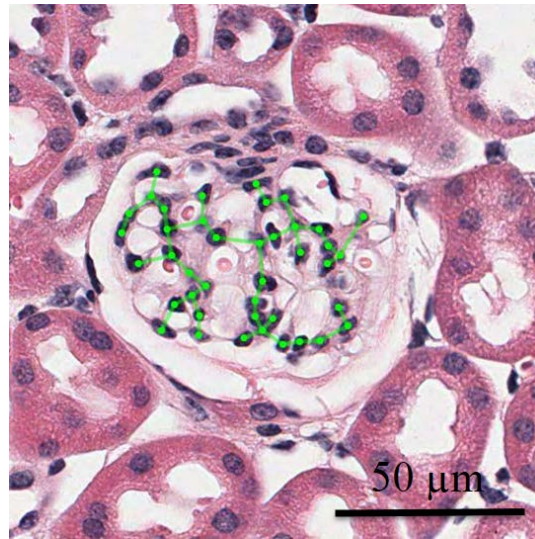
Figure 1. Example of MST constructed between nuclei in a selected mouse glomerulus.
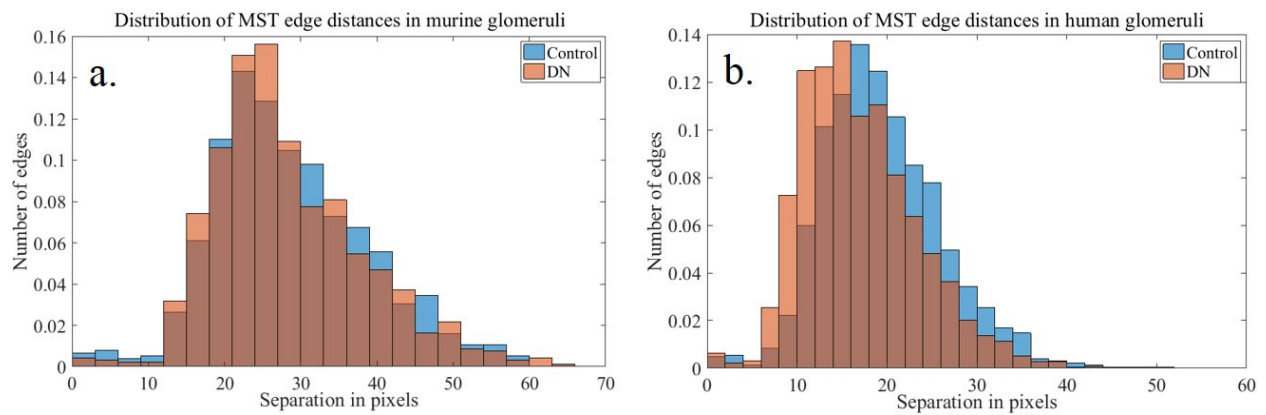

Figure 2. Results for MST edge distances only using hand-picked nuclei. a) Mouse; b) Human.
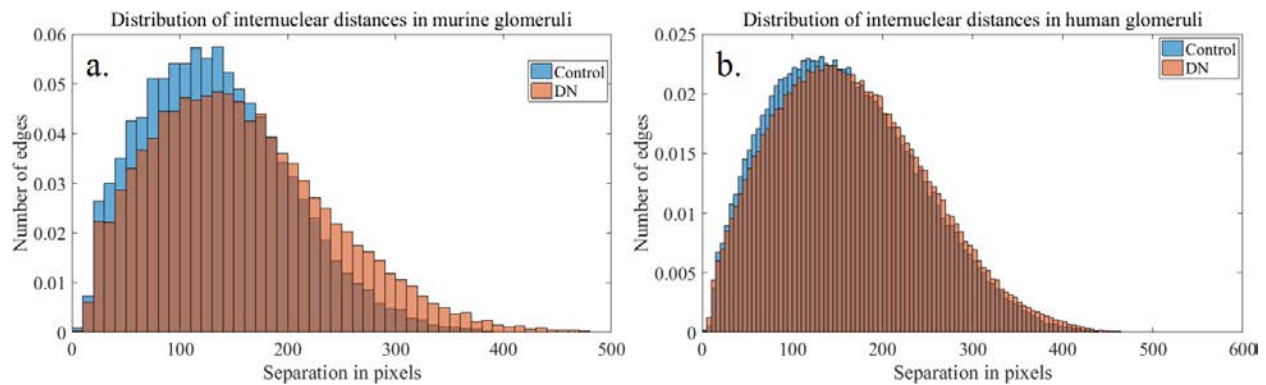

Figure 3. Results for all distances using hand-picked nuclei. a) Mouse; b) Human.

Table 1. Number of nuclei found in each of the four classes studied, for the three different methods used (hand-selected nuclei, semi-automated pipeline, and no-mask pipeline; n = 20 glomeruli for each class.

	Total nuclei	Mean per image	SD
Human, normal (hand-selected)	2260	113	35
Human, DN (hand-selected)	3010	150.5	52.1
Human, normal (semi-automated)	2538	126.9	41.5
Human, DN (semi-automated)	2707	135.4	49.6
Human, normal (no-mask)	7485	374.3	64.7
Human, DN (no-mask)	7856	392.8	40.5
Mouse, control (hand-selected)	775	38.8	13.0
Mouse, DN (hand-selected)	936	46.8	20.2

Table 2. Net shift in the mean values of the distributions for all inter-nuclear distances and for MST edges only, going from normal/control glomeruli to DN/STZ glomeruli. For all methods used, the shift is consistently rightward (positive) for the distribution of all distances, and leftward (negative) for the distribution of MST edges. Kolmogorov-Smirnov test p-values show high statistical significance for all but the mouse-MST shift.

	Net distribution shift (pixels)	p value (K-S test)
Mouse, all distances (hand-selected)	22.0169	5.046×10^{-104}
Mouse, MST edges (hand-selected)	-0.4735	0.15466
Human, all distances (hand-selected)	6.6338	2.2203×10^{-81}
Human, MST edges (hand-selected)	-2.4521	1.8721×10^{-38}
Human, all distances (semi-automated)	4.6965	1.6477e-41
Human, MST edges (semi-automated)	-0.4506	9.2839e-05
Human, all distances (no-mask)	0.9820	1.0441e-19
Human, MST edges (no-mask)	-0.5830	3.2984e-08
Human, MST edges (no-mask, larger image set)	-1.4661	2.1798e-32

Having found that distance metrics such as the full graph and the MST have some potential to distinguish between normal and DN glomeruli, we next sought to speed up the process. Pinpointing nuclear centers one by one by hand is particularly time-intensive and laborious, given that a single glomerular image may contain hundreds of nuclei, which themselves may be irregular in shape or faint or blurred. We therefore settled on a semi-automated approach, whereby the nuclei could be automatically detected within a user-defined region, followed by MST generation.

Our proposed pipeline is outlined in Fig. 4a, and an example of the result at each step is shown in Fig. 4b-g. In summary, the pipeline involves five main steps. First, the glomerular image is subjected to color deconvolution[17] to make the nuclei distinct, using stain vectors extracted from nuclear (hematoxylin) and cytoplasmic (eosin) locations from the images of interest. This is followed by creation of a user-specified mask with MATLAB's *imfreehand* tool to define the boundaries of the glomerulus, so that non-glomerular nuclei can be excluded. The output is thresholded at a value determined by trial and error (0.55 for human and mouse). Next, morphological operations are carried out to clarify likely nuclei, in this case erosion with a disk-shaped structuring element (1 pixel radius for human, 2 for mouse) to help separate overlapping nuclei and size filtering (*bwareaopen* with a cutoff of 10 pixels) to remove stray pixels and fragments. Next, the centroids of the remaining connected objects are identified and an MST is generated between them. Finally, as previously, distance distribution histograms are generated along with statistical testing.
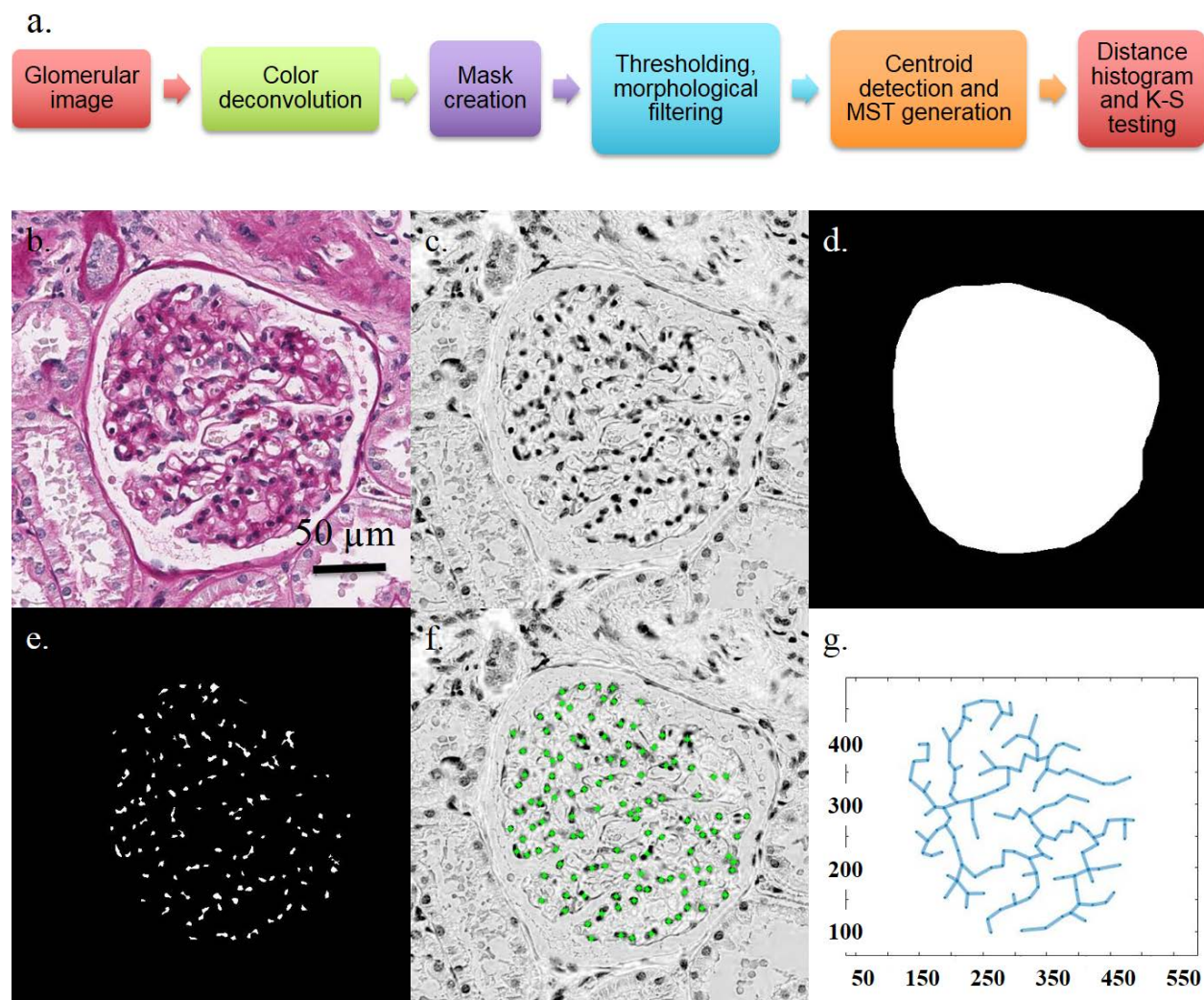

Figure 4. Overview of the semi-automated pipeline for glom internuclear distance analysis. (a) Pipeline. (b-g) Illustration of successive results of our semi-automated pipeline. b) Initial image of a human glomerulus. c) Result

of color deconvolution to make nuclei distinguishable. d) Hand-drawn mask to define area of glomerulus. e) Result of intensity thresholding, size filtering, and morphological erosion. f) Locations of predicted nuclear centroids. g) Minimum spanning tree drawn between the centroids.

Nuclear overlap was a minor issue in the pipeline cases, with the vast majority of nuclei showing clear separation due to the erosion step. Since overlapping nuclei would almost certainly be directly connected in the MST algorithm anyway, we do not believe this affects the overall properties of the resulting MSTs significantly.

Numerical results of the semi-automated approach are shown in Tables 1-2. Table 1 shows that the nuclear count in the semi-automated method was higher in DN than in control, though the difference was somewhat smaller than when nuclei were chosen by hand. Table 2 shows that the semi-automated method achieves a lower significance level and shift size between the DN and control distributions, but that it is nevertheless highly significant. In all cases, we found no significant differences in the distribution of node degree in the MSTs between control and DN glomeruli.

We wondered whether it was necessary to define the boundary of the glomerulus within each sample image in order to obtain a statistically meaningful difference between the DN and the control glomeruli. This was mainly because this step still requires user input to hand-draw the glomerular mask, and therefore is the most labor-intensive remaining part of the process; removing it would effectively make our pipeline completely automatic and hence allow considerably higher throughput. There is also some justification in recent research for considering nuclei outside the glomerulus, as it is increasingly suspected that DN may not be a purely glomerular-based disease, with pathologic changes in the proximal tubules playing a major part in disease's course[3]. Hence, the regions surrounding the glomerulus may conceivably contain useful information about disease progression.

Applying this "no-mask" variation, we found that the glomerular MST is almost always preserved as a subset of the no-mask MST, as is visible in Fig. 5. This fact increased our confidence that the no-mask MST retains the necessary information to detect glomerular changes in DN. Interestingly, analysis of the no-mask MSTs of human glomeruli images by K-S testing showed slightly higher significance than for the semi-automated MSTs as well as a larger shift size, while the full-graph distances showed a smaller effect and decreased significance (Table 2). All semi-automated and no-mask results showed the same previous trend of a DN-driven leftward (negative) shift in the MST distance distribution, and a rightward (positive) shift in the full-graph distribution.
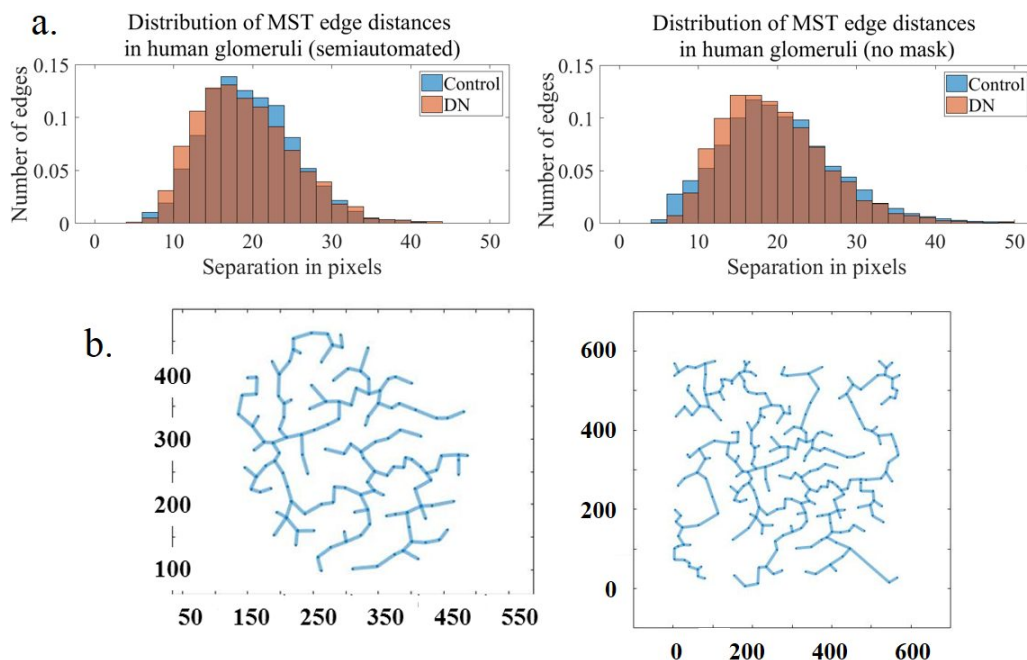

Figure 5. Results of applying the semi-automated pipeline with and without hand-drawn masks. a) In human

glomeruli, MSTs made both with and without masks show the same subtle shift between DN and control as found in the hand-picked nuclei, with control distances on average being slightly larger. b) MSTs derived with hand-drawn glom segmentation masks are typically incorporated intact within the MST derived from using no mask at all, likely explaining the continued though reduced effect even in no-mask MSTs.

Since the fully automatic no-mask arrangement made it practical to test considerably larger sets of glomeruli that our initial group, and appeared even to show slightly improved significance for the MST case than the semi-automated, we decided to try a larger test set of glomerular images. We chose the images from a set of 519 professionally scored images of human glomeruli, and placed them in two groups according to their DN stage. The first group consisted of 49 images of glomeruli with a DN stage of 1 (normal), while the second group consisted of 47 glomeruli with a stage of 4 (severe DN). The results of applying the no-mask MST pipeline to these two groups are shown in the last row of Table 2. It is evident that the effect size and significance has increased considerably with the larger number of images, adding further support to the possibility that the MST distance histogram may be useful as an indicator of DN, even without a glomerular segmentation step.

The glomerulus is not a planar structure, but an intricate 3-dimensional object. Therefore, given that the 2D MST is capable of detecting some DN-related changes in the glomerulus, we wondered about the potential of a 3D MSTs to provide even better characterization of glomerular changes in DN. We therefore used MATLAB's *imregister* function with gradient descent to rigidly register successive slices of the same glomerulus from six serial sections (Fig. 6a), then used a similar method to our semi-automated pipeline to locate nuclear centers. Next, we used the slice number as a z-coordinate in order to generate 3D MSTs (Fig. 6b).
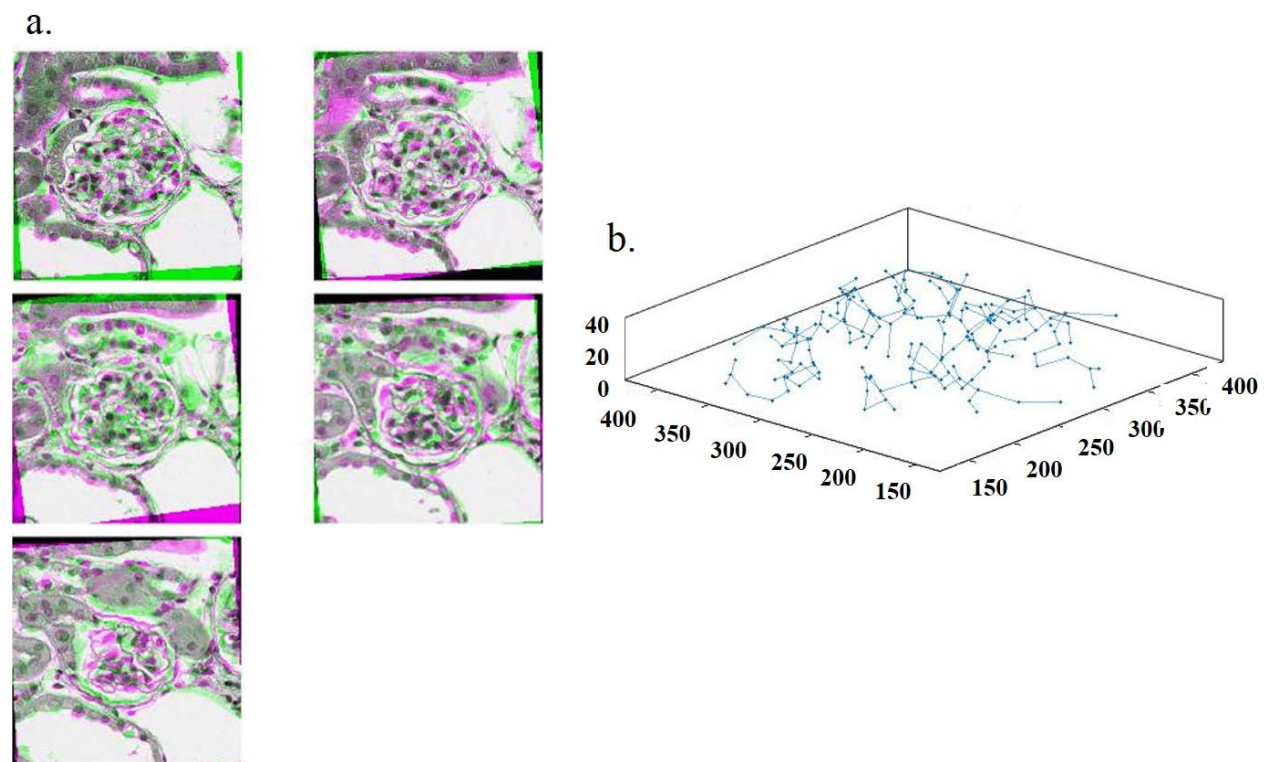

Figure 6. Examples of glom 3D MST generation. a) Result of rigid registration of 6 successive slices of the same glomerulus. b) Stereo view of the resulting 3D MST.

Examining the distance distribution of the 3-dimensional MST in comparison to the aggregated distances of the 2D MSTs generated within each slice, it is very much apparent that the 2D MST badly misrepresents the actual nuclear distance distribution within a glomerulus (Fig. 7). Such differences lend strong preliminary support to the idea that 3D MSTs may provide much richer information about structural properties of the glomerulus in health and disease than the 2D MST.
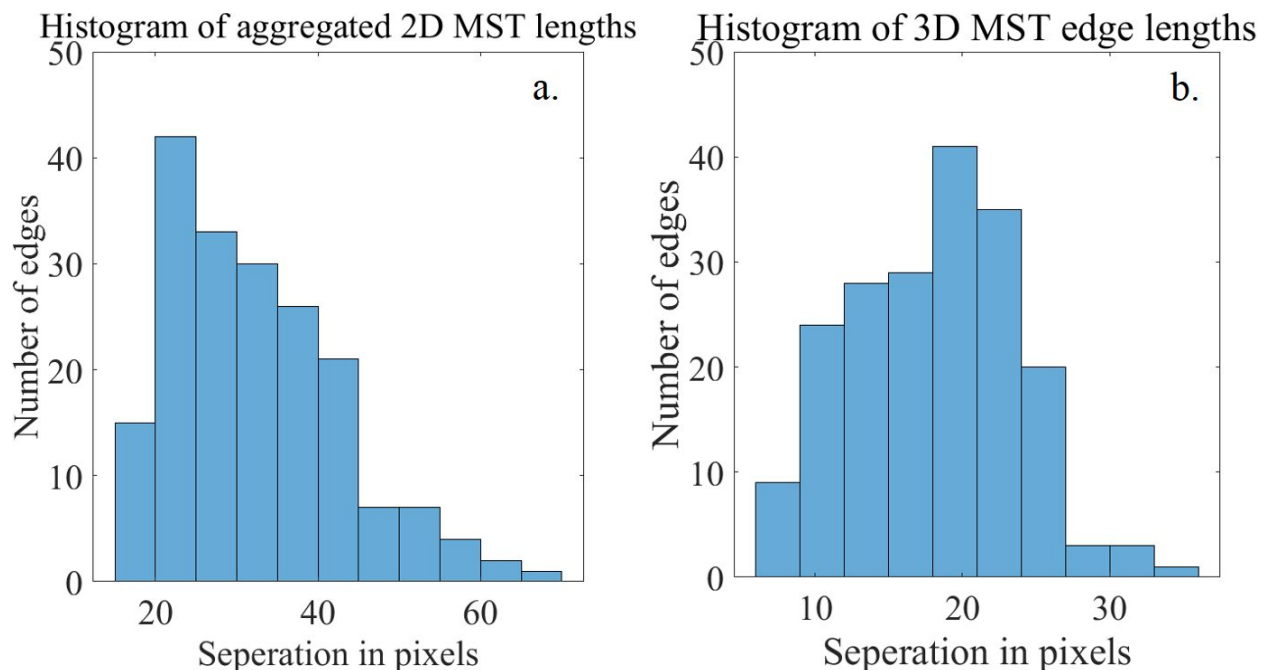
Figure 7. Comparison between interpreting the same glomerulus "stack" of 6 images as an aggregate of 2D MSTs, and a single 3D MST. Notable differences are seen between node degree and average distance.

DISCUSSION

This work represents a first study of the potential of graph-theory based analytics for the study of disease progression in diabetic nephropathy and is, to our knowledge, the first application of minimal spanning trees to glomerular nuclear distributions in diabetic nephropathy. We successfully detected noteworthy statistical differences between normal and DN glomeruli on the basis of intra-glomerular nuclear locations. In particular, we find evidence that human DN is marked by a subtle but significant and consistent decrease in the average values of the MST edges, yet an increase in the average distances among all nuclei. Additionally, we detect substantial increases in number of nuclei in DN in all cases, though these reached higher significance in mouse.

We speculate that the distance-shift effect seen in the graph of all nuclei may be related to the overall expansion of the glomerulus from increased internal pressure as capillary lumens constrict and sclerose, which could increase overall distances slightly despite the increased number of nuclei. In contrast, closer packing of nuclei from increased cell proliferation in the relatively confined region of the mesangium would nonetheless reduce the MST distances. Since the full-graph and MST shifts differ contrariwise in significance between mouse and human, it is plausible that overall expansion is more critical in mouse, while nuclear number and confinement is dominant in the human. We also speculate the differences may be due to the fact that human glomeruli are considerably larger and contain many more nuclei, and perhaps also to differences between the STZ model of DN used in the mouse and the clinical DN present in the human samples.

In general the full graph tended to show more significant differences between control and DN than the MST. Given n nuclei, the full graph contains $n(n-1)/2$ distinct distances in total, while the MST contains only $(n-1)$, meaning that in some sense the full graph provides more raw information for the distance-based classification. The drawback of this is that when dealing with larger numbers of glomeruli or with no-mask cases where all nuclei in an image are utilized, the full graph is both more difficult to conceptualize and can use up computational and particularly memory resources, making the MST more practical. The possibility that the MST may offer a more parsimonious or "compressed" description of the positional changes underway in DN is particularly appealing.

We were further able to fully automate the process of MST generation and distance classification in this work, greatly increasing the potential for application over hand-picked nuclear approaches[13]. We find that changes

in inter-nuclear distances can be consistently found in DN even without specifically masking the glomeruli, but given that this no-mask approach achieved somewhat lower significance and shift size than either the hand-picked or the semi-automated method with hand-drawn masks, and given that the nuclei within the glomerulus are consistently closer packed than nuclei outside and separated by the gap created by the Bowman's space, we believe that graph-cut approaches may offer an effective method for glomerular segmentation to delineate the glomerular region of the MST and, in turn, improve the performance of this method.

Nonetheless, although consistent, the overall effect sizes here were relatively small. The best approach may thus be to combine the MST distances with other geometrical or graphical indicators, such as the Voronoi diagram or the Delaunay triangulation, to create a stronger aggregate disease marker. This kind of multi-feature geometric approach has been used in other areas[11, 12, 18] with significant successes.

Alternatively, we also have presented preliminary results in creation of 3D MSTs, by an extension of the pipeline for 2D MSTs. We believe this to be a useful step particularly for the use of MSTs in DN characterization because, whereas the full distance graph may be taken to differ between 2D slices and 3D in a somewhat predictable way owing to basic geometric considerations, the MST is likely to be substantially different since taking a 2D slice artificially "forces" the MST to be constructed between nuclei within an essentially arbitrary plane that may not be nearest neighbors in 3D space at all.

Furthermore, given the intricately involuted 3-dimensional structure of the glomerular basement membrane, proximity of nuclei tends to outline the "skeleton" of the glomerulus, i.e. the vascular and mesangial tuft where nuclei are concentrated. A three-dimensional MST therefore seems likely to be far more successful in describing this "skeleton", which may better encapsulate clues to the development and pathology of the glomerulus and in turn lead to a much more powerful classifier and effect size.

In conclusion, we have found promising potential for the use of the minimum spanning tree and full graph of inter-nuclear distances for detecting diabetic nephropathy in sets of glomerular images. Future work will involve repeating these studies with larger sample sizes; investigation of other graph features of the MST, such as clustering coefficients, degree divergence, and tree heirarchy overlap[10]; inclusion of larger collections of feature types previously mentioned; and representations of the three-dimensional distribution of nuclei. These, in turn, could evne motivate phenomenological modeling of glomerular nuclear distribution and deep learning using MSTs as training data.

ACKNOWLEDGEMENT

REFERENCES

[1]   Magee, C., Grieve, D. J., Watson, C. J. *et al.*, "Diabetic Nephropathy: a Tangled Web to Unweave," Cardiovasc Drugs Ther, 31(5-6), 579-592 (2017).

[2]   Liu, Z. Z., Bullen, A., Li, Y. *et al.*, "Renal Oxygenation in the Pathophysiology of Chronic Kidney Disease," Front Physiol, 8, 385 (2017).

[3]   Zeni, L., Norden, A. G. W., Cancarini, G. *et al.*, "A more tubulocentric view of diabetic kidney disease," J Nephrol, 30(6), 701-717 (2017).

[4]   National Chronic Kidney Disease Fact Sheet, 2014 Available from:
        https://www.cdc.gov/diabetes/pubs/pdf/kidney_factsheet.pdf

[5]   Papadopoulou-Marketou, N., Chrousos, G. P. and Kanaka-Gantenbein, C., "Diabetic nephropathy in type 1 diabetes: a review of early natural history, pathogenesis, and diagnosis," Diabetes/Metabolism Research and Reviews, 33(2), e2841-n/a (2017).

[6]   Tervaert, T. W., Mooyaart, A. L., Amann, K. *et al.*, "Pathologic classification of diabetic nephropathy," J Am Soc Nephrol, 21(4), 556-63 (2010).

[7]   Kitada, M., Ogura, Y. and Koya, D., "Rodent models of diabetic nephropathy: their utility and limitations," Int J Nephrol Renovasc Dis, 9, 279-290 (2016).

[8]   Dussert, C., Rasigni, G., Rasigni, M. *et al.*, "Minimal spanning tree: A new approach for studying order and disorder," Physical Review B, 34(5), 3528-3531 (1986).

[9]   Rainbolt, J. L. and Schmitt, M., "The use of minimal spanning trees in particle physics," Journal of Instrumentation, 12(02), P02009 (2017).

[10]   Tewarie, P., van Dellen, E., Hillebrand, A. *et al.*, "The minimum spanning tree: An unbiased method for brain network analysis," NeuroImage, 104, 177-188 (2015).

[11]   Wan, T., Cao, J., Chen, J. *et al.*, "Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features," Neurocomput., 229(C), 34-44 (2017).

[12]   Doyle, S., Agner, S., Madabhushi, A. *et al.*, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features." 496-499.

[13]   Coleman, K., van Diest, P. J., Baak, J. P. *et al.*, "Syntactic structure analysis in uveal melanomas," Br J Ophthalmol, 78(11), 871-4 (1994).

[14]   Choi, H.-K., Bengtsson, E., Jarkrans, T. *et al.*, "Minimum spanning trees (MST) as a tool for describing tissue architecture when grading bladder carcinoma." 615-620.

[15]   Tesch, G. H. and Allen, T. J., "Rodent models of streptozotocin-induced diabetic nephropathy," Nephrology (Carlton), 12(3), 261-6 (2007).

[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]   Ruifrok, A. C. and JOHnston, D. A., "Quantification of histochemical staining of color deconvolution," Anal. Quant. Cytol. Histol., 23(4), 291-299 (2001).

[18]   Doyle, S., Hwang, M., Shah, K. *et al.*, "Automated grading of prostate cancer using architectural and textural image features." 1284-1287.