

Reproducibility of the NEPTUNE descriptor-based scoring system on whole-slide images and histologic and ultrastructural digital images

Laura Barisoni^{1,17}, Jonathan P Troost², Cynthia Nast^{3,17}, Serena Bagnasco⁴, Carmen Avila-Casado⁵, Jeffrey Hodgin⁶, Matthew Palmer⁷, Avi Rosenberg⁸, Adil Gasim⁹, Chrysta Liensziewski¹⁰, Lino Merlino¹, Hui-Ping Chien¹¹, Anthony Chang¹², Shane M Meehan¹³, Joseph Gaut¹⁴, Peter Song¹⁵, Lawrence Holzman¹⁶, Debbie Gibson², Matthias Kretzler¹⁰, Brenda W Gillespie¹⁵ and Stephen M Hewitt^{8,17}

¹Department of Pathology, University of Miami, Miller School of Medicine, Miami, FL, USA; ²Department of Pediatrics, Division of Pediatric Nephrology, University of Michigan, Ann Arbor, MI, USA; ³Department of Pathology, Cedars-Sinai Medical Center, Los Angeles, CA, USA; ⁴Department of Pathology, Johns Hopkins University, Baltimore, MD, USA; ⁵Department of Pathology, University of Toronto, Ontario, Canada, USA; ⁶Department of Pathology, University of Michigan, Ann Arbor, MI, USA; ⁷Department of Pathology, University of Pennsylvania, Philadelphia, PA, USA; ⁸Laboratory of Pathology, National Cancer Institute, Bethesda, MD, USA; ⁹Department of Pathology, University of North Carolina, Chapel Hill, NC, USA; ¹⁰Department of Medicine, Division of Nephrology, University of Michigan, Ann Arbor, MI, USA; ¹¹Department of Pathology, Keelung Chang Gung Memorial Hospital, Keelung, Taiwan; ¹²Department of Pathology, University of Chicago, Chicago, IL, USA; ¹³Department of Pathology, Sharp Memorial Hospital, San Diego, CA, USA; ¹⁴Department of Pathology, Washington University, St. Louis, MO, USA; ¹⁵Biostatistics Department, School of Public Health, University of Michigan, Ann Arbor, MI, USA and ¹⁶Department of Medicine, Division of Nephrology, University of Pennsylvania, Philadelphia, PA, USA

The multicenter Nephrotic Syndrome Study Network (NEPTUNE) digital pathology scoring system employs a novel and comprehensive methodology to document pathologic features from whole-slide images, immunofluorescence and ultrastructural digital images. To estimate inter- and intra-reader concordance of this descriptor-based approach, data from 12 pathologists (eight NEPTUNE and four non-NEPTUNE) with experience from training to 30 years were collected. A descriptor reference manual was generated and a webinar-based protocol for consensus/cross-training implemented. Intra-reader concordance for 51 glomerular descriptors was evaluated on jpeg images by seven NEPTUNE pathologists scoring 131 glomeruli three times (Tests I, II, and III), each test following a consensus webinar review. Inter-reader concordance of glomerular descriptors was evaluated in 315 glomeruli by all pathologists; interstitial fibrosis and tubular atrophy (244 cases, whole-slide images) and four ultrastructural podocyte descriptors (178 cases, jpeg images) were evaluated once by six and five pathologists, respectively. Cohen's kappa for inter-reader concordance for 48/51 glomerular descriptors with sufficient observations was moderate ($0.40 < \text{kappa} \leq 0.60$) for 17 and good ($0.60 < \text{kappa} \leq 0.80$) for 8, for 52% with moderate or better kappas. Clustering of glomerular descriptors based on similar pathologic features improved concordance. Concordance was independent of years of experience, and increased with webinar cross-training. Excellent concordance was achieved for interstitial fibrosis and tubular atrophy. Moderate-to-excellent concordance was achieved for all ultrastructural podocyte descriptors, with good-to-excellent concordance for descriptors commonly used in clinical practice, foot process effacement, and microvillous transformation. NEPTUNE digital pathology scoring system enables novel morphologic profiling of renal

Correspondence: Dr L Barisoni, MD, Department of Pathology, University of Miami, Miller School of Medicine, 1400 NW 12th Avenue, Room 4079, Miami, FL 33136, USA or B Gillespie, CSCAR, 3550 Rackham Building, 915 E Washington St., University of Michigan, Ann Arbor, MI 48109 USA.

E-mail: lbarisoni@med.miami.edu or bgillesp@umich.edu

¹⁷Neptune leading pathologists for digital imaging and cases collection in the Neptune Digital Pathology Repository. Received 3 November 2015; revised 12 February 2016; accepted 13 February 2016; published online 22 April 2016

structures. For all histologic and ultrastructural descriptors tested with sufficient observations, moderate-to-excellent concordance was seen for 31/54 (57%). Descriptors not sufficiently represented will require further testing. This study proffers the NEPTUNE digital pathology scoring system as a model for standardization of renal biopsy interpretation extendable outside the NEPTUNE consortium, enabling international collaborations. *Modern Pathology* (2016) 29, 671–684; doi:10.1038/modpathol.2016.58; published online 22 April 2016

The challenge of inter-reader concordance on individual morphologic features of diagnostic renal biopsies is well documented and is highlighted in large collaborative studies.^{1–5} As the complexity of morphologic characterization and the number of features increase, it becomes more difficult to ensure intra- and inter-reader concordance. As one feature may show poor performance, a related and potential surrogate feature may show excellent performance and thus be preferable for routine diagnostic use. In the future, conventional interpretative diagnoses may be revised to include combined morphologic and molecular signatures.^{1,6,7} With these changes in the pathology practice, it is important to assess the performance of individual metrics. The past approach has been to develop metrics that demonstrate high intra-pathologist concordance and high to good inter-pathologist concordance.

The availability of digital whole-slide images allows nephropathologists to overcome limitations of conventional light microscopy analysis, and to address concordance.^{8,9} Recent studies have demonstrated the high concordance and reliability of whole-slide images compared with conventional light microscopy evaluation for diagnoses of renal allograft rejection, as well as for individual Banff morphologic criteria.^{10,11} Morphologic analysis of annotated peritubular capillaries on whole-slide images in Fabry's disease suggests that by pre-selecting specific structures to be scored, achievable only by digital imaging, concordance is increased.¹²

The multicenter Nephrotic Syndrome Study Network (NEPTUNE) exemplifies a new model of systematic digital pathology review. The NEPTUNE Digital Pathology Protocol documents the whole-slide images-based scoring protocol, including selection of specific structures (eg, glomeruli) and the application of the NEPTUNE Digital Pathology Scoring System for comprehensive scoring of glomerular, vascular and tubulointerstitial morphologic features (descriptors).^{13,14}

This study aimed to assess inter- and intra-reader concordance, and the effect of consensus review and training sessions on the NEPTUNE Digital Pathology Scoring System. The ultimate goal is to establish new models for standardization of renal biopsy morphologic profiling, and to test validated descriptors as potential predictors of diagnosis, prognosis, and response to treatment.

Materials and methods

Digital Infrastructure

Pathology material was obtained from the NEPTUNE Digital Pathology Repository, where whole-slide images (from glass slides scanned at 40× on Hamamatsu and Aperio scanners), immunofluorescence, electron microscopy (EM) images, electronic copies of de-identified original pathology reports from cases of focal segmental glomerulosclerosis, minimal change disease, and membranous nephropathy are stored.^{15,13}

Preparation and Training for Scoring

Descriptor reference manual and image library. A reference manual was generated and refined by webinar consensus meetings by NEPTUNE pathologists (Supplementary Figure 1). Descriptors were evaluated for clarity prior to initiation of the concordance tests. The manual was posted in the NEPTUNE digital pathology repository (see Table 1 for descriptors used in this study and Supplementary Table 6 for the comprehensive descriptor reference manual). A library of representative images was created and posted in the NEPTUNE digital pathology repository for independent review prior to initiation of the study, and then removed during the trial.

Electronic scoring documents, material, and test instructions. Separate electronic scoring templates were generated for tubulointerstitial, ultrastructural, and glomerular scoring. The electronic matrix templates were pre-populated with '0' (absent) scoring, so reviewers needed only to select descriptors applicable to a given case/image; semiquantitative or quantitative scores required clicking on a dropdown list. For better visualization, the color of the selected cell automatically changed when a value other than 0 was selected ('0' = blue to '1' = red) (Supplementary Figure 2).

For glomerular scoring, electronic scoring templates included the list of glomeruli, and jpeg images of glomeruli were provided to all pathologists. Separate electronic scoring sheets with the lists of cases to access in the NEPTUNE digital pathology repository were provided to test tubulointerstitial descriptors such as interstitial fibrosis/tubular atrophy and ultrastructural podocyte features. Specific instructions for each of the metrics were made

Table 1 Post study revised definitions of descriptors included in the current study**WSI histology****Glomerular damage**

Glomerular descriptors listed below are scored as present (1) or absent (0).

No (minimal) changes: none of the lesions below is present.

Global sclerosis with hyalinosis: sclerosis involves 100% of the glomerular tuft glomerular size is preserved, or compared with the glomeruli obtained in the same biopsy, increased or decreased by not > 50%.

Global sclerosis without hyalinosis: sclerosis involves 100% of the glomerular tuft, with no accompanying hyalinosis. Glomerular size is preserved, or compared with the glomeruli obtained in the same biopsy, increased or decreased by not > 50%.

Global deflation: global wrinkling and folding of the GBM ($\geq 80\%$ of the tuft) without epithelial cell (podocyte) hypertrophy and hyperplasia (formerly known as ischemic type of collapse). The urinary space is patent. The wrinkling is generally made by small regular folds of the GBM.

Global capillary collapse: wrinkling and folding of the GBM involving $\geq 80\%$ of the tuft with occlusion or subocclusion of capillary lumina. Collapse is generally accompanied by hypertrophy and hyperplasia of overlying epithelial cells (pseudo-crescents). Epithelial cell (podocyte) hypertrophy and hyperplasia if present are marked separately as individual descriptors. The wrinkling is generally made by small and/or big irregular folds of the GBM.

Obsolescent glomeruli: glomeruli are small and globally sclerotic without hyalinosis. Bowman's capsule is completely or partially absent and there is no periglomerular fibrosis. Obsolescent glomeruli are defined when glomerular size is decreased > 50% compared with all other glomeruli in the same biopsy.

Global mesangial sclerosis: a generalized global increase (100%) of mesangial matrix is present with or without mesangial cell hypercellularity and hypertrophy of overlying epithelial cells (podocytes).

Segmental perihilar sclerosis (vascular pole): segmental solidification of the glomerular tuft is present with increased extracellular matrix in continuity with the vascular pole. If hyalinosis, foam cells, hypertrophy of overlying epithelial cells (podocytes), halo and adhesion of the tuft to the Bowman's capsule is present, they should be marked as separate descriptors.

Extended segmental perihilar sclerosis (vascular pole): segmental solidification of the glomerular tuft with increased extracellular matrix in continuity with the vascular pole and extends beyond the middle line of the tuft, with or without involving the tip. If hyalinosis, foam cells, hypertrophy of overlying epithelial cells (podocytes), halo and synechia/adhesion of the tuft to the Bowman's capsule is present, they should be marked as separate descriptors. (This lesion includes segmental solidification known as 'approaching' global sclerosis).

Segmental sclerosis away from vascular and tubular poles: segmental solidification of the tuft with increased extracellular matrix. If hyalinosis, foam cells, hypertrophy of overlying epithelial cells (podocytes), halo and adhesion of the tuft to the Bowman's capsule is present, these features should be marked as separate descriptors.

Segmental sclerosis cannot determine location: None of the above. Vascular or tubular pole cannot be seen in section. If hyalinosis, foam cells, hypertrophy of overlying epithelial cells (podocytes), halo and adhesion of the tuft to the Bowman's capsule is present, they should be marked as separate descriptors.

Cellular tip lesion: foam cells with or without other intracapillary cells within the glomerular tuft at the tubular pole, accompanied by hypertrophy of glomerular epithelial cells (podocytes) exclusive of tubular epithelium, and/or bridging to the Bowman's capsule/proximal tubule take off area. The presence of foam cells or inflammatory cells needs to be marked separately as individual descriptors when present.

Sclerosing tip lesion: solidification of the tuft at the tubular pole with increased extracellular matrix with or without adhesion to Bowman's capsule. Glomerular epithelial cells (podocytes) may be hypertrophic and attached to the epithelium at the tubular pole.

Extended cellular tip lesion: foam cells with or without other intracapillary cells within the glomerular tuft at the tubular pole. The process extends through a large portion of the glomerulus (> 1/2 of the tuft) but does not involve the vascular pole, accompanied by hypertrophy of epithelial cells (podocytes) and/or bridging to the Bowman's capsule/proximal tubule take off area. The presence of foam cells or inflammatory cells needs to be marked separately as individual descriptors when present. (This lesion includes segmental solidification not involving the vascular pole but 'approaching' global sclerosis).

Extended sclerosing tip lesion: solidification of the tuft at the tubular pole with increased extracellular matrix and adhesion to Bowman's capsule, which extends through a large portion of the glomerulus but does not involve the vascular pole. No foam cells are present. Glomerular epithelial cells (podocytes) are hypertrophic and attached to epithelial cells at the tubular pole. (This lesion includes segmental solidification not involving the vascular pole but 'approaching' global sclerosis).

Mid-tuft/central location of segmental sclerosis: located neither at the tip, the perihilum, or the periphery of the tuft (no adhesion to the Bowman's capsule).

Cellular lesions—non-tip: endocapillary hypercellularity with epithelial cell (podocyte) hypertrophy. Hypercellularity may be due to foam cells and/or endocapillary cells with or without karyorrhexis and is not at the tip of the glomerulus. The presence of foam cells, karyorrhexis, or inflammatory cells need to be marked separately as individual descriptors when present.

Segmental capillary collapse: wrinkling and folding of the GBM involving at least one glomerular lobule and < 80% of the tuft, with occlusion or subocclusion of capillary lumina. Collapse is generally accompanied by hypertrophy and hyperplasia of overlying epithelial cells (podocytes); epithelial cell (podocyte) hypertrophy and hyperplasia if present need to be marked separately as individual descriptors. The wrinkling is generally made by small and/or big irregular folds of the GBM.

Segmental deflation: wrinkling and folding of the capillaries without epithelial cell (podocyte) hyperplasia (formerly called ischemic type of collapse) involving < 80% of the glomerular tuft. The wrinkling is generally made by small regular folds of the GBM.

Periglomerular fibrosis: circumferential fibrosis in the interstitium surrounding the Bowman's capsule.

Glomerular foam cells: intracapillary foam cells in the presence or absence of segmental or global sclerosis.

Hyaline droplets in epithelial cell (podocyte): protein droplets are present in glomerular epithelial cells (podocytes). These cells usually are also hypertrophic (if so, both descriptors apply).

Hyalinosis at the vascular pole: hyalinosis is defined as glassy acidophilic, PAS-positive, silver-negative material.

Hyalinosis at the tubular pole: hyalinosis is defined as glassy acidophilic, PAS-positive, silver-negative material. Solidification of the tuft and/or foam cells may be present.

Hyalinosis away from the vascular and tubular poles: hyalinosis is defined as glassy acidophilic, PAS-positive, silver-negative material. Both the vascular and the tubular pole are present in the glomerular cross-section.

Hyalinosis cannot determine location: hyalinosis is defined as glassy acidophilic, PAS-positive, silver-negative material and can occur with or without adhesion to the Bowman's capsule in a location that is not the vascular pole of the tip of the glomerulus. The vascular and/or the tubular poles cannot be identified.

Synechiae: continuity of glomerular tuft basement membrane to the Bowman's capsule with continuity of epithelial cell lining. Note: a synechia generally includes 1–2 capillaries at the most and it may or may not be associated with segmental sclerosis, hyalinosis or foam cells. Larger adhering section of the glomerular tuft to the Bowman's capsule in the presence of significant hyalinosis and/or sclerosis is not considered a synechia but an adhesion part of the segmental sclerosis.

Table 1 (Continued)

- Segmental epithelial cell (podocyte) hypertrophy:** hypertrophy is defined as enlarged cytoplasm or enlarged nuclei with prominent nucleoli or both. Segmental hypertrophy is defined when enlarged epithelial cells (podocytes) overlying the GBM involve < 50% of the glomerular tuft.
- Global epithelial cell (podocyte) hypertrophy:** hypertrophy is defined as enlarged cytoplasm or enlarged nuclei with prominent nucleoli or both. Global hypertrophy is defined when enlarged epithelial cells (podocytes) overlying the GBM involve \geq 50% of the glomerular tuft.
- Segmental epithelial cell (podocyte) hyperplasia:** \geq 2 layers of epithelial cells (podocytes) overlying the GBM are present, involving < 50% of the glomerulus. Hyperplasia may occur with or without hypertrophy.
- Global epithelial cell (podocyte) hyperplasia:** \geq 2 layers of epithelial cells (podocytes) overlying the GBM are present, involving \geq 50% of the glomerulus. Hyperplasia may occur with or without hypertrophy.
- Halo (detachment of overlying podocytes):** detachment of epithelial cells (podocytes) from original underlying GBM is present with intervening new loose basement membrane material (pale on HE, PAS, trichrome, or silver stain).
- Segmental mesangial hypercellularity:** > 3 mesangial cells per mesangial lobule involving < 50% of the visible mesangial regions in a glomerulus.
- Global mesangial hypercellularity:** > 3 mesangial cells per mesangial lobule involving \geq 50% of the visible mesangial regions in a glomerulus.
- Segmental presence of spikes on silver stain:** spikes are defined as silver positive stain with an irregular profile on the outer side of the GBM and involving < 50% of the glomeruli.
- Global presence of spikes on silver stain:** spikes are defined as silver positive stains with an irregular profile on the outer side of the GBM involving \geq 50% of the glomerulus.
- Infiltrating leukocytes:** the presence of leukocytes in glomerular capillaries is recorded when \geq 1 inflammatory cell is present in capillary lumina.
- Segmental endocapillary hypercellularity:** hypercellularity owing to increased number of cells within glomerular capillary lumina causing narrowing of the lumina involving < 50% of the glomerulus.
- Global endocapillary hypercellularity:** hypercellularity owing to increased number of cells within glomerular capillary lumina causing narrowing of the lumina involving > 50% of the glomerulus.
- Segmental GBM Duplication:** is defined as a double contour of the GBM involving < 50% of the glomerular tuft, with or without endocapillary hypercellularity (endocapillary hypercellularity is independent variable).
- Global GBM duplication:** is defined as a double contour of the GBM involving > 50% of the glomerular tuft, with or without endocapillary hypercellularity (endocapillary hypercellularity is independent variable).
- Segmental increased mesangial matrix:** Defined as an increase in the extracellular material in the mesangium such that the width of the interspace exceeds two mesangial cell nuclei in at least one glomerular lobule but < 50% of the glomerular tuft.
- Global increased mesangial matrix:** defined as an increase in the extracellular material in the mesangium such that the width of the interspace exceeds two mesangial cell nuclei in \geq 50% of the glomerular tuft.
- Karyorrhexis:** presence of apoptotic, pyknotic, and/or fragmented nuclei.
- Necrosis:** defined as disruption of the GBM with fibrin exudation and karyorrhexis.
- Very segmental extracapillary cellular proliferation (cellular crescent):** extracapillary cell proliferation of > 2 cell layers with > 50% of the lesion occupied by cells, involving < 25% of the Bowman's space.
- Very segmental extracapillary fibrocellular proliferation (fibrocellular crescent):** defined as part of the circumference of Bowman's capsule covered by a combination of cells and extracellular matrix, with < 50% cells and < 90% matrix involving < 25% of the Bowman's space. This lesion is often associated with disruption of Bowman's capsule. Ischemic, obsolescent glomeruli should be excluded.
- Very segmental extracapillary fibrosis (fibrous crescent):** defined as > 10% of the circumference of Bowman's capsule covered by a lesion composed of > 90% extracellular matrix involving < 25% of the Bowman's space.
- Extensive extracapillary cellular proliferation (cellular crescent):** extracapillary cell proliferation of more than two cell layers with > 50% of the lesion occupied by cells, involving > 25% of the Bowman's space.
- Extensive extracapillary fibrocellular proliferation (fibrocellular crescent):** defined as part of the circumference of Bowman's capsule covered by a combination of cells and extracellular matrix, with < 50% cells and < 90% matrix involving > 25% of the Bowman's space. This lesion is often associated with disruption of Bowman's capsule. Ischemic, obsolescent glomeruli should be excluded.
- Extensive extracapillary fibrosis (fibrous crescent):** defined as > 10% of the circumference of Bowman's capsule covered by a lesion composed of > 90% extracellular matrix involving > 25% of the Bowman's space.

Tubulointerstitial damage

Tubular atrophy: small tubules with thick tubular basement membranes lined by small cuboidal or flat cells. Generally accompanied by fibrosis. Includes 'thyroidization' of the parenchyma. (0 = absent, 1 = present).

Interstitial fibrosis: the interstitium is expanded by the presence of collagen that stain blue on trichrome. Tubules are not back to back, but rather separated by fibrosis and can be atrophic. (0 = absent, 1 = present).

Electron microscopy

A minimum of five electron micrographs are reviewed.

Foot process effacement: Loss of foot processes. % of glomerular capillary surface area affected by effacement is recorded as a semiquantitative value. (0 = 0–10%; 1 = 11–25%; 2 = 26–50%; 3 = 51–75%; 4 = 76–100% of the outer GBM surface).

Condensation of the actin-based cytoskeleton: electron-dense cytoskeleton is reorganized and condensed at the GBM aspect of epithelial cell (podocyte) foot processes. Percentage of glomerular capillary surface area affected by effacement is recorded as semiquantitative value. (0 = 0–5%; 1 = \leq 50%; 2 = > 50% of the outer GBM surface).

Microvillous transformation: cytoplasmic projections into the urinary space that emanate from the luminal side of epithelial cell (podocyte) membrane are present. Percentage of glomerular capillary surface area affected by effacement is recorded as semiquantitative value (0 = 0–5%; 1 = \leq 50%; 2 = > 50% of the outer GBM surface).

Loss of primary processes: epithelial cell (podocyte) body sits directly on underlying GBM. This is generally accompanied by complete effacement (loss of foot processes). It is recorded as present or absent (0 = present – normal; 1 = absent – loss).

GBM, Glomerular Basement Membranes.

Note: The Kappa values are included in the Tables and Supplementary Tables for each descriptor.

available. Training for data entry on the electronic scoring sheets was done during webinar meetings prior to the concordance tests.

Concordance Study Protocol

Image selection. For glomerular histologic descriptors, jpeg (joint photographic experts group) images (stained with hematoxylin and eosin, periodic acid Shift, trichrome, and silver) were obtained from both annotated whole-slide images from the NEPTUNE digital pathology repository and images previously used in a concordance study for the Columbia classification.¹⁶ For tubulointerstitial and ultrastructural podocyte descriptors, whole-slide images and EM jpeg images stored in the NEPTUNE digital pathology repository were used. All images were from previously anonymized whole-slide images or EM digital images collected in the NEPTUNE digital pathology repository following Institutional Review Board guidelines and upon approval in each participating center.

A total of 315 images of glomeruli were hand selected based on quality of the image and representation of descriptors; these images included classic examples as well as more controversial lesions. Interstitial fibrosis and tubular atrophy scoring was tested on whole-slide images from 244 cases including minimal change disease, focal segmental glomerulosclerosis, and membranous nephropathy and podocyte descriptors on 178 ultrastructural images (minimum of five EM images/case) from the minimal change disease/focal segmental glomerulosclerosis cohort.

Participating pathologists. Twelve pathologists participated in the scoring, including eight NEPTUNE pathologists (P1–8, seven of whom participated in glomerular scoring, five in interstitial fibrosis and tubular atrophy scoring, and five in podocyte scoring) and four pathologists recruited outside the NEPTUNE consortium (non-NEPTUNE pathologists) (P9–12, of whom three participated in glomerular scoring and one in interstitial fibrosis and tubular atrophy scoring). The level of experience varied between fellowship level (P7 and P9) to >10 years of experience in renal pathology (Supplementary Table 1).

Glomerular descriptor concordance tests. To assess intra- and inter-reader concordance and the effect of cross-training/consensus review on inter-reader concordance, 131 images of glomeruli were scored three times (Test I, II, and III) by seven NEPTUNE pathologists (Supplementary Figure 1). Webinar reviews occurred 2–4 weeks after each test. Washout intervals between tests varied from 2.5 to 4 months. To increase the number of inter-reader observations, 184 additional images were added to Test II for NEPTUNE pathologists. The 315 images were also scored once by three non-NEPTUNE pathologists

who had one webinar training session. Images of the 131 glomeruli for Tests I and III and 315 glomeruli for Test II were reviewed during consensus webinar meetings to increase concordance in descriptor recognition.

Intra-reader concordance of glomerular descriptors was estimated by comparing each pathologist's scores from Test I vs Test II and Test II vs Test III. These estimates of concordance may be reduced as a result of webinar training; ie, gained knowledge about scoring may reduce consistency with previous scoring.

Inter-reader concordance of descriptors was estimated separately for each Test (I, II, and III), and involved computing concordance for each pair of pathologists, and pooling these estimates over all possible pairs. In addition to the overall estimate of inter-reader concordance, we were interested in four research questions: (a) whether continuous cross-training improved concordance, (b) whether concordance differed by the pathologist's experience, (c) whether concordance was higher using clusters of descriptors sharing similar features than for individual descriptors, and (d) whether concordance was maintained outside the NEPTUNE investigators.

Tubulointerstitial descriptor concordance tests. To test concordance of non-glomerular parameters, we considered the most clinically relevant tubulointerstitial parameters,^{17–22} the percentage (0–100%) of cortex involved by interstitial fibrosis and tubular atrophy, for 244 cases. Conventional pathology practice includes semiquantitative assessment of interstitial fibrosis and tubular atrophy. Therefore, interstitial fibrosis and tubular atrophy scoring was not preceded by webinar training and was performed only once by six pathologists.

Podocyte descriptor concordance tests. Although ultrastructural evaluation of podocyte morphology is common in pathology practice, estimates of some ultrastructural parameters are often not reported. Thus, the podocyte descriptor test was preceded by a webinar session to review definitions reflecting effacement, condensation of actin-based cytoskeleton, microvillous transformation, and loss of primary processes. Ultrastructural podocyte descriptors were scored by five pathologists with 1 to >10 years of experience on 178 cases (minimal change disease/focal segmental glomerulosclerosis) as follows: foot process effacement: 0 = 1–10%, 1+ = 11–25%, 2+ = 26–50%, 3+ = 51–75%, and 4+ = >75%; condensation of actin-based cytoskeleton and microvillous transformation: 0 = not observed, 1+ = segmental ($\leq 50\%$), 2+ = global ($> 50\%$); loss of primary processes was scored as absent (0) or present (1+).

Statistical Methods

For the (dichotomous) glomerular descriptors, intra-reader agreement was assessed by both Cohen's kappa and pathologist-specific counts of the number

of descriptors that the pathologist rated the same way in two consecutive readings. Inter-reader agreement between pairs of pathologists was also estimated using Cohen's kappa, but Fleiss' kappa²³ was used to estimate inter-reader agreement pooled across all pathologists. The variability in kappa values across pairs of pathologists for each descriptor is shown using boxplots.

Scoring was also performed for clusters of glomerular descriptors sharing morphologic similarities. A cluster was judged to be present if at least one descriptor of the cluster was present, and Fleiss' kappa was used to assess pooled inter-reader agreement among pathologists. The kappa statistic ranges from -1 (perfect disagreement) to 1 (perfect agreement), with a value of 0 indicating agreement expected by chance alone. Kappa statistics were categorized and interpreted as: >0.80 (excellent); 0.61–0.80 (good); 0.41–0.60 (moderate); 0.21–0.40 (fair); 0–0.20 (poor); and <0 (no agreement) (<http://healthcare-economist.com/2011/11/02/kappa-statistic>). Because kappa is smaller with lower prevalence of the finding under observation, we report the range over pathologists of the number of glomeruli in which each descriptor was observed. Although we calculated kappa statistics for all descriptors with at least one pathologist rating, some results exclude descriptors with insufficient observations, defined as the maximum over all pathologists of the number of glomeruli in which the descriptor was observed being less than five.

We investigated the four research questions listed above as follows: (a) to assess whether inter-reader concordance could improve with cross-training we evaluated the number of descriptors that increased in concordance between Tests I and II, and between Tests I and III; (b) to assess whether inter-reader concordance depended on the pathologist's years of experience, we compared the kappas from all pathologists with the kappas excluding the trainees; (c) to assess the effect of scoring descriptor clusters, we visually compared cluster concordance with individual concordance estimates for each descriptor in the cluster; and d) to assess whether concordance was maintained outside NEPTUNE investigators we compared concordance among the three non-NEPTUNE pathologists and among the seven NEPTUNE pathologists using the 315 glomerular images from Test II. The Neptune and non-Neptune summary kappas were compared by paired *t*-test.

For the continuous interstitial fibrosis and tubular atrophy scores, inter-reader agreement was estimated using Pearson's correlation coefficient on all pairs of pathologists (the pathologist with more vs less years of experience). For the ordinal podocyte descriptors, Kendall's coefficient of concordance was used to assess inter-reader agreement for pairs of pathologists.

Results

Intra-reader Concordance for Glomerular Descriptors

When comparing glomerular intra-reader concordance Test I vs Test II and Test II vs Test III, the average intra-reader concordance for glomerular descriptors increased with cross-training/consensus webinars. (Supplementary Table 2 and Supplementary Table 3). When comparing glomerular concordance Test II versus Test III, there were four descriptors for which all pairs of readers had good concordance, and 11 descriptors where all pairs had at least moderate concordance (Supplementary Table 3). Interestingly, inconsistent intra-reader concordance was noted for lesions of segmental sclerosis corresponding to 'perihilar' and 'not otherwise specified' variants of the Columbia classification.²⁴ At least moderate intra-reader agreement was found for most of the descriptors commonly associated with segmental sclerosis or collapse, such as various form of hyalinosis, podocyte hypertrophy, foam cells or periglomerular fibrosis. Unexpected inconsistency in intra-reader agreement was noted for basic lesions such as global sclerosis, although other forms of global damage (obsolescence, global collapse, deflation and spikes) were more consistently recognized.

Inter-reader Concordance for Glomerular Descriptors

For the 315 glomeruli (Test II), 48/51 glomerular descriptors had sufficient data for evaluation. The kappa statistics from the combined NEPTUNE and non-NEPTUNE pathologists represent our current best summaries of this investigation. Based on these results, 8/48 descriptors had good inter-reader concordance; these included descriptors indicating global lesions (global spikes, deflation, collapse, and obsolescence) and segmental lesions (foam cells, cellular tip lesion, segmental deflation necrosis). An additional 17/48 descriptors had moderate concordance for a total of 52% of descriptors tested having an inter-reader Cohen's kappa ≥ 0.40 (Table 3). Concordance between pairs of pathologists varied widely by pair and by descriptor, but most had moderate or better concordance (Figure 1a and b).

The overall inter-reader concordance increased with cross-training from Test I through Test III among NEPTUNE pathologists in the set of 131 glomerular images. Of the 51 glomerular descriptors tested, 19 were not sufficiently represented to evaluate inter-reader concordance. For 32 descriptors with sufficient data for comparison, 56% had improved kappas between Tests I and II, and 63% between Tests I and III. Five descriptors improved the initial kappas of moderate to good or excellent, including global lesions (such as global deflation), segmental lesions (mid-glomerular segmental

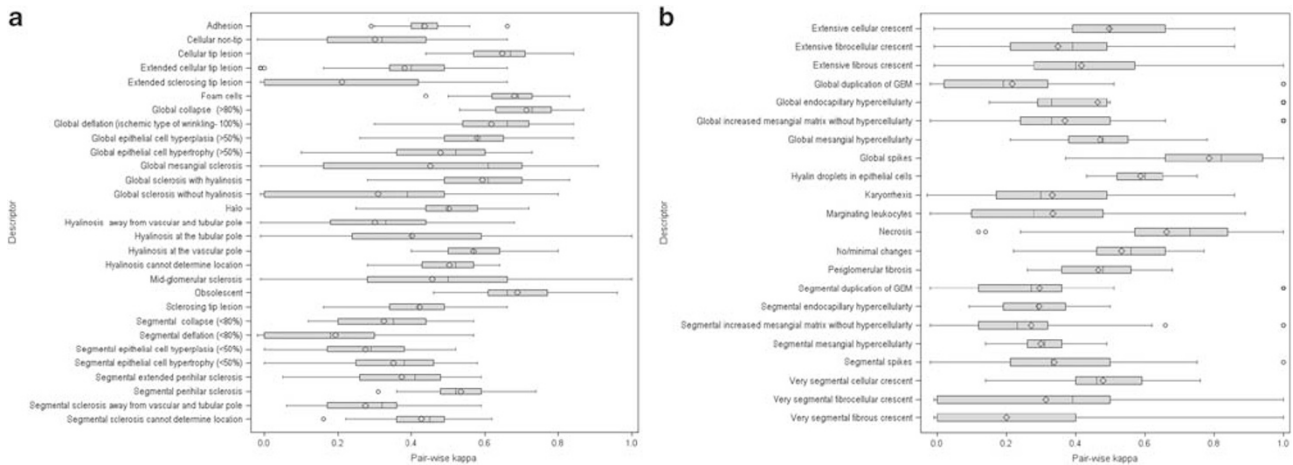


Figure 1 For each of 51 descriptors, an inter-reader kappa statistic was calculated for each of the 45 pairs of 10 pathologists. The boxplots show the distributions of these sets of 45 kappa statistics. (a) glomerular descriptors indicating sclerosis or associated with sclerosis.²⁹ (b) All other glomerular descriptors.²²

sclerosis and hyalinosis at the vascular pole), and the descriptor indicating no abnormalities. An additional three descriptors (cellular non-tip lesions, periglomerular fibrosis, and global podocyte hyperplasia) increased performance from fair/poor to moderate or good. (Table 1, Table 2, Figure 1a and b).

As expected, better concordance was achieved in most cases by clustering descriptors together. Compared with the cluster kappas, most component kappas are substantially smaller. However, for five of the clusters, a single component kappa was larger than the cluster kappa, showing that clustering often, but not always, leads to optimum concordance. Concordance improved when selected descriptors for sclerosing/obliterating lesions or for epithelial cell (podocytes) damage were combined (Table 3).

Concordance was independent of years of experience; analysis excluding the data generated by the trainees did not change significantly the overall concordance (data not shown). NEPTUNE and non-NEPTUNE pathologists had comparable overall inter-reader kappas (mean difference between kappas = 0.015, paired *t*-test *P* = 0.502).

Inter-reader Concordance for Tubulointerstitial Parameters

Excellent concordance was seen for both interstitial fibrosis and tubular atrophy, independent of years of experience (Figure 2; Figure 3d and e; Supplementary Table 4). In addition, overall concordance for interstitial fibrosis and tubular atrophy scoring remained consistently excellent when analyzed separately for each disease (minimal change disease, focal segmental glomerulosclerosis, and membranous nephropathy; data not shown).

Inter-reader Concordance for Podocyte Descriptors

Concordance was excellent for foot process effacement and good for microvillous transformation and condensation of the actin cytoskeleton, and moderate for loss of primary processes. (Figures 3f–i and 4; Supplementary Table 5).

Descriptor Reference Manual Revision

At the end of the study the descriptor reference manual was revised during several consensus webinar sessions that included NEPTUNE pathologists as well as pathologists outside the consortium, and language was added to improve clarity of definitions (Supplementary Table 6).

Discussion

To take advantage of and coordinate with new findings being discovered in molecular nephrology, renal pathologists must identify methodologies and approaches that allow for better integration of morphologic evaluation creating more compelling diagnostic paradigms.¹⁴ Furthermore, it is critical to design and implement classification systems for clinical research that are more meaningful with regard to novel renal biomarkers, prognosis, and treatment approaches.¹ The use of such morphologic observations requires concordance of pathologic analysis across diseases, level of training and experience. One goal of the NEPTUNE consortium is to identify reproducible morphologic variables that can be implemented in clinical practice by creating a new taxonomy of renal diseases. Toward that goal, we carried out a study testing intra- and inter-pathologist concordance using a set of 51

Table 2 Inter-reader concordance of all 51 glomerular descriptors by Test (I, II, and III) and NEPTUNE/non-NEPTUNE affiliation

Glomerular Descriptors	131 glomeruli					315 glomeruli			
	NEPTUNE			Non-NEPTUNE pathologists	NEPTUNE & Non-NEPTUNE pathologists	NEPTUNE pathologists	Non-NEPTUNE pathologists	NEPTUNE & Non-NEPTUNE pathologists	
	Test I (T=1) Kappa	Test II (T=2) Kappa	Test III (T=3) Kappa	Test II (T=1) Kappa	# glomeruli with descriptor seen (min-max over pathologists)	Test II (T=2) Kappa	Test II (T=1) Kappa	Test II (T=1 or 2) Kappa	# glomeruli with descriptor seen (min-max over pathologists)
No/minimal changes	0.54	0.49	0.81	0.24	1-2 ^l	0.58	0.42	0.55	3-12
Global sclerosis with hyalinosis ^a	0.58	0.52	0.60	0.71	0-8	0.55	0.72	0.60	3-11
Global sclerosis without hyalinosis ^a	0.18	0.14	0.22	0.66	1-8	0.31	0.62	0.36	1-22
Global deflation ^a	0.53	0.72	0.79	0.63	2-9	0.64	0.61	0.62	4-14
Global collapse ^{a,c}	0.83	0.85	0.85	0.79	6-12	0.69	0.82	0.72	17-34
Obsolescent ^a	0.61	0.68	0.70	0.80	2-12	0.63	0.71	0.63	0-15
Global mesangial sclerosis	0.00	0.04	0.06	0.01	0-2 ^l	0.11	0.12	0.10	0-16
Segmental perihilar sclerosis ^b	0.36	0.35	0.36	0.54	6-16	0.49	0.63	0.53	14-28
Segmental extended perihilar sclerosis ^b	0.25	0.36	0.28	0.56	3-9	0.33	0.48	0.38	8-18
Segmental sclerosis away from vascular & tubular pole ^b	0.43	0.41	0.40	0.13	2-16	0.30	0.08	0.28	5-33
Segmental sclerosis cannot determine location ^b	0.44	0.45	0.57	0.51	18-56	0.40	0.48	0.43	45-89
Cellular tip lesion ^d	0.67	0.61	0.69	0.81	8-14	0.57	0.82	0.64	12-27
Sclerosing tip lesion ^d	0.66	0.50	0.53	0.11	2-5	0.46	0.35	0.42	5-21
Extended cellular tip lesion ^d	NS	0.11	0.00	0.01	0-2 ^l	0.05	0.00	0.02	0-5
Extended sclerosing tip lesion ^d	0.00	0.08	0.00	0.32	0-4 ^l	0.07	0.28	0.08	0-4 ^l
Mid-glomerular sclerosis ^b	0.53	0.57	0.66	0.24	0-6	0.57	0.16	0.48	0-8
Cellular non-tip	0.38	0.46	0.47	0.22	0-10	0.31	0.39	0.33	4-23
Segmental collapse ^{b,c}	0.27	0.35	0.37	0.20	1-9	0.33	0.32	0.33	4-28
Segmental deflation ^b	0.07	0.05	0.03	0.02	0-6	0.07	0.02	0.61	0-10
Periglomerular fibrosis	0.37	0.53	0.52	0.43	0-34	0.52	0.37	0.44	10-50
Foam cells ^b	0.76	0.70	0.76	0.72	16-28	0.66	0.70	0.68	34-71
Hyaline droplets in epithelial cells (podocytes) ^f	0.59	0.53	0.51	0.61	7-21	0.57	0.58	0.58	25-70
Hyalinosis at the vascular pole ^{b,e}	0.46	0.57	0.63	0.55	4-18	0.60	0.64	0.58	14-26
Hyalinosis at the tubular pole ^{b,e}	0.00	0.01	0.00	0.01	0-3 ^l	0.27	0.02	0.20	0-14
Hyalinosis away from vascular and tubular pole ^{b,e}	0.26	0.34	0.27	0.26	0-12	0.31	0.26	0.33	1-21
Hyalinosis cannot determine location ^{b,e}	0.48	0.54	0.55	0.54	7-32	0.51	0.50	0.51	13-43
Synechiae (adhesion) ^b	0.42	0.43	0.45	0.43	30-82	0.42	0.45	0.44	67-152

Table 2 (Continued)

Segmental epithelial cell (podocyte) hypertrophy ^{f,g,i}	0.52	0.48	0.48	0.25	21-52	0.40	0.24	0.36	43-132
Global epithelial cell (podocyte) hypertrophy ^{f,h,j}	0.61	0.60	0.63	0.21	15-70	0.57	0.26	0.43	18-170
Segmental epithelial cell (podocyte) hyperplasia ^{f,i,l}	0.26	0.30	0.31	0.36	5-46	0.25	0.30	0.28	12-100
Global epithelial cell (podocyte) hyperplasia ^{f,k,j}	0.37	0.57	0.61	0.23	1-11	0.60	0.58	0.58	7-28
Halo (detachment of podocytes) ^g	0.45	0.58	0.53	0.60	7-20	0.50	0.57	0.49	10-31
Segmental mesangial hypercellularity	0.37	0.41	0.33	0.35	4-18	0.30	0.31	0.30	17-63
Global mesangial hypercellularity ^f	0.48	0.48	0.44	0.63	2-8	0.43	0.57	0.48	8-33
Segmental spikes	NS	0.00	0.11	0.00	0-1 ^l	0.14	0.12	0.16	0-9
Global spikes	NS	0.87	1.00	0.85	2-3 ^l	0.85	0.62	0.76	5-21
Intracapillary inflammatory cells	NS	0.00	0.00	0.13	0-5 ^l	0.24	0.23	0.23	0-41
Segmental endocapillary hypercellularity	NS	0.00	0.00	0.02	0-15	0.29	0.21	0.28	5-61
Global endocapillary hypercellularity	NS	NS	NS	0.01	0-2 ^l	0.37	0.27	0.29	0-8
Segmental duplication of GBM	0.00	0.14	0.00	0.16	0-23	0.29	0.06	0.15	0-59
Global duplication of GBM	NS	0.00	NS	0.01	0-3 ^l	0.00	0.06	0.03	0-12
Segmental increased mesangial matrix	0.00	0.08	0.05	0.01	0-3 ^l	0.20	0.16	0.21	0-12
Global increased mesangial matrix	0.00	0.00	0.00	0.01	0-5	0.38	0.01	0.27	0-5
Karyorrhexis	NS	NS	NS	NS	0-0 ^l	0.30	0.33	0.33	1-16
Necrosis	NS	0.00	NS	0.00	0-1 ^l	0.64	0.80	0.69	2-14
Very segmental cellular crescent	0.00	NS	0.00	NS	0-1 ^l	0.47	0.53	0.47	4-11
Very segmental fibrocellular crescent	NS	0.00	1.00	0.50	0-1 ^l	0.16	0.37	0.12	0-4 ^l
Very segmental fibrous crescent	0.17	0.16	0.11	NS	0-1 ^l	0.12	0.00	0.06	0-3 ^l
Extensive cellular crescent	NS	NS	NS	NS	0-0 ^l	0.50	0.46	0.50	1-6
Extensive fibrocellular crescent	NS	NS	NS	0.01	0-2 ^l	0.34	0.55	0.38	1-11
Extensive Fibrous crescent	NS	NS	NS	0.01	0-3 ^l	0.43	0.16	0.36	0-5

The number of training webinars (*T*) that pathologists experienced prior to the test is given for each column. Shading denotes categories of kappa values. Lightest grey (0.41–0.60; moderate), medium grey (0.61–0.80; good) and darkest grey (>0.80; excellent). NS = not seen (descriptor selected as absent by all pathologists). Membership in clusters of descriptors (see Table 3):

^aGlobal obliteration.

^bCollapse segmental and global.

^cSegmental obliteration.

^dTip lesions cellular and sclerosing.

^eEpithelial cell abnormalities.

^fHyalinosis (all locations).

^gEpithelial cell hypertrophy segmental and global.

^hSegmental epithelial cell hypertrophy/hyperplasia.

ⁱGlobal epithelial cell hypertrophy/hyperplasia.

^jEpithelial cell hyperplasia segmental.

^kMesangial hypercellularity segmental and global and global.

^lMaximum number of times seen by a pathologists < 5; these descriptors were omitted from most analyses (applied separately to the groups of 131 and 315).

Table 3 Inter-reader agreement (Cohen's kappa) of NEPTUNE pathologists for clusters of glomerular descriptors assessed in Test I, II, and III (131 glomeruli)

Clusters of Glomerular Descriptors	Test I Kappa	Test II Kappa	Test III Kappa	Cluster Component Kappas (Test III)
Global obliteration ^a	0.79	0.76	0.83	0.60, 0.22, 0.79, 0.85 , 0.70
Segmental obliteration ^b	0.69	0.63	0.73	0.36, 0.28, 0.40, 0.57, 0.66, 0.37, 0.03, 0.76 , 0.63, 0.00, 0.27, 0.55, 0.45
Collapse, segmental and global ^c	0.78	0.75	0.78	0.85 , 0.37
Tip lesions, cellular and sclerosing ^d	0.76	0.70	0.77	0.69, 0.53, 0.00, 0.00
Hyalinosis (all locations) ^e	0.67	0.69	0.72	0.63, 0.00, 0.27, 0.55
Mesangial hypercellularity, segmental and global ^f	0.59	0.64	0.54	0.33, 0.44
Epithelial cell abnormalities ^g	0.70	0.58	0.66	0.51, 0.48, 0.63, 0.31, 0.61, 0.53
Epithelial cell hypertrophy, segmental and global ^h	0.66	0.64	0.60	0.48, 0.63
Epithelial cell hyperplasia, segmental and global ⁱ	0.44	0.47	0.42	0.31, 0.61
Segmental epithelial cell hypertrophy/hyperplasia ^j	0.48	0.48	0.52	0.48, 0.31
Global epithelial cell hypertrophy/hyperplasia ^k	0.60	0.64	0.67	0.63, 0.61

Each cluster is composed of descriptors sharing morphologic similarities, and was judged to be present if at least one descriptor of the cluster was identified as present. Kappa statistics for each component of each cluster are given in the last column; any component kappa larger than the Test III cluster kappa is shown in bold.

Clusters and corresponding descriptors:

^aGlobal obliteration = global sclerosis with hyalinosis, global sclerosis without hyalinosis, global deflation, global collapse, obsolescent;

^bSegmental obliteration = segmental perihilar sclerosis, segmental extended perihilar sclerosis, segmental sclerosis away from vascular and tubular pole, segmental sclerosis cannot determine location, mid-glomerular sclerosis, segmental collapse, segmental deflation, foam cells, hyalinosis at vascular pole, hyalinosis at tubular pole, hyalinosis away from vascular and tubular pole, hyalinosis cannot determine location, adhesion;

^cCollapse, segmental, and global = global collapse, segmental collapse;

^dTip lesions, cellular, and sclerosing = cellular tip lesion, sclerosing tip lesion, extended cellular tip lesion, extended sclerosing tip lesion;

^eHyalinosis (all locations) = hyalinosis at vascular pole, hyalinosis at tubular pole, hyalinosis away from vascular and tubular pole, hyalinosis cannot determine location;

^fMesangial hypercellularity, segmental, and global = segmental mesangial hypercellularity, global mesangial hypercellularity;

^gEpithelial cell abnormalities = hyaline droplets in epithelial cells (podocytes), Segmental epithelial cell (podocyte) hypertrophy, Global epithelial cell (podocyte) hypertrophy, Segmental epithelial cell (podocyte) hyperplasia, Global epithelial cell (podocyte) hyperplasia, Halo (detachment of podocytes);

^hEpithelial cell hypertrophy segmental and global = segmental epithelial cell (podocyte) hypertrophy, global epithelial cell (podocyte) hypertrophy;

ⁱEpithelial cell hyperplasia, segmental, and global = segmental epithelial cell (podocyte) hyperplasia, global epithelial cell (podocyte) hyperplasia;

^jSegmental epithelial cell hypertrophy/hyperplasia = segmental epithelial cell (podocyte) hypertrophy, segmental epithelial cell (podocyte) hyperplasia;

^kGlobal epithelial cell hypertrophy/hyperplasia = global epithelial cell (podocyte) hypertrophy, global epithelial cell (podocyte) hyperplasia.

^{a-k}The descriptors associated with a cluster can be found in Table 2 as the descriptors with the same superscript as the cluster above.

glomerular, two tubulointerstitial and four ultra-structural features.

The first critical step toward a robust morphologic evaluation was the establishment of well defined morphologic criteria documented in a reference manual. The NEPTUNE digital pathology scoring system reference manual is comprehensive of features included by other classification systems and we referred to previously published criteria for some of the descriptors;^{5,25} however, many of the descriptors listed in the NEPTUNE digital pathology scoring system, although used in clinical practice to some degree, were not thoroughly defined by consensus

and organized in a comprehensive reference manual prior this study.

An innovative contribution of this study is the development of a protocol exploiting digital pathology technology. The introduction of digital pathology into large-scale glomerular disease research has enabled simultaneous remote access of multiple users.^{1,11,13,17,26,27} The application of digital technology, and of software for annotation of glomeruli, offers the opportunity to systematically eliminate glomerular selection bias, providing the basis for potentially higher inter-observer concordance.¹² Although it is intuitive that there are minimal

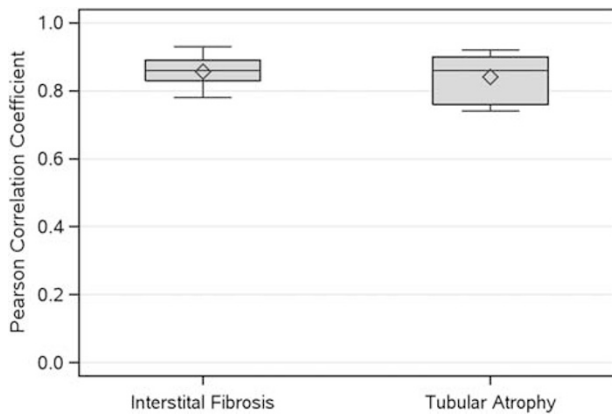


Figure 2 For each of the interstitial fibrosis and tubular atrophy descriptors, Pearson correlation coefficients were calculated for all 10 pairs of 5 pathologists on 244 cases. The boxplots show the distributions of these sets of correlation coefficients.

differences in concordance when scoring interstitial fibrosis and tubular atrophy by conventional light microscopy or whole-slide images, recognizing the value of specifically selecting structures to be evaluated, a recent concordance study was conducted using single digital images of glomeruli to identify the five patterns of focal segmental glomerulosclerosis (Columbia classification). This strategy, eliminating the glomerular selection bias, resulted in an overall good agreement among the six pathologists.¹⁶ In our study, we partially mimic the strategy utilized by Meehan *et al*¹⁶ by capturing digital images of individual annotated glomeruli from the whole-slide images of the 400 cases stored in the NEPTUNE digital pathology repository. By controlling the modality of the image review, the observations made, while under the control of the pathologist, were consistent with regard to image quality and to some extent magnification between reviewers. Using this approach, we were able to apply an 'object oriented' evaluation of performance, rather than a specimen-based approach.

Concordance of individual descriptors and factors contributing to concordance: most concordance studies are based on a one-time assessment. In our study, we demonstrated that concordance is modifiable by cross-training over time. This approach was tested in a study on thymic epithelial neoplasms, and resulted in post-webinar training improved concordance, confirming the value of digital pathology as an educational tool.²⁷ Although the inter-reader discrepancies in our study may appear significant, the total number of parameters involved for which pathologists needed cross-training, compared with a single diagnosis of epithelial neoplasia in the Wang's study, was much greater. Intra-reader concordance also improved with cross-training and webinar-based consensus as more detailed and objective criteria were provided to the participants, lessening individual reluctance in changing internal/subjective criteria. Thus, we still consider our observations encouraging for the systematic

application of webinar cross-training to increase intra- and inter-reader concordance.

The best performance was obtained by the interstitial fibrosis and tubular atrophy score, with overall excellent inter-reader concordance despite the lack of previous webinar training. Similar high concordance was obtained in the Oxford classification study.⁵ We hypothesize that this excellent performance is a consequence of the routine scoring of interstitial fibrosis and tubular atrophy in renal biopsy practice. Similarly, concordance was proportional to the frequency the ultrastructural podocyte descriptors are used in routine renal pathology assessment of biopsies; the highest concordance was recorded for the most commonly used parameter (foot process effacement) and the lowest for the descriptor used only experimentally (loss of primary process).²⁸ These data raise the question of whether descriptors for which familiarity and training are inadequate should be used and included in future studies. Developing robust training tools and metrics of performance is critical, as these infrequently assessed lesions may demonstrate correlation with clinical or molecular parameters and may add value to morphologic analysis or classifications. The continuous cross-training approach may ultimately prevent future classification systems from excluding morphologic criteria initially not performing well, but that may still have great potential as predictors of outcome. This concept may alter the current approach to generating classifications, which currently select for morphologic features based on concordance, to including initially less reproducible but valuable observational data by introducing post-training amendment and adjustment options. Should this occur, greater use of such features in routine clinical practice would then increase familiarity and automatically improve concordance.

The uneven level of concordance of some glomerular histologic descriptors is not easily explained. Although we eliminated the glomerular selection bias and provided a prefilled electronic scoring sheet listing all possible descriptors, lack of reproducibility for some descriptors may derive by failure to see or forgetting to mark a specific lesion among others affecting the same glomerulus, whereas for descriptors that are present in isolation, such as global spikes, it may have been easier to maintain the focus. Whereas global collapse or capillary wall spikes had expected high concordance, variable concordance was observed for subtypes of global or segmental sclerosis, although when consolidated under global or segmental obliteration, overall performance increased. The high concordance of segmental obliteration as an overall category confirms the data obtained by the Oxford classification study, where segmental sclerosis was defined as solidification/obliteration involving any part of the tuft and not broken down in subtypes based on location or cellularity.⁵ The lack of consistency in recognizing the type of segmental sclerosis may

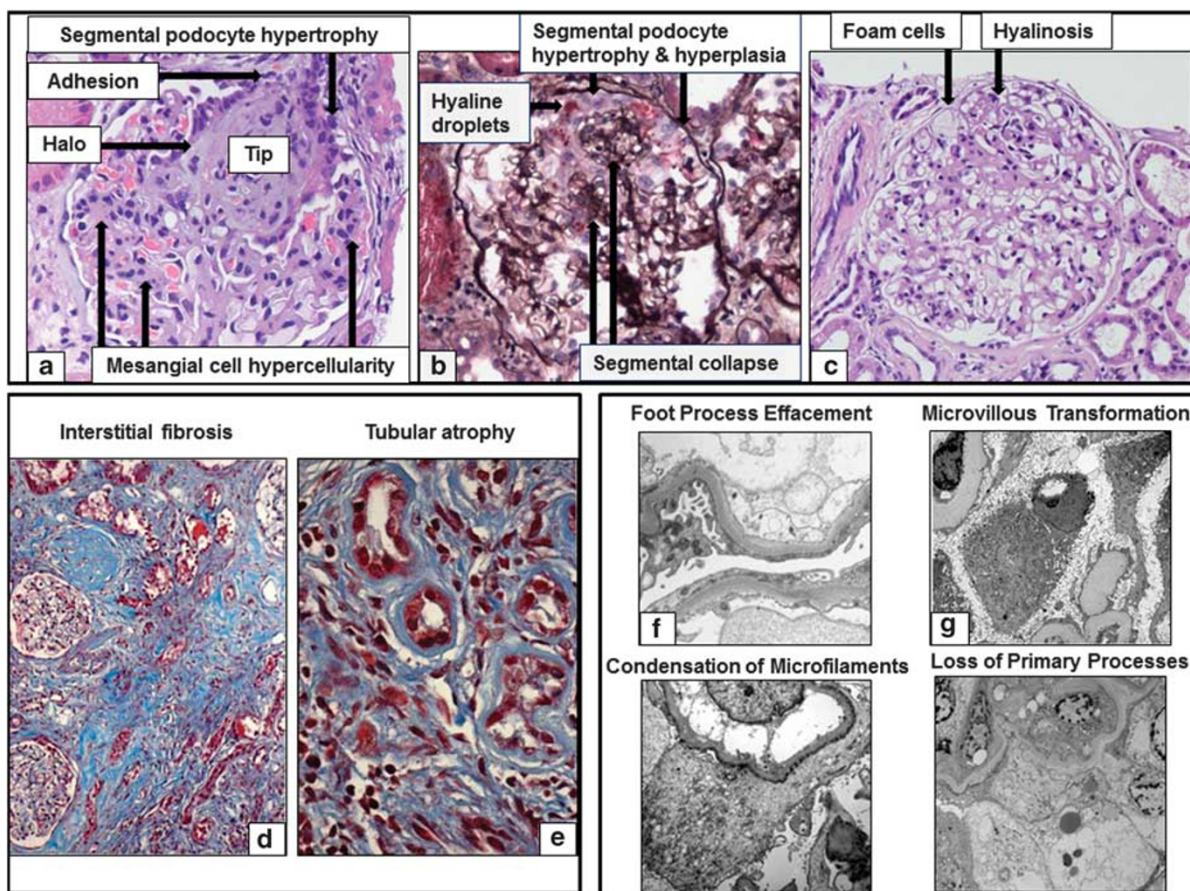


Figure 3 Glomerular and tubulointerstitial histologic descriptors: (a) A segmental obliteration of the glomerular tuft at the tip of the glomerulus is here represented. Additional descriptors applicable to this glomerular image are segmental epithelial cell hypertrophy, halo, adhesion, and global mesangial cell hypercellularity (Hematoxylin & Eosin). (b) The glomerulus here represented is morphologically profiled by segmental epithelial cell (podocyte) hypertrophy and hyperplasia, hyaline droplets and segmental collapse (Silver). (c) Foam cells, adhesion and hyalinosis are noted in the absence of increase matrix (sclerosis) (Hematoxylin & Eosin). (d) Trichrome stain showing increased deposition of collagen in between the tubules (interstitial fibrosis) and thickening on the tubular basement membranes in atrophic tubules (tubular atrophy). (e) Trichrome stain showing small tubules with thickened tubular basement membranes lined by small cuboidal epithelial cells. Ultrastructural Podocyte descriptors. (f) In this electron micrograph there is extensive foot process effacement with loss or widening of individual foot processes. (g) Aberrant formation of numerous slender cellular projections resembling microvilli or vesicle-like structures is seen along the apical surface of podocytes. (h) Condensation of electron-dense filaments against the sole of podocytes is present. (i) Loss of primary processes was recorded when epithelial (podocyte) cell bodies were in direct contact with glomerular basement membranes without interposition of primary processes.

appear to challenge the value of the conventional classification system of focal segmental glomerulosclerosis.²⁵ While low concordance is experienced when using individual descriptors defining the subtypes of segmental sclerosis, the application of the Columbia classification system at the glomerular level may have better concordance.¹⁶ The paradox that summary diagnostic approaches, rather than lesion-driven diagnostic paradigms, have better performance in concordance studies suggest that pathologists are using the totality of the histopathology to arrive at a diagnosis. This 'holistic' approach may be diagnostically powerful, but may limit prognostic utility, which is better elucidated by feature-based criteria. In addition, although all participants recognized epithelial cell (podocyte) injury, there were features that were inconsistently identified across reviewers, with the

greatest difficulty in differentiating segmental vs global lesions and hyperplasia vs hypertrophy. When segmental and global or hypertrophy and hyperplasia were combined, concordance increased. Good concordance was obtained by combining all podocyte abnormalities. It also appeared that the challenge in identifying segmental vs global lesions is not limited to podocytes but also applicable to mesangial cell proliferation. Again, by combining segmental and global mesangial cell proliferation, the kappa coefficient increased in the 315 glomeruli study to 0.64, confirming that the overall mesangial cell proliferation has adequate concordance to be included in classification systems.⁵ The poor concordance of these features suggests that they require additional refinement and evaluation before inclusion in classification systems where, for example, the recognition of segmental vs global damage/

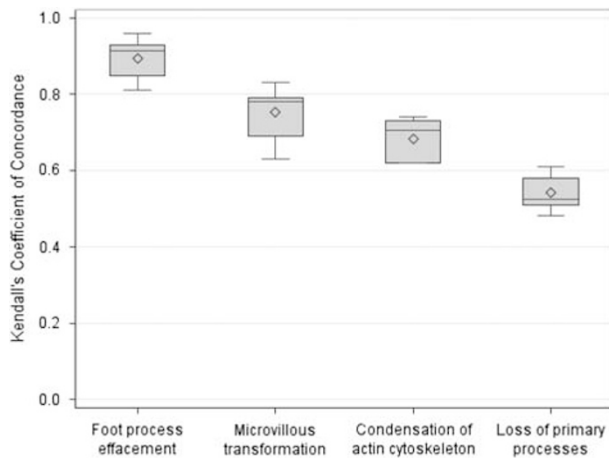


Figure 4 For each of the podocyte descriptors, pairwise Kendall's coefficient of concordance was calculated for all 10 pairs of five pathologists on 178 cases. The boxplots show the distributions of these sets of correlation coefficients.

proliferation may drive therapeutic choices.²⁹ Additional studies, currently in process, have been developed with the goals of (a) re-testing this approach provided more training, (b) testing reproducibility in the context of a European-based (EUREnOmics) and Chinese-based (NEPTUNE-China) study by a different set of reviewing pathologists applying the NEPTUNE digital pathology scoring system, (c) testing all NEPTUNE descriptors using different metrics (for example continuous vs dichotomous), and (d) applying other statistical methods.

When comparing data from NEPTUNE pathologists after several training sessions to non-NEPTUNE pathologists, the overall concordance was in favor of the NEPTUNE pathologists, although on 315 glomeruli the number of descriptors with a good or excellent concordance was greater for non-NEPTUNE pathologists. Several factors may have contributed to this result, including variability among pathologists.

This study also addressed whether concordance depended on years of experience in clinical practice. The overall coefficient of concordance did not change with the exclusion of pathologists in training. Trainees are accustomed to individual feature recognition as part of the learning process compared to experienced pathologists who are used to pattern recognition summarizing individual features into a diagnosis line.

After post-study revision of the reference manual to add clarity to the descriptor definitions (Supplementary Table 6), the NEPTUNE digital pathology scoring system and protocol were shared and implemented by other multicenter consortia with the generation of an INTEGRATE (INTErnational diGital nephRopAthology nEtwork) between pathologists from North America (NEPTUNE), Europe (EUREnOmics) and Asia (China-DiKip).

In conclusion, the NEPTUNE digital pathology scoring system provides comprehensive analysis of

renal structures with good-to-excellent concordance for many parameters. Although previous classification systems have eliminated poorly performing descriptors,⁵ here we provide an alternative model that maintains the original scoring metrics, but applies summary measures of clustered features and recommends continuing cross/training and consensus meetings. As metrics should ultimately be measured against their contribution to outcome and to guiding therapy, the rationale in favor of improving performance in contrast to dropping descriptors is that these descriptors have potential for important clinical value. Thus, this novel protocol for continuous improvement may serve as a model with potential to modify current classification systems, applicable across multiple international consortia, enabling world-wide collaboration and compilation of permanently recordable granular observational data suitable for correlation with clinical and molecular profiling of glomerular diseases.

Acknowledgments

The Nephrotic Syndrome Study Network Consortium (NEPTUNE) is a part of the National Center for Advancing Translational Sciences (NCATS), the Rare Disease Clinical Research Network (RDCRN), and is supported through a collaboration between the Office of Rare Diseases Research (ORDR), NCATS, and the National Institute of Diabetes, Digestive, and Kidney Diseases. RDCRN is an initiative of ORDR and NCATS. Additional funding and/or programmatic support for this project has also been provided by the University of Michigan, NephCure Kidney International, and the Halperin Foundation. This study was supported by a NEPTUNE pilot award. We thank Dr Charles Jennette for his contribution to creating the descriptor manual and participation in the training webinar sessions, and Dr William Smoyer for critical review of the manuscript.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

- Adam B, Randhawa P, Chan S *et al*. Banff initiative for quality assurance in transplantation (BIFQUIT): reproducibility of polyomavirus immunohistochemistry in kidney allografts. *Am J Transplant* 2014;14:2137–2147.
- Haas M, Sis B, Racusen LC *et al*. Banff 2013 meeting report: inclusion of c4d-negative antibody-mediated rejection and antibody-associated arterial lesions. *Am J Transplant* 2014;14:272–283.
- Roberts CA, Beitsch PD, Litz CE *et al*. Interpretive disparity among pathologists in breast sentinel lymph node evaluation. *Am J Surg* 2003;186:324–329.

- 4 Roberts JM, Jin F, Thurloe JK *et al*. High reproducibility of histological diagnosis of human papillomavirus-related intraepithelial lesions of the anal canal. *Pathology* 2015;47:308–313.
- 5 Working Group of the International Ig ANN, the Renal Pathology S, Roberts IS, Cook HT *et al*. The Oxford classification of IgA nephropathy: pathology definitions, correlations, and reproducibility. *Kidney Int* 2009;76:546–556.
- 6 Barisoni L, Schnaper HW, Kopp JB. A proposed taxonomy for the podocytopathies: a reassessment of the primary nephrotic diseases. *Clin J Am Soc Nephrol* 2007;2:529–542.
- 7 Barisoni L, Schnaper HW, Kopp JB. Advances in the biology and genetics of the podocytopathies: implications for diagnosis and therapy. *Arch Pathol Lab Med* 2009;133:201–216.
- 8 Polley MY, Leung SC, Gao D *et al*. An international study to increase concordance in Ki67 scoring. *Mod Pathol* 2015;28:778–786.
- 9 Polley MY, Leung SC, McShane LM *et al*. An international Ki67 reproducibility study. *J Natl Cancer Inst* 2013;105:1897–1906.
- 10 Jen KY, Olson JL, Brodsky S *et al*. Reliability of whole slide images as a diagnostic modality for renal allograft biopsies. *Hum Pathol* 2013;44:888–894.
- 11 Ozluk Y, Blanco PL, Mengel M *et al*. Superiority of virtual microscopy versus light microscopy in transplantation pathology. *Clin Transplant* 2012;26:336–344.
- 12 Barisoni L, Jennette JC, Colvin R *et al*. Novel quantitative method to evaluate globotriaosylceramide inclusions in renal peritubular capillaries by virtual microscopy in patients with fabry disease. *Arch Pathol Lab Med* 2012;136:816–824.
- 13 Barisoni L, Nast CC, Jennette JC *et al*. Digital pathology evaluation in the multicenter Nephrotic Syndrome Study Network (NEPTUNE). *Clin J Am Soc Nephrol* 2013;8:1449–1459.
- 14 Nast CC, Lemley KV, Hodgin JB *et al*. Morphology in the digital age: integrating high-resolution description of structural alterations with phenotypes and genotypes. *Semin Nephrol* 2015;35:266–278.
- 15 Gadegbeku CA, Gipson DS, Holzman LB *et al*. Design of the Nephrotic Syndrome Study Network (NEPTUNE) to evaluate primary glomerular nephropathy by a multidisciplinary approach. *Kidney Int* 2013;83:749–756.
- 16 Meehan SM, Chang A, Gibson IW *et al*. A study of interobserver reproducibility of morphologic lesions of focal segmental glomerulosclerosis. *Virchows Arch* 2013;462:229–237.
- 17 Ford SL, Polkinghorne KR, Longano A *et al*. Histopathologic and clinical predictors of kidney outcomes in ANCA-associated vasculitis. *Am J Kidney Dis* 2014;63:227–235.
- 18 Lemley KV. Diabetes and chronic kidney disease: lessons from the Pima Indians. *Pediatr Nephrol* 2008;23:1933–1940.
- 19 Miettinen J, Helin H, Pakarinen M *et al*. Histopathology and biomarkers in prediction of renal function in children after kidney transplantation. *Transpl Immunol* 2014;31:105–1.
- 20 Mise K, Hoshino J, Ubara Y *et al*. Renal prognosis a long time after renal biopsy on patients with diabetic nephropathy. *Nephrol Dial Transplant* 2014;29:109–118.
- 21 Mise K, Hoshino J, Ueno T *et al*. Clinical and pathological predictors of estimated GFR decline in patients with type 2 diabetes and overt proteinuric diabetic nephropathy. *Diabetes Metab Res Rev* 2015;31:572–581.
- 22 Mise K, Hoshino J, Ueno T *et al*. Impact of tubulointerstitial lesions on anaemia in patients with biopsy-proven diabetic nephropathy. *Diabet Med* 2015;32:546–5.
- 23 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
- 24 Thomas DB, Franceschini N, Hogan SL *et al*. Clinical and pathologic characteristics of focal segmental glomerulosclerosis pathologic variants. *Kidney Int* 2006;69:920–926.
- 25 D'Agati VD, Fogo AB, Bruijn JA *et al*. Pathologic classification of focal segmental glomerulosclerosis: a working proposal. *Am J Kidney Dis* 2004;43:368–382.
- 26 Gavrielides MA, Conway C, O'Flaherty N *et al*. Observer performance in the use of digital and optical microscopy for the interpretation of tissue-based biomarkers. *Anal Cell Pathol (Amst)* 2014.
- 27 Wang H, Sima CS, Beasley MB *et al*. Classification of thymic epithelial neoplasms is still a challenge to thoracic pathologists: a reproducibility study using digital microscopy. *Arch Pathol Lab Med* 2014;138:658–663.
- 28 Barisoni L, Kriz W, Mundel P *et al*. The dysregulated podocyte phenotype: a novel concept in the pathogenesis of collapsing idiopathic focal segmental glomerulosclerosis and HIV-associated nephropathy. *J Am Soc Nephrol* 1999;10:51–61.
- 29 Weening JJ, D'Agati VD, Schwartz MM *et al*. The classification of glomerulonephritis in systemic lupus erythematosus revisited. *J Am Soc Nephrol* 2004;15:241–250.

Supplementary Information accompanies the paper on Modern Pathology website (<http://www.nature.com/modpathol>)