

Cure Glomerulonephropathy Pathology Classification and Core Scoring Criteria, Reproducibility, and Clinicopathologic Correlations

Matthew B. Palmer^a Virginie Royal^b J. Charles Jennette^c Abigail R. Smith^d
Qian Liu^d Josephine M. Ambruzs^e Nicole K. Andeen^f Vivette D. D'Agati^g
Agnes B. Fogo^h Joseph Gautⁱ Rasheed A. Gbadegesin^j
Larry A. Greenbaum^k Jean Hou^l Margaret E. Helmuth^d
Richard A. Lafayette^m Helen Liapisⁿ Bruce Robinson^d Michael B. Stokes^g
Katherine Twombly^o Hong Yin^p Cynthia C. Nast^l
for the CureGN Consortium

^aDepartment of Pathology, University of Pennsylvania, Philadelphia, PA, USA; ^bDepartment of Pathology, Hospital Maisonneuve-Rosemont, University of Montreal, Montreal, QC, Canada; ^cDepartment of Pathology, University of North Carolina Kidney Center, Chapel Hill, NC, USA; ^dArbor Research Collaborative for Health, Ann Arbor, MI, USA; ^eArkana Laboratories, Little Rock, AR, USA; ^fDepartment of Pathology, Oregon Health and Science University, Portland, OR, USA; ^gDepartment of Pathology, Columbia University, New York, NY, USA; ^hDepartment of Pathology, Vanderbilt University, Nashville, TN, USA; ⁱDepartment of Pathology, Washington University, St. Louis, MO, USA; ^jDivision of Nephrology, Duke Children's Hospital Medical Center, Durham, NC, USA; ^kDivision of Nephrology, Department of Pediatrics, Emory University, Atlanta, GA, USA; ^lDepartment of Pathology, Cedars-Sinai Medical Center, Los Angeles, CA, USA; ^mDivision of Nephrology, Department of Medicine, Stanford University, Stanford, CA, USA; ⁿNephrology Center, Ludwig Maximilian University, Munich, Germany; ^oDivision of Nephrology, Department of Pediatrics, Medical University of South Carolina, Charleston, SC, USA; ^pPathology, Children's Healthcare of Atlanta, Atlanta, GA, USA

Keywords

Cure Glomerulonephropathy · Kidney biopsy · Reproducibility · Pathology scoring · Clinical-pathologic correlation

Abstract

Introduction: Cure Glomerulonephropathy (CureGN) is an observational cohort study of patients with minimal change disease (MCD), focal segmental glomerulosclerosis (FSGS),

membranous nephropathy (MN), or IgA nephropathy. We developed a conventional, consensus-based scoring system to document pathologic features for application across multiple pathologists and herein describe the protocol, reproducibility, and correlation with clinical parameters at biopsy. **Methods:** Definitions were established for

Matthew B. Palmer and Virginie Royal contributed equally to this work.

glomerular, tubular, interstitial, and vascular lesions evaluated semiquantitatively using digitized light microscopy slides and electron micrographs, and reported immunofluorescence. Cases with curated pathology materials as of April 2019 were scored by a randomly assigned pathologist, with at least 10% of cases scored by a second pathologist. Scoring reproducibility was assessed using Gwet's agreement coefficient (AC)1 statistic and correlations with clinical variables were performed. **Results:** Of 800 scored biopsies (134 MCD, 194 FSGS, 206 MN, 266 IgA), 94 were scored twice (11.8%). Of 60 pathology features, 46 (76.7%) demonstrated excellent ($AC1 > 0.8$), and 12 (20.0%) had good ($AC1 0.6-0.8$) reproducibility. Mesangial hypercellularity scored as absent, focal, or diffuse had moderate reproducibility ($AC1 = 0.58$), but good reproducibility ($AC1 = 0.71$) when scored as absent or focal versus diffuse. The percent glomeruli scored as no lesions had fair reproducibility ($AC1 = 0.34$). Strongest correlations between pathologic features and clinical characteristics at biopsy included interstitial inflammation, interstitial fibrosis, and tubular atrophy with estimated glomerular filtration rate, foot process effacement with urine protein/creatinine ratio, and active crescents with hematuria. **Conclusions:** Most scored pathology features showed excellent reproducibility, demonstrating consistency for these features across multiple pathologists. Correlations between certain pathologic features and expected clinical characteristics show the value of this approach for future studies on clinicopathologic correlations and biomarker discovery.

© 2023 The Author(s).
Published by S. Karger AG, Basel

Introduction

Cure Glomerulonephropathy (CureGN) is an international, multi-center, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)-funded consortium ultimately recruiting approximately 2,400 participants with a biopsy-documented diagnosis of minimal change disease (MCD), focal segmental glomerulosclerosis (FSGS), membranous nephropathy (MN), or immunoglobulin A nephropathy (IgAN), including IgA vasculitis (IgAV) [1]. A common protocol was developed to enroll patients based in part on pathology review, to prospectively collect clinical and laboratory data, to procure biosamples for biomarker discovery, and to generate pathology data based on standardized light (LM), immunofluorescence (IF), and electron (EM) microscopy parameters. A group of expert nephropathologists (CureGN pathology committee) developed kidney biopsy pathology inclusion and ex-

clusion criteria for disease cohort assignment as MCD, FSGS, MN, or IgAN using widely accepted pathologic criteria. Pathology criteria for C1q nephropathy were established to identify this subgroup of MCD and FSGS. Separately, the CureGN Core Scoring Working Group created a comprehensive system for evaluating lesions of glomeruli, tubules, interstitium, and vessels. Kidney biopsy materials from enrolled patients, including LM glass slides scanned into high-resolution whole slide images (WSIs) and EM digital images, were collected and uploaded into a digital pathology repository, establishing an infrastructure for digital pathology evaluation [2].

The standardized scoring system developed by the Core Scoring Working Group allows for conventional pathology scoring to qualitatively and semiquantitatively document patterns of kidney injury, providing scores based on explicitly defined terminology and gradations, and applied in a consistent way across multiple scoring pathologists. While this consensus-based approach was designed to reflect how pathologists currently practice in the clinical setting and provide information necessary for conventional disease classifications [3, 4], the scoring system also provides the basis for future CureGN studies requiring standardized pathology data.

Advantages of the CureGN core scoring protocol include a mechanism for evaluation of inter-scorer reproducibility, as a proportion of cases were reviewed and scored by two pathologists. Assuring reproducibility in kidney biopsy scoring is a well-recognized challenge in large collaborative studies [5-7]. Applying widely used conventional definitions and scoring parameters facilitates data reproducibility and reliability. Another advantage with this conventional approach is that study conclusions based on correlation of routinely used pathology features with clinical and biological data can be easily translated to clinical use. Moreover, these data will serve as a standard baseline for comparing novel scoring and classification approaches that align better with clinical, molecular, or biomarker data accrued by CureGN.

In the current study, we sought to validate the CureGN pathology system (1) for inclusion and exclusion classification criteria for cohort assignment, and (2) for pathology scoring to establish foundational pathology data for future CureGN studies. We also sought to identify pathologic features that are (1) reproducible across pathologists using a digital platform and (2) correlate with a limited set of clinical characteristics at the time of biopsy. Ongoing and future CureGN studies will determine how well these kidney biopsy pathologic

features correlate with clinical outcomes, biomarkers of disease activity and chronicity, and novel predictors of disease progression.

Materials and Methods

Study Design

CureGN is a multi-center prospective observational study of patients with glomerular disease. Children and adults with MCD, FSGS, MN, or IgAN (including IgAV determined clinically) were eligible if their first diagnostic kidney biopsy was within 5 years of study enrollment [1, 8]. Patients were excluded at the time of screening if they had end-stage kidney disease or a specified number of systemic diseases such as diabetes mellitus, systemic lupus erythematosus, HIV, or active hepatitis infection [1]. Demographics and clinical characteristics were collected at study enrollment and clinical data were collected at study visits aligned with clinical care. Race and ethnicity were self-reported or reported by parents of children as required by the funder. A specified set of retrospective laboratory data from the time of biopsy was collected. For each participant, the pathology report was reviewed by a study pathologist to determine eligibility for enrollment and cohort assignment. If a pathology report was not sufficient for enrollment, LM slides and EM images were reviewed. If further assessment was needed, evaluation by an additional pathologist was performed. The pathology criteria for study enrollment and cohort assignment have been published previously and are summarized in online supplementary Table S1 (for all online suppl. material, see <https://doi.org/10.1159/000534755>) [1].

Pathology Materials and Core Scoring

The majority (97%) of kidney biopsy slides from CureGN-enrolled patients were sent to NIH/NCI for central scanning at $\times 40$ magnification using an Aperio CS (Leica Microsystems, Vista, CA, USA) scanning system. The remaining slides (3%) were scanned locally at their originating institution. All available slides from each case (minimum of 2 slides including H&E [3440] and at least one stained with periodic acid-Schiff [2964], Masson's trichrome [1952], or Jones methenamine silver [1645]) were scanned into WSIs and uploaded to a SlidePath server (Leica Microsystems) for scoring pathologist review. Original pathology reports and EM images were also uploaded for scoring.

The CureGN Core Scoring Working Group, composed of 11 nephropathologists, selected kidney pathologic features for scoring (37 on LM, 14 on IF, PLA2R by IF or immunohistochemistry as available, and 8 on EM) based on those used in routine clinical practice for diagnosis of the four cohort diseases and assessed qualitatively or semiquantitatively. Definitions of the features, semiquantitative assessment parameters and scoring prevalence are provided in the Supplemental materials (online suppl. Table S2–S6). These parameters allowed determination of the Columbia FSGS variants, Oxford IgAN scores, and Ehrenreich and Churg MN stages. Definitions of the LM and EM pathologic features were based on approaches used in routine clinical practice and reviewed by the CureGN pathology committee. Webinar training of all scoring pathologists using example lesions was completed prior to initiation of scoring. For LM and EM scoring, all WSIs and EM images were evaluated. IF findings were abstracted from the pa-

thology reports using a method to transform different terminology into standardized scores amenable to data analysis (e.g., mild = 1+, moderate = 2+, severe = 3+). A series of pilot scoring was conducted, with all 11 pathologists scoring the same cases (6 cases in each of two pilot tests). Data from the pilot tests were used for preliminary reproducibility assessments, followed by further discussion and webinar training to improve reproducibility by refining the list of scoring parameters and their definitions. An option to request a second opinion by the scorer for specific questionable lesions (see examples in online suppl. Fig. 1) was implemented. Scores from these 12 pilot test cases were not included in the formal reproducibility analysis described in the next section, nor were there any cases with second opinion requests included. Eligible cases were randomly assigned to one of 11 pathologists for scoring.

An electronic CRF was utilized for pathology data collection. The CRF included the detailed definition for each feature to increase reproducibility. In each case, when present, a single example of specified pathologic features (including mesangial hypercellularity, endocapillary hypercellularity, fibrinoid necrosis, karyorrhexis, cellular, fibrocellular and fibrotic crescents, and FSGS variants) was identified by hidden annotation on the WSI and the slide level noted on the CRF to allow future review if needed. Scores initially were left blank and not set to 0, requiring all items to be completed by the scorer. After selecting all scoring features in the LM, IF, and EM sections, the CRF program determined and indicated completeness of each section, and the final cohort assignment was verified. After detailed pathology scoring, a small number of cases required removal from the study ($n = 4$) due to not meeting enrollment criteria, while a larger number ($n = 14$) required cohort reassignment (13 MCD to FSGS due to segmental sclerosing lesion identified during scoring including tip, perihilar, NOS, and cellular variants; 1 FSGS to IgA due to data entry error).

Reproducibility

To assess reproducibility, a minimum of 10% of all cases were scored a second time by a randomly assigned different pathologist. All possible pairs of pathologists are represented in this reproducibility analysis. Cases for second reads were distributed in two batches, both included in the reproducibility analysis. The first batch ($N = 12$, different from the 12 cases used in the pilot scoring activities described above) was specifically selected to contain a wide array of morphologic features. The second batch ($N = 88$) was randomly chosen but stratified by diagnosis to include a higher proportion of IgAN and FSGS cases and a lower proportion of MCD cases because the former were expected to have a greater number and variety of positive scorable lesions; however, all disease cohorts were included for generalizability of results. Cases assigned for second read were distributed together with single-read cases so that pathologists were blinded to whether they were completing a first or second read.

Clinical-Pathologic Correlation

Correlations between pathology parameters and demographic and clinical variables at biopsy, or as close in time as available, were assessed. These data include age at biopsy, sex, weight status (underweight [BMI <18.5 or BMI percentile <5 for children], normal weight [BMI 18.5–<25 or BMI percentile 5–<85 for children], overweight [BMI 25–<30 or BMI percentile 85–<95 for children], obese [BMI ≥ 30 or BMI percentile ≥ 95 for children]), and blood pressure status (normal, elevated blood pressure, stage 1

Table 1. Clinical and demographic characteristics by disease diagnosis (core scoring cohort)

Mean (SD) or median (Q1–Q3) or <i>N</i> (%) ^a	Overall (<i>n</i> = 800)	MCD (<i>n</i> = 134)	FSGS (<i>n</i> = 194)	MN (<i>n</i> = 206)	IgAN (<i>n</i> = 204)	IgAV (<i>n</i> = 62)
Age at biopsy, years	37.9 (21.4)	28.3 (23.4)	37.2 (22.4)	50.2 (17.0)	37.0 (17.1)	23.4 (17.9)
<18, <i>n</i> (%)	184 (23)	60 (45)	48 (25)	9 (4)	32 (16)	35 (56)
≥18, <i>n</i> (%)	616 (77)	74 (55)	146 (75)	197 (96)	172 (84)	27 (44)
Male, <i>n</i> (%)	452 (57)	69 (51)	93 (48)	121 (59)	132 (65)	37 (60)
Hispanic ^b , <i>n</i> (%)	118 (15)	20 (15)	37 (19)	23 (11)	29 (14)	9 (15)
Race ^c , <i>n</i> (%)		15 (12)	15 (8)			
Asian, <i>n</i> (%)	78 (10)	17 (14)	56 (31)	20 (10)	26 (13)	2 (3)
Black, <i>n</i> (%)	121 (16)	7 (6)	6 (3)	35 (1)	11 (6)	2 (3)
Other, <i>n</i> (%)	27 (4)	86 (69)	102 (57)	5 (3)	7 (4)	2 (3)
White	536 (70)	103.8 (78.7–123.7)	61.8 (35.6–95.4)	140 (70)	156 (78)	52 (90)
eGFR at biopsy ^d , mL/min per 1.73 m ²	77.1 (46.5–105.8)	6.0 (1.1–10.1)	4.0 (1.9–7.7)	85.2 (55.8–107.8)	60.3 (40.0–92.0)	82.1 (58.9–110.0)
UPCR at biopsy ^f , g/g	3.5 (1.2–7.1)	2.4 (1.7–3.3)	3.1 (2.1–3.8)	5.6 (3.0–8.7)	1.4 (0.7–2.7)	1.7 (0.7–5.3)
Serum albumin at biopsy ^f , g/dL	3.2 (2.2–3.8)	28.3 (23.4)	37.2 (22.4)	2.5 (2.0–3.2)	3.8 (3.4–4.2)	3.3 (2.9–3.8)

SD, standard deviation; IQR, interquartile range; MCD, minimal change disease; FSGS, focal segmental glomerulosclerosis; MN, membranous nephropathy; IgAN, IgA nephropathy; IgAV, IgA vasculitis; eGFR, estimated glomerular filtration rate; UPCR, urine protein-to-creatinine ratio (measured on 24-hour urine, first morning void, or spot/random urine in hierarchy as available). ^aPercent reported among nonmissing observations. ^b≤1% missing. ^c1–5% missing. ^d5–10% missing. ^e10–20% missing. ^f20–25% missing.

hypertension, or stage 2 hypertension) at study enrollment (as these were not collected at biopsy), and specified laboratory test results at biopsy (dipstick urine blood, urine protein/creatinine ratio (UPCR), serum albumin, and estimated glomerular filtration rate (eGFR). eGFR was estimated using serum Cr by the CKiD under 25 (U25) formula for participants younger than 25 years old, and CKD-EPI formula without race coefficient for participants 25 years and older [9, 10].

Statistical Analysis

Inter-pathologist reproducibility of pathology parameters was assessed using intraclass correlation coefficient for the only continuous parameter (the greatest number of glomeruli identified on any one of the scanned slides for a case), and using Gwet's agreement coefficient (AC) statistic for all other parameters [11]. For semiquantitative (ordinal) parameters, linear weights were used so that scores that are exactly the same receive a weight of 1; and weights decrease evenly until scores that are furthest apart receive a weight of 0. Gwet's AC, instead of Fleiss' kappa, was used because Fleiss' kappa calculation of chance agreement depends on the distribution of ratings from each rater, and performs differently for skewed and balanced samples with the same percent agreement. Gwet's AC is less sensitive to prevalence of features and was chosen due to the low prevalence of some features in this study. With a sample size of 94 cases with second reads, estimates of Gwet's AC were accurate within 0.05–0.2, with the majority of features able to be estimated within 0.15.

Correlations were assessed for continuous, ordinal, and binary pathology parameters. Categorical pathology parameters (*n* = 10) were also excluded, as correlations cannot be assessed with categorical parameters. For continuous pathology parameters, Pearson, polyserial, and point-biserial correlation coefficients were calculated to assess associations with continuous, ordinal, and binary variables, respectively. For an ordinal pathology parameter, polyserial, polychoric, and point-biserial correlation coefficients were calculated to

assess associations with continuous, ordinal, and binary variables, respectively [12–14]. For a binary pathology parameter, point-biserial, polychoric, and tetrachoric correlation coefficients were calculated to assess its association with continuous, ordinal, and binary variables, respectively [12].

Reproducibility analyses were conducted using R software, version 4.0.3 (R Development Core Team, Vienna). All other statistical analyses were conducted using SAS software, version 9.4 (SAS Institute, Cary, NC, USA).

Results

A total of 800 participants had biopsy features scored as of March 2021, including 134 MCD, 194 FSGS, 206 MN, and 266 IgAN (including 62 participants with IgAV) (Table 1). Adequate pathology materials were available for LM scoring in 789 (including 2 cases with 2 WSI each and 17 cases with 3 WSI each), IF scoring in 793, and EM scoring in 762. Mean (standard deviation) age at biopsy was 38 (21) years, with 23% children (<18 years) and 77% adults. Fifty-seven percent of the participants were male, 70% were white, and 85% were non-Hispanic. At biopsy, median (interquartile range [IQR]) eGFR was 77.1 (46.5–105.8) mL/min per 1.73 m²; median (IQR) UPCR was 3.5 (1.2–7.1) g/g; and median (IQR) serum albumin was 3.2 (2.2–3.8) g/dL. Characteristics by disease diagnosis are presented in Table 1. Compared to the overall CureGN cohort (*N* = 2,492, Table 2), those who had core scoring done were older (mean age 38 vs. 32 years), more often adults

Table 2. Clinical and demographic characteristics by disease diagnosis (entire CureGN cohort)

Mean (SD) or median (Q1–Q3) or N (%) ^a	Overall (n = 2,492)	MCD (n = 571)	FSGS (n = 644)	MN (n = 559)	IgAN (n = 545)	IgAV (n = 173)
Age at biopsy, years	31.9 (21.6)	21.4 (21.2)	32.0 (21.1)	47.4 (18.3)	31.1 (17.5)	18.2 (15.0)
<18, n (%)	922 (37)	350 (61)	226 (35)	54 (10)	174 (32)	118 (68)
≥18, n (%)	1,570 (63)	221 (39)	418 (65)	505 (90)	371 (68)	55 (32)
Male, n (%)	1,413 (57)	309 (54)	339 (53)	334 (60)	327 (60)	104 (60)
Hispanic ^b , n (%)	326 (13)	63 (11)	91 (14)	65 (12)	86 (16)	21 (12)
Race ^c , n (%)		50 (9)	33 (5)			
Asian, n (%)	209 (9)	99 (18)	177 (29)	46 (8)	70 (13)	10 (6)
Black, n (%)	16 (393)	27 (5)	32 (5)	89 (16)	24 (5)	4 (2)
Other, n (%)	97 (4)	374 (68)	377 (61)	14 (3)	19 (4)	5 (3)
White	1,710 (71%)	107.7 (81.8–126.8)	71.0 (40.6–102.7)	393 (73%)	417 (79%)	149 (89%)
eGFR at biopsy ^e , mL/min per 1.73 m ²	85.7 (55.1–112.4)	4.9 (0.8–9.5)	4.0 (2.0–8.0)	88.6 (64.3–109.8)	71.5 (44.4–101.6)	95.8 (69.4–113.5)
UPCR at biopsy ^f , g/g	3.3 (1.1–7.2)	2.4 (1.8–3.2)	3.0 (2.1–3.7)	5.5 (2.7–8.5)	1.4 (0.7–3.0)	1.8 (0.7–4.6)
Serum albumin at biopsy ^f , g/dL	3.0 (2.2–3.8)	2.14 (21.2)	32.0 (21.1)	2.6 (2.0–3.2)	3.8 (3.4–4.2)	3.4 (2.9–3.8)

SD, standard deviation; IQR, interquartile range; MCD, minimal change disease; FSGS, focal segmental glomerulosclerosis; MN, membranous nephropathy; IgAN, IgA nephropathy; IgAV, IgA vasculitis; eGFR, estimated glomerular filtration rate; UPCR, urine protein-to-creatinine ratio (measured on 24-hour urine, first morning void, or spot/random urine in hierarchy as available). ^aPercent reported among nonmissing observations. ^b≤1% missing. ^c1–5% missing. ^d5–10% missing. ^e10–20% missing. ^f20–30% missing.

(77% vs. 63%) due to pediatric pathology materials being obtained at a slower rate compared to adults, and had a lower median eGFR (77.1 vs. 85.7 mL/min per 1.73 m²). They were similar with regard to other characteristics, including race and ethnicity, UPCR, and serum albumin at biopsy.

There were 194 biopsies showing FSGS that underwent core scoring including 34 collapsing, 42 tip lesion, 7 cellular, 22 perihilar, and 86 not otherwise specified variants by the Columbia classification. Three biopsies could not be classified due to the diagnostic lesion(s) not being present in the LM slides scored; however, by the pathology report, FSGS was identified in the LM, IF, or EM material that was not available for core scoring evaluation. One case had a discrepancy between the report description and WSIs, which on review revealed the incorrect biopsy report had been included with the case. This error was subsequently corrected and, in concert with cohort changes described above due to pathology findings, highlights the value of core scoring as an additional QC step to ensure data quality. Clinical characteristics by FSGS variant are in Table 3.

Reproducibility

The number of WSIs available for evaluation was similar between the cases with second reads (mean 10.6, range 3–20) and all core-scored cases (mean 10.1, range 2–23). Among *N* = 100 cases (11.8% of 800 total) assigned

second reads, *N* = 94 were included in the reproducibility analysis (*N* = 6 excluded due to at least one read not completed due to digital pathology repository unavailability for technical reasons), none of which had second opinions performed. These 94 cases included 7 MCD, 37 FSGS, 18 MN, 26 IgAN, and 6 IgAV. Of the 60 pathology features, 46 (76.7%) demonstrated excellent reproducibility (Gwet's AC >0.8), and 12 (20.0%) had good reproducibility (Gwet's AC >0.6) (Fig. 1). Mesangial hypercellularity scored as absent, focal, or diffuse had only moderate reproducibility (AC = 0.58), although when scored as absent or focal versus diffuse (M0 vs. M1 in Oxford IgA scoring), it had good reproducibility (AC = 0.71). The percent glomeruli scored as having no lesions had only fair reproducibility (AC = 0.34), and thus was not included in the clinical-pathologic correlation analysis. Given the precision of Gwet's AC in this sample, 76% of parameters with at least good agreement (AC >0.6) would still be estimated to have good agreement if the true AC was at the lower bound of the 95% confidence interval.

Clinical-Pathologic Correlation

Clinical parameters were correlated with several pathologic features (Fig. 2). Age was positively correlated (*r* range 0.40–0.60) with global glomerular sclerosing/obliterative lesions, global glomerular capillary wall wrinkling, interstitial fibrosis and tubular atrophy, and

Table 3. Clinical and demographic characteristics by FSGS variant

Mean (SD) or median (Q1–Q3) or <i>N</i> (%) ^a	Collapsing (<i>n</i> = 34)	Tip (<i>n</i> = 42)	Cellular (<i>n</i> = 7)	Perihilar (<i>n</i> = 22)	All others (<i>n</i> = 86)	Diagnostic lesion not present in WSIs (<i>n</i> = 3)
Age at biopsy, years	35.7 (22.4)	39.5 (23.6)	46.9 (17.2)	35.0 (18.8)	36.5 (23.4)	32.7 (22.2)
<18, <i>n</i> (%)	8 (24)	12 (29)	0 (0)	3 (14)	24 (28)	1 (33)
≥18, <i>n</i> (%)	26 (76)	30 (71)	7 (100)	19 (86)	62 (72)	2 (67)
Male, <i>n</i> (%)	14 (41)	19 (45)	3 (43)	15 (68)	39 (45)	3 (100)
Hispanic ^c , <i>n</i> (%)	7 (21)	6 (14)	1 (14)	6 (27)	17 (20)	0 (0)
Race ^d , <i>n</i> (%)	2 (6)	3 (8)	1 (14)	2 (11)	6 (7)	1 (33)
Asian, <i>n</i> (%)	17 (55)	7 (18)	1 (14)	3 (17)	28 (35)	0 (0)
Black, <i>n</i> (%)	2 (6)	1 (3)	0 (0)	0 (0)	3 (4)	0 (0)
Other, <i>n</i> (%)	10 (32)	28 (72)	5 (71)	13 (72)	44 (54)	2 (67)
White	31.9 (14.5–64.9)	79.6 (39.5–102.1)	46.3 (37.7–99.1)	72.6 (43.1–93.0)	61.4 (36.7–100.0)	99.6 (55.9–132.7)
eGFR at biopsy ^d , mL/min per 1.73 m ²	5.4 (1.5–8.9)	5.2 (3.7–8.2)	5.1 (4.1–5.2)	3.1 (1.3–4.9)	3.3 (1.6–5.6)	6.7 (1.8–8.7)
UPCR at biopsy ^e , g/g	2.2 (1.6–3.1)	2.3 (1.7–3.0)	2.9 (2.2–3.9)	4.1 (3.2–4.3)	3.6 (2.8–4.0)	1.6 (1.4–4.4)
Serum albumin at biopsy ^e , g/dL	35.7 (22.4)	39.5 (23.6)	46.9 (17.2)	35.0 (18.8)	36.5 (23.4)	32.7 (22.2)

FSGS, focal segmental glomerulosclerosis; SD, standard deviation; IQR, interquartile range; eGFR, estimated glomerular filtration rate; UPCR, urine protein-to-creatinine ratio (measured on 24-hour urine, first morning void, or spot/random urine in hierarchy as available). ^aPercent reported among nonmissing observations. ^b≤1% missing. ^c1–5% missing. ^d5–10% missing. ^e10–20% missing.

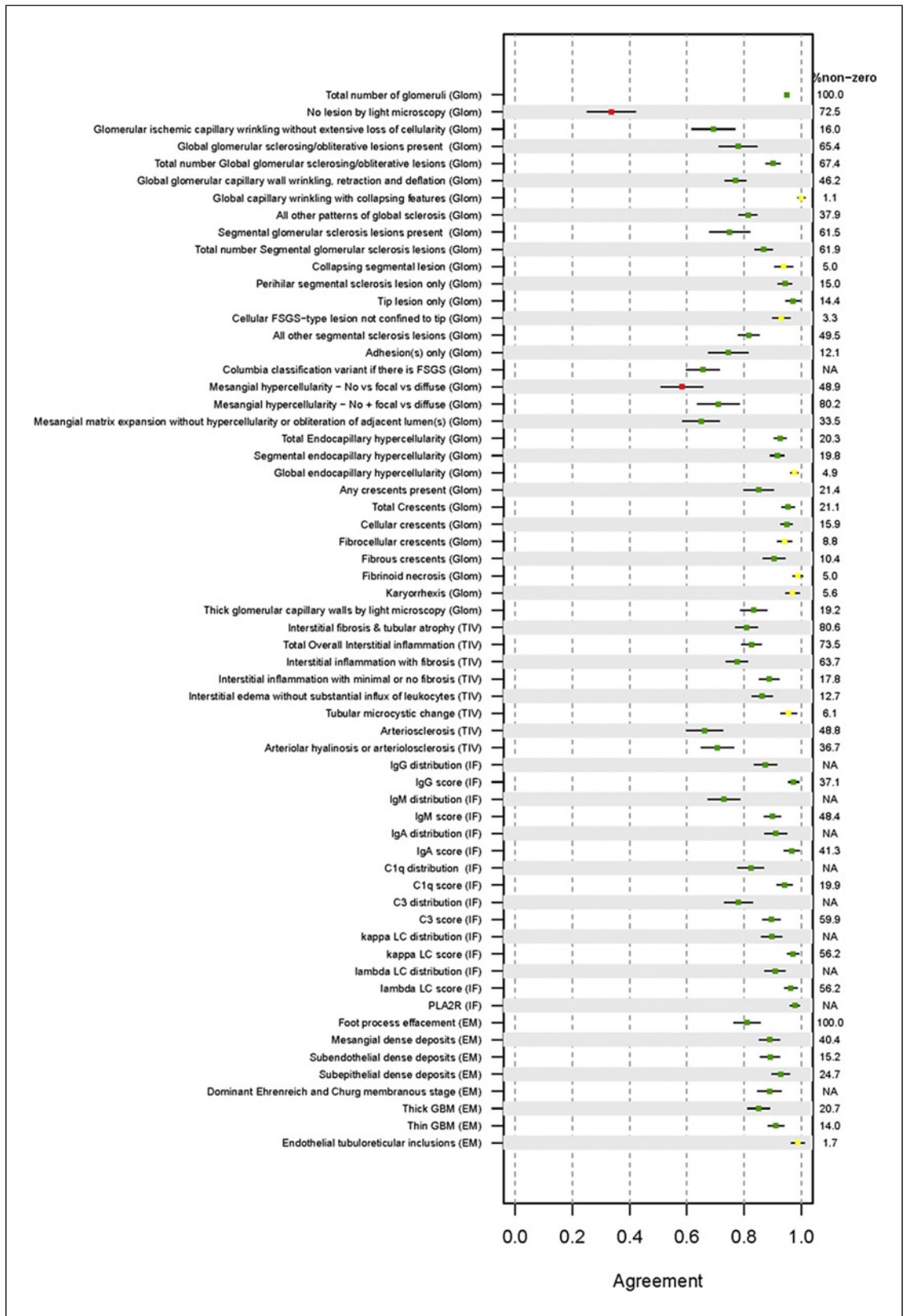
vascular sclerosis, and positively correlated (r range 0.31–0.39) with inflamed interstitial fibrosis, ischemic glomerular capillary wall wrinkling, podocyte foot process effacement (FPE), subepithelial electron dense deposits, and thick glomerular basement membranes. Age was negatively correlated with global endocapillary hypercellularity ($r = -0.45$). Hypertension was positively correlated (r range 0.26–0.34) with vascular sclerosis, global glomerular sclerosing/obliterative lesions, global glomerular capillary wall wrinkling, tubular atrophy and interstitial fibrosis, inflamed interstitial fibrosis, and thick glomerular basement membranes. Sex and weight status were not correlated with the pathologic features.

Hematuria was positively correlated with several glomerular features, including mesangial and endocapillary hypercellularity ($r = 0.51$ and 0.53 , respectively), active (cellular and fibrocellular) crescents ($r = 0.62$ and 0.55 , respectively), necrosis ($r = 0.46$), and karyorrhexis ($r = 0.39$). Hematuria was also positively correlated with IgA, C3, and lambda light chain deposition by IF (r range 0.41–0.58), and with mesangial and subendothelial electron dense deposits ($r = 0.52$ and 0.38 , respectively). UPCR was negatively correlated with fibrous crescents ($r = -0.36$) and positively correlated with FPE ($r = 0.39$). Serum albumin was negatively correlated with FPE

($r = -0.62$) and was positively correlated with IgA deposition and mesangial electron dense deposits ($r = 0.42$ and 0.41 , respectively). eGFR was negatively correlated with patterns of global and segmental glomerulosclerosis (r range -0.34 to -0.70), interstitial inflammation, interstitial fibrosis, and tubular atrophy (r range -0.58 to -0.62), tubular microcystic change ($r = -0.51$), and vascular sclerosis (r range -0.53 to -0.55).

Discussion

We have described and implemented a system to document pathologic features in the CureGN study. The intent is to establish baseline pathology features that can be useful as a foundation for further experimental approaches, including the search for biomarkers that correlate with the presence and severity of pathologic processes, and may elucidate pathogenic mechanisms and novel targets for therapy. The approach described addresses challenges inherent in CureGN and other large multi-institutional studies of kidney disease, including: (1) a large dataset of digital images, (2) a diverse set of contributing sites with differences in processes for tissue preparation, (3) different disease categories with both



(For legend see next page.)

unique and overlapping lesions, (4) a large number of scoring pathologists trained for congruent scoring, and (5) the need to design pathology methods that can be readily applied by the larger community of renal pathologists to use in practice. We show that this approach is practical, reproducible, and effective for clinicopathologic correlations in the CureGN study, and thus can serve as a model for pathologic classification and scoring in similarly designed studies. The definitions and procedures explicitly described here are available and may be useful as a representation of the conventional pathology approach to consortia studying glomerular diseases, including CureGN itself, who may benefit from having a conventional pathology benchmark against which more experimental approaches can be tested. CureGN Core Scoring serves in this way as a pathology baseline, as well as a database of reproducibly scored parameters, so that novel scoring systems can be applied in the setting of the CureGN cohort and compared against it as a way to assess their added value to prognosis or other clinical questions.

This initial analysis of cases was performed to assess our scoring system with respect to interobserver reproducibility and correlation with clinical variables. The cohort of scored cases is generally representative of the larger study by demographic and clinical characteristics, with the exception that the scored cohort is slightly older with a lower median eGFR. The difference in eGFR is likely due to an underrepresentation of pediatric cases in the materials available for earlier scoring.

The majority of pathologic features showed excellent or very good reproducibility, justifying our approach for accommodating a large number of scoring pathologists working with materials originating from multiple hospitals. We attribute the encouraging reproducibility results in part to holding a series of pilot scoring tests followed by training calls and video conferences among scorers, which enabled us to refine specific definitions. Another likely contributor to facilitating reproducibility was our reliance on broad consensus for these definitions and features, giving priority to practices a majority of scorers already employ in their clinical work. This approach also has the intended practical effect of rendering any findings based on these features more translatable to existing clinical practices.

Fig. 1. Inter-agreement statistics using ICC or Gwet's AC for 94 cases included in reproducibility analysis. The numbers on the right side of the plot are percent of nonzero values to show the prevalence of the ordinal pathology parameters (NA is shown for categorical parameters because percent of nonzero is not applicable). Standard error bars are shown for Gwet's AC statistics. Agreement statistics ≥ 0.6 are

One feature with lower reproducibility was mesangial hypercellularity. After follow-up discussions with the scoring group, we attribute this variation in part to scorers using different stains to score this feature as they do in their clinical pathology practice. In other studies, such as the Oxford classification of IgAN, a specific stain such as PAS was specified for the purpose of scoring mesangial cellularity to reduce variability [7]. We expect that clarifying the use of a specific stain, such as the PAS, to use for this feature will address this problem. A feature that was only fairly reproducible was the number of glomeruli without any lesions; this has been shown to be problematic as exemplified by the Oxford IgA study [7]. For our study, we believe this reflects insufficient clarification in the definition and initial training. The definition was "glomeruli with an absence of all of the LM lesions described below." However, in reviewing the data, it appeared some lesions and the stain used to evaluate them, such as global capillary wall thickening (e.g., in MN) and using silver versus PAS or H&E stains, were inconsistently applied to glomeruli without any lesions. Therefore, as with the above mesangial scoring, specificity of the definition, further training, and possibly restriction to specific stains are expected to clarify and alleviate the issue.

We performed correlations between pathologic and clinical variables in this scoring cohort and the results provide confidence that this approach will be useful for future clinicopathologic and biomarker correlations. For example, the finding that age correlates with global glomerulosclerosis and with interstitial fibrosis is in keeping with other reports [15–19]. The negative correlation between eGFR and interstitial fibrosis is also well established and is nicely recapitulated in this initial assessment of clinical and pathologic features in the CureGN study [20–22]. Similarly, podocyte FPE correlates with proteinuria, as expected from previous data [18, 23]. Since CureGN includes several different glomerular disease entities, the application of a single scoring system of conventional parameters might be questioned. The prognostic value of some conventional pathology lesions has been demonstrated across a wide range of kidney disease [24], including association of interstitial fibrosis and glomerulosclerosis with

plotted in green and those < 0.6 are plotted in red. Agreement statistics for parameters agreement statistics ≥ 0.6 and $< 10\%$ nonzero values are in yellow. The two features with agreement statistics < 0.6 both had $> 10\%$ nonzero values. Glom, glomerular; TIV, tubulointerstitial and vascular; IF, immunofluorescence; EM, electron microscopy; ICC, intraclass correlation coefficient.

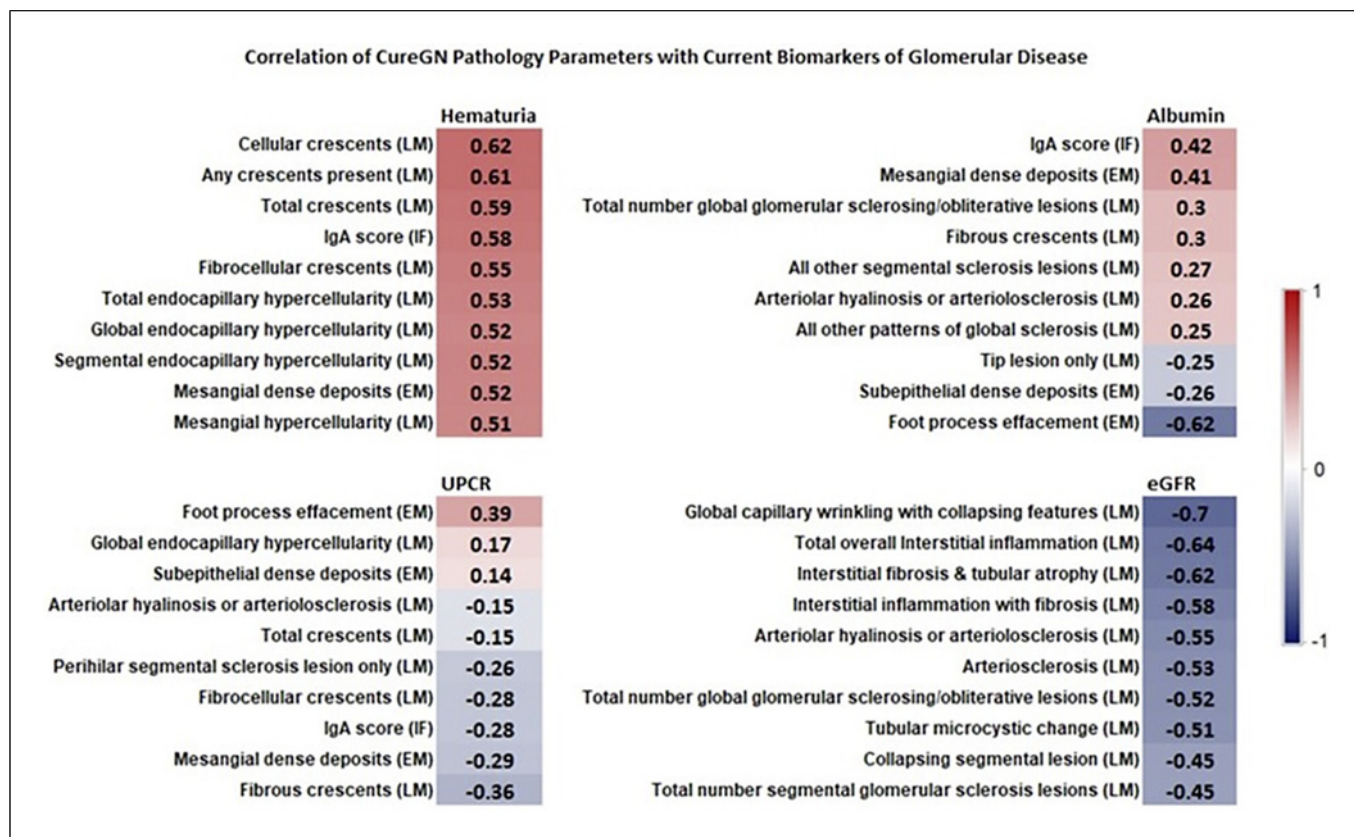


Fig. 2. Pathology parameters with strongest correlation with baseline or clinical characteristics. IF, immunofluorescence; EM, electron microscopy; IgA score, immunoglobulin A intensity by immunofluorescence; LM, light microscopy; UPCR, urine protein-to-creatinine ratio; eGFR, estimated glomerular filtration rate; albumin, serum albumin.

progression. Similarly, our clinicopathologic correlations as noted above hold despite the different underlying disease etiologies represented in CureGN, justifying the use of a single system of conventional lesions broadly applied. Many of the correlations result from differences in clinical features of different disease entities, such as the expected greater hematuria and relatively less proteinuria of IgAN compared to MN, leading to changes in correlations between mesangial proliferation and IgAN score (for example) to those clinical markers. Thus, there should be great comfort that the clinicopathologic correlations shown here are largely expected and moderately strong.

The aim of this study was to score the kidney biopsies for currently used conventional pathology features to generate reproducible pathology data in the CureGN study. These data can then be used to evaluate correlations with outcomes of these diseases, and look for disease biomarkers, clinical measures, or

biological findings that may account for disease activity and risk. Future studies can determine how to optimally use these scores in predicting treatment response and prognosis, and understand disease biology (such as through omics). These data will also provide a benchmark for determining the value of conventional and novel pathologic features in predicting precision treatments, and disease remission and progression in patients with the CureGN glomerular diseases.

Statement of Ethics

This study protocol was reviewed and approved by the Salus IRB which serves as the single IRB per NIH policy [#IRB00013544]. Written informed consent was obtained from all adult patients to participate in the study. For patients under 18 years of age, written informed consent was obtained from participant's parent/legal guardian/next of kin to participate in the study.

Conflict of Interest Statement

The authors declare that they have no relevant financial interests or conflicts of interest.

Funding Sources

Funding for the CureGN consortium is provided by U24DK100845 (formerly UM1DK100845), U01DK100846 (formerly UM1DK100846), U01DK100876 (formerly UM1DK100876), U01DK100866 (formerly UM1DK100866), and U01DK100867 (formerly UM1DK100867) from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). Patient recruitment is supported by NephCure Kidney International. Dates of funding for first phase of CureGN were September 16, 2013 to May 31, 2019. The funders of this study had no role in study design; collection, analysis, and interpretation of data; writing the report; or the decision to submit the report for publication.

References

- Mariani LH, Bomback AS, Canetta PA, Flessner MF, Helmuth M, Hladunewich MA, et al. CureGN study Rationale, design, and methods: establishing a large prospective observational study of glomerular disease. *Am J Kidney Dis.* 2019;73(2):218–29.
- Barisoni L, Nast CC, Jennette JC, Hodgin JB, Herzenberg AM, Lemley KV, et al. Digital pathology evaluation in the multicenter nephrotic syndrome study network (NEPTUNE). *Clin J Am Soc Nephrol.* 2013;8(8):1449–59.
- D'Agati VD, Fogo AB, Bruijn JA, Jennette JC. Pathologic classification of focal segmental glomerulosclerosis: a working proposal. *Am J Kidney Dis.* 2004;43(2):368–82.
- Trimarchi H, Barratt J, Catran DC, Cook HT, Coppo R, Haas M, et al. Oxford classification of IgA nephropathy 2016: an update from the IgA nephropathy classification working group. *Kidney Int.* 2017;91(5):1014–21.
- Barisoni L, Troost JP, Nast C, Bagnasco S, Avila-Casado C, Hodgin J, et al. Reproducibility of the NEPTUNE descriptor-based scoring system on whole-slide images and histologic and ultrastructural digital images. *Mod Pathol.* 2016;29(7):671–84.
- Bellur SS, Roberts ISD, Troyanov S, Royal V, Coppo R, Cook HT, et al. Reproducibility of the Oxford classification of immunoglobulin A nephropathy, impact of biopsy scoring on treatment allocation and clinical relevance of disagreements: evidence from the VALidation of IGA study cohort. *Nephrol Dial Transplant.* 2019;34(10):1681–90.
- Working Group of the International IgA Nephropathy Network an, d the Renal Pathology Society, Roberts ISD, Cook HT, Troyanov S, Alpers CE, Amore A, et al. The Oxford classification of IgA nephropathy: pathology definitions, correlations, and reproducibility. *Kidney Int.* 2009;76(5):546–56.
- Selewski DT, Ambruzs JM, Appel GB, Bomback AS, Matar RB, Cai Y, et al. Clinical characteristics and treatment patterns of children and adults with IgA nephropathy or IgA vasculitis: findings from the CureGN study. *Kidney Int Rep.* 2018;3(6):1373–84.
- Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N Engl J Med.* 2021;385(19):1737–49.
- Pierce CB, Muñoz A, Ng DK, Warady BA, Furth SL, Schwartz GJ. Age- and sex-dependent clinical equations to estimate glomerular filtration rates in children and young adults with chronic kidney disease. *Kidney Int.* 2021;99(4):948–56.
- Gwet K. Handbook of inter-rater reliability: the definitive guide to measuring the extent of agreement among raters. 4th ed. Gaithersburg, MD: Advanced Analytics; 2014.
- DF. Polychoric and polyserial correlations. In: Kotz LNJ, editor. Encyclopedia of statistical sciences. 7. New York: Wiley; 1988. p. 69–74.
- KHC. Biserial correlation. In: KS, editor. Encyclopedia of statistical sciences. 1. New York: Wiley; 1982. p. 276–9.
- Olsson UDF, Drasgow F, Dorans NJ. The polyserial correlation coefficient. *Psychometrika.* 1982;47(3):337–47.
- Denic A, Glasscock RJ, Rule AD. Structural and functional changes with the aging kidney. *Adv Chronic Kidney Dis.* 2016;23(1):19–28.
- Glasscock RJ, Rule AD. Aging and the kidneys: anatomy, physiology and consequences for defining chronic kidney disease. *Nephron.* 2016;134(1):25–9.
- Hommoss MS, Glasscock RJ, Rule AD. Structural and functional changes in human kidneys with healthy aging. *J Am Soc Nephrol.* 2017;28(10):2838–44.
- Palmer MB, Abedini A, Jackson C, Blady S, Chatterjee S, Sullivan KM, et al. The role of glomerular epithelial injury in kidney function decline in patients with diabetic kidney disease in the TRIDENT cohort. *Kidney Int Rep.* 2021;6(4):1066–80.
- Ricaurte Archila L, Denic A, Mullan AF, Narasimhan R, Bogojevic M, Thompson RH, et al. A higher foci density of interstitial fibrosis and tubular atrophy predicts progressive CKD after a radical nephrectomy for tumor. *J Am Soc Nephrol.* 2021;32(10):2623–33.
- Ellis RJ, Kalma B, Del Vecchio SJ, Aliano DN, Ng KL, Dimeski G, et al. Chronic kidney cortical damage is associated with baseline kidney function and albuminuria in patients managed with radical nephrectomy for kidney tumours. *Pathology.* 2019;51(1):32–8.
- Li P, Gupta S, Mothi SS, Rennke HG, Leaf DE, Waikar SS, et al. Histopathologic correlates of kidney function: insights from nephrectomy specimens. *Am J Kidney Dis.* 2021;77(3):336–45.
- Mariani LH, Martini S, Barisoni L, Canetta PA, Troost JP, Hodgin JB, et al. Interstitial fibrosis scored on whole-slide digital imaging of kidney biopsies is a predictor of outcome in proteinuric glomerulopathies. *Nephrol Dial Transplant.* 2018;33(2):310–8.
- Royal V, Zee J, Liu Q, Avila-Casado C, Smith AR, Liu G, et al. Ultrastructural characterization of proteinuric patients predicts clinical outcomes. *J Am Soc Nephrol.* 2020;31(4):841–54.
- Srivastava A, Palsson R, Kaze A, Chen M, Palacios P, Sabbiseti V, et al. The prognostic value of histopathologic lesions in native kidney biopsy specimens: results from the Boston Kidney Biopsy Cohort study. *J Am Soc Nephrol.* 2018;29(8):2213–24.

Author Contributions

Research idea and study design: M.B.P., V.R., J.C.J., A.R.S., R.L., C.C.N., N.A., H.Y., L.A.G., B.R., K.T., and R.A.G.; data acquisition: M.B.P., V.R., J.C.J., J.M.A., V.D.D., A.B.F., J.G., J.H., H.L., M.B.S., C.C.N., N.A., and H.Y.; and statistical analysis: A.R.S., Q.L., and M.H. Each author contributed important intellectual content during manuscript drafting or revision and accepts accountability for the overall work by ensuring that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved; all authors approved the final version of the manuscript.

Data Availability Statement

The data that support the findings of this study are available in the supplemental Tables, and any additional data may be obtained from the CureGN consortium following data use procedures available in the CureGN website (<https://www.dev.curegn.org>).