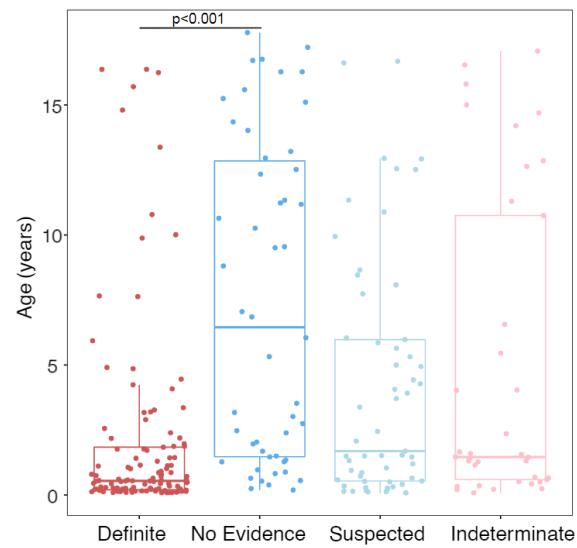
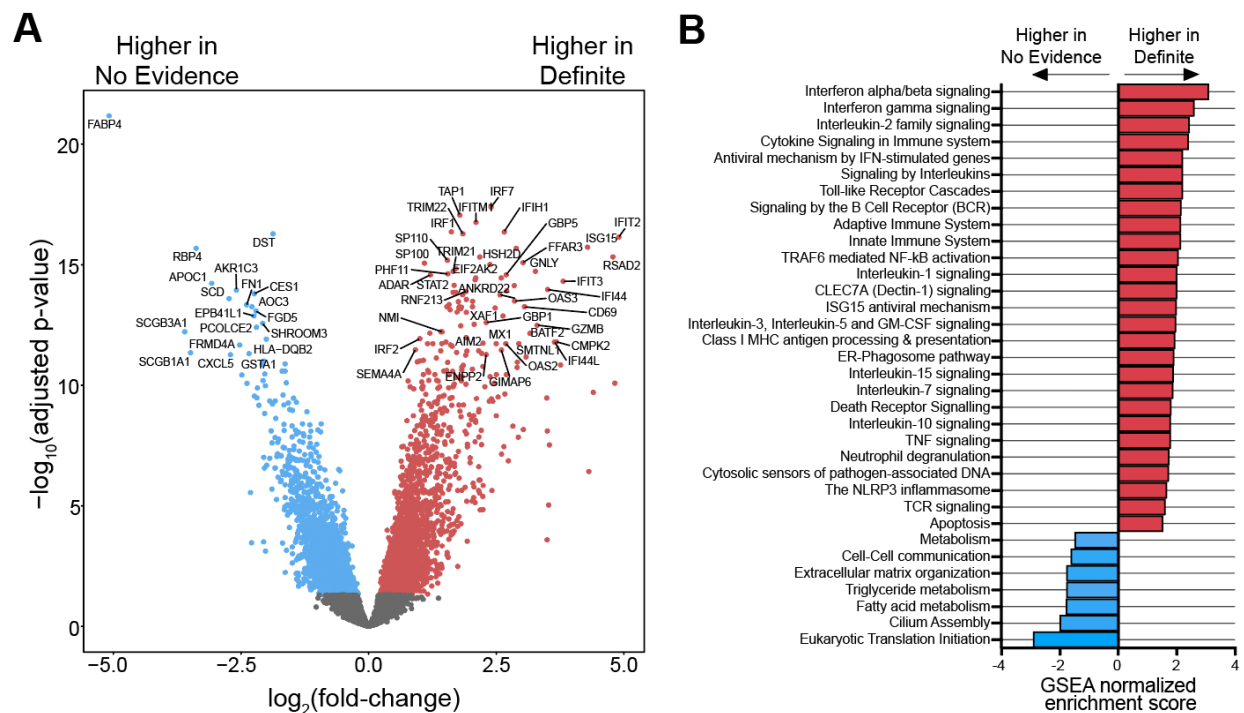


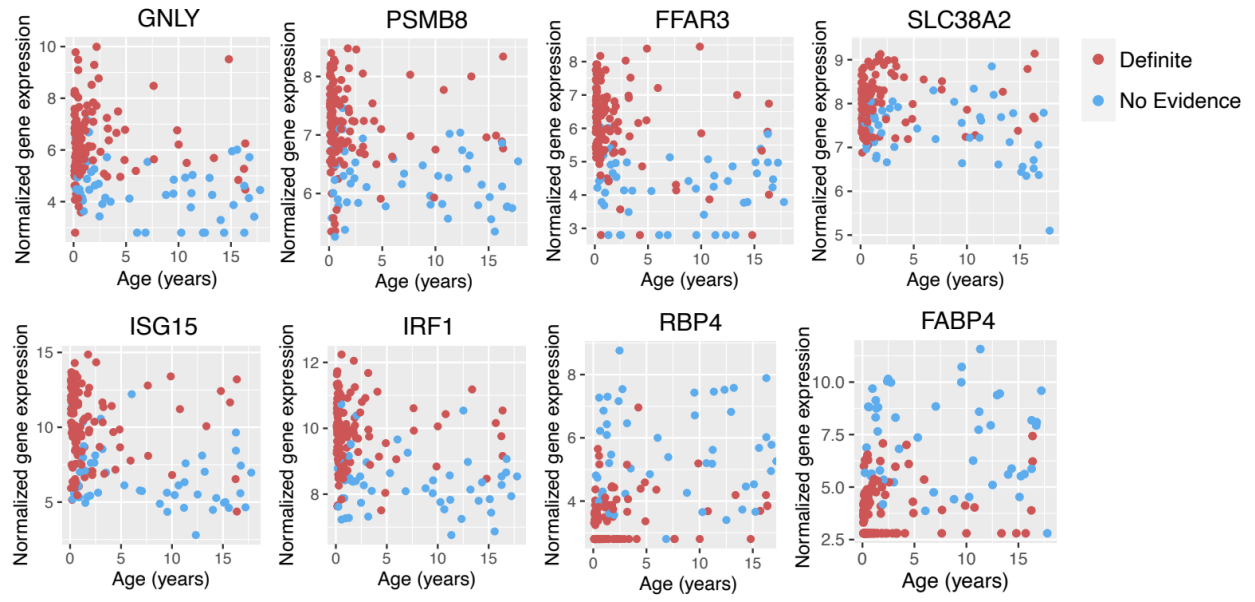
Supplemental Figures



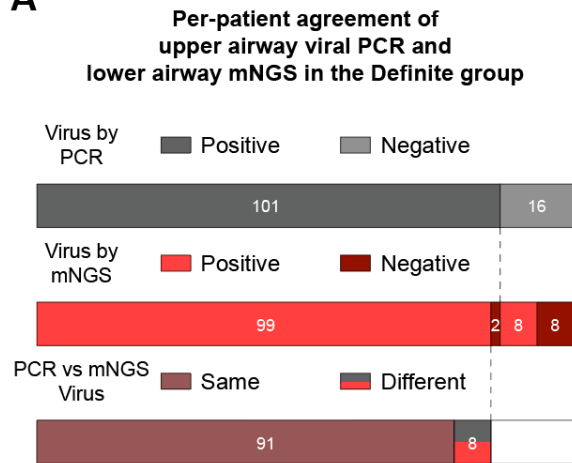
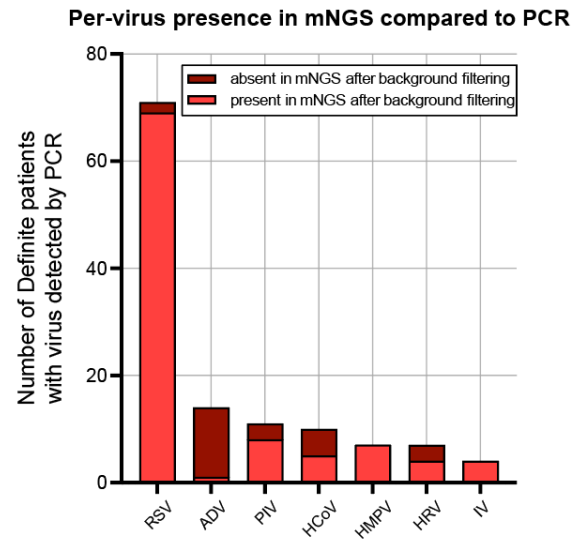
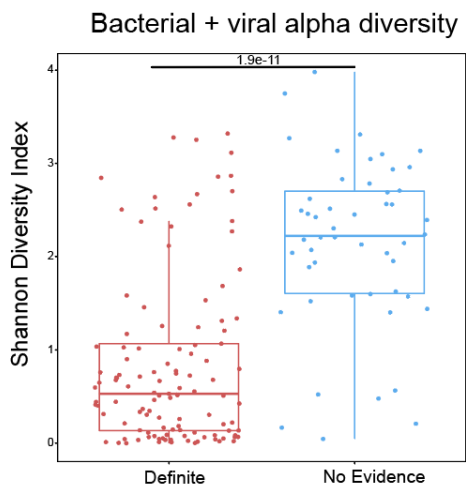
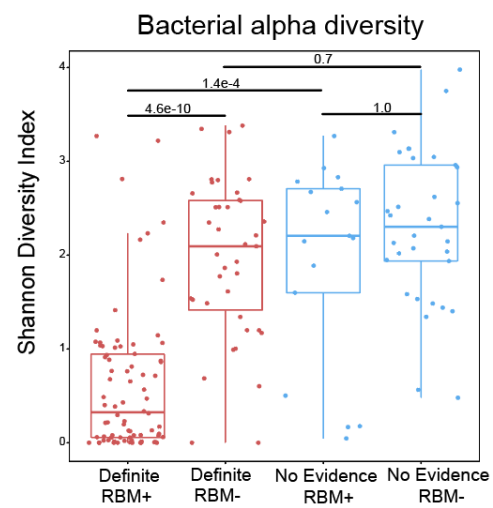
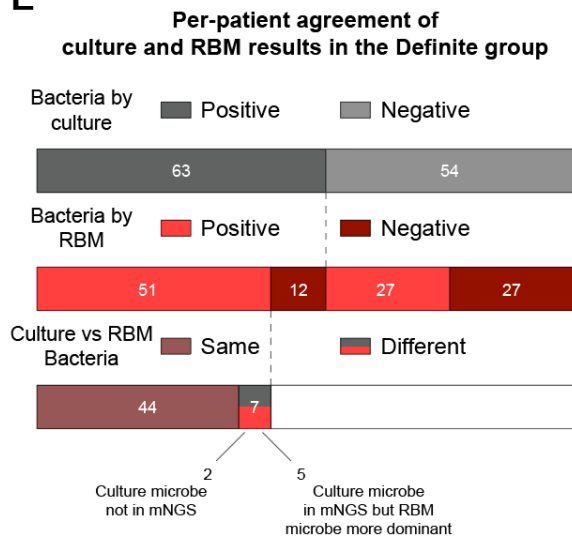
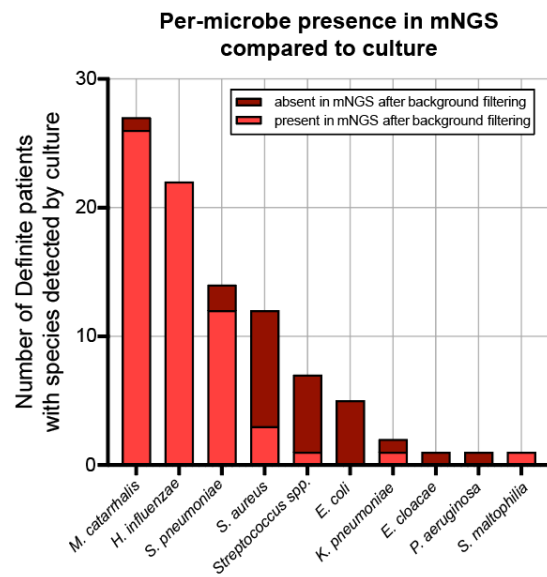
Supplemental Figure 1: Age distribution across the four LRTI status groups. P-value for the comparison between Definite and No Evidence patients was calculated using a Mann-Whitney test.



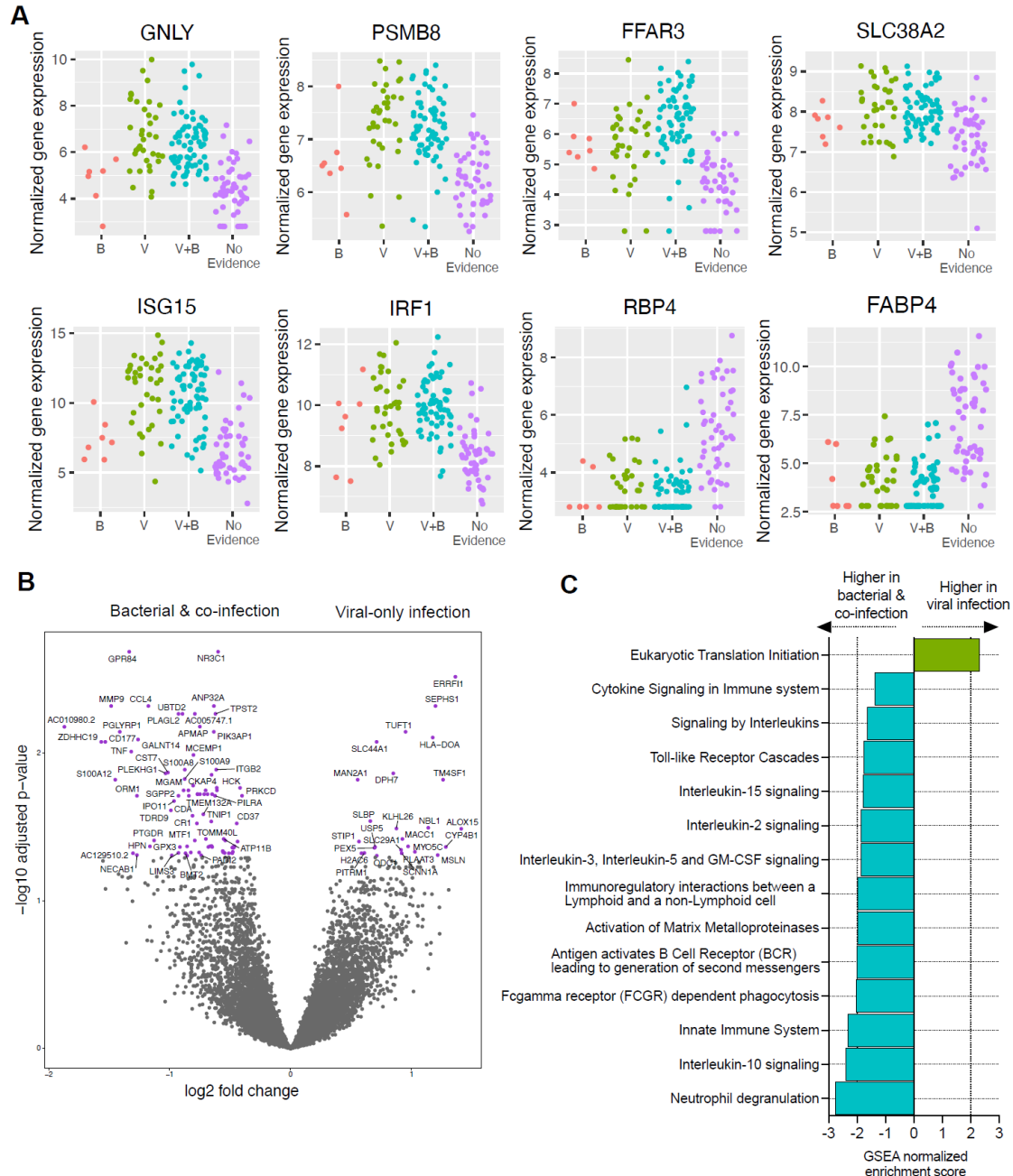
Supplemental Figure 2: A) Volcano plot highlighting genes differentially expressed (DE) between Definite and No Evidence patients. Colored genes reached statistical significance (adjusted p-value < 0.05). **B)** Normalized enrichment scores of selected REACTOME pathways that reached statistical significance (adjusted p-value < 0.05) in the GSEA based on the DE results between Definite and No Evidence patients.



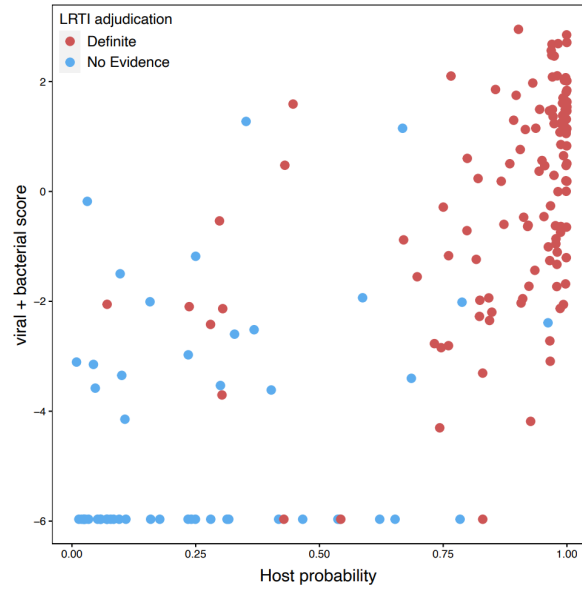
Supplemental Figure 3: Expression of eight host classifier genes as a function of age in Definite (red) and No Evidence (blue) patients.

A**B****C****D****E****F**

Supplemental Figure 4: **A)** Diagram depicting the agreement at the patient level between clinical upper respiratory PCR viral testing and lower airway mNGS detection after background filtering in the Definite group. Agreement between the two methods in a patient was defined as at least one virus identified by both. **B)** Bar plot showing the number of cases of each virus detected by clinical upper respiratory PCR testing and the proportion that was also present by mNGS after background filtering. RSV, respiratory syncytial virus; ADV, adenovirus; PIV, parainfluenza virus; HCoV, human coronavirus; HMPV, human metapneumovirus; HRV, human rhinovirus; IV, influenza virus. **C)** Boxplots of bacterial+viral microbiome alpha diversity, measured by the Shannon index, in Definite and No Evidence patients. Horizontal lines denote the median, box hinges represent the interquartile range (IQR), and whiskers extend to the most extreme value no greater than 1.5*IQR from the hinges. **D)** Boxplots of bacterial-only alpha diversity measured by the Shannon index. Definite and No Evidence patients are split by whether a potential pathogen was identified by the RBM. P-values in C) and D) were calculated by a Mann-Whitney test with Bonferroni correction. **E)** Diagram depicting the agreement at the patient level between clinical culture and the results of the RBM in the Definite group. Agreement between the two methods in a patient was defined as at least one species identified by both. **F)** Bar plot showing the number of cases of each species detected by clinical culture and the proportion that was also present by mNGS after background filtering. *Streptococcus spp.*, *Streptococcus* species other than *S. pneumoniae*.



Supplemental Figure 5: A) Expression of eight host classifier genes in Definite patients with only bacterial pathogens identified by the RBM (B; n=7), only viral pathogens detected by mNGS (V; N=36), viral and bacterial pathogens (V+B; n=71), and the No Evidence patients (n=50) for comparison. Three patients from the Definite group are not shown because they did not have any pathogens identified by mNGS. One No Evidence sample in the plot of *SLC38A2* was omitted since it was an extreme outlier. **B)** Volcano plot highlighting genes differentially expressed (DE) between Definite patients with any bacterial infection (bacterial-only + co-infection) and viral-only infection. Genes colored in purple reached statistical significance (adjusted p-value < 0.05). **C)** Normalized enrichment scores of selected REACTOME pathways that reached statistical significance (adjusted p-value < 0.05) in the GSEA based on the DE results between Definite patients with any bacterial infection (bacterial-only + co-infection) and viral-only infection.



Supplemental Figure 6: Scatterplot of the host LRTI probability (x-axis) and the sum of the log₁₀-transformed microbial scores (y-axis) in the Definite and No Evidence patients.

Supplemental Tables

Supplemental Table 1: Host genes selected by lasso logistic regression in each of the 5 cross-validation train/test splits, and the area under the receiver operating characteristic curve (AUC) of a random forest classifier using the selected genes.

Test fold	Gene ID	Gene symbol	Regression coefficient	AUC
1	(Intercept)	NA	-2.0280997	0.996
1	ENSG00000115523	<i>GNLY</i>	0.40435424	
1	ENSG00000114737	<i>CISH</i>	0.29736499	
1	ENSG00000185897	<i>FFAR3</i>	0.2334893	
1	ENSG00000134294	<i>SLC38A2</i>	0.23225034	
1	ENSG00000204264	<i>PSMB8</i>	0.14737796	
1	ENSG00000133101	<i>CCNA1</i>	0.13698008	
1	ENSG00000125347	<i>IRF1</i>	0.08788657	
1	ENSG00000152766	<i>ANKRD22</i>	0.01618487	
1	ENSG00000103356	<i>EARS2</i>	-0.005208	
1	ENSG00000272168	<i>CASC15</i>	-0.0129449	
1	ENSG00000163710	<i>PCOLCE2</i>	-0.0141204	
1	ENSG00000196139	<i>AKR1C3</i>	-0.0145091	
1	ENSG00000163735	<i>CXCL5</i>	-0.0202633	
1	ENSG00000164631	<i>ZNF12</i>	-0.0296153	
1	ENSG00000162929	<i>KIAA1841</i>	-0.031873	
1	ENSG00000102962	<i>CCL22</i>	-0.0470626	
1	ENSG00000196305	<i>IARS1</i>	-0.0541444	
1	ENSG00000259094	<i>AC013457.1</i>	-0.0642868	
1	ENSG00000203865	<i>ATP1A1-AS1</i>	-0.0785439	
1	ENSG00000115414	<i>FN1</i>	-0.0963792	
1	ENSG00000132300	<i>PTCD3</i>	-0.1094826	
1	ENSG00000273173	<i>SNURF</i>	-0.1236364	
1	ENSG00000138207	<i>RBP4</i>	-0.1867265	
1	ENSG00000170323	<i>FABP4</i>	-0.2260672	
1	ENSG00000182141	<i>ZNF708</i>	-0.2407744	
2	(Intercept)	NA	-0.4183313	0.953
2	ENSG00000134294	<i>SLC38A2</i>	0.21115084	
2	ENSG00000185897	<i>FFAR3</i>	0.17219249	
2	ENSG00000115523	<i>GNLY</i>	0.15221918	
2	ENSG00000204264	<i>PSMB8</i>	0.0974774	
2	ENSG00000152766	<i>ANKRD22</i>	0.06800369	
2	ENSG00000272821	<i>U62317.2</i>	0.05245633	
2	ENSG00000187608	<i>ISG15</i>	0.03953433	
2	ENSG00000162929	<i>KIAA1841</i>	-0.0224986	
2	ENSG00000196139	<i>AKR1C3</i>	-0.0991982	
2	ENSG00000170323	<i>FABP4</i>	-0.1720699	

2	ENSG00000253729	<i>PRKDC</i>	-0.2010562	
2	ENSG00000138207	<i>RBP4</i>	-0.2940594	
3	(Intercept)	NA	0.27758103	0.986
3	ENSG00000185507	<i>IRF7</i>	0.29760005	
3	ENSG00000133101	<i>CCNA1</i>	0.22848885	
3	ENSG00000115523	<i>GNLY</i>	0.20275662	
3	ENSG00000185897	<i>FFAR3</i>	0.09319602	
3	ENSG00000134294	<i>SLC38A2</i>	0.05523801	
3	ENSG00000272821	<i>U62317.2</i>	0.03959668	
3	ENSG00000196189	<i>SEMA4A</i>	0.03111383	
3	ENSG00000136231	<i>IGF2BP3</i>	0.01793655	
3	ENSG00000114737	<i>CISH</i>	0.01066616	
3	ENSG00000117143	<i>UAP1</i>	-0.0043278	
3	ENSG00000163735	<i>CXCL5</i>	-0.0087013	
3	ENSG00000113068	<i>PFDN1</i>	-0.0087115	
3	ENSG00000149021	<i>SCGB1A1</i>	-0.0088169	
3	ENSG00000232629	<i>HLA-DQB2</i>	-0.0158939	
3	ENSG00000164631	<i>ZNF12</i>	-0.0351792	
3	ENSG00000273173	<i>SNURF</i>	-0.0502438	
3	ENSG00000196139	<i>AKR1C3</i>	-0.0532949	
3	ENSG00000170324	<i>FRMPD2</i>	-0.0607793	
3	ENSG00000259094	<i>AC013457.1</i>	-0.0637096	
3	ENSG00000272660	<i>AC090425.2</i>	-0.1024448	
3	ENSG00000008226	<i>DLEC1</i>	-0.1276673	
3	ENSG00000272168	<i>CASC15</i>	-0.1499889	
3	ENSG00000170323	<i>FABP4</i>	-0.4076148	
4	(Intercept)	NA	-5.181784	0.954
4	ENSG00000175073	<i>VCPIP1</i>	0.50212823	
4	ENSG00000115523	<i>GNLY</i>	0.23413097	
4	ENSG00000135604	<i>STX11</i>	0.23226792	
4	ENSG00000168394	<i>TAP1</i>	0.13854581	
4	ENSG00000185897	<i>FFAR3</i>	0.13639215	
4	ENSG00000185885	<i>IFITM1</i>	0.07866599	
4	ENSG00000133106	<i>EPSTI1</i>	0.05899777	
4	ENSG00000158769	<i>F11R</i>	-0.0055638	
4	ENSG00000163710	<i>PCOLCE2</i>	-0.0206461	
4	ENSG00000149021	<i>SCGB1A1</i>	-0.0263191	
4	ENSG00000182141	<i>ZNF708</i>	-0.1006603	
4	ENSG00000138207	<i>RBP4</i>	-0.1413799	
4	ENSG00000170323	<i>FABP4</i>	-0.2005039	
5	(Intercept)	NA	0.5726104	0.967
5	ENSG00000204264	<i>PSMB8</i>	0.27278452	
5	ENSG00000115523	<i>GNLY</i>	0.0762971	

5	ENSG00000175073	<i>VCPIP1</i>	0.05579582	
5	ENSG00000185897	<i>FFAR3</i>	0.04184826	
5	ENSG00000188820	<i>CALHM6</i>	0.03110363	
5	ENSG00000133106	<i>EPSTI1</i>	0.02410902	
5	ENSG00000187608	<i>ISG15</i>	0.00687972	
5	ENSG00000163710	<i>PCOLCE2</i>	-0.0191247	
5	ENSG00000151914	<i>DST</i>	-0.029031	
5	ENSG00000163735	<i>CXCL5</i>	-0.1149234	
5	ENSG00000170323	<i>FABP4</i>	-0.389078	

Supplemental Table 2: Genes selected for the final host classifier by lasso logistic regression applied to all the Definite and No Evidence patients, with their regression coefficients. The number of times the gene was selected across the 5 cross-validation (CV) splits is also indicated.

Gene ID	Gene symbol	Gene product	Regression coefficient	Times selected in CV
ENSG00000115523	<i>GNLY</i>	Granulysin	0.257	5
ENSG00000204264	<i>PSMB8</i>	Proteasome subunit beta 8	0.249	3
ENSG00000185897	<i>FFAR3</i>	Free fatty acid receptor 3	0.224	5
ENSG00000134294	<i>SLC38A2</i>	Solute carrier family 38 member 2	0.214	3
ENSG00000187608	<i>ISG15</i>	ISG15 ubiquitin-like modifier	0.070	2
ENSG00000125347	<i>IRF1</i>	Interferon regulatory factor 1	0.027	1
ENSG00000162929	<i>KIAA1841</i> (also known as <i>SANBR</i>)	SANT and BTB domain regulator of class switch recombination	-0.014	2
ENSG00000272660	<i>AC090425.2</i>	Long non-coding RNA, antisense to <i>ACTL6A</i>	-0.016	1
ENSG00000196139	<i>AKR1C3</i>	Aldo-keto reductase family 1 member C3	-0.019	3
ENSG00000163735	<i>CXCL5</i>	C-X-C motif chemokine ligand 5	-0.019	3
ENSG00000080546	<i>SESN1</i>	Sestrin 1	-0.033	0
ENSG00000163710	<i>PCOLCE2</i>	Procollagen C-endopeptidase enhancer 2	-0.033	3
ENSG00000138207	<i>RBP4</i>	Retinol binding protein 4	-0.167	3
ENSG00000170323	<i>FABP4</i>	Fatty acid binding protein 4	-0.297	5
(Intercept)			-3.112	

Supplemental Table 3: Differential expression results for the 14 final classifier genes comparing: **A)** No Evidence patients under four years old (n=23; median age 1.3 years) versus over four years old (n=27; median age 12.5), and **B)** Definite patients under four years old (n=100; median age 0.4) versus No Evidence patients under four years old (n=23; median age 1.3).

A

Gene symbol	Log ₂ fold-change	P-value	Adjusted P-value
<i>GNLY</i>	-1.14	0.01	0.72
<i>PSMB8</i>	-0.33	0.21	0.82
<i>FFAR3</i>	-0.43	0.33	0.87
<i>SLC38A2</i>	-0.57	0.07	0.75
<i>ISG15</i>	-1.68	0.01	0.72
<i>IRF1</i>	-0.30	0.25	0.83
<i>KIAA1841</i>	0.14	0.59	0.94
<i>AC090425.2</i>	1.48	0.07	0.76
<i>AKR1C3</i>	0.81	0.10	0.76
<i>CXCL5</i>	0.24	0.67	0.95
<i>SESN1</i>	0.27	0.47	0.91
<i>PCOLCE2</i>	0.54	0.21	0.81
<i>RBP4</i>	-0.04	0.95	0.99
<i>FABP4</i>	-0.56	0.45	0.90

B

Gene symbol	Log ₂ fold-change	P-value	Adjusted P-value
<i>GNLY</i>	2.73	2.11E-08	1.46E-06
<i>PSMB8</i>	1.11	7.78E-08	4.22E-06
<i>FFAR3</i>	2.93	9.70E-12	4.25E-09
<i>SLC38A2</i>	0.60	2.50E-05	4.33E-04
<i>ISG15</i>	3.49	3.71E-09	3.89E-07
<i>IRF1</i>	1.52	5.18E-12	2.88E-09
<i>KIAA1841</i>	-0.89	2.84E-05	4.76E-04
<i>AC090425.2</i>	-0.18	7.50E-01	8.56E-01
<i>AKR1C3</i>	-2.48	2.24E-12	1.42E-09
<i>CXCL5</i>	-2.63	4.62E-09	4.52E-07
<i>SESN1</i>	-0.53	4.92E-02	1.38E-01
<i>PCOLCE2</i>	-2.06	2.60E-09	2.93E-07
<i>RBP4</i>	-3.64	1.51E-17	1.00E-13
<i>FABP4</i>	-5.54	4.02E-26	5.36E-22

Supplemental Table 4: Comparison of mNGS viral detection in TA samples with PCR viral detection in nasopharyngeal (NP) swabs or in the same TA samples in a subset of patients.

Definite patients with matched NP swab and TA viral PCR testing (n=21)	Agreement of mNGS with NP swab PCR	Concordance of mNGS and NP swab PCR	Agreement of mNGS with TA PCR	Concordance of mNGS and TA PCR
All viruses	22/34 = 64.7%	22/37 = 59.5%	23/24 = 95.8%	23/26 = 88.5%
Respiratory syncytial virus	11/12 = 91.7%	11/12 = 91.7%	10/10 = 100%	10/11 = 90.9%
Rhinovirus	4/7 = 57.1%	4/10 = 40%	6/6 = 100%	6/7 = 85.7%
Adenovirus	0/6 = 0%	0/6 = 0%	0/0 = 100%	0/0 = 100%
Coronavirus	2/3 = 66.7%	2/3 = 66.7%	2/2 = 100%	2/2 = 100%
Human metapneumovirus	3/3 = 100%	3/3 = 100%	3/3 = 100%	3/3 = 100%
Parainfluenza virus	2/3 = 66.7%	2/3 = 66.7%	2/3 = 66.7%	2/3 = 66.7%

Agreement reflects the number of viruses detected by mNGS out of the total number of viruses detected by PCR.

Concordance reflects the number of viruses detected by both mNGS and PCR out of the total number of viruses detected by at least one method.

Supplemental Table 5: Per-fold area under the curve (AUC) values for the integrated host/microbe logistic regression classifier.

Test fold	AUC
1	1.000
2	0.953
3	0.986
4	0.963
5	0.988

Supplemental Table 6: A) mNGS and clinical respiratory microbiology results for the Definite and No Evidence patients whose integrated LRTI classification was inconsistent with their adjudication. **B)** Primary diagnoses of the No Evidence patients whose integrated LRTI classification was inconsistent with their adjudication.

A

Patient	LRTI adjudication	Host P(LRTI)	Clinical respiratory microbiology results	mNGS viruses	Viral Σ rpM	mNGS RBM hits	Dominance of RBM hits	Integ. P(LRTI)
P2	Definite	0.30	<i>S. aureus</i> , HRV		0	<i>S. aureus</i>	0.62	0.10
P3	Definite	0.30	HRV, ADV	HRV C	2.15		0	0.22
P6	Definite	0.24	PIV, HCoV	PIV 4, HCoV NL63	87.61		0	0.29
P124	Definite	0.28	HCoV, <i>S. aureus</i> , <i>S. viridans</i>	HCoV 229E	41.58		0	0.41
P185	Definite	0.07	<i>S. maltophilia</i> , <i>S. pneumoniae</i> , <i>K. pneumoniae</i>		0	<i>S. maltophilia</i> , <i>S. pneumoniae</i>	0.75	0.01
P189	Definite	0.30	<i>S. aureus</i>	CMV	0.38	<i>S. aureus</i>	0.75	0.43
P195	Definite	0.54	<i>S. viridans</i> , <i>E. coli</i>		0		0	0.23
P218	Definite	0.43	<i>M. catarrhalis</i>		0		0	0.16
P1	No Evidence	0.67	No testing performed	HCoV NL63	44.94	<i>P. aeruginosa</i>	0.31	0.95
P30	No Evidence	0.35	No culture performed, negative PCR	HRV C	22.15	<i>M. catarrhalis</i>	0.85	0.68
P75	No Evidence	0.96	No culture performed, negative PCR	RSV	44.53		0	1.00
P166	No Evidence	0.79	No testing performed		0	<i>S. aureus</i>	0.82	0.92
P175	No Evidence	0.69	No testing performed	HCoV NL63	4.33		0	0.78
P250	No Evidence	0.78	Negative culture, no PCR performed		0		0	0.62

rpM, reads-per-million; RBM, rules-based model.

ADV, adenovirus
HCoV, human coronavirus
CMV, cytomegalovirus
HRV, human rhinovirus
PIV, parainfluenza virus
RSV, respiratory syncytial virus

B

Patient	LRTI adjudication	Primary diagnosis
P1	No Evidence	Neurological
P30	No Evidence	Trauma
P75	No Evidence	Non-infectious respiratory distress
P166	No Evidence	Ingestion (drug/toxin)
P175	No Evidence	Ingestion (drug/toxin)
P250	No Evidence	Seizures

Supplemental Table 7: mNGS results for the No Evidence patients with a primary diagnosis of non-pulmonary sepsis.

Patient	LRTI adjudication	Host <i>P</i> (LRTI)	mNGS viruses	Viral Σ rpM	mNGS RBM hits	Dominance of RBM hits	Integ. <i>P</i> (LRTI)
P34	No Evidence	0.05		0	<i>E. coli</i>	0.02	0.00
P40	No Evidence	0.62		0		0	0.20
P51	No Evidence	0.10	RSV	0.13	<i>M. catarrhalis</i>	0.22	0.01
P116	No Evidence	0.24		0		0	0.03
P178	No Evidence	0.65		0		0	0.35
P243	No Evidence	0.18		0		0	0.01

rpM, reads-per-million; RBM, rules-based model.

RSV, respiratory syncytial virus

Supplemental Data Files

Supplemental Data File 1. Basic sample metadata.

Supplemental Data File 2. Differential expression (DE) analyses between: i) Definite and No Evidence patients; ii) Definite patients with any bacterial LRTI and with purely viral LRTI.

Supplemental Data File 3. Gene set enrichment analysis (GSEA) results from the DE between: i) Definite and No Evidence patients; ii) Definite patients with any bacterial LRTI and with purely viral LRTI.

Supplemental Data File 4. Pathogens identified in Definite patients by clinical testing and by mNGS.