

Deep learning-based prediction of multiple Actionable Genomic Alterations and HER2 over-expression in NSCLC

Sylvana Hassanieh¹, Jack Rawson¹, Eliza Naismith¹, Inbal Gazy², Jonathan Zalach², Ross J. Hill¹, Anne Shah³, Espen Walker¹, Yishan Chuang³, Fernando Lopez-Rios⁴

¹AstraZeneca, Cambridge/GB, ² Imagene AI Inc., USA, ³AstraZeneca, Gaithersburg/MD/USA, ⁴Hospital Universitario 12 de Octubre, Universidad Complutense de Madrid, Research Institute Hospital Universitario 12 de Octubre (imas12). CIBERONC, Madrid/ES

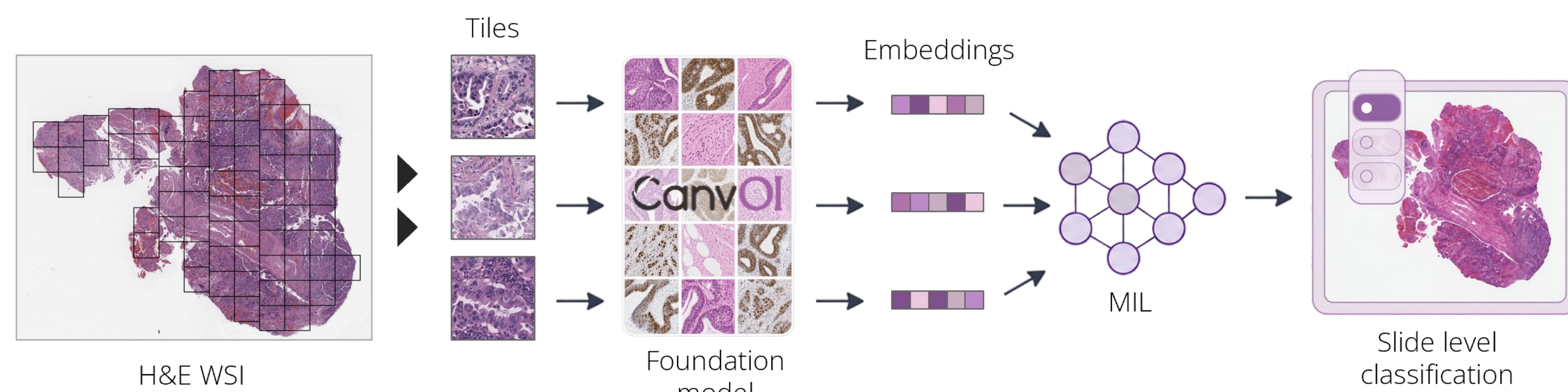
INTRODUCTION

- Molecular profiling for actionable genomic alterations is essential for selecting optimal treatment in non-small cell lung cancer (NSCLC).
- Testing is often hindered by access, insufficient tissue samples, testing failure and long turnaround times, resulting in suboptimal treatment.
- Rapid AI-based prediction of biomarkers directly from routine H&E-stained whole-slide images (WSIs) could help address these challenges¹.
- In this study we developed and validated AI-classifiers to predict NSCLC genomic alterations directly from WSIs of H&E-stained samples.

METHODS

- The lung panel validated in this study includes AI-classifiers for the following NSCLC biomarkers: *EGFR* mutations, *ALK* fusions, *BRAF* V600E, *KRAS* G12C, *ERBB2* (HER2) mutations, *ROS1* fusions, *RET* fusions and *MET* exon 14 skipping.
- In addition, a HER2 expression classifier was developed as part of an initial feasibility. HER2 IHC scoring using 4B5 antibody followed the gastric scoring criteria for surgical specimens. Scores 0/1+ were labeled negative and 2+/3+ overexpressed.

Figure 1: Schematic overview of the AI-based biomarker prediction



H&E whole-slide images (WSIs) are divided into tiles, which are encoded into embeddings using the CanvOI foundation model² and serve as input to downstream Multiple Instance Learning (MIL) algorithms to generate slide-level biomarker predictions.

DISCUSSION AND CONCLUSION

- AI-based prediction from H&E WSIs is feasible in both NSCLC routine and clinical trial samples.
- The observed performance suggests that this approach has the potential to improve access to optimal therapeutic strategies and support patient identification for trial enrollment across diverse clinical settings.

RESULTS

- External validation was performed leveraging two AstraZeneca Phase III trials (POSEIDON and NEPTUNE) and an independent commercial cohort (TICA) (Table 1).

Table 1. Cohorts Characteristics

	POSEIDON Late-stage n = 545	NEPTUNE Late-stage n = 948	TICA Early-stage n = 592	All n = 2,085
Histological subtypes (n)				
Non-squamous	284	589	333	1,026
Squamous	242	349	228	819
Genomic alteration cohorts (n)				
<i>EGFR</i> mutations (positives, n)	12	52	85	149 (7.6%)
<i>ALK</i> fusions (positives, n)	11	19	0	30 (1.5%)
<i>BRAF</i> V600E (positives, n)	6	11	2	19 (1.0%)
<i>KRAS</i> G12C (positives, n)	48	68	76	192 (9.7%)
<i>ERBB2</i> (HER2) mutations (positives, n)	19	32	0	51 (2.6%)
<i>ROS1</i> fusions (positives, n)	7	8	0	15 (0.8%)
<i>RET</i> fusions (positives, n)	10	15	0	25 (1.3%)
<i>MET</i> ex. 14 skip. (positives, n)	7	26	2	35 (1.8%)
Expression cohort (n)				
HER2 expression (positives, n)	40	55	158	253 (20.7%)
HER2 Antibody	CAL27	CAL27	4B5	

Table 2. Performance of the AI-biomarker prediction models

	AUROC	Sensitivity	Specificity
<i>EGFR</i> mutations	0.852	0.796	0.790
<i>ALK</i> fusion	0.926	0.533	0.980
<i>BRAF</i> V600E	0.918	0.684	0.943
<i>KRAS</i> G12C	0.804	0.724	0.752
<i>ERBB2</i> (HER2) mutations	0.842	0.571	0.906
<i>ROS1</i> fusion	0.737	0.429	0.868
<i>RET</i> fusion	0.771	0.400	0.829
<i>MET</i> ex. 14 skip.	0.763	0.429	0.875
HER2 expression †	0.649	0.671	0.548

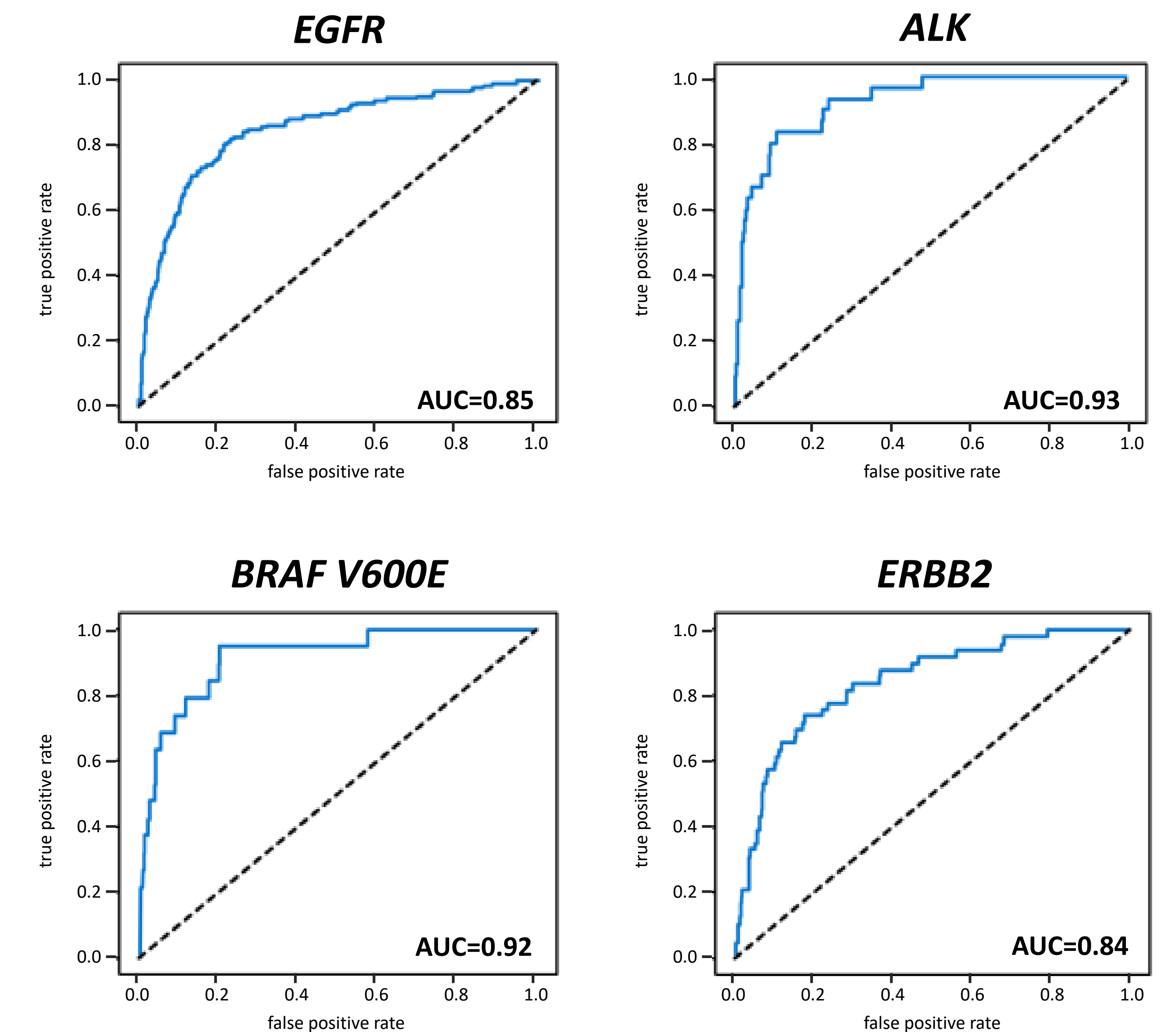
Analysis was performed on the biomarker-evaluable population, including samples from all cohorts.

† Preliminary results for the HER2 expression classifier, developed using 4B5 as part of an initial feasibility study and currently undergoing further development.

Sensitivity and specificity calculated at optimal threshold by Youden's J statistics.
AUCROC = area under the receiver operating characteristic curve

- The AI-classifiers demonstrated consistent performances across key genomic alterations.
- Notably, all investigated genomic alteration biomarkers achieved AUCs exceeding 0.737 (Table 2). Representative ROC curves are presented in Figure 2.
- Additionally, the model effectively predicted HER2 overexpression from H&E-stained images. Further development of the HER2 classifier is ongoing. Notably, higher interobserver agreement in HER2 scoring was associated with improved evaluated AUC, supporting future assessment of the test sets using scores from multiple pathologists.

Figure 2: ROC curves of representative biomarker, highlighting the four highest-performing classifiers



REFERENCES

- Waissengrin B. *et al.* J Clin Pathol. 76:790-2, 2023.
- Zalach et al. arXiv, 2024. 2409.02885