Assistance of an artificial intelligence–powered PD-L1 analyser reduces interobserver variation in pathologic reading of tumour proportion score in non-small cell lung cancer

Sangjoon Choi[1], Soo Ick Cho[2], Seokhwi Kim[3], Hyojin Kim[4], Minuk Ma[2], Seonwook Park[2], Sergio Pereira[2], Brian Jaehong Aum[2], Seunghwan Shin[2], Kyunghyun Paeng[2], Donggeun Yoo[2], Wonkyung Jung[2], Chan-Young Ock[2], Se-Hoon Lee[5], Yoon-La Choi[1], Jin-Haeng Chung[4], Tony S. Mok[6]

[1]Department of Pathology and Translational Genomics, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea, [2]Lunit Inc., Seoul, Republic of Korea, [3]Department of Pathology, Ajou University School of Medicine, Suwon, Korea, [4]Department of Pathology, Seoul National University Bundang Hospital, Seongnam, Republic of Korea, [5]Division of Hematology-Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea, [6]State Key Laboratory of Translational Oncology, Department of Clinical Oncology, Chinese University of Hong Kong, Hong Kong, China

Abstract #2440

E-Poster 1805P

*The presenting author has no conflicts of interest to declare

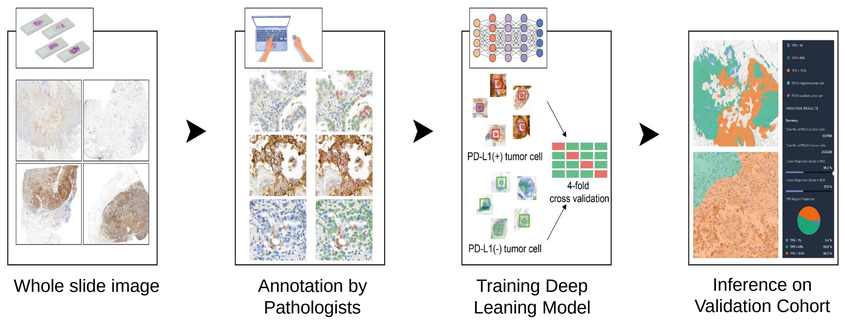
Background

- Programmed death ligand 1 (PD-L1) expression is the standard biomarker in advanced NSCLC. However, manual evaluation of PD-L1 tumor proportion score (TPS) by pathologists has practical limitation associated with interobserver bias.
- We developed an artificial intelligence (AI)-powered TPS analyzer, namely Lunit SCOPE PD-L1, for objective annotation of tumor cell PD-L1 expression for prediction of ICI response in advanced NSCLC, and explored whether an AI-powered TPS analyser could reduce interobserver variation.

Methods

- Lunit SCOPE PD-L1 was developed by a total of 393,565 tumor cells annotated by board-certified pathologists for PD-L1 expression in 802 whole-slide images (WSI) stained by 22C3 pharmDx immunohistochemistry.
- A 4-fold cross validation approach was used to find the best hyper-parameters. After excluding the in-house control tissue regions, the WSI were divided into patches, from which a deep learning-based model detected the location and PD-L1 positivity of tumor cells. The patch-level cell predictions were aggregated for TPS estimation.
- Three independent board-certified pathologists labelled PD-L1 TPS in an external dataset of 479 NSCLC tumour slides. The TPS of each pathologist for a slide was divided into three groups (TPS <1%, 1%–49%, and ≥50%). Consensus was defined in terms of whether the three pathologists scored the same TPS group (all agreed or two agreed). For cases of disagreement between each pathologist and the AI model, the pathologists were asked to revise the TPS grade assisted by the AI model.
- Finally, we compared the concordance rate of the three pathologists with or without AI assistance and analysed the effect on clinical outcome of immune checkpoint inhibitors (ICIs).

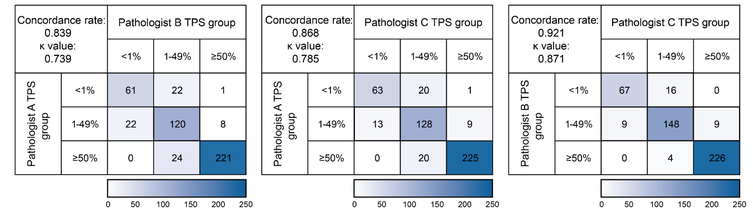
Development of Lunit SCOPE PD-L1 model



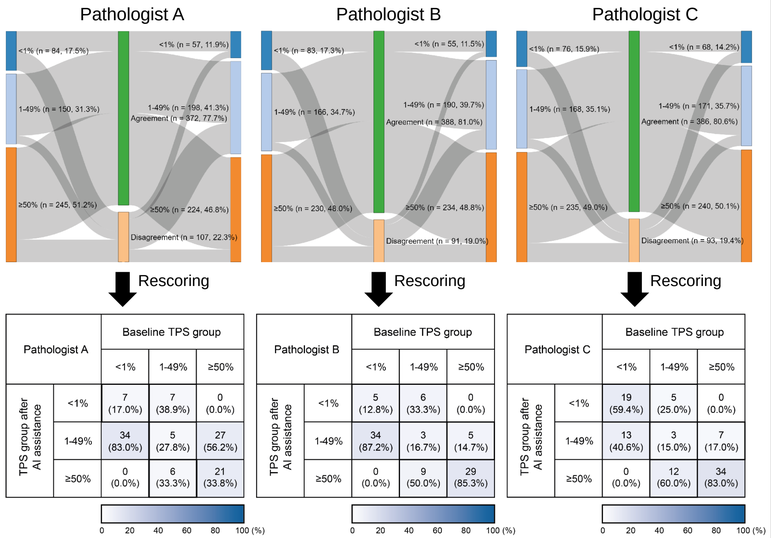
Whole slide image → Annotation by Pathologists → Training Deep Leaning Model → Inference on Validation Cohort

Results

1. Concordance among pathologists and AI model

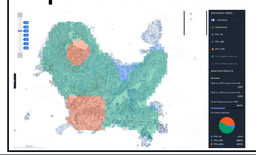


- The concordance and Cohen's κ between two of the three pathologists were 92.1% and 0.871 in the highest case, and 83.9% and 0.739 in the lowest case, respectively



- Each pathologist revised their baseline TPS group for the human-AI disagreement cases (N = 91, 93, and 107, respectively.)

AI-powered PD-L1 analyzer Demo



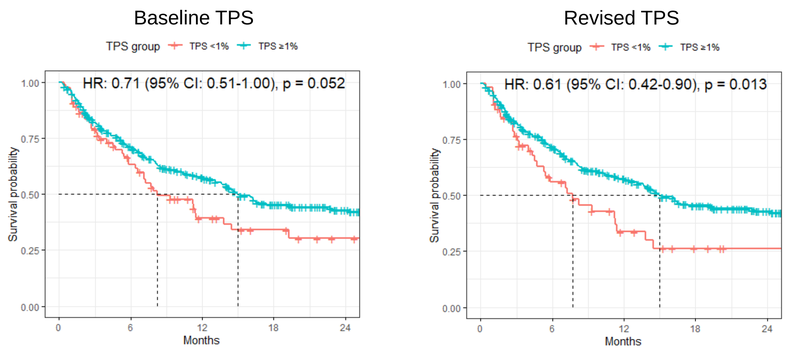
Copies of this e-Poster obtained through QR, AR and/or text key codes are for personal use only and may not be reproduced without written permission of the author

2. The baseline and revised consensus of three pathologists

	Baseline consensus of three pathologist		Revised consensus after LUNIT SCOPE PD-L1 assistance	
	Concordant (Three agreed)	*Discordant (Two agreed)*	*Concordant (Three agreed)*	*Discordant (Two agreed)*
All	390 **(81.4%)**	89 **(18.6%)**	432 **(90.2%)**	47 **(9.8%)**
TPS <1% group	55 (67.9%)	26 **(32.1%)**	52 (89.6%)	6 **(10.4%)**
TPS 1-49% group	117 (72.2%)	45 **(27.8%)**	162 (86.2%)	26 **(13.8%)**
TPS ≥50% group	218 (92.4%)	18 **(7.6%)**	218 (93.6%)	15 **(6.4%)**

- After AI assistance, the overall concordance rate increased to 90.2% (N = 432/479, κ = 0.890) from 81.4% (N = 390/479, κ = 0.798).
- Interobserver variability was substantially improved in the subgroups of TPS<1% (18.6%→9.8%) and 1-49% (32.1%→10.4%)

3. OS according to PD-L1 status before and after AI assistance



- Two TPS groups (<1% and ≥1%) showed no significant difference when OS was predicted based on the baseline consensus of 3 pathologists (HR: 0.71, 95% CI: 0.51-1.00, P = 0.052) before AI assistance.
- Revised TPS showed significant difference of OS betwwen two groups(HR: 0.61, 95% CI: 0.42-0.90, P = 0.013).

Conclusion

- TPS interpretation assisted by AI can reduce the heterogeneity between pathologists, and more accurately predict clinical outcome of ICIs.