TITLE:

Lower airway dysbiosis affects lung cancer progression

Jun-Chieh J. Tsay[1,2], Benjamin G. Wu[1], Imran Sulaiman[1], Katherine Gershner[3], Rosemary Schluger[1], Yonghua Li[1], Ting-An Yie[1], Peter Meyn[4], Evan Olsen[1], Luisannay Perez[1], Brendan Franca[1], Joseph Carpenito[1], Tadasu Iizumi[1] Mariam El-Ashmawy[5], Michelle Badri[7], James T. Morton[6], Nan Shen[8], Linchen He[9], Gaetane Michaud[1], Samaan Rafeq[1], Jamie L. Bessich[1], Robert L. Smith[2], Harald Sauthoff[2], Kevin Felner[2], Ray Pillai[1], Anastasia-Maria Zavitsanou[10], Sergei B. Koralov[10], Valeria Mezzano[10], Cynthia A. Loomis[10], Andre L. Moreira[10], William Moore[11], Aristotelis Tsirigos[10], Adriana Heguy[4,10], William N. Rom[1], Daniel H. Sterman[1], Harvey I. Pass[13], Jose C. Clemente[8], Huilin Li[9], Richard Bonneau[6,7,12], Kwok-kin Wong[14], Thales Papagiannakopoulos[10], and Leopoldo N. Segal[1]*

[1]Division of Pulmonary and Critical Care Medicine, New York University School of Medicine, NY
[2]Division of Pulmonary and Critical Care Medicine, VA New York Harbor Healthcare System, NY
[3]Section of Pulmonary, Critical Care, Allergy and Immunology, Wake Forest School of Medicine, NC
[4]NYU Langone Genomic Technology Center, New York University School of Medicine, NY
[5]Department of Medicine, New York University School of Medicine, NY
[6]Flatiron Institute, Center for Computational Biology, Simons Foundation, NY
[7]Department of Biology, New York University, NY
[8]Department of Genetics and Genomic Sciences and Immunology Institute, Icahn School of Medicine at Mount Sinai, NY.
[9]Department of Population Health, New York University School of Medicine, NY
[10]Department of Pathology, New York University School of Medicine, NY
[11]Department of Radiology, New York University School of Medicine, NY
[12]Center for Data Science, New York University School of Medicine, NY
[13]Department of Cardiothoracic Surgery, New York University School of Medicine, NY
[14]Division of Hematology and Oncology, New York University School of Medicine, NY

Jun-Chieh J. Tsay, MD	jun-chieh.tsay@nyumc.org
Benjamin G. Wu, MD	Benjamin.Wu@nyumc.org
Imran Sulaiman, MD, PhD	SheikMohammadImran.Sulaiman@nyumc.org
Katherine Gershner, DO	k.gershner@wakehealth.edu
Rosemary Schluger	rosemary.schluger@nyumc.org
Yonghua Li, MD, PhD	Yonghua.Li@nyumc.org
Ting-An Yie	ting-an.yie@nyumc.org
Peter Meyn	peter.meyn@nyumc.org
Evan Olsen	evan.olsen@nyumc.org
Luisanny Perez	luisanny.perez@nyumc.org
Brendan Franca	Brendan.franca@nyumc.org
Joseph Carpenito	Joseph.carpenito@nyumc.org
Tadasu Iizumi MD, PhD	Tadasu.Iizumi@nyumc.org
Mariam El-ashmawy MD, PhD	Mariam.el-ashmawy@nyumc.org
Michelle H. Badri	mhb383@nyu.edu
James T. Morton, PhD	jmorton@flatironinstitute.org
Nan Shen	nan.shen@icahn.mssm.edu
Linchen He	lh1790@nyu.edu
Gaetane Michaud, MD	gaetane.michaud@nyumc.org
Samaan Rafeq, MD	Samaan.Rafeq@nyulangone.org
Jamie Bessich, MD	Jamie.Bessich@nyulangone.org

Robert L. Smith, MD robert.smith4@va.gov
Harald Sauthoff, MD harald.sauthoff@va.gov
Kevin Felner, MD kevin.felner@va.gov
Ray Pillai, MD ray.pillai@nyulangone.org
Anastasia-Maria Zavitsanou amz354@nyu.edu
Sergei Koralov, PhD Sergei.koralov@nyulangone.org
Valeria Mezzano, MD PhD valeria.mezzanorobinson@nyulangone.org
Cynthia A. Loomis MD PhD cindy.loomis@nyulangone.org
Andre L. Moreira, MD Andre.Moreira@nyulangone.org
William Moore, MD William.moore@nyulangone.org
Aristotelis Tsirigos, PhD Aristotelis.Tsirigos@nyulangone.org
Adriana Heguy, PhD Adriana.heguy@nyumc.org
William N. Rom, MD William.rom@nyumc.org
Daniel H. Sterman, MD Daniel.sterman@nyumc.org
Harvey Pass, MD Harvey.pass@nyumc.org
Jose C. Clemente, PhD jose.clemente@mssm.edu
Huilin Li, PhD Huilin.Li@nyulangone.org
Richard Bonneau, PhD rb133@nyu.edu
Kwok-Kin Wong, MD, PhD Kwok-Kin.Wong@nyulangone.org
Thales Papagiannakopoulos, PhD Thales.Papagiannakopoulos@nyulangone.org
Leopoldo N. Segal, MD[1] Leopoldo.Segal@nyumc.org

*Corresponding Author/ Address for Reprints:
Leopoldo N. Segal, MD[1] Leopoldo.Segal@nyumc.org

NYU School of Medicine
462 First Ave 7N21
New York, NY 10016
Tel: (212) 263-6479
Fax: (212) 263-8441

Authors Contributions:

JC.J.T. and L.N.S. conceived of and designed the study. Data was obtained by JC.J.T., B.G.W, I.S., K.G., R.S., Y.L., T.A.Y., P.M., E.O., L.P., B.F., J.C., T.I., M.E., M.H.B, J.M, N.S., L.H, W.M. J.C.C., H.L, R.B, R.P., A.Z., V. M., led by L.N.S. Data were analyzed by JC.J.T., B.G.W, I.S., K.G., M.H.B, J.M, N.S., L.H, W.M. J.C.C., H.L, R.B., R.P., S.B.K., C.A.L led by L.N.S. The first draft of the manuscript was written by JC.J.T. and L.N.S. All authors read, critically revised and approved the final manuscript.

Research support funding:

R37 CA244775 (LNS, NIH/NCI); PACT grant (LNS, FNIH); K23 AI102970 (LNS, NIH/NIAD); EDRN 5U01CA086137-13 (WNR); DoD W81XWH-16-1-0324 (JJT); Research supported by the 2018 AACR-Johnson & Johnson Lung Cancer Innovation Science Grant Number 18-90-52-ZHAN (HP/LNS); A Breath of Hope Foundation (JJT), Simons Foundation (RB); CTSI Grant #UL1 TR000038 (LNS); The Genome Technology Center is partially supported by the Cancer Center Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center (AH, AT); T32 CA193111 (BGW); UL1TR001445 (BGW); FAMRI Young Clinical Scientist Award (BGW), Stony Wold-Herbert Fund Grant-in-Aid/Fellowship (BGW, IS, & KG), R01 HL125816 (LNS, SBK, NIH/NHLBI); R01 DK110014 (HL, LH).

Acknowledgement:

We would like to thank the Genome Technology Center (GTC) for expert library preparation and sequencing, and the Applied Bioinformatics Laboratories (ABL) for providing bioinformatics support and helping with the analysis and interpretation of the data. Experimental Pathology Research Laboratory for histopathology services and imaging. GTC and ABL are shared resources partially supported by the Cancer Center Support Grant P30CA016087 at the Laura and Isaac Perlmutter Cancer Center. This work has used computing resources at the NYU School of Medicine High Performance Computing Facility (HPCF). Financial support for the PACT project is possible through funding support provided to the FNIH by: AbbVie Inc., Amgen Inc., Boehringer-Ingelheim Pharma GmbH & Co. KG, Bristol-Myers Squibb, Celgene Corporation, Genentech Inc., Gilead, GlaxoSmithKline plc, Janssen Pharmaceutical Companies of Johnson & Johnson, Novartis Institutes for Biomedical Research, Pfizer Inc., and Sanofi.

Running Head: Lung Microbiome and Lung Cancer Prognosis

Word Count: 4869
Body of the manuscript:

Financial Disclosure: None
Key words: microbiome, bronchoscopy, lung cancer

Abstract

Word Count: 146

Abstract:

In lung cancer, enrichment of the lower airway microbiota with oral commensals commonly occurs and ex vivo models support that some of these bacteria can trigger host transcriptomic signatures associated with carcinogenesis. Here, we show that this lower airway dysbiotic signature was more prevalent in group IIIB-IV TNM stage lung cancer and is associated with poor prognosis, as shown by decreased survival among subjects with early stage disease (I-IIIA) and worse tumor progression as measured by RECIST scores among subjects with IIIB-IV stage disease. In addition, this lower airway microbiota signature was associated with upregulation of IL-17, PI3K, MAPK and ERK pathways in airway transcriptome, and we identified *Veillonella parvula* as the most abundant taxon driving this association. In a KP lung cancer model, lower airway dysbiosis with *V. parvula* led to decreased survival, increased tumor burden, IL-17 inflammatory phenotype and activation of checkpoint inhibitor markers.

Statement of Significance (50 word limit)

Multiple lines of investigations have shown that the gut microbiota affects host immune response to immunotherapy in cancer. Here we support that the local airway microbiota modulates the host immune tone in lung cancer affecting tumor progression and prognosis.

Introduction

Lung cancer has remained the leading cause of cancer deaths worldwide. In this past year alone, lung cancer occurred in approximately 2.1 million patients and was responsible for 1.7 million deaths(1). Targeting certain somatic mutations has improved survival but this is only applicable to ~30% of subjects with lung adenocarcinoma(2,3). More recently, immunotherapy that targets inhibitory checkpoint molecules, such as programmed death 1 (PD-1), has been shown to affect the responses of T-cells to neoantigens and improve survival in lung cancer(4-8). However, 40-60% of patients will not respond to or will develop resistance to immunotherapy(7). Recent investigations have identified gut microbiota signatures that are associated with augmenting anti-tumor immunity and responding to PD-1 blockade in murine models and in prospective analyses of immunotherapy-responsive cancer cohorts(9-11). For example, modulation of the microbiota in germ-free mice can enhance anti-tumor immunity and augment effects of checkpoint blockade(12,13). Matson *et al.* found that in patients with melanoma, anti-PD-1 treatment responders had a higher abundance of *B. longum, C. aerofaciens, and E. faecium* compared to non-responders(11). Gopalakrishnan *et al*. demonstrated that patients with higher bacterial diversity and increased relative abundance of *Ruminococcaceae* in the gut had enhanced systemic and anti-tumor immune responses(10). Routy *et al.* identified that the relative abundance of *A. muciniphila* was associated with a favorable clinical response to immunotherapy(9). While most investigations have focused on the gut microbiome, no human studies have studied the lower airway microbiota and lung cancer prognosis despite growing evidence supporting the role of the lung microbiota in lower airway inflammation(14-16).

Our understanding of the role of lung microbiota in health and disease is rapidly evolving with evidence that some phenotypic characteristics of the local lung immune tone appears to be more closely correlated to the lung microbiome than to the gut microbiome(14). Culture-independent techniques show that the lower airways of normal individuals commonly harbor oral

bacteria such as *Prevotella* and *Veillonella*(15,17-19). Our group has described that lower airway dysbiosis characterized by enrichment with oral commensals is associated with increased host inflammatory tone in the lung of healthy individuals(15,19). This same lower airway dysbiotic signature was found to differentiate between subjects with lung cancer and subjects with benign lung nodules(16). Importantly, we have shown in humans and in *ex vivo* experimental models that this dysbiotic signature likely triggers transcriptomic signatures (PI3K and MAPK) previously described in non-small cell lung cancer (NSCLC)(16,20), including the p53 mutation pathway(21). In order to explore the clinical implications of the lower airway microbiota in lung cancer, we utilized a prospective human cohort and a preclinical model to identify lower airway dysbiotic signatures that may affect the prognosis in this disease.

Results

Lung Cancer Cohort

Between March 2013 and October 2018, we recruited 148 subjects with lung nodules from the NYU Lung Cancer Biomarker Center who underwent clinical bronchoscopy for diagnostic purposes in whom lower airway brushes were obtained for research (**Supplementary Figure 1**). Fifteen subjects had non-lung primary tumors (metastasis), 12 had benign lung nodules and 38 subjects had other non-malignant diagnosis and were excluded. The remaining 83 subjects had a final diagnosis of lung cancer and were included for this project. Among these subjects, all had microbiome 16S rRNA gene sequencing data, 70/83 had transcriptomic data, and 75/83 had greater than six months of follow-up clinical data. **Supplementary Table 1** describes the demographics and clinical characteristics of this cohort: 91% were current or former smokers with a mean history of 46 pack-years. Eighty-nine percent had a diagnosis of NSCLC, of which 65% had adenocarcinoma and 49% was found to have stage IIIB-IV. The median survival was 2.1 years; 54% received chemotherapy, 30% received radiation therapy, 24% received surgery, and 14% received immunotherapy. All bio-specimens were obtained prior to treatment. Using the Cox Proportional Hazards model we determined that surgical treatment and stage IIIB-IV were significantly associated with overall survival (**Supplementary Table 2**).

Microbiomic signatures associated with stage and prognosis

In addition to lower airway brushings, we obtained buccal brushes and bronchoscope background control samples that were included in the 16S rRNA gene sequencing analysis. As compared with background controls, the bacterial load were ~10 times higher in lower airway brushing samples and ~10,000 times higher in the upper airways (buccal) ($p<0.001$, **Supplementary Figure 2**). Alpha diversity based on the Shannon Index showed greater diversity among lower airway samples than upper airway and background control samples ($p<0.001$, **Supplementary Figure 3a**). Principal coordinate analysis (PCoA) based on Bray Curtis Dissimilarity Index showed significant compositional differences across sample types

(**Figure 1a**, PERMANOVA p<0.001). Across lower airway samples, there were also compositional differences between small cell lung cancer and NSCLC (PERMANOVA p=0.01). Among NSCLC samples, there were no statistically significant differences in α- diversity and β-diversity between squamous cell carcinoma and adenocarcinoma.

We then evaluated microbial differences in lower airway samples based on the clinical NSCLC stage, grouped as I-IIIA and IIIB-IV of TNM classification. The selection of this cut point for TNM classification allowed for dichotomized analyses and we support based on prior prognosis/survival data and cancer management guidelines related to surgical management of NSCLC (22-25). Alpha diversity was similar across staging groups of NSCLC (**Supplementary Figure 3b** for comparison across individual stages and **Supplementary Figure 3c** for two-group comparisons of stages IIIB-IV vs. stage I-IIIA). Compositional differences between the I-IIIA vs. IIIB-IV groups lung cancer were noted based on β-diversity analysis (**Figure 1b**, left panel, p=0.005), where stage IIIB-IV lung cancer was compositionally more similar to buccal samples than I-IIIA stage lung cancer samples (**Figure 1b,** right panel). Compositional difference comparing all individual stages (I-IV) were also noted base on β-diversity analysis (**Supplementary Figure 4a**, p=0.047), where lower airway samples from more advanced stages had a greater similarity to buccal samples than lower airway samples from earlier stage subjects (**Supplementary Figure 4b**). MiRKAT analysis showed that differences in the microbial community profiles noted between stage I-IIIA and IIIB-IV NSCLC were not due to differences in location of the samples. Interestingly, sub-analysis on patient samples where tumor PD-L1 expression was available (n=39) shows that subjects with high PD-L1 expression (≥80%, n=12) had a lower airway microbiota with greater similarity to upper airway microbiota vs. the disease of similarity found among patients with lower tumor PD-L1 expression (0%, n=16 and 1-79%, n= 11), (p<0.05, **Supplementary Figure 5**).

Compositional differences based on 6-months and 1-year survival were also identified in □□

diversity analysis (**Figures 1c,** left panel) where samples from subjects with decreased survival were associated with greater compositional similarity to buccal samples than samples from subjects with better outcomes (**Figures 1c,** right panel). Shannon index showed decreased □□ diversity among samples from subjects with <6 months survival in both stage I-IIIA and IIIB-IV but this difference was not statistically significant at 1 year (**Supplementary Figure 6**). Multi-variate PERMANOVA analysis demonstrate that the association between microbial community composition and 6-month/1year mortality was independent of TNM staging (**Supplementary Figure 7**). No statistically significant differences were noted in □ or □ diversity analyses of buccal microbiota between subjects with different stages or mortality.

DESeq analyses was then performed to evaluate for taxonomic differential enrichment between SC vs. NSCLC and between the I-IIIA vs. IIIB-IV groups of NSCLC (**Supplementary Figure 8a**). Importantly, lower airway samples from patients in IIIB-IV stage group were enriched with many Operational Taxonomic Units (OTUs), which annotated to the genera *Moraxella*, *Fusobacterium*, *Pseudomonas*, and *Haemophilus*, and were decreased in abundance of *Actinomycetales* (**Supplementary Figure 8b**, **Supplementary File Table 1**). Using a mixed effect model that adjust for sample location, we report the top 20 OTUs ranked by their absolute coefficients estimates as having a differential abundance between the I-IIIA vs. IIIB-IV stage groups (**Supplementary Table 3**). Once again, stage IIIB-IV lung cancer was enriched with OTUs recognized as oral commensals, such as *Haemophilus*, *Fusobacterium*, *Gemella*, *Prevotella*, and *Granulicatella.*

Among stage I-IIIA and IIIB-IV subgroups, multiple OTUs were differentially enriched when worse vs. better survival groups were compared (both at 6 and 12 months). Several of the OTUs annotated to the genera *Veillonella, Prevotella*, and *Streptococcus* were found to be enriched in samples from subjects with worse prognosis (**Supplementary Figure 9a-d, Supplementary File Table 2-5**). In order to further explore for taxonomic associations with mortality while considering TNM staging we constructed Beta diversity biplots that allows for co-

location of lower airway samples and taxa driving the spatial distribution. Using a multivariate analysis adjusted by TNM stage, **Supplementary Figure 10** shows that poor prognosis was associated with enrichment of the lower airway microbiota with oral commensals (such as *Streptococcus*, *Prevotella* and *Veillonella*). When analysis was repeated only considering the lower airway samples with closest proximity to the cancer, similar results were found (**Supplementary Figure 11**). Using a mixed effect model adjusted by smoking status, stage (I-IIIA/IIIB-IV), and treatment type, we identified top OTUs associated with overall survival. **Supplementary Table 4** reports the top 20 OTUs ranked by absolute coefficient estimates associated with overall survival. Poor prognosis was associated with enrichment with OTUs recognized as oral commensals that belong to the genera *Prevotella*, *Streptococcus*, *Lactobacillus*, and *Gemella*.

Utilizing a Dirichlet Multinomial Model (DMM), we established that samples can be divided into two clusters: cluster one consists of all the upper airway samples and ~60% of lower airway samples and cluster two consists of all the bronchoscope background control samples and ~40% of the lower airway samples (**Supplementary Figure 12a,b**). Thus, similar to previously published data(15), our cohort consists of one cluster of lower airway samples enriched with background predominant taxa (BPT), such as *Flavobacterium* and *Pseudomonas*, while the second cluster was enriched with supraglottic predominant taxa (SPT), such as *Veillonella*, *Streptococcus*, *Prevotella*, and *Haemophilus* (**Supplementary Figure 12c and Supplementary File Table 6**). **Supplementary table 5** shows that we did not identify statistically significant differences in demographic or clinical characteristics, other than stage IV TNM staging ($p<0.05$), between subjects with a lower airway microbiota that clustered as BPT vs. SPT. Applying decontam(26) approach to this data, an analytical pipeline that accounts for taxa most likely to be contaminants, we identified Flavobacterium as a background contaminant (also most prevalent and abundant OTU in background controls) while oral commensals, such as *Veillonella* and *Streptococcus*, as most representatives of lower airway microbiota

(**Supplementary Figure 13**).

We then used the DMM grouping to evaluate whether the prevalence of SPT/BPT was different among stage I-IIIA and IIIB-IV NSCLC and/or associated with prognosis. The percentage of SPT was higher in lower airway samples from subjects with IIIB-IV stage NSCLC group compared to lower airway samples from I-IIIA stage NSCLC group (**Figure 1d,** p=0.006). Importantly, the Kaplan-Meier survival analysis shows that among subjects with stage I-IIIA NSCLC, the SPT-pneumotype was associated with worse survival than the BPT-pneumotype (**Figure 1e**, p=0.047). In stage IIIB-IV NSCLC, there were no statistically significant differences in survival between the SPT- vs. BPT- pneumotypes, although the overall mortality was much worse with a median survival of less than one year as found in the above analysis. To further evaluate microbial signatures associated with treatment response, we analyzed a subset of stage IIIB-IV NSCLC patients (thus non-surgical) with available longitudinal imaging which allowed us to calculate the Response Evaluation Criteria In Solid Tumors (RECIST)(27). Correlation analysis between delta RECIST score and diversity dissimilarity between upper and lower airways showed a significant inverse correlation (**Figure 1f**, Spearman r = -0.48, p=0.03). Thus, although overall mortality was not associated with pneumotypes categorization in IIIB-IV stage group, having a positive delta RECIST score, indicating tumor progression, was associated with having a lower airway microbiota more similar to that of upper airways. Taxonomic differences between a dichotomized RECIST score showed lower airway samples from patients with tumor progression (RECIST = Progressive Disease or Stable Disease) were enriched with *Veillonella, Streptococcus*, *Prevotella*, and *Rothia* when compared with lower airway samples from patients with tumor regression (RECIST = Complete Response or Partial Response; **Supplementary Figure 14** and **Supplementary File Table 7**).

Transcriptomic signatures associated with stage, prognosis and microbiota

After quality control, RNA-Seq data was obtained on 70 lower airway samples from 70 subjects

with NSCLC. We then compared global transcriptomic differences between stage I-IIIA and IIIB-IV NSCLC with PCoA based on the Bray Curtis dissimilarity index. In contrast to microbiota data, there were no statistically significant differences in □□ diversity between these two groups. DESeq analysis showed that there were only 20 genes differentially regulated in stage IIIB-IV compared with stage I-IIIA NSCLC (**Supplementary Figure 15**, **Supplementary File Table 8**). Similarly, very few transcripts were found differentially expressed when comparing better vs. worse outcomes at 6-month and 1-year survival (**Supplementary File Table 8)**.

We then used DESeq to compare transcriptomic signatures associated with a distinct lower airway microbiota base on DMM and found that there were 209 genes up-regulated and 88 genes down-regulated in airway brushes of subjects with SPT lower airway microbiota vs. BPT lower airway microbiota (**Figure 2a**, **Supplementary File Table 9**, FDR<0.25). Sub-analysis of the transcriptomic data among stage I-IIIA and IIIB-IV NSCLC showed the most significant differences for SPT vs. BPT within stage I-IIIA NSCLC. Functional enrichment analysis (Ingenuity Pathway Analysis) of differentially expressed genes between SPT vs. BPT (all samples or stage I-IIIA NSCLC samples) showed that SPT was associated with upregulation of the following top canonical pathways: p53 mutation, PI3K/PTEN, ERK, and IL-6/IL-8 (**Figure 2b**).

Multi-omic Analysis

To better characterize host/microbe interaction in lung cancer we used a multi-omic analytical framework that evaluates for associations between co-occurring taxa and host RNA transcriptome signatures. We estimated co-occurrence probabilities between taxa and the host transcripts adding the probability ranks for the taxa being associated with stage I-IIIA or IIIB-IV lung cancer using MMvec[27,28]. Based on the predicted microbe-transcript co-occurrences, there were two distinct clusters of taxa (**Figure 2c**, interactive figure available at https://segalmicrobiomelab.github.io). The first cluster consisted of SPT-associated taxa

(belonging to the genera *Veillonella*, *Prevotella* and *Streptococcus*) that had a high probability of being observed in subjects with stage IIIB-IV. The second cluster consisted of BPT-associated taxa (such as *Flavobacterium* genus) that had a high probability of being observed in subjects with stage I-IIIA stage NSCLC; however it is important to note that many of the high abundant genera in this cluster (stage I-IIIA) likely represent background taxa as identified by *decontam* (**Supplementary Figure 13**) and not true lower airway taxa. Among SPT-associated taxa, a *Veillonella* taxon (OTU#585419) had the highest relative abundance and a high probability of being found in subjects with stage IIIB-IV lung cancer. This taxon was also highly associated with cell adhesion molecules, IL-17, cytokines and growth factors, chemokine signaling pathway, TNF, Jak-STAT, and PI3K-Akt signaling pathway (**Supplementary File Table 10**). Using BLAST(28), the sequence of this OTU most closely aligned with *Veillonella parvula*.

Lung dysbiosis Preclinical model

To evaluate the causal effects of lower airway dysbiosis on lung cancer progression, we tested the effects of lower airway dysbiosis induced by *Veillonella parvula* in a preclinical lung cancer model (KP mice, **Figure 3a**). We selected this bacterium since we have found it to be a good marker for SPT, it was consistently associated with NSCLC,(16) and it was the taxa with the highest relative abundance identified in our multi-omic analysis as associated with stage IIIB-IV and transcriptomic signatures. Of note, lower airway dysbiosis induced by other oral commensals, such as *Streptococcus mitis* and *Prevotella melaninogenica*, also led to increased lower airway inflammation but at a lesser degree than *V. parvula* (**Supplementary Figure 16-17a,b**). Thus, as a proof of concept, we chose *Veillonella parvula* as our lower airway dysbiosis model for the KP lung cancer mice.

Dysbiosis was induced once KP seeding was determined. Induction of lower airway dysbiosis with *V. parvula* in WT mice did not affect the mice's survival or weight gain. In contrast, within

KP lung cancer mice, exposure to dysbiosis (KP+Dys) led to decreased survival, weight loss, and increased tumor burden (**Figure 3a,b, Supplementary Figure 18a,b**). The experiment was repeated at an early sac time-point (3 weeks post induction of dysbiosis) to evaluate the immune response to dysbiosis with host transcriptomics, T-cell profiling, and cytokine measurements. PCoA analysis of host transcriptomics showed clear differences between the four experimental conditions, where dysbiosis led to greater compositional changes than lung cancer alone (**Supplementary Figure 19a**)**.** Characterization of immune cell subsets inferred from bulk transcriptomics (CIBERSORT) identified clear clustering by condition where lower airway dysbiosis led to an increase in Th1 cells and activation of dendritic cells (**Supplementary Figure 19b**). IPA analysis showed that dysbiosis led to upregulation of PI3k/Akt, ERK/MAPK, IL-17A, IL-6/IL-8, and Inflammasome pathways (**Figure 3c**). Comparisons between transcriptomic signatures induced by lower airway dysbiosis in the NSCLC mouse model and those identified in SPT among subjects with NSCLC showed concordant signatures related to IL-17 signaling, Chemokine, Toll-like receptor, PD-L1 signaling, and PI3K-Akt signaling, among others (**Supplementary Figure 20a,b**). While there are notable differences between transcriptomic signatures in human and mice data, these findings provide a promising direction for follow up. Lastly, lung dysbiosis induced by V. parvula led to the recruitment of Th17 cells, with increased levels of IL-17 production, increased expression of PD-1^{+} T-cells, and recruitment of neutrophils (**Figure 3d, Supplementary Figure 21**). Spatial analysis with immunohistochemistry (IHC) targeting $CD4^{+}$, $CD8^{+}$, and neutrophils show that the increase of these inflammatory cells in response to dysbiosis occurred predominately in tumor-spared lung tissue (**Figure 3e, Supplementary Figure 22a**). Interestingly, in the tumor there was a decrease in $CD4^{+}$ T-cells associated with lower airway dysbiosis.

To further assess the functional importance of dysbiotic-induced IL-17 activation in lung tumorigenesis, dysbiotic-KP mice were treated with monoclonal antibodies against IL-17 or isotype antibody control for two weeks after tumor initiation (**Figure 4a**). Tumor luminescence

data showed that IL-17 blockade led to a decrease in tumor burden over the second week compared with Isotype-control (p=0.0059, **Figure 4b**). Immune profiling evaluated at day 14 after IL-17 blockade showed that treatment with anti-IL-17 antibodies was associated with decreased RORγt^{+} CD4^{+} T-cells, neutrophils, and a non-statistically significant trend towards lower IL-17^{+} CD4^{+} and IL17^{+}TCRγδ^{+} T-cells (**Figure 4c**). Histological assessment with IHC shows that IL-17 blockade led to a decrease in CD4^{+}, CD8^{+}, and Neutrophils in the spared non-tumor lung tissue but not in the tumor itself (**Figure 4d, Supplementary Figure 22b**). Overall, these data suggest that lower airway dysbiosis contributes to a tumor inflammatory microenvironment characterized by an increase in the Th1 and Th17 phenotype, activation of dendritic cells with potential antigen presentation capacity, and an increase in checkpoint inhibitor markers within the surrounding lung tissue.

Discussion

The lower airway microbiota, whether in health or disease state, is mostly affected by aspiration of oral secretions and the lower airway microbial products are in constant interaction with the host immune system(15,19,29-31). In this study, we are the first to demonstrate that a lower airway dysbiotic signature present in lung cancer patients affects tumor progression and clinical prognosis, likely due to alteration in stage I-IIIA immune tone promoting inflammation and checkpoint inhibition. First, patients with stage IIIB-IV NSCLC are more likely to have enrichment of the lower airway microbiota with oral commensals compared to patients with stage I-IIIA disease. In addition, this dysbiotic signature was associated with: a) worse outcome at six-month and one-year (for both groups I-IIIA and IIIB-IV stage disease); b) overall survival in group I-IIIA stage disease; and c) tumor progression in IIIB-IV stage disease. Our preclinical data using a NSCLC mouse supports a model in which aspiration of oral commensals (identified in our human cohort) affect the lower airway inflammatory tone and promote tumor cell proliferation. Dysbiosis in these mice led to upregulation of ERK/MAPK, IL-1, IL-6, and

inflammasome signaling pathways. Immune profiling showed that lung dysbiosis led to a substantial increase in Th17 cells and PD-1^{+} cells. Previous preclinical models of cancer have shown the association between lung dysbiosis and lung inflammation but have limited human microbiome data to support the clinical relevance (especially considering that the human and murine microbiota differs) (32-35). Our data identified that enrichment of the lower airway microbiota with human oral commensals, such as *Veillonella*, contribute to a local pro-tumor immune tone leading to progression of NSCLC suggesting that micro-aspiration and/or impaired airway clearance likely affect the pathogenesis of this disease(36).

Several lines of investigations have shown that increased inflammation and decreased immune surveillance, characterized by IL-17 tone and checkpoint inhibition, are associated with poor prognosis in NSCLC. Increased local and systemic IL-17(37,38), systemic IL-6(39), and higher neutrophil-to-T-cell ratio(40) are associated with a poor prognosis in lung cancer. PD-L1, the ligand for PD-1, is induced in non-lymphoid cells and tumor cells under inflammatory conditions triggered by several cytokines, such as IFN-□ and pathogen-associated molecular patterns (PAMPs)(41-43). In addition, many signaling molecules (e.g., NF-□B, MAPK, PI3K, mTOR, and JAK/STAT) that affect proliferation, apoptosis, and cell survival induce PD-L1 expression(44,45). In a bi-transgenic mouse model expressing a conditional IL-17A allele and a conditional Kras^{G12D}, increased IL-17 production was associated with accelerated lung tumor growth, decreased responsiveness to checkpoint inhibition and decreased survival(46). In many cancer models (breast cancer, gastric carcinoma, and lung cancer), inflammasome activation, through IL-1β signaling, leads to an inflammatory response characterized by decreased anti-tumor immune surveillance(47-49). In the current investigation we show that the increase in IL-17 inflammatory tone triggered by lower airway dysbiosis can be blunted by anti-IL-17 blocking antibodies which seemed to led to a decrease in the tumor burden. More experiments are obviously needed to further characterize the phenotypic inflammatory profile in the tumor and surrounding tissue, to understand the molecular mechanisms by which lower airway

inflammatory cells respond to lower airway dysbiosis, and to better characterize how these factors affect tumor burden and survival. However, the above discussed investigations and the data presented in the current paper supports that the balance between Th17 inflammation and immune surveillance affects NSCLC pathogenesis, and, thus, future investigations are warranted to explore the role of IL-17 blockade in this disease.

Immune checkpoint molecules, such as PD-1, mediate the response of T-cells to neoantigens and are now first line therapy for advanced NSCLC(4-8). However, 40-60% of patients will not benefit from these therapies, and existing biomarkers (e.g., expression of PD-1 ligand) have limited capacity to predict efficacy(7,50). Different gut microbiota signatures have been identified as associated with augmenting anti-tumor immunity and a PD-1 blockade response(9-11). In the gut, higher α-diversity and enrichment of *Ruminococcaceae* were associated with a favorable response to anti-PD-1 treatment in melanoma patients(10,51); and modulation of the microbiota in germ-free mice can enhance antitumor immunity and augment effects of checkpoint blockade(12,13). In germ-free or antibiotic-treated mice, lung adenocarcinoma (Kras mutation/p53 deletion) development is decreased compared to specific pathogen-free mice(32). In this model, lung microbiota activates IL-1β and IL-23 cytokines from myeloid cells and induces IL-17 producing γδ T-cells. Thus, while most studies have focused on the effect of the gut microbiome on cancer development and progression, there is increasing evidence to suggest that the local lung microbiota plays a pivotal role in lung cancer pathogenesis and treatment. Multiple lines of investigations have shown that the lower airway microbiota is a major determinant of the airway immune tone in health and many disease states. For example, recent preclinical models have shown that lower airway mucosal inflammation is primarily associated with the composition of the lower airway microbiota rather than the composition of the gut or upper airway microbiota(14). In humans, we have shown that pneumotype$_{SPT}$ is associated with increased local inflammatory cells and the Th17 phenotype(15,52), and the lower airway microbial metabolism can be modulated by, for example, chronic macrolide

therapy leading to release of microbial metabolites with anti-inflammatory effects(53,54). Anaerobes are commonly found in the lower airways and can survive oxygen stress by forming multicellular complexes within the hypoxic environment present in biofilms(55,56). Short chain fatty acids (SCFAs) produced by fermentation, such as butyrate, are one energy source for anaerobes(57), and we have shown that their presence in the lower airways is higher in pneumotype$_{SPT}$ and regulates IFN-γ and IL-17A production in $CD4^+/CD8^+$ lymphocytes(58). In NSCLC, we recently demonstrated that pneumotype$_{SPT}$ is associated with several inflammatory cancer-related pathways, such as ERK/MAPK and PI3K/AKT(16), that can lead to chronic inflammation, altered Treg/Th17 balance(59-61), augmented Th17 differentiation(62,63), and induction of PD-L1 expression(44,45). Our current findings expand the above observations by demonstrating that a dysbiotic signature characterized by enrichment of the lower airway microbiota with oral commensals can contribute to the progression of disease.

Among the limitations pertinent to this study we should point out that there is a significant degree of disease heterogeneity and the appropriate sub-analyses could only be explored with a much larger cohort. For example, we decided to focus on NSCLC because there were few cases of small cell lung cancer. Further, within NSCLC there were several pathological subtypes, driver mutation status, PD-L1 status, etc. The small subsample size prevents us from conducting the appropriate sub-analysis. However, our analysis and models were stratified and adjusted by staging (dichotomized as I-IIIA and IIIB-IV stage group and adjusted by individual TNM stages) which is a very significant covariate associated with prognosis and treatment modality. Interestingly, we found a few host transcriptomic signatures associated with a disease stage while there were much more transcriptomic signatures associated with lower airway microbiota subtype (SPT/BPT). It is possible that the histological heterogeneity within NSCLC will affect these results and a larger cohort may allow to control for this. Other potential confounders related to patient's clinical condition, such as swallowing and deglutition problems, cannot be fully accounted in the current cohort but may have significant impact on our results.

Given our finding of the enrichment of the lower airway microbiota with oral commensal as associated with prognosis, future investigations that include evaluation of swallowing functions are warranted. Low biomass samples are subjected to contamination with background DNA (coming from the reagents, bronchoscopy or sequencing noise)(64,65). To address concerns regarding DNA contamination during sample collection and preparation, we applied decontam(26) analysis and showed that *Flavobacterium*, a taxon identified in the multi-omic analysis and dominant of BPT, is likely a background contaminant. This is consistent with prior data showing no growth from lower airway samples characterized as BPT(16). We therefore induced airway dysbiosis in our mice model with Veillonella and compared it with PBS (which 16S rRNA gene sequencing composition most resemble BPT) rather than a separate bacterium as control. Our investigation supports the hypothesis that the lower airway microbiota contributes to a local pro-tumor immunity, however, we did not investigate the systemic inflammatory response in this model. Further support for the relevance of this mechanism will need to focus on blocking the immune response to the microbial exposure in the setting of lung cancer and evaluating the effects of induced lower airway dysbiosis during immunotherapy. In the current investigation, we did not explore the association between lung microbiota and response to immunotherapy because this treatment was applied in a relatively small fraction of patients (16%) and the vast majority of the samples were collected before this therapy became standard of care. Also, while we identified a taxonomic signature associated with inflammatory tone and prognosis in lung cancer, we cannot determine the molecular signatures present in the microbial community that may be responsible for this association. Future investigations that exploit novel functional microbiomic approaches (e.g. metagenome, metatranscriptome, metabolome) should focus on molecular markers with significant immunomodulatory activity. In our preclinical model we tested whether *Veillonella parvula* was sufficient to induce lower airway inflammation and worsening of tumor progression. Other oral commensal, when present in the lower airways, may also be contributing to this process and may need to be further evaluated as

key components of lower airway dysbiosis in isolation or in complex microbial communities. Although the lower airway microbiota was associated with staging and survival, other dysbiotic signatures in other mucosae could also have significant associations. Even though we did not identify significant microbiota signatures in the buccal samples, future investigations should include gut samples as well to establish the relative role of the microbiota of different mucosae niches to the pathogenesis of lung cancer. Finally, further validation of the results presented here will require a second cohort where sampling approach and design are customized to overcome some of the limitations here described.

This study has broad clinical implications regarding lung cancer pathogenesis and treatment response. Identification of lower airway dysbiotic signatures associated with lung cancer prognosis may be important to personalize approaches for lung cancer treatment and prognosis. Fecal microbiota transplant (FMT), a strategy with proven efficacy in difficulty-to-treat *Clostridium difficile* infection and inflammatory bowel disease(66,67), can influence the susceptibility to anti-PD-1 cancer immunotherapy(9,10), and its clinical impact is now being tested in humans within ongoing clinical trials. Despite the evidence that the local microbiota affects the local inflammatory tone of the lung, there are no human trials aiming to modify the lung microbiome in the setting of malignancy. The data presented here suggest that lower airway dysbiosis induced by microaspiration of oral commensals affect lung tumorigenesis by promoting an IL-17 driven inflammatory phenotype, a pathway amenable for targeted therapy that may have potential implications in this disease. A better understanding of the microbial host interaction in the lower airways will be needed to uncover how the lung cancer-associated microbiota could be modulated to affect prognosis and response to immunotherapies.

Methods

Subjects

All subjects signed written informed consents to participate in this study that was approved by the Institutional Review Board of New York University. Participants included patients who had suspicious nodules on chest imaging and who underwent clinical bronchoscopy. Lung cancer sub-type, somatic mutation, stage was recorded after histopathological confirmation. We excluded subjects with a prior history of cancer or recent (less than 1 month) antibiotic use. Response Evaluation Criteria In Solid Tumors (RECIST)(27) score was analyzed at the 6-12 month time point after diagnosis of lung cancer, where this data was most consistently available.

Bronchoscopic Procedure

Both background and supraglottic (buccal) samples were obtained prior to the procedure as previously described(16). The background samples were obtained by passing sterile saline through the suctioning channel of the bronchoscope prior to the procedure. For this project, we obtained multiple lower airway samples from different locations, including 82 from the right mainstem, 59 from the airways leading to the lung cancer lesion (involved segments), and 69 from the airways spared of disease on the contralateral lung. A detailed description of the number of samples and the analyses performed in them is provided in **Supplementary Table 6**.

Bacterial 16S rRNA-encoding genes sequencing

High-throughput sequencing of bacterial 16S rRNA-encoding gene amplicons (V4 region)(68) was performed. Reagent control samples and mock mixed microbial DNA were sequenced and analyzed in parallel (**Supplementary Figure 23**). The obtained 16S rRNA gene sequences were analyzed with the Quantitative Insights Into Microbial Ecology (QIIME,RRID:SCR_008249) 1.9.1 package(69). Operational taxonomic units (OTU) were not removed from upstream

571 analysis. PERMANOVA testing was used to compare the compositional differences of groups. A
572 prevalence-based method using the R package *decontam* (v1.6.0)(26) was used to identify
573 potential contaminants from the sequencing datasets. In this process, all reads from
574 background bronchoscope control samples were identified as negative controls and, thus,
575 possible source of contaminants. No OTU was removed from the analyses performed and data
576 from the 16S microbiome for this manuscript is available (data available at Sequence Read
577 Archive, RRID:SCR_001370 : #PRJNA592147).

578 Sample clustering of meta-communities was based on Dirichlet-Multinomial mixtures (DMM)
579 modeling(70).

580 *Transcriptome of bronchial epithelial cells*

581 RNA-Seq was performed on bronchial epithelial cells obtained by airway brushing, as
582 described(71-73), using the Hi-seq/Illumina platform at the NYU Langone Genomic Technology
583 Center (data available at Sequence Read Archive: # PRJNA592149). KEGG(74,75) annotation
584 was summarized at levels 1 to 3. Genes with an FDR-corrected adjusted p-value <0.25 were
585 considered significantly differentiated, unless otherwise specified. Pathway analysis using
586 differentially regulated genes (FDR<0.25) was done using Ingenuity Pathway Analysis,
587 RRID:SCR_008653 (QIAGEN Inc.)(76). Gene Set Enrichment Analysis (GSEA) was performed
588 with differential genes (FDR<0.25) for dataset comparison, R package fgsea v1.4.1(77).

589 *Experimental Mouse Model:*

590 The mice utilized in this experiment were 5 week-old females at the time of use. The strain was
591 B6(Cg)-Tyrc-2J/J mice purchased via vendor (Jackson Laboratory; Bar Harbor, ME, USA
592 Cat#000058). The mice were kept in Skirball Animal Facility and were kept under controlled
593 conditions with cycles of 12-hour daylight and 12-hour darkness. Mice were euthanized by
594 carbon dioxide asphyxiation followed by cardiac puncture. Blood, skin swabs, oral swabs, lung

lavage, lung tissue, humerus bone marrow, cecum, terminal ileum, and fecal pellets were collected for study. The Institutional Animal Care and Use Committee of the New York University School of Medicine approved all procedures and experiments were carried out following their guidelines (IACUC# s16-00032).

KP Model Lung Cancer:

The KP model of lung cancer histopathologically resembles that of human cancers and has been used to study translational models of lung cancer in mice(78). The KP model of lung cancer is based on $Kras^{LSL\text{-}G12D/+}$;$p53^{fl/fl}$ Non-small cell Lung Cancer models require induction by use of replication-deficient adenoviruses expression Cre (Ad-Cre) to induce transient Cre expression in the lungs of mice. Once tumor burden is increased in the mice, the lungs are harvested and the KP lung cancer cells grown in cell culture(79). Cell culture lines of KP lung cancer cells are grown in DMEM 10%FBS plus gentamicin under aerobic conditions with 5% carbon dioxide at 37°C. Cells were harvested from the cell culture when 90% congruent. The goal was to grow cells to 3,000,000 KP Cells/mL (or 150,000 cells / 50 uL). To detect *in vivo* luminescence, images were acquired using the IVIS spectrum (PerkinElmer) after intraperitoneal injection of Luciferin (Promega). We then proceed to intra-tracheal inoculation of KP cells. The mice were anesthetized utilizing isoflurane until sedated. The mice were then placed on an intubation platform and with blunt forceps, their tongue was gently pulled ventrally until the pharynx was exposed.(78) Then, an Exel Safelet catheter (Exel International Inc.; St. Petersburg, FL, USA Cat# 26746) was introduced through the vocal cords of the mice, and a 50 µL inoculum of lung cancer ($1.5x10^5$ KP cells) was placed into the catheter. The mice were then removed from the intubation platform to recover from anesthesia on a heat pad.

Creation of Veillonella parvula inoculum:

The following human oral commensals were obtained: *Veillonella parvula*, *Prevotella melaninogenica*, and *Streptococcus mitis* (ATCC; Manassas, VA, USA). These bacteria were

grown in anaerobic conditions (Bactron 300, Shel Labs, Cornelius, OR), then stored in 20% glycerol tryptic soy broth at -80°C. To prepare the oral commensal challenges the bacteria strains were thawed and streaked on anaerobic PRAS-Brucella Blood agar plates (Anaerobe Systems, Morgan Hill, CA). The plates were incubated at 37°C in an oxygen-free environment (tri-mix: 5% carbon dioxide, 5% hydrogen, and 90% nitrogen) in an anaerobic chamber for 24-48 hours. The colonies were collected from the plate and re-suspended in 1 ml of sterile PBS. The OD620 was measured to calculate the approximate concentration prior to use.

Intra-tracheal microbial and control challenge:

Mice were assigned to receive the microbial challenge with *Veillonella parvula* twice a week via intra-tracheal inoculation starting 2 weeks after the inoculation with lung cancer. First, mice were sedated with the use of isoflurane anesthesia. The mice were then suspended by their dorsal incisors upon an elastic cord; a blunt pair of forceps was used to ventrally pull the tongue forward to expose the larynx. Then, a pneumatic otoscope (Welch-Allyn; Shaneateles Falls, NY, USA Cat#71000C) with a 2mm ear specula was advanced until the vocal cords were visualized. Using a gel loading tip, a 50 µL volume of the *Veillonella parvula* was deployed into the trachea of the mouse. These exposures occurred twice a week, spaced 3-4 day/week apart. Mice were monitored during this process; no mice died due to the inoculation procedure. A control procedure to inoculate mice with PBS was performed in the same manner.

Immune inhibition experiment:

Two weeks after KP cell inoculation, mice were challenged intra-tracheal with *Veillonella parvula* similar to above. At this time mice were randomized 1:1 to receive anit-IL-17 (1mg/mL; Bio X Cell Lebanon, NH, USA), anti-IL-17 iso-type control (2mg/mL; Bio X Cell Lebanon, NH, USA). Antibody dose was diluted in 100µl and given via intraperitoneal injection twice a week for a total of 2 weeks.

644

645 *Organization and measurements on mice:*

646 Once lung tumor development was detected by IVIS (2 weeks post inoculation) mice were
647 randomized according to tumor burden to receive either PBS or dysbiosis with *V. parvula* while
648 maintaining co-house conditions. For the KP mice, those with median lumens of $8x10^5$-$7x10^6$
649 photon-flux (photons/s/cm2/steradian) at 2 weeks were utilized for the experiments. Wild type
650 mice from the same strain and no KP exposure were used as control mice and were exposed to
651 sterile PBS or *V. parvula*. Thus, in all experiments, mice were organized to the following groups:
652 1) Wild type with PBS control (WT), 2) Wild type with dysbiosis with *V. parvula* (Dys), 3) KP lung
653 cancer with PBS control (LC), and, 4) KP Lung cancer with *V. parvula* (LC + Dys). Imaging the
654 mice utilizing luciferins expression (lumens) occurred 2 weeks after inoculation with KP lung
655 cancer cells. The platform we used to image the mice was Perkin-Elmer IVIS Spectrum (Perkin-
656 Elmer; Waltham, MA, USA Cat# 124262). 1.5mg of Luciferin (Perkin Elmer, Xeno-Light D-
657 Luciferin Potassium Salt, cat# 122799) was given intraperitoneally. Mice received 50µL of their
658 respective inoculum with the *Veillonella* condition receiving $1.5x10^6$ cfu/mL. The mice were
659 organized into groups based upon their median lumens to establish experimental groups of
660 mice with the same luminosity for a baseline. The imaging of the mice occurred twice every
661 week on the day prior to inoculation. For the survival experiment we utilized 60 mice that were
662 followed for six weeks after initiation of microbial challenge or PBS control. Forty additional mice
663 were divided in same four conditional groups for immune phenotyping on lung homogeneate,
664 including lung transcriptomics, flow cytometry and cytokine measurement. For this experiment,
665 mice were sacrificed after two weeks post initiation of microbial or PBS exposure. For host RNA
666 transcriptome, flash frozen lung samples were defrosted and then homogenized utilizing a hand
667 TissueRuptor II on the 2nd lowest setting (Qiagen, Hilden, Germany). Then samples were spun
668 down on a table-top centrifuge 14,000 rpm for 2 minutes and the pellet was collected and sent
669 for RNA processing. RNA was extracted from collected supernatant using the Qiagen

miRNeasy Mini Kit (Qiagen, Hilden, Germany Cat#74135). Quality control was established with RNA integrity number (RIN) cut-off >6. RNA sequencing was performed using HiSeq (Illumina, San Diego, CA) at the NYU Genomic Technology Center. RNA-Seq library preps were made using Illumina TruSeq® Stranded mRNA LT kit (Illumina, San Diego, CA Cat#RS-1222-2101) on a Beckman Biomek FX instrument, using 250 ng of total RNA as input, amplified by 12 cycles of PCR, and run on an Illumina 2500 (v4 chemistry), as single-read 50bp. Sequences from the murine lung homogenate were aligned against the murine ensemble reference genome utilizing STAR, RRID:SCR_015899 (v2.5) aligner(80). Gene counting of each sample was performed using featureCounts, RRID:SCR_012919 (v1.5.3) (81,82). FACS was performed on single cell suspension derived from lung homogenate. First, lung samples were minced and dissociated utilizing Liberase (Hoffmann-La Roche, Basel, Switzerland) for 35 minutes in a 37°C water bath and followed by mechanical disruption through a 70-micron filter. Liberase was used at a concentration of 0.5 mg/mL in DMEM supplemented with 10% fetal bovine serum (FBS). For intra-cellular cytokine staining, the cells were treated with a cell stimulation and protein transport inhibition cocktail containing PMA, Ionomycin, Brefeldin A, and Monensin (500x eBioscience Affymetrix, Santa Clara, CA) for 4 hours. The cells were surface stained, fixed in 2% PFA, and permeabilized with 0.5% saponin. Cell staining with fluorochrome-conjugated antibodies was performed targeting: $CD3^+$, $CD4^+$, $CD8^+$, $CD69^+$, $PD1^+$, $IL17^+$ (Thermo-Fischer, Waltham, MA) and measurement were performed on a BD LSR II flow cytometer (BD Bioscience, Franklin Lakes, NJ). Acquired data was analyzed using FlowJo, RRID:SCR_008520 version 10.3 (Tree Star Inc., Ashland, OR). Cytokines and Chemokines were measured using Luminex (Murine Cytokine Panel II, EMD Millipore, Burlington, MA). Lung homogenates were thawed and processed according to recommended protocol using the Murine Cytokine/Chemokine Magnetic Bead Panel # MCYTMAG-70K-PXkl32). All cytokines/chemokines concentrations were normalized by the gram of lung homogenate and included those with dynamic range: G-CSF,

Eotaxin, IFN-g, IL-1a, Il-1b, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9 IL-10, IL-12p40, IL-12p70, LIF, IL-17, IP-10, KC, MCP-1, MIP-1a, MIP-1b, M-CSF, MIP-2, MIG, RANTES, VEGF, and TNF-a.

Multiplex immunostaining:

Five-micron sections of paraffin embedded preserved lung tissue were stained with Akoya Biosciences® Opal™ multiplex automation kit reagents unless stated otherwise. Automated staining was performed on Leica BondRX® autostainer. The protocol was performed according to manufacturers' instructions with the antibodies specified in **Supplementary Table 7**. Briefly, all slides underwent sequential epitope retrieval with Leica Biosystems epitope retrieval 1 (ER1, citrate based, pH 6.0, Cat. AR9961) and 2 solution (ER2, EDTA based, pH9, Cat. AR9640), primary and secondary antibody incubation and tyramide signal amplification (TSA) with Opal® fluorophores (**Supplementary Table 7**). Primary and secondary antibodies were removed during epitope retrieval steps while fluorophores remain covalently attached to the epitope.

Image acquisition and analysis:

Semi-automated image acquisition was performed on a Vectra® Polaris multispectral imaging system. After whole slide scanning at 20X the tissue was manually outlined to select fields for spectral unmixing and analysis using InForm® version 2.4.10 software from Akoya Biosciences. Fields of view for analysis were separated as containing tumor only or areas of pulmonary parenchyma where tumor was not apparent. For each field of view, cells were segmented based on nuclear signal (DAPI). Cells were phenotyped after segmentation using inForm's trainable algorithm based on glmnet(83) package in R. Four algorithms were created to classify cell as Ly6g+ (Neutrophils) or 'other', CD4+ or 'other', CD8+ or 'other' and F4/80+ or 'other'. Phenotypes were reviewed for different samples during training iterations. Data was exported as text containing sample names, field of acquisition coordinates, individual cell information including coordinates and identified phenotype. Each image was analyzed with all four

algorithms so that every cell was classified four times. Concatenation of all phenotyping information was performed in R using the Phenoptr Reports package (Kent S Johnson (2020). phenoptr: inForm Helper Functions. R package version 0.2.7. https://akoyabio.github.io/phenoptr/) in RStudio software [RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.] Statistical analysis (Mann-Whitney U test) was run for the following groups: lung cancer vs. lung cancer + dysbiosis (n=4 and 8 mice respectively, **Figure 3e)**, and lung cancer + dysbiosis vs. lung cancer + dysbiosis + anti-IL-17 (n=8 and 6 mice respectively, **Figure 4d**), taking each field as an independent value.

Statistical and Multi-omic Analysis:

In **Supplementary Table 2**, the categorical variables were presented as frequencies and percentages and their distribution difference between groups with Dead or Alive overall survival (OS) status were assessed by the Fisher's exact test. The Cox Proportion Hazards models(84) were used to evaluate each variable's marginal association with the time to death. Hazard ratio (HR) and p-value were reported.

The microbiome regression-based kernel association test (MiRKAT) (85) was used to investigate whether the community level microbial profile among lower airway samples were different between any paired samples from right main, involved, or non-involved locations, and between stage I-IIIA and IIIB-IV while adjusting for smoking status within each location samples. The survival version of MiRKAT test: MiRKAT-S(86) was used to investigate whether the community level microbial profile is associated with the overall survival (OS) while adjusting for smoking status, stage and surgery within each location samples. The paired Bray-Curtis dissimilarity was used in all tests.

For the taxonomic level analysis, we used the linear mixed effect model on the arcsine square root transformed relative abundance at genus level for their associations with stage (I-IIIA/ IIIB-IV, **Supplementary Table 3**). In the model, the subject was set as the random effect to take care of the correlation among three location samples from the same subjects. The stage was set as fixed effect while adjusting for smoking status. We used the two stage linear mixed effect model(87) on the arcsine square root transformed relative abundance at genus level for their associations of the overall survival (**Supplementary Table 4**) while adjusting for smoking status, stage, and surgery. In the first stage, the linear mixed effect model was used to take care of the correlation among three location samples from the same subjects. The random intercept estimates from the first stage were used in the Cox proportional hazards model in the second stage to investigate their association with the overall survival.

Since the distributions of microbiome data are non-normal, and no distribution-specific tests are available, we used non-parametric tests of association. For association with discrete factors, we used either the Mann-Whitney test (in the case of 2 categories) or the Kruskal Wallis ANOVA (in the case of >2 categories). For tests of association with continuous variables, we used the non-parametric Spearman correlation tests. False discovery rate (FDR) was used to control for multiple testing(88). To evaluate for taxonomic or transcriptomic differences between groups, we utilized DESeq2(89).

Differential abundance of microbes related to lung cancer stage (IIIB-IV vs. I-IIIA) were calculated using Songbird as previously described.(90) Then we computed the microbe-transcript co-occurrence probability (probability of observing a transcriptomic pathway when a microbe is observed) using mmvec.(91) A probability matrix of the top 10 transcriptome related pathways for each microbe was generated and used to create a network based on the Fruchterman-Reingold force-directed algorithm using R package ggnet v 0.1.0. (reference: https://cran.r-project.org/web/packages/GGally/index.html). Microbe nodes were colored based on differential analysis of stage IIIB-IV versus I-IIIA non-small cell lung cancer.

769 *Data Storage*

770 Sequencing data is available at Sequence Read Archive(92,93) under accession number 16S
771 Microbiome PRJNA592147, Human RNASeq PRJNA600487, and Murine RNASeq
772 PRJNA600489. Codes utilized for the analyses presented in the current manuscript are
773 available at *https://github.com/segalmicrobiomelab/reviewer_copy (acct: reviewermicrobiome,*
774 *password: sunshine888manatee).*

Reference:

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. **2019**;69(1):7-34 doi 10.3322/caac.21551.
2. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature **2014**;511(7511):543-50 doi 10.1038/nature13385.
3. Rosell R, Karachaliou N. Large-scale screening for somatic mutations in lung cancer. Lancet **2016**;387(10026):1354-6 doi 10.1016/S0140-6736(15)01125-3.
4. Patel SP, Kurzrock R. PD-L1 Expression as a Predictive Biomarker in Cancer Immunotherapy. Mol Cancer Ther **2015**;14(4):847-56 doi 10.1158/1535-7163.MCT-14-0983.
5. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ*, et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science **2015**;348(6230):124-8 doi 10.1126/science.aaa1348.
6. Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS*, et al.* Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. Nature **2014**;515(7528):563-7 doi 10.1038/nature14011.
7. Gandhi L, Rodriguez-Abreu D, Gadgeel S, Esteban E, Felip E, De Angelis F*, et al.* Pembrolizumab plus Chemotherapy in Metastatic Non-Small-Cell Lung Cancer. N Engl J Med **2018**;378(22):2078-92 doi 10.1056/NEJMoa1801005.
8. Reck M, Rodriguez-Abreu D, Robinson AG, Hui R, Csoszi T, Fulop A*, et al.* Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med **2016**;375(19):1823-33 doi 10.1056/NEJMoa1606774.
9. Routy B, Le Chatelier E, Derosa L, Duong CPM, Alou MT, Daillere R*, et al.* Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science **2018**;359(6371):91-7 doi 10.1126/science.aan3706.
10. Gopalakrishnan V, Spencer CN, Nezi L, Reuben A, Andrews MC, Karpinets TV*, et al.* Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science **2018**;359(6371):97-103 doi 10.1126/science.aan4236.
11. Matson V, Fessler J, Bao R, Chongsuwat T, Zha Y, Alegre ML*, et al.* The commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. Science **2018**;359(6371):104-8 doi 10.1126/science.aao3290.
12. Sivan A, Corrales L, Hubert N, Williams JB, Aquino-Michaels K, Earley ZM*, et al.* Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. Science **2015**;350(6264):1084-9 doi 10.1126/science.aac4255.
13. Vetizou M, Pitt JM, Daillere R, Lepage P, Waldschmitt N, Flament C*, et al.* Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. Science **2015**;350(6264):1079-84 doi 10.1126/science.aad1329.
14. Dickson RP, Erb-Downward JR, Falkowski NR, Hunter EM, Ashley SL, Huffnagle GB. The Lung Microbiota of Healthy Mice are Highly Variable, Cluster by Environment, and Reflect Variation in Baseline Lung Innate Immunity. Am J Respir Crit Care Med **2018** doi 10.1164/rccm.201711-2180OC.
15. Segal LN, Clemente JC, Tsay JC, Koralov SB, Keller BC, Wu BG*, et al.* Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. Nat Microbiol **2016**;1:16031 doi 10.1038/nmicrobiol.2016.31.
16. Tsay JJ, Wu BG, Badri MH, Clemente JC, Shen N, Meyn P*, et al.* Airway Microbiota Is Associated with Upregulation of the PI3K Pathway in Lung Cancer. Am J Respir Crit Care Med **2018**;198(9):1188-98 doi 10.1164/rccm.201710-2118OC.
17. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A*, et al.* Topographical continuity of bacterial populations in the healthy human respiratory tract. American journal of respiratory and critical care medicine **2011**;184(8):957-63 doi 10.1164/rccm.201104-0655OC.

18. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB*, et al.* Bacterial Topography of the Healthy Human Lower Respiratory Tract. MBio **2017**;8(1) doi 10.1128/mBio.02287-16.
19. Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Chen H*, et al.* Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. Microbiome **2013**;1(1):19 doi 10.1186/2049-2618-1-19.
20. Gustafson AM, Soldi R, Anderlind C, Scholand MB, Qian J, Zhang X*, et al.* Airway PI3K pathway activation is an early and reversible event in lung cancer development. Sci Transl Med **2010**;2(26):26ra5 doi 10.1126/scitranslmed.3000251.
21. Greathouse KL, White JR, Vargas AJ, Bliskovsky VV, Beck JA, von Muhlinen N*, et al.* Interaction between the microbiome and TP53 in human lung cancer. Genome Biol **2018**;19(1):123 doi 10.1186/s13059-018-1501-6.
22. Yoon SM, Shaikh T, Hallman M. Therapeutic management options for stage III non-small cell lung cancer. World J Clin Oncol **2017**;8(1):1-20 doi 10.5306/wjco.v8.i1.1.
23. Fan H, Shao ZY, Xiao YY, Xie ZH, Chen W, Xie H*, et al.* Incidence and survival of non-small cell lung cancer in Shanghai: a population-based cohort study. BMJ Open **2015**;5(12):e009419 doi 10.1136/bmjopen-2015-009419.
24. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, Rami-Porta R*, et al.* The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. J Thorac Oncol **2007**;2(8):706-14 doi 10.1097/JTO.0b013e31812f3c1a.
25. Ettinger DS, Wood DE, Aggarwal C, Aisner DL, Akerley W, Bauman JR*, et al.* NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 1.2020. J Natl Compr Canc Netw **2019**;17(12):1464-72 doi 10.6004/jnccn.2019.0059.
26. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome **2018**;6(1):226 doi 10.1186/s40168-018-0605-2.
27. Schwartz LH, Litiere S, de Vries E, Ford R, Gwyther S, Mandrekar S*, et al.* RECIST 1.1-Update and clarification: From the RECIST committee. Eur J Cancer **2016**;62:132-7 doi 10.1016/j.ejca.2016.03.081.
28. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res **2008**;36(Web Server issue):W5-9 doi 10.1093/nar/gkn201.
29. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Beck JM, Huffnagle GB*, et al.* Spatial Variation in the Healthy Human Lung Microbiome and the Adapted Island Model of Lung Biogeography. Annals of the American Thoracic Society **2015**;12(6):821-30 doi 10.1513/AnnalsATS.201501-029OC.
30. Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB*, et al.* Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. mBio **2015**;6(2):e00037 doi 10.1128/mBio.00037-15.
31. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL*, et al.* Comparison of the Respiratory Microbiome in Healthy Non-Smokers and Smokers. Am J Respir Crit Care Med **2013**;187(10):1067-75 doi 10.1164/rccm.201210-1913OC.
32. Jin C, Lagoudas GK, Zhao C, Bullman S, Bhutkar A, Hu B*, et al.* Commensal Microbiota Promote Lung Cancer Development via gammadelta T Cells. Cell **2019**;176(5):998-1013 e16 doi 10.1016/j.cell.2018.12.040.
33. Gui QF, Lu HF, Zhang CX, Xu ZR, Yang YH. Well-balanced commensal microbiota contributes to anti-cancer response in a lung cancer mouse model. Genet Mol Res **2015**;14(2):5642-51 doi 10.4238/2015.May.25.16.

34. Cheng M, Chen Y, Wang L, Chen W, Yang L, Shen G*, et al.* Commensal microbiota maintains alveolar macrophages with a low level of CCL24 production to generate anti-metastatic tumor activity. Sci Rep **2017**;7(1):7471 doi 10.1038/s41598-017-08264-8.
35. Cheng M, Qian L, Shen G, Bian G, Xu T, Xu W*, et al.* Microbiota modulate tumoral immune surveillance in lung through a gammadeltaT17 immune cell-dependent mechanism. Cancer Res **2014**;74(15):4030-41 doi 10.1158/0008-5472.can-13-2462.
36. Lee SH, Sung JY, Yong D, Chun J, Kim SY, Song JH*, et al.* Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions. Lung Cancer **2016**;102:89-95 doi http://dx.doi.org/10.1016/j.lungcan.2016.10.016.
37. Chen X, Wan J, Liu J, Xie W, Diao X, Xu J*, et al.* Increased IL-17-producing cells correlate with poor survival and lymphangiogenesis in NSCLC patients. Lung Cancer **2010**;69(3):348-54 doi 10.1016/j.lungcan.2009.11.013.
38. Xu C, Hao K, Yu L, Zhang X. Serum interleukin-17 as a diagnostic and prognostic marker for non-small cell lung cancer. Biomarkers **2014**;19(4):287-90 doi 10.3109/1354750X.2014.908954.
39. Liao C, Yu Z, Guo W, Liu Q, Wu Y, Li Y*, et al.* Prognostic value of circulating inflammatory factors in non-small cell lung cancer: a systematic review and meta-analysis. Cancer Biomark **2014**;14(6):469-81 doi 10.3233/CBM-140423.
40. Tomita M, Shimizu T, Ayabe T, Nakamura K, Onitsuka T. Elevated preoperative inflammatory markers based on neutrophil-to-lymphocyte ratio and C-reactive protein predict poor survival in resected non-small cell lung cancer. Anticancer Res **2012**;32(8):3535-8.
41. Loke P, Allison JP. PD-L1 and PD-L2 are differentially regulated by Th1 and Th2 cells. Proc Natl Acad Sci U S A **2003**;100(9):5336-41 doi 10.1073/pnas.0931259100.
42. Liu J, Hamrouni A, Wolowiec D, Coiteux V, Kuliczkowski K, Hetuin D*, et al.* Plasma cells from multiple myeloma patients express B7-H1 (PD-L1) and increase expression after stimulation with IFN-{gamma} and TLR ligands via a MyD88-, TRAF6-, and MEK-dependent pathway. Blood **2007**;110(1):296-304 doi 10.1182/blood-2006-10-051482.
43. Qian Y, Deng J, Geng L, Xie H, Jiang G, Zhou L*, et al.* TLR4 signaling induces B7-H1 expression through MAPK pathways in bladder cancer cells. Cancer Invest **2008**;26(8):816-21 doi 10.1080/07357900801941852.
44. Lee SK, Seo SH, Kim BS, Kim CD, Lee JH, Kang JS*, et al.* IFN-gamma regulates the expression of B7-H1 in dermal fibroblast cells. J Dermatol Sci **2005**;40(2):95-103 doi 10.1016/j.jdermsci.2005.06.008.
45. Chen J, Feng Y, Lu L, Wang H, Dai L, Li Y*, et al.* Interferon-gamma-induced PD-L1 surface expression on human oral squamous carcinoma via PKD2 signal pathway. Immunobiology **2012**;217(4):385-93 doi 10.1016/j.imbio.2011.10.016.
46. Akbay EA, Koyama S, Liu Y, Dries R, Bufe LE, Silkes M*, et al.* Interleukin-17A Promotes Lung Tumor Progression through Neutrophil Attraction to Tumor Sites and Mediating Resistance to PD-1 Blockade. J Thorac Oncol **2017**;12(8):1268-79 doi 10.1016/j.jtho.2017.04.017.
47. Guo B, Fu S, Zhang J, Liu B, Li Z. Targeting inflammasome/IL-1 pathways for cancer immunotherapy. Sci Rep **2016**;6:36107 doi 10.1038/srep36107.
48. Zhong FL, Mamai O, Sborgi L, Boussofara L, Hopkins R, Robinson K*, et al.* Germline NLRP1 Mutations Cause Skin Inflammatory and Cancer Susceptibility Syndromes via Inflammasome Activation. Cell **2016**;167(1):187-202 e17 doi 10.1016/j.cell.2016.09.001.
49. Kolb R, Phan L, Borcherding N, Liu Y, Yuan F, Janowski AM*, et al.* Obesity-associated NLRC4 inflammasome activation drives breast cancer progression. Nat Commun **2016**;7:13007 doi 10.1038/ncomms13007.
50. Maleki Vareki S, Garrigos C, Duran I. Biomarkers of response to PD-1/PD-L1 inhibition. Crit Rev Oncol Hematol **2017**;116:116-24 doi 10.1016/j.critrevonc.2017.06.001.

51. Blacher E, Levy M, Tatirovsky E, Elinav E. Microbiome-Modulated Metabolites at the Interface of Host Immunity. J Immunol **2017**;198(2):572-80 doi 10.4049/jimmunol.1601247.
52. Pradhan D, Segal LN, Kulkarni R, Chung S, Rom WN, Weiden MD*, et al.* Bronchial Reactivity In Early Emphysema May Be Associated With Local Neutrophilic Inflammation. Am J Respir Crit Care Med **2013**:A1110.
53. Segal LN, Clemente JC, Wu BG, Wikoff WR, Gao Z, Li Y*, et al.* Randomised, double-blind, placebo-controlled trial with azithromycin selects for anti-inflammatory microbial metabolites in the emphysematous lung. Thorax **2017**;72(1):13-22 doi 10.1136/thoraxjnl-2016-208599.
54. Dickson RP, Morris A. Macrolides, inflammation and the lung microbiome: untangling the web of causality. Thorax **2017**;72(1):10-2 doi 10.1136/thoraxjnl-2016-209180.
55. Lone AG, Atci E, Renslow R, Beyenal H, Noh S, Fransson B*, et al.* Staphylococcus aureus induces hypoxia and cellular damage in porcine dermal explants. Infect Immun **2015**;83(6):2531-41 doi 10.1128/IAI.03075-14.
56. Williamson KS, Richards LA, Perez-Osorio AC, Pitts B, McInnerney K, Stewart PS*, et al.* Heterogeneity in Pseudomonas aeruginosa biofilms includes expression of ribosome hibernation factors in the antibiotic-tolerant subpopulation and hypoxia-induced stress response in the metabolically active population. J Bacteriol **2012**;194(8):2062-73 doi 10.1128/JB.00022-12.
57. Bourriaud C, Robins RJ, Martin L, Kozlowski F, Tenailleau E, Cherbut C*, et al.* Lactate is mainly fermented to butyrate by human intestinal microfloras but inter-individual variation is evident. J Appl Microbiol **2005**;99(1):201-12 doi 10.1111/j.1365-2672.2005.02605.x.
58. Segal LN, Clemente JC, Li Y, Ruan C, Cao J, Danckers M*, et al.* Anaerobic Bacterial Fermentation Products Increase Tuberculosis Risk in Antiretroviral-Drug-Treated HIV Patients. Cell Host Microbe **2017**;21(4):530-7 e4 doi 10.1016/j.chom.2017.03.003.
59. Barbi J, Pardoll D, Pan F. Metabolic control of the Treg/Th17 axis. Immunol Rev **2013**;252(1):52-77 doi 10.1111/imr.12029.
60. Okkenhaug K, Patton DT, Bilancio A, Garcon F, Rowan WC, Vanhaesebroeck B. The p110delta isoform of phosphoinositide 3-kinase controls clonal expansion and differentiation of Th cells. J Immunol **2006**;177(8):5122-8.
61. Sauer S, Bruno L, Hertweck A, Finlay D, Leleu M, Spivakov M*, et al.* T cell receptor signaling controls Foxp3 expression via PI3K, Akt, and mTOR. Proc Natl Acad Sci U S A **2008**;105(22):7797-802 doi 10.1073/pnas.0800928105.
62. Kurebayashi Y, Nagai S, Ikejiri A, Ohtani M, Ichiyama K, Baba Y*, et al.* PI3K-Akt-mTORC1-S6K1/2 axis controls Th17 differentiation by regulating Gfi1 expression and nuclear translocation of RORgamma. Cell Rep **2012**;1(4):360-73 doi 10.1016/j.celrep.2012.02.007.
63. Liu H, Yao S, Dann SM, Qin H, Elson CO, Cong Y. ERK differentially regulates Th17- and Treg-cell development and contributes to the pathogenesis of colitis. Eur J Immunol **2013**;43(7):1716-26 doi 10.1002/eji.201242889.
64. Erb-Downward JR, Falkowski NR, D'Souza JC, McCloskey LM, McDonald RA, Brown CA*, et al.* Critical Relevance of Stochastic Effects on Low-Bacterial-Biomass 16S rRNA Gene Analysis. mBio **2020**;11(3) doi 10.1128/mBio.00258-20.
65. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF*, et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol **2014**;12:87 doi 10.1186/s12915-014-0087-z.
66. Moayyedi P, Surette MG, Kim PT, Libertucci J, Wolfe M, Onischi C*, et al.* Fecal Microbiota Transplantation Induces Remission in Patients With Active Ulcerative Colitis in a Randomized Controlled Trial. Gastroenterology **2015**;149(1):102-9 e6 doi 10.1053/j.gastro.2015.04.001.
67. Paramsothy S, Kamm MA, Kaakoush NO, Walsh AJ, van den Bogaerde J, Samuel D*, et al.* Multidonor intensive faecal microbiota transplantation for active ulcerative colitis: a randomised placebo-controlled trial. Lancet **2017**;389(10075):1218-28 doi 10.1016/S0140-6736(17)30182-4.

68. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N*, et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J **2012**;6(8):1621-4 doi 10.1038/ismej.2012.8.
69. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK*, et al.* QIIME allows analysis of high-throughput community sequencing data. Nat Methods **2010**;7(5):335-6 doi 10.1038/nmeth.f.303.
70. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS One **2012**;7(2):e30126 doi 10.1371/journal.pone.0030126.
71. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods **2008**;5(7):621-8 doi 10.1038/nmeth.1226.
72. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I*, et al.* Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature **2008**;453(7199):1239-43 doi 10.1038/nature07002.
73. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M*, et al.* A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science **2008**;321(5891):956-60 doi 10.1126/science.1160342.
74. Tanabe M, Kanehisa M. Using the KEGG database resource. Curr Protoc Bioinformatics **2012**;Chapter 1:Unit1 12 doi 10.1002/0471250953.bi0112s38.
75. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res **2012**;40(Database issue):D109-14 doi 10.1093/nar/gkr988.
76. Kramer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics **2014**;30(4):523-30 doi 10.1093/bioinformatics/btt703.
77. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A **2005**;102(43):15545-50 doi 10.1073/pnas.0506580102.
78. DuPage M, Dooley AL, Jacks T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. Nat Protoc **2009**;4(7):1064-72 doi 10.1038/nprot.2009.95.
79. Romero R, Sayin VI, Davidson SM, Bauer MR, Singh SX, LeBoeuf SE*, et al.* Keap1 loss promotes Kras-driven lung cancer and results in dependence on glutaminolysis. Nat Med **2017**;23(11):1362-8 doi 10.1038/nm.4407.
80. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S*, et al.* STAR: ultrafast universal RNA-seq aligner. Bioinformatics **2013**;29(1):15-21 doi 10.1093/bioinformatics/bts635.
81. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res **2013**;41(10):e108 doi 10.1093/nar/gkt214.
82. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **2014**;30(7):923-30 doi 10.1093/bioinformatics/btt656.
83. Friedman JH, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. 2010 **2010**;33(1):22 doi 10.18637/jss.v033.i01.
84. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B (Methodological) **1972**;34(2):187-220.
85. Zhao N, Chen J, Carroll IM, Ringel-Kulka T, Epstein MP, Zhou H*, et al.* Testing in Microbiome-Profiling Studies with MiRKAT, the Microbiome Regression-Based Kernel Association Test. Am J Hum Genet **2015**;96(5):797-807 doi 10.1016/j.ajhg.2015.04.003.
86. Plantinga A, Zhan X, Zhao N, Chen J, Jenq RR, Wu MC. MiRKAT-S: a community-level test of association between the microbiota and survival times. Microbiome **2017**;5(1):17 doi 10.1186/s40168-017-0239-9.

87. Sayers A, Heron J, Smith A, Macdonald-Wallis C, Gilthorpe MS, Steele F*, et al.* Joint modelling compared with two stage methods for analysing longitudinal data and prospective outcomes: A simulation study of childhood growth and BP. Stat Methods Med Res **2017**;26(1):437-52 doi 10.1177/0962280214548822.
88. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics **2003**;19(3):368-75.
89. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol **2014**;15(12):550 doi 10.1186/s13059-014-0550-8.
90. Morton JT, Marotz C, Washburne A, Silverman J, Zaramela LS, Edlund A*, et al.* Establishing microbial composition measurement standards with reference frames. Nature communications **2019**;10(1):2719 doi 10.1038/s41467-019-10656-5.
91. Morton JT, Aksenov AA, Nothias LF, Foulds JR, Quinn RA, Badri MH*, et al.* Learning representations of microbe-metabolite interactions. Nature methods **2019**;16(12):1306-14 doi 10.1038/s41592-019-0616-3.
92. Kodama Y, Shumway M, Leinonen R, International Nucleotide Sequence Database C. The Sequence Read Archive: explosive growth of sequencing data. Nucleic Acids Res **2012**;40(Database issue):D54-6 doi 10.1093/nar/gkr854.
93. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database C. The sequence read archive. Nucleic Acids Res **2011**;39(Database issue):D19-21 doi 10.1093/nar/gkq1019.

Figure Legends:

Figure 1. Lung microbiota in lung cancer and cancer survival. **a.** Principal coordinate analysis (PCoA) of airway samples show a difference in β-diversity (p=0.01, PERMANOVA) between small cell lung cancer and non-small cell lung cancer (n=83). **b.** Among patients with NSCLC (n=74), PCoA shows a difference in β-diversity (p=0.005, PERMANOVA) between stage IIIB-IV and I-IIIA NSCLC (**Left Panel**); lower airway microbiota of stage IIIB-IV was more similar to buccal microbiota than lower airway microbiota of stage I-IIIA (**Right Panel,** p<0.0001, Bray Curtis Distance). **c. Left Panel.** PCoA based on cancer stage and survival at 6-months and 1-year shows difference in β-diversity (p<0.05, PERMANOVA). **c. Right Panel.** Lower airway microbiota in lung cancer and worse survival at 6-months or 1-year was more similar to buccal microbiota than with better survival both in the stage IIIB-IV (n=36) and in the stage I-IIIA (n=37) groups (p<0.05, Bray Curtis Distance). **d.** Stage IIIB-IV lung cancer was associated with having a higher proportion of subjects whose lower airway microbiota was classified as enriched with oral taxa (supraglottic predominant taxa, **SPT**) vs. background tax (background predominant taxa, **BPT**), p=0.006. **e.** Enrichment of the lower airway with $\text{Pneumotype}_{\text{SPT}}$ was associated with better survival in stage I-IIIA cancer than enrichment with $\text{Pneumotype}_{\text{BPT}}$, p<0.05; there was no difference in stage IIIB-IV cancer. **f.** Bray Curtis Dissimilarity Index between lower airway and buccal samples was inversely associated with delta RECIST score for stage IIIB-IV NSCLC measured at 6-12 months (Spearman r = -0.48, p=0.03).

Figure 2. Airway transcriptome in NSCLC lung cancer based on lung microbiota. Comparisons between microbiome and host transcriptomic signatures were conducted using samples where paired matched data was available (n=70). **a**. Volcano plot of differentially expressed genes (FDR<0.25) between **$\text{Pneumotype}_{\text{SPT}}$** vs. **$\text{Pneumotype}_{\text{BPT}}$** in all, stage I-IIIA only, or stage IIIB-IV only lower airway samples**. b.** Unsupervised hierarchical heat-map of canonical pathway analysis based on Ingenuity Pathway Analysis (IPA, RRID:SCR_008653)

using the airway transcriptome of all subjects and those with stage I-IIIA comparing **Pneumotype$_{SPT}$** vs. **Pneumotype$_{BPT}$** groups. Sub-analysis using samples from patients with stage IIIB-IV disease is not presented given the paucity of differentially expressed genes between the groups. Orange shows up-regulation of pathway, blue shows down-regulation of pathway. **c.** Network analysis based on conditional co-occurrence probability of microbiome and transcriptome data; Microbiome nodes (circles) are colored red for stage IIIB-IV lung cancer, green for stage I-IIIA lung cancer (based on a gradient) and sized by relative abundance. Edges connect microbiome nodes to pathway nodes and edge width is based on their conditional probability.

Figure 3. Pre-clinical model of lung dysbiosis in lung cancer and cancer survival. a. Experimental condition and Kaplan Meier survival showing decrease survival in mice with lung cancer and dysbiosis (**LC + Dys,** n=22) compared with **LC** (n=20) alone ($p<0.001$). **Dys** did not affect mice survival in wild-type control (n=10 for each group). **b.** Quantitative data of tumor burden (measured as lumens prior to death or sacrifice normalized to baseline lumens) showing that **LC + Dys** mice had increased tumor burden ($p<0.05$, n=5 for each experimental condition). **c.** IPA analysis was used to identify dysregulated transcriptomic pathways. **d.** Immune profiling of lung tissue by FACS and cytokine measurement demonstrates that lower airway dysbiosis induces Th17 and PD-1 T-cell phenotype in the lung. e. Immunohistochemistry analysis comparing LC and LC+dys shows increase in $CD4^+$, $CD8^+$, neutrophils in the non-tumor region after dysbiosis. Minimal difference in immune response was seen within the tumor itself (n=4 (LC) vs. n=8 (LC+dys) mice/group, each dot represents different regions analyzed color-coded by mice).

Figure 4. IL-17 blockade during lung dysbiosis in lung cancer preclinical model. a. Experimental conditions (anti-IL-17 or isotype Ab control) were administered to KP mice with

lower airway dysbiosis induced by *Veillonella parvula*. **b.** anti-IL-17 therapy was associated with decreased tumor burden change during the 2 weeks of antibody injections as compared with isotype control. **c.** Immune profiling of lung tissue by FACS demonstrates that IL-17 blockade of KP mice with lower airway dysbiosis decrease RORyt^{+} and Neutrophils (n=6-7 for each experimental condition). **d**. Immunohistochemistry analysis shows that IL-17 blockade of KP mice with lower airway dysbiosis decrease CD4^{+}/CD8^{+} T-cells and neutrophils in the non-tumor region ($p<0.0001$ and $p=0.0002$, respectively). However, a minimal difference in immune response was seen within the tumor itself (n=8 (LC+dys) vs. n=6 (LC+dys+anti-IL-17) mice/group, each dot represents different region analyzed color-coded by mice).

Figure 1

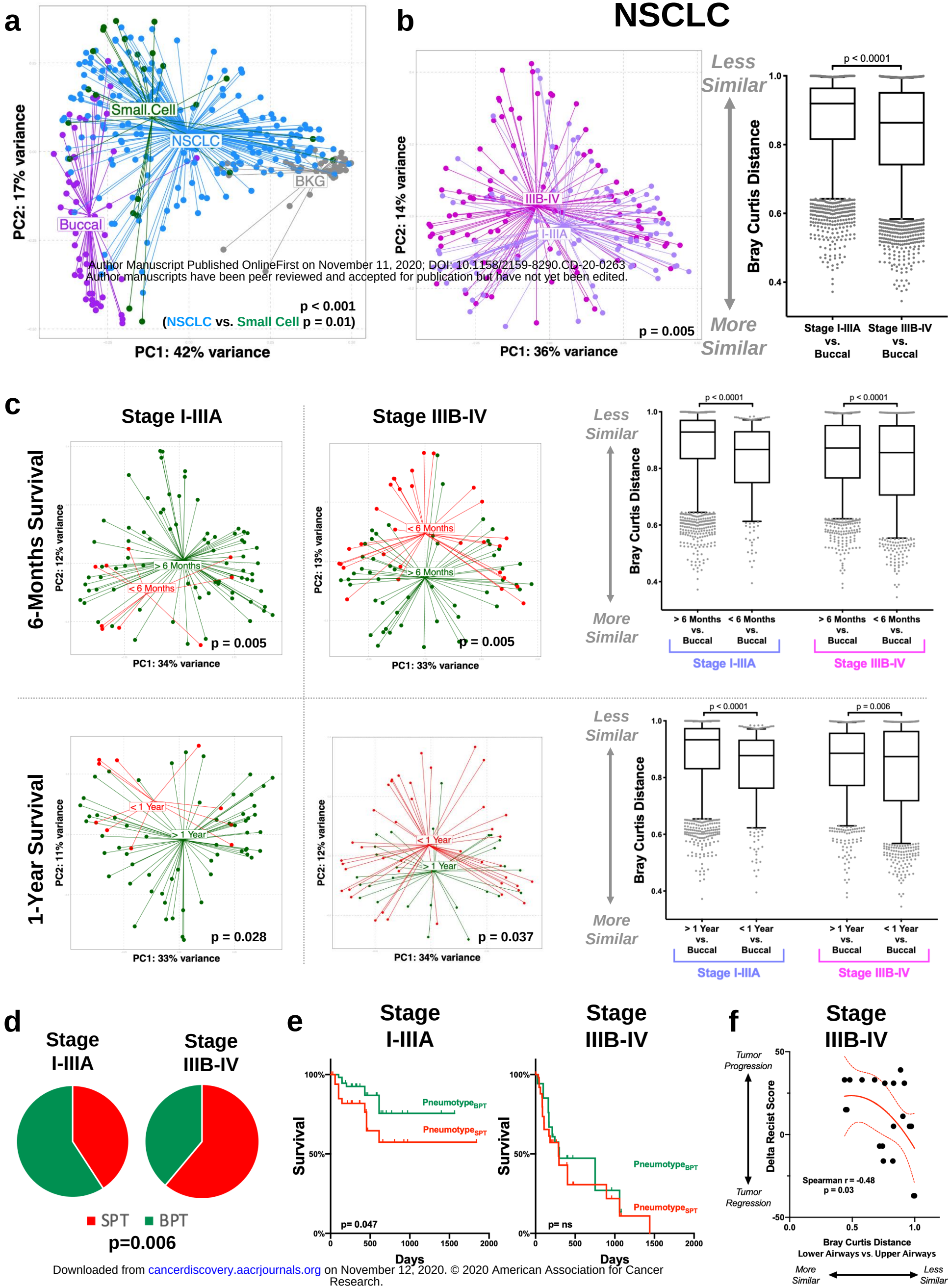


Figure 2

Figure 3

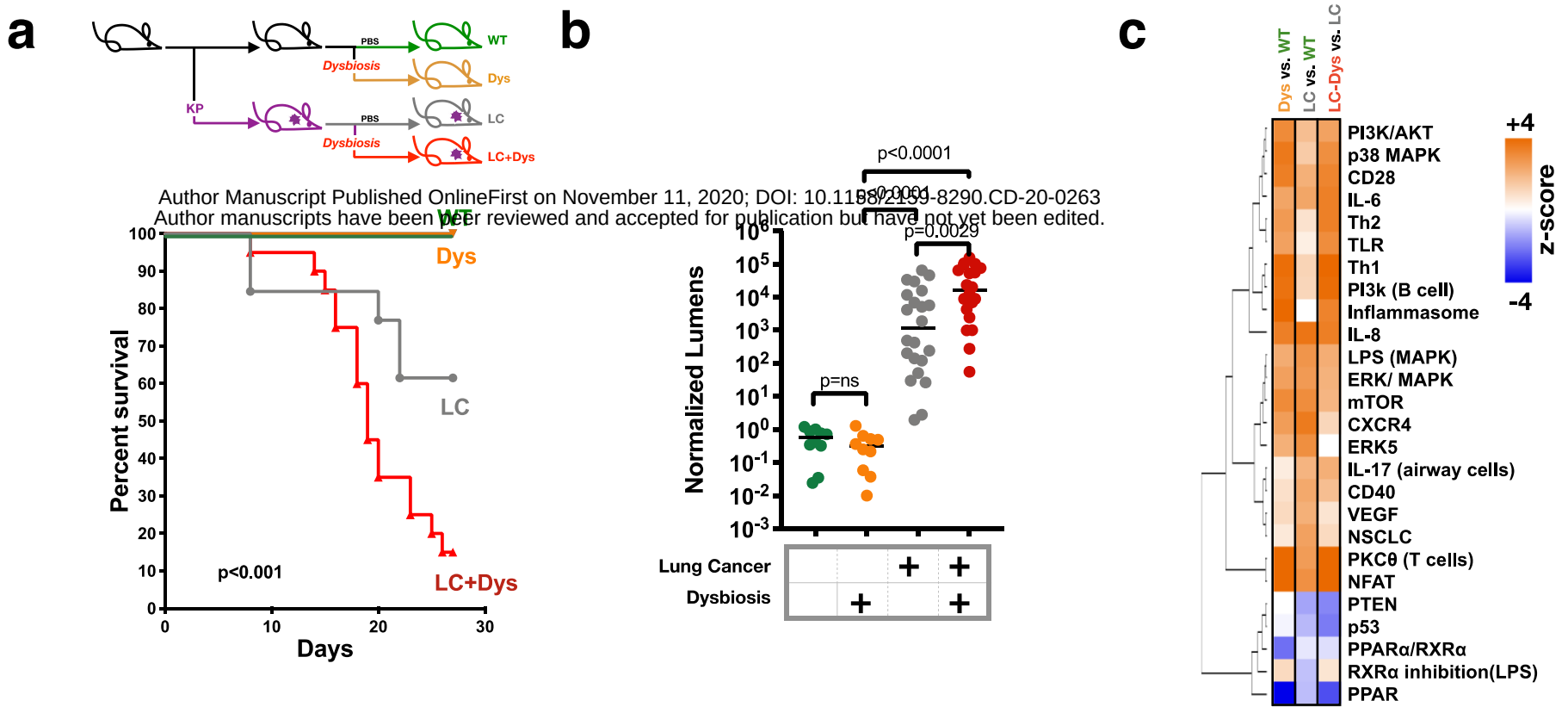
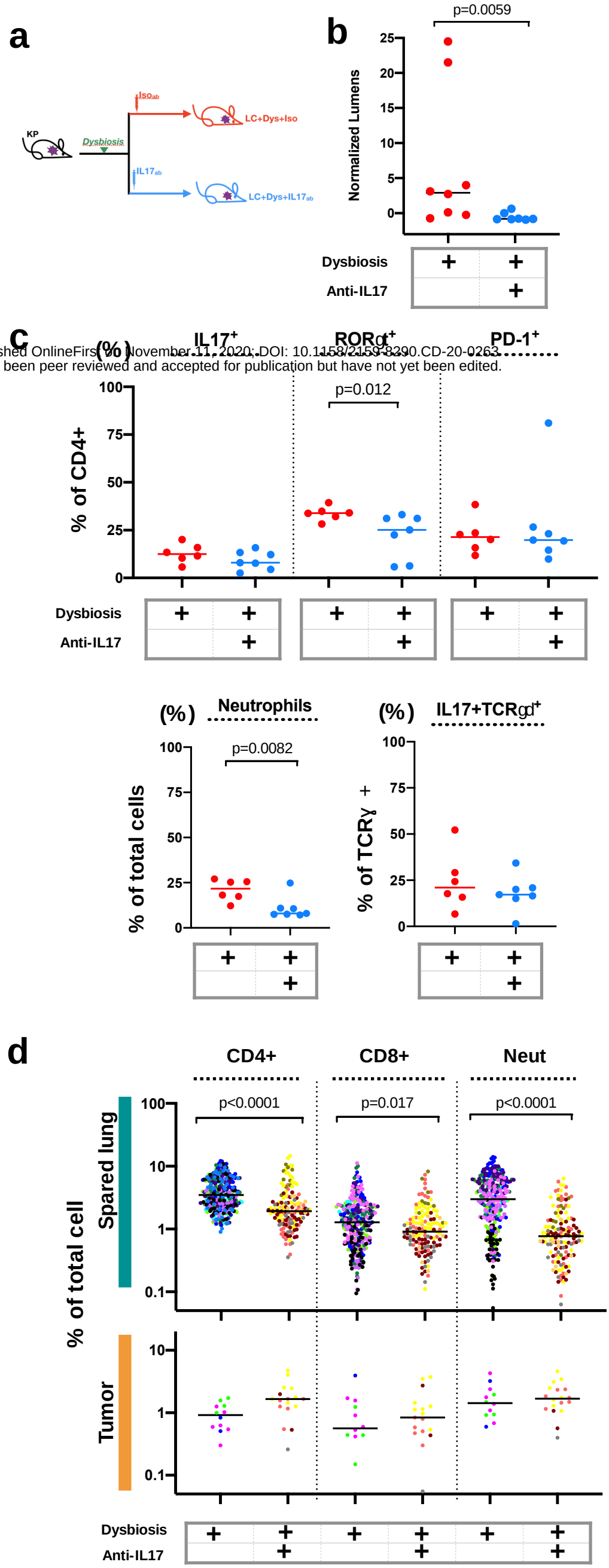


Figure 4



AACR American Association for Cancer Research

CANCER DISCOVERY

Lower airway dysbiosis affects lung cancer progression

Jun-Chieh J Tsay, Benjamin G Wu, Imran Sulaiman, et al.

Cancer Discov Published OnlineFirst November 11, 2020.