

## **SUPPLEMENTARY MATERIAL**

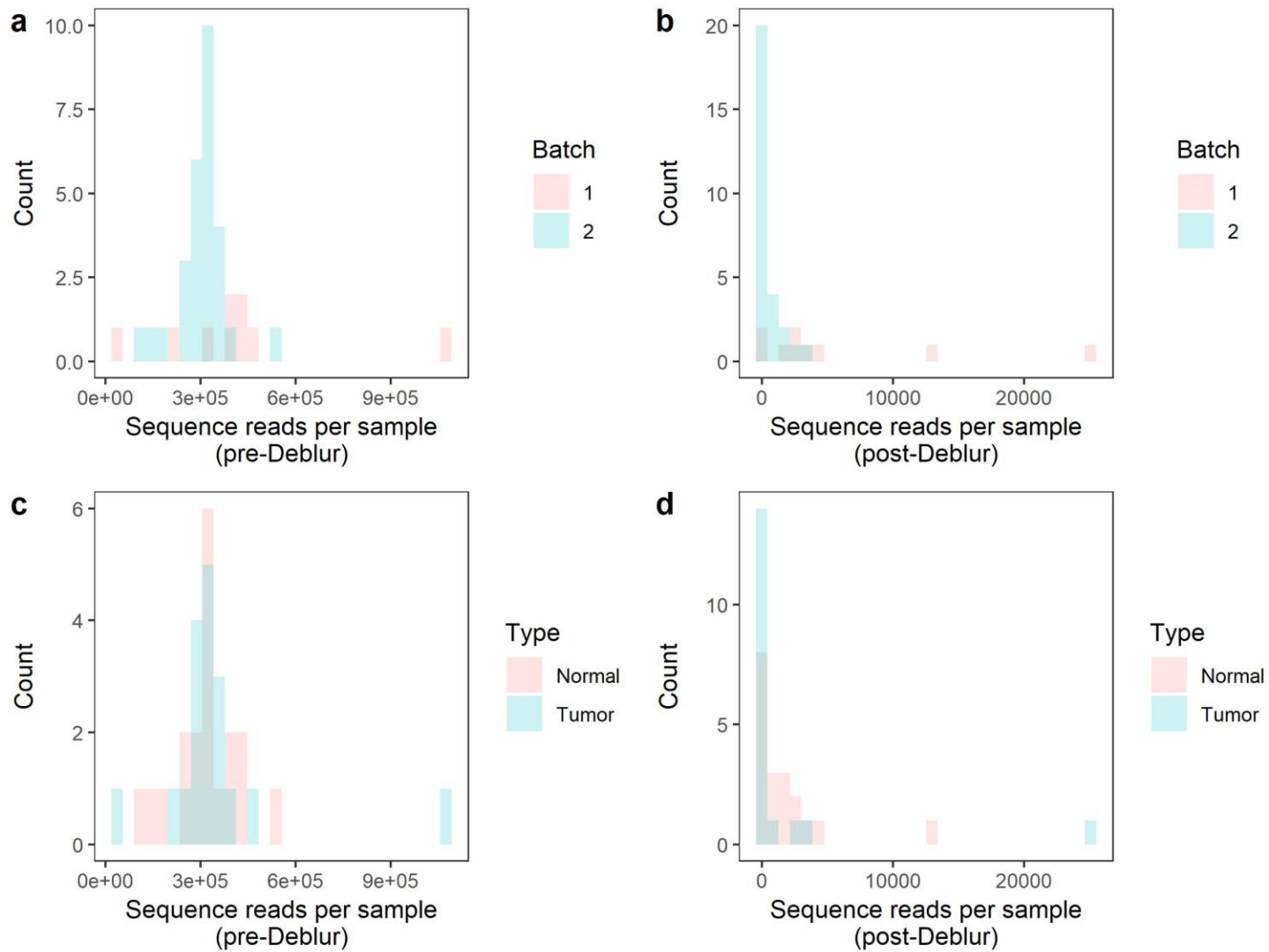
### **The microbiome in lung cancer tissue and recurrence-free survival**

Brandilyn A. Peters,<sup>a</sup> Richard B. Hayes,<sup>a,b</sup> Chandra Goparaju,<sup>c</sup> Christopher Reid,<sup>c</sup> Harvey I. Pass,<sup>b,c</sup> Jiyoun Ahn<sup>a,b</sup>

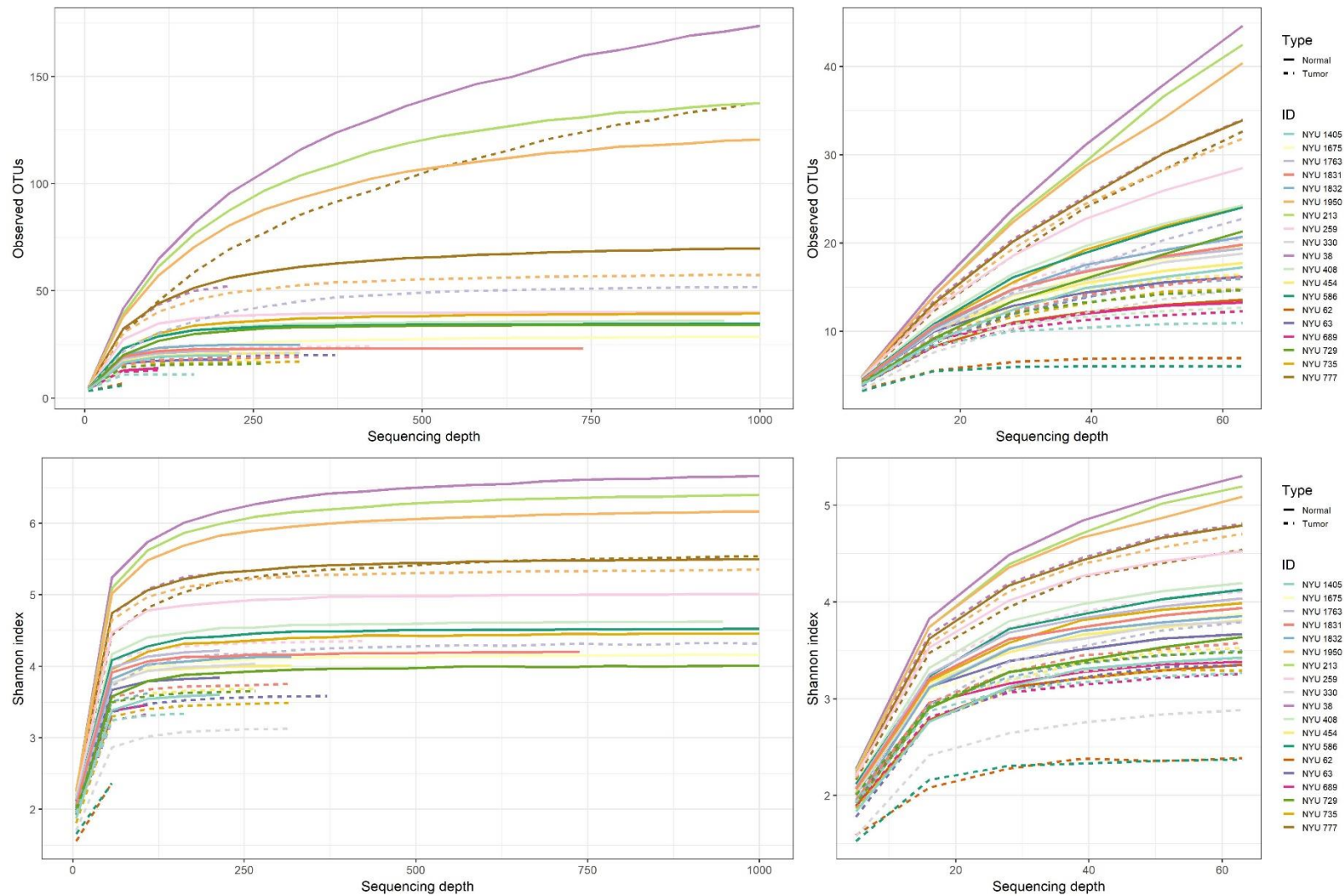
<sup>a</sup>Department of Population Health, NYU School of Medicine, New York, NY, USA

<sup>b</sup>NYU Perlmutter Cancer Center, New York, NY, USA

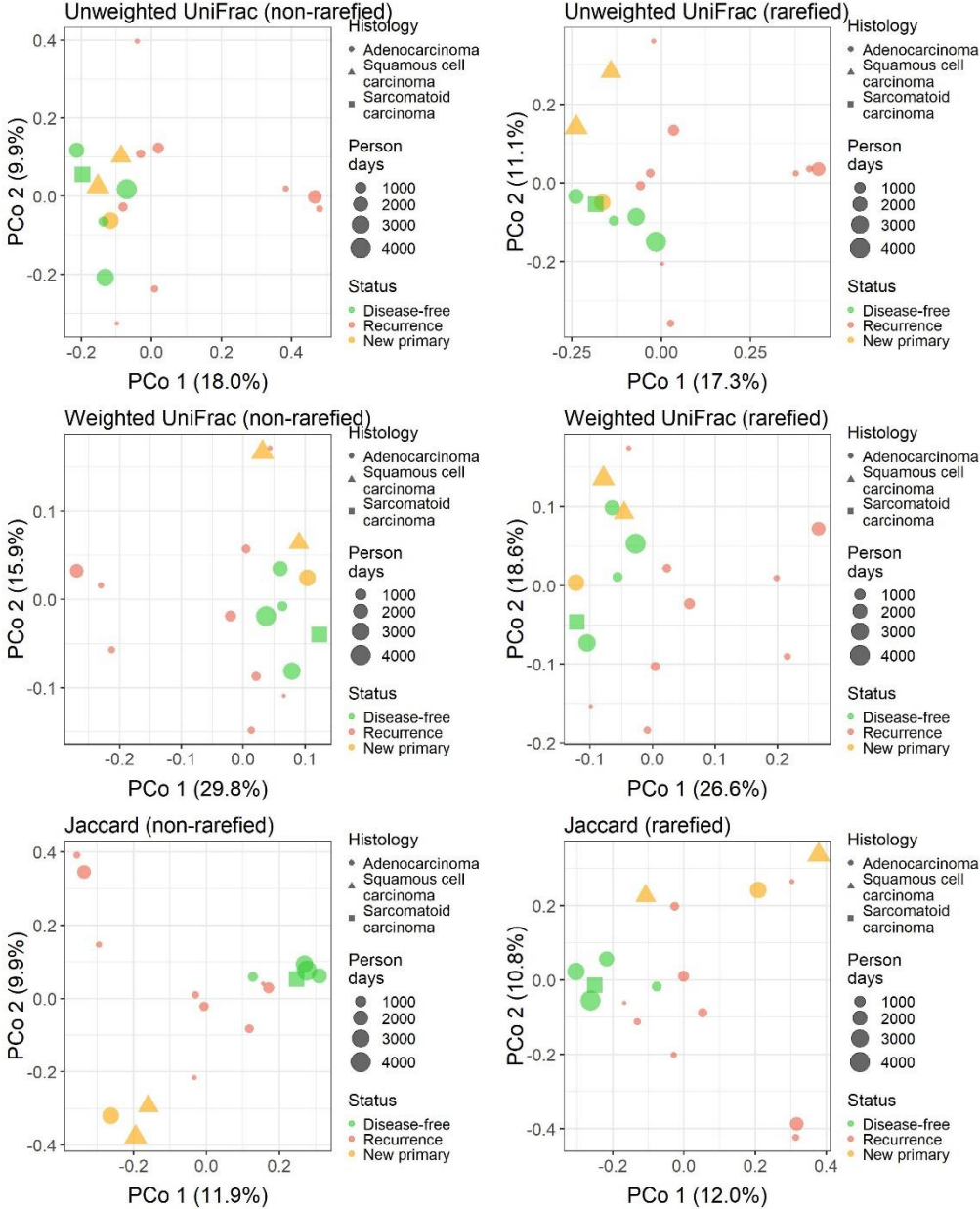
<sup>c</sup>Department of Cardiothoracic Surgery, NYU School of Medicine, New York, NY, USA



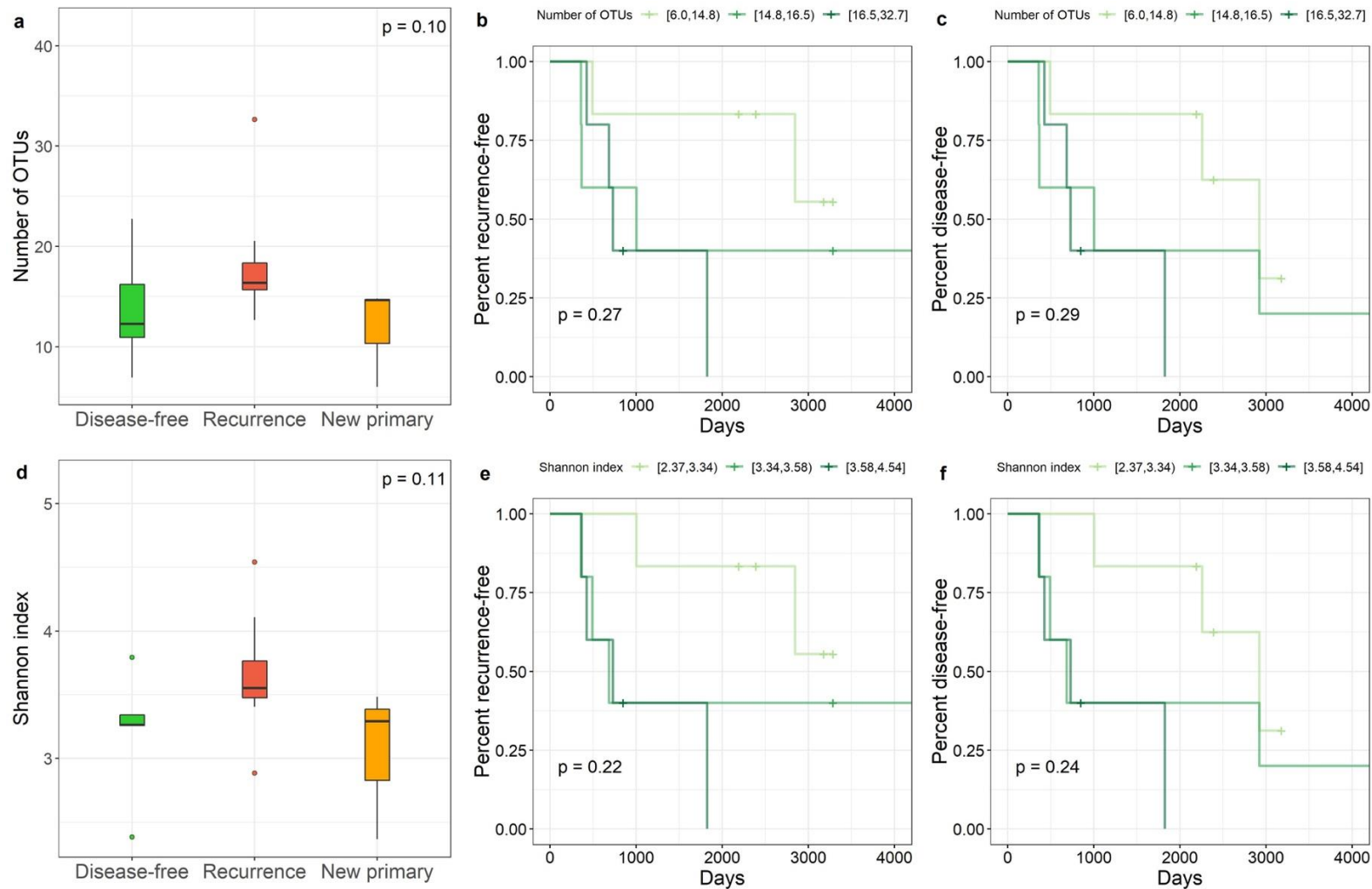
**Supplementary Figure 1.** Histograms of the number of sequence reads per sample (a-b) by batch pre- and post-Deblur and (c-d) by sample tissue type pre- and post-Deblur.



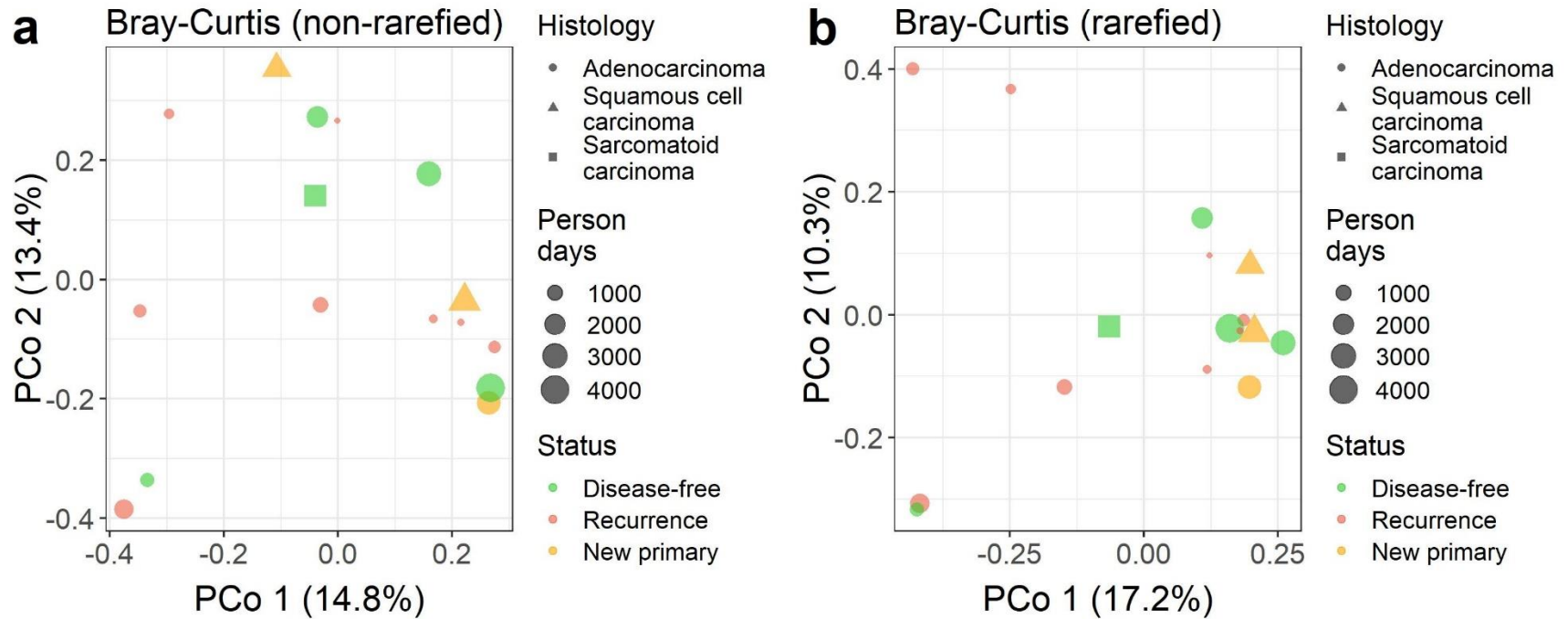
**Supplementary Figure 2.** Rarefaction curves of number of observed OTUs (top) and Shannon index (bottom) in paired lung tumor and normal samples. Left panel shows depth up to 1000 sequence reads per sample. Right panel shows depth up to 63 sequences per sample (lowest observed depth among the samples).



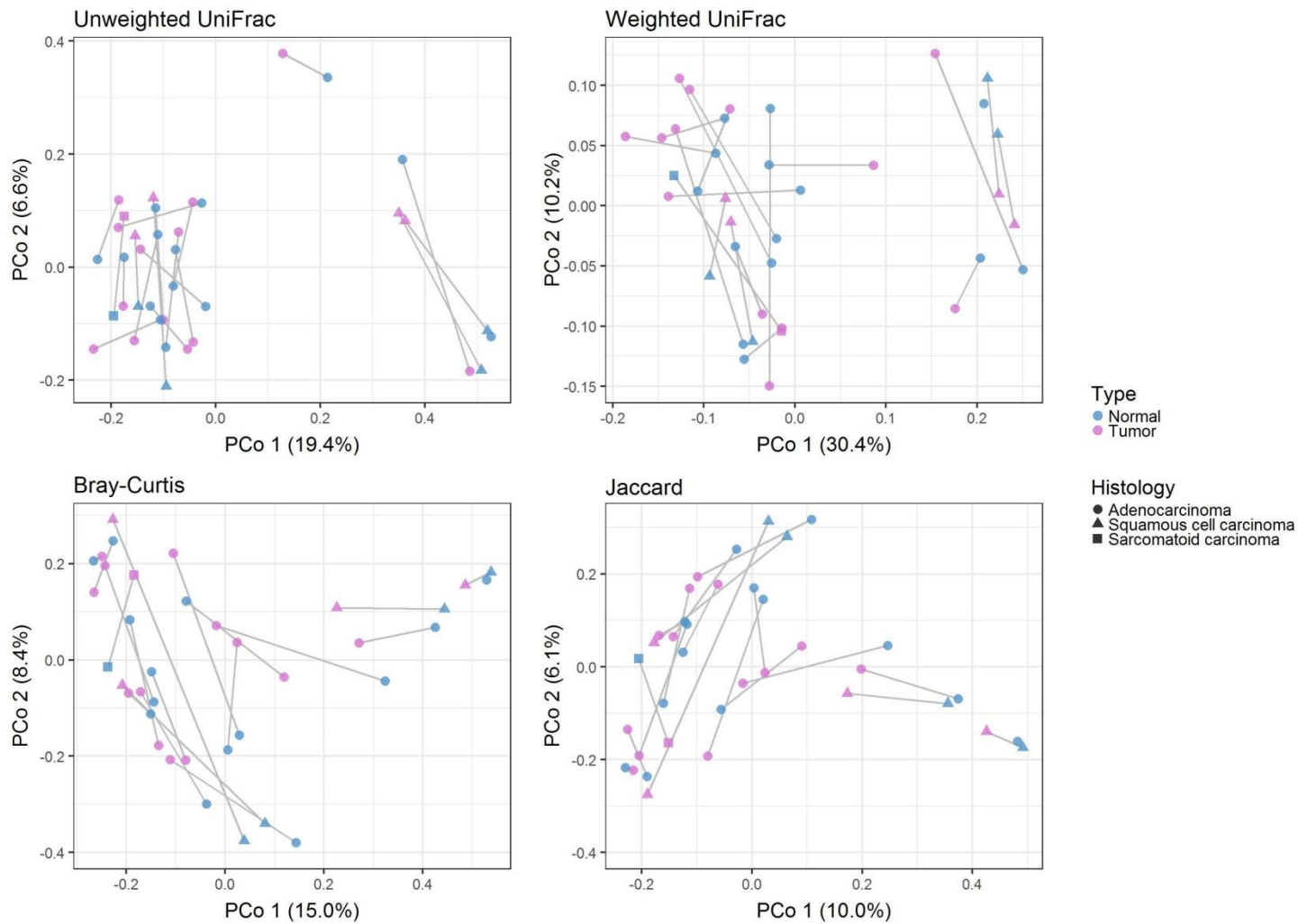
**Supplementary Figure 3.  $\beta$ -diversity in normal lung tissue and survival.** Principal coordinate analysis of the unweighted UniFrac distance, weighted UniFrac distance, and Jaccard index (non-rarefied and rarefied to an even depth of 63 sequence reads per sample), with samples annotated according to recurrence status, histology, and person days.



**Supplementary Figure 4.  $\alpha$ -diversity in tumor lung tissue and survival.** Boxplots show distribution of (a) number of OTUs and (d) Shannon index at an even depth of 63 sequence reads per sample in tumor lung tissue by disease status of patients ( $p$ -values are from Kruskal-Wallis tests). Kaplan-Meier plots show recurrence-free and disease-free survival curves for patients grouped in tertiles of (b-c) number of OTUs and (e-f) Shannon index at an even depth of 63 sequence reads per sample in tumor lung tissue ( $p$ -values are from log-rank tests for trend).



**Supplementary Figure 5.  $\beta$ -diversity in tumor lung tissue according to lung cancer recurrence status.** Principal coordinate analysis of the Bray-Curtis dissimilarity, with samples annotated according to recurrence status, histology, and person days. (a) non-rarefied, (b) rarefied to an even depth of 63 sequence reads per sample.



**Supplementary Figure 6.** Principal coordinate analysis of unweighted and weighted UniFrac distances, Bray-Curtis dissimilarity, and Jaccard index. Gray lines connecting normal (blue) and tumor (purple) samples indicate samples from the same individual.

**Supplementary Table 1.** Number of sequence reads<sup>a</sup> in paired lung tumor and normal samples.

ID	Recurrence or new primary	Histology	Extraction batch <sup>c</sup>	Number of sequence reads	
				Normal	Tumor
NYU 330	Recurrence	Adenocarcinoma	2	308	366
NYU 454	Recurrence	Adenocarcinoma	2	350	302
NYU 1832	Recurrence	Adenocarcinoma	2	357	131
NYU 1831	Recurrence	Adenocarcinoma	2	773	332
NYU 408	Recurrence	Adenocarcinoma	2	968	107
NYU 259	Recurrence	Adenocarcinoma	2	1106	471
NYU 586	New primary	Squamous cell carcinoma	2	1601	63
NYU 729	New primary	Squamous cell carcinoma	2	1617	271
NYU 777	Recurrence	Adenocarcinoma	1	1907	25140
NYU 735	New primary	Adenocarcinoma	2	2360	351
NYU 213	Recurrence	Adenocarcinoma	1	3306	NA
NYU 1675	Recurrence	Adenocarcinoma	1	4279	253
NYU 62	No	Adenocarcinoma	2	95	103
NYU 689	No	Adenocarcinoma	2	155	124
NYU 1763	No	Adenocarcinoma	2	218	3033
NYU 63	No	Adenocarcinoma	2	232	406
NYU 1405	No	Sarcomatoid carcinoma	2	245	175
NYU 1950	Missing <sup>b</sup>	Squamous cell carcinoma	1	2415	2379
NYU 38	Missing <sup>b</sup>	Squamous cell carcinoma	1	12666	262

<sup>a</sup>Number of reads determined after Deblur workflow.

<sup>b</sup>No follow up due to postoperative death.

<sup>c</sup>Samples were extracted/amplified in two batches (batch 1 = 10 samples and batch 2 = 28 samples); all samples were sequenced in one batch on the Illumina MiSeq.



**Supplementary Table 2.** Relation of lung tissue  $\alpha$ -diversity<sup>a</sup> with lung cancer recurrence-free and disease-free survival.

	Disease-free (n=5)	Median		P-value <sup>b</sup>	Recurrence-free survival		Disease-free survival	
		Recurrence (n=9)	New primary (n=3)		HR <sup>c</sup> (95% CI)	P-value <sup>c</sup>	HR <sup>c</sup> (95% CI)	P-value <sup>c</sup>
<b>Normal tissue</b>								
Number of OTUs	16.19	20.73	24.04	0.01	1.07 (0.99, 1.16)	0.08	1.09 (1.01, 1.17)	0.03
Shannon index	3.43	3.94	3.99	0.06	3.60 (1.10, 11.81)	0.03	3.97 (1.24, 12.70)	0.02
<b>Tumor tissue<sup>d</sup></b>								
Number of OTUs	12.27	16.37	14.65	0.10	1.04 (0.96, 1.14)	0.35	1.05 (0.96, 1.14)	0.28
Shannon index	3.26	3.55	3.29	0.11	1.87 (0.62, 5.67)	0.27	2.08 (0.72, 6.05)	0.18

<sup>a</sup>Number of OTUs and Shannon index were obtained by averaging over 100 iterations at a common depth of 63 sequences per sample.

<sup>b</sup>Kruskal-Wallis test comparing  $\alpha$ -diversity index between disease-free and recurrence or new primary groups.

<sup>c</sup> $\alpha$ -diversity index used as predictor in Cox proportional hazards model.

<sup>d</sup>One subject with recurrence was missing tumor data; results for normal tissue were similar when excluding that subject.

**Supplementary Table 3.** Sensitivity analyses for association of  $\alpha$ - and  $\beta$ -diversity with recurrence-free and disease-free survival.

	Adenocarcinoma only <sup>a</sup>		Batch 2 only <sup>b</sup>		Non-smokers only <sup>c</sup>		Stage I-II only <sup>a</sup>		$\geq 124$ reads/sample only <sup>d</sup>	
	RFS	DFS	RFS	DFS	RFS	DFS	RFS	DFS	RFS	DFS
<b>Normal tissue</b>										
Shannon index <sup>e</sup>	2.49 (0.13)	2.73 (0.09)	9.67 (0.05)	11.75 (0.03)	4.10 (0.04)	4.68 (0.02)	5.53 (0.02)	6.20 (0.01)	2.51 (0.08)	2.65 (0.06)
Weighted UniFrac <sup>f</sup>	0.09	0.08	0.10	0.07	0.01	0.02	0.02	0.03	0.05	0.07
Bray-Curtis <sup>f</sup>	0.10	0.03	0.15	0.04	0.09	0.01	0.11	0.02	0.19	0.13
<b>Tumor tissue</b>										
Shannon index <sup>e</sup>	1.82 (0.30)	1.93 (0.25)	2.04 (0.41)	2.55 (0.26)	2.65 (0.23)	2.79 (0.20)	1.62 (0.48)	1.91 (0.31)	1.65 (0.42)	1.66 (0.42)
Weighted UniFrac <sup>f</sup>	0.79	0.73	0.47	0.56	0.80	0.72	0.82	0.63	0.79	0.68
Bray-Curtis <sup>f</sup>	0.99	0.99	0.79	0.89	0.97	0.97	0.74	0.78	0.99	0.99

<sup>a</sup>Normal tissue n=14, tumor tissue n=13.

<sup>b</sup>Normal tissue n=14, tumor tissue n=14.

<sup>c</sup>Normal tissue n=15, tumor tissue n=14.

<sup>d</sup>Normal tissue n=16, tumor tissue n=13.

<sup>e</sup>Values are HR (p) from Cox model; Shannon index obtained by averaging over 100 iterations at a common depth of 63 (or 124 for  $\geq 124$  reads/sample analysis) sequences per sample.

<sup>f</sup>Values are P-values from the community-level test of association between the microbiota and survival times (MiRKAT-S). Bray-Curtis dissimilarity based on non-rarefied data.

**Supplementary Table 4.** Relation of lung tissue  $\beta$ -diversity<sup>a</sup> with lung cancer recurrence-free and disease-free survival.

	Recurrence-free survival P-value	Disease-free survival P-value
<b>Normal tissue (non-rarefied)</b>		
MiRKAT-S (unweighted UniFrac)	0.02	0.04
MiRKAT-S (weighted UniFrac)	0.03	0.04
MiRKAT-S (Bray-Curtis)	0.09	0.02
MiRKAT-S (Jaccard)	0.08	0.02
OMiSA	0.20	0.12
<b>Normal tissue (rarefied<sup>b</sup>)</b>		
MiRKAT-S (unweighted UniFrac)	0.005	0.02
MiRKAT-S (weighted UniFrac)	0.02	0.03
MiRKAT-S (Bray-Curtis)	0.22	0.09
MiRKAT-S (Jaccard)	0.19	0.06
OMiSA	0.07	0.07
<b>Tumor<sup>c</sup> tissue (non-rarefied)</b>		
MiRKAT-S (unweighted UniFrac)	0.70	0.66
MiRKAT-S (weighted UniFrac)	0.73	0.64
MiRKAT-S (Bray-Curtis)	0.95	0.95
MiRKAT-S (Jaccard)	0.95	0.95
OMiSA	0.64	0.48
<b>Tumor<sup>c</sup> tissue (rarefied<sup>b</sup>)</b>		
MiRKAT-S (unweighted UniFrac)	0.54	0.57
MiRKAT-S (weighted UniFrac)	0.60	0.56
MiRKAT-S (Bray-Curtis)	0.56	0.65
MiRKAT-S (Jaccard)	0.48	0.60
OMiSA	0.56	0.39

<sup>a</sup> $\beta$ -diversity was assessed by unweighted and weighted UniFrac phylogenetic distances, Bray-Curtis dissimilarity, and the Jaccard index. P-values were obtained by the community-level test of association between the microbiota and survival times (MiRKAT-S) and the optimal microbiome-based survival analysis test (OMiSA), all with 10,000 permutations.

<sup>b</sup>OTU table rarefied to an even depth of 63 sequences per sample prior to calculating distances.

<sup>c</sup>One subject with recurrence was missing tumor data; results for normal tissue were similar when excluding that subject.

**Supplementary Table 6.**  $\alpha$ - and  $\beta$ -diversity in relation to lung tumor and normal tissue type.

	Median		P-value <sup>a</sup>
	Normal (n=18)	Tumor (n=18)	
$\alpha$ -diversity <sup>b</sup>			
Number of OTUs	20.73	16.08	0.0001
Shannon index	3.94	3.49	<0.0001
$\beta$ -diversity <sup>c</sup> (non-rarefied)			
Unweighted UniFrac	NA	NA	0.17
Weighted UniFrac	NA	NA	0.26
Bray-Curtis	NA	NA	0.07
Jaccard	NA	NA	0.07
$\beta$ -diversity <sup>c</sup> (rarefied <sup>d</sup> )			
Unweighted UniFrac	NA	NA	0.16
Weighted UniFrac	NA	NA	0.58
Bray-Curtis	NA	NA	0.43
Jaccard	NA	NA	0.45

<sup>a</sup>Wilcoxon signed-rank paired test was used for comparing  $\alpha$ -diversity index between paired tumor and normal tissue; permutational multivariate analysis of variance with tissue pair as strata and 10,000 permutations was used for  $\beta$ -diversity.

<sup>b</sup>Number of OTUs and Shannon index were obtained by averaging over 100 iterations at a common depth of 63 sequences per sample.

<sup>c</sup> $\beta$ -diversity was assessed by unweighted and weighted UniFrac phylogenetic distances, Bray-Curtis dissimilarity, and the Jaccard index.

<sup>d</sup>OTU table rarefied to an even depth of 63 sequences per sample prior to calculating distances.