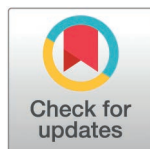RESEARCH ARTICLE

Leveraging transcriptome-wide association studies identifies the relationship between upper respiratory flora and cell type-specific gene expression in severe respiratory disease

Lei Xu[1‡], Ran He[1‡], Xiangyu Ye[1], Yifan Wang[2*], Shirong Hui[1], Haochang Li[1], Hongbo Chen[2*], Peng Huang[1*]

1 Depatment of Epidemiology, Center for Global Health, School of Public Health, National Vaccine Innovation Platiorm, Nanjing Medical University, Nanjing, China, 2 Department of Infectious Disease, Jurong Hospital Affiliated to Jiangsu University, Jurong, Jiangsu, China

‡ These authors share first authorship on this work.
* 18752942998@163.com (YW); chb2180@126.com (HC); huangpeng@njmu.edu.cn (PH)

Abstract

Objectives

The upper respiratory tract flora may influence host immunity and modulate susceptibility to viral respiratory infections. This study aimed to investigate the associations between upper respiratory tract flora and immune cells in severe ILI, identify specific microbial taxa and immune response pathways contributing to disease severity, and elucidate how flora influences ILI progression by modulating immune cell functions.

Methods

Heritability of GWAS summary data was estimated using LDSC (v1.0.1). Gene-level genetic associations were analyzed with MAGMA. scRNA-seq data were integrated with genetic association data using scDRS. FUSION was used to construct cell type-specific expression quantitative trait locus models based on genotypes and scRNA-seq data from the onek1k project, which were combined with flora abundance-related GWAS data for a transcriptome-wide association study.

Results

From the LDSC analysis, data from 1195 severe ILI-associated GWASs with upper respiratory flora($h2 > 0.1$) were included in subsequent analysis. TWAS identified 19 significant association pairs ($P_{adj} < 0.05$), and 1226 differentially expressed genes between mild and severe ILI patients ($P_{adj} < 0.05$ and $|\log_2 FC| > 0.25$). Functional enrichment analyses using GO, KEGG, and Reactome databases revealed that immune cells, such as CD4+T effector memory cells, cDCs, NK cells, were enriched in multiple biological processes or pathways.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: ILI = influenza like illness, Treg = regulatory T cells, COVID-19 = novel coronavirus pneumonia, QC = quality control, KO = Kyoto Encyclopedia of Genes and Genomes, BH = Benjamini & Hochberg, FDR = false discovery rate, COPD = chronic obstructive pulmonary disease, GWAS = genome-wide association studies, CNGBdb = China National Gene Bank database, WGS = whole genome sequencing, SNPs = single nucleotide polymorphisms, MAF = minor allele frequency, EAS = East Asian, EUR = European, scRNA-seq = single-cell RNA sequencing, GEO = Gene Expression Omnibus, PBMC = peripheral blood mononuclear cell, UMAP = uniform manifold approximation and projection, PLT = platelet, RBC = red blood cell, MC = Monte Carlo, LD = linkage disequilibrium, eQTL = expression quantitative trait loci, TWAS = transcriptome-wide association analysis, GO = Gene Ontology, FC = fold change, NK = natural killer cells, cDC = conventional dendritic cells, Prolif = proliferation cells.

Conclusions

This study identified associations between severe ILI-related upper respiratory tract flora and cell type-specific gene expression, potentially explaining how differential flora influences ILI progression. CD16＋monocytes exhibited the most differentially expressed genes, followed by proliferating cells and cDCs, highlighting the significant role of immune cell-enriched pathways in ILI progression.

1. Introduction

Influenza-like illness (ILI) is a global public health problem that kills hundreds of thousands of people each year [1–3]. Acute pulmonary inflammation and multiple organ failure associated with severe ILI are the leading causes of death after ILI infection [4–6]. Vaccination is the most effective way to prevent influenza. However, vaccine effectiveness has been low to moderate in recent years [7,8], and vaccine coverage remains low, particularly in low- and middle-income countries where pathogens are susceptible to mutation and protection is suboptimal [9]. Consequently, ILI infection remains a significant public health problem with enormous economic and social burdens globally. The study of ILI-organism interactions and mechanisms is important for infection control.

As more research supports the critical role of the flora in regulating host immunity [10–12], exploring whether these effects extend to influenza risk may provide new perspectives and approaches for refining prevention strategies. Animal models and human studies have shown that nasopharyngeal respiratory surfaces are colonized by large commensal flora [11–13]. They can alter susceptibility to viral respiratory infections, including influenza, by modulating host immunity. For example, flora can potentially inhibit the overgrowth of pathogens and prevent their transmission to the lungs, thereby serving as a "protective barrier" in maintaining respiratory health.[14] In recent years, it has been suggested that there is a relationship between the diversity of the upper respiratory tract flora and respiratory infections [15]. For example, one study reported that host susceptibility to ILI was closely related to the abundance of Streptococcus and Bacillus salivarius [16]. The abundances of Haemophilus, Prevotella, Clostridium and Neisseria in the oral cavity and nasopharynx were associated with the length of hospital stay in infants hospitalized with respiratory syncytial virus infection in one study [17]. A recent study of patients with novel coronavirus pneumonia (COVID-19) revealed significant differences in the diversity of the upper respiratory tract flora between COVID-19 patients and controls, with Corynebacterium, Lactobacillus, and other related flora consistently and significantly lower in COVID-19 patient samples, whereas Neisseria was consistently and significantly greater [18]. The epithelial cells of the upper and lower respiratory tracts are the primary target cells for influenza virus infection and replication. These epithelial cells are surrounded by a complex flora that may directly or indirectly interact with influenza viruses to mediate infection risk. Symbiotic strains can prevent infection by modulating the host's innate and adaptive

immune responses [10,11]. However, relatively few studies have been conducted on the relationship between the upper respiratory flora and the progression of ILI, and its mechanism in the progression of ILI has not been clarified. In-depth investigation and elucidation of the interactions and regulatory mechanisms between ILI and the host upper respiratory flora, lung injury, and systemic immune responses may provide new intervention strategies for ILI and help prevent and ameliorate infection-induced lung injury.

The antigen-specific immune response is an important means for the body to recognize and kill pathogens and has a significant impact on infection control and clinical resolution. Excessive immune killing can cause severe histopathological damage, leading to worsening of symptoms and even death [19–21]. In the intervening years, with the development of biotechnology, the understanding of immune cell function has deepened, and the role of different immune subtypes in the process of pathogen infection has been gradually recognized. For example, CD8＋effector memory T cells generated after atypical pneumonia can provide durable protection against coronaviruses. In addition, during respiratory syncytial virus infection, this subtype of T cells is associated with the development of pathological immune damage [22,23]. In COVID-19 patients, tissue-resident Th17 cells coordinate the killing effects of lung macrophages and cytotoxic CD8＋T cells and may be associated with a protective role against lung injury [24].

On the basis of the above findings, this study hypothesized that the upper respiratory tract flora may influence the course of ILI infection by modulating the recognition and killing effects of immune cells on pathogens after an organism is infected with ILI. To this end, We employed the method of Genome-Wide Association Study (GWAS), a well-established and efficient approach widely used to identify genetic loci associated with common diseases or traits [25]. The method aims to explore the relationship between specific traits or diseases and single nucleotide polymorphisms (SNPs), which represent variations in a single nucleotide within the genome. Typically, GWAS analyses rely on high-throughput geno-typing technologies, such as SNP arrays, and incorporate statistical methods (e.g., logistic regression, mixed-effects models) to assess the association between each SNP and the trait of interest. By comparing genotype differences between individuals, GWAS can uncover genetic variations related to specific traits or disease risks. In recent years, the scale of GWAS has expanded to include over 10 million SNPs [26]. Since these SNPs are typically inherited in linkage disequilibrium blocks, certain representative or tag SNPs can capture the majority of variations within each block. Fur-thermore, with the increasing demand for explaining the genetic basis of more diseases or traits, as well as the sub-stantial increase in sample sizes [27], the statistical power and reliability of the results have been significantly enhanced [28].Through Transcriptome-Wide Association Studies (TWAS), we integrated GWAS data with expression quantitative trait loci (eQTL) data [29]. TWAS not only focuses on the direct association between genotypes (such as SNPs) and traits, but also considers how gene expression levels in specific tissues or cell types influence the occurrence of traits or diseases. Although thousands of genomic loci related to complex traits have been identified through GWAS, most of the trait-associated variants are located in non-coding regions, the functionality of which remains difficult to interpret [30]. Moreover, identifying variants with small or moderate effects typically requires large sample sizes to achieve adequate statistical power. By aggregating multiple eQTLs to capture gene regulatory effects and directly measuring these effects at the transcriptome level, TWAS effectively overcomes these challenges, enabling the establishment of more biolog-ically interpretable associations between genes and complex traits or diseases [31]. We first calculated the heritability of the upper respiratory tract flora data obtained. Second, PBMC samples from patients with severe influenza were categorized into two groups, ILI mild and ILI severe, on the basis of symptoms, to identify the types of upper respiratory tract flora associated with these conditions. Third, after performing quality control on the data from these two groups, we conducted horizontal genetic association analysis via MAGMA software to integrate the flora SNP signals at the gene level. The scDRS (single-cell disease relevance score) was subsequently used to calculate genetic signals associated with the flora to identify the relevant cell types. Fourth, cell type-specific eQTL models were constructed for each identi-fied microbiome-associated cell type to identify specific genes. Finally, pathway enrichment analysis was conducted on the identified gene sets via R software.

2. Materials and methods

2.1 Collection of flora data

All adult Chinese individuals in this cohort were recruited for the multi-omics study, This cohort consists of 2,984 healthy individuals from the Chinese population. From 2017 to 2018, approximately 3,932 oral samples were newly collected from this cohort for whole metagenome sequencing, including 2,017 tongue dorsum samples and 1,915 saliva samples [32]. For the saliva sample, a double concentration of stabilizing reagent kit was used and 2ml saliva was collected. The samples from Yunnan Province were self-collected using a commercial kit (Catalog No. 401103, Zeesan, Xiamen, China) [33].

2.2 Data acquisition and processing for genome-wide association studies

(1) Data Acquisition for Genome-Wide Association Study of Upper Respiratory Tract Flora: GWAS data of upper respiratory tract flora were downloaded from the China National Gene Bank database (CNGBdb, project data No. CNP0001664) [32]. A total of 1,566 dorsal tongue flora and 1,547 salivary flora GWAS summary datasets were included in this project (genome version GRCh38). The full GWAS summary data were downloaded for this study. Notably, owing to some of the differences in the definitions of some of the floras between this study and the project, this study referred to the LPSN database (https://lpsn.dsmz.de/) for the GWAS summary data that had the same or most similar definitions to each of the respiratory tract flora [34].

(2) Data quality control: The following QC processes were also performed on the downloaded GWAS summary data related to upper respiratory tract flora: i) SNPs located on autosomes were retained; ii) SNPs detected in less than 70% of the samples were excluded; iii) SNPs with alleles other than "A", "C", "G", or "T" were excluded; iv) SNPs with anomalies in the variance of the summary statistic (≤ 0); v) SNPs with abnormal standard error (SE ≤ 0) were excluded; and vi) SNPs with allele frequency (MAF) < 0.001.

(3) Heritability calculation: The heritability (h2) of the included GWAS summary data was estimated via LDSC software (v.1.0.1), with 504 East Asian (EAS) populations from the 1000 Genomes Project Phase III as the reference panel [35,36]. For GWAS data related to the upper respiratory tract flora, only those with heritability estimates (h2) greater than 0.1 were included in subsequent analyses.

2.3 Data acquisition and processing for single-cell sequencing

(1) Data acquisition for single-cell sequencing: In this study, two scRNA-seq datasets containing peripheral blood mononuclear cell (PBMC) samples from patients with severe ILI were acquired to identify cell types associated with the upper respiratory tract flora. The original study of the first scRNA-seq dataset (GSE149689) included a total of 20 PBMC samples from asymptomatic, mildly symptomatic, and severely symptomatic COVID-19 patients; patients with severe influenza; and healthy controls [37]. We selected 16 samples totaling 53,387 cells from patients hospitalized with COVID-19 and severe influenza for subsequent analysis. A second source study of scRNA-seq data (GSE164948) included 20 PBMC samples from patients hospitalized with COVID-19, influenza, and CAP infections with other pathogens, as well as healthy controls [38]. Similarly, 16 samples from hospitalized patients with a total of 13,999 cells were selected for subsequent analysis. Five samples from asymptomatic and mildly symptomatic COVID-19 patients were defined as having mild ILI, and the remaining 27 samples from patients with severe COVID-19, critically hospitalized patients and hospitalized patients with influenza or CAP were defined as having severe ILI.

(2) Data quality control:First, i) scDblFinder (v.1.9.4) was used to identify and remove potential doubles [39]; ii) cells expressing fewer than 200 genes, with a total UMI count ≥20,000, or expressing a percentage of mitochondrial genes ≥15% of the total UMI count were excluded; and iii) genes expressed in fewer than 3 cells were excluded. Finally, 51,243 cells and 23,442 genes were retained for subsequent analyses.

(3) Sample merging, dimensionality reduction and clustering: i) Seurat (v.4.4.0) was first employed to perform *SCTransform* normalization on each sample, in which a regression was performed on the percentage of mitochondrial gene expression in each cell [40,41]; ii) all samples were merged, 3,000 highly variable genes were selected, and linear dimensionality reduction was performed via principal component analysis; iii) batch correction of PCs was conducted on a per-sample basis via Harmony (v.1.2.0) [42]; iv) the top 28 Harmony-corrected PCs that cumulatively explained up to 90% of the variance were selected to construct a K-nearest neighbor graph structure and were analysed by clustering via Louvain's algorithm; and v) uniform manifold approximation and projection (UMAP) with the same number of PCs for visualization.

(4) Cell type annotation: To ensure the accuracy of cell type annotation, we performed cell type annotation on the sub-populations obtained by clustering on the basis of widely used marker genes as well as those used in the two original studies of the data (Table 1). A total of 51,243 cells were finally divided into 15 subtypes. Of these, two subtypes, platelets (PLTs, n = 3,636) and red blood cells (RBCs, n = 829), were not the cell types.

(5) Differential expression analysis: The differentially expressed genes (DEGs) between the severe and mild ILI groups in each cell type were analysed via the Wilcoxon rank sum test. Among them, genes whose $|\log_2(\text{fold change})|(|\log_2\text{FC}|) > 0.25$ and BH-adjusted $P < 0.05$ were defined as differentially expressed genes.

2.4 Analysis of flora-associated cell types

The enrichment of flora-related genetic signals in 29,149 cells from the severe ILI samples was calculated via the single-cell disease relevance score (*scDRS*, v.1.0.3) to identify each flora-associated cell type [43]. In brief, given the processed scRNA-seq data and gene–disease association results output from *MAGMA*, *scDRS* first constructs a putative disease gene set G with each g in G weighted by its GWAS *MAGMA* z score (ω_g) and then quantifies the aggregate expression of the putative disease gene set in each cell to generate cell–specific raw disease scores S_c:

$$S_c = \frac{\sum_{g \in G} \omega_g \sigma_{tech, \ g}^{-1} X_{cg}}{\sum_{g \in G} \omega_g \sigma_{tech, \ g}^{-1}}$$

Table 1. Cell types defined marker gene.

Cell Subtype	Cell Population	Marker Gene
T Cell	15715	CD2、CD3E
CD4 + naive T cell	5821	CD3E、CD4、CCR7、SELL
CD4 + Effector T Cell	360	CD4、GNLY、NKG7
CD4 + Effector Memory T Cell	716	CD4、CTLA4、GPR183
CD8 + Naive T Cell	760	CD3E、CD8A、CD8B、CCR7、SELL
CD8 + Effector T Cell	5022	CD8A、CD8B、GNLY、NKG7
CD8 + Effector Memory T Cell	3036	CD8A、CD8B、GZMK、GPR183
NKT	1382	CD3G、KLRF1、KLRD1
NK	6378	KLRF1、KLRD1
B Cell	6231	CD79A、MS4A1
Monocyte	16029	CD14、CD163、CD68
CD14 Mono	14377	CD14、S100A12
CD16 Mono	1652	CD14、FCGR3A
cDC	272	FCER1A、CD1C
Proliferating Cell	771	MKI67、TYMS
PLT	3636	PF4、PPBP
RBC	829	HBD、HBM

where $\sigma_{tech,\ g}$ indicates the estimated gene-specific technical noise of gene g and where X_{cg} is the normalized expression of cell c on gene g. scDRS also uses Monte Carlo (MC) to sample matched control gene sets and generate a set of raw control scores. scDRS normalizes these raw scores and computes cell-level P values on the basis of the empirical distribution of the pooled normalized control scores for each cell c:

$$P_c = \frac{1 + \sum_{C'=1}^{n_{cell}} \Sigma_{b=1}^{B} \Pi \left(S_c \leq S_{c'b}^{ctrl} \right)}{1 + n_{cell}B}$$

where B is the number of MC samples. In this study, we draw the top 1,000 genes from *MAGMA* ranked by the z score to generate the disease gene set and set the MC sampling times to 1000.

2.5 Identification of flora-associated cell type-specific genes

For the identified cell types associated with each flora, the cell type-specific genes associated with each flora were identified by constructing a cell type-specific gene expression quantitative trait locus (eQTL) model and applying the analytical framework of FUSION software to perform TWAS analysis [29]. Only cis-eQTL relationships (cis-eQTLs) were considered in this study.

2.5.1 Cell type-specific gene expression quantitative trait genome models.

(1) Genotyping Data Processing: The converted genotyping data were subjected to a series of QCs with reference to the original study [44], and genotypes were populated via minimac4 (v.4.1.6) software. After excluding SNPs with a fill score < 0.7 or an MAF < 0.01, a final set of 5,807,466 SNPs from 1034 individuals was obtained [35].

(2) Single-cell sequencing data processing: The QC steps applied to the scRNA-seq data included the following: i) exclusion of genes whose expression was detected in fewer than 3 cells; ii) exclusion of 18,959 cells from 15 subjects that were not included in the genotype resampling; and iii) considering the heterogeneity of populations, experimental conditions, and cell type definitions, a total of 883,855 cells from 19 subtypes were reclassified into 14 subtypes on the basis of the similarity of gene expression in the flora-associated cell type analysis section. (**Table 2**).

(3) Model construction: Construct the cis-eQTL model via FUSION.compute_weights pipeline from FUSION [29]. The cis-eQTL model construction process consists of i) extracting SNPs located in the genetic cis-regulatory regions of the target genes (cis-SNPs) from the genotypic data of the study subjects via PLINK (v.1.9.0) software [45]; ii) a kinship matrix was constructed on the basis of the cis-SNPs of the target gene, and its heritability was calculated via GCTA (v.1.93.2 beta) in combination with the individual expression levels of the gene (cis-h2) [46]; iii) for genes with significant heritability, each cis-eQTL method was fitted in the form of cross-validation to estimate the effect of fitting the model via each method; and iv) each cis-eQTL method was fitted via all samples to obtain the final result of each cis-SNP gene expression weight.

2.5.2 Identification of flora-associated cell type-specific genes.
On the basis of the constructed cell type-specific cis-eQTL model, genomic data from the EAS population at 1000 Genomes Project were used as a reference panel, and TWAS analysis was performed on each flora-associated GWAS summary dataset via the FUSION framework to identify the cell type-specific expressed genes associated with each flora. Considering that the eQTL model was constructed from the European(EUR)population and that the flora-associated GWAS summary data were from the EAS population, this part of the study focused only on nominally significant genes, i.e., those with uncorrected P < 0.05.

2.6 Functional enrichment analysis

For a given set of genes, pathway enrichment analysis was performed via the clusterProfiler (v.4.2.2) and ReactomePA (v.1.38.0) software packages on the basis of the Gene Ontology (GO), KEGG, and Reactome pathway databases [47,48].

Table 2. onek1k Summary of single-cell data quality control.

onek1k Cell Subtype	Cell Count		Target Cell Subtype	Target Cell Count
	Before Filtering	**After Filtering**		
CD4 Naive	259012	255130	CD4 + Naive T Cell	255130
CD4 CTL	17993	17849	CD4 + Effector T Cell	17849
CD4 TEM	31261	30879	CD4 + Effector Memory T Cell	30879
CD8 Naive	52538	51487	CD8 + Naive T Cell	51487
CD8 TCM	16409	16175	CD8 + Effector T Cell	16175
CD8 TEM	161051	159634	CD8 + Effector Memory T Cell	159634
NK	163202	160465	NKT	160465
NK_CD56bright	7006	6922	NK	6922
B naive	65702	64321	B memory	123261
B intermediate	29889	29466	B memory	
B memory	30234	29474	B memory	
CD14 Mono	36130	35560	CD14 + Monocyte	35560
CD16 Mono	15743	15502	CD16 + Monocyte	15502
cDC1	114	114	cDC	4515
cDC2	4456	4401	cDC	
CD4 Proliferating	773	762	Proliferating Cell	2770
CD8 Proliferating	305	304	Proliferating Cell	
NK Proliferating	1731	1704	Proliferating Cell	

https://doi.org/10.1371/journal.pone.0322864.t002

3. Results

3.1 Analysis of GWAS data related to the flora of the upper respiratory tract and identification of associated cell types

A total of 1,566 GWAS summary statistics of the dorsal tongue flora and 1,547 GWAS summary statistics of the salivary flora were included in this part of the study. After QC, 5,345,095 SNPs were obtained. On the basis of the results of LDSC, 1195 GWASs of upper respiratory tract flora with h2 > 0.1 were included in the subsequent analyses, of which 489 were salivary flora-associated GWASs and 706 were dorsal tongue flora-associated GWASs. On the basis of MAGMA genetic association analysis, we identified 27 genes that were significantly associated with at least one flora. The top five genes, ranked in ascending order of P value, are *SLC2A9* ($P = 1.37$E-09), *SLC2A9* ($P = 3.82$E-08), *MDGA2* ($P = 4.59$E-08), *NUDT8* ($P = 1.01$E-07), and *BBOX1* ($P = 1.40$E-07) (Table 3 and An additional file shows this in more detail [see S1 Fig]). Notably, the *SLC2A9* gene was significantly associated with two flora in the GWAS, both of which belong to the genus Oribacterium within the phylum Firmicutes in the tongue dorsum microbiome.

For the included ILI-associated single-cell data, a total of 51,243 PBMCs were obtained after QC for subsequent analysis. Downscaling, fractionation and cell type annotation yielded a total of 15 immune cell subtypes. (Fig 1, An additional file shows this in more detail [see S2-S3 Figs]). A total of 4,465 cells from two subtypes, PLTs (n = 3,636) and RBCs (n = 829), were further excluded. A total of 29,149 cells from 13 cell types from severe ILI samples were ultimately included in the flora-associated cell type analysis.

According to the results of the scDRS analysis, a total of nine taxa associated with ILI at the "phylum" level were identified. The phylum Firmicutes was the most abundant phylum with 6,305 instances, representing 40.59% of the total, followed by Bacteroidota with 1,846 instances (11.88%), Actinobacteriota with 1,820 instances (11.72%), Proteobacteria with 1,287 instances (8.28%), and Fusobacteriota with 975 instances (6.26%). In total, 205 species were found to be statistically significant, with an adjusted p value (P_{adj}) < 0.05. (**Table 4**).Thirteen immune cell types associated with severe ILI were

Table 3. Genetic association analysis based on MAGMA gene level.

Gene	Chromo-some	N$_{SNPs}$	Ns$_{ample}$	P	Sample	Bacterial community
SLC2A9	4	808	1996	1.37E-09	Tongue	k__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__Oribacterium; s__unclassified_mgs_1215
SLC2A9	4	808	1991	3.82E-08	Tongue	k__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__Oribacterium; s__unclassified_mgs_489
MDGA2	14	1636	2000	4.59E-08	Tongue	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Aerococcaceae; g__Granulicatella; s__unclassified_mgs_2134
NUDT8	11	64	1898	1.01E-07	Saliva	k__Bacteria; p__Firmicutes; c__Clostridia; o__TANB77; f__CAG-508; g__CAG-793; s__unclassified_mgs_3234
BBOX1	11	205	1985	1.40E-07	Tongue	k__Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Actinomycetales; f__Actinomycetaceae; g__Pauljensenia; s__unclassified_mgs_3571
NDUFV1	11	50	1898	1.57E-07	Saliva	k__Bacteria; p__Firmicutes c__Clostridia; o__TANB77; f__CAG-508; g__CAG-79;|s__unclassified_mgs_3234
GJB4	1	78	1998	3.08E-07	Tongue	k__Bacteria; p__Spirochaetota; c__Spirochaetia; o__Treponematales; f__Treponemataceae; g__Treponema_D; s__unclassified_mgs_1124
SPDYE18	7	838	1900	3.20E-07	Saliva	k__Bacteria; p__Firmicutes; c__Negativicutes; o__Veillonellales; f__Veillonellaceae; g__F0422; s__unclassified_mgs_1482
KANSL3	2	44	1866	3.35E-07	Saliva	k__Bacteria; p__Patescibacteria; c__Saccharimonadia; o__Saccharimonadales; f__Saccharimonadaceae; g__unclassified_mgs_2626
EDEM1	3	86	1987	3.45E-07	Tongue	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Aerococcaceae; g__Granulicatella; s__unclassified_mgs_3116
SPRTN	1	86	1996	3.53E-07	Tongue	k__Bacteria; p__Fusobacteriota; c__Fusobacteriia; o__Fusobacteriales; f__Leptotrichiaceae; g__unclassified_mgs_2348
TBX10	11	80	1898	3.65E-07	Saliva	k__Bacteria; p__Firmicutes; c__Clostridia; o__TANB77; f__CAG-508; g__CAG-793; s__unclassified_mgs_3234
EGLN1	1	225	1996	5.04E-07	Tongue	k__Bacteria; p__Fusobacteriota; c__Fusobacteriia; o__Fusobacteriales; f__Leptotrichiaceae; g__unclassified_mgs_2348
ATP8A1	4	554	1904	5.13E-07	Saliva	k__Bacteria; p__Firmicutes; c__Negativicutes; o__Selenomonadales; f__Selenomonadaceae; g__Selenomonas; s__unclassified_mgs_2794
ZNF277	7	198	1905	5.27E-07	Saliva	k__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Prevotella; s__Prevotella_oulorum_mgs_3240
GJB5	1	74	1998	5.42E-07	Tongue	k__Bacteria; p__Spirochaetota; c__Spirochaetia; o__Treponematales; f__Treponemataceae; g__Treponema_D; s__unclassified_mgs_1124
MCHR2	6	96	1983	7.05E-07	Tongue	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus; s__unclassified_mgs_449
LRIF1	1	49	1894	7.97E-07	Saliva	k__Bacteria; p__Fusobacteriota; c__Fusobacteriia; o__Fusobacteriales; f__Fusobacteriaceae; g__Fusobacterium; s__unclassified_mgs_2180
EDEM1	3	86	1976	8.33E-07	Tongue	k__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Porphyromonadaceae; g__Porphyromonas; s__unclassified_mgs_2117
ANG	14	70	1902	9.27E-07	Saliva	k__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__Oribacterium; s__unclassified_mgs_388
IFNE	9	52	1894	1.08E-06	Saliva	k__Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Actinomycetales; f__Actinomycetaceae; g__Pauljensenia; s__unclassified_mgs_3546
RNASE4	14	81	1902	1.08E-06	Saliva	k__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__Oribacterium; s__unclassified_mgs_388
EXOC8	1	27	1996	1.15E-06	Tongue	k__Bacteria; p__Fusobacteriota; c__Fusobacteriia; o__Fusobacteriales; f__Leptotrichiaceae; g__unclassified_mgs_2348
FZD3	8	248	1905	1.31E-06	Saliva	k__Bacteria; p__Campylobacterota; c__Campylobacteria; o__Campylobacterales; f__Campylobacteraceae; g__Campylobacter_A; s__Campylobacter_A_showae_A_mgs_1682

(Continued)

| Gene	Chromo-some	N$_{SNPs}$	Ns$_{ample}$	*P*	Sample	Bacterial community
RAI14	5	540	1995	1.44E-06	Tongue	k__Bacteria; p__Firmicutes; c__Bacilli; o__Erysipelotrichales; f__Erysipelotricha-ceae; g__Solobacterium; s__unclassified_mgs_3577
KCNK9	8	237	1991	1.45E-06	Tongue	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus; s__unclassified_mgs_3318
MUC7	4	212	1899	2.02E-06	Saliva	k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus; s__unclassified_mgs_1910
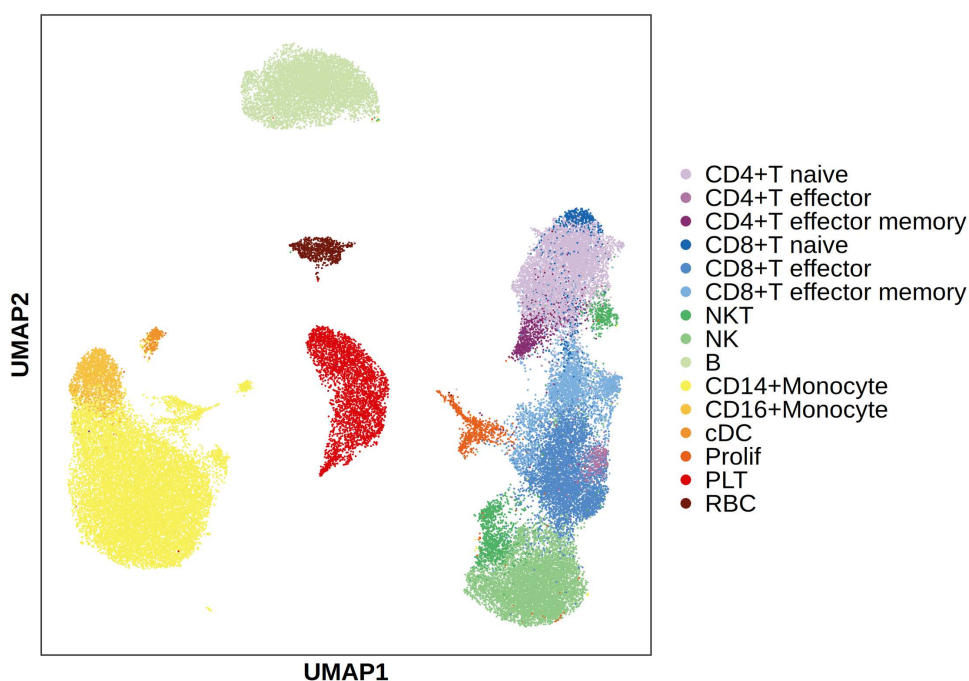
Fig 1. UMAP plot of downscaled scRNA-seq data from severe ILI colored by cell type. Description of data: NK: natural killer cells; cDC: conventional dendritic cells; Prolif: proliferation cells; PLT: platelets; RBC: red blood cells.

Table 4. Based on the "phylum" classification level, the statistical analysis of severe influenza-like illness (ILI) upper respiratory flora.

phylum	Total (%)	Saliva (%)	Tongue (%)	P_sig (%)
Firmicutes	6305(40.59)	2223(35.27)	4082(64.73)	91(1.44)
Bacteroidota	1846(11.88)	975(52.82)	871(47.18)	13(0.70)
Actinobacteriota	1820(11.72)	858(47.14)	962(52.86)	19(1.04)
Proteobacteria	1287(8.28)	507(39.40)	780(60.60)	16(1.24)
Fusobacteriota	975(6.26)	416(42.67)	559(57.33)	10(1.03)
Others	3289(21.17)	1378(41.90)	1911(58.10)	56(1.00)

associated with at least one flora, whereas 205 flora were associated with at least one immune cell type. (An additional file shows this in more detail [see **S4 Fig**]). On the basis of the proportion of significant flora nominally significantly associated cells, CD16 + monocytes were associated with metagenomics-defined Campylobacter_A_concisus_F_mgs_1384 in saliva (29.49%, P_{adj}=0.013), unclassified_mgs_803 in the genus Prevotella salivarius (28.9%, P_{adj}=0.013), unclassified_mgs_1198 in the genus Haemophilus salivarius (23.43%, P_{adj}=0.006), and Lachnoanaerobaculum_sp000296385_mgs_3102 in the genus Lachnoanaerobaculum (27.74%, P_{adj}=0.006). The abundance of these flora was significantly correlated with that of other cell types. In this study of the upper respiratory tract flora associated with ILI, Firmicutes was the taxon with the highest proportion at the "phylum" level of categorization, which is consistent with other similar studies [49,50]. Thus, in the study of Firmicutes, 13 immune cell types were associated with at least one flora, and 91 flora were associated with at least one immune cell type.(**Fig 2**) In addition to the significant correlation between Lachnoanaerobaculum_sp000296385_mgs_3102 and the abundance of CD16 + monocytes in the genus Lachnoanaerobaculum described above, this strain was also significantly correlated with the abundance of cDCs (24.48%, P_{adj}=0.006). CD4 + T effector cells (27.75%, P_{adj}=0.004), CD8 + T effector cells (21.54%, P_{adj}=0.004), and NK cells (23.43%, P_{adj}=0.004) were significantly associated with the abundance of unclassified_mgs_1428 in the genus Catonella from the tongue dorsum.

3.2 Identification of cell type-specific genes associated with the upper respiratory tract flora

For the 205 flora identified by scDRS with P_{adj}<0.05 and their associated 13 cell types, cell-specific eQTL models were further constructed and analysed by TWAS to identify severe ILI-associated cell type-specific genes associated with the flora.

A total of 19 significantly associated pairs (P_{adj}<0.05) that were closely associated with severe ILI and involved 9 types of immune cells were identified via TWAS analysis. In particular, CD14 + monocyte cells, CD16 + monocyte cells, and CD4 + T effector memory cells had greater numbers of significant associations. Among the 8947 nominally significant association pairs (uncorrected P<0.05), 3854 salivary flora association pairs and 5093 dorsal tongue flora association pairs were identified (Fig 3 and An additional file shows this in more detail [see S1 File]). To further explore their potential roles in ILI, we conducted differential expression analysis on cell type-specific genes associated with particular flora and reported that 1226 genes presented significantly different expression between patients with mild and severe ILI (P_{adj}<0.05 and|$\log_2$FC|>0.25) (An additional file shows this in more detail [see S2 File]). The differential expression analysis results for the 9 cell types significantly correlated with the ILI-associated pairs revealed that *TMEM41B, ERAP2*, and *CCR1* in CD14 + monocyte cells were associated with the metagenomic definition unclassified_mgs_2621 within the genus Granulicatella salivarius ($P_{adj\ max}$<0.001, $\log_2$FC$_{max}$=-0.321). Compared with other immune cells, CD16 + monocyte cells presented the greatest number of DEGs and were significantly associated with 26 types of flora, followed by proliferating cells and cDCs. CD16 + monocyte cells were significantly associated with genes such as *L1TD1, MTHFR*, and *ARL6IP5* related to unclassified_mgs_1057 in the genus Streptococcus from the tongue dorsum ($P_{adj\ max}$=0.021, $\log_2$FC$_{max}$=-0.257). In CD4 + T effector memory cells, the expression levels of genes such as *CCT3, SEPT2*, and *EFTUD2*, which are associated with Aggregatibacter actinomycetemcomitans_mgs_301 within the genus Aggregatibacter ($P_{adj\ max}$=0.003, $\log_2$FCmax=-0.362), were significantly lower in the severe ILI group. Conversely, the expression levels of the gene *SKAP2* associated with Neisseria_sicca_A_mgs_986 in the genus Neisseria from saliva ($P_{adj\ max}$<0.001, $\log_2$FC$_{max}$=0.473) in CD14 + monocyte cells and the expression levels of genes such as *CFD* and *PSME2* ($P_{adj\ max}$<0.001, $\log_2$FC$_{max}$=27.591) associated with unclassified_mgs_1057 in the genus Streptococcus from the tongue dorsum and the expression levels of genes such as *TMEM70 and XBP1* ($P_{adj\ max}$<0.001, $\log_2$FC$_{max}$=3.703) associated with unclassified_mgs_1956 in the genus Prevotella from the tongue dorsum in CD16 + monocyte cells were significantly greater in the severe ILI group.

3.3 Functional enrichment analysis of cell type-specific DEGs associated with the flora of the upper respiratory tract

Functional enrichment analysis based on the GO, KEGG, and Reactome pathway databases was conducted for the identified flora-associated, cell type-specific genes that were significantly differentially expressed between patients with
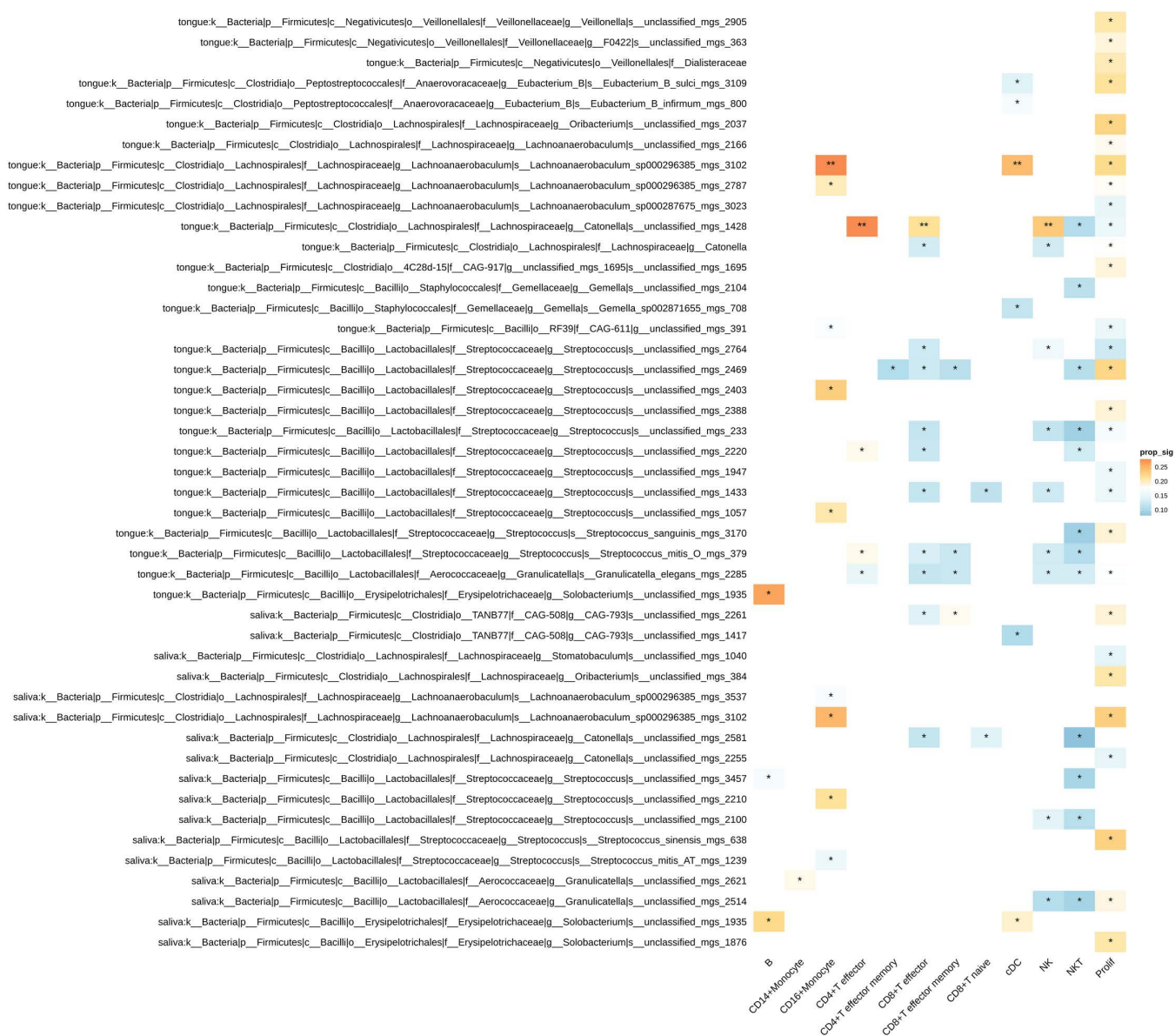
Fig 2. Heatmap of cell type associations with the Firmicutes "phylum" based on scDRS analysis. Description of data: Colors represent the proportion of cells significantly associated with the flora in each cell type, with asterisks indicating significance. *: BH-corrected $P < 0.05$.

https://doi.org/10.1371/journal.pone.0322864.g002

mild and severe ILI. The enrichment results indicated that in CD14 + monocyte cells from ILI patients, differential genes associated with the metagenomic definition unclassified_mgs_2621 (saliva_pheno.2184) in the genus Granulicatella were significantly enriched in processes such as protein carrier chaperone activity, aminopeptidase activity, and intramembrane lipid transporter activity (Enrichment ratio$_{min}$ = 0.167, FDR_{max} = 0.011), whereas differential genes associated with Neisseria_sicca_A_mgs_986 (saliva_pheno.307) in the genus Neisseria were significantly enriched in biological processes such as the regulation of blood vessel remodelling, negative regulation of extracellular matrix organization, and leucine zipper domain binding (Enrichment ratio$_{min}$ = 0.2, FDR_{max} = 0.024). The DEGs associated with unclassified_mgs_1939 (tongue_pheno.2645) of the genus Campylobacter_A from the tongue dorsum in CD4 + T effector memory cells were significantly enriched in the processes of the cytoplasmic vesicle lumen, vesicle lumen, and secretory granule lumen (Enrichment
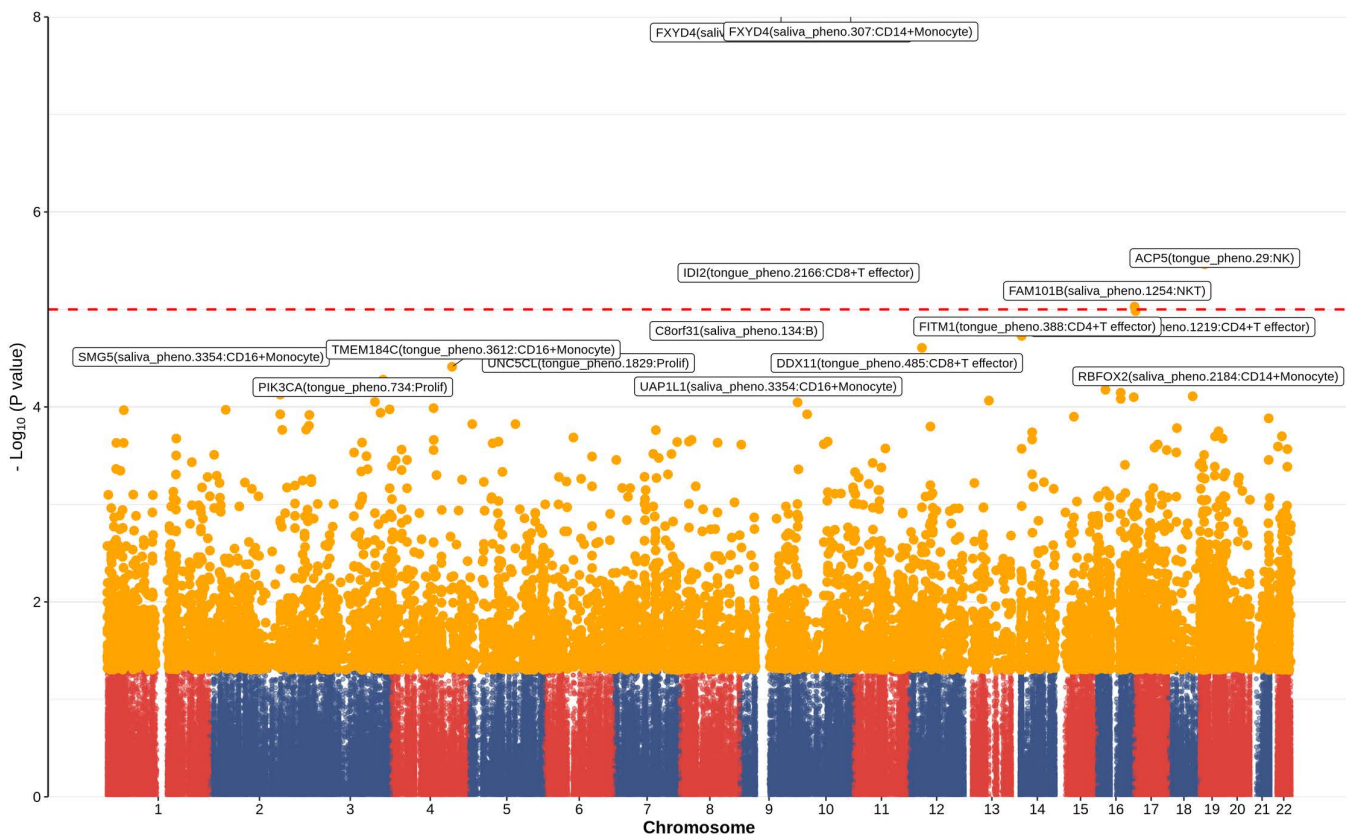
Fig 3. Manhattan plot of cell-specific gene expression and flora associations based on TWAS analysis. Description of data: Red dashed line represents nominal significance level ($P = 0.05$).

ratio$_{min}$ = 0.667, FDR_{max} < 0.001). Additionally, DEGs associated with Aggregatibacter_actinomycetemcomitans_mgs_301 (tongue_pheno.485) of the genus Aggregatibacter from the tongue dorsum were significantly enriched in the biological processes of the regulation of establishment of protein localization to telomeres, the regulation of protein localization to the Cajal body, and the positive regulation of protein localization to the Cajal body (Enrichment ratio$_{min}$ = 0.250, FDR_{max} < 0.001). Differential genes in cDC associated with unclassified_mgs_1417 (saliva_pheno.1286) from the genus saliva CAG-793 were enriched in processes such as mitochondrial protein-containing complexes, oxidoreductase complexes, and the mitochondrial matrix (Enrichment ratio$_{min}$ = 0.375, FDR_{max} < 0.001). Additionally, differential genes associated with Eubacterium_B_infirmum_mgs_800 (tongue_pheno.71) from the dorsal tongue Eubacterium were significantly enriched in biological processes related to hydrolase activity, hydrolysis of O-glycosyl compounds, and other pathways involving the degradation of glycosyl bonds and sugars. In contrast, NK cells were significantly enriched for biological processes associated with differential gene enrichment, with both saliva Prevotella_loescheii_mgs_2547 (saliva_pheno.1016) and tongue dorsum Catonella (tongue_pheno.29), and Prevotella_loescheii_mgs_2547 mgs_2547 were enriched for biological processes such as regulation of leukocyte-mediated cytotoxicity and regulation of cell killing (Enrichment ratio$_{min}$ = 0.500, FDR_{max} = 0.007). The genus Catonella was enriched for tubulin binding, Polycomb repressive complex and other processes (Enrichment ratio$_{min}$ = 0.400, FDR_{max} = 0.015). The unclassified_mgs_1397 (saliva_pheno.2783)-related DEGs in the genus Saliva Campylobacter_A were enriched mainly in biological processes such as glutamate receptor binding, spindle pole and dopaminergic synapse; pathways related to neurodegeneration; multiple diseases; and other biological pathways (Enrichment ratio$_{min}$ = 0.333, FDR_{max} = 0.041). Differential genes associated with saliva Eikenella

2545 unclassified_mgs_2545 (saliva_pheno.2390) were enriched in proliferating cells for motor proteins, microtubule motor activity, vasopressin-regulated water reabsorption and other biological processes (Enrichment ratio$_{min}$ = 0.500, FDR_{max} = 0.005). Differential genes associated with saliva unclassified_mgs_2225 (saliva_pheno.2774) and tongue dorsal Campylobacter_A_showae_A_mgs_3545 (tongue_pheno.760) in the genus Campylobacter_A were associated with CD16 + monocyte cells and CD8 + T effector cells enriched in the following biological processes: ficolin-1-rich granules, ficolin-1-rich granule lumen (Enrichment ratio$_{min}$ = 0. 214, FDR_{max} = 0.002), the response to starvation, the cellular response to nutrient levels, and the cellular response to starvation (Enrichment ratio$_{min}$ = 0.600, FDR_{max} < 0.001).(Figs 4 and 5, Table 5, and An additional file shows this in more detail [see S3 File]).

4. Discussion

This study integrated upper respiratory tract flora-related GWAS data and severe ILI-related transcriptomic data to explore the potential associations between upper respiratory tract flora and immune cells in severe ILI and to investigate the possible mechanisms influencing severe ILI from the perspectives of metabolism and the immune response.

The study employed a stratified design to integrate multi-source heterogeneous data: the upper respiratory tract flora GWAS data were derived from a homogeneous cohort of healthy Han Chinese individuals, including 2017 tongue dorsum samples and 1915 saliva samples, while transcriptomic data related to severe ILI were integrated from independent cohorts of Southeast Asian and European populations. Previous studies have shown that host genetic factors exhibit long-term stability in the relative abundance of heritable flora taxa, with minimal disruption by environmental factors [51]. Therefore, this study effectively reduced the interference of host heterogeneity and environmental confounding factors
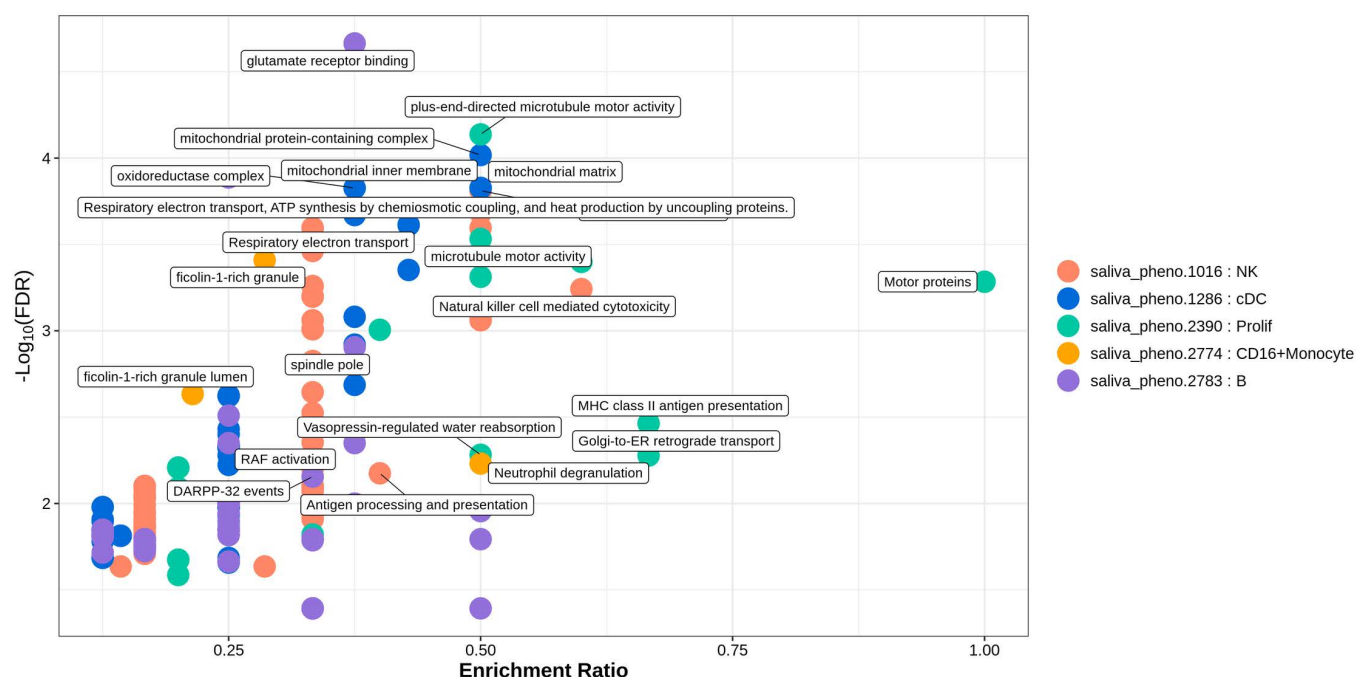

Fig 4. Bubble plot of functional enrichment analysis for salivary flora-associated cell type-specific differentially expressed genes. Description of data: Results of the top 5, arranged in ascending order by q-value; saliva_pheno.1016: Prevotella_loescheii_mgs_2547 within the genus Prevotella in the salivary flora; saliva_pheno.1286: unclassified_mgs_1417 within the genus CAG-793 in the salivary flora; saliva_pheno.2390: unclassified_mgs_2545 within the genus Eikenella in the salivary flora; saliva_pheno.2774: unclassified_mgs_2225 within the genus Campylobacter_A in the salivary flora; saliva_pheno.2783: unclassified_mgs_1397 within the genus Campylobacter_A in the salivary flora.

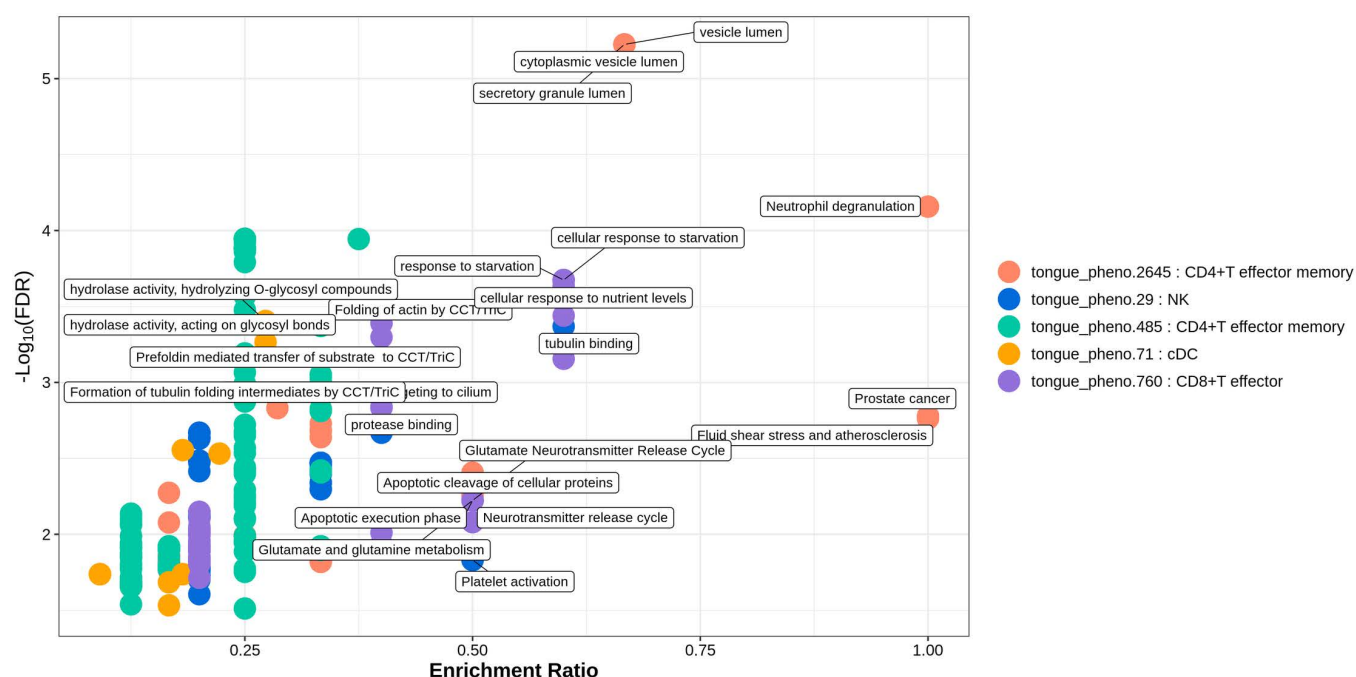https://doi.org/10.1371/journal.pone.0322864.g004

Fig 5. Bubble plot of functional enrichment analysis for dorsal tongue flora-associated cell type-specific differentially expressed genes.
Description of data: Results of the top 5, arranged in ascending order by q-value; tongue_pheno.2645: unclassified_mgs_1939 within the genus Campylobacter_A in the dorsal tongue flora; tongue_pheno.29: the genus Catonella in the dorsal tongue flora; tongue_pheno.485: Aggregatibacter_actinomycetemcomitans_mgs_301 within the genus Aggregatibacter in the dorsal tongue flora; tongue_pheno.71: Eubacterium_B_infirmum_mgs_800 within the genus Eubacterium_B in the dorsal tongue flora; tongue_pheno.760: Campylobacter_A_showae_A_mgs_3545 within the genus Campylobacter_A in the dorsal tongue flora.

https://doi.org/10.1371/journal.pone.0322864.g005

in the flora-immune association analysis by including a cohort with consistent genetic background, strict health status screening, and standardized sampling sites. Further research has shown that the diversity of human immune genes is driven by human contact with pathogens [52,53], and there is a significant correlation between genetic variability and global pathogen diversity [54]. Some of these correlations may be confounded by variables such as climate, diet, or lifestyle. Thus, incorporating samples with diverse genetic backgrounds can reveal the heterogeneity of immune responses across different populations.

The primary objective of this study is to investigate the common immunological mechanisms underlying influenza-like illness (ILI) and their associations with the upper respiratory flora and host immune response, rather than to elucidate pathogen-specific mechanisms. Moreover, the early symptoms of COVID-19 substantially overlap with those of ILI, and in severe cases, both conditions manifest with upper respiratory inflammation and immune dysregulation. Therefore, we regard COVID-19, in essence, as a form of ILI, and it is classified as part of ILI in public health surveillance. To encompass a broader range of ILI and a wider spectrum of ILI samples, we included patients with ILI caused by various pathogens. This design not only reveals the common features of upper respiratory immune responses and alterations in the flora but also represents the broad clinical phenotype of ILI.

Through flora-associated cell type analysis, we identified a total of 205 flora associated with 13 immune cell subtypes, with each immune cell subtype identifying 1–4 flora. Notably, statistically significant associations were identified between 205 species of flora and at least one immune cell types, which may reveal the potential regulatory roles of flora in modulating immune responses.In the gastrointestinal tract, the characteristic feature of acute mucosal infections is dysbiosis, accompanied by significant changes in flora and bacterial dominance. This dysbiotic state may enhance

Table 5. The functional enrichment analysis of cell type-specific expression genes related to flora community.

Bacterial community	Cell Type	Data Base	Pathway/Biological Process	Enrichment Ratio	*FDR*
			aminopeptidase activity	0.167	0.010
			intramembrane lipid transporter activity	0.167	0.010
			calcium-dependent phospholipid binding	0.167	0.011
			chemokine binding	0.167	0.008
saliva:k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Burkholderiales; f__Neisseriaceae; g__Neisseria; s__Neisseria_sicca_A_mgs_986	CD14+Monocyte	GO	regulation of blood vessel remodeling	0.2	0.023
			negative regulation of extracellular matrix organization	0.2	0.023
			negative regulation of collagen metabolic process	0.2	0.023
			regulation of extracellular matrix disassembly	0.2	0.023
			cotranslational protein targeting to membrane	0.2	0.024
			glycoprotein catabolic process	0.2	0.024
			leucine zipper domain binding	0.2	0.011
			LRR domain binding	0.2	0.011

Bacterial community	Cell Type	Data Base	Pathway/Biological Process	Enrichment Ratio	*FDR*
tongue:k__Bacteria; p__Campylobacterota; c__Campylobacteria; o__Campylobacterales; f__Campylobacteraceae; g__Campylobacter_A; s__unclassified_mgs_1939	CD4+T effector memory	GO	secretory granule lumen	0.7	5.95E-06
			cytoplasmic vesicle lumen	0.7	5.95E-06
			vesicle lumen	0.7	5.95E-06
		KEGG	Prostate cancer	1	0.002
			Fluid shear stress and atherosclerosis	1	0.002
saliva:k__Bacteria;p__Campylobacterota; c__Campylobacteria; o__Campylobacterales; f__Campylobacteraceae; g__Campylobacter_A; s__unclassified_mgs_1397	B	GO	glutamate receptor binding	0.375	2.17E-05
			spindle pole	0.375	0.001
		KEGG	Pathways of neurodegeneration - multiple diseases	0.5	0.041
			Dopaminergic synapse	0.333	0.041
			Oocyte meiosis	0.333	0.041
			Fluid shear stress and atherosclerosis	0.333	0.041
			Adrenergic signaling in cardiomyocytes	0.333	0.041
			Cellular senescence	0.333	0.041
			Autophagy - animal	0.333	0.041

Bacterial community	Cell Type	Data Base	Pathway/Biological Process	Enrichment Ratio	*FDR*
saliva:k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Burkholderiales; f__Neisseriaceae; g__Eikenella; s__unclassified_mgs_2545	Proliferating cell	GO	plus-end-directed microtubule motor activity	0.5	7.29E-05
			microtubule motor activity	0.5	< 0.001
		KEGG	Motor proteins	1	0.001
			Vasopressin-regulated water reabsorption	0.5	0.001
saliva:k__Bacteria; p__Firmicutes; c__Clostridia; o__TANB77; f__CAG-508; g__CAG-793; s__unclassified_mgs_1417	cDC	GO	mitochondrial protein-containing complex	0.5	9.60E-05
			oxidoreductase complex	0.375	< 0.001
			mitochondrial matrix	0.5	< 0.001
			mitochondrial inner membrane	0.5	< 0.001
			inner mitochondrial membrane protein complex	0.375	< 0.001

Bacterial community	Cell Type	Data Base	Pathway/Biological Process	Enrichment Ratio	*FDR*

(Continued)

Table 5. (Continued)

Bacterial community	Cell Type	Data Base	Pathway/Biological Process	Enrichment Ratio	FDR
tongue:k__Bacteria; p__Proteobacteria; c__Gammaproteobacteria; o__Enterobacterales; f__Pasteurellaceae; g__Aggregatibacter; s__Aggregatibacter_actinomycetemcomitans_ mgs_301	CD4 + T effector memory	GO	positive regulation of establishment of protein localization to telomere	0.25	< 0.001
			regulation of establishment of protein localization to telomere	0.25	< 0.001
			regulation of protein localization to Cajal body	0.25	< 0.001
			positive regulation of protein localization to Cajal body	0.25	< 0.001
			regulation of establishment of protein localization to chromosome	0.25	< 0.001
			protein localization to nuclear body	0.25	< 0.001
			positive regulation of protein localization to chromosome, telomeric region	0.25	< 0.001
			protein localization to Cajal body	0.25	< 0.001
saliva:k__Bacteria;p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Prevotella; s__Prevotella_loescheii_mgs_2547	NK	GO	regulation of leukocyte mediated cytotoxicity	0.5	< 0.001
			regulation of cell killing	0.5	< 0.001
		KEGG	Natural killer cell mediated cytotoxicity	0.6	0.001
			Antigen processing and presentation	0.4	0.007

Bacterial community	Cell Type	Data Base	Pathway/Biological Process	Enrichment Ratio	FDR
tongue:k__Bacteria; p__Campylobacterota; c__Campylobacteria; o__Campylobacterales; f__Campylobacteraceae; g__Campylobacter_A; s__Campylobacter_A_showae_A_mgs_3545	CD8 + T effector	GO	cellular response to starvation	0.6	< 0.001
			response to starvation	0.6	< 0.001
			cellular response to nutrient levels	0.6	< 0.001
saliva:k__Bacteria; p__Campylobacterota; c__Campylobacteria; o__Campylobacterales; f__Campylobacteraceae; g__Campylobacter_A; s__unclassified_mgs_2225	CD16 + Monocyte	GO	ficolin-1-rich granule	0.288	< 0.001
			ficolin-1-rich granule lumen	0.214	0.002
tongue:k__Bacteria;; p__Firmicutes; c__Clostridia; o__Peptostreptococcales; f__Anaerovoracaceae; g__Eubacterium_B; s__Eubacterium_B_infirmum_mgs_800	cDC	GO	hydrolase activity, hydrolyzing O-glycosyl compounds	0.273	< 0.001
			hydrolase activity, acting on glycosyl bonds	0.273	0.001
		KEGG	Other glycan degradation	0.222	0.003

Bacterial community	Cell Type	Data Base	Pathway/Biological Process	Enrichment Ratio	FDR
tongue:k__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__Catonella	NK	GO	tubulin binding	0.6	< 0.001
			protease binding	0.4	0.001
		KEGG	Polycomb repressive complex	0.5	0.015
			Platelet activation	0.5	0.015

saliva_pheno.2184: The stran identified as unclassified_mgs_2621 within the genus Granulicatella, isolated from the saliva, was defined through metagenomic analysis; saliva_pheno.307: The stran identified as unclassified_mgs_986 within the genus Neisseria, isolated from the saliva, was defined through metagenomic analysis; tongue_pheno.2645:The stran identified as unclassified_mgs_1939 within the genus Campylobacter_A, isolated from the dorsal surface of the tongue, was defined through metagenomic analysis; saliva_pheno.2783:The stran identified as unclassified_mgs_1397 within the genus Campylobacter_A, isolated from the saliva, was defined through metagenomic analysis; saliva_pheno.2390:The stran identified as unclassified_mgs_2545 within the genus Eikenella, isolated from the saliva, was defined through metagenomic analysis; saliva_pheno.1286:The strain unclassified_mgs_1417, defined by 16S rRNA gene sequencing, within the genus associated with saliva CAG-793; tongue_pheno.485:Within the genus Aggregatibacter, Aggregatibacter_actinomycetemcomitans_mgs_301 is a rod-shaped stran isolated from the dorsal surface of the tongue; saliva_pheno.1016: Within the genus Prevotella, Prevotella_loescheii_mgs_2547 is a strain isolated from saliva; tongue_pheno.760:Within the genus Campylobacter_A, Campylobacter_A_showae_A_mgs_3545 is a strain isolated from the dorsal surface of the tongue;saliva_pheno.2774:The stran identified as unclassified_mgs_2225 within the genus Campylobacter_A, isolated from the saliva, was defined through metagenomic analysis; tongue_pheno.71:Within the genus Eubacterium_B, Eubacterium_B_infirmum_mgs_800 is a strain isolated from the dorsal surface of the tongue; tongue_pheno.29: Within the genus Catonella, isolated from the dorsal surface of the tongue.

https://doi.org/10.1371/journal.pone.0322864.t005

pathogen invasiveness and local inflammatory responses, thereby exacerbating inflammation and tissue damage [55]. Previous studies have shown that interactions within the flora, mediated by metabolites that modulate innate immune responses, regulate synergistic mechanisms that drive the stability of flora. Therefore, the correlation between multiple flora and at least one immune cell type may help explain why the upper respiratory tract flora does not rely on a single or few "key flora," but rather exerts its effects through a complex regulatory network formed by the overall composition of the community, thereby reflecting the collective function of the flora [56].The surface components of the Granulicatella flora can activate CD14 + monocyte cells via the Toll-like receptor, which recognizes pathogens through pathogen-associated molecular patterns. ligation of Toll-like receptor - pathogen-associated molecular patterns initiates Toll-like receptor signalling, which is critical for the innate immune cascade that induces cytokine or chemokine secretion and cell recruitment for pathogen clearance [57]. In addition, these CD14 + monocytes eliminate bacteria via phagocytosis and initiate further immune responses by differentiating into macrophages or dendritic cells [57]. This flora influences the phenotype of ILI by modulating the response of the body's immune cells. Another study revealed that NK cells play a critical role in antiviral and antitumour immunity. The function of NK cells is regulated by their surface CD16 receptors, a process that can be influenced by different bacterial populations. In HIV-1-infected individuals, the number of CD16 + cells increases [58], and the interaction between Campylobacterota flora and NK cells may influence the host immune response by modulating NK cell activity and quantity. This study revealed that these associations may be partially related to their roles in phagocytosis and antiviral activities in CD14 + monocyte cells, NK cells, and CD16 + monocyte cells, as well as in regulating NK cell activity. Campylobacter_A_showae_A_mgs_3545 from the genus Campylobacter_A on the dorsal tongue and unclassified_mgs_2621 from the genus Granulicatella in saliva Infection with the two flora can result in local and systemic immune responses, including the activation and differentiation of CD4 + naive T cells, which, upon encountering Campylobacter_A_showae antigens displayed by antigen-presenting cells, further differentiate into helper T cells (Th1, Th2, Th17) and regulatory T cells that regulate the immune response. As previous studies have shown, the genus Streptococcus is a common flora in the oral cavity and stimulates naive CD8 + T cells through antigen-presenting cells, leading to their activation and differentiation into cytotoxic T lymphocytes, which, in the local environment of the oral cavity, can directly kill infected epithelial cells and control the spread of bacterial infection and coinfection with viruses and can directly kill infected epithelial cells and control the spread of bacterial infection and coinfection with viruses, thereby reducing the bacterial load, which peaks after 7 days of ILI infection [59,60]. In this setting, other associated immune cells show reduced phagocytosis and increased levels of oxidative free radicals, associated with the suppression of proinflammatory cytokine secretion. In the present study, the Streptococcus genus flora was hypothesized to effectively activate CD8 + naive T cells via an antigen presentation mechanism, prompting them to differentiate into highly efficient CD8 + T cells, which actively participate in the viral clearance process. The extent of ILI infection was significantly controlled, effectively preventing its progression to more severe stages. Furthermore, these CD8 + T cells further differentiate into CD8 + T effector memory cells, which may provide long-lasting protection for future immune defense, enhancing the body's ability to fight against similar pathogens. Studies have shown that the Firmicutes phylum not only plays a critical role in maintaining intestinal homeostasis but also affects CD16 + monocyte function through its metabolites. These metabolites can enhance the anti-inflammatory and immunomodulatory functions of monocytes by regulating their phenotype, thereby providing protection against infections and chronic inflammatory diseases [61]. Among all significantly identified flora in this study, seven phyla were found to be correlated with NK cells, CD16 + monocytes, CD14 + monocyte cells, and the metabolites of the flora may affect the regulation of the phenotype of mononuclear immune cells and their anti-inflammatory and immunomodulatory functions. In patients with severe ILI, changes in the abundance of the Firmicutes phylum in the blood may affect the expression of contact-dependent genes in CD16 + monocyte cells by regulating the expression of inflammation-related proteins such as CX3CR1-CX3CL1 and IL-1β, thereby influencing the body's immune recovery capacity [62]. The abundance of the Firmicutes and Campylobacterota phyla was particularly notable

In addition, certain strains of the Prevotella genus have complex relationships with human health, with some strains being associated with inflammation, metabolic diseases, and immune responses [63]. Prevotella flora stimulate epithelial cells to produce IL-8, IL-6, and CCL20, thereby promoting mucosal Th17 immune responses and neutrophil recruitment. Prevotella genus-mediated mucosal inflammation leads to the systemic dissemination of inflammatory mediators, bacteria, and bacterial products, which in turn may influence systemic disease outcomes [64]. The associations found in the present study between these strains and various responses and inflammatory factors may be related to their effects on the activity of specific B cells, CD4＋T effector cells, and NK cells, as well as the expression of specific genes or proteins. Further research is needed to confirm these findings.

In the present study, certain genes associated with the genera Streptococcus CD14＋monocyte cells, NK cells and CD16＋monocyte cells also exhibited significant differential expression between the severe and mild ILI groups. Functional enrichment analysis of these DEGs revealed that in the genus Streptococcus, the DEGs associated with the dorsum of the tongue metagenome, unclassified_mgs_1057, were significantly enriched in pathways related to nuclear androgen receptor binding and transcription coactivator activity. These findings suggest that the genus Streptococcus tongue dorsalis may influence the progression of ILI by initiating the transcription of specific genes in the nucleus of CD16＋monocyte cells that interact with the cellular regulatory RNA polymerase II complex. Similarly, certain genes associated with Eubacterium in cDCs were similarly and significantly differentially expressed between the severe and mild ILI groups. Further functional enrichment analysis of these differentially expressed genes revealed that those associated with Eubacterium B infirmum mgs 800 on the dorsum of the tongue were significantly enriched in processes related to the azurophil granule membrane, lysosomal lumen, hydrolase activity, glycosidase activity, and transcription initiation factor binding. This finding suggests a mechanism by which the strain may influence the course of ILI disease by regulating enzyme activity, activating transcription factors, and regulating gene expression in combination with specific DNA sequences. The present study also revealed that the dorsal tongue unclassified_mgs_1939 of the genus Campylobacter_A was associated with CD4＋T effector memory in patients with mild ILI with differential gene expression. TWAS and pathway enrichment analysis revealed that these DEGs were enriched in biological processes such as the secretory granule lumen, prostate cancer, fluid shear stress and atherosclerosis. These findings suggest that the effect of the dorsal tongue metagenomics definition unclassified_mgs_1939 on the body's infection-associated immune response may be related to its effect on regulatory signalling in CD4＋T effector memory cells or the production of immunosuppressive molecules in CD4＋T effector memory cells. However, "enrichment" does not necessarily imply that these genes exhibit higher expression levels in severe patients; rather, it reflects a stronger association or activation trend of these genes within specific immune responses or signaling pathways. Regarding the differential regulation of immune cell gene expression by the same associated flora, the upper respiratory tract flora can directly regulate immune cell gene expression through its metabolites, thereby leading to differences in gene expression within the same immune cell type. Severe ILI patients typically exhibit a more pronounced immune response, this immune state is closely associated with the interaction between the host and the upper respiratory tract flora. The same flora may exert distinct biological effects under different immune conditions, ultimately contributing to differences in immune cell gene expression.

The main limitations of the present study are as follows. First, the integrative analysis for identifying flora-associated cell types was based on the assumption that interindividual differences in the abundance of upper respiratory tract flora could be attributed to genetic variation to some extent. Although the study acquired all the data of upper respiratory tract flora-associated genome-wide association studies from the National Gene Bank of China database and only the GWAS summary data with h2＞0.1 were included in the integration analysis, the GWAS revealed only the association between genetic variation and phenotype rather than direct causality, while the mechanism of the effect of fewer heritable flora on ILI remained unclear. Second, the relatively small sample size (n＝2984) of the GWAS pooled data included in this study may be one of the reasons for the unsatisfactory signals of the integrated analysis and the TWAS analysis. In addition, in the identification of flora-associated cell type-specific expressed genes, the eQTL model used was constructed from the

EUR population, and although genomic data from different races were used as reference panels in the model construction (EUR) and TWAS analysis (EAS), differences in genetic structure between races may still be an important reason why fewer significant genes were identified in this part of the study than in similar studies. Finally, this part of the study investigated only the associations between flora and immune cells from a transcriptional and metabolic perspective, and additional epigenetic and protein studies may be considered to gain a more comprehensive understanding of the underlying mechanisms involved in "upper respiratory flora-immunity-ILI progression". Moreover, the conclusions of the bioinformatics analysis should be further verified by experimental and population studies.

In summary, a total of 205 flora associated with 13 cell types were analysed in this study, and cell-specific eQTL models were further constructed and analysed via TWAS. On the basis of this analysis, 1226 flora-associated cell type-specific genes with significantly different expression levels between patients with mild and severe ILI were identified ($P_{adj} < 0.05$ and $|\log_2 FC| > 0.25$). Finally, by analysing the genes specifically expressed in cell types associated with the significantly differentially expressed flora, pathways or biological processes related to ILI progression were enriched.

Supporting information

S1 Supporting information: This zip file contains supplementary files, figures to the study.
(ZIP)

Author contributions

Data curation: Xiangyu Ye.

Resources: Xiangyu Ye.

Supervision: Ran He, Yifan Wang, Shirong Hui, Haochang Li, Hongbo Chen, Peng Huang.

Writing – original draft: Lei Xu.

References

1. Paules C, Subbarao K. Influenza. Lancet. 2017;390(10095):697–708. https://doi.org/10.1016/S0140-6736(17)30129-0 PMID: 28302313
2. Iuliano AD, Roguski KM, Chang HH, Muscatello DJ, Palekar R, Tempia S, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet. 2018;391(10127):1285–300. https://doi.org/10.1016/S0140-6736(17)33293-2 PMID: 29248255
3. Eisenstein M. Towards a universal flu vaccine. Nature. 2019;573(7774):S50–2. https://doi.org/10.1038/d41586-019-02751-w PMID: 31534251
4. Sellers SA, Hagan RS, Hayden FG, Fischer WA 2nd. The hidden burden of influenza: A review of the extra-pulmonary complications of influenza infection. Influenza Other Respir Viruses. 2017;11(5):372–93. https://doi.org/10.1111/irv.12470 PMID: 28745014
5. Medina RA, Garcia-Sastre A. Influenza A viruses: new research developments. Nat Rev Microbiol. 2011;9(8):590–603.
6. Moreno G, Rodríguez A, Reyes LF, Gomez J, Sole-Violan J, Díaz E, et al. Corticosteroid treatment in critically ill patients with severe influenza pneumonia: a propensity score matching study. Intensive Care Med. 2018;44(9):1470–82. https://doi.org/10.1007/s00134-018-5332-4 PMID: 30074052
7. Fuse Brown EC, Sarpatwari A. Removing ERISA's Impediment to State Health Reform. N Engl J Med. 2018;378(1):5–7. https://doi.org/10.1056/nejmp1709167 PMID: 29298152
8. Belongia EA, Simpson MD, King JP, Sundaram ME, Kelley NS, Osterholm MT, et al. Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. Lancet Infect Dis. 2016;16(8):942–51. https://doi.org/10.1016/S1473-3099(16)00129-8 PMID: 27061888
9. Krammer F, Smith GJD, Fouchier RAM, Peiris M, Kedzierska K, Doherty PC, et al. Influenza. Nat Rev Dis Primers. 2018;4(1):3.
10. Ichinohe T, Pang IK, Kumamoto Y, Peaper DR, Ho JH, Murray TS, et al. Microbiota regulates immune defense against respiratory tract influenza A virus infection. Proc Natl Acad Sci U S A. 2011;108(13):5354–9. https://doi.org/10.1073/pnas.1019378108 PMID: 21402903
11. Abt MC, Osborne LC, Monticelli LA, Doering TA, Alenghat T, Sonnenberg GF, et al. Commensal bacteria calibrate the activation threshold of innate antiviral immunity. Immunity. 2012;37(1):158–70. https://doi.org/10.1016/j.immuni.2012.04.011 PMID: 22705104
12. Levy M, Blacher E, Elinav E. Microbiome, metabolites and host immunity. Curr Opin Microbiol. 2017;35:8–15. https://doi.org/10.1016/j.mib.2016.10.003 PMID: 27883933
13. Thibeault C, Suttorp N, Opitz B. The microbiota in pneumonia: From protection to predisposition. Sci Transl Med. 2021;13(576):eaba0501. https://doi.org/10.1126/scitranslmed.aba0501 PMID: 33441423

14. Robinson MR, Kleinman A, Graff M, Vinkhuyzen AAE, Couper D, Miller MB, et al. Genetic evidence of assortative mating in humans. Nature Human Behaviour. 2017;1(1):0016.

15. Ramos-Sevillano E, Wade WG, Mann A, Gilbert A, Lambkin-Williams R, Killingley B, et al. The effect of influenza virus on the human oropharyngeal microbiome. Clin Infect Dis. 2019;68(12):1993–2002. https://doi.org/10.1093/cid/ciy821 PMID: 30445563

16. Tsang TK, Lee KH, Foxman B, Balmaseda A, Gresh L, Sanchez N, et al. Association between the respiratory microbiome and susceptibility to influenza virus infection. Clin Infect Dis. 2020;71(5):1195–203. https://doi.org/10.1093/cid/ciz968 PMID: 31562814

17. Roggiani S, Zama D, D'Amico F, Rocca A, Fabbrini M, Totaro C, et al. Gut, oral, and nasopharyngeal microbiota dynamics in the clinical course of hospitalized infants with respiratory syncytial virus bronchiolitis. Front Cell Infect Microbiol. 2023;13:1193113. https://doi.org/10.3389/fcimb.2023.1193113 PMID: 37680746

18. Rosas-Salazar C, Kimura KS, Shilts MH, Strickland BA, Freeman MH, Wessinger BC, et al. Upper respiratory tract microbiota dynamics following COVID-19 in adults. Microb Genom. 2023;9(2):mgen000957. https://doi.org/10.1099/mgen.0.000957 PMID: 36820832

19. Jansen JM, Gerlach T, Elbahesh H, Rimmelzwaan GF, Saletti G. Influenza virus-specific CD4+ and CD8+ T cell-mediated immunity induced by infection and vaccination. J Clin Virol. 2019;119:44–52. https://doi.org/10.1016/j.jcv.2019.08.009 PMID: 31491709

20. Pizzolla A, Wakim LM. Memory T cell dynamics in the lung during influenza virus infection. J Immunol. 2019;202(2):374–81. https://doi.org/10.4049/jimmunol.1800979 PMID: 30617119

21. Koutsakos M, Illing PT, Nguyen THO, Mifsud NA, Crawford JC, Rizzetto S, et al. Human CD8+ T cell cross-reactivity across influenza A, B and C viruses. Nat Immunol. 2019;20(5):613–25. https://doi.org/10.1038/s41590-019-0320-6 PMID: 30778243

22. Schmidt ME, Knudson CJ, Hartwig SM, Pewe LL, Meyerholz DK, Langlois RA, et al. Memory CD8 T cells mediate severe immunopathology following respiratory syncytial virus infection. PLoS Pathog. 2018;14(1):e1006810. https://doi.org/10.1371/journal.ppat.1006810 PMID: 29293660

23. Channappanavar R, Fett C, Zhao J, Meyerholz DK, Perlman S. Virus-specific memory CD8 T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus infection. J Virol. 2014;88(19):11034–44. https://doi.org/10.1128/JVI.01505-14 PMID: 25056892

24. Zhao Y, Kilian C, Turner JE, Bosurgi L, Roedl K, Bartsch P, et al. Clonal expansion and activation of tissue-resident memory-like Th17 cells expressing GM-CSF in the lungs of severe COVID-19 patients. Science immunology. 2021;6(56).

25. <Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.pdf>.

26. <A haplotype map of the human genome.pdf>.

27. Manolio TA, Brooks LD, Collins FS. A HapMap harvest of insights into the genetics of common disease. J Clin Invest. 2008;118(5):1590–605. https://doi.org/10.1172/JCI34772 PMID: 18451988

28. MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017;45(D1):D896–901. https://doi.org/10.1093/nar/gkw1133 PMID: 27899670

29. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48(3):245–52. https://doi.org/10.1038/ng.3506 PMID: 26854917

30. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017;101(1):5–22.

31. Cao C, Ding B, Li Q, Kwok D, Wu J, Long Q. Power analysis of transcriptome-wide association study: Implications for practical protocol choice. PLoS Genet. 2021;17(2):e1009405. https://doi.org/10.1371/journal.pgen.1009405 PMID: 33635859

32. Liu X, Tong X, Zhu J, Tian L, Jie Z, Zou Y, et al. Metagenome-genome-wide association studies reveal human genetic impact on the oral microbiome. Cell Discov. 2021;7(1):117. https://doi.org/10.1038/s41421-021-00356-0 PMID: 34873157

33. Zhu J, Tian L, Chen P, Han M, Song L, Tong X, et al. Over 50,000 metagenomically assembled draft genomes for the human oral microbiome reveal new taxa. Genomics Proteomics Bioinformatics. 2022;20(2):246–59.

34. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. Int J Syst Evol Microbiol. 2020;70(11):5607–12. https://doi.org/10.1099/ijsem.0.004332 PMID: 32701423

35. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74. https://doi.org/10.1038/nature15393 PMID: 26432245

36. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47(3):291–5. https://doi.org/10.1038/ng.3211 PMID: 25642630

37. Lee JS, Park S, Jeong HW, Ahn JY, Choi SJ, Lee H, et al. Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. Sci Immunol. 2020;5(49):eabd1554. https://doi.org/10.1126/sciimmunol.abd1554 PMID: 32651212

38. Youngs J, Provine NM, Lim N, Sharpe HR, Amini A, Chen Y-L, et al. Identification of immune correlates of fatal outcomes in critically ill COVID-19 patients. PLoS Pathog. 2021;17(9):e1009804. https://doi.org/10.1371/journal.ppat.1009804 PMID: 34529726

39. Germain PL, Lun A, Garcia Meixide C, Macnair W, Robinson MD. Doublet identification in single-cell sequencing data using scDblFinder. F1000Res. 2021;10:979.

40. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;20(1):296. https://doi.org/10.1186/s13059-019-1874-1 PMID: 31870423

41. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048 PMID: 34062119

42. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289–96. https://doi.org/10.1038/s41592-019-0619-0 PMID: 31740819

43. Zhang MJ, Hou K, Dey KK, Sakaue S, Jagadeesh KA, Weinand K, et al. Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. Nat Genet. 2022;54(10):1572–80. https://doi.org/10.1038/s41588-022-01167-z PMID: 36050550

44. Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. Science. 2022;376(6589):eabf3041. https://doi.org/10.1126/science.abf3041 PMID: 35389779

45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.

46. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76–82. https://doi.org/10.1016/j.ajhg.2010.11.011 PMID: 21167468

47. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2(3):100141. https://doi.org/10.1016/j.xinn.2021.100141 PMID: 34557778

48. Yu G, He Q-Y. ReactomePA: An R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016;12(2):477–9. https://doi.org/10.1039/c5mb00663e PMID: 26661513

49. 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. Ann Am Thorac Soc. 2015;12(6):821–30. https://doi.org/10.1513/AnnalsATS.201501-029OC PMID: 25803243

50. Faner R, Sibila O, Agusti A, Bernasconi E, Chalmers JD, Huffnagle GB, et al. The microbiome in respiratory medicine: current challenges and future perspectives. Eur Respir J. 2017;49(4).

51. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe. 2016;19(5):731–43. https://doi.org/10.1016/j.chom.2016.04.017 PMID: 27173935

52. Key FM, Teixeira JC, de Filippo C, Andrés AM. Advantageous diversity maintained by balancing selection in humans. (1879-0380 (Electronic)).

53. Quintana-Murci L. Human immunology through the lens of evolutionary genetics. Cell. 2019;177(1):184–99. https://doi.org/10.1016/j.cell.2019.02.033 PMID: 30901539

54. Prugnolle F, Manica A, Charpentier M, Guégan JF, Guernier V, Balloux F. Pathogen-driven selection and worldwide HLA class I diversity. (0960-9822 (Print)).

55. Egan CE, Cohen SB, Denkers EY. Insights into inflammatory bowel disease using Toxoplasma gondii as an infectious trigger. Immunol Cell Biol. 2012;90(7):668–75. https://doi.org/10.1038/icb.2011.93 PMID: 22064707

56. Levy M, Thaiss CA, Zeevi D, Dohnalová L, Zilberman-Schapira G, Mahdi JA, et al. Microbiota-modulated metabolites shape the intestinal microenvironment by regulating NLRP6 inflammasome signaling. Cell. 2015;163(6):1428–43. https://doi.org/10.1016/j.cell.2015.10.048 PMID: 26638072

57. Didierlaurent A, Goulding J, Hussell T. The impact of successive infections on the lung microenvironment. Immunology. 2007;122(4):457–65. https://doi.org/10.1111/j.1365-2567.2007.02729.x PMID: 17991012

58. Milush JM, Long BR, Snyder-Cappione JE, Cappione AJ,3rd, York VA, Ndhlovu LC, et al. Functionally distinct subsets of human NK cells and monocyte/DC-like cells identified by coexpression of CD56, CD7, and CD4. Blood. 2009;114(23):4823–31.

59. van der Sluijs KF, Nijhuis M, Levels JH, Florquin S, Mellor AL, Jansen HM, et al. Influenza-induced expression of indoleamine 2,3-dioxygenase enhances interleukin-10 production and bacterial outgrowth during secondary pneumococcal pneumonia. J Infect Dis. 2006;193(2):214–22.

60. Sun K, Metzger DW. Inhibition of pulmonary antibacterial defense by interferon-gamma during recovery from influenza infection. Nat Med. 2008;14(5):558–64. https://doi.org/10.1038/nm1765 PMID: 18438414

61. Soderquest K, Powell N, Luci C, van Rooijen N, Hidalgo A, Geissmann F, et al. Monocytes control natural killer cell differentiation to effector phenotypes. Blood. 2011;117(17):4511–8. https://doi.org/10.1182/blood-2010-10-312264 PMID: 21389319

62. Roy-Chowdhury E, Brauns N, Helmke A, Nordlohne J, Brasen JH, Schmitz J, et al. Human CD16+ monocytes promote a pro-atherosclerotic endothelial cell phenotype via CX3CR1-CX3CL1 interaction. Cardiovasc Res. 2021;117(6):1510–22.

63. Larsen JM. The immune response to Prevotella bacteria in chronic inflammatory disease. Immunology. 2017;151(4):363–74. https://doi.org/10.1111/imm.12760 PMID: 28542929

64. Doerflinger SY, Throop AL, Herbst-Kralovetz MM. Bacteria in the vaginal microbiome alter the innate immune response and barrier properties of the human vaginal epithelia in a species-specific manner. J Infect Dis. 2014;209(12):1989–99.