

Supplementary

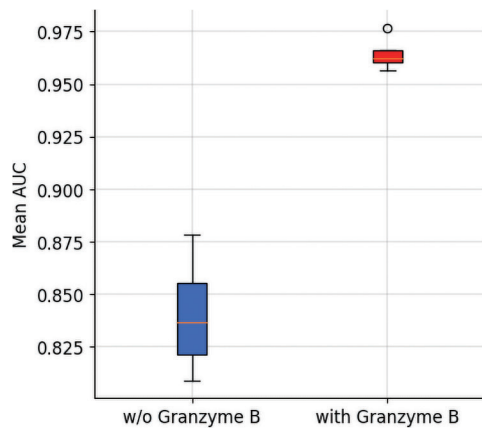


Figure S1 Comparison of the mean AUC of the four machine learning models with and without granzyme B. AUC, area under the curve.

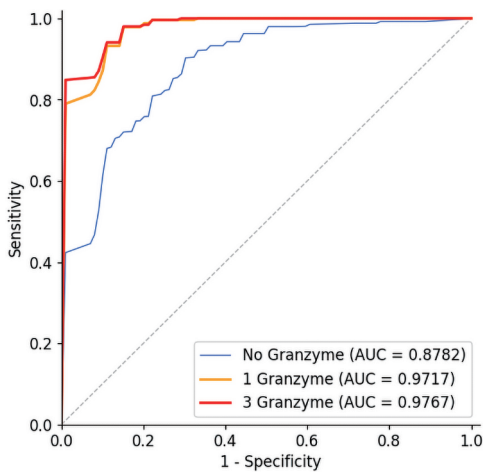


Figure S2 Comparison of ROC curve in Random Forest by the number of granzyme B. ROC, receiver operating characteristics.

Table S1 Performance evaluation of CEA and proGRP for lung cancer detection

| | CEA | ProGRP |
|------------------------------|------|--------|
| Sensitivity, % | 0.30 | 0.75 |
| Specificity, % | 0.79 | 0.15 |
| Positive predictive value, % | 0.60 | 0.32 |
| Negative predictive value, % | 0.52 | 0.75 |

CEA, carcinoembryonic antigen; pro-GRP, pro-gastrin-releasing peptide.

Table S2 Correlation of CRP, ANC, ALC, and NLR with granzyme B

| | r | P value |
|---------------|--------|---------|
| CRP | -0.584 | 0.563 |
| Procalcitonin | 0.321 | 0.181 |
| ANC | 0.193 | 0.240 |
| ALC | -0.041 | 0.805 |
| NLR | 0.117 | 0.479 |

CRP, C-reactive protein; ANC, absolute neutrophil count; ALC, absolute lymphocyte count; NLR, neutrophil-to-lymphocyte ratio.

Table S3 The importance of additional features with granzyme B by comparing AUC

| Feature | AUC |
|------------------|-------|
| Granzyme B alone | 0.949 |
| Pack-years | 0.975 |
| Smoking | 0.974 |
| Sex | 0.952 |
| Age | 0.949 |
| CRP | 0.959 |
| Procalcitonin | 0.949 |
| NLR | 0.956 |
| ANC | 0.935 |
| ALC | 0.948 |
| CEA | 0.954 |

CRP, C-reactive protein; NLR, neutrophil-to-lymphocyte ratio; ANC, absolute neutrophil count; ALC, absolute lymphocyte count; CEA, carcinoembryonic antigen; AUC, area under the curve.

Table S4 Baseline characteristics of propensity score matching (PSM) data

| | Healthy controls | Lung cancer | P value |
|-----------------------------|------------------|-------------|---------|
| No. of patients | 21 | 21 | |
| Mean age, y | 63.9 | 65.3 | 0.564 |
| Sex (male) | 20 (95.2%) | 20 (95.2%) | 1.000 |
| Currently smoking (current) | 7 (33.3%) | 8 (38.1%) | 1.000 |
| Mean pack-years | 27.2 | 38.0 | 0.063 |
| Stage | | | |
| 1 | | 3 (11.9%) | |
| 2 | | 4 (7.1%) | |
| 3 | | 3 (7.1%) | |
| 4 | | 6 (14.3%) | |

Table S5 Sensitivity, specificity, positive predictive value, negative predictive value after PSM

| | PSM data |
|------------------------------|----------|
| Sensitivity, % | 1.00 |
| Specificity, % | 0.91 |
| Positive predictive value, % | 0.95 |
| Negative predictive value, % | 1.00 |
| Cut-off | 19.8 |

Table S6 Comparison of performance evaluation of Random Forest by the number of granzyme B

| Metric | 1 Granzyme B | 3 Granzyme B |
|------------------------------|--------------|--------------|
| Sensitivity, % | 0.920 | 0.920 |
| Specificity, % | 0.913 | 0.922 |
| Positive predictive value, % | 0.909 | 0.918 |
| Negative predictive value, % | 0.922 | 0.926 |
| Accuracy, % | 0.914 | 0.920 |
| ROC AUC | 0.972 | 0.977 |

ROC, receiver operating characteristics; AUC, area under the curve.