

Novel Quantitative Method to Evaluate Globotriaosylceramide Inclusions in Renal Peritubular Capillaries by Virtual Microscopy in Patients With Fabry Disease

Laura Barisoni, MD*; J. Charles Jennette*, MD; Robert Colvin*, MD; Sheela Sitaraman, PhD; Alexander Bragat, BSc; Jeff Castelli, PhD; Dan Walker, BA; Pol Boudes, MD

• **Context.**—Assessing the amount of globotriaosylceramide inclusions in renal peritubular capillaries by a semiquantitative approach is a standard and useful measure of therapeutic efficacy in Fabry disease, achievable by light microscopy analysis.

Objective.—To describe a novel virtual microscopy quantitative method to measure globotriaosylceramide inclusions (Barisoni Lipid Inclusion Scoring System [BLISS]) in renal biopsies from patients with Fabry disease.

Design.—Plastic embedded 1- μ m-thick sections from kidney biopsies from 17 patients enrolled in a Fabry disease clinical trial were evaluated using a standard semiquantitative methodology and BLISS to compare sensitivity. We also tested intrareader and interreader variability of BLISS and compared results from conventional light microscopy analysis with a virtual microscopy-based methodology. Peritubular capillaries were first annotated on digital images of whole slides by 1 pathologist and then scored

for globotriaosylceramide inclusions by 2 additional pathologists.

Results.—We demonstrated that (1) quantitative analysis by BLISS results in detection of small amount of globotriaosylceramide inclusions even when by semiquantitative analysis the score is 0, (2) application of BLISS combined with conventional light microscopy results in low intrareader and interreader variability, and (3) BLISS combined with virtual microscopy results in significant reduction of intrareader and interreader variability compared with BLISS–light microscopy.

Conclusions.—BLISS is a simpler and more sensitive scoring system compared to the semiquantitative approach. The virtual microscopy–based methodology increases accuracy and reproducibility; moreover, it provides a permanent record of retrievable data with full transparency in clinical trials.

(*Arch Pathol Lab Med.* 2012;136:816–824; doi: 10.5858/arpa.2011-0350-OA)

Fabry disease (FD) is an X-linked lysosomal storage disorder resulting from mutations in *GLA*, the gene that encodes the lysosomal hydrolase, α -galactosidase A. In FD, reduced or absent activity of α -galactosidase A results in accumulation of globotriaosylceramide (GL-3) and related glycosphingolipids in various tissues. Residual enzyme activity can still occur if the genetic defect causes an unstable, misfolded α -galactosidase. Although

FD is classically more severe in male patients, females can also develop severe disease. Life expectancy for FD patients is reduced in both females and males.¹

Enzyme replacement therapy (ERT) with agalsidase alfa and agalsidase beta is available for the treatment of FD. The US Food and Drug Administration approved agalsidase beta in 2001 following a clinical study in which a reduced amount of GL-3 inclusions in endothelial cells of renal peritubular capillaries (PTCs) was demonstrated after a period of intermittent intravenous treatment with this recombinant enzyme.^{2,3} Although the efficacy of ERT in FD has been measured using numerous criteria,^{3,4} levels of GL-3 cytoplasmic inclusions in the endothelium of PTCs are still considered of prime importance by the FDA.

Previous trials of ERT used a semiquantitative approach to measure GL-3 inclusions in PTCs by light microscopy (LM),² but limitations of this methodology have been identified. Although clinical symptoms are qualitatively comparable,^{1,5} recent studies showed a smaller amount of GL-3 inclusions in female compared to male patients.⁶ For females, or for males with residual enzyme activity, a more sensitive methodology is required to quantify small amounts of GL-3 in PTCs and to better assess disease status and response to treatment. Moreover, it has been recently suggested that early detection of renal disease

Accepted for publication December 7, 2011.

* Authors have equally contributed to create BLISS-VM methodology.

From the Department of Pathology and Medicine, New York University Langone Medical Center, New York (Dr Barisoni); the Department of Pathology, University of North Carolina Medical Center, Chapel Hill (Dr Jennette); the Department of Pathology, Massachusetts General Hospital, Boston (Dr Colvin); Amicus Therapeutics, Inc, Cranbury, New Jersey (Ms Sitaraman, Messrs Bragat and Castelli, and Dr Boudes); and Aperio, Vista, California (Mr Walker).

Dr Colvin is a paid consultant for Amicus Therapeutics, Inc. Ms Sitaraman, Mr Bragat, Mr Castelli, and Dr Boudes are employees of Amicus Therapeutics, Inc, and have stock options in the company. Mr Walker is an employee of Aperio. The other authors have no relevant financial interest in the products or companies described in this article.

Reprints: Laura Barisoni, MD, Department of Pathology and Medicine, New York University, 560 First Ave, New York, NY 10016 (e-mail: barisl01@med.nyu.edu or Laura.Barisoni@nyumc.org).

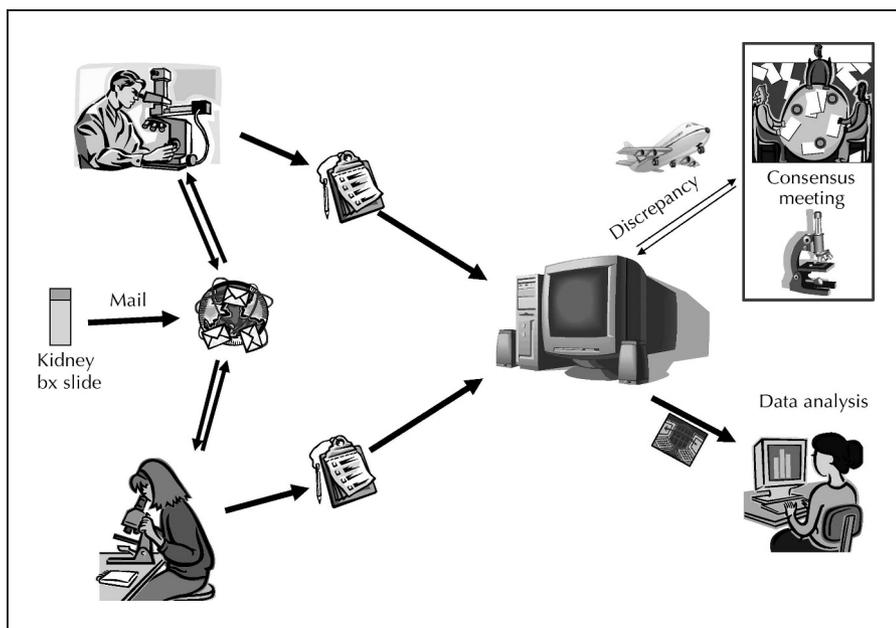


Figure 1. Conventional light microscopy protocol. Glass slides are mailed from pathologist to pathologist for evaluation. Each pathologist scores glass slides by light microscopy and records scores on a spreadsheet, which is mailed to the data center. Analysis for discrepancies is then performed and pathologists resolve disagreement and reach consensus face to face in a consensus meeting. Abbreviation: bx, biopsy.

followed by treatment improves outcome, suggesting that new methodologies with higher sensitivity may be clinically helpful.⁷ To improve the ability to detect GL-3 PTC inclusions in females or males with residual enzyme activity, a novel quantitative methodology to measure GL-3 inclusions in PTCs was designed and was named the Barisoni Lipid Inclusions Scoring System (BLISS). The BLISS system is more sensitive compared to previous semiquantitative scoring systems. Moreover, we introduced the use of virtual microscopy (VM) in the analysis of renal biopsies to increase reproducibility.

MATERIALS AND METHODS

Renal Biopsies

Seventeen patients (8 males and 9 females) with symptomatic FD were enrolled in 3 phase 2 studies of AT1001 (miglstat hydrochloride; Amicus Therapeutics, Cranbury, New Jersey), a pharmacologic chaperone in clinical development for the treatment of FD. Pretreatment biopsies were analyzed.

Morphologic Analysis

Renal tissue was processed according to standard protocols. A fragment of renal parenchyma containing cortex was fixed in glutaraldehyde and processed for ultrastructural analysis. Sections (1 μm thick) were stained with methylene blue–azure II according to previous protocols in FD.² All slides were given a unique number and all patient identifiers were removed. Analysis of glass slides stained with methylene blue–azure II was performed by conventional LM as well as VM to measure the amount of GL-3 inclusions in PTCs. Staining intensity of GL-3 granules in podocytes was used as reference for PTC inclusions; only PTC inclusions stained at a similar intensity to podocyte GL-3 inclusions were counted. To assure uniformity of assessments, capillaries that were smaller than an adjacent tubular cell nucleus or with any one internal diameter more than 4 times longer than any other internal diameter were excluded. This was an arbitrary cutoff with the rationale that for capillaries cut longitudinally, the total amount of endothelium surface and cytoplasm is much greater compared to those PTCs with a cross section. This is particularly important in clinical trials, for example, where comparison of average

number of GL-3 inclusions per capillary between pretreatment and posttreatment needs to be assessed.

Protocols

Slides were scored using 2 different protocols:

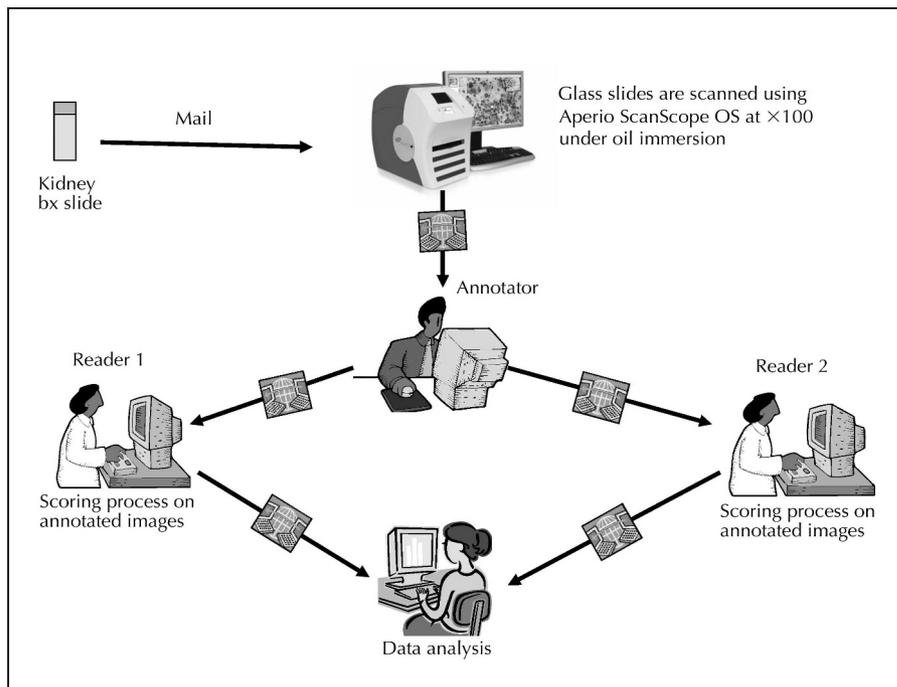
Conventional LM Protocol.—Two pathologist-readers were involved (L.B. and J.C.J.). Glass slides were mailed to pathologist 1, who read slides at $\times 100$ under oil immersion and recorded GL-3 scores in a spreadsheet. Glass slides were then mailed to pathologist 2, who recorded GL-3 scores in a separate spreadsheet. Both pathologists scored 50 to 100 PTCs in the same glass slides, but not necessarily the same PTCs. The GL-3 scores were entered into an electronic spreadsheet and e-mailed for data analysis (Figure 1).

VM protocol.—A pathologist-annotator and 2 pathologist-readers were involved (L.B., J.C.J., and R.C.). Each pathologist served as annotator in one-third of the cases and as reader in the remaining two-thirds of the cases. After glass slide preparation was completed, slides were scanned with the Scanscope OS (Aperio, Vista, California) at $\times 100$ under oil immersion into whole-slide digital images. Virtual images were then posted on a secure server where they could be accessed simultaneously from different locations. The annotator was provided with a unique password to access the virtual images on the server. The annotator identified 50 to 100 PTCs and marked each with an arrow. The system automatically assigned a unique number (1–100) to each marked PTC. This number was captured in an annotation sheet linked to the virtual image. Once annotation was completed, 2 identical copies were created from the master annotated image. Then the other 2 pathologists, reader 1 and reader 2, received a unique password that allowed them to access only 1 of the 2 copies of the identical annotated image. The scoring process occurred by each pathologist-reader recalling each annotated PTC and assigning a score for that PTC on the annotation sheet. As scores were assigned to the annotated PTC, the score/number appeared on the screen, near the arrow. Both pathologist-readers scored exactly the same PTCs. Data assigned to each specific PTC were available on the annotation sheet linked to the image, eliminating the need of the transcription step (Figures 2 and 3).

Scoring Systems

Thurberg System.—According to this semiquantitative scoring system,² each capillary received a score based on the below-described criteria:

Figure 2. Novel virtual microscopy protocol. Glass slides are mailed to an imaging center where they are scanned (Aperio ScanScope OS; Aperio, Vista, California) into virtual images. The annotator pathologist annotates peritubular capillaries on digitalized images. Annotated images are duplicated and distributed to reader-pathologists for scoring. Abbreviation: bx, biopsy.



- 0 = ≤ 2 granules
- 1 = ≥ 3 granules
- 2 = ≥ 1 nonbulging aggregate
- 3 = ≥ 1 bulging aggregate

The final score for each biopsy was based on the majority of capillaries with a given score. For example, if 49% of capillaries have a score of 2, 30% a score of 3, 30% a score of 1, and 11% a score of 0, the biopsy score is 2. Examples of bulging and nonbulging aggregates are shown in Figure 4, A and B.

BLISS System.—The number of GL-3 inclusions per PTC was recorded. The final score for each biopsy was reported as the average number of inclusions per PTC.

Methodologies for PTC GL-3 Inclusion Quantification

The combination of the 2 scoring systems (Thurberg and BLISS) with the 2 glass reading protocols resulted in 3 different methodologies used for analysis.

Thurberg-LM Scoring.—Two pathologists each separately scored the 17 biopsies by conventional LM (Figure 1), using the published semiquantitative Thurberg scoring system.

BLISS-LM Scoring.—The same 2 pathologists each separately scored the 17 biopsies by conventional LM (Figure 1) using the quantitative BLISS approach.

BLISS-VM Scoring.—The VM protocol was used in combination with BLISS. After the annotation process was completed by the annotator pathologist, the 2 pathologist-readers independently scored for GL-3 inclusions on identical duplicates of the annotated virtual images, using the quantitative approach, BLISS (Figures 2 and 3).

Variables Tested and Statistical Analysis

The ability of the scoring system to detect GL-3 inclusions was tested by comparing Thurberg-LM versus BLISS-LM (Table).

Scoring reproducibility for both BLISS-LM and BLISS-VM was assessed using regression analysis and a method adapted from Bland and Altman⁸ for assessing agreement between 2 measurements. Data

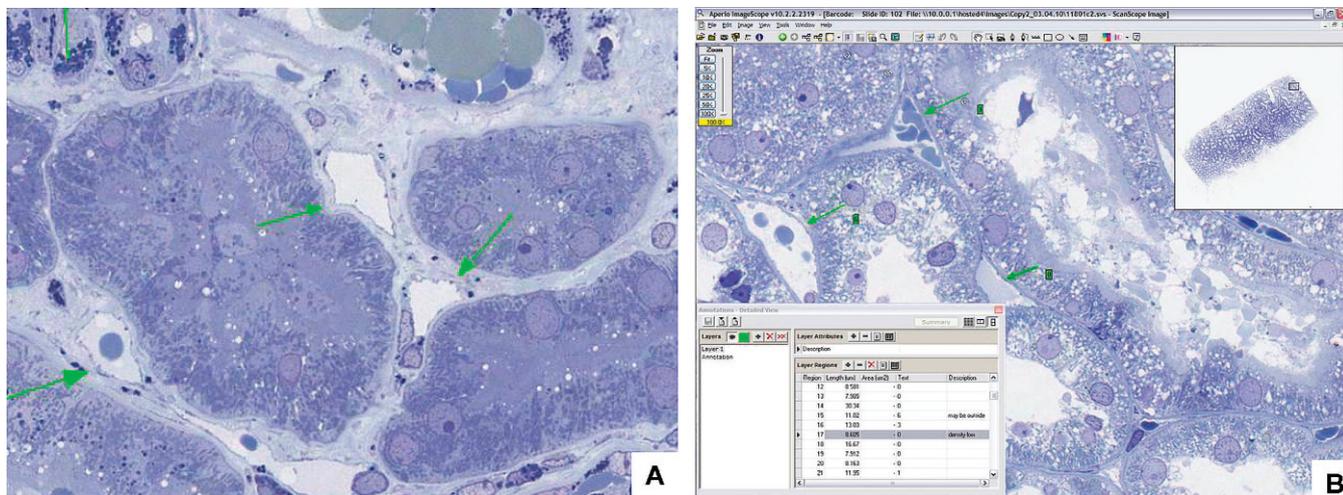


Figure 3. Spectrum screen shot. A, Digital images are first annotated (green arrows) by the annotator. B, The pathologist-readers score each of the peritubular capillaries for globotriaosylceramide inclusions.

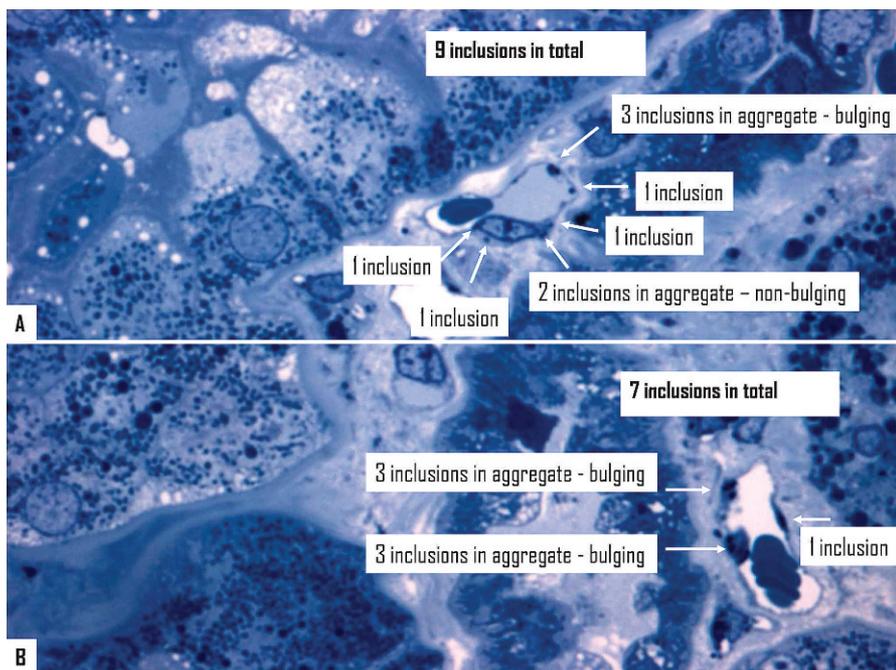


Figure 4. A, Peritubular capillary with a total of 9 globotriaosylceramide (GL-3) inclusions. According to the Barisoni Lipid Inclusion Scoring System (BLISS), the score for this capillary is 9. The same capillary has 1 bulging aggregate and 1 nonbulging aggregate in addition to isolated single small GL-3 inclusions and has a score of 3 according to the Thurberg semiquantitative scoring system. B, A peritubular capillary with a total of 7 GL-3 inclusions. According to BLISS, the score for this capillary is 7. The same capillary has 3 bulging aggregates in addition to a single small GL-3 inclusion and has a score of 3 according to the Thurberg semiquantitative scoring system.

analyses were performed in SAS-JMP 8.0 (SAS Institute, Cary, North Carolina) using bivariate plots and matched-pairs analysis. To prevent reader bias, slides were deidentified and scores were not shared between readers. In addition, slides for the intrareader assessment were renumbered before the second scoring to maintain reader objectivity.

Intrareader variability was assessed by plotting 2 sets of quantitative scores for a set of biopsy slides and performing linear regression to estimate relative accuracy of the second score versus the first score from the same reader. The degree of agreement between the 2 sets of scores was then assessed by plotting each slide's score difference versus the mean and by calculating the limits of agreement as the 95% confidence intervals of the differences. A narrow range in the limits of agreement that included zero within the upper and lower bounds was interpreted as a measure of good agreement between 2 sets of scores.

Similarly, interreader variability was assessed by plotting each score from the second reader against the first reader and performing regression analysis to estimate relative accuracy of the scores by the second reader versus the first reader. The degree of agreement was also assessed by plotting score differences between readers versus the mean and calculating the limits of agreement as the 95% confidence interval of the differences.

RESULTS

Ability of Scoring System to Detect GL-3 Inclusions

The scores assigned by the 2 pathologist-readers by both the Thurberg-LM and the BLISS-LM method are presented in the Table. In general, the semiquantitative Thurberg-LM score and the corresponding quantitative BLISS-LM indicated low amounts of GL-3 inclusions. However, in some of the cases the BLISS-LM method provided evidence of GL-3 inclusions even when the Thurberg-LM score was 0.

Intrareader Reproducibility

BLISS-LM method.—The intrareader reproducibility of the BLISS-LM method is shown in Figure 5. There was a

relative bias (constant error) between the read scores. The slope of 0.6 indicates that the second read tended to have lower values (Figure 5, A). The 95% confidence intervals of the differences between the 2 readings were -0.68 and 0.05 , with a mean difference of -0.31 ($n = 17$) (Figure 5, B). Although zero is within the confidence bounds, most of the difference values were negative, indicating a tendency for the second set of scores to be lower than the first.

BLISS-VM method.—The intrareader reproducibility of the BLISS-VM method is shown in Figure 6. Scores were highly reproducible. A slope of 1.02 indicates a lack of bias in the repeat measurements (Figure 6, A). The 95% confidence intervals of the differences between 2 results were -0.34 and 0.49 , with a mean difference of 0.07 ($n = 12$) (Figure 6, B). A systematic error or bias between the 2 scores was therefore very minimal, confirming good agreement and reproducibility.

Compared to BLISS-LM, the BLISS-VM results showed an improvement in the intrareader reproducibility. Although the BLISS-LM Bland-Altman plots showed 95% confidence intervals that marginally included zero, the BLISS-VM confidence intervals were more equidistant from zero. This indicated that either score (first or second read) in BLISS-VM was equally representative of the true value as estimated by the average.

The relative accuracy of the second read versus the first read scores, as estimated by the regression slope, was also improved in the BLISS-VM. The BLISS-LM slope of 0.6 indicated an estimated proportional error of approximately 40%. On the other hand, the BLISS-VM slope of 1.02 indicated a negligible estimated proportional error and closer to 100% relative accuracy.

Interreader Reproducibility

BLISS-LM method.—The interreader reproducibility of the BLISS-LM method is shown in Figure 7. The regression line, with a slope of 1.32, seems to indicate a

Categoric Versus Quantitative Scoring Systems: Scoring of Globotriaosylceramide Inclusion in Peritubular Capillaries Using Thurberg–Light Microscopy (Thurberg-LM) and Barisoni Lipid Inclusion Scoring System–Light Microscopy (BLISS-LM)^a

Subject	Sex	Thurberg-LM		BLISS-LM	
		Reader 1	Reader 2	Reader 1	Reader 2
1	M	1	1	2.3	2.3
2	M	1	1	4.4	7.4
3	M	1	1	2.3	3.5
4	M	1	1	0.5	0.2
5	M	1	1	0.7	2.0
6	M	1	1	2.4	2.7
7	M	0	1	3.5	2.3
8	M	1	0	0.4	0.2
9	F	1	0	0.7	0.2
10	F	1	1	0.2	0.2
11	F	1	0	0.5	0.1
12	F	1	0	0.3	0.1
13	F	1	0	0.2	0.1
14	F	0	1	0.2	0.2
15	F	1	1	0.2	0.2
16	F	0	0	0.2	0.2
17	F	0	0	0.1	0.1

^a In previously implemented scoring methods (Thurberg-LM), although reported results were numeric values (eg, 0, 1), the scoring assignment was categoric (e.g. ≤ 2 granules, ≥ 3 granules). The BLISS-LM method, on the other hand, is fully quantitative with continuous variable results ranging from 0.1 to 7.4.

tendency of the second reader to assign higher scores (Figure 7, A). However, one high value for the second reader is likely influencing the slope. The differences between scores (Reader 2 – Reader 1) for all matched pairs were plotted against the mean score for each slide (Figure 7, B). The 95% confidence intervals of the differences between the 2 readings were -0.34 and 0.61 , with a mean difference of 0.13 ($n = 17$). Zero lies midway within the narrow 95% confidence intervals, indicating reproducibility of scores between readers.

BLISS-VM method.—The interreader reproducibility of the scores in the BLISS-VM method are shown in Figure 8. Scores were highly reproducible. A slope of 0.93 indicates a lack of significant bias between readers (Figure 8, A). The 95% confidence intervals of the differences between 2 results were -0.21 and 0.35 , with a mean difference of 0.07 ($n = 17$) (Figure 8, B). Because zero lies midway within the confidence limits, a systematic error or bias between 2 readers does not exist, indicating good agreement and reproducibility. Both the BLISS-LM and BLISS-VM showed acceptable interreader reproducibility, with 95% confidence intervals that were similarly equidistant from zero. However, the BLISS-VM showed improvement by having a narrower spread of the 95% confidence interval. The BLISS-LM 95% confidence interval range was 0.95 , whereas the BLISS-VM range was 0.56 . This indicated that in BLISS-VM the individual scores (reader 1 and reader 2) were equally representative of the true value (average) and were also closer to each other. The relative accuracy of the reader 2 versus reader 1 scores, as estimated by the regression slope, was also improved in the BLISS-VM. The BLISS-LM slope of 1.32 indicated an estimated proportional error of approximately 32%. On the other hand, the BLISS-VM slope of 0.93 indicated a smaller estimated proportional error (7%) and closer to 100% relative accuracy.

COMMENT

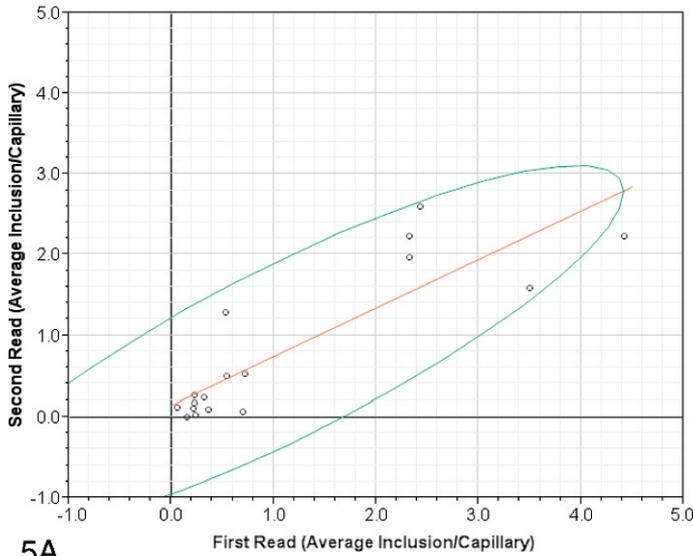
This study was intended to compare (1) 2 different scoring systems, semiquantitative (Thurberg)² and quan-

titative (BLISS), to quantify specific pathologic changes and (2) 2 different protocols, LM- and VM-based, for revision of histologic parameters in 17 kidney biopsies from patients with Fabry disease. Because the kidney biopsies were from patients with residual enzyme activity, one of the key issues was to evaluate the best scoring method when the amount of GL-3 storage in endothelial cells of PTCs is low.

The use of a semiquantitative scoring system in clinical trials or research protocols is common. However, the dependence on skilled operators, issues with the reproducibility of results, and poor documentation and archiving of findings are significant disadvantages. In addition, semiquantitative scoring systems do not always provide sufficient precision to capture small differences between patients with a given disease or difference in parameters between pretherapy and posttherapy. In Fabry disease, a semiquantitative system has been previously used in a cohort of patients with significant amount of renal storage of GL-3. Our study provides evidence that in Fabry patients with a moderate level of storage by morphologic analysis, a more sensitive scoring system is required. Most of patients enrolled in phase 2 trials of AT1001 scored 0 by the semiquantitative approach, Thurberg-LM (Table), whereas the BLISS-LM allowed the detection of GL-3 inclusions in all patients' PTCs, indicating BLISS-LM had higher sensitivity.

Pathologic evaluations for which scoring of glass slides is performed are used in clinical trials, validation of morphologic classification systems, and other research protocols. Morphologic evaluation and scoring of glass slides have been performed using a conventional LM approach (Figure 1). However, the LM-based approach has limitations, including the possible loss or damage of slides during mailing, high intrareader and interreader variability, the fading of staining with time, and the process's being altogether more tedious. When using conventional LM, pathologists look under the microscope and search for a specific structure and morphologic

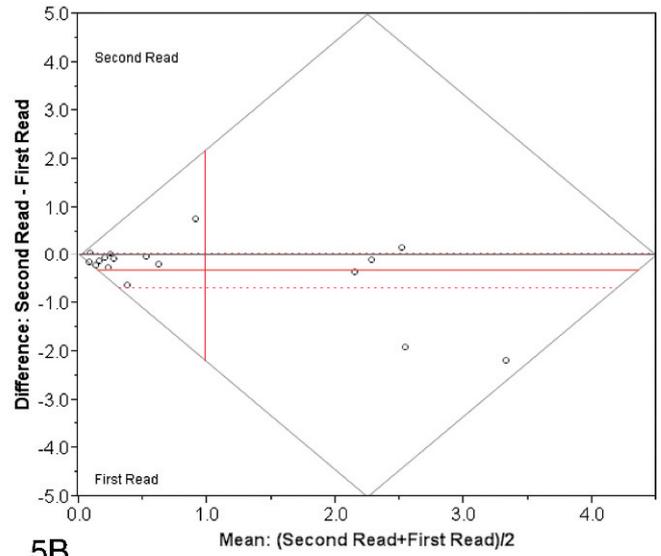
Second Read vs. First Read Scores



5A

Slope: 0.60
Intercept: 0.14
R²: 0.74

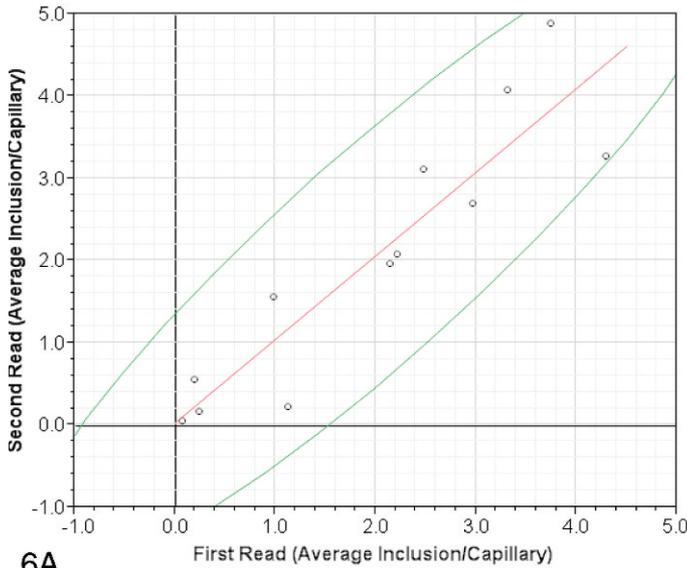
Read Difference vs. Mean Score



5B

Mean difference: -0.31 (N=17)
95% CI: -0.68 to 0.05
Correlation: 0.86

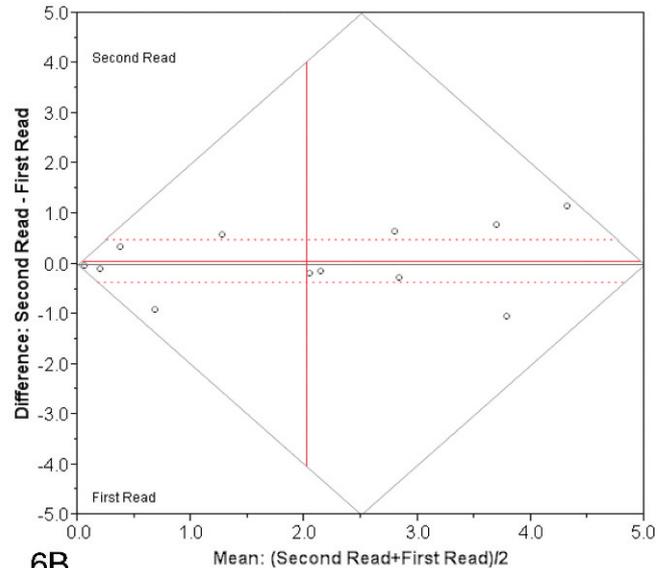
Second Read vs. First Read Scores



6A

Slope: 1.02
Intercept: 0.04
R²: 0.84

Read Difference vs. Mean Score



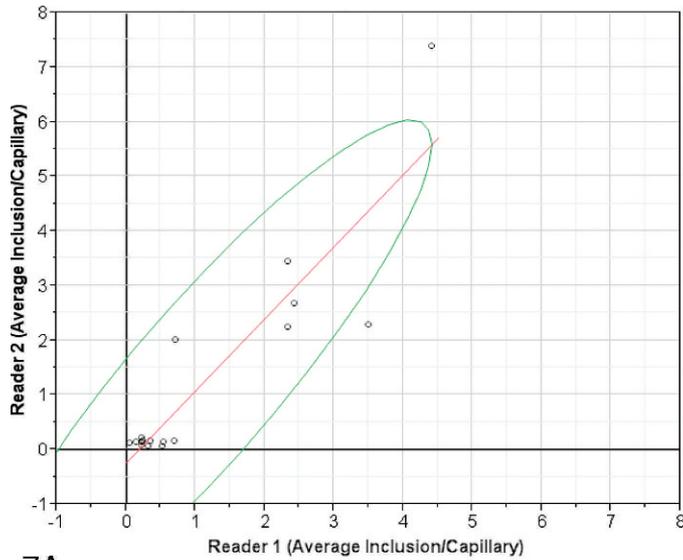
6B

Mean difference: 0.07 (N=12)
95% CI: -0.34 to 0.49
Correlation: 0.91

Figure 5. Intrareader variability—Barisoni Lipid Inclusion Scoring System by light microscopy (BLISS-LM). A, Regression plot of second read versus first read scores within readers using the BLISS-LM (left). B, Bland-Altman plot of the difference between second and first read scores versus the average score in BLISS-LM (right). The mean difference is shown as the red horizontal solid line (-0.31), with the upper and lower 95% confidence intervals (CI) shown as red dotted lines. The gray horizontal line at zero is within limits but is very close to the border of the upper bound, indicating that the mean difference is not significantly different from zero at the 0.05 level but with a consistent negative bias for the second read. The mean of the mean of pairs is shown by the red vertical solid line.

Figure 6. Intrareader variability—Barisoni Lipid Inclusion Scoring System by virtual microscopy (BLISS-VM). A, Regression plot of second read versus first read scores within reader using BLISS-VM (left). B, Bland-Altman plot of the difference between second and first read scores versus the average score in BLISS-VM (right). The mean difference is shown as the red horizontal solid line (0.07), with the upper and lower 95% confidence intervals shown as dotted lines. The confidence region includes the horizontal line at zero, indicating that the mean difference is not significantly different from zero at the 0.05 level. The mean of the mean of pairs is shown by the red vertical line.

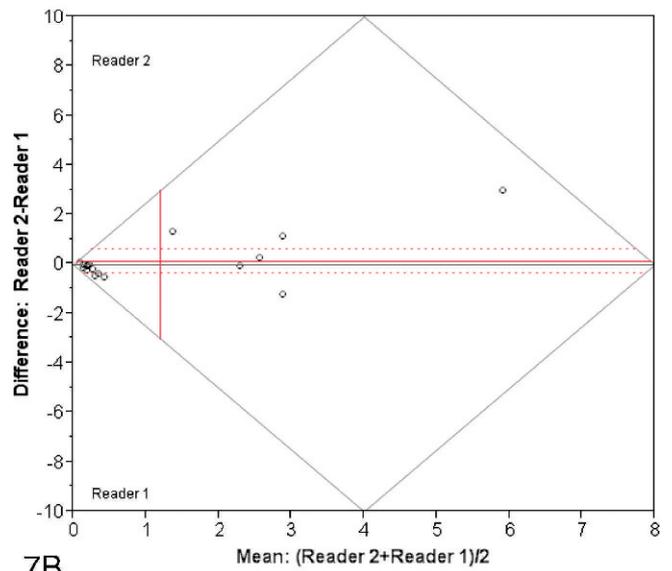
Reader 2 vs. Reader 1 Scores



7A

Slope: 1.32
Intercept: -0.23
R²: 0.82

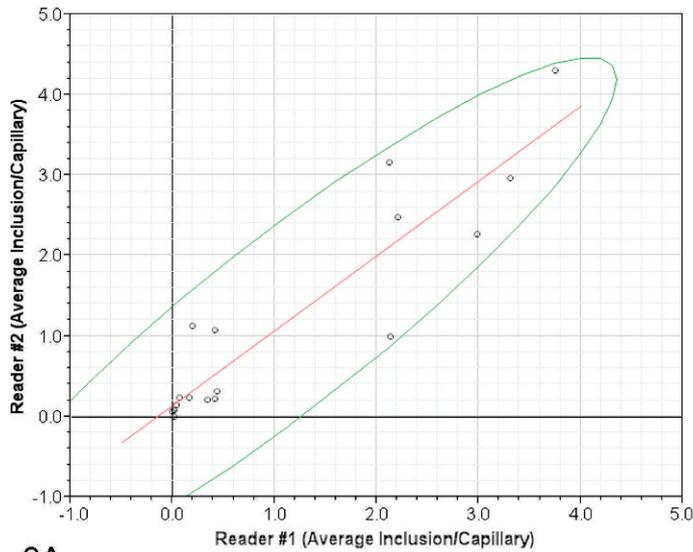
Reader Difference vs. Mean Score



7B

Mean difference: 0.13 (N=17)
95% CI: -0.34 to 0.61
Correlation: 0.91

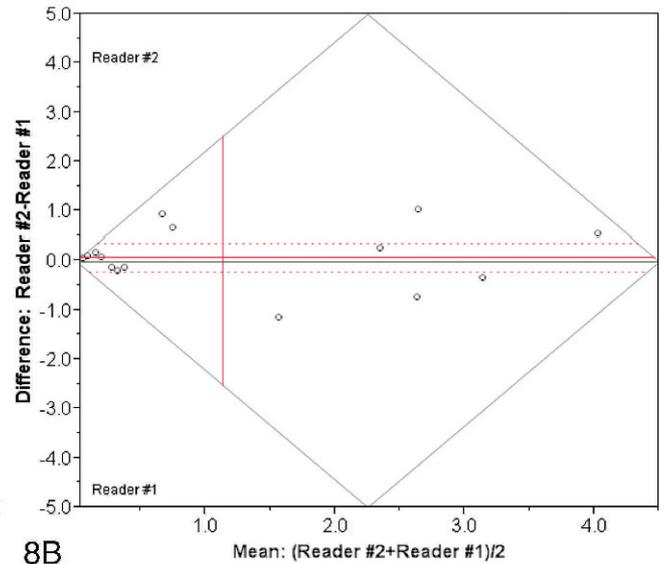
Reader 2 vs. Reader 1 Scores



8A

Slope: 0.93
Intercept: 0.14
R²: 0.84

Reader Difference vs. Mean Score



8B

Mean difference: 0.07 (N=17)
95% CI: -0.21 to 0.35
Correlation: 0.92

Figure 7. Interreader variability—Barisoni Lipid Inclusion Scoring System by light microscopy (BLISS-LM). A, Regression plot of reader 2 versus reader 1 scores using BLISS-LM (left). B, Bland-Altman plot of the difference between reader 2 and reader 1 scores versus the average score in BLISS-LM (right). The mean difference is shown as the horizontal line (0.13), with the upper and lower 95% confidence intervals (CI) shown as dotted lines. The confidence region includes the horizontal line at zero, indicating that the mean difference is not significantly different from zero at the 0.05 level. The mean of the mean of pairs is shown by the vertical line.

Figure 8. Interreader variability—Barisoni Lipid Inclusion Scoring System by virtual microscopy (BLISS-VM). A, Regression plot of reader 2 versus reader 1 scores using BLISS-VM (left). B, Bland-Altman plot of the difference between reader 2 and reader 1 scores versus the average score in BLISS-VM (right). The mean difference is shown as the horizontal line (0.07), with the upper and lower 95% confidence intervals shown as dotted lines. The confidence region includes the horizontal line at zero, indicating that the mean difference is not significantly different from zero at the 0.05 level. The mean of the mean of pairs is shown by the vertical line.

findings to score, and then physically record data on a spreadsheet kept near the microscope. This is time consuming, and transcription errors are likely if several structures need to be scored within the same slide, such as in the case of GL-3 inclusions in PTCs. Searching for a specific structure a hundred times within the same glass slide and recording of the data is itself susceptible to mistakes due to fatigue. Moreover, density of GL-3 inclusions may vary from capillary to capillary, resulting in high interreader variability if the 2 pathologist-readers are looking at different capillaries within the same glass slide. Similarly, intrareader variability can also be high. Agreement among pathologists on specific findings has always been problematic. Previous studies have shown that expert renal pathologists may disagree on multiple parameters, including such basic features as the total number of glomeruli per slide.⁹ An additional issue with conventional LM methodologies is the need of consensus meetings to resolve discrepancies associated with high interreader variability. Consensus meetings are time consuming as well as expensive, especially if international travel is required.

The recent introduction of virtual microscopy has made possible digital exploration of histologic slides, archiving of virtual slides, and immediate access to this information, enabling an “anywhere, anytime” model. The digital image or virtual slide is built from sequential captures of microscopic fields.^{10,11} Virtual slides are high-resolution images that are easy to access and to interpret from any computer. Virtual microscopy is widely used in education^{12,13} and has been introduced in many research protocols in pathology.¹⁴ Its potential use for diagnostic purposes and clinical care is currently under evaluation.^{15–19}

This study is, to our knowledge, the first application of VM in the renal pathology field for scoring of specific morphologic features. A key benefit of VM over LM, independent from training and experience of the pathologists or the accuracy of any scoring system, is that the interreader and intrareader variability is improved. This has always been an issue with LM, notably in clinical registration trials where regulatory agencies require high accuracy and reproducibility.

The BLISS-VM method is based on a 2-step process: annotation of PTCs and then scoring of GL-3 inclusions in each PTC. The annotation step is critical to assure that both scoring pathologists are not only scoring the same glass slide, but also scoring exactly the same capillary. If the process needs to be repeated, the same pathologist can rescore exactly the same set of previously annotated PTCs. The present study shows that results were generally consistent when GL-3 inclusions were rescored by the same reader or by 2 different readers using BLISS-LM methodology. However, when the slides were scored using BLISS-VM methodology, there was a measureable improvement in relative accuracy and reproducibility and dramatic improvement of intrareader and interreader variability. The linear regression analyses for both sets of scores were much closer to the line of identity for the VM-based protocol compared to the LM-based protocol, indicating better relative accuracy. Also, the 95% confidence intervals of differences were more equidistant from zero, indicating better reproducibility. These improved performances of VM are essential when it is important to detect lower levels of inclusions.

One explanation for the improved reproducibility using the VM-based protocol is related to reduction in pathologist fatigue. Fatigue is reduced for the following reasons. First, the annotator has already searched for the structures to score and marked them visibly on the screen. The reader thus only has to score the inclusions within the selected PTC, not search and score. Second, the scoring occurs directly on the screen, thus eliminating distractions from the microscope to focus on the spreadsheet, and then returning back to the microscope. Third, the scoring can be interrupted any time, as the data can be saved. With LM, the pathologist cannot be interrupted, or he or she has to start over from the beginning. It is difficult to find the same capillaries again by conventional LM and remember which ones have been previously scored.

The use of the VM approach also affects study budgets. Although the VM approach requires an initial investment for the scanner, the greater the number of pathologists participating in the study, the easier it is to amortize the expense. Moreover, the number of study coordinators involved in the study and the mailing expenses can be dramatically reduced. We also experienced a decrease in time spent to evaluate each case with the VM approach compared with conventional LM. In addition, the VM-based protocol eliminates the need for consensus meetings, as pathologists have the ability to access the same virtual images from different locations in PTCs simultaneously, and can resolve discrepancies electronically, dramatically reducing traveling costs. The VM-based system also provides a permanent record of data with full documentation for regulatory agencies. It could also be applicable for validating pathologic classification or other histology-based research protocols. Although the VM-based protocol appears to be superior to conventional LM, it has some limitations. It does not provide depth of field. It adds some levels of complexity, as more steps are involved, each of them requiring additional quality control procedures to be put in place, and the initial capital investment.

In summary, VM-BLISS is a simpler and more sensitive scoring system to measure GL-3 inclusions in PTCs when compared to semiquantitative methodologies. The use of VM allows annotation of PTCs, thus providing better accuracy and decreasing intrareader and interreader variability. The application of VM to histopathologic studies also allows capturing of scores/data on the virtual image, providing a permanent and retrievable record of the images, annotations, and scores.

We thank the Mt Sinai Hospital Department of Pathology, Electron Microscopy Laboratory (New York, New York), Aperio (Vista, California), and Quintiles CEVA (Durham, North Carolina) for technical support.

References

1. Hughes DA, Barba Romero MÁ, Hollak CE, Giugliani R, Deegan PB. Response of women with Fabry disease to enzyme replacement therapy: comparison with men, using data from FOS—the Fabry Outcome Survey. *Mol Genet Metab*. 2011;103(3):207–214.
2. Thurberg BL, Rennke H, Colvin RB, et al. Globotriaosylceramide accumulation in the Fabry kidney is cleared from multiple cell types after enzyme replacement therapy. *Kidney Int*. 2002;62(6):1933–1946.
3. Schaefer RM, Tytki-Szymanska A, Hilz MJ. Enzyme replacement therapy for Fabry disease: a systematic review of available evidence. *Drugs*. 2009;69(16):2179–2205.
4. Schiffmann R, Brady RO. New prospects for the treatment of lysosomal storage diseases. *Drugs*. 2002;62(5):733–742.
5. Hughes DA, Barba Romero MA, Hollak CE, Giugliani R, Deegan PB. Response of women with Fabry disease to enzyme replacement therapy:

comparison with men, using data from FOS—the Fabry Outcome Survey. *Mol Genet Metab*. 2011;103(3):207–214.

6. Najafian B, Svarstad E, Bostad L, et al. Progressive podocyte injury and globotriaosylceramide (GL-3) accumulation in young patients with Fabry disease. *Kidney Int*. 2011;79(6):663–670.

7. Warnock DG, Ortiz A, Mauer M, et al. Renal outcomes of agalsidase beta treatment for Fabry disease: role of proteinuria and timing of treatment initiation [published online ahead of print July 29, 2011]. *Nephrol Dial Transplant*. doi: 10.1093/ndt/gfr420.

8. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307–310.

9. Roberts IS, Cook HT, Troyanov S, et al. The Oxford classification of IgA nephropathy: pathology definitions, correlations, and reproducibility. *Kidney Int*. 2009;76(5):546–556.

10. Weinstein RS, Graham AR, Richter LC, et al. Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Hum Pathol*. 2009;40(8):1057–1069.

11. Afework A, Beynon MD, Bustamante F, et al. Digital dynamic telepathology—the Virtual Microscope. *Proc AMIA Symp*. 1998:912–916.

12. Lee SH. Virtual microscopy: applications to hematology education and training. *Hematology*. 2005;10(suppl 1):151–153.

13. Paulsen FP, Eichhorn M, Brauer L. Virtual microscopy—the future of teaching histology in the medical curriculum? *Ann Anat*. 2010;192(6):378–382.

14. Arnes JB, Begin LR, Stefansson I, et al. Expression of epidermal growth factor receptor in relation to BRCA1 status, basal-like markers and prognosis in breast cancer. *J Clin Pathol*. 2009;62(2):139–146.

15. Furness P. A randomized controlled trial of the diagnostic accuracy of internet-based telepathology compared with conventional microscopy. *Histopathology*. 2007;50(2):266–273.

16. Furness P, Rashbass J. The virtual double-headed microscope: telepathology for all? *Histopathology*. 2000;36(2):182–183.

17. Grabe N, Lahrmann B, Pommerencke T, von Knebel Doeberitz M, Reuschenbach M, Wentzensen N. A virtual microscopy system to scan, evaluate and archive biomarker enhanced cervical cytology slides. *Cell Oncol*. 2010; 32(1–2):109–119.

18. Lopez AM, Graham AR, Barker GP, et al. Virtual slide telepathology enables an innovative telehealth rapid breast care clinic. *Hum Pathol*. 2009; 40(8):1082–1091.

19. Nielsen PS, Lindebjerg J, Rasmussen J, Starklint H, Waldstrom M, Nielsen B. Virtual microscopy: an evaluation of its validity and diagnostic performance in routine histologic diagnosis of skin tumors. *Hum Pathol*. 2010;41(12):1770–1776.