

**SUPPORTING INFORMATION****Robotic bronchoscopic needle based confocal laser endomicroscopy to diagnose peripheral lung nodules**

Christopher J. Manley<sup>1</sup>, Tess Kramer<sup>2</sup>, Rohit Kumar<sup>1</sup>, Yulan Gong<sup>3</sup>, Hormoz Ehya<sup>3</sup>, Eric Ross<sup>4</sup>, Peter I. Bonta<sup>2</sup>, Jouke T. Annema<sup>2</sup>

<sup>1</sup> Department of Pulmonary and Critical Care, Fox Chase Cancer Center, Philadelphia, PA, USA

<sup>2</sup> Department of Respiratory Medicine, Amsterdam UMC, Amsterdam, Netherlands

<sup>3</sup> Department of Pathology, Fox Chase Cancer Center, Philadelphia, PA, USA

<sup>4</sup> Department of Biostatistics and Bioinformatics, Fox Chase Cancer Center, Philadelphia, PA, USA

**Appendix S1-METHODS**

In approximately 15 minutes, five raters (two pulmonologists, two pathologists and one researcher) were trained in nCLE malignancy and airway/lung parenchyma criteria. No raters had been involved in the data acquisition or analysis. One rater was experienced with the nCLE malignancy and airway/lung parenchyma criteria (expert), 2 raters were experienced with probe-based CLE-imaging (in the thorax and bladder) but inexperienced with nCLE-malignancy and airway/lung parenchyma (semi-expert) and 2 raters were completely nCLE-naïve (non-expert).

Of the 54 obtained nCLE videos, 6 nCLE videos were used for training purposes and 26 videos (n=15 airway/lung parenchyma, n=11 malignancy) for validation purposes (32/54, 59%). Blinded raters distinguished nCLE-videos of malignancy from airway/lung parenchyma with an overall sensitivity and specificity of 91.9% respectively (Table 1). Of the nCLE videos scored with good confidence (n=157, 60%), 97% were rated correctly, compared to 88% and 72% in videos with moderate (n=74, 28%) or poor confidence (n=29, 11%). The experienced rater had an overall accuracy of 96.2%, compared to 93.3% with semi-expert raters and 88.5% with non-expert raters. The IOA and IOR between the five raters were substantial (Table 2). All individual nCLE malignancy and airway/lung parenchyma criteria were recognized with moderate or substantial agreement (Table 2). The nCLE malignancy criteria 'enlarged pleomorphic cells' and 'dark clumps' were present in 42% and 40% of the validation videos, compared to 12% for 'directional streaming'. Of the nCLE airway/lung parenchyma criteria, 'still image' was present in 31% of the validation videos, compared to 20% for 'elastin fibers', 16% for 'bronchial epithelium' and 11% for 'alveoli'.

**Table S1. Scoring performances of blinded raters (n=5) for the distinction of malignant nCLE-videos from airway/lung parenchyma.**

	Accuracy	Sensitivity	Specificity	NPV	PPV
All raters (n=5)	91.9	91.9	91.9	93.8	89.5
Expert (n=1 rater)	96.2	90.9	100	93.8	100
Semi-expert (n=2 raters)	93.3	93.2	93.3	94.9	91.1
Non-expert (n=2 raters)	88.5	91.1	86.4	92.7	83.7

In 20 patients, 26 nCLE-videos were selected and scored twice by five raters (total of 260 ratings). nCLE, needle based confocal laser endomicroscopy; NPV, negative predictive value; PPV, positive predictive value

**Table S2. Interobserver agreement and intra-observer reliability for the final diagnosis and nCLE malignancy and airway/lung parenchyma criteria**

Final diagnosis	K-value
IOA	0.80 (0.71-0.81)
IOR (SD)	0.80±0.05
nCLE malignancy criteria	
Enlarged pleomorphic cells	0.79 (0.71-0.87)
Dark clumps	0.70 (0.62-0.78)
Directional streaming	0.63 (0.54-0.73)
nCLE airway/lung parenchyma criteria	
Alveoli	0.46 (0.37-0.55)
Elastin fibers	0.68 (0.60-0.77)
Still image	0.76 (0.68-0.84)
Bronchial epithelium	0.72 (0.63-0.81)

In 20 patients, 26 nCLE-videos were selected and scored twice by five raters (total of 260 ratings). The data are presented as the IOA  $\kappa$  (95%CI) unless specified otherwise. Landis-Koch interpretation system: poor <0.2, fair 0.21-0.4, moderate 0.41-0.6, substantial 0.61-0.8 and excellent 0.81-1. IOA, inter-observer agreement; IOR, intra-observer reliability; nCLE, needle based confocal laser endomicroscopy.