**Supplementary Table 1**: Demographics and Clinical Characteristics of the Lung Cancer Cohort

|  |  |
| --- | --- |
|  | **Microbiome Cohort** |
|  | **n = 83** |
| **Age** | 68.2 [47.8-74.5] |
| **Sex (Male)** | 54 (65%) |
| **Race** |  |
| Caucasian | 42 (51%) |
| African American | 19 (23%) |
| Hispanic | 6 (7%) |
| Asian | 9 (11%) |
| Other | 7 (8%) |
| **Smoking Status** |  |
| Former | 53 (64%) |
| Current | 22 (27%) |
| Never | 7 (8%) |
| Unknown | 1 (1%) |
| **Pack Years (avg)** | 35.5 [5.0-51.2] |
|  |  |
| **NSCLC** | 74 (89%) |
| ***Histology:*** |  |
| Adenocarcinoma | 48 (65%) |
| Squamous | 26 (35%) |
| **TNM Stage:** |  |
| I | 7 (10%) |
| II | 12 (16%) |
| IIIA | 18 (24%) |
| IIIB | 7 (10%) |
| IV | 29 (39%) |
| Unknown | 1 (1%) |
| **Treatment:** |  |
| Surgery | 20 (27%) |
| Chemotherapy | 38 (51%) |
| Radiation Therapy | 21 (28%) |
| Immunotherapy | 12 (16%) |
| No Treatment | 12 (16%) |
| No Follow up | 6 (8%) |
| **Median Survival (years)** | 2.1 |
| **Small Cell** | 9 (11%) |
| **Stage:** |  |
| Limited | 4 (44%) |
| Extensive | 5 (56%) |
| **Treatment:** |  |
| Chemotherapy | 7 (78%) |
| Radiation Therapy | 4 (44%) |
| No Treatment | 1 (11%) |
| No Follow up | 1 (11%) |
| **Median Survival (yrs)** | 0.8 |

Data expressed as Median [IQR] or counts (%) unless otherwise specified.

**Supplementary Table 2**: Clinical characteristics and overall survival of NSCLC patients (n=73)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Categories | Alive | Dead | Fisher’s | Cox PH model | |
|  |  |  |  | p value | HR | p value |
| **Smoking** | Current | 14 (0.34) | 7 (0.22) | 0.48 | 1 | 0.44 |
|  | Former | 23 (0.56) | 22 (0.69) |  | 1.68 |  |
|  | Never | 4 (0.10) | 3 (0.09) |  | 1.19 |  |
| **RACE** | Asian | 5 (0.12) | 4 (0.12) | 0.20 |  | 0.16 |
|  | Black (AA) | 8 (0.20) | 9 (0.28) |  | 1.59 |  |
|  | Hisp/Latino | 6 (0.15) | 0 (0) |  | 1.82E-08 |  |
|  | White | 18 (0.44) | 17 (0.53) |  | 1.48 |  |
|  | Other | 4 (0.10) | 2 (0.06) |  | 1.47 |  |
| **Treatment** |  |  |  |  |  |  |
| Chemo/ RT | Yes | 25 (0.61) | 18 (0.56) | 0.81 | 0.69 | 0.30 |
|  | No | 16 (0.39) | 14 (0.44) |  |  |  |
| Surgery | Yes | 18 (0.44) | 3 (0.09) | 0.0015 | 0.19 | 8.45E-04 |
|  | No | 23 (0.56) | 29 (0.91) |  |  |  |
| Immunotherapy | Yes | 11 (0.27) | 3 (0.09) | 0.076 | 0.39 | 0.08 |
|  | No | 30 (0.73) | 29 (0.91) |  |  |  |
| **Stage** |  |  |  |  |  |  |
| Advanced | Yes | 14 (0.34) | 25 (0.78) | 3.21E-04 | 4.36 | 1.44E-04 |
|  | No | 27 (0.66) | 7 (0.22) |  |  |  |

**Supplementary Table 3**: Top 20 OTUs associate with the stage (I-IIIA vs. IIIB-IV) ranked by absolute coefficient estimates with raw p value at 0.1 significance level.

|  |  |  |  |
| --- | --- | --- | --- |
| OTU | coef | 95% CI | p-val |
| Haemophilus\_865469 | 0.048 | (0.007, 0.091) | 0.026 |
| Fusobacterium\_938948 | 0.043 | (0.009, 0.077) | 0.016 |
| Gemellaceae(u.g.)\_1074210 | 0.038 | (0.007, 0.070) | 0.020 |
| Streptophyta(u.g.)\_769222 | 0.022 | (0.003, 0.041) | 0.027 |
| Capnocytophaga\_987790 | 0.020 | (0.006, 0.034) | 0.007 |
| Prevotella[Paraprevotellaceae]\_269907 | 0.017 | (-0.002, 0.035) | 0.078 |
| Granulicatella\_1059729 | 0.016 | (0.002, 0.030) | 0.032 |
| Pasteurellaceae(u.g.)579141 | 0.013 | (0.002, 0.025) | 0.029 |
| Streptophyta(u.g.)\_1787644 | 0.013 | (0.0001, 0.026) | 0.051 |
| Porphyromonas(u.g.)\_530164 | 0.011 | (-0.001, 0.022) | 0.070 |
| Flavobacterium\_960076 | -0.058 | (-0.127, 0.010) | 0.099 |
| Methylobacteriaceae(u.g.)\_785526 | -0.029 | (-0.051, -0.008) | 0.010 |
| Corynebacterium\_441265 | -0.024 | (-0.052, 0.004) | 0.091 |
| Prosthecobacter\_516569 | -0.021 | (-0.043, 0.0004) | 0.058 |
| Methylobacterium\_926160 | -0.020 | (-0.037, -0.004) | 0.016 |
| Cytophagaceae(u.g.)\_304640 | -0.018 | (-0.034, -0.002) | 0.031 |
| Prosthecobacter\_4468331 | -0.015 | (-0.028, -0.001) | 0.036 |
| Fluviicola\_650001 | -0.014 | (-0.026, -0.002) | 0.028 |
| Candidatus Rhodoluna\_810672 | -0.012 | (-0.025, 0.002) | 0.089 |
| Limnohabitans\_1108275 | -0.011 | (-0.024, 0.001) | 0.071 |

A positive coefficient indicates the advanced group having higher relative abundance after arcsine square root transformation. 95% CI represents the 95% confidence interval for estimated coefficients.

**Supplementary Table 4**: Associations between microbial profile and OS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| OTU | coef | HR | 95% CI | p-val |
| Prevotella\_530206 | 0.595 | 1.81 | (1.29, 2.54) | 0.0005 |
| Methylobacterium\_574655 | 0.485 | 1.62 | (1.16, 2.27) | 0.004 |
| Lactobacillus\_1107027 | 0.497 | 1.64 | (1.17, 2.31) | 0.004 |
| Lautropia\_1084417 | 0.433 | 1.54 | (1.13, 2.10) | 0.006 |
| Sphingomonadaceae(u.g.)\_820086 | 0.512 | 1.67 | (1.12, 2.49) | 0.012 |
| Proteobacteria(u.g.)\_52576 | 0.425 | 1.53 | (1.09, 2.15) | 0.014 |
| Streptococcus\_1083194 | 0.456 | 1.58 | (1.09, 2.28) | 0.015 |
| Gemellaceae\_1065569 | 0.446 | 1.56 | (1.08, 2.25) | 0.017 |
| Bacillus\_1051517 | 0.427 | 1.53 | (1.05, 2.24) | 0.027 |
| Streptococcus\_4297222 | 0.450 | 1.57 | (1.05, 2.34) | 0.027 |
| Alphaproteobacteria(u.g.)\_929489 | 0.423 | 1.53 | (1.04, 2.25) | 0.032 |
| Kaistobacter\_790545 | 0.452 | 1.57 | (1.03, 2.39) | 0.035 |
| Caulobacteraceae(u.g.)\_310003 | 0.416 | 1.52 | (1.02, 2.26) | 0.040 |
| Sphingobacteriales(u.g.)\_1114591 | -0.976 | 0.38 | (0.15, 0.95) | 0.038 |
| Armatimonas\_141815 | -0.720 | 0.49 | (0.24, 1.01) | 0.052 |
| Sphingobacteriales(u.g.)\_40351 | -0.686 | 0.50 | (0.24, 1.04) | 0.064 |
| Neisseriaceae(u.g.)\_933546 | -0.811 | 0.44 | (0.19, 1.06) | 0.067 |
| Sphingomonadaceae(u.g.)\_813418 | -0.696 | 0.50 | (0.23, 1.06) | 0.072 |
| Neisseriaceae(u.g.)\_3393186 | -1.025 | 0.36 | (0.12, 1.10) | 0.072 |
| Coxiellaceae(u.g.)\_334554 | -0.750 | 0.47 | (0.19, 1.15) | 0.098 |

A positive coefficient indicates a worse prognosis and a negative coefficient indicates a protective effect. In addition, HR is the estimated hazard ratio and 95% CI is 95% confidence interval for HR.

**Supplementary Table 5**: Comparisons of Demographics and Clinical Characteristics base on lung microbiome sub-type- Background Pneumontype (BPT) vs. Supraglottic Pneumontype (SPT)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **BPT**  (n=37) | **SPT** (n=37) | p-value |
| **Age** | 70 [64-77] | 65 [59-73] | NS |
| **Sex (Male)** | 24 (65%) | 26 (79%) | NS |
| **Race** |  |  |  |
| Caucasian | 20 (54%) | 16 (43%) | NS |
| African American | 8 (22%) | 8 (22%) | NS |
| Hispanic | 3 (8%) | 3 (8%) | NS |
| Asian | 4 (11%) | 5 (14%) | NS |
| Other | 2 (5%) | 5 (14%) | NS |
| **BMI** | 25.9 [21.5-28.6] | 23.7 [20.9-27.5] | NS |
| **Smoking Status** |  |  | NS |
| Former | 20 (54%) | 27 (73%) |  |
| Current | 10 (27%) | 9 (24%) |  |
| Never | 7 (19%) | 1 (3%) |  |
| **Pack Years** | 48.9 [21.6-51.5] | 39.8 [25.0-50.0] | NS |
| **Medical** |  |  |  |
| DM | 3 (8%) | 5 (14%) | NS |
| GERD | 8 (22%) | 9 (24%) | NS |
| COPD | 10 (27%) | 10 (27%) | NS |
| OSA | 0 | 3 (8%) | NS |
| ETOH | 15 (38%) | 9 (24%) | NS |
| Neurological Disorder | 2 (5%) | 2 (5%) | NS |
| Oral Disorder | 2 (5%) | 2 (5%) | NS |
| **NSCLC** |  |  |  |
| Adenocarcinoma | 23 (62%) | 25 (68%) | NS |
| Squamous | 14 (38%) | 12 (32%) | NS |
| TNM Stage: |  |  |  |
| I | 5 (14%) | 2 (5%) | NS |
| II | 6 (16%) | 6 (16%) | NS |
| IIIA | 12 (32%) | 6 (16%) | NS |
| IIIB | 3 (8%) | 4 (11%) | NS |
| IV | 11 (30%) | 18 (49%) | \* |
| Unknown | 0 | 1 (3%) | NS |

Data expressed as Median [IQR] or counts (%) unless otherwise specified.

**Supplementary Table 6:** Lung Cancer subjects and number/type of airway samples used for microbiome and transcriptome

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Subjects** | **16s (samples)** | **RNASeq (samples)** |
| **Buccal** | 83 | 54 | - |
| **Background** | 83 | 51 | - |
| **Airway brushes** | 83 | 210 | 84 |
| **Histology** |  |  |  |
| NSCLC | 74 | 186 | 70 |
| Small cell | 9 | 24 | 12 |

**Supplementary Table 7:** Antibodies for multiplex immunostaining.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Antigen** | **Company** | **Host** | **Clone** | **Cat #** | **Dilution** | **Fluorophore** | **Retrieval** |
| CC10 | Abcam | Rabbit | Poly | AB40873 | 3000 | 690 | No retrieval |
| Ly6g | BD | Rat | 1A8 (RUO) | 551459 | 100 | 570 | ER1 |
| CD4 | CST | Rabbit | D7D2Z | 25229S | 500 | 620 | ER2 |
| CD8 | CST | Rabbit | D4W2Z | 98941 | 200 | 520 | ER2 |
| F4/80 | CST | Rabbit | D2S9R | 70076S | 500 | 480 | ER2 |