Clinical Bronchoscopy (03/13-10/18):
N = 3573

Recruited for Research:
n = 148

Excluded:
- n = 15, metastatic cancer
- n = 12, benign nodule
- n = 38, non-cancer diagnosis

Primary Lung Cancer Cohort:
n = 83

- Transcriptome data: n = 70
- RECIST score: n = 40
- 6 month follow up: n = 75
- 12 month follow up: n = 64
Supp. Figure 1. Consortium diagram.
Supp Figure 2

16S rRNA gene Copies per μL

BKG  Small Cell  NSCLC  Buccal

p=ns
Supp. Figure 2. Bacterial load. Droplet digital quantitative PCR targeting the 16S rRNA gene was performed on background, buccal and lower airway samples in lung cancer.
Supp Figure 3

(a) Comparison of Shannon Index between BKG, Buccal, NSCLC, and Small cell samples. There is a significant difference between NSCLC and Small cell samples (p < 0.001). No significant difference was observed between NSCLC vs. Small cell (p = ns).

(b) Comparison of Shannon Index among different stages: Stage I, Stage II, Stage IIIA, Stage IIIB, and Stage IV. No significant difference was observed among the stages (p = ns).

(c) Comparison of Shannon Index between Stage I-III and Stage III-IV. No significant difference was observed (p = ns).
Supp. Figure 3. Alpha diversity measurement. a. Comparison based on Shannon Index between different sample types and between NSCLC vs. small cell lung cancer. b,c. Comparisons between different stages (p=ns for all comparisons).
Supp Figure 4

PC1: 36% variance

PC2: 14% variance

Stage.I
Stage.II
Stage.IIIA
Stage.IIIB
Stage.IV

Bray Curtis Distance to UA

Less Similar
More Similar

p=0.047

Supp Figure 4
Supp. Figure 4. Differences in global microbial community parameters among TNM stages in NSCLC. 

a. β-diversity comparisons (Bray Curtis) showed differences based on lung cancer stages (PERMANOVA p=0.047). 

b. Bray Curtis Dissimilarity index between lower airway and buccal samples shows greater degree of similarity among higher TNM stages (*p<0.05)
Supp Figure 5

![Box plot showing Bray-Curtis distance to UA against PD-L1 %](image)

- **Less Similar**
- **More Similar**

**Y-axis:** Bray-Curtis Distance to UA

**X-axis:** PD-L1 %

Legend:
- *: Significant difference

Data points are distributed across different PD-L1 % ranges (0-1-79, 80-100).
Supp. Figure 5. Relationship between PD-L1 tumor expression and degree of similarity between lower airway and buccal samples. Bray Curtis Dissimilarity index between lower airway and buccal samples shows greater degree of similarity for samples with higher PD-L1 immunohistochemistry (≥80%) as compared with lower PD-L1 groups (* = p<0.05)
Supp Figure 6

6-Months Survival

Stage I-IIIA

Shannon Index

Stage IIIB-IV

Shannon Index

1-Year Survival

> 6 Months  < 6 Months

> 1 Year  < 1 Year

p = 0.02

p = 0.017

p = ns

p = ns

Supp Figure 6
Supp. Figure 6. Alpha diversity measurement among Stage I-III A and Stage III B-IV NSCLC comparing survival. Shannon index showed that poor prognosis at 6-months (but not at 1 year) was associated with diversity in both Stage I-III A and Stage III B-IV NSCLC groups (p=0.02 and 0.017, respectively).
Supp Figure 7
Supp. Figure 7. Multi-variate analysis of microbial composition (diversity) and prognosis adjusted by TNM stage. Analysis show diversity distribution of samples based on Bray Curtis Dissimilarity index color coded by TNM stage. PERMANOVA test evaluating for associations between microbial community and prognosis at 6 month (top panel) and 1 year (bottom panel) adjusted by TNM stage.
Supp. Figure 8. Taxonomic differences between lung cancer types and stage. a. Volcano plot based on DESseq analysis shows taxonomic differences in lung microbiota between small cell lung cancer (SC) and non-small cell lung cancer (NSCLC). b. Volcano plot based on DESseq analysis shows taxonomic differences in lung microbiota between stage I-IIIA NSCLC and stage IIIB-IV NSCLC.
Supp Figure 9

(a) stage I-IIIA

6-Months Survival

Fold-change (log2)

>6 month survival <6 month survival

(b) stage IIIB-IV

6-Months Survival

Fold-change (log2)

>6 month survival <6 month survival

(c) 1-Year Survival

Fold-change (log2)

>1 year survival <1 year survival

(d) 1-Year Survival

Fold-change (log2)

>1 year survival <1 year survival

Relative Abundance

- 0.01
- 0.05
- 0.10
- 0.15
Supp. Figure 9. Taxonomic differences between different mortality among NSCLC. a and b. Volcano plot shows statistically significant differences at 6 months mortality for stage I-III A and IIIB-IV NSCLC. c and d. Volcano plot shows statistically significant differences at 1 year mortality for stage I-III A and IIIB-IV NSCLC.
Supp Figure 10

6-Month Mortality

NMDS1: 65% variance

NMDS2: 35% variance

β-Diversity: $R^2 = 0.034$, $p=0.002$
TNM Stage: $R^2 = 0.087$, $p=0.002$

1-Year Mortality

NMDS1: 65% variance

NMDS2: 35% variance

β-Diversity: $R^2 = 0.035$, $p=0.013$
TNM Stage: $R^2 = 0.097$, $p=0.003$
Supp. Figure 10. Biplot analysis of taxa and samples among Stage I-IIIA and Stage IIIB-IV NSCLC comparing survival. Biplot shows the distribution of all lower airway samples and taxa across the NMDS plot, and their association with prognosis at 6 months (Top panel) and 1 year (Bottom panel). Multi-variate PERMANOVA adjusting for TNM stage.
Supp Figure 11

6-Month Mortality

NMDS1: 59% variance

NMDS2: 41% variance

β-Diversity: R^2 = 0.034, p = 0.002
TNM Stage: R^2 = 0.087, p = 0.369

1-Year Mortality

NMDS1: 61% variance

NMDS2: 39% variance
Supp. Figure 11. Biplot analysis of taxa and single sample most proximal to tumor lesion among Stage I-IIIA and Stage IIIB-IV NSCLC comparing survival. Biplot shows the distribution of lower airway sample most proximal to tumor lesion and taxa across the NMDS plot, and their association with prognosis at 6 months (Top panel) and 1 year (Bottom panel). Multivariate PERMANOVA adjusting for TNM stage.
Supp. Figure 12. Airway microbiota clustering. 

a. Dirichlet Multinomial Model (DMM) of samples showed that 2 clusters have best fitness for the data. 

b. PCoA analysis based on Bray Curtis Dissimilarity Index and sample colored label based on DMM showed that all buccal samples clustered together in addition to a subset of lower airway samples (SPT: supraglottic predominant taxa) while all background samples also clustered together with a subset of lower airway samples (BPT: background predominant taxa). 

c. Volcano plot based on DESeq analysis shows taxonomic differences of lower airway microbiota based on DMM clusters comparing SPT with BPT.
Supp. Figure 13. Identification of lower airway contaminants and non-contaminants.

Contaminants in lower airway samples were identified using decontam. XY plot shows the prevalence of OTUs identified as lower airway OTUs in airway brushings (true sample) vs. prevalence of OTUs identified as background contamination coming from bronchoscopic controls (negative control). Node size is proportional to relative abundance, green nodes identify background contaminants while red nodes identify true lower airway OTUs.
Stage IIIB-IV

Supp Figure 14

Enriched in Disease Regression

Enriched in Disease Progression

Relative Abundance

0.01

0.05

0.15
**Supp. Figure 14. Taxonomic differences between tumor progression and disease regression.** Volcano plot based on DESseq analysis shows taxonomic differences in lower airway samples of stage IIIB-IV NSCLC between those that had a negative delta RECIST score (tumor progression) vs. those that had a positive RECIST score (disease regression). This data showed that lower airway samples from subjects that had tumor progression were enriched with *Veillonella*, *Streptococcus*, *Prevotella*, and *Rothia* when compared with lower airway samples from patients with tumor regression.

*RECIST Tumor progression = Progressive Disease and Stable Disease*

*RECIST Disease regression = Complete Response or Partial Response*
Stage I-III A vs. III B-IV

Supp Figure 15
Supp. Figure 15 Transcriptomic differences based on stage of NSCLC. Volcano plot comparing gene expression profiles of lower airway samples from stage I-IIIA vs. stage IIIB-IV NSCLC.
**Supp. Figure 16. Immune response post lower airway dysbiosis with oral commensals.**

Measurement of Th1, Th17, Tregs, CD4^+PD1^+, and CD8^+PD1^+ in mice lung 14 days after induced dysbiosis by three oral commensals found to be associated with Pneumotype_{SPT} (Streptococcus mitis, Prevotella melaninogenica and Veillonella parvula). Data show that *V. parvula* induced the highest immune response when compared with the other two taxa.
Supp Figure 17

(a) Diagram showing the comparison of CD4:CD8 Ratio, CD4+ activation, CD8+ activation, and Th1 between Single and Multiple Aspiration.

(b) Diagram showing Veillonella concentration and its effect on CD4/CD8 ratio, IL17, and RORγt.
Supp. Figure 17. Dose-dependent immune response to microbial aspiration challenge. a. Multiple (2 times per week for 2 weeks) challenges with *V. parvula* leads to higher immune response, including Th17 and checkpoint inhibitor markers in T cells, compared to single dose challenge. b. Higher concentration (4.6e8 cfu/ml per mouse) of Veillonella challenge leads to higher CD4/CD8 and higher levels of Th17 cells compared to median concentration (1.5e8 cfu/ml per mouse).
Supp Figure 18

(a) Weight loss (%)

-40 -30 -20 -10 0 10 20

Lung Cancer Dysbiosis
+ +

Dysbiosis
+ +

(b) Weight Loss (%)

-40 -30 -20 -10 0 10 20

Dysbiosis
+ +

Anti-IL17
Supp. Figure 18. Weight measurements. **a.** Quantitative data of weight loss (calculated as the percent change of weight prior to death or sacrifice compared to baseline weight) showing that LC + Dys mice had increased weight loss (p<0.05). **b.** Quantitative data of weight loss (calculated as the percent change of weight prior to death or sacrifice compared to baseline weight) showing no difference in weight change after anti-IL-17 treatment as compared with isotype Ab (p=ns, n=8-10 mice in each group).
Supp. Figure 19. Transcriptomic data from lung cancer and dysbiosis in mice. a. Principal coordinate analysis of host transcriptomic data (RNA Seq) on lung homogeneates shows significant compositional differences based on experimental conditions. b. Immune cell annotation of transcriptomic data using CIBERSORT showed significant unsupervised hierarchical clustering between conditions.
Supplementary Figure 20

Mouse 2710

Human 127

- Notch
- Neurotrophin
- GnRH
- VEGF
- PI3K
- Akt
- p53
- Phospholipase D
- Jak–STAT
- Fanconi anemia
- RIG–I–like receptor
- C–type lectin receptor
- NOD–like receptor
- TNF–PD–L1
- T cell receptor
- B cell receptor
- Cytosolic DNA–sensing
- NF–kappa B
- Toll–like receptor
- Chemokine
- NF–kappa B
- Cytosolic DNA–sensing
- B cell receptor
- T cell receptor
- PD–L1
- TNF
- NOD–like receptor
- C–type lectin receptor
- RIG–I–like receptor
- Fanconi anemia
- Jak–STAT
- Phospholipase D
- p53
- PI3K–Akt
- VEGF
- GnRH
- Neurotrophin
- Notch
- Adipocytokine
- Toll and Imd
- Rap1
- mTOR
- ErbB
- MAPK
- Thyroid hormone
- Relaxin
- Insulin
- Wnt
- Estrogen
- Oxytocin
- Stem Cell Pluri potency
- TGF–beta
- cAMP
- Apelin
- Hippo
- cGMP–PKG
- Longevity regulating
- Glucagon
- FoxO
- AMPK
- AGE–RAG (DM)
- Fc epsilon RI
- PPAR
- Calcium
- HIF–1
- Ras

Median Log Fold Change

Human
Mouse

Supplmentary Figure 20
Supp. Figure 20. Transcriptomic signatures associated with lower airway dysbiosis in human and mouse. a. GSEA analysis comparing transcriptomic signatures found in lower airway samples of SPT vs. BPT and those found in lung samples of cancer + dysbiosis vs. cancer control. In the mouse data, 2710 differentially expressed genes were found while in the human data 127 differentially expressed genes were found, with 75 genes that overlapped (GSEA FDR<0.25). b. Comparison between summarized transcriptomic signatures found with dysbiosis in human and mouse data showed concordant signatures in large amount of transcripts such as those involved in IL-17, cytokine, Toll-like receptor, and PD-L1 pathways.
All Cytokine lung homogenate data

Supp Figure 21
Supp. Figure 21. Cytokine/chemokine profile of mice lung homogeneate. Heatmap with clustering (based on Bray Curtis Dissimilarity Index) shows clear clustering between the experimental conditions with cleft marking dysbiotic exposure as the predominant driver for the clustering. Wild-type (WT), lung dysbiosis (Dys), lung cancer (LC), and lung cancer + lung dysbiosis (LC/Dys)
Supp Figure 22

**a**

LC

![Image](LC.png)

LC+Dys

![Image](LC_Dys.png)

**b**

Iso

![Image](Iso.png)

anti-IL17

![Image](anti_IL17.png)
Supp. Figure 22. Immunohistochemistry of lung cancer and surrounding spared lung tissue in KP dysbiotic preclinical model. Representative lung tissue images. **a.** Hematoxylin and eosin staining paired with immunohistochemistry for the lung cancer dysbiosis model. **b.** Hematoxylin and eosin staining paired with immunohistochemistry for IL-17 blockade of lung cancer dysbiosis model. **SL**= Spared lung, non-tumor. **T**= tumor.
**Supp. Figure 23. Taxonomic composition of background controls.** Unsupervised Hierarchical clustering based on Bray Curtis of taxonomic data (Top 50 genera) across background samples including bronchoscope and PBS used during preclinical experiments.