**Supplementary Table S4.** Multivariate linear regression analysis of mutation count in patients stratified by smoking status on the base of validation set (GLCI).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Smoking status****Variables** |  | **Never smoking** |  | **Current/Ever smoking** |
|  | **regression coefficients** | **p** |  | **regression coefficients** | **p** |
| **Gender (male vs female)a** |  | 0.189 (-0.154~0.533) | 0.966 |  |  |  |
| **Age (≥65 vs <65)** |  | 0.007 (-0.328~0.342) | 0.267 |  | -0.029 (-0.296~0.238) | 0.827 |
| **EGFR (mutation vs WT)** |  | -0.033 (-0.387~0.321) | 0.849 |  | -0.357 (-0.627~-0.087) | **0.011\*** |
| **KRAS (mutation vs WT)** |  | 0.393 (0.043~0.743) | **0.029\*** |  | 0.031 (-0.235~0.298) | 0.814 |
| **TP53 (mutation vs WT)** |  | 0.345 (0.010~0.680) | **0.044\*** |  | 0.459 (0.187~0.731) | **0.001\*** |
| **STK11 (mutation vs WT)b** |  |  |  |  | -0.022 (-0.281~0.237) | 0.866 |

Dependent Variable: **mutation count**

a no female patient in current/ever smoking subgroup; b no patient with STK11 mutation in never smoking subgroup. **\*** indicated significant difference.