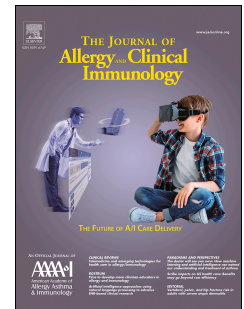


Journal Pre-proof

Lung virome: new potential biomarkers for asthma severity and exacerbation

Sungmi Choi, PhD, Kyoung-Hee Sohn, MD, PhD, Jae-Woo Jung, MD, PhD, Min-Gyu Kang, MD, PhD, Min-Suk Yang, MD, PhD, Sujeong Kim, MD, Jeong-Hee Choi, MD, PhD, Sang-Heon Cho, MD, PhD, Hye-Ryun Kang, MD, PhD, Hana Yi, PhD



PII: S0091-6749(21)00473-5

DOI: <https://doi.org/10.1016/j.jaci.2021.03.017>

Reference: YMAI 15037

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 2 November 2020

Revised Date: 8 March 2021

Accepted Date: 11 March 2021

Please cite this article as: Choi S, Sohn K-H, Jung J-W, Kang M-G, Yang M-S, Kim S, Choi J-H, Cho S-H, Kang H-R, Yi H, Lung virome: new potential biomarkers for asthma severity and exacerbation, *Journal of Allergy and Clinical Immunology* (2021), doi: <https://doi.org/10.1016/j.jaci.2021.03.017>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Inc. on behalf of the American Academy of Allergy, Asthma & Immunology.

Lung virome: new potential biomarkers for asthma severity and exacerbation

Sungmi Choi, PhD^{1,2}, Kyoung-Hee Sohn, MD, PhD^{3,4}, Jae-Woo Jung, MD, PhD⁵, Min-Gyu Kang, MD, PhD⁶, Min-Suk Yang, MD, PhD⁷, Sujeong Kim, MD⁸, Jeong-Hee Choi, MD, PhD^{9,10}, Sang-Heon Cho, MD, PhD^{3,11}, Hye-Ryun Kang, MD, PhD^{3,11†}, and Hana Yi, PhD^{1,2,12‡}

¹Institute for Biomaterials, Korea University, Seoul, Korea

²Interdisciplinary Program in Precision Public Health, Korea University, Seoul, Korea

³Institute of Allergy and Clinical Immunology, Seoul National University Medical Research Center, Seoul National University College of Medicine, Seoul, Korea

⁴Department of Internal Medicine, Kyung Hee University Hospital, Seoul, Korea

⁵Department of Internal Medicine, Chung-Ang University College of Medicine, Seoul, Korea

⁶Department of Internal Medicine, Chungbuk National University College of Medicine, Chungbuk National University Hospital, Cheongju, Korea

⁷Department of Internal Medicine, SMG-SNU Boramae Medical Center, Seoul, Korea

⁸Department of Internal Medicine, School of Medicine, Kyungpook National University, Kyungpook National University Hospital, Daegu, Korea

⁹Department of Pulmonology and Allergy, Hallym University Dongtan Sacred Heart Hospital, Hwaseong, Korea

¹⁰Allergy and Clinical Immunology Research Center, Hallym University College of Medicine, Chuncheon, Korea

¹¹Department of Internal Medicine, Seoul National University Hospital, Seoul, Korea

¹²School of Biosystems and Biomedical Sciences, Korea University, Seoul, Korea

[‡]Correspondence:

Hana Yi, PhD

School of Biosystems and Biomedical Sciences, Korea University

145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea

Tel. 82-2-3290-5644 Fax: 82-2-940-2849

E-mail: hanayi@korea.ac.kr

Hye-Ryun Kang, MD, PhD

Seoul National University Hospital

101 Daehak-ro, Jongno-Gu, Seoul 110-744, Korea

Tel. 82-2-2072-0820 Fax: 82-2-742-3291

E-mail: helenmed@snu.ac.kr

Total Words: 3,258

Conflict of Interest: The authors declare no conflict of interest.

This research was supported by a fund (2016ER670400 and 2017ER670100) from the Research of Korea Centers for Disease Control and Prevention.

Abstract

Background: Although some respiratory virus infections are known to contribute to the development and exacerbation of asthma, commensal virome in airway have not been extensively studied due to technical challenges.

Objectives: We investigated the characteristics of the virome in asthmatic airways.

Methods: Both the bacteriome and virome profiles in sputum from 12 healthy individuals, 15 patients with non-severe asthma, and 15 patients with severe asthma were analyzed and assessed for the association with clinical characteristics such as severity, exacerbation, Asthma Control Test (ACT), and lung function.

Results: While analysis of the 16S rRNA bacteriome in the airway showed no differences, clear contrasts in the diversity and composition of airway virome were observed between healthy controls and patients with asthma. Herpesviruses were the most abundant type of virus in the asthma group ($44.6\% \pm 4.6\%$), mainly with cytomegalovirus (CMV) and Epstein-Barr virus (EBV) accounting for $24.5\% \pm 3.3\%$ and $16.9\% \pm 3.5\%$ respectively in contrast to those in the healthy controls ($5.4\% \pm 2.5\%$ and $7.1\% \pm 3.0\%$, respectively). CMV and EBV were more abundant in asthmatics who experienced exacerbation and their abundance showed correlation with more severe asthma, lower ACT score, and lower lung function. On the contrary, bacteriophage abundant in healthy controls was severely reduced in asthma patients in the order of non-severe and severe asthma and presented significant positive correlation with ACT and FEV_1/FVC .

Conclusion: Lung virome, especially, CMV, EBV, and bacteriophage may be potential biomarkers of asthma severity and exacerbation.

Clinical Implications: The abundance of human herpesvirus such as cytomegalovirus and Epstein Barr virus or bacteriophages in lung may be considered as a new potential biomarker of severity and exacerbation in asthmatic patients.

Capsule Summary: Lung virome alteration such as increased cytomegalovirus and Epstein Barr virus as well as reduced bacteriophage were observed in asthma patients and presented good positive or negative correlation with severity and symptoms score of asthma, and lung function.

Keywords: asthma, herpesvirus, CMV, EBV, virome, metagenomics, microbiome

Abbreviations: CMV, cytomegalovirus; EBV, Epstein-Barr virus; RSV, respiratory syncytial virus; AHR, airway hyper-responsiveness; FEV₁, forced expiratory volume in 1s; ACQ, Asthma Control Questionnaire; ACT, Asthma Control Test; PCR, polymerase chain reaction; PCoA, principal coordinate analysis; LEfSe, linear discriminant analysis effect size; HSV, herpes simplex virus; FVC, forced vital capacity; HHV, human herpesvirus; BALF, bronchoalveolar lavage fluid; COPD, chronic obstructive pulmonary disease

INTRODUCTION

Asthma, characterized by cough, dyspnea, and wheezing as a result of chronic airway inflammation, is a heterogeneous disorder caused by various pathophysiological mechanisms.¹ Asthma results from a combination of genetic and environmental factors, and immunologic responses to allergens such as house dust mites, pollen, and animal fur are the key mechanisms underlying asthma pathogenesis.

Viruses are also believed to play important roles in the development and exacerbation of asthma. In particular, many studies on respiratory virus infections and allergic responses have reported that respiratory viruses are associated with 80% of asthma exacerbation episodes.^{1,2} Among the various viruses, rhinoviruses are known to have a major influence in the development of asthma; other viruses known to be associated with virus-induced asthma include influenza virus and respiratory syncytial virus (RSV).³ Notably, based on the hygiene hypothesis, frequent infections during childhood are expected to lead to the development of protective immunity such as Th1 immunity. Therefore, exposure to commensal or nonpathogenic viruses is considered to contribute to a reduced risk for the development of asthma and allergies.⁴ However, the presence or alterations of the commensal airway viral flora and their roles in the development of asthma are not clearly understood.

The virome refers to the collection of all viruses, including those infecting eukaryotic and prokaryotic cells, that reside in and on the human body and constitute a large part of the human microbiome.⁵ Although few studies have assessed the virome owing to its unculturability, recent virome studies demonstrated that the viral flora is an important factor in host health and diseases.⁶⁻⁸ Unknown members of the virome are referred to as viral dark matter.⁹ Because of the technical challenges involved in virome experiments and analyses, the virome has not been extensively studied compared with the bacteriome, and only a few

studies have investigated the associations between the virome and asthma. The first virome analysis performed in patients with asthma was carried out in 2009, which evaluated DNA viral communities in patients with and without cystic fibrosis.¹⁰ Additionally, a 2018 study analyzed the virome and bacteriome of nasopharyngeal swabs from 134 children with asthma and reported that rhinovirus-C, Bolarra virus-1, and RSV-B were the predominant species.¹¹ However, these are the only reports of virome analyses in patients with asthma.

Therefore, in the current study, we aimed to identify and characterize the virome using induced sputum collected from healthy participants and asthma patients. Our findings suggest that herpesviruses may be used as biomarkers for asthma severity.

METHODS

Sample collection

In total, 42 patients (12 healthy individuals, 15 patients with non-severe asthma, and 15 patients with severe asthma) were recruited. This study was approved by the Seoul National University Hospital Institutional Review Board (IRB approval number: 1607-148-778) and was conducted in accordance with Declaration of Helsinki. Informed consent was obtained from all patients. Asthma was diagnosed according to the presence of chronic respiratory symptoms such as cough, sputum, dyspnea, and wheezing as well as airway reversibility or airway hyper-responsiveness (AHR). Airway reversibility was defined as an increase in the forced expiratory volume in 1s (FEV₁) of more than 12% and 200 mL or greater when measured 15 min after inhalation of a short-acting β 2 agonist. AHR was defined as positive if less than 16 mg/mL methacholine in a bronchial provocation test induced a 20% reduction in FEV₁. Severe asthma was defined as an Asthma Control Questionnaire (ACQ) score > 1.50 or an FEV₁ that was 40–80% of the predicted value despite the use of a daily inhaled

corticosteroid dose equivalent to $\geq 1,000$ μg fluticasone. All patients were nonsmokers or ex-smokers (less than 20 pack-years). Individuals who were treated with antibiotics or experienced acute exacerbation within the past three months were excluded from the study. The baseline characteristics of the enrolled patients were investigated, including pulmonary function tests, skin tests, serum total IgE, Asthma Control Test™ (ACT) score, and peripheral blood cell counts. Asthma exacerbation was defined according to the American Thoracic Society/European Respiratory Society Statement as including at least one of the following: (1) systemic use of corticosteroid for consecutive three days, (2) asthma-specific emergency department visits or hospitalization and, (3) change in $\text{FEV}_1 > 20\%$ compared to personal best baseline.¹²

Construction of the virome sequencing library

Whole sputum samples were pretreated with dithiothreitol to reduce viscosity; sequentially prefiltered using syringe filters with pore sizes of 5, 2.2, and 0.45 μm ; and then concentrated using a 20 mL Vivaspin® 20 ultrafiltration device (Sartorius Stedim Biotech, Göttingen, Germany) equipped with a 30,000 kDa pore-sized polyethersulfone membrane. The centrifugation time at 10,000 g varied between 1–2 h depending on the sample. Consequently, samples of approximately 6 mL were concentrated about 12-fold to a final volume < 500 μL . The concentrated samples were treated with DNase I and RNase A to remove free nucleic acids and then concentrated to 200 μL using a 2 mL Vivaspin® 2 ultrafiltration device (Sartorius Stedim Biotech). Both DNA and RNA were extracted using a QIAamp MinElute Virus Spin Kit (Qiagen, Valencia, CA). First-strand cDNA was synthesized using a SuperScript III First-Strand Synthesis System (Invitrogen Inc., Carlsbad, CA) and V-N random primer (5'-GTTTCCCAGTCACGATCNNNNNN-3'). Second-strand cDNA was

synthesized using a large Klenow fragment (Invitrogen Inc.) and V-N random primers. To obtain sufficient DNA for sequencing, a polymerase chain reaction (PCR) was performed using the synthesized cDNA as the template. V primer (5'-GTTTCCCAGTCACGATC-3') was used for the PCR. The constructed virome amplicon was sent to a commercial sequencing service provider (Chunlab, Seoul, Korea) and sequenced as 150 bp paired-end reads using a HiSeq3000 instrument (Illumina, San Diego, CA). A detailed description of the experimental procedures is given in Supplementary text 1 and Fig. 1.

Virome sequence analysis

The sequencing raw reads were processed in the sequence of quality-filtering, pair-end merging, length-filtering, and human sequence filtering. The resultant non-human reads were identified using BLASTn (e-value cutoff: $1E-10$) against an in-house viral database which is composed of sequences downloaded (August 2019) from NCBI Viral Genomes (<https://www.ncbi.nlm.nih.gov/genome/viruses/>) and non-metagenomic viral sequences from RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>). Among the identified viral sequences, only sequences with $\geq 50\%$ of alignment length coverage were selected as verified viral sequences and used for further analyses. In addition, if the alignment coverage was lower than 90%, the ambiguous identification was marked by adding '-like' at the end of the taxon name (e.g., MERS-related coronavirus-like). Finally, to confirm the accuracy of viral reads identified as CMV or EBV, sequencing reads from an individual sample were extracted and mapped to both reference genomes using the CLC workbench (Qiagen). A detailed description is given in Supplementary text 1.

Bacteriome sequencing and analysis

Bacterial nucleic acids were extracted from 1 mL of induced sputum using a FastDNA SPIN Kit (MP Biomedicals, Santa Ana, CA). The V3–V4 region of the 16S rRNA gene was amplified to construct the sequencing library and 300-bp paired-end sequencing was conducted using an Illumina MiSeq v3 platform (Illumina). The sequencing raw data files were preprocessed using the bioinformatics pipeline described in Supplementary text 1.

Statistical Analysis

Virome beta diversity was assessed by calculating Bray-Curtis distance using viral species abundance and plotted using principal coordinate analysis (PCoA) using the stats package in R version 3.5.1. The significance of grouping in the ordination was evaluated using permutational multivariate analysis of variance (PERMANOVA) using R program. Log-transformed relative abundance was regressed on the experimental groups after adjusting for the effect of age and medication (oral corticosteroid and antibiotics) using Rex version 3.0.3.¹³ Any missing values for covariates were replaced with their mean. Inhaled corticosteroid was not included in the adjustment due to multicollinearity since it had a very high correlation with disease severity ($r = 0.947$, $P < 0.0001$). Pearson's Chi-square test for categorical variables and Student's t-test or Fisher's exact test for continuous variables were used to analyze baseline characteristics of subjects. Correlation analysis was performed to analyze the associations between virome composition and clinical variables. Pearson correlation tests were applied to assess the relationships of ACT score and FEV₁/FVC with virome abundance. Correlations between the abundance of a given virus in the total virus reads and clinical variables were calculated using Spearman's rank correlation coefficient and the Mann-Whitney U test. The software package GraphPad Prism 8.0 (GraphPad Software, San Diego, CA) was used for data analysis and preparation of graphs. Results with $P < 0.05$

205 were considered significant.

Journal Pre-proof

RESULTS

Patient characteristics and clinical parameters

Table 1 summarizes the characteristics of the 12 healthy individuals, 15 patients with non-severe asthma, and 15 patients with severe asthma enrolled in this study. Compared with the patients with non-severe asthma, those with severe asthma had worse symptom scores (ACT, ACQ), lower lung function (FEV_1 , FEV_1 /forced vital capacity [FVC]), and a higher frequency of acute exacerbation.

Sequencing statistics

The sequencing reads statistics for individual samples over the stages of the data processing pipeline are presented in Table E1. On average, approximately 40×10^6 raw reads were produced per sample (Table 2). After data processing, approximately 13×10^6 length filtered reads per sample were obtained. The length filtered reads were filtered through human genome mapping and approximately 0.4×10^6 non-human reads per sample were obtained and were subjected to a BLAST search against the in-house viral database. On average, the BLAST identified 8,400 reads per sample as viral sequences. The identified viral reads were further filtered using sequence alignment coverage (threshold $\geq 50\%$) and finally defined as verified viral reads. The number of verified viral reads varied across samples (range: 30–5,300 sequences). In summary, the length-filtered reads were identified as human (96.853%), viruses (0.063%), and others (bacteria and unknown, 3.083%). The ratio of identified (verified) viral reads accounted for 0.051% (0.004%), 0.072% (0.006%), and 0.064% (0.004%) of the sequences in the control, non-severe asthma, and severe asthma groups, respectively; no statistical differences were found between groups.

Virome in induced sputum

The virome structures in the healthy, non-severe asthma, and severe asthma groups were clearly distinguished based on beta-diversity analyses ($P < 0.05$) (Fig. 2a).

The full description for the virome composition of individual samples is provided in Table E2 and Fig. 3a and the group averages are summarized in Fig. 2c. Among all the viruses in the total sample, viruses of prokaryotes were the most abundant at $62.0\% \pm 4.5\%$, followed by herpesviruses at $35.7\% \pm 4.2\%$. Herpesviruses accounted for $13.4\% \pm 5.2\%$ of the viruses in the healthy individuals and $44.6\% \pm 4.6\%$ in the asthma patients. When analyzed in terms of herpesvirus groups, herpes simplex virus 1 (HSV-1; $3.1\% \pm 1.0\%$), cytomegalovirus (CMV; $24.5\% \pm 3.3\%$), and Epstein Barr virus (EBV, $16.9\% \pm 3.5\%$) of the alpha, beta, and gamma subfamilies, respectively, accounted for the majority of all herpesviruses in the asthma patients. Other eukaryotic viruses like as Betapapillomavirus were detected also, but they showed no differences among groups.

To verify the accuracy of CMV and EBV identification, the identified reads from individual samples were extracted and aligned to reference virus genomes (Fig. E1). As a result, the reads identified as CMV were mapped only to CMV genome but not to EBV genome, and vice versa. The results demonstrated the accuracy of viral sequence identification and also indicated the presence of large amounts of herpesviruses in the respiratory samples by showing high genome mapping coverage and depth.

The full description for the phage composition of individual samples is provided in Table E3. Bacteriophage was the majority of the virome in the healthy group, but it tended to decline in the asthma group in the order of healthy ($86.4\% \pm 5.3\%$), non-severe asthma ($70.3 \pm 4.5\%$), and severe asthma group ($34.1\% \pm 5.8\%$). Among the 20 bacteriophages identified in the samples, *Streptococcus* phage was the most abundant virus, constituting an average of

25.9% $\pm$ 5.0% of the total sample (Fig. 2c). The abundance of *Streptococcus* phage was significantly lower in the asthma group (11.1% $\pm$ 3.2% in non-severe, 5.5% $\pm$ 2.7% in severe) than in the control group (70.0% $\pm$ 7.2%) (Table E3). While *Streptococcus* phage decreased in asthma group, *Klebsiella* phage and *Escherichia* phages increased and became the most abundant bacteriophage in the non-severe and severe asthma groups, respectively.

Bacteriome in induced sputum

The bacteriome structures of the healthy, non-severe asthma, and severe asthma groups were not distinguished based on beta-diversity analyses (Fig. 2b). The most abundant bacteria were *Streptococcus* species (26.3%; Fig. 2d). Other dominant bacteria constituting 5% or more of the total samples were, in descending order, *Prevotella* (16.0%), *Neisseria* (12.8%), *Veillonella* (8.5%), and *Haemophilus* (7.6%), followed by *Porphyromonas*, *Fusobacterium*, *Leptotrichia*, *Actinomyces*, and *Alloprevotella*. In contrast to virome analysis, no bacteria showed differences in abundance among groups (Kruskal-Wallis tests, LEfSe analysis).

The results of the bacteriome and phage virome analyses were inconsistent. Whereas the virome analysis revealed differences in the abundances of *Streptococcus*, *Escherichia*, and *Klebsiella* phages depending on the presence or absence of asthma and severity, the bacteriome analysis showed similar abundances for these three bacteria with no differences among the groups.

Detection rate for individual virus among sputum samples

To know the occurrence rate of a specific virus among population, the proportion of virus-detected samples out of all the samples was calculated for each group (Fig. 3b). Human beta herpesviruses and human gamma herpesviruses, primarily CMV and EBV respectively,

were detected in 58% of sputum samples from healthy individuals while relatively more were identified in sputum samples from asthma patients in the order of non-severe (60-87%) and severe asthma (93-100%). On the contrary, *Streptococcus* phage was detected in 100% in healthy individuals and less frequently found in patients with non-severe asthma (60%), down to 47% in those with severe asthma.

Association between clinical factors and lung virome

Our virome analysis showed that herpesviruses were associated with asthma severity; therefore, we further explored this group of viruses based on clinical factors. The proportions of human beta herpesviruses and human gamma herpesviruses, primarily CMV and EBV respectively, significantly increased with asthma severity (Fig. 4a). Notably, the CMV abundance in the healthy group ($5.4\% \pm 2.5\%$) was significantly different from those in the non-severe and severe asthma groups ($18.1\% \pm 3.8\%$ and $31.0\% \pm 5.1\%$, respectively; $P = 0.009$). Similarly, EBV had different abundances among three groups ($P = 0.047$) in the order of healthy ($7.1\% \pm 3.1\%$), non-severe asthma groups ($9.3\% \pm 4.9\%$), and severe asthma group ($24.6\% \pm 4.3\%$).

Furthermore, CMV and bacteriophage abundances affected the occurrence of acute exacerbation during the one-year period in asthma patients ($n = 30$). The exacerbation positive rate was different between the high ($\geq 50^{\text{th}}$ percentile) and low ($< 50^{\text{th}}$ percentile) CMV or bacteriophage groups (Fig. 4b). In terms of subjective parameters of asthma, CMV and EBV revealed moderately negative correlations with ACT scores reflecting greater asthma control ($n = 30$, $P = 0.031$ and 0.072 , respectively, Fig. 4c). Moreover, the abundance of CMV was negatively associated with the FEV_1/FVC ratio, a marker of airway obstruction ($n = 42$, Fig. 4d). Oral corticosteroids or antibiotics usage did not show correlation with any

of the viruses identified in this study (Table E4). The CMV abundances and EBV abundances in samples were not correlated (Fig. E2), indicating the independence of the two herpesviruses in the lung. Unlike CMV and EBV, human alpha herpesviruses (HSV-1) did not show an association with any of the clinical factors.

In contrast to CMV and EBV, the proportions of bacteriophage decreased with asthma severity ($P < 0.001$) and exacerbation (Fig. 4a-b). In addition, it had a positive correlation with ACT and FEV₁/FVC ratio ($P = 0.003$ and <0.001 , respectively) (Fig. 4c-d). By analyzing bacteriophage by host, we found that *Streptococcus* phage was responsible for those phage-clinical factor correlations. The abundance of *Streptococcus* phage significantly decreased with disease severity ($P < 0.003$) and exacerbation (Fig. E3a-b). The proportion of *Streptococcus* phage was positively correlated with ACT score and FEV₁/FVC ($P = 0.005$ and < 0.001 , respectively) (Fig. E3c-d). The other phages including *Klebsiella* phage and *Escherichia* phage did not show any significant correlation with clinical features.

DISCUSSION

In this study, we demonstrated for the first time that the virus species and composition differed greatly depending on the presence or absence of asthma. In particular, the abundance of human beta and gamma herpesviruses was found to be associated with asthma symptoms.

Herpesviruses are DNA viruses that have infected humans with life-long latency. There are eight species of human herpesviruses (HHVs). In our study, we identified HHV-1 (HSV-1), an alpha-herpesvirus; HHV-5 (CMV) and HHV-6A/B, beta-herpesviruses; and HHV-4 (EBV), a gamma-herpesvirus. Herpesviruses, especially CMV and EBV, are commonly detected in the general population. EBV seropositive rate in the general population is known to be >90%.¹⁴ The global seroprevalence of CMV is estimated as 83% with a regional variation of 66-90%.¹⁵ We also found high proportion (58%) of both viruses in healthy individuals, though the virome-based positive rate was lower than serologic test-based rate due to the technical difference. We think the inconsistency of positive rates between previous reports and our report is because seropositive rate includes past infection history but virome analyses detect current presence of viral DNA which may reflect the reactivation of latent herpesviruses. Therefore, herpesviruses are considered ubiquitous latent viruses with relatively less importance than other pathogenic respiratory viruses. However, there are few reports on the association between herpesviruses and asthma. For example, only three of 63 articles on asthma exacerbation and viruses have reported HSV-1 and CMV as major viruses.¹⁶ This may be attributed to underestimation on the importance of herpesviruses present in the airway and mainstream studies primarily focused on well-known pathogenic viruses in the past.

As virome studies have been accumulating lately, many herpesvirus-related studies are also being conducted. Importantly, CMV DNA was found in the blood of adult patients with asthma, and its copy number was correlated with an increased risk of asthma.¹⁷ The CMV

DNA copy number was relatively higher in the bronchoalveolar lavage fluid (BALF) of 111 pediatric patients with asthma than in healthy children.¹⁸ Consistent with this, several studies have reported that a herpesvirus infection causes pathological changes in the lungs and leads to chronic respiratory disease.¹⁹⁻²¹ Here, we found that asthma severity and exacerbation increased as the abundance of human herpesviruses (particularly CMV and EBV) increased. In addition, CMV and EBV showed significant inverse correlations with ACT scores and FEV₁/FVC, suggesting that they may be related to the worsening of symptoms and decreased lung function. We hypothesize that reactivation of latent herpesviruses may be related to diverse pathological immune dysregulations or medication into the asthmatic lung, though the exact molecular mechanisms that leads to reactivation cannot be explained here.

The correlation between CMV and asthma severity appears to arise from the relationship between HHVs and the immune system. Several studies have demonstrated an association between herpesviruses and the immune system,²² including their roles in interferon- α production through toll-like receptor 9²³ and the regulation of cytokines associated with interleukin-10.²⁴ Additionally, several population-based studies have suggested that asthma may be a risk factor for herpes zoster, human alpha-herpesvirus 3 infection.^{25, 26} Although the relationship between herpesviruses and asthma may be associated with changes in the immune system, the causal relationship between herpesviruses and respiratory diseases has not yet been elucidated; therefore, further study on this topic is needed.

In this study, shotgun sequencing and 16s rRNA amplicon sequencing were performed on the same sample, and correlational analyses were carried out between virome and bacteriome data. Virome analysis showed a very high abundance of *Streptococcus* phages in the control group (70.0%) and markedly lower abundance in the asthma group, whereas the bacteriome analysis revealed no differences in *Streptococcus* abundance between the groups. Correlation

analysis also found no significant correlation between the abundance of *Streptococcus* phages in the virome and that of *Streptococcus* in the bacteriome. However, the bacteriophage abundance may appear to be relatively lower in patients with asthma than in healthy individuals because of the large increase in the abundance of eukaryotic viruses in asthma patients rather than because of an actual reduction in bacteriophage abundance. A previous report on the respiratory virome in chronic obstructive pulmonary disease (COPD) noted a reduced abundance of bacteriophages in nasopharyngeal swabs from patients with COPD,²⁷ similar to our results demonstrating a reduced abundance of bacteriophages in patients with asthma.

Viruses, rather than bacteria, constitute the largest proportion of the human microbiota. Therefore, in various diseases and health states, the virome has attracted considerable attention. However, virome research remains limited owing to technical challenges, including low biomass and a lack of universal molecular markers and standardized virus databases. To overcome the low biomass problem in respiratory samples, which have low viral loads, virus particles are usually collected using cesium chloride density gradient centrifugation. However, achieving widespread use of this method in the diagnostic environment is difficult because it requires high technical proficiency. To address this issue, we used a simple and convenient approach by applying commercially available sample concentration tubes (2- and 20-mL tubes) to collect the viral particles. The viral sequence proportion among the non-human sequences obtained through this approach was 2.79% (0.04 - 30.16% range depending on samples when NCBI Virus database was used for sequence identification; data not shown). This viral ratio is equivalent to the value observed in other sputum virome study which used cesium chloride ultracentrifugation (3.34%).¹⁰ Thus, the newly developed mini-concentrator approach may enable standardization and increase the reproducibility of experimental

procedures for virome analysis.

Airway virome research faces additional challenges because of the difficulties in sampling airway virome compared to that in sampling intestinal virome. Despite conducting sufficient oral washing before inducing sputum, contamination by the upper airways and saliva can occur in the process of collecting samples. Even BALFs, which are known to have a relatively low chance of contamination, are not free from contamination.²⁸ However, according to previous reports, the signal that consistently correlates with indices of lung disease is not entirely obscured despite the noise introduced by contamination from the upper airways.²⁹ Accordingly, we believe the asthma-virus correlation observed in this study is meaningful despite the risk of upper airway contamination.

The possible underestimation of RNA viruses is one limitation of this study. The proportion of eukaryotic viruses excluding bacteriophages was 38.0%, of which most were DNA viruses, whereas RNA viruses accounted for less than 1%. The low abundance of RNA viruses is likely due to the excessive removal of RNA viruses during the virus isolation and nucleic acid extraction processes, as reported previously.^{30,31} In particular, the use of RNase to remove free RNAs could have affected the RNA fraction of the virome.³² Considering that bacteriome studies often have large sample sizes, the low number of virome samples is a limitation also. To fully discover the viral dark matter in asthma, further studies with large sample sizes and enforced RNA viral sequencing are required. The other limitation of this study is that the contribution of inhaled corticosteroid on the abundance of commensal viruses could not be analyzed independently with disease severity due to the multicollinearity of these two factors. Further analysis such as longitudinal data analysis which measures virome before/after the medication is required to adjust the effect of inhaled corticosteroid.

411 Conclusions

412 Our findings uncovered significant association between virome and clinical variable of
413 asthma severity, which suggest that herpesviruses and bacteriophages from airway may have
414 potential future applications as a biomarker of asthma.

References

1. Beasley R, Semprini A, Mitchell EA. Risk factors for asthma: is prevention possible? *Lancet* 2015; 386:1075-85.
2. Busse WW, Lemanske RF, Jr., Gern JE. Role of viral respiratory infections in asthma and asthma exacerbations. *Lancet* 2010; 376:826-34.
3. Rantala A, Jaakkola JJ, Jaakkola MS. Respiratory infections precede adult-onset asthma. *PLoS One* 2011; 6:e27912.
4. Bloomfield SF, Stanwell-Smith R, Crevel RW, Pickup J. Too clean, or not too clean: the hygiene hypothesis and home hygiene. *Clin Exp Allergy* 2006; 36:402-25.
5. Wylie KM, Weinstock GM, Storch GA. Emerging view of the human virome. *Transl Res* 2012; 160:283-90.
6. Popgeorgiev N, Temmam S, Raoult D, Desnues C. Describing the silent human virome with an emphasis on giant viruses. *Intervirology* 2013; 56:395-412.
7. Zarate S, Taboada B, Yocupicio-Monroy M, Arias CF. Human virome. *Arch Med Res* 2017; 48:701-16.
8. Cadwell K. The virome in host health and disease. *Immunity* 2015; 42:805-13.
9. Krishnamurthy SR, Wang D. Origins and challenges of viral dark matter. *Virus Res* 2017; 239:136-42.
10. Willner D, Furlan M, Haynes M, Schmieder R, Angly FE, Silva J, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One* 2009; 4:e7370.
11. Romero-Espinoza JA, Moreno-Valencia Y, Coronel-Tellez RH, Castillejos-Lopez M, Hernandez A, Dominguez A, et al. Virome and bacteriome characterization of children with pneumonia and asthma in Mexico City during winter seasons 2014 and

2015. PLoS One 2018; 13:e0192878.
12. Reddel HK, Taylor DR, Bateman ED, Boulet LP, Boushey HA, Busse WW, et al. An official American Thoracic Society/European Respiratory Society statement: Asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice. *Am J Respir Crit Care Med* 2009; 180:59-99.
 13. RexSoft. Rex: Excel-based statistical analysis software. 2018. <http://rexsoft.org/>.
 14. Boppana SB, Fowler KB. Persistence in the population: epidemiology and transmission. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, et al., editors. *Human herpesviruses: biology, therapy, and immunoprophylaxis*. Cambridge; 2007.
 15. Zuhair M, Smit GSA, Wallis G, Jabbar F, Smith C, Devleesschauwer B, et al. Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis. *Rev Med Virol* 2019; 29:e2034.
 16. Zheng XY, Xu YJ, Guan WJ, Lin LF. Regional, age and respiratory-secretion-specific prevalence of respiratory viruses associated with asthma exacerbation: a literature review. *Arch Virol* 2018; 163:845-53.
 17. Kowalski ML, Wardzynska A, Studzinska M, Pawelczyk M, Lesnikowski ZJ, Paradowska E. Cytomegalovirus DNA is highly prevalent in the blood of patients with asthma and is associated with age and asthma traits. *Allergy* 2017; 72:2035-8.
 18. Sun HM, Li SX, Yan YD, Chen ZR, Wang YQ, Hao CL, et al. Associations between patient clinical characteristics and the presence of cytomegalovirus DNA in the bronchoalveolar lavage fluid of children with recurrent wheezing. *BMC Infect Dis* 2018; 18:458.
 19. Sawant A, Spoletini G, Whitaker P, Etherington C, Clifton I, Peckham D.

- Cytomegalovirus-associated pulmonary exacerbation in patients with cystic fibrosis. ERJ Open Res 2018; 4:00111-2017.
20. Nenna R, Zhai J, Packard SE, Spangenberg A, Sherrill DL, Martinez FD, et al. High cytomegalovirus serology and subsequent COPD-related mortality: a longitudinal study. ERJ Open Res 2020; 6:00062-2020.
21. Reuter S, Lemmermann NAW, Maxeiner J, Podlech J, Beckert H, Freitag K, et al. Coincident airway exposure to low-potency allergen and cytomegalovirus sensitizes for allergic airway disease by viral activation of migratory dendritic cells. PLoS Pathog 2019; 15:e1007595.
22. Sehrawat S, Kumar D, Rouse BT. Herpesviruses: harmonious pathogens but relevant cofactors in other diseases? Front Cell Infect Microbiol 2018; 8:177.
23. Hochrein H, Schlatter B, O'Keeffe M, Wagner C, Schmitz F, Schiemann M, et al. Herpes simplex virus type-1 induces IFN-alpha production via Toll-like receptor 9-dependent and -independent pathways. P Natl Acad Sci USA 2004; 101:11416-21.
24. Dreyfus DH. Herpesviruses and the microbiome. J Allergy Clin Immunol 2013; 132:1278-86.
25. Kwon HJ, Bang DW, Kim EN, Wi CI, Yawn BP, Wollan PC, et al. Asthma as a risk factor for zoster in adults: A population-based case-control study. J Allergy Clin Immunol 2016; 137:1406-12.
26. Esteban-Vasallo MD, Dominguez-Berjon MF, Gil-Prieto R, Astray-Mochales J, de Miguel AG. Sociodemographic characteristics and chronic medical conditions as risk factors for herpes zoster: A population-based study from primary care in Madrid (Spain). Hum Vacc Immunother 2014; 10:1650-60.
27. van Rijn AL, van Boheemen S, Sidorov I, Carbo EC, Pappas N, Mei H, et al. The

respiratory virome and exacerbations in patients with chronic obstructive pulmonary disease. PLoS One 2019; 14:e0223952.

28. Schneeberger PHH, Prescod J, Levy L, Hwang D, Martinu T, Coburn B. Microbiota analysis optimization for human bronchoalveolar lavage fluid. Microbiome 2019; 7:141.
29. Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. Lancet 2014; 384:691-702.
30. Garmaeva S, Sinha T, Kurilshikov A, Fu J, Wijmenga C, Zhernakova A. Studying the gut virome in the metagenomic era: challenges and perspectives. BMC Biol 2019; 17:84.
31. Virgin HW. The virome in mammalian physiology and disease. Cell 2014; 157:142-50.
32. Adriaenssens EM, Farkas K, Harrison C, Jones DL, Allison HE, McCarthy AJ. Viromic analysis of wastewater input to a river catchment reveals a diverse assemblage of RNA viruses. mSystems 2018; 3:e00025-18.

501 **Table 1. Demographic characteristics of patients with asthma enrolled in this study**

	Control (n = 12)	Non-severe asthma (n = 15)	Severe asthma (n = 15)	P value*
Sex, female (%)	75.0%	73.3%	46.7%	0.218
Age (years)	46.8 ± 5.4	60.2 ± 7.1	56.6 ± 6.1	0.001
BMI (kg/m ²)	22.8 ± 2.8	24.3 ± 2.7	25.2 ± 2.9	0.107
Baseline FVC (L)	3.3 ± 0.7	3.1 ± 0.6	2.7 ± 0.9	0.001
Baseline FVC (%)	101.8	100.3	78.9	0.001
Baseline FEV1 (L)	2.8 ± 0.6	2.3 ± 0.4	1.7 ± 0.7	0.001
Baseline FEV1 (%)	101.6	89.5	63.2	0.012
FEV1/FVC (%)	86.1	75.0	64.0	0.001
Exacerbation/year (yes, %)	-	6.7	60	0.021
Allergic rhinitis (%)	-	53.3	73.3	0.006
ACT score	-	22.6	17.1	0.001
ACQ score	-	3.9	14.5	0.001
Daily ICS [†]	-	600.0 ± 207.0	964.3 ± 129.1	0.001
Periostin (pg/mL)	205.0 ± 371.0	119.6 ± 253.3	249.0 ± 406.6	0.177
25(OH)Vitamin D (ng/mL)	21.5 ± 11.7	23.6 ± 10.7	23.1 ± 9.9	0.801
Sputum eosinophilia (%)	-	71.4%	81.8%	0.424

Values are n (%) or mean ± standard deviation.

BMI, body mass index; FVC, forced vital capacity; FEV1, forced expiratory volume; ACT, Asthma Control Test; ACQ, Asthma Control Questionnaire; ICS, inhaled corticosteroid eosinophilic, blood eosinophil count > 300 or sputum eosinophil percentage > 3%.

*Fisher exact test was used for categorical variables; Kruskal-Wallis tests with Tukey's post-hoc

analysis was used for continuous variables.

†Fluticasone equivalent

Table 2. Statistics of virome sequencing data in the different experimental groups.

	Control	Non-severe	Severe asthma
Number of reads	(average; n = 12)	(average; n = 15)	(average; n = 15)
Raw reads	40,329,874	40,658,010	40,101,714
Quality-filtered reads	34,115,597	34,035,614	32,831,945
Pair-end reads	14,182,197	14,323,577	13,269,595
Length-filtered reads	13,794,395	13,626,956	12,742,862
Human reads	13,302,901	13,263,632	12,353,993
Non-human reads	491,494	363,324	388,869
Identified viral reads [*]	6,897	9,675	8,283
Verified viral reads [#]	524	808	473

^{*} Non-human reads identified as viral sequences as a result of BLAST search.

[#] Identified viral reads further verified as true viral sequences based on sequence alignment coverage ($\geq 50\%$).

Figure Legends

Figure 1. Experimental procedure to collect and concentrate virus-like particles from the induced sputum. DTT, dithiothreitol; PBS, phosphate-buffered saline; VLP, virus-like particle.

Figure 2. Virome and bacteriome in the sputum of healthy control, non-severe asthma, and severe asthma groups. Principal coordinates analysis (PCoA) of the beta-diversity of the virome (a) and bacteriome (b) showing the variation of the microbiome structure depending on the experimental groups. The geometric partitioning of variation across a multivariate data cloud was evaluated using permutational multivariate analysis of variance (PERMANOVA) and the p-value is indicated. The ellipses represent the 95% confidence interval for each group. The average compositions of the virome (c) and bacteriome (d) are shown for each experimental group. HSV-1, herpes simplex virus 1; CMV, cytomegalovirus; EBV, Epstein–Barr virus.

Figure 3. Virome composition in individual samples and virus detection rate depending on disease groups. (a) The virus composition in individual samples. (b) The detection rate indicates the proportion of virus-positive samples out of all the samples in the same condition. The p-value indicate chi-square for linear trend. HSV-1, herpes simplex virus 1; CMV, cytomegalovirus; EBV, Epstein–Barr virus.

Figure 4. Associations between the virome and clinical data. (a) Abundance of major viruses depending on severity of asthma (n = 42). (b) Exacerbation positive rate depending on the high ($\geq 50^{\text{th}}$ percentile) and low ($< 50^{\text{th}}$ percentile) virus groups in asthma patients (n = 30).

531 Exacerbation means that the patient experienced asthma exacerbation one or more time
532 during the last one year. Correlation between herpesviruses and (c) ACT scores in asthma
533 patients (n = 30) or (d) FEV₁/FVC ratios in all participants (n = 42). Values are n (%) or mean
534 $\pm$ standard error. HSV-1, herpes simplex virus 1; CMV, cytomegalovirus; EBV, Epstein–Barr
535 virus.



Lung virome in asthma reflecting asthma severity and exacerbation

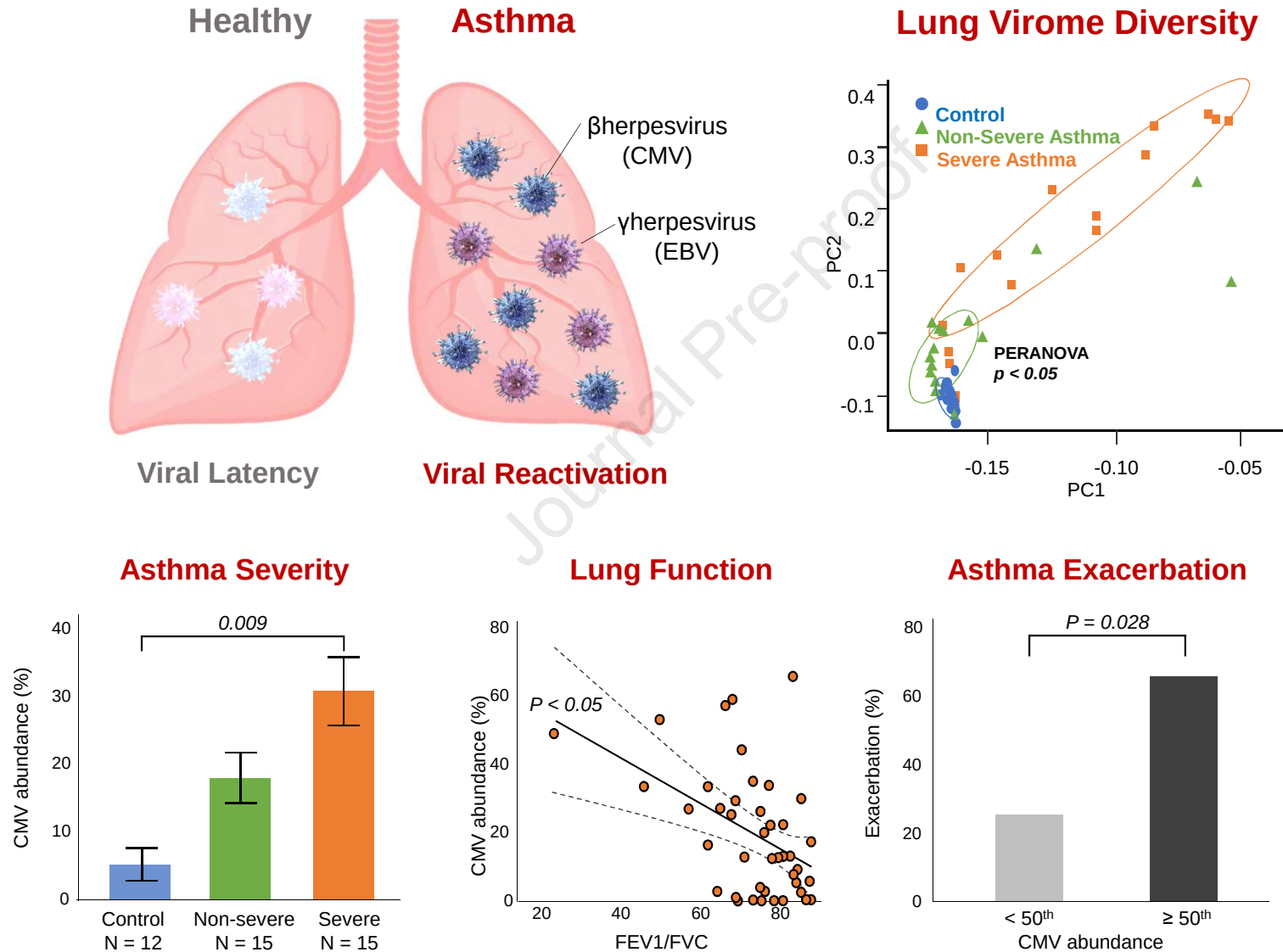


Fig 1.

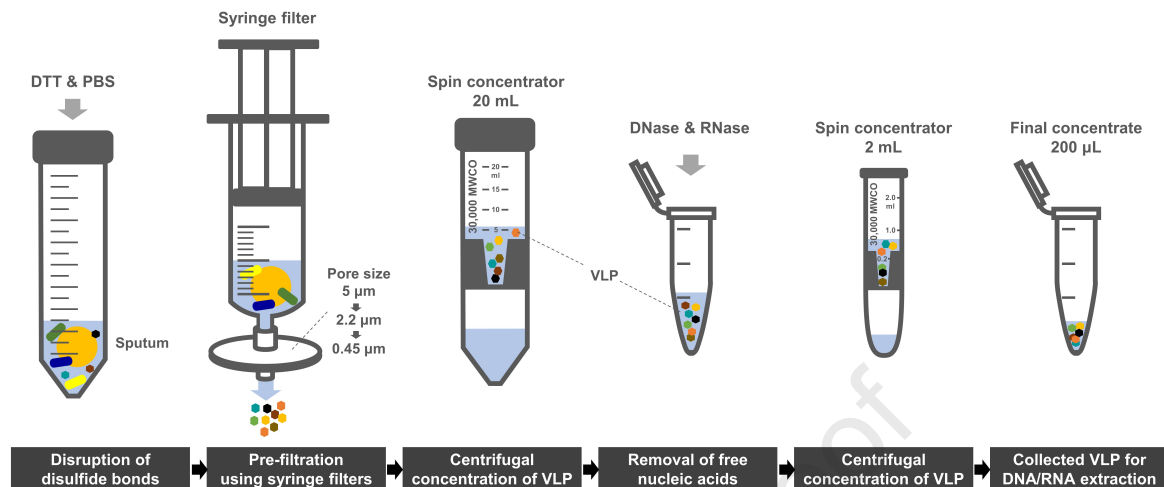


Fig 2.

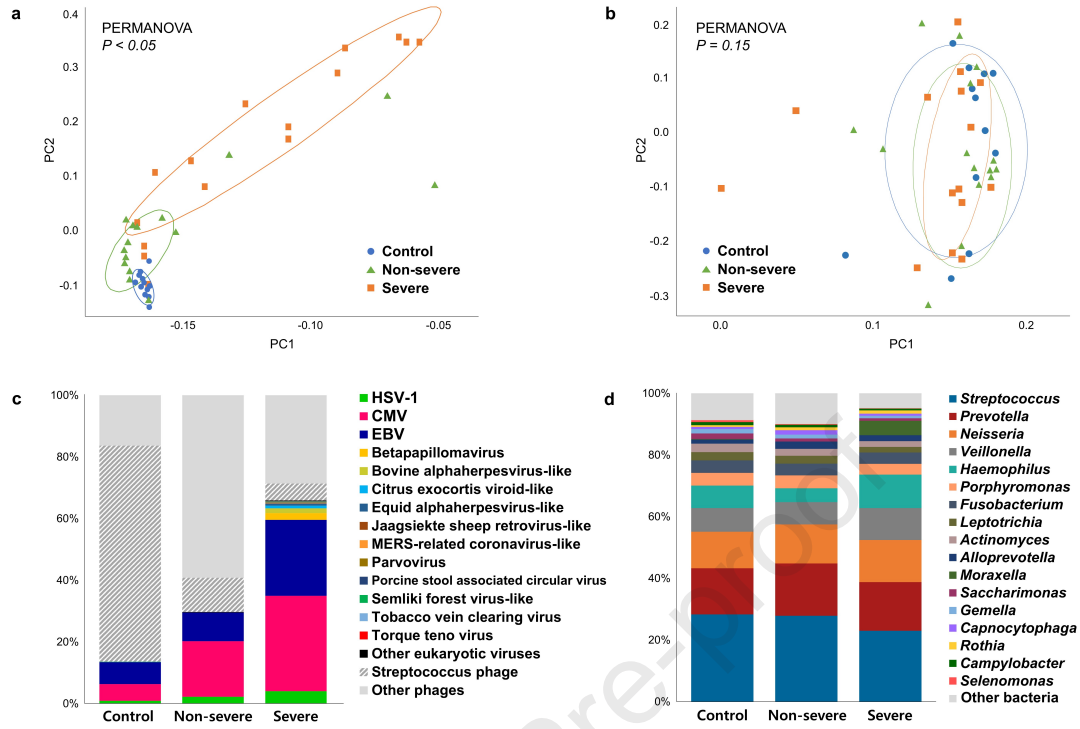


Fig 3.

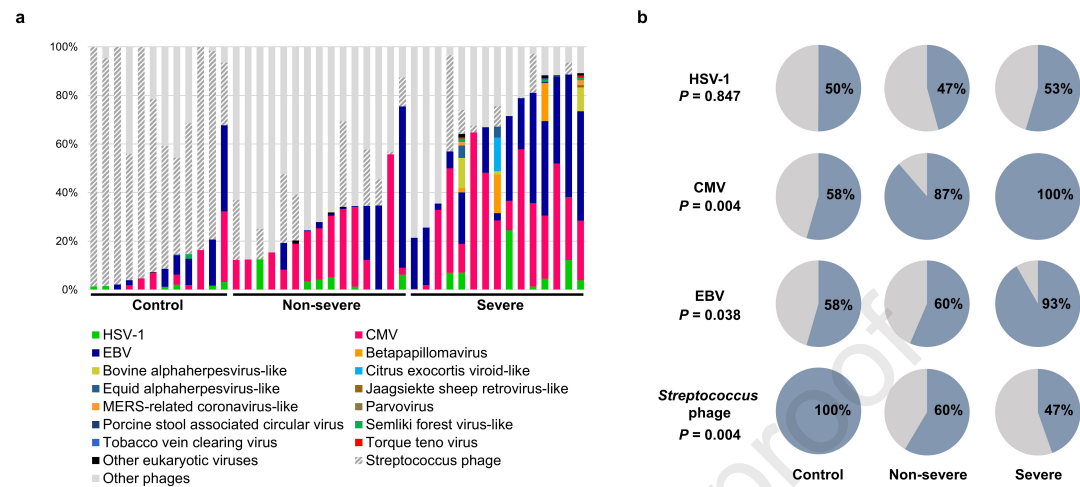
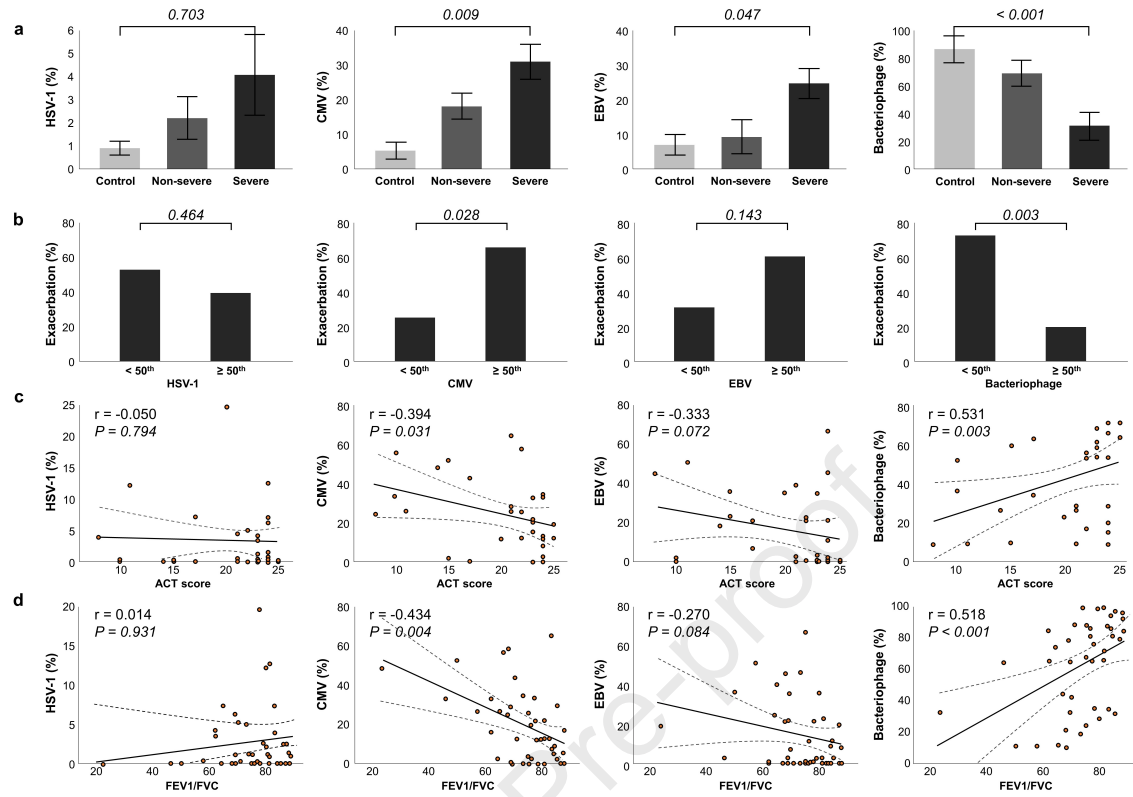


Fig 4.



Online Repository Materials

Lung virome: new potential biomarkers for asthma severity and exacerbation

Sungmi Choi, PhD^{1,2}, Kyoung-Hee Sohn, MD, PhD^{3,4}, Jae-Woo Jung, MD, PhD⁵, Min-Gyu Kang, MD, PhD⁶, Min-Suk Yang, MD, PhD⁷, Sujeong Kim, MD⁸, Jeong-Hee Choi, MD, PhD^{9,10}, Sang-Heon Cho, MD, PhD^{3,11}, Hye-Ryun Kang, MD, PhD^{3,11‡}, and Hana Yi, PhD^{1,2,12‡}

¹Institute for Biomaterials, Korea University, Seoul, Korea

²Interdisciplinary Program in Precision Public Health, Korea University, Seoul, Korea

³Institute of Allergy and Clinical Immunology, Seoul National University Medical Research Center, Seoul National University College of Medicine, Seoul, Korea

⁴Department of Internal Medicine, Kyung Hee University Hospital, Seoul, Korea

⁵Department of Internal Medicine, Chung-Ang University College of Medicine, Seoul, Korea

⁶Department of Internal Medicine, Chungbuk National University College of Medicine, Chungbuk National University Hospital, Cheongju, Korea

⁷Department of Internal Medicine, SMG-SNU Boramae Medical Center, Seoul, Korea

⁸Department of Internal Medicine, School of Medicine, Kyungpook National University, Kyungpook National University Hospital, Daegu, Korea

⁹Department of Pulmonology and Allergy, Hallym University Dongtan Sacred Heart Hospital, Hwaseong, Korea

¹⁰Allergy and Clinical Immunology Research Center, Hallym University College of Medicine,
Chuncheon, Korea

¹¹Department of Internal Medicine, Seoul National University Hospital, Seoul, Korea

¹²School of Biosystems and Biomedical Sciences, Korea University, Seoul, Korea

[‡]Correspondence:

Hana Yi, PhD

School of Biosystems and Biomedical Sciences, Korea University

145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea

Tel. 82-2-3290-5644 Fax: 82-2-940-2849

E-mail: hanayi@korea.ac.kr

Hye-Ryun Kang, MD, PhD

Seoul National University Hospital

101 Daehak-ro, Jongno-Gu, Seoul 110-744, Korea

Tel. 82-2-2072-0820 Fax: 82-2-742-3291

E-mail: helenmed@snu.ac.kr

Repository text 1.*Collection of virus-like particles from induced sputum*

Sputum samples contain an abundance of protein and exhibit high viscosity; thus, pretreatment was required before nucleic acid extraction. A mixture of 5 mL induced sputum, 5 mL phosphate-buffered saline, and 100 μ L dithiothreitol (Sigma-Aldrich, St. Louis, MO, USA) was shaken at 200 rpm and 25 °C for 1 h using a shaking incubator to break the disulfide bonds of mucus components. To filter out human cells and microorganisms such as bacteria and fungi derived from the human body while retaining the viruses, polyethersulfone syringe filters (Sartorius Stedim Biotech, Göttingen, Germany), with pore sizes of 5, 2.2, and 0.45 μ m, were sequentially applied. The prefiltered samples were concentrated using a 20-mL Vivaspin 20 ultrafiltration device (Sartorius Stedim Biotech) equipped with a 30,000-kDa pore-sized polyethersulfone membrane. The centrifugation time at 10,000 g varied between 1–2 h depending on the sample. Samples of approximately 6 mL were concentrated about 12-fold to a final volume of less than 500 μ L. To remove free nucleic acids, 2 μ L DNase I (Qiagen, Valencia, CA, USA) was added to the concentrated sample and incubated at 25 °C for 30 min. Next, 2 μ L RNase A (Qiagen) was added and incubated at 37 °C for 30 min. Subsequently, the reaction was stopped by denaturing the degrading enzyme at 80 °C for 15 min. The resulting sample was further concentrated to 200 μ L using a 2-mL Vivaspin 2 centrifugal concentrator (Sartorius Stedim Biotech) at 10,000 rpm for 30 min. The volume of the final concentrate was 200 μ L or less.

Reverse transcription and amplification

Both DNA and RNA were extracted from the concentrated samples using a QIAamp MinElute Virus Spin Kit (Qiagen). First-strand cDNA was synthesized using SuperScript III First-Strand Synthesis (Invitrogen Inc., Carlsbad, CA, USA) and V-N random primer (5'-GTTTCCCAGTCACGATCNNNNNN-3'). For the 20- μ L reverse transcription reaction, 5 μ L nucleic acid eluent was used. Reverse transcription was performed for 50 min at 50 °C, and the enzyme was inactivated for 5 min at 85 °C. Second-strand cDNA was synthesized using Large Klenow Fragment (Invitrogen Inc.) and V-N random primer at 37°C for 1 h. Following enzyme inactivation, reaction tubes were incubated at 65 °C for 5 min. To obtain sufficient DNA for sequencing, PCR was performed using the synthesized cDNA as the template. V primer (5'-GTTTCCCAGTCACGATC-3') was used, and the PCR cycling conditions were as follows: 5 min at 95 °C; 30 cycles of 30 s at 95 °, 45 s at 55 °, and 30 s at 94 °; 3 min at 72 °; and holding at 4 °C. The constructed virome amplicon was sent to a commercial sequencing service provider (Chunlab, Seoul, Korea) and sequenced as 150 bp paired-end reads using a HiSeq3000 instrument (Illumina, San Diego, CA).

Virome sequence analysis

The poor-quality sequencing reads from both ends were removed from the raw sequencing data using Trimmomatic.¹ Subsequently, paired-end reads were merged and short sequences (<100 nt) were filtered using PEAR. Sequences derived from the host genome sequences were removed using the Bowtie2 program.² Finally, the non-humans sequences were identified using BLASTn (e-value cutoff 1E-10) against our custom-made database to find the most similar sequences. The NCBI Viral Genomes database (<https://www.ncbi.nlm.nih.gov/genome/viruses/>) and RefSeq database

(<https://www.ncbi.nlm.nih.gov/refseq/>) were downloaded (accessed March 2019) and combined to construct the custom-made database. Among the sequences downloaded from RefSeq, only sequences satisfying the following two conditions were extracted and used for database construction; 1) sequences containing the keyword 'virus' or 'phage' in the sequence title were selected and 2) sequences originated from 'uncultured' or 'metagenome' were excluded. The identified viral sequences were further examined based on sequence alignment length coverage. Among the identified viral sequences, only sequences with $\geq 50\%$ of alignment length coverage were selected as verified viral sequences and presented in the manuscript. The taxonomic assignment for sequencing reads was determined by the taxonomy ID of the top BLAST results. If the alignment coverage was lower than 90%, the ambiguous identification were marked by adding '-like' at the end of the taxon name (eg. MERS-related coronavirus-like).

Verification of herpesvirus identification

To confirm the accuracy of viral reads identified as herpesviruses, sequencing reads identified as CMV or EBV were extracted and mapped to reference genomes using the CLC workbench (Qiagen). The identified viral reads from an individual sample were extracted and mapped to both reference genomes of CMV (NC006273) and EBV (NC007605). For example, the viral reads from sample 16M-CAU001 were mapped to both CMV and EBV reference genomes. As a result, the reads identified as CMV were mapped only to CMV genome but not to EBV genome showing accuracy of identification (Fig. E1). In the same manner, the reads identified as EBV were exclusively mapped to EBV genome.

Bacteriome (16S rRNA gene amplicon) sequencing and analysis

To decrease viscosity, 20 μ L beta-mercaptoethanol was added to 1 mL induced sputum, and the sample was shaken at 200 rpm for 0.5–1 h. The pretreated sample was subjected to DNA extraction using a FastDNA SPIN Kit for soil DNA extraction (MP Biomedicals, Santa Ana, CA, USA). For 16S rRNA gene amplification, 2 $\times$ KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Resnova, Ancona, Italy) and primers (318F: 5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3' and 806R: 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3') were used. The PCR cycling conditions were as follows: 3 min at 95 °C; 25 cycles of 30 s at 95 °C, 30 s at 55 °C, and 30 s at 72 °C; 5 min at 72 °C; and holding at 4 °C. The PCR products were purified using AMPure XP beads (Beckman Coulter, High Wycombe, UK). The amplicon sequencing library was constructed according to the 16S metagenomic sequencing library preparation methods (Illumina),³ and 300-bp paired-end sequencing was conducted using an Illumina MiSeq v3 platform (Illumina).

The raw sequencing data files were preprocessed using bioinformatics tools before downstream data analysis. To remove low-quality sequences, the sequencing raw reads were prefiltered using PRINSEQ.⁴ Adapter sequences were removed using Cutadapt,⁵ and ambiguous bases were further trimmed. Paired-end reads were merged using PEAR⁶ and then further filtered using PRINSEQ. Chimeric sequences and singletons were screened and reduced using USEARCH⁷. To equalize the sequencing depth throughout the samples, 20,000 reads were randomly sampled. Downstream data analysis was carried out using QIIME⁸ with the Ezbiocloud 16S rRNA gene sequence database.⁹ Operational taxonomic units were defined as clusters of sequences with 97% or greater identity. Alpha- and beta-diversity indices were calculated using the core_diversity_analyses option. The weighted normalized UniFrac distance¹⁰ was calculated

using QIIME and plotted using principal coordinate analysis (PCoA) using the R version 3.5.1. The significance of grouping in the ordination was evaluated using permutational multivariate analysis of variance (PERMANOVA) using R program. Kruskal-Wallis tests and linear discriminant analysis effect size (LEfSe)¹¹ were performed on the three experimental groups to identify particular bacterial taxa that were more abundant in the asthma groups than in the control group.

Table E1. Statistics of virome sequencing data for individual samples.

	Sample	Raw reads	Quality filtered	Paired-end	Length filtered	Human	Non-human	Identified viral reads	Verified viral reads
Control	16M_SNU014	39,819,534	34,030,610	14,039,369	12,859,275	12,837,007	22,268	1,850	144
	18M_SNU008	44,371,644	38,130,202	15,902,534	14,973,490	14,942,839	30,651	1,511	131
	16M_SNU018	33,810,330	27,903,796	8,894,356	13,785,119	10,714,488	3,070,631	7,595	140
	18M_SNU004	42,260,022	35,970,550	14,907,826	13,915,823	13,835,083	80,740	10,235	2,823
	18M_SNU005	38,985,146	33,457,122	14,492,292	13,697,157	13,665,528	31,629	13,201	2,076
	18M_SNU009	40,287,950	33,457,722	13,768,316	12,951,331	12,883,771	67,560	3,410	410
	18M_SNU003	44,166,114	37,652,678	16,307,434	15,424,907	15,288,905	136,002	6,842	82
	18M_SNU006	41,246,774	34,779,390	15,070,707	14,259,314	14,197,986	61,328	2,355	338
	18M_SNU001	37,584,890	31,847,584	14,222,213	13,624,298	12,625,705	998,593	7,812	102
	18M_SNU007	38,707,282	32,641,728	12,875,740	12,110,418	11,051,472	1,058,946	21,531	49
	16M_SNU026	42,007,568	34,892,118	14,461,532	13,363,937	13,325,849	38,088	3,213	58
	18M_SNU002	40,711,234	34,623,660	15,244,045	14,567,669	14,266,176	301,493	3,211	31
	Average	40,329,874	34,115,597	14,182,197	13,794,395	13,302,901	491,494	6,897	532
Non-severe asthma	16M_SNU044	39,358,588	33,473,518	13,443,913	13,707,267	13,110,724	596,543	452	73
	18M_SNU011	35,778,215	30,523,412	14,326,862	13,326,833	13,037,162	289,671	14,231	169
	18M_SNU013	46,844,758	37,845,028	17,121,002	13,778,891	13,587,212	191,679	10,293	167
	16M_CAU011	41,364,812	34,479,066	13,609,143	13,570,239	13,506,689	63,550	2,011	52
	18M_SNU020	41,965,970	38,790,194	17,679,986	17,480,339	17,248,304	232,035	13,228	338
	18M_SNU019	34,392,844	29,405,991	11,593,046	11,458,320	11,293,468	164,852	1,233	73
	16M_CAU009	38,409,766	32,911,468	12,145,683	12,078,373	9,482,598	2,595,775	30,341	5,343
	16M_CAU012	39,861,756	34,096,000	14,081,076	13,129,092	13,036,742	92,350	17,036	893
	16M_CAU005	38,995,884	21,640,968	9,145,654	8,754,840	8,562,912	191,928	1,899	274
	16M_SNU063	37,231,238	36,513,561	15,901,163	15,321,120	15,029,310	291,810	7,322	832
	16M_CAU008	44,499,362	37,466,308	15,476,105	15,408,942	15,227,169	181,773	8,932	235
	18M_SNU014	45,559,082	33,417,314	14,851,706	13,158,033	13,012,397	145,636	32,881	2,773
	18M_SNU015	45,408,538	39,207,128	18,806,294	17,799,588	17,740,222	59,366	1,843	215
	16M_SNU001	39,905,494	33,614,906	14,225,330	13,408,964	13,323,283	85,681	2,503	269
	18M_SNU021	40,293,843	37,149,342	12,446,692	12,023,503	11,756,293	267,210	922	420
	Average	40,658,010	34,035,614	14,323,577	13,626,956	13,263,632	363,324	9,675	808
Severe asthma	16M_SNU040	33,491,992	24,940,301	9,320,196	9,102,341	9,100,020	2,321	809	626
	18M_CBU003	38,576,268	27,233,522	12,087,731	11,627,095	11,610,392	16,703	4,930	258
	18M_SNU016	40,141,644	35,682,630	9,677,098	9,669,500	9,657,098	12,402	754	76
	18M_KBU002	41,854,960	35,477,674	13,469,794	12,698,590	11,698,400	1,000,190	24,320	231
	18M_BSN002	42,845,824	31,149,636	17,807,871	17,117,709	17,002,912	114,797	3,021	212
	16M_CBU002	44,566,153	39,216,730	14,165,342	13,847,372	12,732,966	1,114,406	6,183	852
	16M_HYU001	39,994,356	33,652,390	15,900,374	14,948,372	14,809,495	138,877	432	139
	16M_CAU001	41,797,344	34,319,568	15,731,682	12,717,747	11,948,430	769,317	16,912	1,432
	16M_HYU002	45,876,206	28,302,168	13,319,800	13,305,019	13,083,720	221,299	20,341	1,380
	16M_CAU021	41,797,344	36,412,878	12,222,058	12,111,553	10,113,620	1,997,933	11,754	1,231
	18M_HRU006	40,131,530	30,317,204	13,612,715	13,460,491	13,210,293	250,198	25,932	143
	18M_BSN001	31,523,566	42,304,112	14,127,967	14,059,293	14,004,127	55,166	3,021	154
	16M_SNU066	41,591,582	35,465,774	11,601,220	11,174,815	11,112,301	62,514	1,309	129
	18M_CAU003	44,426,550	31,411,572	14,895,556	14,210,228	14,203,002	7,226	504	123
	16M_SNU041	32,910,391	26,593,012	11,104,521	11,092,811	11,023,118	69,693	4,021	102
	Average	40,101,714	32,831,945	13,269,595	12,742,862	12,353,993	388,869	8,283	473

Table E2. Eukaryotic virus composition in individual samples.

	Sample	HSV-1	CMV	EBV	Betapapillomavirus	Bovine alphaherpesvirus-like	Cirrus exocortis viroid- like	Equid alphaherpesvirus- like	Jaagsiekte sheep retrovirus-like	MERS-related coronavirus-like	Parvovirus	Porcine stool associated circular virus	Semliki forest virus-like	Tobacco vein clearing virus	Torque teno virus	Other eukaryotic viruses
Control	16M_SNU014	1.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU008	1.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16M_SNU018	0.00	0.00	2.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU004	0.00	1.84	2.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00
	18M_SNU005	0.00	4.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU009	0.00	7.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24
	18M_SNU003	1.23	0.00	7.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU006	2.07	4.14	7.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.00	0.00
	18M_SNU001	0.00	1.96	10.78	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.96	0.00	0.00	0.00
	18M_SNU007	0.00	16.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16M_SNU026	1.72	0.00	18.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU002	3.23	29.03	35.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Average	0.93	5.42	7.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.16	0.02	0.00	0.02
Non-severe asthma	16M_SNU044	0.00	12.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU011	0.00	12.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU013	12.57	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16M_CAU011	0.00	15.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU020	0.00	8.28	10.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU019	0.00	18.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.35
	16M_CAU009	3.44	20.68	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16M_CAU012	4.14	21.16	2.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16M_CAU005	5.07	25.36	0.72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.72
	16M_SNU063	0.24	32.97	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.60
	16M_CAU008	1.28	32.77	0.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU014	0.00	12.23	22.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU015	0.00	0.00	34.74	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16M_SNU001	0.00	55.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU021	6.19	2.86	66.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Average	2.20	18.08	9.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18
Severe asthma	16M_SNU040	0.00	0.32	21.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_CBU003	0.00	1.94	23.64	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_SNU016	0.00	32.89	2.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_KBU002	6.96	43.04	6.96	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_BSN002	7.08	11.79	21.23	1.89	12.26	0.00	5.19	0.00	1.42	0.00	0.00	1.42	0.00	0.47	1.42
	16M_CBU002	0.00	64.79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16M_HYU001	0.00	48.20	18.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16M_CAU001	0.21	28.35	3.00	15.92	1.33	13.83	4.54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16M_HYU002	24.49	12.10	34.93	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16M_CAU021	0.00	57.84	21.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_HRU006	1.40	34.27	45.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	18M_BSN001	4.55	25.97	38.96	14.94	0.00	0.00	0.00	0.00	0.65	0.00	0.65	0.65	0.65	0.00	1.30
	16M_SNU066	0.00	51.94	35.66	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.78	0.00	0.00	0.00	0.00
	18M_CAU003	12.20	26.02	50.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	16M_SNU041	3.92	24.51	45.10	0.00	9.80	0.00	0.00	0.98	1.96	0.00	0.00	0.98	0.00	0.98	0.98
	Average	4.05	30.93	24.59	2.18	1.56	0.92	0.65	0.07	0.27	0.00	0.09	0.20	0.04	0.10	0.25

Table E3. Phage composition in individual samples.

	Sample	Streptococcus phage	Klebsiella phage	Escherichia phage	Acinetobacter phage	Actinomyces phage	Aggregatibacter phage	Bacillus phage	Burkholderia phage	Enterobacteria phage	Gordonia phage	Haemophilus phage	Lactobacillus phage	Propionibacterium phage	Pseudomonas phage	Other phages	
Control	16M_SNU014	98.61	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	18M_SNU008	93.89	0.00	0.00	0.00	1.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.05	
	16M_SNU018	97.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	18M_SNU004	52.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.00	0.00	0.00	0.00	39.99	
	18M_SNU005	95.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	18M_SNU009	71.29	0.00	0.00	0.00	7.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	14.36	
	18M_SNU003	50.62	0.00	0.00	4.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	35.80	
	18M_SNU006	39.94	0.00	0.00	0.00	45.56	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	18M_SNU001	53.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	31.37	
	18M_SNU007	83.67	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	16M_SNU026	77.59	0.00	0.00	0.00	1.72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	18M_SNU002	25.81	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.45	0.00	0.00	0.00	0.00	
	Average	70.04	0.00	0.00	0.41	4.66	0.00	0.00	0.00	0.00	0.00	0.87	0.00	0.00	0.00	10.38	
Non-severe asthma	16M_SNU044	24.66	45.21	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.74	0.00	6.85	8.22	
	18M_SNU011	0.00	37.87	49.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	18M_SNU013	12.57	54.49	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.60	10.78	0.60	8.38	
	16M_CAU011	0.00	84.62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	18M_SNU020	28.40	1.48	48.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.66	
	18M_SNU019	18.92	21.62	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	22.97	0.00	16.22	
	16M_CAU009	0.00	75.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	
	16M_CAU012	0.00	72.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	16M_CAU005	0.00	28.26	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.36	0.00	0.00	0.00	39.49	
	16M_SNU063	35.24	24.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.39	0.00	0.00	3.94	
	16M_CAU008	0.00	65.53	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	18M_SNU014	23.23	30.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.69	0.69	0.00	10.57	
	18M_SNU015	10.80	45.54	0.00	0.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.47	0.00	0.00	7.98	
	16M_SNU001	1.12	20.45	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.74	0.00	13.38	8.55	
	18M_SNU021	11.90	12.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.00	0.00	0.00	
		Average	11.12	41.30	6.53	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.52	2.30	1.39	7.07
	Severe asthma	16M_SNU040	0.00	0.00	78.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16
18M_CBU003		0.00	0.00	74.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
18M_SNU016		0.00	0.00	64.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
18M_KBU002		39.57	0.00	0.00	0.00	3.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
18M_BSN002		9.91	0.00	15.57	0.47	3.30	0.47	0.00	0.00	0.47	0.00	0.00	0.00	0.00	0.00	5.66	
16M_CBU002		2.82	0.00	32.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
16M_HYU001		0.00	0.00	24.46	0.00	0.00	0.00	0.00	8.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
16M_CAU001		8.52	0.00	21.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.86	
16M_HYU002		0.00	0.00	28.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
16M_CAU021		0.24	0.00	20.88	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
18M_HRU006		16.08	0.00	0.00	0.00	1.40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.40	
18M_BSN001		0.00	0.00	7.79	0.00	1.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.60	
16M_SNU066		0.00	0.00	11.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
18M_CAU003		4.88	0.00	0.00	0.00	2.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.07	
16M_SNU041		0.00	0.00	0.00	0.98	3.92	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5.88	
		Average	5.47	0.00	25.33	0.10	1.06	0.03	0.00	0.58	0.03	0.00	0.00	0.00	0.00	0.00	1.51

Table E4. Summary of regression analysis of relative abundance of viruses against the clinical variables. P-values were calculated from regression of log-transformed relative abundance on the experimental groups.

	HSV-1	CMV	EBV	Bacteriophage	<i>Streptococcus</i> phage
Group	0.7034	0.0091	0.0472	0.0001	0.0029
Age	0.5514	0.9300	0.1105	0.2418	0.7902
Oral corticosteroid	0.6344	0.8628	0.5331	0.6802	0.3258
Antibiotics	0.9597	0.6789	0.2273	0.9102	0.9279

Figure E1. Genome area coverage plot of CMV and EBV viral reads detected in individual sputum samples. (a) Sequencing reads identified as CMV in the 16M-CAU001 sample were mapped against the reference genome (NC_006273). (b) Sequencing reads identified as EBV in the 16M-CAU003 sample were mapped against the reference genome (NC_007605). The y-axis represents the number of reads at each nucleotide position. Blue features represent forward reads; dark blue areas represent reverse reads.

The mapped reads from sample 16M-CAU001 covered 99.2% of the total genome length of CMV with 20-fold coverage, whereas those of 16M-CAU003 covered 98.1% of EBV with 8-fold coverage. These results indicate that, despite the short virome read lengths, the reads were accurately identified as herpesvirus sequences. In addition, this demonstrated that herpesviruses were present in large amounts in the respiratory samples, and the whole viral genome could be recovered via viral metagenome sequencing.

Figure E2. Associations between the CMV abundance and the EBV abundance in samples.

Correlation analysis was performed to analyze the associations between the abundance of CMV and the abundance of EBV in individual samples (n=42).

Figure E3. Associations between the *Streptococcus* phage and clinical data. (a) Abundance of *Streptococcus* phage depending on severity of asthma (n = 42). (b) Exacerbation positive rate depending on the high ($\geq 50^{\text{th}}$ percentile) and low ($< 50^{\text{th}}$ percentile) *Streptococcus* phage groups in asthma patients (n = 30). Exacerbation (+) means the patient experienced exacerbation one or more time during the last one year. (c) Correlation between *Streptococcus* phage and (c) ACT scores in asthma patients (n = 30) or (d) FEV₁/FVC ratios in all participants (n = 42). Values are n (%) or mean $\pm$ standard error.

References

1. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; 30:2114-2120.
2. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012; 9:357-359.
3. 16S metagenomic sequencing library preparation. Available from https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf.
4. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 2011; 27:863-864.
5. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:3
6. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 2014; 30:614-620.
7. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010; 26:2460-2461.
8. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010; 7:335-3356.
9. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 2017; 67:1613-1617.
10. Hamady M, Lozupone C, Knight R. Fast UniFrac: facilitating high-throughput phylogenetic analyses of microbial communities including analysis of pyrosequencing and PhyloChip data. *ISME J* 2010; 4:17-27.
11. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol* 2011; 12:R60.