

Supplementary Figures

Enhancing Lung Cancer Classification Through Integration of Liquid Biopsy Multi-Omics Data with Machine Learning Techniques

Hyuk-Jung Kwon^{1,2†}, Ui-Hyun Park^{1†}, Chul Jun Goh¹, Dabin Park¹, Yu Gyeong Lim¹, Isaac Kise Lee^{1,2,3}, Woo-Jung Do¹, Kyoung Joo Lee¹, Hyojung Kim¹, Seon-Young Yun¹, Jounghsu Joo¹, Na Young Min¹, Sunghoon Lee¹, Sang-Won Um⁴, Min-Seob Lee^{1,5*}

¹ R&D Department, Eone-Diagnomics Genome Center, Inc., 291 Harmony-ro, Yeonsu-gu, Incheon, 22014, Republic of Korea

² Department of Computer Science and Engineering, Incheon National University (INU), Incheon, 22012, South Korea,

³NGENI Foundation, California, USA

⁴Division of Pulmonary and Critical Care Medicine, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul, 06351, Republic of Korea

⁵Diagnomics, Inc., 5795 Kearny Villa Rd. San Diego, CA 92123, USA

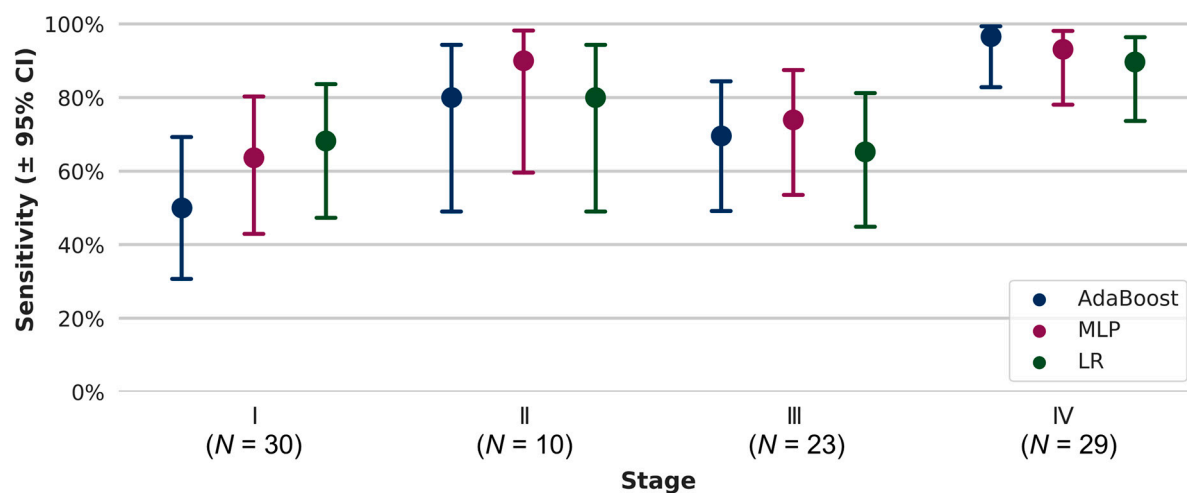
[†]These authors contributed equally

*Correspondence: Min Seob Lee, Ph.D.

Eone-Diagnomics Genome Center, Inc., 291 Harmony-ro, Yeonsu-gu, Incheon, 22014, Republic of Korea

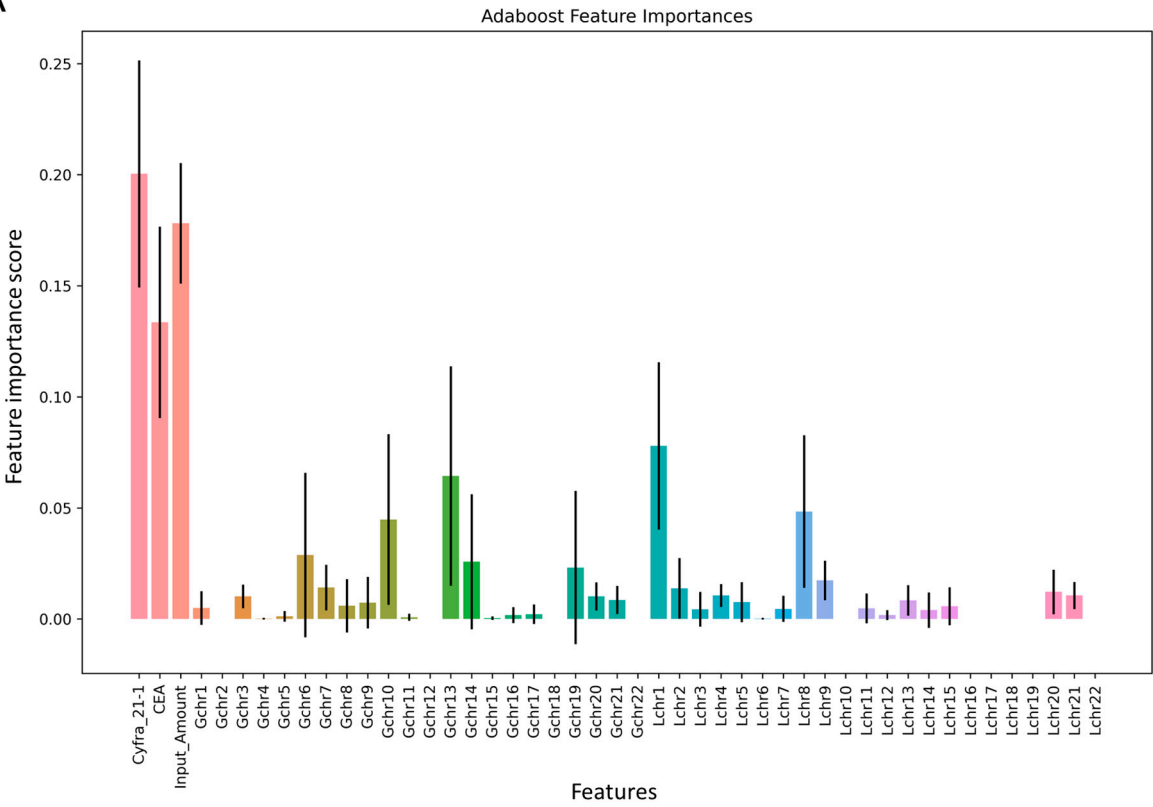
Phone No: +82-10-3080-1393

Email: mlee@edgc.com

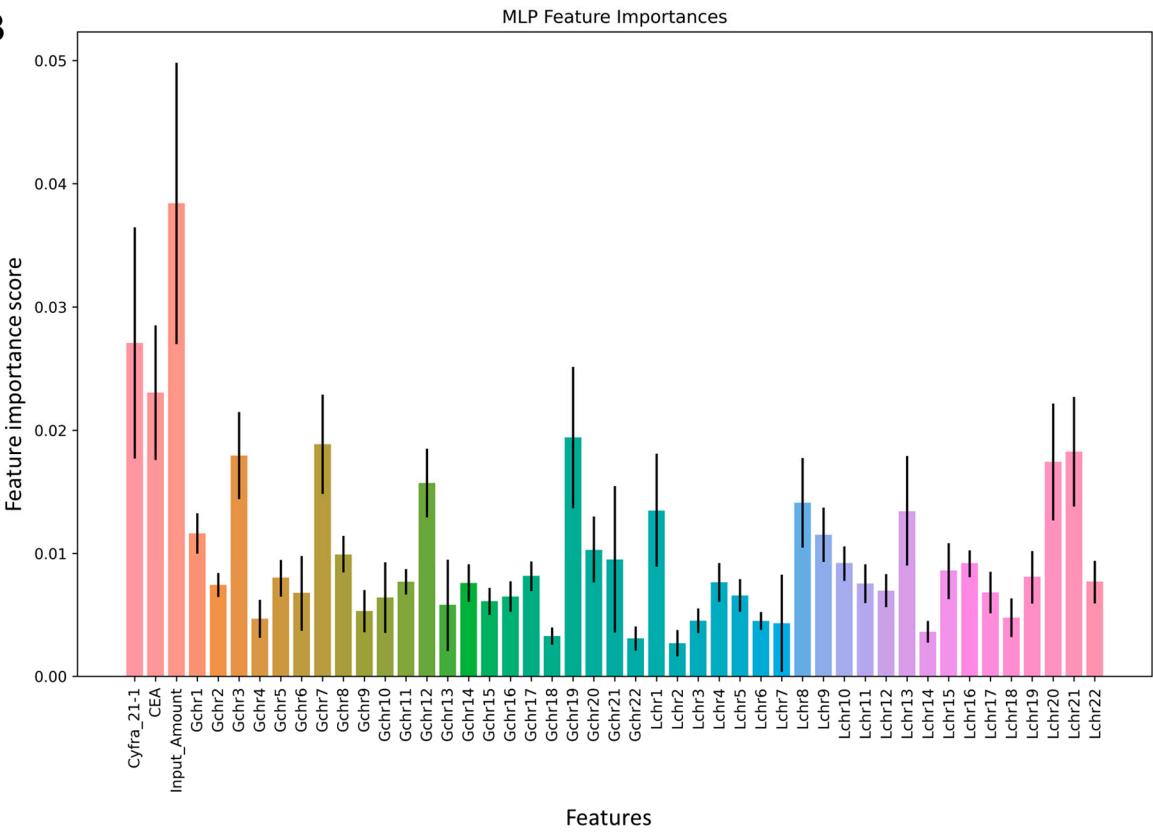


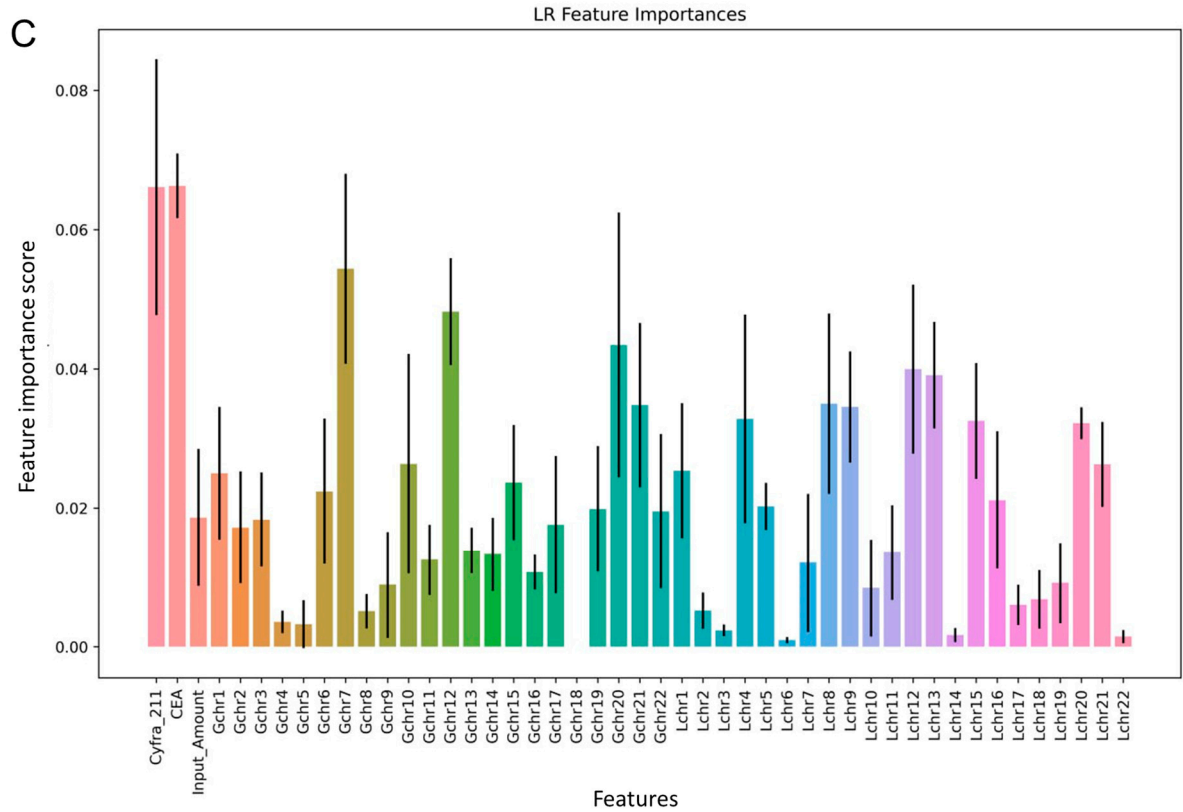
Supplementary Figure S1. Sensitivity plot representing stage-wise cancer detection performance of each machine learning model. The sensitivity values of the three machine learning models, AdaBoost (blue dots), MLP (dark red dots), and LR (dark green dots), are presented using a dot plot. The x-axis shows the stages of lung cancer and 'N' indicates the number of patients. The y-axis represents the sensitivity values. The lines extending above and below each dot represent the 95% confidence interval for that value, considering a specificity of 90%.

A



B





Supplementary Figure S2. Feature importance score of combined multi-omics data in AdaBoost model. Feature importance, encompassing all multi-omics data, for the AdaBoost (A), MLP (B), and LR (C) models in lung cancer prediction, quantified as the mean absolute values. The legend indicates the significance of the blood cancer markers (Cyfra21-1 and CEA), cfDNA concentration (Input_amount), and copy number variation (CNV, Gchr1~Lch22).