# Supplementary Information: Determination of Lung Cancer Exhaled Breath Biomarkers Using Machine Learning – A New Analysis Framework

## Methodology

Calibration curves for o-cymene and hexadecane were constructed using eight-point serial dilutions. Both standards were obtained in liquid form and diluted using micropipettes. A fixed volume (10 µL) of each diluted standard was spiked into 0.5 L Tedlar bags pre-filled with exhaled breath from control participants (not included in the main study) to mimic the real sample matrix. These spiked breath samples were used to determine the sensitivity, as well as the limits of detection (LOD) and quantification (LOQ) for each compound. To assess reproducibility, four replicate runs of the lowest concentration of o-cymene and hexadecane standard were analyzed.

**Table S1.** Summary of calibration curve parameters, LOD, LOQ and RSD (%) for o-cymene and hexadecane. LOD and LOQ were calculated based on the standard deviation of the intercept and the slope, following IUPAC guidelines.

| Compound | Slope (a.u./ppm) | Intercept (a.u.) | $R^2$ | LOD (ppm) | LOQ (ppm) | RSD(%) | Calibration equations |
|---|---|---|---|---|---|---|---|
| O-Cymene | 8548.06 | -7934.82 | 0.998 | 4.89 | 14.83 | 2.16 | y= 8548.06x - 7934.82 |
| Hexadecane | $1.23 \times 10^9$ | $-6.75 \times 10^7$ | 0.997 | 0.08 | 0.25 | 3.49 | $y=1.23 \times 10^9 x - 6.75 \times 10^7$ |

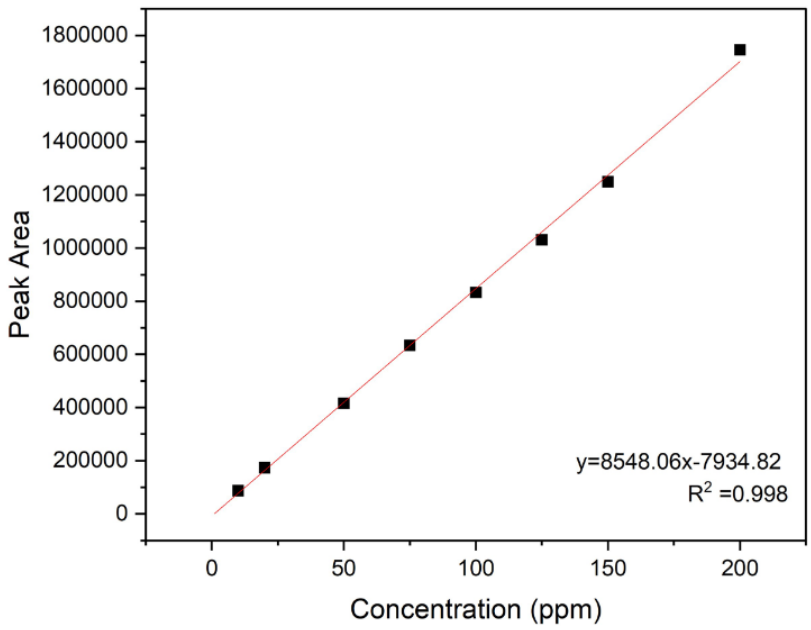


**Figure S1.** Calibration Curve of O-Cymene standard

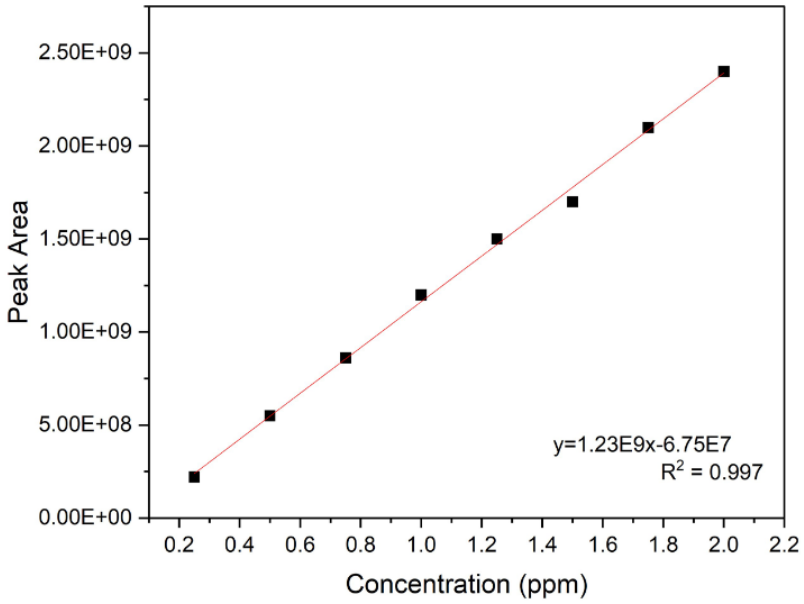


**Figure S2.** Calibration curve of Hexadecane standard

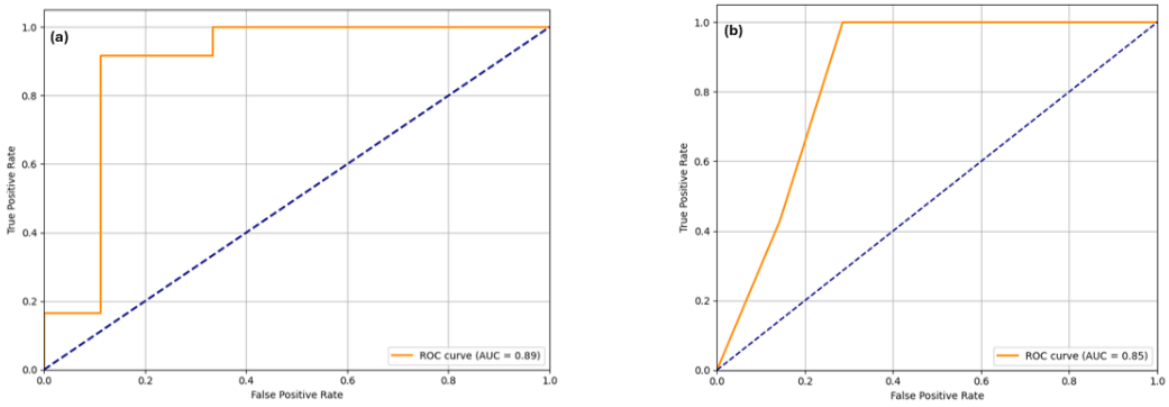


**Figure S3.** Receiver Operating Characteristics (ROC Curve) for **(a)** Support Vector Machine (SVM) and **(b)** K-Nearest Neighbor (KNN) classification models

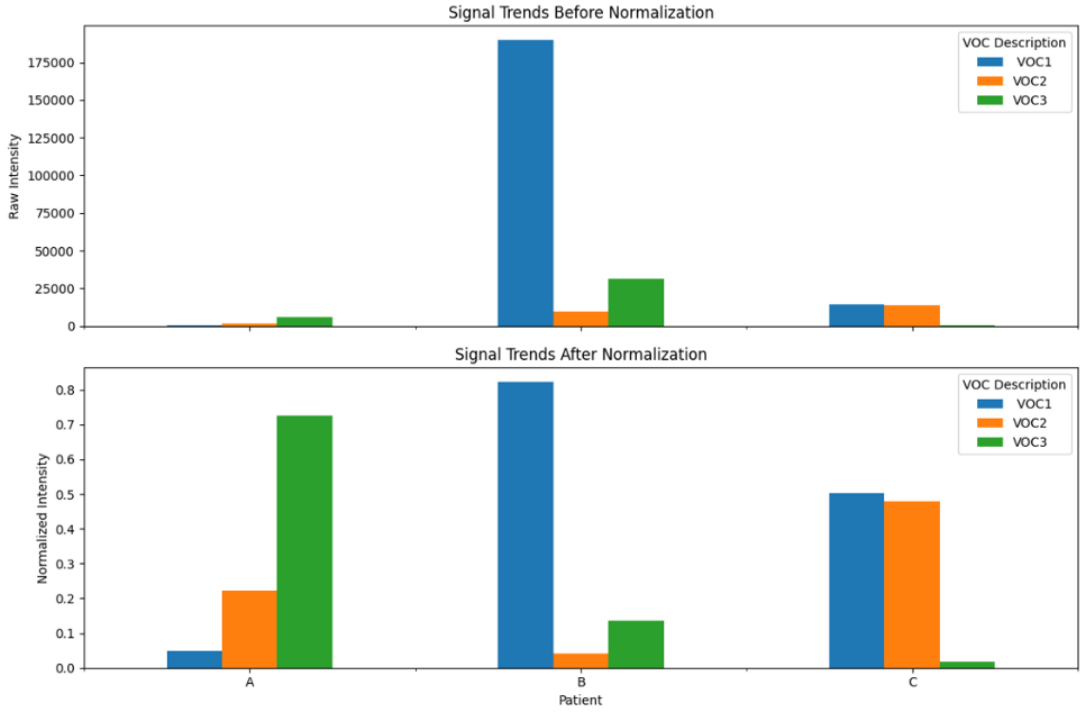


**Figure S4.** Comparison of signal intensities for selected VOCs across three samples before (top) and after (bottom) 1-norm normalization. The normalization process reduces sample-to-sample variation in total signal intensity, allowing consistent comparison of VOC profiles.