SUPPLEMENTAL MATERIAL

A) Seasonal time frames for recruitment of cross-sectional and longitudinal subjects and controls

Cross-sectional and longitudinal cohorts, as well as, all controls in both cohorts were all enrolled between Fall 2012 and Fall of 2015, covering three consecutive years and RSV seasons (August - April).

Developmental microbiota changes over time were controlled for by age matched in longitudinal cohort and matched group mean and standard deviation in the cross-sectional cohort. Average age for cases were 2.45 (± 1.24) months and 3.18 (± 2.24) months for longitudinal and cross-sectional respectively. Average age for controls were 2.58 (± 1.23) months and 3.01 (± 2.19) months for longitudinal and cross-sectional respectively. Thus, ages are similar across both cohorts.

Specifically, for the longitudinal cohort control samples:

1. 78% of healthy control samples were collected during one of the three RSV seasons (August - April)
2. 67% of healthy control samples were collected during the same season as sick samples Fall (Sept) through Spring (May)
3. 50% of healthy control samples were collected during the same months as sick samples (Oct - Apr)

Specifically, for the cross-sectional cohort control samples:

1. 83% of RSV negative control samples were collected during one of the three RSV seasons (August - April)
2. 78% of RSV negative control samples were collected during the same season as RSV positive samples Fall (Sept) through Spring (May)
3. 50% of RSV negative control samples were collected during the same months as RSV positive samples (Nov - Apr)

Longitudinal Cohort

Percentage of Samples within an RSV Season

78% of Healthy Control samples are during and RSV season (August – April).

Cohort	Non-Season	Season 1	Season 2	Season 3
Healthy	22%	0%	50%	28%
Sick	0%	58%	42%	0%

Percentage of samples within a given season

Seasons defined as:

- Spring is March 1 – May 31
- Summer is June 1 – August 31
- Fall is September 1 – November 30
- Winter is December 1 – February 28

67% of Healthy Control samples were collected during Fall through Spring (Sept – May).

Cohort	Spring	Summer	Fall	Winter
Healthy	11%	33%	36%	19%
Sick	33%	0%	8%	58%

Percentage of samples within each month

50% of Healthy Control samples were collected during the same months as Sick samples (Oct – Apr).

Cohort	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Healthy	3%	3%	3%	6%	3%	11%	8%	14%	14%	8%	14%	14%
Sick	28%	14%	28%	6%	0%	0%	0%	0%	0%	3%	6%	17%

Cross-sectional Cohort

Percentage of Samples within an RSV Season

83% of RSV Negative Control samples are during and RSV season (August – April).

RSV	Non-Season	Season 1	Season 2	Season 3	Season 4
Negative	17%	12%	41%	23%	8%
Positive	0%	12%	44%	44%	0%

Percentage of samples within a given season

Seasons defined as:

- Spring is March 1 – May 31
- Summer is June 1 – August 31
- Fall is September 1 – November 30
- Winter is December 1 – February 28

78% of RSV Negative Control samples were collected during Fall through Spring (Sept – May).

RSV	Spring	Summer	Fall	Winter
Negative	11%	22%	27%	40%
Positive	18%	0%	2%	80%

Percentage of samples within each month

50% of RSV Negative Control samples were collected during the same months as RSV Positive samples (Nov – Apr).

| RSV	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Negative	22%	8%	2%	6%	3%	4%	10%	8%	8%	15%	5%	11%
Positive	25%	36%	17%	1%	0%	0%	0%	0%	0%	0%	2%	19%

B) Selection criteria for cross-sectional and longitudinal subjects and controls

Control samples and subjects for the cross-sectional cohort were selected by age group mean and standard deviation to minimize population level differences in age at the time of sampling, gestational age at birth, and mode of delivery.

For the longitudinal cohort, samples were collected from the RSV group at approximately one month of age, during acute RSV infection, and approximately one month after illness, and at corresponding timepoints from the healthy controls. Control subjects were selected to match on an individual basis by sex, mode of delivery, gestational age at birth, and samples selected to match by age.

For both cohorts, subjects were eligible as controls if they had no respiratory illness between birth and at least ten days after the last sample and had not been exposed to antibiotics within 30 days of enrollment. Patients and controls in the longitudinal cohort were not exposed to antibiotics within 30 days of the time points studied.

Supplemental Figure 1
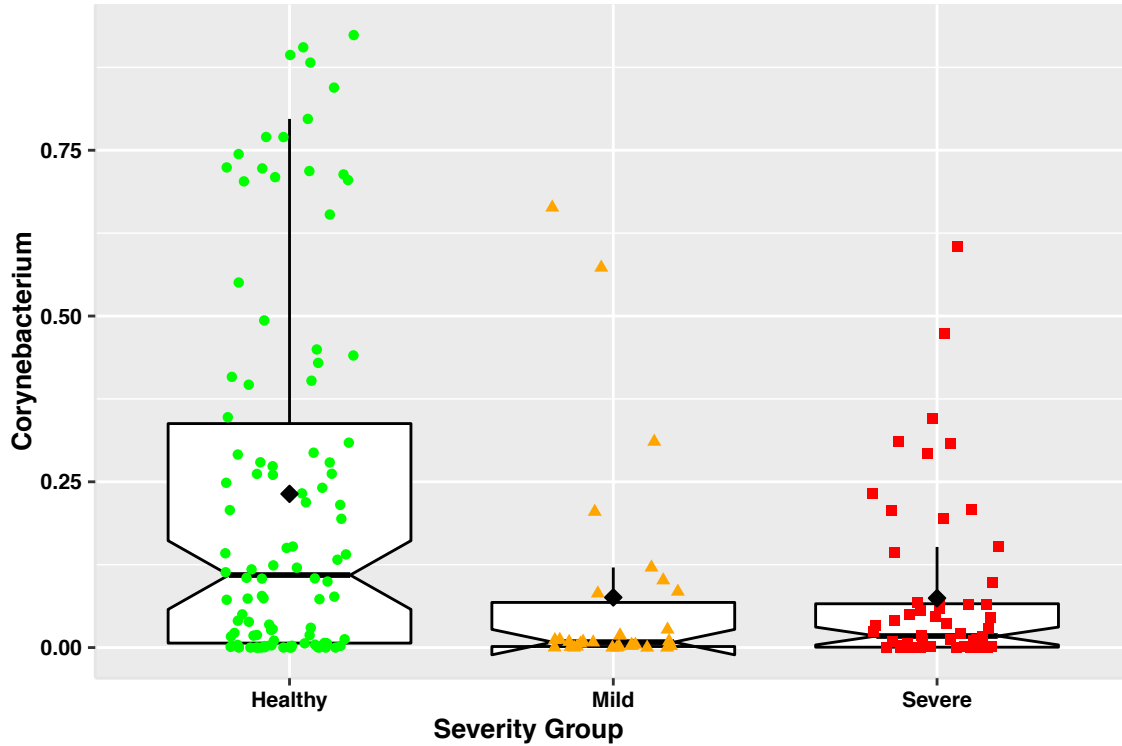
Supplemental Figure 1. Relative abundances (y-axes) of Corynebacterium significantly associated with more severe disease in the cross-sectional cohort, with samples grouped by dichotomizing illness based on severity into mild and severe groups (x-axes), using a severity score threshold of 3.5. Each colored point represents the relative abundance of Corynebacterium in a single individual, with columns (left to right), shapes (circle, triangle, square), and colors (green, orange, red) distinguishing between healthy, mild illness, and severe illness groups, respectively. The black diamonds indicate the group mean for each group. Box plots are overlaid on each group, centered on the group median, with notches indicating an approximately 95% confidence interval, boxes indicating boundaries of the first and third quartiles, and whiskers extending to the largest and smallest values no further than 1.5*(inter-quartile range) from the boxes. Points beyond the whiskers are commonly considered outliers, which in this case would suggest that many of the observed associations between Corynebacterium relative abundance and illness severity are driven primarily by outliers, or that Corynebacterium abundance in severely ill infants comprises more than one underlying distribution.

Supplemental Figure 2
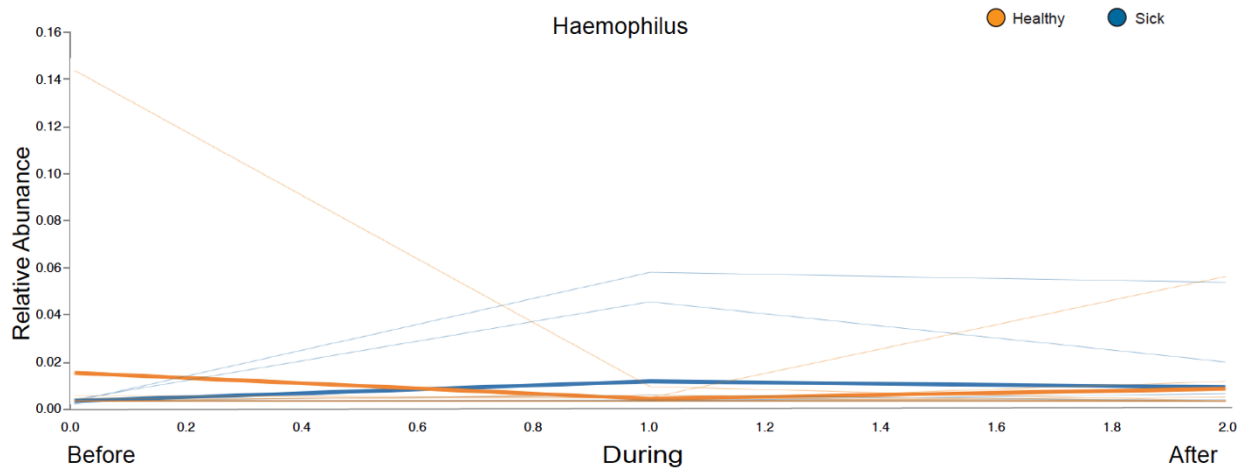
Supplemental Figure 2. Relative abundances (y-axes) of *Hemophilus* at all three time points (x-axes) in the longitudinal cohort. Each thin line corresponds to the abundance of *Haemophilus* in a particular individual, while the thick lines show the mean abundance of each group at each time point. Members of the healthy group are orange and members of the group that developed infection are blue. The temporal trends with respect to illness indicate that *Haemophilus* is significantly more abundant in the infected group during (p < 0.001) illness and minimally abundant in both groups before and after.