# Machine-Learning-Quantified Lupus Nephritis Histological Features Correlate with NIH Activity and Chronicity Index Subscores

Poster # TH-PO567

## STUDY BACKGROUND

Lupus nephritis (LN) is the most common cause of morbidity and mortality in patients with systemic lupus erythematosus (SLE).[1,2] Between 10%-30% of patients with progressive LN will require kidney replacement therapy, due to kidney failure, within 5 years of diagnosis.[1,2] The National Institutes of Health (NIH) LN activity and chronicity indices, which are based on histological evaluation and scoring of renal biopsies, are used for diagnosis, to confirm the extent of disease, and inform treatment decisions.[1,2,3] However, interobserver variability and poor quantitation limit the utility of histology-based metrics for precision medicine.[2] To mitigate these challenges, and to develop new insight into LN pathobiology, machine learning (ML)-based models that quantify histologic features in LN were co-developed by Genentech and PathAI.

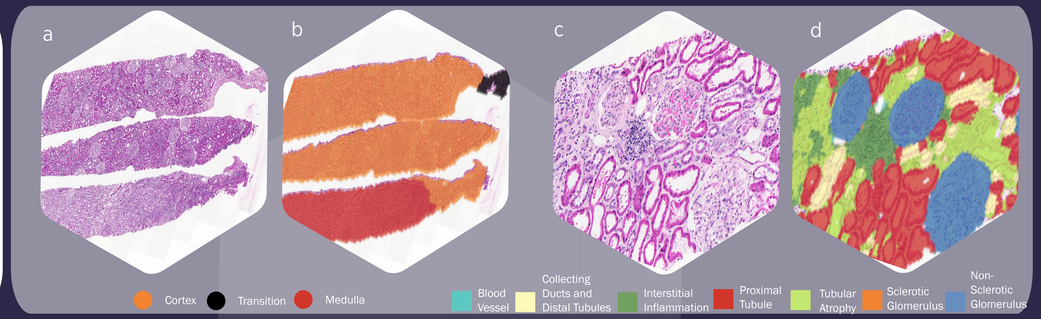



Figure 2. Examples of H&E WSI and resulting ML model predictions depicted as colored overlays. H&E WSI input and Regional Tissue Model predictions are shown in a and b, and H&E input and Feature Tissue Model in c and d.

## CONCLUSIONS

This proof-of-concept study showed that ML models can be trained to quantify regions and features on WSI of LN H&E biopsies with accuracy that is in concordance with pathologists.

Furthermore, ML-predicted HIFs correlate with NIH disease index subscores and kidney function metrics. The utility of this approach in predicting treatment response is being evaluated.

## METHODS

### Model Training

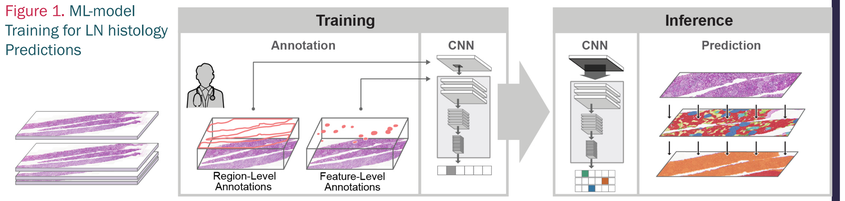
A dataset containing 441 hematoxylin and eosin (H&E)-stained whole-slide images (WSI) of non-LN kidney and LN biopsies (Table 1) was used to train models, based on deep convolutional neural networks (CNN), to predict histologic features of kidney tissue and LN pathology (Figure 1, Figure 2). WSI were annotated by expert pathologists to label histology at the region-level (e.g. cortex, medulla), and feature-level (e.g. glomerulus, tubules) as listed in Table 2, and used as inputs for model training. Annotations of sclerotic and non-sclerotic glomeruli were further used to segment individual glomeruli and thresholds were set to classify each glomerulus as Non Sclerotic (<25% sclerotic area), Partially Sclerotic (25-75% sclerotic area), or Globally Sclerotic (≥75% sclerotic area).

Table 1. Distribution of WSI in datasets used for Model Training and Testing

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Model Training (N=441) | 32 | 171 | - | 238 |
| Model Performance Testing (N=206) | 99 | - | 16 | 91 |

Sources: 1= University of Geneva; 2= randomized clinical trial in proliferative LN (NOBILITY, NCT02550652) baseline samples; 3= NCT02550652 post-treatment, week 52 samples; 4=The Cancer Genome Atlas and commercial data

Figure 1. ML-model Training for LN histology Predictions




ML models were trained using pathologist annotations of H&E stained WSI to predict Region-Level and Feature Level histology relevant to LN

### Feature Extraction

Model predictions were used to generate Human Interpretable Features (HIFs) that quantify and describe relationships between renal features and regions. HIFs included quantification of tissue areas, area proportions, slide-level and individual glomeruli features.

### Evaluation of Model Performance

Model performance was assessed with a held-out test dataset that contained 206 WSI (Table 1). ML-based region and feature predictions were compared with consensus annotations from a group of expert pathologists within a defined area or frame of a WSI using F1 scores as a measure of accuracy. Correlations between manual revised NIH LN activity and chronicity index (CI) subscores and model predictions were investigated from biopsies collected at baseline and week 52 from the NOBILITY trial (Table 1) using Spearman correlation statistics.

## RESULTS

### Model Predictions Compared to Pathologist Consensus Annotations

1. ML-region-level predictions showed high accuracy, performing better on average than a pathologist in identifying cortex and medulla compared to consensus annotations from 3 pathologists across the entire tissue present on 96 WSI (**Region frames**) (Table 2).

2. Feature-level ML-predictions were compared with consensus annotations from 5 pathologists within 143 375 µm x 375 µm frames (**Feature frames**). Although overall performance was strong, the feature-level model was slightly less accurate than the average pathologist, except for interstitial inflammation (F1= 0.70) which is a key indicator of disease severity, prognosis, and treatment approach (Table 2).

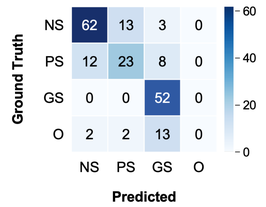
3. **Glomeruli Frames** compared ML predictions of sclerosis in a glomerulus with the consensus label of 3 pathologists in 190 frames that were selected to be centered on a single glomerulus. Models show moderate accuracy in predicting globally, partially, and non-sclerotic glomeruli, with confusion mainly between globally and partially sclerotic predictions (Figure 3).

Table 2. Model Performance Evaluated by Region- and Feature-Level Frames

| Results from Region frames evaluation* | Model | Pathologist |
|---|---|---|
| Cortex | 0.86 | 0.77 [0.74-0.81] |
| Medulla | 0.76 | 0.66 [0.62-0.72] |
| Transition Zone | 0.35 | 0.52 [0.46-0.58] |
| Other (non-renal region) | 0.50 | 0.62 [0.52-0.76] |
| Other than above specified features | 1.0 | 1.0 [1.0-1.0] |
| **Results from Feature frames evaluation#** | **Model** | **Pathologist** |
| Glomerulus | 0.89 | 0.92 [0.90-0.93] |
| Proximal Tubule | 0.85 | 0.88 [0.86-0.89] |
| Collecting Ducts Distal Tubule | 0.56 | 0.61 [0.59-0.64] |
| Tubular Atrophy | 0.47 | 0.51 [0.48-0.54] |
| Interstitial Inflammation | 0.70 | 0.64 [0.61-0.67] |
| Blood Vessel | 0.62 | 0.68 [0.65-0.71] |
| Other (non-renal region) | 0.17 | 0.13 [0.088-0.16] |
| Other than above specified features | 0.66 | 0.68 [0.66-0.69] |

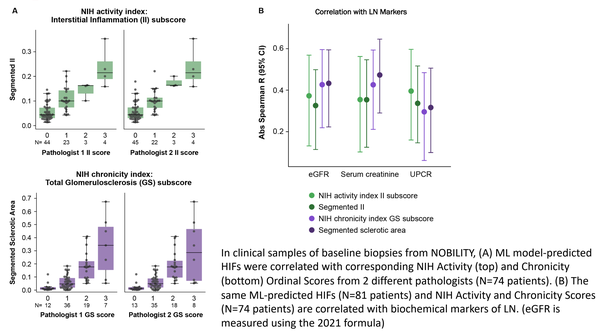
Model column indicates the model's predicted F1 score versus pathologist consensus; Pathologist column indicates each individual pathologist against the remaining consensus. Ranges are 5th and 95th percentile. Consensus from *3 or #5 pathologists

Figure 3. Model Performance Evaluated by Glomeruli Frames




NS: Non Sclerotic; PS: Partially Sclerotic; GS: Globally Sclerotic; O: Other

Figure 4. Correlation of ML-based features with LN disease indices

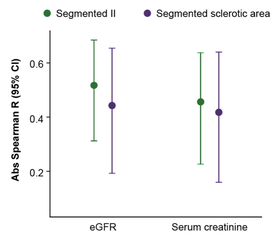



In clinical samples of baseline biopsies from NOBILITY, (A) ML model-predicted HIFs were correlated with corresponding NIH Activity (top) and Chronicity (bottom) Ordinal Scores from 2 different pathologists (N=74 patients). (B) The same ML-predicted HIFs (N=81 patients) and NIH Activity and Chronicity Scores (N=74 patients) are correlated with biochemical markers of LN. (eGFR is measured using the 2021 formula)

To determine how model predicted features correlate with clinically relevant pathologist measurement, selected HIFs were correlated with pathologist-derived NIH activity and chronicity subscores in the NOBILITY cohort.
The continuous ML-predicted area proportion of interstitial inflammation in the cortex (Segmented II) correlated with ordinal scoring of interstitial inflammation (r=0.638, p<0.0001), and ML-predicted proportion of sclerotic area in total glomerulus area (segmented sclerotic area) correlated with ordinal scoring of glomerulosclerosis (r=0.702, p<0.0001) by two different pathologists in each case (Figure 4A).

Figure 5. ML-based features correlate with LN disease indices in a Held-Out Clinical Cohort

In a held-out clinical cohort consisting of post-treatment biopsies from NOBILITY (N=12) plus samples from U. Geneva (N=50), ML-predicted HIFs were correlated with biochemical markers of LN, eGFR and serum creatinine (eGFR is measured using the 2021 formula).




HIF and corresponding NIH subscore pairs correlated with biochemical serum and urine markers, eGFR, creatinine, and UPCR (r=0.3-0.47, p<0.01), that are early indicators of LN (Figure 4B). This trend of moderate correlation between histologic and biochemical markers was still observed in a held-out dataset (Figure 5).

## AUTHORS


Solange Moll[4], Cary D. Austin[2], Balazs Toth[2], Sergio Dellepiane[2*], Filip Kos[1], Stephanie Hennek[1], Christina Jayson[1], Brian H. Baker[1], Benjamin Trotter[1], Webster U. Lincoln[2], Jacqueline A. Brosnan-Cashman[1], Thomas Schindler[3], Murray Resnick[1], Anthony Chang[5], Ilan Wapinski[1], Michael Montalto[1], Jay P. Garg[2], Wei Tew[2], Marco Prunotto[3]

* Presenting Author

1. PathAI, Boston, MA, United States

2. Genentech Inc, South San Francisco, CA, United States

3. F Hoffmann-La Roche AG, Basel, Basel-Stadt, Switzerland

4. Université de Geneve, Geneve, GE, Switzerland

5. University of Chicago Pritzker School of Medicine, Chicago, IL, United States

Contact: christina.jayson@pathai.com


## REFERENCES


1.Yu F, Haas M, Glassock R, Zhao MH. Redefining lupus nephritis: clinical implications of pathophysiologic subtypes. Nat Rev Nephrol. 2017 Aug;13(8):483-495

2.Parikh SV, Almaani S, Brodsky S, Rovin BH. Update on Lupus Nephritis: Core Curriculum 2020. Am J Kidney Dis. 2020 Aug;76(2):265-281.

3.Bajema IM, Wilhelmus S, Alpers CE, Bruijn JA, Colvin RB, Cook HT, D'Agati VD, Ferrario F, Haas M, Jennette JC, Joh K, Nast CC, Noël LH, Rijnink EC, Roberts ISD, Seshan SV, Sethi S, Fogo AB. Revision of the International Society of Nephrology/Renal Pathology Society classification for lupus nephritis: clarification of definitions, and modified National Institutes of Health activity and chronicity indices. Kidney Int. 2018 Apr;93(4):789-796..


## ACKNOWLEDGMENTS


We thank the patients, families, and participating investigators in the clinical trials used in this study. All trials were sponsored and funded by Genentech and data analysis was performed by PathAI.
We thank Bioscience Communications for assistance with figure design. This poster template was developed by SciStories LLC. https://scistories.com


PathAI