

eTable 1: Comparison of Alpha Diversity p-trend results from conditional and unconditional logistic regression models, among lung cancer cases and controls, in two prospective cohorts of never-smokers.

Alpha diversity index metric	All 113 case-control pairs		Excluding 3 pairs unsuccessfully matched on Antibiotic Use	
	Conditional Logistic Regression	Unconditional Logistic Regression¥	Conditional Logistic Regression	Unconditional Logistic Regression¥
Shannon	0.07	0.04	0.08	0.04
Shannon*	0.02	0.02	0.03	0.02
Simpson	0.04	0.03	0.05	0.03
Simpson*	0.03	0.03	0.04	0.03
Observed species	0.96	0.64	0.97	0.55
Observed species*	0.07	0.05	0.08	0.05

¥adjusted for matching factors (sex, age, sample collection time, menopause status (for women), and antibiotic use); *adjusted for sequencing reads per sample.

eTable 2: Alpha Diversity between lung cancer cases and controls, in two prospective cohorts of never-smokers[†].

α -diversity metric	Medium vs Low Diversity				High vs Low Diversity				P _{trend}
	Odds Ratio	95%CI		P	Odds Ratio	95%CI		P	
		Lower	Upper			Lower	Upper		
All subjects (114 controls vs 113 cases)									
Shannon	0.62	0.32	1.21	0.16	0.73	0.37	1.46	0.38	0.05
Simpson	0.55	0.28	1.07	0.08	0.74	0.38	1.44	0.37	0.04
Observed Species	0.81	0.41	1.60	0.55	1.06	0.55	2.08	0.85	0.64
Only subjects with no antibiotic use in 7 days prior to sample collection (103 controls vs 99 cases)[†]									
Shannon	0.59	0.29	1.20	0.15	0.73	0.36	1.49	0.39	0.07
Simpson	0.47	0.23	0.96	0.04	0.71	0.35	1.42	0.33	0.04
Observed Species	0.86	0.42	1.77	0.69	1.22	0.60	2.46	0.58	0.96
Only subjects with no antibiotic use in 7 days prior to sample collection, and sample collected ≥ 2 years before cancer (103 controls vs 86 cases)[†]									
Shannon	0.61	0.29	1.28	0.19	0.73	0.34	1.57	0.43	0.09
Simpson	0.47	0.22	1.02	0.05	0.72	0.34	1.52	0.39	0.06
Observed Species	0.81	0.38	1.73	0.59	1.11	0.53	2.33	0.79	0.84

[†]Adjusted for sex, age, sample collection time, menopause status (among women), education, and antibiotic use (when appropriate).

eTable 3: Alpha Diversity, stratified by sex/cohort, between lung cancer cases and controls, in two prospective cohorts of never-smokers†.

α-diversity metric†	Men / Shanghai Men's Health Study									Women / Shanghai Women's Health Study								
	Medium vs Low Diversity				High vs Low Diversity					Medium vs Low Diversity				High vs Low Diversity				
	OR ¹	95%CI		P	OR	95%CI		P	P _{trend}	OR	95% CI		P	OR	95% CI		P	P _{trend}
LCI ²		UCI ³	LCI			UCI	LCI				UCI	LCI			UCI			
All subjects																		
24 controls vs 23 cases										90 controls vs 90 cases								
Shannon	0.41	0.08	1.99	0.27	0.31	0.04	2.67	0.29	0.23	0.60	0.27	1.30	0.19	0.80	0.38	1.71	0.57	0.10
Simpson	0.48	0.09	2.44	0.38	1.23	0.23	6.63	0.81	0.23	0.48	0.22	1.04	0.06	0.65	0.31	1.38	0.26	0.06
Observed species	1.20	0.28	5.10	0.81	1.08	0.15	7.88	0.94	0.55	0.79	0.36	1.75	0.57	1.14	0.55	2.39	0.72	0.98
Only subjects with no antibiotic use in 7 days prior to sample collection																		
24 controls vs 20 cases										79 controls vs 79 cases								
Shannon	0.41	0.08	1.99	0.27	0.31	0.04	2.67	0.29	0.23	0.58	0.25	1.32	0.19	0.82	0.37	1.81	0.62	0.15
Simpson	0.48	0.09	2.44	0.38	1.23	0.23	6.63	0.81	0.23	0.42	0.18	0.97	0.04	0.63	0.29	1.38	0.25	0.07
Observed species	1.20	0.28	5.10	0.81	1.08	0.15	7.88	0.94	0.55	0.83	0.35	1.95	0.67	1.33	0.61	2.91	0.48	0.65

†Adjusted for age, sample collection time, menopause status (for women), education, and antibiotic use (excluding adjustment for antibiotic use in the stratified analysis); ¹Odds Ratio, ²Lower bound of 95% confidence interval (CI); ³Upper bound of 95% confidence interval (CI).

eTable 4: Beta Diversity¹ between lung cancer cases and controls, in two prospective cohorts of never-smokers †.

Model	P Value										
	PCoA1	PCoA2	PCoA3	PCoA4	PCoA5	PCoA6	PCoA7	PCoA8	PCoA9	PCoA10	MiRKAT
All (114 controls vs 113 cases)	0.53	0.06	0.45	0.62	0.93	0.19	0.41	0.44	0.07	0.98	0.30
Male (24 controls vs 23 cases)	0.81	0.13	0.27	0.26	0.63	0.68	0.89	0.68	0.66	0.30	0.61
Female (90 controls vs 90 cases)	0.52	0.18	0.13	0.38	0.97	0.09	0.20	0.36	0.08	0.52	0.22
Age ≤ 61 (59 controls vs 59 cases)	0.89	0.30	0.83	0.78	0.49	0.98	0.78	0.62	0.02	0.99	0.88
Age > 61 (55 controls vs 54 cases)	0.23	0.16	0.82	0.89	0.88	0.07	0.27	0.64	0.77	1.00	0.35
No Antibiotic (103 controls vs 99 cases)	0.64	0.03	0.46	0.68	0.71	0.17	0.55	0.58	0.07	0.85	0.19

¹ Bray Curtis distance matrix

† MiRKAT tests the association between each covariate and each beta diversity, adjusted for sex, age, sample collection time, menopause status (for women), education, and antibiotic use (menopause, sex, and antibiotic use excluded in stratified analysis if not applicable). The percentage of variance explained by each PCoA: PCoA1 (13.7 %), PCoA2 (11.6 %), PCoA3 (6.7 %), PCoA4 (5.8 %), PCoA5 (5.0 %), PCoA6 (3.4 %), PCoA7 (3.1 %), PCoA8 (2.9 %), PCoA9 (2.5 %), and PCoA10 (2.3 %).

Table 5: Abundance of taxa associated with risk of lung cancer[†] in two prospective cohorts of never-smokers, stratified by antibiotic use.

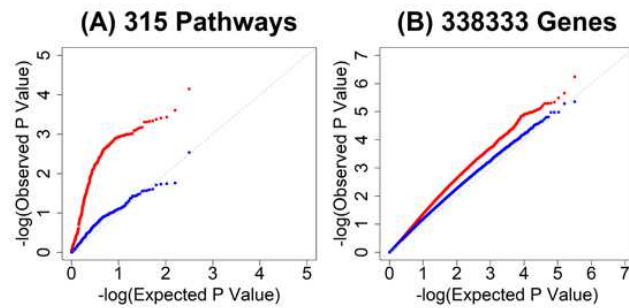
Taxa [‡]	Medium vs Low Abundance				High vs Low Abundance				P _{trend}	Medium vs Low Abundance				High vs Low Abundance				
	OR ¹	95%CI		P	OR	95%CI		P		OR	95% CI		OR	95% CI		P	P _{trend}	
		LCI ²	UCI ³			LCI	UCI				LCI	UCI		LCI	UCI			
All subjects										Only subjects with no antibiotic use in 7 days prior to sample collection								
114 controls vs 113 cases										103 controls vs 99 cases								
<i>Spirochaetes</i> (p), <i>Spirochaetia</i> (c)	0.61	0.32	1.18	0.14	0.42	0.21	0.85	0.02	0.01	0.68	0.34	1.37	0.28	0.40	0.19	0.85	0.02	0.02
<i>Bacteroidetes</i> (p), <i>Bacteroidetes</i> (c)	0.66	0.35	1.25	0.21	0.31	0.15	0.64	0.002	0.002	0.74	0.38	1.45	0.38	0.25	0.11	0.57	<0.01	<0.01
<i>Bacteroidetes</i> (p), <i>Bacteroidetes</i> (c), <i>Bacteroidetes</i> (o)	0.66	0.35	1.25	0.21	0.31	0.15	0.64	0.002	0.002	0.74	0.38	1.45	0.38	0.25	0.11	0.57	<0.01	<0.01
<i>Bacteroidetes</i> (p), <i>Bacteroidetes</i> (c), <i>Bacteroidetes</i> (o), <i>Bacteroidetes</i> (f)	0.66	0.35	1.25	0.21	0.31	0.15	0.64	0.002	0.002	0.74	0.38	1.45	0.38	0.25	0.11	0.57	<0.01	<0.01
<i>Firmicutes</i> (p), <i>Bacilli</i> (c)	1.49	0.73	3.08	0.28	2.40	1.18	4.87	0.02	0.01	2.01	0.94	4.29	0.07	2.76	1.30	5.85	<0.01	<0.01
<i>Firmicutes</i> (p), <i>Bacilli</i> (c), <i>Lactobacillales</i> (o)	2.15	1.03	4.47	0.04	3.25	1.58	6.70	0.001	0.002	2.60	1.20	5.62	0.02	3.61	1.67	7.78	<0.01	<0.01

[†]Adjusted for age, sex, sample collection time, menopause status (for women), education, and antibiotic use (excluding adjustment for antibiotic use in the stratified analysis when not applicable);

¹Odds Ratio, ²Lower bound of 95% confidence interval (CI); ³Upper bound of 95% confidence interval (CI).

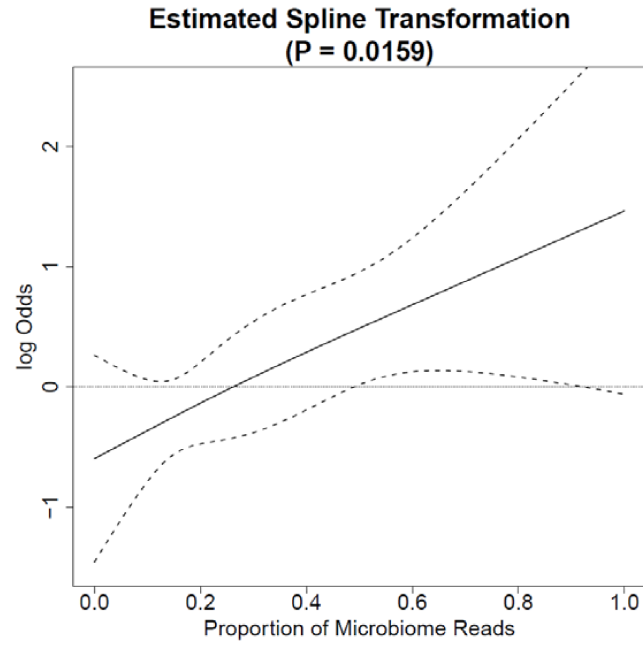
[‡]Indicated by phylogenetic tree attributes (p=phylum; c=class; o=order; f=family)

eFigure 1: QQ plots of P-values from logistic models of (A) relative abundance of 315 pathways and (B) Presence/absence of 338,333 genes with a prevalence of >0.1 associated with risk of lung cancer in two prospective cohorts of never-smokers[†].



[†]Red points: Adjusted for sex, age, sample collection time, menopause status (for women), education, and antibiotic use; Blue points: additionally adjusted the microbiota reads rate.

eFigure 2: Proportion of microbiota reads associated with risk of lung cancer in two prospective cohorts of never-smokers[†].



[†]Adjusted for sex, age, sample collection time, menopause status (for women), education, and antibiotic use