Evaluation of Interobserver Agreement between Renal Pathologists in the Classification of Lupus Nephritis Using Digital Pathology Image Sets

Sung-Eun Choi
Bundang CHA General Hospital, Korea, Republic of

Objective
ISN/RPS classification is used worldwide for diagnosing lupus nephritis. However, it is well known that the interobserver variability is substantial among nephropathologists by a number of studies. In the era of digital pathology, there is a growing attempt to utilize artificial intelligence in the classification of diagnosis of lupus nephritis. Nevertheless, it is not easy to train aritifical intelligence without the "gold standard". We aimed to evaluate utilize digital pathology images to confirm intra- and interobserver variability among pathologists for the accurate diagnosis and classification of lupus nephritis, and investigate the changes of agreement through learning to achieve an improvement in the quality of diagnosis and provide a gold standard for future AI research.

Methods
After obtaining standard, training, and validation image sets for lupus nephritis using digital pathology images, we aimed to investigate the intra- and interobserver variability among renal pathologists from domestic and international institutions specializing in renal pathology through and internet survey. The participants will be given training images and will undergo a 5-month online consensus meeting several times for learning. Subsequently, they will be asked to diagnose standard and validation images, and we will evaluate intra- and interobserver variability, and differences of variability after consensus meeting.

Results
Thirty-one members of Korean renal pathology society participated. Among them, 16 were experts, who had experienced more than 10 years of diagnosing renal pathology. Twenty-five computer images of a representative glomerulus from lupus nephritis cases were selected from multicenter and designated as standard image set. Overall Fleiss kappa values for each image ranged from 0.184 to 0.612. Overall agreement for normal, mesangial hypercellularity, fibrous crescent, and double contour was slight. The Fleiss kappa values were 0.082, 0.128, 0.101, and 0.127, respectively. The experts showed slight agreement for mesangial hypercellularity, fibrous crescent, and double contour and the Fleiss kappa values were 0.122, 0.037, and 0.160. Their overall Fleiss kappa values for each image ranged from 0.176 to 0.678.